We sincerely thank Reviewer #2 for the constructive and thoughtful suggestions. We have provided point-by-point responses to the comments in blue and improved our manuscript accordingly. In the revised manuscript with track changes, the changes are highlighted in red.

The major revisions are summarized below:

1) We have clarified the idealized assumptions and limitations of the OSSEs, and moderated claims regarding inversion performance and sectoral attribution throughout the manuscript. The revised discussion states that the current OSSE cannot test sectoral separation, as the synthetic truth preserves the prior sectoral composition, or quantify the errors of the real-data estimates.

2) We have added covariance-based sectoral averaging-kernel estimates and expanded the discussion of prior and posterior intersectoral correlations. We also report YRD-integrated prior uncertainties and normalized innovation chi-squared diagnostics to assess sectoral coupling and the consistency of the combined prior uncertainty and model-data mismatch assumptions.

3) We have clarified the inversion implementation and observation-screening procedures, including background corrections, window initialization, and the effects of wind-speed filtering. We have also shortened the results sections, corrected technical descriptions and references.

The point-by-point responses are provided below:

The manuscript presents an inversion of methane emissions over East Asia, with a focus on the Yangtze River Delta region, using the CTDAS ensemble Kalman smoother coupled to the WRF-GHG (part of WRF-Chem) model and assimilating in-situ observations. The authors first test the ability of the system to estimate CH_4 emissions in an idealized setup using OSSEs over shorter periods and then apply the system to estimate emissions for the full year of 2022.

China is a highly relevant methane emitter (it has been the largest emitter since the 2000s), with large anthropogenic emissions from coal mining, rice cultivation, livestock, and waste treatment. Quantifying CH_4 emissions is therefore of great importance and this study presents a valuable contribution to this goal.

The authors present a well-designed inversion study for the real-world case (see comment below regarding the OSSEs), with extensive analyses, including independent validation sites, and place their results in the context of other bottom-up and top-down estimates for the same region. Overall, the manuscript is well written and easy to follow, and I enjoyed reading it. The manuscript fits well within the scope of ACP, and I recommend it for publication after revision. I was initially torn between recommending minor and major revisions, but I ultimately recommend a major revision. My concerns relate primarily to the design of the OSSEs, not to the sensitivity inversions themselves, but specifically to the way the "truth" was selected. In their current form, I believe that the OSSEs do not provide a meaningful assessment of the performance of the inversion system and, in particular, do not provide sufficient support for the claims regarding sectoral separation in the real-world inversion results.

I would therefore strongly encourage the authors to expand the OSSEs with newly designed experiments that test the various aspects of the system, especially its ability to separate the different sectors, in a more realistic and meaningful way.

As a minimum alternative, the discussion and conclusions regarding the OSSE results, particularly those concerning sectoral separation, could be revised to more clearly acknowledge the limitations of the current setup. However, I think this would be a missed opportunity to use the OSSE framework to test these aspects of the inversion more thoroughly.

Apart from this point, I find the inversion study convincing, and the manuscript well suited for publication.

General Comments

Basic design choice of OSSE:

The choice of using the true emissions as 2× the prior represents an extreme case for an inversion in several respects. First, setting the truth to twice the prior is a very strong deviation, which translates into a very strong signal in concentration space.

Second, using a uniform scaling of the prior to obtain the truth masks many of the challenges an inversion would encounter under more realistic circumstances. For example, sectoral or spatial misattribution would be largely masked, as everything is scaled by a factor of 2 anyway. Also, it makes it impossible to assess the ability to distinguish different sectors. While this choice provides some first insights into the system's performance, I think it should only be a starting point for the OSSEs.

An OSSE is already an idealized system that represents the absolute upper limit of what an inversion system could achieve (under the unrealistic assumption of perfect representation of meteorology), but the setup presented here makes it even more unrealistic. The presented numbers on how much of the true emissions the system is capable of recovering are therefore impossible to interpret.

In addition, as I understand it, no measurement noise was added to the pseudo-observations. To obtain more realistic results, measurement errors would need to be added (and ideally atmospheric transport would also need to be perturbed) to mimic measurement errors (and imperfect transport in a CTE).

RE: Thank you for this detailed assessment. We agree that uniformly doubling the prior emissions, combined with noise-free pseudo-observations and an identical transport configuration, creates a highly idealized experiment. This design evaluates the response to a prescribed emission increase but does not establish sectoral discrimination or quantify recovery under realistic observational and transport errors.

We have revised Sect. 2.3.1, Sect. 3.1, Sect. 3.2, and Sect. 4 accordingly. Recovery estimates are now explicitly restricted to the prescribed OSSE conditions, and claims about sectoral attribution that are not supported by the experimental design have been moderated or removed. We have also shortened the OSSE discussion and placed greater emphasis on the real-data inversion, while acknowledging its attribution uncertainties.

In the revised manuscript, we have acknowledged the limitations of the present OSSE (Sect. 4) and made its interpretation and conclusions more cautious accordingly. Reviewer 1 raised a similar concern, and we have addressed both comments by following one of Reviewer 1's suggested options, i.e., reducing the OSSE section and adjusting the interpretation of the sectoral attribution. We agree that more realistic OSSEs would strengthen the assessment, but additional simulations are beyond the scope and computational capacity of this study. We will take the reviewer's suggestions forward in our ongoing inversion work combining satellite and surface observations.

The specific revisions are provided below:

(Sect. 2.3.1, Page 14, Lines 320–326):

"*The synthetic truth is prescribed as twice the prior emissions for every optimized flux category and grid cell ($\boldsymbol{X}_{true} = 2\boldsymbol{X}_{prior}$; Wecht et al., 2012; Yu et al., 2021). It therefore preserves the prior spatial patterns and relative sectoral composition. No random noise is added to the pseudo-observations. The same WRF-Chem configuration and meteorological forcing are used to generate the pseudo-observations and perform the inversion, so this OSSE does not test the effects of transport-model mismatch. It evaluates recovery of the prescribed scaling and sensitivity to inversion settings, but cannot assess sectoral separation under a different true sectoral composition, or the errors of the real-data inversion."*

(Sect. 3.1.2, Page 24, Lines 499–515):

"Because the prior sector perturbations are independent by construction, their true intersectoral correlations are zero. The weak period-mean intersectoral correlations in Table B1 ($|C^b|$<0.10) are consistent with finite-ensemble sampling noise. In the posterior ensemble, stronger negative correlations between rice cultivation and natural emissions (-0.38) and between rice cultivation and waste (-0.33) indicate compensating variations in the estimated sectoral emissions, indicating that these sectors cannot be constrained fully independently. The estimated averaging kernel (AK) matrix provides complementary information on these sectoral responses (Appendix C; Table C1). Natural emissions have the largest diagonal element (0.39), indicating the strongest response within the same sector among the six categories, consistent with their relatively large posterior adjustment. The off-diagonal averaging-kernel elements show substantial coupling between natural emissions and both rice and waste. The natural-to-rice and natural-to-waste sensitivities are 0.27 and 0.19, respectively, while the corresponding reverse sensitivities are 0.19 and 0.21, indicating that these sectoral signals are not fully separable. The livestock and "Other sectors show weaker self-sensitivity, with diagonal elements of 0.068 and 0.093, respectively. The coupling between sectors is asymmetric, as natural emissions exert little influence on the livestock and "Other estimates (0.014 and 0.007), whereas "Other emissions have a considerably stronger influence on the natural-sector estimate (0.25). Overall, these diagnostics indicate limited separation of the six emission sectors by the observing network."

(Sect. 3.2.2, Page 32, Line 605–610):

"Overall, the posterior estimates indicate a stronger summer emission peak, with much of the seasonal adjustment assigned to rice cultivation and natural emissions. This allocation depends on the prescribed prior uncertainties and intersectoral coupling, as examined in Sect. 3.2.3. Further evaluation of the sectoral attribution would require OSSEs with distinct perturbations for individual sectors. The seasonal estimates also retain the prescribed soil CH_4 sink, whose seasonal variability is included but not adjusted by the inversion."

(Sect. 3.2.3, Page 34, Line 629–642):

"*These reductions in ensemble spread are accompanied by changes in intersectoral dependence. Pearson correlations between YRD-integrated sector totals become more negative after assimilation (Appendix B; Table B2), most notably between rice cultivation and natural emissions (-0.29), between rice cultivation and waste (-0.21), and between natural emissions and waste (-0.16). These negative correlations indicate that different combinations of sectoral emissions remain compatible with the*

observational constraints, so the sectors cannot be fully constrained individually.

The estimated averaging kernel (AK) matrix further characterizes these sectoral responses (Appendix C; Table C2). Its period-mean diagonal elements are largest for waste (0.42), natural emissions (0.32), and rice cultivation (0.30), indicating relatively greater sensitivity of their posterior estimates to changes in the same sectors. The smaller values for livestock (0.080) and "Other (0.097) are consistent with their modest uncertainty reductions. Off-diagonal responses, including those from natural emissions to rice cultivation (0.20) and from rice cultivation to natural emissions (0.16), also indicate coupling between sectors. These diagnostics suggest that reduced emission uncertainty within the YRD coexists with uncertainty in the allocation of emissions among individual sectors."

(Sect. 4, Page 41, Lines 755–769):

"Several limitations should be considered when interpreting these results. The OSSE uses uniformly doubled prior emissions and noise-free pseudo-observations generated with the same transport configuration used in the inversion. Therefore, it evaluates emission adjustments under idealized conditions, while independent sectoral recovery and sensitivity to transport and background errors require further testing. Its short duration and unexplored interactions among inversion settings (e.g., prior uncertainty and the MDM) also limit the generalizability of the sensitivity results. Sectoral estimates remain influenced by the prior sectoral emission magnitudes and spatial distributions, and by intersectoral coupling. The extent to which the sectoral responses in these idealized OSSEs carry over to real-data inversions remains uncertain. ... Further evaluation of CTDAS-WRF should therefore examine independently perturbed sectoral priors and truths, background optimization, longer analysis periods, and transport error. Such tests, potentially complemented by satellite observations, would provide a stronger basis for interpreting the real-data inversions."

Further setup of the OSSE:

o It is a pity that the OSSE does not test the optimization of different sectors in a setup with differently perturbed sector priors (or, conversely, different sector truths). Although an OSSE is limited in its implications for real-world applications, it provides a relatively simple, cheap, and easy "playground" to test such aspects, and I would recommend doing this. Such tests would also be necessary to support the real-world interpretations as they are made now.

o Another missed opportunity is that neither perturbations nor the optimization of the background were included in the OSSEs.

o The OSSEs are performed for a relatively short period of 40 days. In the presented setup this seems sufficient, but in a more realistic set-up it would be useful to test a longer period.

o The selected a priori uncertainties are not consistent with the way the truth is defined (2x prior). In this setup, the a priori uncertainties are too conservative.

RE: Thank you for these constructive suggestions. We agree that OSSEs with different perturbations for individual sectors would provide a more direct test of sectoral attribution. Perturbing and optimizing the background, and extending the experiment beyond the current 40-day period, would also help assess the inversion under more realistic conditions. These are valuable extensions, but the additional simulations are beyond the scope and computational capacity of the present study.

We also agree that the prescribed doubling of emissions exceeds the assigned prior standard deviations. The resulting posterior adjustments therefore reflect a balance between fitting the pseudo-observations and the prior constraint; As noted in our previous response, we have tempered our interpretation of the OSSEs and our claims about sectoral attribution. We now clarify their limitations in Sect. 4 and will address the suggested experiments in our ongoing work combining satellite and surface observations. The specific manuscript revisions are detailed in the response above.

Setup choices in the real-world inversion:

Overall, I think the setup is well designed, and the results look convincing and are in good agreement with other studies, as shown by the authors.

My main concern is the a priori uncertainty, which, although in agreement with other studies such as Wang et al. (2025), is in my opinion chosen to be quite low. An uncertainty of 30%, 40%, or 50% at the grid-cell level results in a quite low uncertainty at the country level. The evaluation of the chosen uncertainty (in conjunction with the chosen MDM) could be done using metrics such as the chi-squared statistic. I am missing this diagnostic in the manuscript. This (or other comparisons with the a priori residuals) would also be useful for evaluating the MDM choices.

RE: Thank you for this comment. We agree that the prior uncertainty of the annual YRD total is small: 8.46 ± 0.16 Tg CH_4 yr^{-1}, or 1.9%, despite the 30%–50% uncertainties assigned at the grid-cell level. We have added this regional estimate to Appendix D (Table D1) and acknowledge that the small aggregated uncertainty may constrain regional adjustments. To evaluate the prior uncertainty and MDM together, we added normalized innovation chi-square diagnostics (Table D2). The network-wide value is 1.42 for 18,551 observations. This suggests that the combined error settings capture the overall scale of the innovations, although the station-level range of 0.74–3.30 indicates that some sites remain less well represented. Because the diagnostic reflects both error components, it does not validate either setting separately. We have added this interpretation to Sect. 4; the SARI-specific MDM sensitivity tests are reported in Supplement Table S3. For ease of reference, we have copied the newly added Appendices C and D to the end of this response document.

(Sect. 4, Page 40–41, Line 743–748):

"*Uncertainty reductions are concentrated in observation-influenced regions and differ among sectors. Within the YRD, estimated averaging kernels indicate relatively greater responses within the waste, natural, and rice cultivation sectors. However, negative posterior intersectoral correlations and off-diagonal AK responses indicate coupling between sectors. Thus, while total and sectoral flux uncertainties are reduced, uncertainty in the allocation of emissions among sectors remains. Further assessment would require OSSEs with distinct perturbations for individual sectors."*

(Sect. 4, Page 41, Line 763–766):

"In the real-data inversion, the chi-squared innovation diagnostics indicate incomplete consistency between the residuals and the R-matrix; neither the empirical SARI MDM settings nor the network-wide incorrect application of the 10 m s^{-1} wind-speed cutoff has been fully evaluated."

"Table S2. Prior χ^2_{innov} diagnostics for the 50 and 100 ppb SARI MDM experiments during 28 June–

27 July 2022.

10-day cycle (2022)	*SARI MDM (ppb)*	*SARI station prior* χ^2_{innov}	*all station prior* χ^2_{innov}
28 June−7 July	*50*	*1.55*	*2.56*
	100	*0.41*	*2.32*
8 July−17 July	*50*	*3.70*	*2.45*
	100	*1.30*	*1.97*
18 July−27 July	*50*	*1.16*	*2.67*
	100	*0.31*	*2.57*

** The prior* χ^2_{innov} *diagnostic is described in detail in Appendix D."*

Specific Comments

Page 2, Line 45–46: These numbers do not say much, as they are highly dependent on the design of the OSSE (see my general comments).

RE: Thank you for this comment. We have removed the recovery percentages and the associated performance claim from the abstract. The revised text explicitly identifies the idealized OSSE setup and reports only the main spatial pattern of posterior adjustments.

(Abstract, Page 2, Lines 44–46):

"Idealized Observing System Simulation Experiments (OSSEs), using a uniform doubling of prior emissions as the synthetic truth, indicate that posterior adjustments are concentrated mainly in the Yangtze River Delta (YRD), South Korea, and northern Japan."

Page 4, Line 22: It is unclear to me why the authors state that Lagrangian models are less efficient or provide less constraint over larger regions than Eulerian models. There are many inversion systems based on Lagrangian models that do exactly this.

RE: Thank you for pointing this out. We agree that Lagrangian models are widely used for regional inversions and that our comparison was not justified. We have removed this statement and revised the text to acknowledge previous tower-based studies in the YRD and introduce our Eulerian framework without suggesting a general advantage over Lagrangian approaches.

(Page 4, Lines 120–122):

"*Hu et al. (2023, 2026) used tower observations to investigate* CH_4 *emissions and seasonal variability in the YRD. As a complementary approach, we use a regional Eulerian inversion framework to examine emission seasonality across the YRD.*"

Table 1: How do the geological emissions from Etiope et al. (2019) look in this region? Geological emissions from this dataset are frequently scaled strongly downward in Italy and Eastern Europe, where the dataset is considered to be unrealistically high. Could this also be the case here, especially in Japan? Could this drive some of the downward corrections in the inversion?

RE: Thank you for raising this concern. We compared the prior and posterior natural CH_4 fluxes (Fig. R1) with the prior natural-source components, including the geological emissions from Etiope et al.

(2019) (Fig. S2). In parts of Japan, substantial downward adjustments coincide with high prior geological fluxes. We consider it likely that the geological prior is too high in these areas and contributes to the adjustments, consistent with your concern. We agree with the reviewer that this issue deserves attention, while keeping the analysis and discussion in the manuscript focused on the YRD emission characteristics.

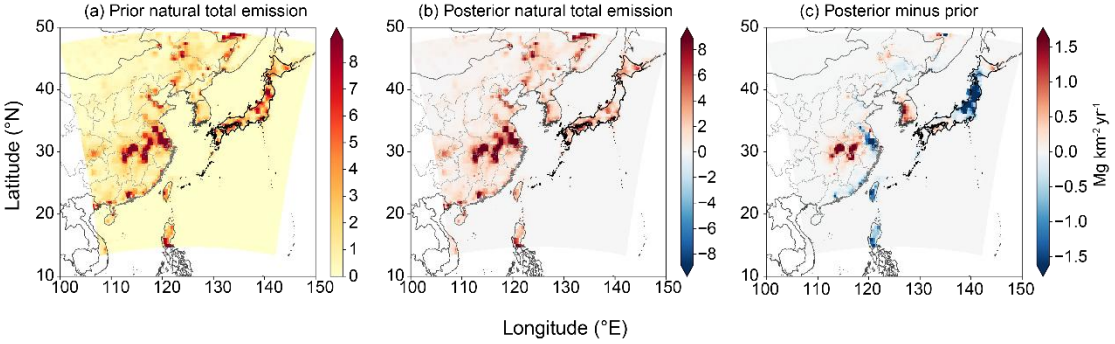
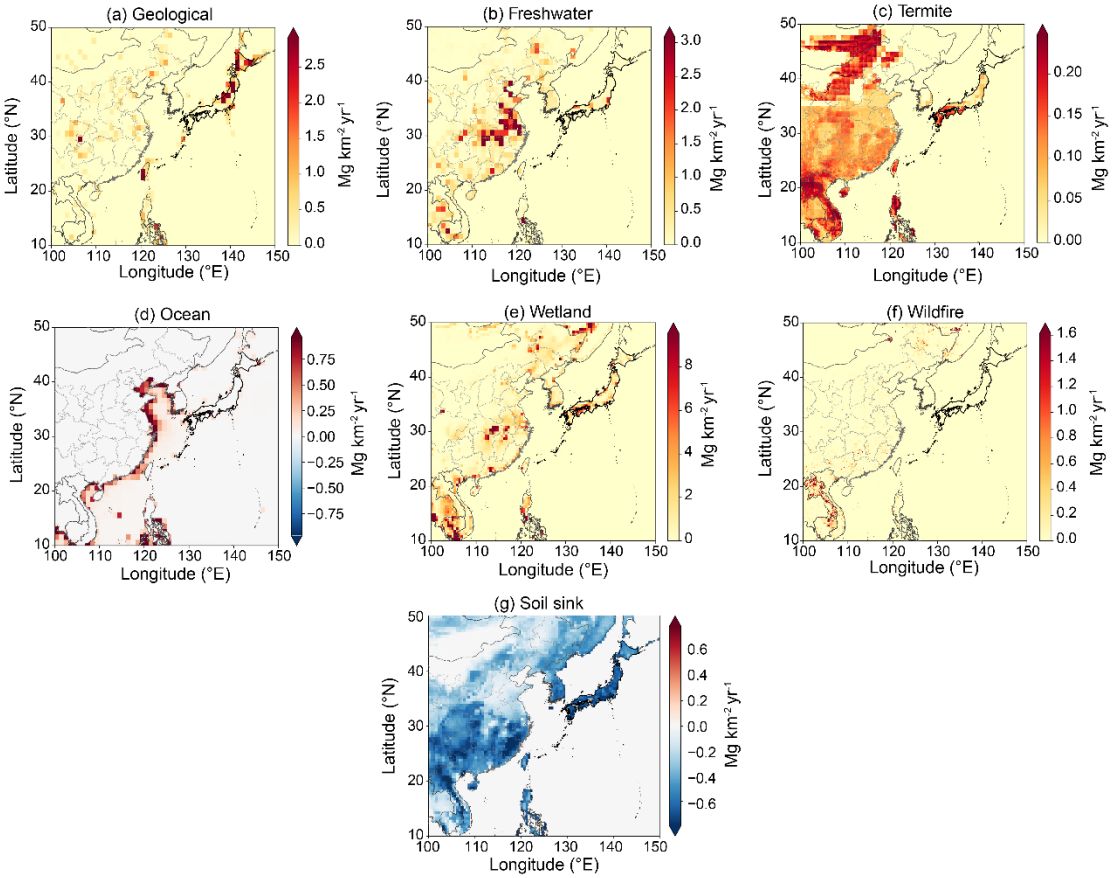


Figure R1: Spatial distribution of annual-averaged (a) prior and (b) posterior natural CH_4 fluxes and (c) their difference (posterior minus prior) over the parent domain (D01) in2022.



"Figure S2: Spatial distribution of annual-averaged prior natural CH4 emissions from (a) geological, (b) freshwater, (c) termite, (d) ocean, (e) wetland, (f) wildfire and (g) soil sink sources over the parent domain (D01) in 2022."

Page 11, Line 267: The number of ensemble members should roughly match the degrees of

freedom in the a priori uncertainty matrix. What is the degree of freedom here?

RE: Thank you for this comment. We estimated the effective degrees of freedom to be 571 for the prescribed prior covariance of a single newly initialized 10-day window, compared with 28,574 state-vector elements. We used 100 ensemble members in this study, based on our sensitivity test and on previous work showing that localization combined with a smaller ensemble can yield results similar to those of a much larger ensemble.

We apply localization to the Kalman gain with a length scale of 900 km (3 × 300 km, the correlation length). Localization limits the projection of sampling noise, which arises from the finite ensemble size, onto the state vector, and thereby focuses the update on the state-vector elements closest to the observation locations. This reduces the number of ensemble members required relative to the degrees of freedom of the prior covariance. Consistent with this, previous studies reported the following:

- Fig. 6 of Peters et al. (2005, https://doi.org/10.1029/2005JD006157) showed that 200 members with localization (three-fold correlation length) produced an annual-mean uncertainty-reduction pattern similar to that of 1,500 members without localization.
- Steiner et al. (2024, https://doi.org/10.5194/acp-24-2759-2024) used 192 members for a prior covariance with 425 effective degrees of freedom. Flux errors were larger with 50 members or without Kalman gain localization, but improved little with 300 members.
- Thanwerdas et al. (2025, https://doi.org/10.5194/gmd-18-1505-2025) reported 261 degrees of freedom per window. With 50 samples, localization raised the mean error reduction from −21.9% to 23.3%.

In our sensitivity test, increasing the ensemble from 100 to 150 members improved the YRD total-flux error reduction from 55.7% to 62.2% (Table 4). Thus, we acknowledge that increasing the ensemble size could improve the results. We have removed the statement that 100 members are sufficient for convergence and added the degree of freedom in the revised manuscript.

(Page 12 Line 279–281):

"Using the effective-dimension measure of Bretherton et al. (1999), we estimated the effective degrees of freedom to be 571 for the prescribed prior covariance per newly initialized window."

Page 13, Line 314: An OSSE framework with the assumption of perfect atmospheric transport can only assess the absolute upper limit of an inversion system's ability to recover the true emissions. As I pointed out in the general comments, with the current setup, the OSSE results represent an unrealistic scenario and do not allow an assessment of how much of the true emissions can really be recovered.

RE: Thank you for this clarification. We have revised the OSSE description and interpretation to acknowledge its idealized assumptions, including identical transport and noise-free pseudo-observations. We now present it as a test of the system's response to prescribed emission scaling and inversion settings, rather than an assessment of emission recovery under real-world conditions.

(Page 14, Line 315–326):

"To evaluate the internal response of the inversion system under controlled conditions, we conduct an Observing System Simulation Experiment (OSSE). WRF-Chem is driven and nudged toward the analyzed meteorological fields and run with a prescribed set of synthetic emissions (referred to as true state hereafter). The resulting CH4 concentration fields are then sampled at the same hour-specific times and locations as the real stations to generate pseudo-observations (Brasseur and Jacob, 2017).

The synthetic truth is prescribed as twice the prior emissions for every optimized flux category and grid cell ($\boldsymbol{X}_{true} = 2\boldsymbol{X}_{prior}$; Wecht et al., 2012; Yu et al., 2021). It therefore preserves the prior spatial patterns and relative sectoral composition. No random noise is added to the pseudo-observations. The same WRF-Chem configuration and meteorological forcing are used to generate the pseudo-observations and perform the inversion, so this OSSE does not test the effects of transport-model mismatch. It evaluates recovery of the prescribed scaling and sensitivity to inversion settings, but cannot assess sectoral separation under a different true sectoral composition, or the errors of the real-data inversion."

Page 15, Line 361: This probably makes the SARI station prone to local contamination, which then results in the large mismatches that the authors need to account for with a very large MDM.

RE: Thank you for this comment. We have revised the description to acknowledge that SARI observations may capture local CH_4 variability that is not fully resolved at the model grid scale, potentially contributing to the larger model-observation mismatches considered when assigning its MDM.

(Page 16, Line 373–376):

"*The SARI station is situated in a densely developed scientific innovation and educational district in Shanghai, providing observations of urban CH4 variability. These observations may capture local enhancements that are not fully resolved at the model grid scale, contributing to model-observation mismatches.*"

Page 16, Line 394: I don't understand why an upper wind-speed limit is used as a rejection threshold for the observations. Why would it be a problem when emissions from a broader area are transported? I would understand a threshold for high wind variability, which is difficult to represent well in a model, but not for high wind speeds in general.

RE: Thank you for pointing out this issue. We acknowledge that we had incorrectly applied to all stations a criterion that Zhang et al. (2022) used only for observations from the Waliguan station (WLG), without validating its applicability at each station. Transport from a broader area alone does not justify rejecting an observation, and we have removed this explanation from Sect. 2.4.

We have quantified the observations excluded specifically by the 10 m s^{-1} upper wind-speed cutoff. At the two stations within the Yangtze River Delta (YRD), SORPES and SARI, this cutoff excluded only one hourly observation in total (none at SORPES and one at SARI). Thus, the cutoff caused negligible observational loss at the two YRD stations.

Outside the YRD, including some island stations, for example, it removed 519 observations at HAT (18.69%) and 566 at YON (21.03%), respectively. These larger exclusions could affect the observational constraints supplied by those stations. We have explicitly acknowledged this limitation in Sect. 4.

The revised Supplement provides the number of observations removed for all nine stations in Table S3.

Table S3. Number of observations excluded by the lower and upper wind-speed thresholds at each

station.

Site ID	*Before wind filtering*	*Wind speed < 1.5 m s^{-1}*	*Wind speed > 10 m s^{-1}*
AMY	*2622*	*129*	*119*
GSN	*2497*	*45*	*330*
HAT	*2876*	*98*	*519*
YON	*2757*	*65*	*566*
ULD	*2620*	*160*	*397*
RYO	*2687*	*129*	*122*
COI	*2864*	*34*	*472*
SORPES	*1568*	*191*	*0*
SARI	*2225*	*74*	*1*
Total	*22716*	*925*	*2526*

(Page 16, Lines 386–389):

"An upper wind-speed cutoff of 10 m s^{-1} was incorrectly applied to all stations (Table S3), following the criterion used by Zhang et al. (2022), who applied it only to observations from the Waliguan station (WLG). The limitations of applying the 10 m s^{-1} wind-speed cutoff across all stations are discussed in Sect. 4. The retained observations are used in the inversion."

(Page 41, Lines 763–765):

"*In the real-data inversion, the chi-squared innovation diagnostics indicate incomplete consistency between the residuals and the R-matrix; neither the empirical SARI MDM settings nor the network-wide incorrect application of the 10 m s^{-1} wind-speed cutoff has been fully evaluated.*"

Figure 4: In the YRD, Northern Japan, and South Korea, there is always one window in which the posterior deviates quite significantly from the other windows and from the true state. These features are not discussed**.**

RE: Thank you for highlighting these features. We have added a discussion of the anomalous windows in all three regions (the YRD, Northern Japan, and South Korea). The scaling factors are not strongly correlated across optimization windows, allowing the posterior adjustments to differ between windows. We also discuss possible contributors to this variability, while making clear that these explanations remain hypotheses and that the causes of the individual departures have not been separately diagnosed.

(Page 22, Lines 471–478):

"At the parent-domain scale, the posterior total is 22.5% higher than the prior. Larger relative adjustments occur in the three focus regions, where the posterior totals are 60.8%, 67.9%, and 92.7% above the prior in the YRD, South Korea, and northern Japan, respectively. These adjustments represent different degrees of recovery toward the prescribed 100% emission increase. Recovery also varies among the 10-day windows; for example, the increase in South Korea ranges from 36.6% to 86.2%, and each regional time series contains a window that remains closer to the prior than the adjacent windows. Although not diagnosed separately, this variability may reflect meteorology-dependent observation-flux sensitivity and differences in observation availability (Chen and Prinn, 2006; Steiner et al., 2024)."

Sect. 3.1.2: Apart from the fact that the assessment of the sectorial emissions cannot be made in a

meaningful way with the current setup, this section completely misses the discussion of the a posteriori covariances between sectors.

RE: Thank you for this constructive comment. We agree that the current OSSE does not demonstrate reliable sectoral separation and that the original section omitted the discussion of intersectoral posterior covariances. We have revised Sect. 3.1.2 to interpret the sectoral results as adjustments conditional on the prescribed prior uncertainties and observation-flux sensitivities. We now explicitly discuss the prior and posterior intersectoral correlations (Table B1), with negative posterior correlations indicating compensating variations between sectors. Together with the low diagonal and nonzero off-diagonal elements of the estimated averaging kernel matrix (Appendix C; Table C1), these diagnostics indicate limited separation of the six sectors. This limitation is now explicitly stated in the revised section. The relevant revisions are reproduced below:

(Page 24, Lines 499–515):

"Because the prior sector perturbations are independent by construction, their true intersectoral correlations are zero. The weak period-mean intersectoral correlations in Table B1 ($|C^b|$<0.10) are consistent with finite-ensemble sampling noise. In the posterior ensemble, stronger negative correlations between rice cultivation and natural emissions (-0.38) and between rice cultivation and waste (-0.33) indicate compensating variations in the estimated sectoral emissions, indicating that these sectors cannot be constrained fully independently. The estimated averaging kernel (AK) matrix provides complementary information on these sectoral responses (Appendix C; Table C1). Natural emissions have the largest diagonal element (0.39), indicating the strongest response within the same sector among the six categories, consistent with their relatively large posterior adjustment. The off-diagonal averaging-kernel elements show substantial coupling between natural emissions and both rice and waste. The natural-to-rice and natural-to-waste sensitivities are 0.27 and 0.19, respectively, while the corresponding reverse sensitivities are 0.19 and 0.21, indicating that these sectoral signals are not fully separable. The livestock and "Other sectors show weaker self-sensitivity, with diagonal elements of 0.068 and 0.093, respectively. The coupling between sectors is asymmetric, as natural emissions exert little influence on the livestock and "Other estimates (0.014 and 0.007), whereas "Other emissions have a considerably stronger influence on the natural-sector estimate (0.25). Overall, these diagnostics indicate limited separation of the six emission sectors by the observing network."

Sect. 3.1 and Sect. 3.2: The two sections are a bit lengthy; consider using more concise language.

RE: Thank you for this suggestion. We have shortened Sects. 3.1 and 3.2 by approximately one-third overall, removing repetitive descriptions and focusing on the main findings while retaining essential methodological explanations and limitations. Most of the paragraphs reproduced below have been reduced in word count by approximately half.

(Sect. 3.1.1 Page 21, Lines 445–451 and Page 22, Lines 471–473):

"While the synthetic truth was defined by a uniform twofold scaling of the prior, the posterior response is spatially heterogeneous (Fig. 5). Coherent positive adjustments occur in observation-rich regions (the YRD, South Korea, and northern Japan) and also over a localized coal-mine hotspot in Shanxi Province (34.57–40.73°N, 110.23–114.55°E). These adjustments weaken toward surrounding areas and are negligible across most of the remaining land domain far from observation sites in both the flux and concentration fields (Fig. S4 and S5), demonstrating that the response to the prescribed perturbation is

spatially confined to areas close to the observations."

"At the parent-domain scale, the posterior total is 22.5% higher than the prior. Larger relative adjustments occur in the three focus regions, where the posterior totals are 60.8%, 67.9%, and 92.7% above the prior in the YRD, South Korea, and northern Japan, respectively."

(Sect. 3.1.2 Page 23, Lines 485–488 and Lines 489–491):

"The total posterior-to-prior ratio shows widespread adjustments across the YRD, but weaker responses in the southern and northwestern areas (Fig. 7a). With SORPES as the only station within the YRD, this spatial contrast is consistent with spatially varying observation-flux sensitivity and the distance-dependent damping of observation-flux covariances by localization (Steiner et al., 2024).

The predefined categories show different posterior responses, with YRD-integrated posterior-to-prior ratios of 2.2, 1.6, 1.5, 1.4, 1.1, and 1.0 for natural emissions, rice cultivation, waste, fuel exploitation, livestock, and "Other", respectively (Fig. 7b–i)."

(Sect. 3.1.3 Page 25, Lines 523–525):

"Table 4 summarizes the reduction in total and sectoral flux errors over the parent domain (D01) and the YRD. In the reference case, C_{mae} ranges from 3.0% to 25.0% across the total flux and individual sectors over D01, whereas the total-flux error reduction reaches 55.7% over the YRD."

(Sect. 3.2.1 Page 28, Lines 552–557):

"Major emission hotspots are present in both the prior and posterior estimates, located in central and eastern China and the Pearl River Delta (Fig. 8a–b). Posterior emissions are lower than the prior over the YRD, Shanxi, and northern Japan (north of 37° N), but higher over western South Korea and eastern Hubei province in China (29.97–31.37°N, 113.68–115°E) (Fig. 8c). Within the YRD, emissions and the largest downward adjustments are concentrated in northern Anhui, southern Jiangsu, and Shanghai. A localized upward adjustment occurs in southwestern Anhui (Fig. 8d–f)."

(Sect. 3.2.2 Page 31, Lines 583–588):

"Figure 11 summarizes the seasonal variability of YRD CH_4 emissions and the contributions of individual sectors to the inferred emission adjustments (posterior minus prior). Prior emissions vary little throughout the year (0.022–0.025 Tg d^{-1}), whereas posterior emissions exhibit a summer maximum of approximately 0.04 Tg d^{-1} in July–August. The seasonal amplitude increases from approximately 0.003 to 0.03 Tg d^{-1}, with posterior emissions exceeding the prior from June to mid-September and remaining lower during the rest of the year."

(Sect. 3.3.2 Page 37–38, Lines 690–696):

"*The inversion reduces total CH_4 emissions to 65.77 ± 0.77 Tg yr^{-1} over the parent domain and 7.53 ± 0.10 Tg yr^{-1} in the YRD, representing decreases of 5.1% and 11.0% relative to the prior, respectively, driven mainly by reductions in fuel exploitation and waste emissions (Fig. 14a, b). The posterior anthropogenic estimate in the YRD is 6.18 Tg yr^{-1}, closest to CEDS-2025, differing by only 1.7%. The YRD posterior anthropogenic estimate is also relatively close to EDGAR_2025 but lower than the other anthropogenic estimates by 14.2–31.2%, while posterior total emissions are 6.7% and 24.2% lower than CT-CH4-2025 and CAMS v25r1, respectively.*"

Sect. 3.2.2: As mentioned in my general comment, this section would greatly benefit from a discussion of the limitations in the sectoral attribution of emissions. Some of this is now placed in Sect. 3.2.3. But overall, I think the whole discussion about the sectorial attribution in this manuscript is overconfident. Sectoral attribution remains a very difficult task for inversion systems, and this would be important to emphasize here. And as mentioned above, this could or should be supported with accordingly designed OSSEs.

Thank you for this comment. We have moderated sectoral attribution claims throughout the manuscript and explicitly discussed the influence of prior uncertainties and intersectoral coupling in Sect. 3.2.2. We have also added an estimated averaging-kernel analysis in Appendix C and incorporated its implications into Sect. 3.2.3. These revisions distinguish the sectoral adjustments inferred by the inversion from independently validated source contributions. Further validation using OSSEs with distinct sector-specific perturbations is identified as a priority for future work.

(Page 31–32, Lines 595–600):

"Natural emissions account for a further 0.35 Tg (41.7%) of the upward adjustments. These summertime increases are consistent with previous studies reporting enhanced rice and wetland CH_4 emissions during the warm season (Chen et al., 2013; Huang et al., 2021; East et al., 2024). However, the partitioning of the inferred summer enhancement between rice cultivation and natural emissions remains sensitive to the prescribed prior uncertainties and intersectoral coupling (Sect. 3.2.3)."

(Page 32, Lines 605–610):

"Overall, the posterior estimates indicate a stronger summer emission peak, with much of the seasonal adjustment assigned to rice cultivation and natural emissions. This allocation depends on the prescribed prior uncertainties and intersectoral coupling, as examined in Sect. 3.2.3. Further evaluation of the sectoral attribution would require OSSEs with distinct perturbations for individual sectors. The seasonal estimates also retain the prescribed soil CH_4 sink, whose seasonal variability is included but not adjusted by the inversion."

Technical Comments

Page 3, Line 96: IMI does not combine TROPOMI and GOSAT observations, but uses the blended TROPOMI+GOSAT product from Harvard.

Page 4, Line 99: Ioannidis et al. (2026) is a study presenting the results of a variety of inversion

systems, many of which are not based on CTDAS.

RE: Thank you for these corrections. We have clarified that IMI 2.0 uses the Harvard blended TROPOMI+GOSAT product, rather than directly combining observations from the two satellites. We have also specified that CTDAS-WRF was one of the systems compared by Ioannidis et al. (2026), without implying that all systems were based on CTDAS.

(Page 3–4, Lines 94–101):

"*For example, Estrada et al. (2025) developed the Integrated Methane Inversion version 2.0 (IMI 2.0) and applied it to estimate 2023 CH_4 emissions over the contiguous United States using the Harvard*

blended TROPOMI+GOSAT product (Balasus et al., 2023), with a native transport-model resolution of approximately 25 × 25 km. In Europe, Steiner et al. (2024) coupled the CarbonTracker Data Assimilation Shell (CTDAS) with ICON-ART to estimate CH_4 emissions from ground-based observations for 2008, 2013, and 2018. Ioannidis et al. (2026) compared multiple inversion systems, including a CTDAS-WRF configuration that optimized anthropogenic and natural CH_4 emissions for the same years using in situ observations."

Page 4, Line 108: Steiner et al. (2024) did not look at different inversions constrained by satellite and surface-only networks. Sicsik-Paré et al. (2026) did this, for example.

RE: Thank you for this correction. We have removed the inappropriate citation to Steiner et al. (2024) and revised the paragraph to cite studies directly comparing satellite- and surface-observation-constrained inversions, including Sicsik-Paré et al. (2026).

(Page 4, Lines 108–113):

"*... Direct comparisons of methane inversions constrained by satellite and in situ observations show that the two observing systems provide complementary information but can yield different posterior emission estimates across regions because of their contrasting coverage and sensitivity (Cressot et al., 2014; Lu et al., 2022; Sicsik-Paré et al., 2026; Sun et al., 2026; Ward et al., 2026). Such comparisons are therefore important for assessing the robustness of regional inversions and interpreting inconsistencies under different observational constraints.****"***

Page 5, Line 151: Nudging to what? Observations or the driving data?

RE: Thank you for pointing out this ambiguity. Spectral nudging is applied toward the driving meteorological fields, not directly toward observations. We have clarified this in the revised text.

(Page 5, Lines 149–151):

"*Spectral nudging is applied to constrain the large-scale components of the simulated meteorological fields toward the driving meteorological fields (Lai and Gan, 2025).****"***

Fig. 1: Generally, the labels are too small. Especially in c), the relevant parts of the plot at the edges are much too small. Also, the grid shows noticeable aliasing, which should be avoided in the figure rendering.

RE: Thank you for this comment. We have enlarged the figure and labels and increased the rendering resolution to reduce grid aliasing. We have also moved the former Fig. 1c to a separate Fig. 2 in Sect. 2.2.1, where the tile concept is first introduced, to improve the readability of details at the edges.

Page 6, Line 173: I think I know what the authors want to say with this sentence, but it is not very clearly formulated.

RE: Thank you for this comment. We have rephrased this passage to distinguish the initialization of successive simulation windows from the application of lateral boundary conditions and to clarify the roles of the specified and relaxation zones.

(Page 6, Lines 172–177):

"*...CTDAS applies additive corrections to the initial and lateral boundary concentrations using eight background parameters. The WRF lateral boundary zone is five grid cells wide (Fig. 1b), comprising a*

one-cell specified zone and a four-cell relaxation zone (Wong et al., 2012). Background corrections vary along the boundaries but remain uniform across this five-cell width. In the initial concentration field, the corrections extend throughout the domain, gradually transitioning from the corresponding boundary offsets to the mean of all eight parameters toward the domain interior. These corrections are applied uniformly across vertical levels."

Page 10, Line 243: Initializing each newly entered window with an unscaled state vector ($\lambda = 1$) deviates from other studies, especially other studies using CTDAS, which propagate information from the previously optimized windows into the new window with a certain weighting factor. There are arguments for each approach; what is the authors' argument for doing it this way?

RE: Thank you for this comment. We agree that weighted propagation of previously optimized scaling factors can help retain information about persistent inventory biases. In our implementation, the flux-scaling-factor ensemble for each newly introduced window is centered on $\lambda = 1$. Because the temporal persistence of inventory corrections has not been established in this study, we retain the prescribed emissions ($\lambda = 1$) for each new period as its prior emissions, without extrapolating the preceding period's posterior correction. This avoids transferring potentially period-specific corrections into the new prior. We acknowledge that this choice also limits the transfer of information about persistent inventory biases. We have added this rationale and trade-off to Sect. 2.2.1.

(Page 11, Lines 258–262):

"The newly introduced window (window 3) is initialized with a flux-scaling-factor ensemble centered on $\lambda = 1$. Because the temporal persistence of inventory corrections has not been established in this study, we retain the prescribed emissions for each new period as its prior emissions rather than extrapolating the preceding period's posterior correction. This avoids directly carrying potentially period-specific corrections into the new prior."

Page 12, Line 282: 0.1%, not 1%.

RE: Thank you for pointing out this error. We have corrected "1%" to "0.1%".

(Page 13, Lines 288–290):

"... the prior uncertainty for each background parameter is constrained to 2 ppb (~0.1%) for the eight boundary regions where CH_4 from the CAMS global product is used to represent the modeled CH_4 mole fraction,"

Page 12, Line 286: "...estimates, combined..." → "...estimates combined ..."

RE: Thank you for pointing this out. We have corrected the comma placement.

(Page 13, Lines 292–293):

*"The **R** matrix is diagonal and consists of reported measurement uncertainties and model-data mismatch (MDM) estimates combined under the assumption that the two sources of error are independent."*

Page 19, Line 416 / Fig. 3: So why are there then adjustments of the ocean fluxes at all? How is this possible when the only emissions there are not part of the state vector?

RE: Thank you for raising this point. The prescribed oceanic CH_4 fluxes are excluded from the state

vector and remain unchanged. The adjustments over ocean grid cells arise from EDGAR shipping and aviation emissions, which are included in the optimized "Other" category. These adjustments are small in absolute terms but can produce noticeable posterior-to-prior ratios in the original Fig. 3b (now Fig. 5b) where prior emissions are low.

Figure 4: The figure labels are too small.

RE: Thank you for pointing this out. We have increased the font size of the labels in Fig. 4 (now Fig. 6) to improve their readability.

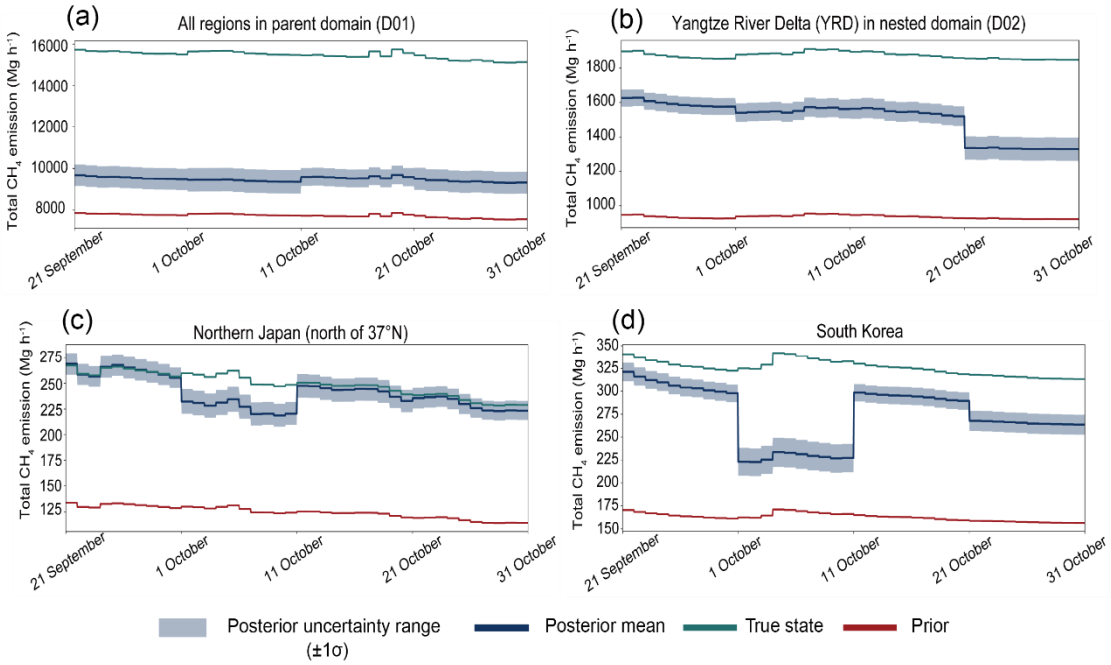


Figure 6: Temporal variation of hourly CH_4 emissions … The shaded posterior uncertainty range represents ±1 standard deviation (1σ) around the posterior mean."

Sect. 3.2: A map showing simply the emissions for the YRD would be useful (similar to Fig. S3, but only for the YRD).

RE: Thank you for this suggestion. We interpret this comment as requesting a map restricted to the YRD that shows the spatial distribution of CH_4 emissions. This information is already presented in the original Fig.7d (now Fig. 8d). If a different emission quantity or form of visualization was intended, we would appreciate further clarification.

Page 29, Line 601: Fig. 3d–f, not Fig. 4d–f.

RE: Thank you for identifying the incorrect figure reference. We have removed the relevant passage to streamline Sect. 3.2.

(Page 28, Lines 565):

"The spatial distributions of sectoral emissions and their adjustments are shown in Fig. S6."

Figure 11: In my opinion, simply showing a histogram of the concentrations would be nicer in the right column. In its current form, the right subfigure is not really intuitive to read.

RE: Thank you for this helpful suggestion. We have replaced the panels in the right column of Fig. 11

(now Fig. 12) with histograms of the concentration distributions and revised the figure caption accordingly.

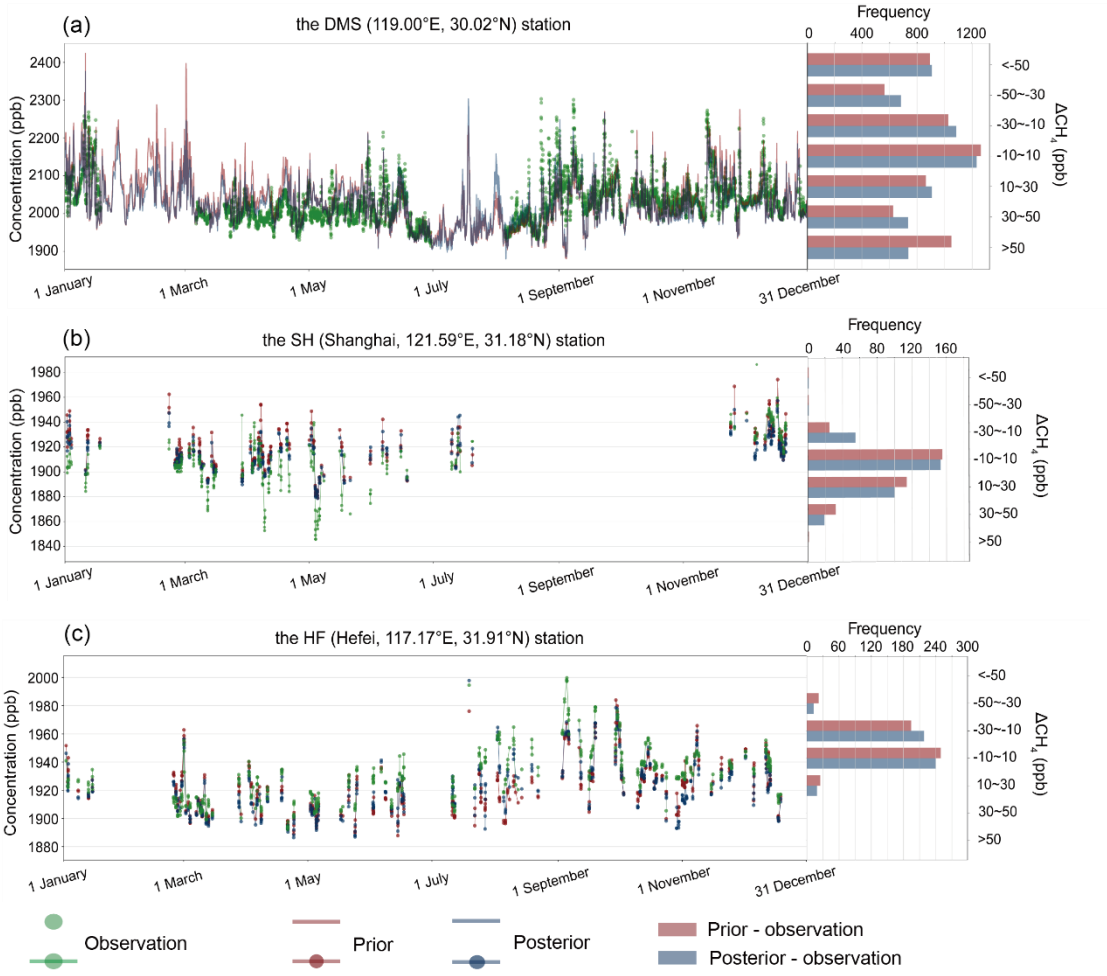


Figure 12: Hourly CH_4 concentrations (ppb) … at the (a) DMS station, (b) SH (Shanghai) station, and (c) HF (Hefei) station in the YRD during 2022."

Figure S3: For readability, the sectors should be indicated in the plot, not only in the caption.

RE: Thank you for this suggestion. We have added sector labels directly to Fig. S3 (now Fig. S6) to improve readability.

Appendix C. Covariance-based sectoral averaging-kernel estimates in the YRD

A sectoral averaging kernel describes the sensitivity of retrieved sector emissions to changes in true emissions. Diagonal entries represent responses within the same sector, whereas off-diagonal entries represent cross-sector responses. We estimate these responses using the covariance expression of Rodgers (2000) and Bruch et al. (2025b), with aggregation analogous to Sun et al. (2026).

The same calculation is used for the OSSE and the real-data inversion. For cycle c, the original prior at lag 1 of cycle (c -1) is paired with the final posterior at lag 0 of cycle c. They represent the same physical emission window and use identical prior emission weights in Eqs. (B7)–(B8). Synthetic true emissions

are not direct inputs to this covariance calculation.

Using the sector totals defined in Appendix B and the ensemble convention in Appendix A, the prior covariance is:

$$[\boldsymbol{P}^{b}_{E,c}]_{pq} = \frac{1}{n-1}\sum_{i=2}^{n}(\boldsymbol{E}^{b}_{c-1,l,p,i} - \boldsymbol{E}^{b}_{c-1,l,p,1}) \times (\boldsymbol{E}^{b}_{c-1,l,q,i} - \boldsymbol{E}^{b}_{c-1,l,q,1}),\ (l = 1) \qquad (C1)$$

The posterior covariance is

$$[\boldsymbol{P}^{a}_{E,c}]_{pq} = \frac{1}{n-1}\sum_{i=2}^{n}(\boldsymbol{E}^{a}_{c,l,p,i} - \boldsymbol{E}^{a}_{c,l,p,1}) \times (\boldsymbol{E}^{a}_{c,l,q,i} - \boldsymbol{E}^{a}_{c,l,q,1}),\ (l = 0) \qquad (C2)$$

Here, n = 100 denotes the ensemble size. The subscripts c, l, p, e, and i denote the assimilation cycle, assimilation lag, emission sector, spatial region element, and ensemble member, respectively. Member i = 1 represents the ensemble mean state. Applying the referenced covariance expression gives:

$$\boldsymbol{A}_c = (\boldsymbol{P}^{b}_{E,c} - \boldsymbol{P}^{a}_{E,c})\,(\boldsymbol{P}^{b}_{E,c})^{-1} = \boldsymbol{I}_6 - \boldsymbol{P}^{a}_{E,c}(\boldsymbol{P}^{b}_{E,c})^{-1}, \qquad (C3)$$

Rows refer to retrieved sectors and columns to sector-total perturbations. Under the linear Bayesian assumptions, the aggregated response interpretation follows the prior covariance structure. For localized ensemble updates, these covariance-based estimates are not assumed to equal exact retrieval derivatives.

Reported means are calculated as:

$$\bar{\boldsymbol{A}} = mean_c(\boldsymbol{A}c), \qquad (C4)$$

Table C1. Period-mean covariance-based averaging kernel estimates for YRD methane emission sectors during 21 September–30 October 2022.

Sector pair	Rice cultivation	Natural	Waste	Fuel exploitation	Livestock	Other
Rice cultivation	0.347	0.272	0.228	0.066	0.074	0.160
Natural	0.193	0.385	0.212	0.018	0.062	0.251
Waste	0.131	0.187	0.246	0.026	0.088	0.076
Fuel exploitation	0.030	0.022	0.042	0.251	0.133	0.078
Livestock	0.010	0.014	0.011	0.012	0.068	0.007
Other	0.004	0.007	0.005	0.002	0.000	0.093

Table C2. Period-mean estimated averaging kernel matrix for the six emission sectors in the YRD in the 2022 real-observation inversion.

Sector pair	Rice cultivation	Natural	Waste	Fuel exploitation	Livestock	Other

Rice cultivation	*0.303*	*0.199*	*0.176*	*0.028*	*0.112*	*0.173*
Natural	*0.164*	*0.320*	*0.123*	*0.035*	*0.068*	*0.143*
Waste	*0.087*	*0.073*	*0.417*	*0.034*	*0.045*	*0.239*
Fuel exploitation	*0.016*	*0.022*	*0.032*	*0.211*	*0.051*	*0.030*
Livestock	*0.002*	*0.001*	*0.002*	*0.003*	*0.080*	*−0.003*
Other	*0.004*	*0.005*	*0.010*	*0.002*	*0.001*	*0.097*

Appendix D. Diagnostics of prior uncertainty and model-data mismatch

The annual prior uncertainties are used to quantify the uncertainty assigned to YRD-integrated emissions by the inversion setup. They are calculated by propagating the prescribed scaling-factor covariance to sector totals and then to annual totals, retaining spatial covariances.

Following Eq. (B7), the annual prior emission of sector p is:

$$E^b_{p,yr} = \sum_c E^b_{c-1,l,p,i}, (l = 1, i = 1) \quad (D1)$$

For each time step, applying covariance propagation to Eq. (B7) gives:

$$cov(E^b_{c-1,l,p}, E^b_{c-1,l,q}) = \sum_{e\in B_p^{YRD}} \sum_{e\prime\in B_q^{YRD}} F^b_{c-1,l,p,e} F^b_{c-1,l,q,e\prime} [\boldsymbol{P}^b_{c-1,l}]_{(p,e),(q,e\prime)}, (l = 1) \quad (D2)$$

Here, c denotes the assimilation cycle and l the assimilation lag; the prior quantities are evaluated at cycle (c-1) and lag l = 1. The indices p and q denote emission sectors, while e and e' identify spatial region elements within sectors p and q, respectively. The emission weights F^b and region sets B_q^{YRD} follow Appendix B. In Eq. (D2), $\boldsymbol{P}^b_{c,l}$ denotes the prescribed prior covariance, whose finite-ensemble representation is given in Eq. (A4).

Because the annual emission is the sum over assimilation cycles and the newly generated prior states of different cycles are assumed to be independent, the cross-cycle covariance terms vanish, and the annual sector covariance reduces to the sum of the per-cycle covariances:

$$cov(E^b_{p,yr}, E^b_{q,yr}) = \sum_c cov(E^b_{c-1,l,p}, E^b_{c-1,l,q}), (l = 1) \quad (D3)$$

Setting p = q gives the annual sectoral variance. The annual total of the six optimized sectors and its variance are:

$$E^b_{total,yr} = \sum_{p=1}^{6} E^b_{p,yr}, \quad (D4)$$

$$[\sigma(E^b_{total,yr})]^2 = \sum_{p=1}^{6} \sum_{q=1}^{6} cov(E^b_{p,yr}, E^b_{q,yr}), \quad (D5)$$

Uncertainties are reported as one standard deviation (1σ). The relative annual prior uncertainty of sector p is $\frac{\sigma(E^b_{p,yr})}{|E^b_{p,yr}|} \times 100\%$.

Table D1. YRD-integrated prior CH_4 emissions and prescribed 1σ uncertainties for 2022.

Emission sector	*Annual prior emission (Tg CH_4)*	*Annual prior 1σ uncertainty (Tg CH_4)*	*Relative 1σ uncertainty (%)*
Rice cultivation	*2.940*	*0.094*	*3.21*
Natural	*1.487*	*0.086*	*5.81*
Waste	*2.138*	*0.068*	*3.17*
Fuel exploitation	*1.208*	*0.063*	*5.25*
Livestock	*0.461*	*0.015*	*3.19*
Other	*0.342*	*0.012*	*3.52*
Total of the six optimized sectors	*8.576*	*0.159*	*1.86*
Fixed flux	*-0.116*	*0.0*	*0.00*
Total	*8.460*	*0.159*	*1.88*

The consistency of the innovations with the combined prior prediction uncertainty and prescribed model–data mismatch is assessed using the innovation chi-squared diagnostic of Steiner et al. (2024, Eq. 1). Using the prior observation-space covariance from Eq. (A8), the annual diagnostic is:

The combined prior prediction uncertainty and MDM are evaluated using the normalized innovation chi-squared diagnostic of Steiner et al. (2024). Using the prior observation-space covariance from Eq. (A8), the annual diagnostic is:

$$\chi^2_{innov} = \frac{\sum_c (y_c^o - \bar{y}_c^b)^T (\boldsymbol{P}^b_{yy,c} + \boldsymbol{R}_c)^{-1} (y_c^o - \bar{y}_c^b)}{\sum_c N_{obs,c}}, \qquad (D6)$$

Here, y_c^o contains the assimilated observations, $\bar{y}_c^b$ denotes the corresponding simulated observations for the prior mean state, and $\boldsymbol{P}^b_{yy,c}$ is evaluated before assimilating these observations in cycle c. The MDM covariance $\boldsymbol{R}_c$ follows Appendix A, and $N_{obs,c}$ is the number of assimilated observations in that cycle.

Table D2. Prior residuals, prescribed MDM, prior ensemble spread, and normalized innovation diagnostics for the 2022 real-observation inversion.

Site ID	*Number of observations*	*Prior bias (ppb)*	*Prior RMSE (ppb)*	$\boldsymbol{\chi^2_{innov}}$	*Posterior RMSE (ppb)*
AMY	*2243*	*-9.54*	*61.15*	*1.67*	*53.63*
COI	*2264*	*-5.20*	*22.20*	*1.45*	*16.85*

GSN	*2043*	*15.93*	*34.12*	*0.86*	*21.43*
HAT	*2221*	*9.59*	*22.44*	*1.36*	*14.88*
NJS	*1374*	*11.65*	*135.75*	*3.30*	*85.80*
RYO	*2385*	*11.14*	*24.70*	*0.93*	*12.07*
SAR	*1888*	*39.60*	*145.40*	*0.74*	*53.49*
ULD	*2034*	*10.74*	*31.32*	*1.59*	*17.11*
YON	*2099*	*15.08*	*24.35*	*1.54*	*15.52*
Full network	*18551*	*10.32*	*66.86*	*1.42*	*37.07*

**Bias is model minus observation.*