Review of

German methane fluxes in 2021 estimated with an ensemble-enhanced scaling inversion based on the ICON–ART model (2025-1464)

Valentin Bruch, Thomas Rösch, Diego Jiménez de la Cuesta Otero, Beatrice Ellerhoff, Buhalqem Mamtimin, Niklas Becker, Anne-Marlene Blechschmidt, Jochen Förstner, and Andrea K. Kaiser-Weiss

For Atmospheric Chemistry and Physics

1 Overview

This paper describes the regional methane inversion system developed by the DWD. It covers multi-sectors of methane emissions over Europe, focusing on Germany. I am more familiar with inversion systems that try to estimate the initial conditions and the fluxes simultaneously, while this system estimates a so-called "far-field" first and followed by the fluxes, the latter of which appear to be represented as "scaling factors". Hence, I had to try hard to understand how this system works.

The system is very complex and the study is clearly very comprehensive, examining how observations can be pre-processed and filtered, how the uncertainties (R-matrix) can be estimated, how the posterior errors are examined. It studies the sensitivities of the posteriors to the truth (using the averaging kernels), and sensitivities to a very wide range of configurations of the system.

The system appears to work and is clearly very useful, but there are some weaknesses of the paper, mainly in how it is presented. I would very much like to see the work published, but I feel there needs to be some major re-writing. In particular, after reading the paper, I feel the following overall points deserve attention.

- The system description is not in the best order. The paper would benefit from a system flow chart of the steps and a more logical order of descriptions.
- There is arguably too much information given in the appendices (there are nine appendices). The work shown in some of the appendices is important, such as additional experiments and sensitivity tests. In fact I think that the paper is too long. I would suggest having a two-part paper, part I for the system description, and part II for the results. I feel that the appendices should just contain things like derivations and supplementary tables. The mixture of results between the main part of the paper and the appendices caused me to lose track of the flow of the paper.
- There appears to be some contradictory statements made in the paper. This may be just my misunderstanding, but an example is the description of how the model uncertainties are found. There seems to be one explanation in Section 2.4, another one in Sect. 4.2, and another in Sect. 5.4.2. See also detailed points 10, 20, and 27. More examples are given in my detailed comments.
- I'm not completely sure how the ensemble comes into the main inversion. I can see how an ensemble is used to help define the model uncertainties in Sect. 2.4, and in some of the supplementary statistics, but I am not sure how it is used in the main inversion (Eq. (1)), given the paper's title suggests that the ensemble is used for that. See also detailed points 8 and 10.
- I am unclear about the structure of the main control variable, s, especially as these are called scaling factors. I cannot find what these factors actually multiply in order to lead to predictions of the model's observations via Hs. See also detailed points 2, 7e, 7f, and 41a.

- I am unclear how individual flux categories can be distinguished from the observations (apart from using information in the ratio of the background error statistics, which divides the posterior fluxes according to how these are prescribed). See also detailed points 29 and 40a.

- It is not clear to me how the ICON model is used in the inversion itself, and whether any chemical processes are simulated as part of the forward model, H. See also detailed points 1, 1b, 12, and 13.

As I mentioned above, the work should be published, as there are many positives of the work, but after some major restructuring and rewriting. I give a detailed list of scientific issues in Sect. 2 and minor points in Sect. 3. Note that although there are a lot of points in my report, I believe that each can be addressed fairly quickly. I don't think any more experiments necessarily need to be run.

In my report the text of the paper is referenced by Lx (line x, where x uses the guide down each page), or by section/figure/equation/table number, and I often quote from the paper to help refer to the part that the comment refers to. Text that I suggest to be added is underlined, and text that I suggest to be removed is scored out.

2 Scientific points

1. L49-51: "Atmospheric transport is simulated using the numerical weather prediction model ICON (Zängl et al., 2015) extended with the module for Aerosol and Reactive Trace gases (ART) (Rieger et al., 2015; Schröter et al., 2018) with a spatial resolution of 6.5 km."
   (a) I'm just wondering how the ICON model is used to provide the transport winds. As well as providing tracer transport winds, is the ICON model actually simulating the winds (on-line) at the same time as advecting the methane, or is the ICON model used to pre-determine winds, which are then used by the inversion system to do the transport? The former would provide very high-temporal resolution winds (and include sub-grid-scale transport), without the need to store them, which would be ideal, but very expensive. If ICON is simulating the winds on-line, then it must be run with a data assimilation scheme to keep them realistic. I think the authors could say a little about how this is done (although see also point 12 below).
   (b) Are any chemical processes (such as the reaction with OH and methane oxidation included? L398-399: "Another potential contribution to the seasonal cycle could arise from neglecting the OH sink of CH_4 in our limited domain." suggests not. Could this be an issue over the timescale of the inversions? The authors should at least say that this is an assumption made by their system

2. L72-74: "The categorized fluxes are scaled to minimize the mismatch between model prediction and observed concentrations. Thus, the inversion result consists of one scaling factor for each flux category. The a priori fluxes multiplied by the scaling factors yield the a posteriori fluxes." By "... one scaling factor for each flux category.", do the authors mean one for all positions and time, or one per flux category per position and per time (obviously with relevant correlation scales)?

3. L91-92: "Each ensemble member uses slightly different but equally likely parametrizations and meteorological initial and boundary conditions." Does each member use slightly different driving winds too?

4. Sect. 2.2 and Appendix B: I'm not really sure what is meant by the far field. I assume it's a smooth correction to the methane field, determined before the inversion, and is a function of space and time. Is that right? Looking at Eqs (1) and (B1), it looks like it exists only at observation locations. See also point 35 below.

5. L100-102: "To minimize potential biases arising from imperfect boundary conditions, we construct a correction field which is added to the modeled far-field concentration in the whole domain after the transport simulation." The terms "far-field" and "whole domain" sound contradictory to me.

6. L102-103: "We require this correction field to be smooth on large length and time scales, chosen in our case as ..." This sounds like a tautology to me: surely "smooth" means "large scale"?

7. Sect. 2.3:

   (a) How is B determined as used in Eq. (1)?

   (b) What are the observations of? Appendix A lists the stations, but I cannot find anywhere whether the observations are of methane or of the flux itself. What are the instrument types and instrument precisions? This would be good to know even though they are claimed to be negligible to the model error (L125). I think it should be stated as early as possible in the paper that the observations are of methane.

   (c) What is the time window and how is information propagated from one time window to the next?

   (d) As the initial conditions of the methane field are not mentioned in the cost function, Eq. (1), it looks like the initial conditions are not adjusted as part of the inversion. Is that right? Perhaps this is the purpose of the far-field correction step? If so, please make this clear. If the initial conditions are sufficiently wrong, then this would have an impact on the quality of the inversion.

   (e) Given that the control variable in Eq. (1) is s (a scaling factor), this suggests that it should multiply something. What is the field that it multiplies? See also point 41a below.

   (f) What is the structure of the field s? Is it represented on a grid, or does it multiply some basis functions? Field s must, presumably, be associated with the different categories of source field (a number >40 is mentioned in D2, e.g.).

8. Eq. (1): Given that the title of the paper talks about an ensemble, is Eq. (1) minimised with respect to each ensemble member? Or is the ensemble just used for the procedure described in Sect. 2.4 (see point 10 below)?

9. L111-112: "In the first term, the vector y of observed concentrations is compared to the model prediction, which consists of the transported fluxes Hs and the modeled far field x^{ff}."

   (a) How can fluxes be transported? Do the authors mean "transported methane emitted by the fluxes"?

   (b) In Eq. (1) and in the above quote, x^{ff} is the "far field". Is this the same as field c in Appendix B? If so, I would recommend that the same symbol is used in all parts of the paper. If not, I would recommend that they are not referred to as the "far field".

10. Sect. 2.4: The R' matrix in Eq. (2) is determined from an ensemble.

    (a) Is this the same ensemble that is (possibly) used in the main inversion (but see point 8 above)? How is it initialised?

    (b) L132: "... added to each observation accounting for any representativity error." I think representativity error is something different, unless sub-grid-scale processes are included in the divergence of the ensemble members.

11. L163-168: "Plumes caused by high emissions in a small area require special treatment to avoid a potential bias in the inversion due to the so-called double penalty issue (Vanderbecken et al., 2023). In cases where our model falsely predicts that the plume reaches an observation site, the inversion will reduce the emissions to improve the agreement with the observation. In the opposite case, when the model fails to predict that a plume reaches the observation, the inversion will not change the plume emission amount but will wrongly increase emissions in other areas instead. This can cause systematic underestimation of fluxes from localized plumes." I very much like this interpretation.

12. L177-178: "The meteorological initial and lateral boundary conditions used to drive our transport model are taken from the archive of DWD's operational numerical weather prediction (NWP), which also employs the ICON model." This statement partially addresses point 1 above, but I am still unsure whether it is ICON itself – or another model that just uses ICON-derived winds – that is used for the H operator in Eq. (1).

13. L184-185: "In contrast to the meteorological fields, the CH_4 concentrations are only transported and never re-initialized." Is it just transportation, or are chemical reactions included too?

14. L188: "We ensured mass conservation when interpolating to our model grid." How is this done? Is it by multiplying the interpolated fluxes by a factor to ensure that the total flux is the same after interpolation?

15. L189-190: "Anthropogenic fluxes excluding LULUCF are split further into 12 GNFR sectors (gridded aggregated NFR, nomenclature for reporting, Veldeman et al. (2013)), ..."
    (a) GNFR is not defined (I assume GNFR = Gridded NFR, but I don't know what NFR stands for).
    (b) I don't really understand the text in the brackets.

16. L197: "These emissions are missing in our a priori estimate." Does this represent a low bias in the a-priori?

17. L219-221: "Observations within the planetary boundary layer are most representative in the afternoon hours whereas measurements at high mountains have less local influence at night time (Bergamaschi et al., 2015). We therefore use the time windows 23 h to 5 h (local mean time) for stations on high mountains and 11 h to 17 h for all other stations." If observations on high mountains have less influence over night, why only use them during 23 h to 5 h?

18. L227-229: "In the last filtering step – step 5 in Table 2 – we exclude data points with extreme mismatch between model and observations of more than 200 ppb. Data points where the observations are more than 20 ppb below the model-predicted far field are also discarded."
    (a) I assume by "model", the authors are referring to the a-priori?
    (b) For the first (extreme) condition is this written mathematically as $\left|y - Hs - x^{\mathrm{ff}}\right| >$ 200ppb or $|y - Hs| >$ 200ppb to discard? Perhaps this is better explained as an explicit inequality?
    (c) For the second condition is this, mathematically $y - x^{\mathrm{ff}} < -20$ppb to discard? Again, perhaps this is better explained as an explicit inequality?

19. Section 4.1.2, point (ii):
    (a) By advecting a tracer, how is it possible to set a lifetime for that tracer? One would know only the tracer concentration in any grid box, not how long it has been since that tracer was released (and anyway is likely to be comprised of tracers that have a variety of emission times).

(b) The statement that tracers have a lifetime and that "no CH_4 is lost" seem contradictory to me.

20. Section 4.2: I'm very confused as there seems to be two prescriptions for how the R-matrix is determined – this section and Sect. 2.4. See also point 27 below.

21. Section 4.3: Is any cycling done between the monthly inversions? That is, does the posterior of one month become the prior of the next?

22. Section 4.4:

    (a) L316: "In each inversion time window, we consider uncorrelated a priori scaling factors ..." and L319-320: "... , and within Germany categories describing the same sector have an a priori uncertainty correlation of 50 %". These two statements seem contradictory.

    (b) L316-320: "In each inversion time window, we consider uncorrelated a priori scaling factors with a two standard deviation (2σ) uncertainty of 80 % for most flux categories, corresponding to a 95 % confidence interval of ±80 %. Throughout this paper, uncertainties will denote two standard deviations or 95 % confidence intervals. Categories resolving emission sectors have a higher prior 2σ uncertainty of ±100 %, and within Germany categories describing the same sector have an a priori uncertainty correlation of 50 % (e.g., ...)" There are (presumably) two quantities represented by percentages here, the confidence intervals (which represent percentage of the PDF volume), and the methane quantity itself (presumably represented as a percentage of some value). If this is correct could there be a better way to describe these to avoid confusion?

23. L333-345: "Additionally, we combine the two variants of inversion (prior-R and posterior-R, see Sect. 2.4.2) by taking the arithmetic mean of the two separate inversion results, arriving at the combined scaling factors." The posterior error statistics should just be a function of the inverse Hessian of the inversion. Why the need to refer to the R-matrices? A similar mention of the R-matrices is made in L4-5 of the caption of Fig. 5.

24. L348-349: "Figure 4 presents an overview of (a) the a priori CH4 fluxes, (c) the resulting scaling factors, and the respective uncertainties (b, d), all accumulated over the year 2021." Presumably "averaged" rather than "accumulated".

25. L416-417: "Other filtering parameters such as the number of sampling heights per station (case 202) and ..." I don't understand what is meant by, "the number of sampling heights per station", given that the height of an observation station is (presumably) fixed.

26. L418-420: "Neglecting extreme outliers has only a small effect (cases 206, 207), but limiting the influence of outliers by increasing their uncertainty has a considerable impact (cases 208, 209)." One would expect that neglecting extreme outliers is just the limiting case of increasing their uncertainty to infinity. It might be counter-intuitive then that increasing their uncertainties has a large impact, then a small impact when their uncertainties are increased further to infinity.

27. L428-429: "... and the uncorrelated additive uncertainty σ_{const} of each data point (cases 309, 310)." Perhaps I have missed it, but I cannot find any reference to σ_{const} in Sects. 2.4 or 4.2. This adds to my confusion about how the R-matrix is determined. See also point 20 above.

28. L471: "~~These~~ Each vector of scaling factors defines ~~the~~ a synthetic truth, ..." (These suggested changes distinguishes considering all scaling factors together in one ensemble-based inversion.)

29. L484-485: "Within Germany, we distinguish agriculture from other emissions. The ability to distinguish sectors can be described by averaging kernel matrices which estimate ..." Putting the B-matrix aside, how can different sectors be distinguished from observations of methane? See also point 40a below.

30. Section 5.6: This section is about simulated transport errors, but I cannot find a description of how transport errors are simulated. Is noise added to the winds e.g.?

31. Figure 10: I'd recommend adding the 1:1 line as a guide. See also point 39c below.

32. L515-516: "Localized sources that cause a strong plume are underestimated by both methods, though the bias is reduced in the posterior-R inversion as predicted ..." I assume it is the posterior fluxes that are plotted in Fig. 10, using the two R-inversion methods, with and without the far-field correction (and not their errors). Figure 10 (and Fig. F2, see point 39c below) therefore compares the posterior values for each relative to the prior. How is it possible therefore to tell whether a strong plume is underestimated and what the biases are?

33. L537: "Firstly, we find that our top-down CH_4 emission estimates are significantly higher than reported for Germany." Although, looking at Fig. 5, the NIR estimate is well within the 95% confidence interval of the posterior.

34. L619-620: "The increasing availability of satellite data is especially interesting for constraining concentrations and emissions in less observed regions, such as near the boundaries of our domain." This seems an obvious extension of the work given the wide range of total column methane retrievals available from satellites. This may require careful tuning though to remove potential biases.

35. Appendix B, Eq.(B2): I am concerned that the formula (B2) is not formally correct in one detail (in the final step). Let's go through the maths to demonstrate

$$\text{penalty}(c) = (x+c-y)^\intercal \, P^\intercal (P\tilde{R}P^\intercal)^{-1} P\,(x+c-y) + c^\intercal P^\intercal (P\tilde{C}P^\intercal)^{-1} Pc.$$

(Incidentally, all other cost functions in the paper have a factor of 1/2 except this one. This does not make any difference to the outcome, but the authors might wish to be consistent by including the 1/2 factor here.) The gradient is

$$\nabla\text{penalty}(c) = 2P^\intercal (P\tilde{R}P^\intercal)^{-1} P\,(x+c-y) + 2P^\intercal (P\tilde{C}P^\intercal)^{-1} Pc = 0 \text{ for min.}$$

This agrees with (B3) of the paper. This can be rearranged:

$$\begin{aligned}
\left(P^\intercal (P\tilde{R}P^\intercal)^{-1} P + P^\intercal (P\tilde{C}P^\intercal)^{-1} P\right) c &= P^\intercal (P\tilde{R}P^\intercal)^{-1} P\,(y-x) \\
P^\intercal \left((P\tilde{R}P^\intercal)^{-1} + (P\tilde{C}P^\intercal)^{-1}\right) Pc &= P^\intercal (P\tilde{R}P^\intercal)^{-1} P\,(y-x)
\end{aligned}$$

We cannot simply cancel $P^\intercal$ from the left, but because $PP^\intercal$ is full rank we can multiply from the left with P and then cancel $PP^\intercal$. Doing this and then developing:

$$\begin{aligned}
\cancel{PP^\intercal}\left((P\tilde{R}P^\intercal)^{-1} + (P\tilde{C}P^\intercal)^{-1}\right) Pc &= \cancel{PP^\intercal}(P\tilde{R}P^\intercal)^{-1} P\,(y-x) \\
(P\tilde{R}P^\intercal)\left((P\tilde{R}P^\intercal)^{-1} + (P\tilde{C}P^\intercal)^{-1}\right) Pc &= P\,(y-x) \\
\left(I + (P\tilde{R}P^\intercal)(P\tilde{C}P^\intercal)^{-1}\right) Pc &= P\,(y-x) \\
\left[P\tilde{C}P^\intercal + P\tilde{R}P^\intercal\right](P\tilde{C}P^\intercal)^{-1} Pc &= P\,(y-x) \\
P(\tilde{C}+\tilde{R})P^\intercal (P\tilde{C}P^\intercal)^{-1} Pc &= P\,(y-x) \\
Pc &= P\tilde{C}P^\intercal \left[P[\tilde{C}+\tilde{R})P^\intercal\right]^{-1} P\,(y-x)\,.
\end{aligned}$$

If we could eliminate P from the left of each side then we would get (B2). We do the trick similar to above: first multiply by $P^\intercal$:

$$P^\intercal P c = P^\intercal P \tilde{C} P^\intercal \left[P[\tilde{C} + \tilde{R}) P^\intercal \right]^{-1} P\,(y - x)\,.$$

Unfortunately, $P^\intercal P$ is not full rank so we cannot obtain a unique equation for c (but we can for Pc). This makes sense because the constraint described by (B1) applies only to those observation points selected by P. Since there is no a-priori constraint (to allow information to be available to not-selected observations), the not-selected observations will have no result from the above analysis. I think this requires some attention in the paper.

36. Appendix C: I didn't follow the reasoning in this appendix. For example, I note the following points.

    (a) The left hand side of (C1) seems strange to me. It is the probability of a real number (the mismatch) being realised. This doesn't make any sense to me. Normally a probability *density* would be used, e.g. $P(\mu = y - Hs^{\text{prior}} - x^{\text{ff}})d\mu$ to mean the probability between values of μ and $\mu + d\mu$.

    (b) What is the variable dP_s and how does one arrive at (C1)?

    (c) The line in (C2) appears to integrate over the posterior. How does this relate to (C1)?

    (d) The term $H\tau + \mu$, which appears in (C2) has the following form, given that $\tau = s - s^{\text{prior}}$ and $\mu = y - Hs^{\text{prior}} - x^{\text{ff}}$:

    $$H\tau + \mu = Hs - Hs^{\text{prior}} + y - Hs^{\text{prior}} - x^{\text{ff}} = Hs - 2Hs^{\text{prior}} + y - x^{\text{ff}}.$$

    This doesn't seem correct to me.

    (e) The rest of the section does contain some useful analysis about the expected value and range of χ^2/N_{dof}, but the appendix should be explained more and a reference added.

37. Appendix D

    (a) L692: "We estimate the model uncertainty using a meteorological ensemble." Which model uncertainty and which quantity? I assume the authors are referring to the uncertainty in s?

    (b) L693: "Stronger emissions lead to stronger spatial gradients in the model concentrations ..." (suggested change to make distinct from gradients in state space).

    (c) L712-715: "When using a priori scaling factors to estimate the model uncertainty, we need only the total concentration $x_i^m(s^{\text{prior}})$ for each ensemble member. Thus, only a single tracer field is required in the ensemble transport simulation. To fully compute $x_i^m(s)$ as function of s, the tracer categories need to be distinguished for each ensemble member, resulting in >40 tracer fields in the ensemble simulation." The only difference between the calculation of $x_i^m(s^{\text{prior}})$ and $x_i^m(s)$ is that s^{prior} is replaced by s. How does this require that >40 different tracer fields are required? I guess that >40 different values of s are chosen. What is the significance of the number 40? The difficulty to understand this may be connected to point 7f above.

    (d) In the above, what does i stand for?

    (e) I don't really understand Eq. (D3). What is the distinction between $x_i^{mg} = x_i^m(P_g s^{\text{prior}})$ and $HP_g s^{\text{prior}}$ (they seem to be the same thing)? Making this distinction is probably the key to understanding (D3).

38. Table E1

(a) ID 101: "no time averaging", given the explanation column, shouldn't this read "change in time averaging"?

(b) Some of the test cases (rows in table) are ambiguous to me. For example, "200, fewer hours of day, use time window 12 h–16 h (0 h–4 h for high mountains)". Does it mean that the reference case used observations at all hours? Then again "201, all hours of day, no filtering by time of day, increase uncertainty inflation by factor 1.5" seems to imply that using observation at all hours is a test case different from the reference. I would recommend going through the entire list and making sure that each test case is unambiguous to the reader.

39. Appendix F

(a) Figure F1: I am confused why the truth (horizontal lines) should change with the test ID.

(b) L755-757: "Next, we test the effect of an underestimation or overestimation of all emissions. In case 20 of Fig. F1, all natural and LULUCF fluxes are reduced by 40 % in the truth, and cases 21 and 22 change all anthropogenic emissions excluding LULUCF by −20 % and +20 %, respectively." Are these experiments simply changing the true emissions (and the synthetic observations and a-priori values) and then repeating the inversion? I would've thought that such an experiment would be expected to have posterior fluxes that are consistent with the truth. Wouldn't a more interesting experiment be one with the underestimation or overestimation of all emissions in the a-priori, but with the truth (and synthetic observations) unchanged?

(c) Fig. F2, and text: It would be useful to draw the 1:1 line in each panel (see also point 31 above). It would then be seen that the when far-field correction is applied the posterior R-inversion does better than the prior R-inversion. I think it is best to describe fully the experiments being done and then show the results in the Figs. In the case of the text around Fig. F2 it is not until the end of the appendix that the reader learns that the "prior is underestimated compared to the synthetic truth."

40. Appendix G

(a) L765-766: "When observations can detect a change in total emissions but cannot distinguish between different emission sectors, the sector-resolving inversion will change the sectoral distribution based on the prior uncertainties." Out of curiosity, how could observations of methane make this distinction? Also mention is made in the main text at L365.

(b) L768: "The a priori probability density for an emission ..."

(c) L771: The symbol $P(e)$ is used to represent the posteriori probability density, but the same symbol is used for the prior density in Eq. (G1). Although this is often done when describing probabilities, one can distinguish the prior from the posterior by its argument. The prior is often written as $P(e)$ and posterior as the conditional density $P(e|y)$. The statement on L770-771 needs to refer to the posterior.

41. Appendix H

(a) L786-787: "We first estimate the sensitivity of the posterior scaling factor to the true emissions under the assumption that the transport model, far field, and the flux pattern within each flux category are perfect." Flux patterns are referred to, but not defined anywhere. I suspect addressing this point will help the reader understand point 7e above also.

(b) L788: $\mu = y - Hs^{\text{prior}} - x^{\text{ff}}$. Let $y = Hs^{\text{truth}} + x^{\text{ff}} + \epsilon^y$ (where ϵ^y is the observation error). Then, $\mu = H(s^{\text{prior}} - s^{\text{truth}}) + \epsilon^y$. The expression given in the paper doesn't have ϵ^y, which suggests that the authors are making the additional assumption that the observations are perfect.

(c) Eq. (H1): Given that the above expression is used to substitute for $y - Hs - x^{\text{ff}}$ (and not $y - Hs^{\text{prior}} - x^{\text{ff}}$ as used in point 41b above), why not simply present the previous expression with $s^{\text{prior}} \to s$?

42. Appendix I

    (a) Fig. 11: There is no explanation of the distinction between the two panels in the caption. The only difference is (a) is labeled $A^{\text{emissions}}$ and (b) is labeled $A^{\text{scaling factors}}$. I assume that (a) is computed from (b)? What are the smaller matrices at the bottom of each panel?

3 Presentational points and very minor points

1. The appendices are referenced in a different order to their placement. For example, appendix B is referenced first. It makes logical sense to place the appendices in the order that they are first referenced in the main text.
2. Table 1: As there are many acronyms, the caption of Table 1 might be a good place to (re)define acronyms. Incidentally, I cannot see where GNFR is defined anywhere.
3. L201: reference to Fig. 3, I would recommend moving Fig. 3 to Fig. 1 as it referenced first. Also in Fig. 3: "A white ~~ellipsis~~ ellipse marks the Upper Silesian Coal Basin, ..."
4. Sections 4.1.1 and 4.1.2: I got very lost trying to follow exactly what was done here. I would recommend that these sections are rewritten, possibly with a table to help the reader see exactly how different regions and sectors are combined, etc.
5. L466, suggested change: "We therefore use experiments with ~~a "synthetic"~~ synthetic data, i.e., define truth and ..."
6. L110: "... ~~penalized~~ penalizes ..."
7. Table E1 caption: "Overall, we see that most tests have an impact ..."
8. L194: "For Germany, the a priori LULUCF fluxes obtained from the Thünen Institute cover ..."
9. L207: "... in model coordinates ~~is~~ are derived from the station sampling heights and the modeled station elevations, depending on the ..."
10. L214: "We start by ~~smoothening~~ smoothing both observations and ..."
11. L372: "In France, Belgium, and Switzerland, the inversion scales flux categories that overlap~~ping~~ multiple countries[2] ."
12. Table E1, caption, L4: "Overall, we see that most tests have an impact ..."
13. L568-569: "Though simulating emissions and transport in a large domain, we can only provide reliable emission estimates for selected countries (compare Fig. 5)." I think there is a slight problem with the beginning of the sentence. Possibly the authors mean either, "Through simulating emissions and transport in a large domain, we can only provide reliable emission estimates for selected countries (compare Fig. 5)." or, "~~Though~~ Although we are simulating emissions and transport in a large domain, we can only provide reliable emission estimates for selected countries (compare Fig. 5)."
14. L579: "... is more likely to increase the far field rather than ~~decreasing~~ to decrease it, leading to a bias towards ..."
15. L588: "Moreover, adjusting fewer degrees of freedom may ..."