Review of manuscript *egusphere-2026-3879*

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 performace, 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).

- **Further setup of the OSSE**:
    - 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.
    - Another missed opportunity is that neither perturbations nor the optimization of the background were included in the OSSEs.
    - 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.
    - 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.
- **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.

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).
- **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.

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

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.
- **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.

- **Page 5, Line 151**: Nudging to what? Observations or the driving data?
- **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.
- **Page 6, Line 173**: I think I know what the authors want to say with this sentence, but it is not very clearly formulated.
- **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?
- **Page 12, Line 282**: 0.1%, not 1%.
- **Page 12, Line 286**: "...estimates, combined..." → "...estimates combined,..."
- **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?
- **Figure 4**: The figure labels are too small.
- **Sect. 3.2**: A map showing simply the emissions for the YRD would be useful (similar to Fig. S3, but only for the YRD).
- **Page 29, Line 601**: Fig. 3d–f, not Fig. 4d–f.
- **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.
- **Figure S3**: For readability, the sectors should be indicated in the plot, not only in the caption.