Author's response #2

Regional and global atmospheric observation-based estimates of SF6 emissions are presented using a Lagrangian inverse modelling system. Emissions trends are derived for the major emitting regions, China, the USA and EU, which are generally consistent with other available regional studies, but mostly higher than reported emissions. Global emissions are also broadly consistent with previous studies.

The article is detailed and meticulous and very well written. The methods are interesting and novel, and the application is important and timely. I think the paper is suitable for publication in Atmospheric Chemistry and Physics, subject to some minor corrections.

We would like to thank reviewer #2 for the detailed and very productive review of our manuscript.

In the response we use 4 different colors. The blue colored text is the general answer to the reviewer's comments. Additionally, we show how the text is changed in the manuscript: The original text is colored grey, removed text is colored red, and new text is colored green.

Main text:

L15: Point 5 in this list is somewhat confusingly worded. Perhaps something like: "Global total SF6 emissions are comparable to previous studies but are sensitive to a priori estimates, because of the poor network sensitivity to some regions (e.g., Africa, South America)""

We changed this, following largely the reviewer's suggestion:

The global total SF6 emissions are captured well by the inversion, however, results are sensitive to the a priori emission estimates, given that substantial biases of these estimates in regions poorly covered by the measurement network (e.g. Africa, South America) can be improved but not entirely corrected. -> Global total SF_6 emissions are comparable to estimates in previous studies but are sensitive to a priori estimates, due to the low network sensitivity in poorly monitored regions.

L22 and L27-29: I suggest deleting the lines beginning "However, this GWP 100 value…" and "Thus, GWPs, which are typically…". I don't think it's accurate to say that the GWP 100 value "underplays the climate impact of this gas". If you wanted to examine the climate impact over longer timescales, you could define a longer-term GWP. It's well known that GWP has several flaws, but I don't think you need to go into them here.

Yes -> done

L43: I'd separate out the part of this sentence on SF6 measurements being used to determine stratospheric OH into its own sentence. The other parts of this list are sources, whereas this is a measurement of atmospheric SF6. You could also add that it has been used as an ocean tracer though.

done -> changed it to:

Furthermore, SF6 finds applications in semiconductor manufacturing, facilitating precise etching processes (Lee et al., 2004) and serves for blanketing or degassing in the magnesium or aluminum metal industry (Maiss and Brenninkmeijer, 1998). Moreover, it is used in medicine (Lee et al., 2017; Brinton and Wilkinson, 2009), photovoltaic manufacturing (Andersen et al., 2014), military applications (Koch, 2004), particle accelerators (Lichter et al., 2023), soundproof glazing (Schwarz, 2005), sports shoes (Pedersen, 2000), car tyres Schwaab, 2000), wind turbines (EPA, 2023) and as a tracer gas in the atmosphere (Martin et al., 2011), in groundwater (Okofo et al., 2022), rivers (Ho et al., 2002), and oceans (Tanhua et al., 2004).

L44 and 54: I suggest removing "developed" and "developing". These terms are not needed here.

done

L70: This statement isn't true, as Rigby et al., 2011 was a global inverse modelling study that used a 3D (Eulerian) model.

Yes, we agree and changed this to:

Up to this point, SF6 inversion studies have exclusively been focusing on specific geographical areas, i.e., using regional inversions only. Although global observation-based box models, such as the AGAGE 12-box model (e.g., Rigby et al., 2013) are considered to be capable of accurately determining the global total emissions, a comprehensive top-down perspective of the global SF6 emission distribution is missing. -> Although global SF6 emissions can be well constrained by global box models, such as the AGAGE 12-box model (e.g., Rigby et al., 2013), and regional inversion systems have been used to estimate SF6 emissions in specific regions, there is no clear link between regional and global emissions and an updated, comprehensive top-down perspective of the global SF6 emission distribution is missing.

L113: Measurement location and time?

Yes - done

L115 – 118 and throughout the following sections: I think you need to be careful with the notation here. In this section, where you define He, e, etc. it implies that these sensitivities are to the grid-scale emissions or mole fraction fields. However, you've used a basis function decomposition of the emissions field in your inversion (but I'm not sure how you're scaling your initial conditions field, see below). Therefore, the matrices and vectors in Equations 2 and 3 are different to those defined here. I think you could make this consistent by stating that e, He, etc are for aggregated groups of grid cells when you define them?

Yes, thank you, that is true!

We added: Note at this point, that we aggregate grid cells of the emission grid for the optimization (see Sec.2.5) and that the just-defined variables (He, e, Hi, yi) refer to aggregated groups of grid cells. For a detailed description please see (Thompson and Stohl (2014).

Figure 2: How have you dealt with the different frequency between the flask and high-frequency data here? Is this the average over all time points, with zeros during times where there are no flask data?

The average is the sum of all sensitivity fields (representing the sensitivity to one observation respectively) divided by the total number of sensitivity fields. Thus, there is a weighting of the sensitivity of different measurement sites according to the measurement frequency. To clarify this, we have added to the figure caption: "Notice that values represent averages over all cases, for which FLEXPART calculations were made. Thus, sites with high-frequency on-line observations are weighted more strongly than sites where only flask measurements are made, or observations from moving platforms."

Section 2.5: Please clarify:

- if emissions and boundary conditions are being *scaled* in the inversion, or if absolute values are being derived. Furthermore, how are grid cells aggregated within the spatial basis functions? Is the spatial pattern of the underlying grid cells preserved, or are emissions spread out uniformly within the aggregated cells?

  Absolute values of emissions are derived, while the boundary conditions are scaled. Emissions in the fine grid are weighted by the ratio of the area of the fine grid to the variable grid, into which it is aggregated. After the inversion, optimized emissions in the variable-resolution coarse grid were redistributed onto the fine grid according to the relative distribution of the a priori emissions.

  We added: "Emissions in the fine grid are thereby weighted according to the ratio of the area of the fine grid to the variable-resolution coarse grid into which it is aggregated. After the inversion, optimized emissions in the variable grid were redistributed onto the fine grid according to the relative distribution of the a priori emissions."

- how the initial conditions are being adjusted. Is the whole field adjusted each month (or, equivalently, are the baseline mole fractions at the stations being adjusted uniformly? Or perhaps adjusted on a per-station basis?), or is there some spatial decomposition?

Yes, the whole field is adjusted every month.

We added: ... , where the whole field is adjusted on a monthly basis.

- Does R contain only "observational errors", as stated? If so, how is this defined (i.e., is it just measurement repeatability)? And if this is the case, what about model (or mismatch) uncertainty? How have you accounted for this critical (but highly uncertain) term? It seems that this term should also be the subject of a sensitivity test.

Yes, we accounted for the model uncertainty.

We added: "FLEXINVERT+ assumes a diagonal observation error covariance matrix R, and therefore, does not account for possible error correlations between different observations. The diagonal elements represent the sum of measurement and model error, where we assume the latter to be dominant. Our error estimates are based on a number of initial inversion runs, where we assessed the model error according to the a posteriori model residuals (difference between observed and a posteriori simulated mole fractions), and such that the reduced chi-square value (the value of the cost function at minimum divided by the number of observations and divided by 2) is close to 1."

- Have the observational data been filtered at all? For example, excluding points under low boundary layer heights, or at night, as is often done due to poorer model performance under these conditions? Furthermore, note that SF6 mole fractions in populated regions show occasional very large events, perhaps linked to equipment failure (see, for example, the note that very large emissions are derived during some months, here: https://assets.publishing.service.gov.uk/media/62d7b9bee90e071e7e59c97e/verification-uk-greenhouse-gas-emissions-using-atmospheric-observations-annual-report-2021.pdf). Do these need to be excluded, since your emissions model assumes constant fluxes (at least during each month)?

Yes, we excluded occasional very large events.

We added: "In addition, we adopted a method by Stohl et al. (2009) to identify observations that cannot be brought into agreement with modeled mole fractions by the inversion, which we removed entirely (in contrast to Stohl et al., 2009, who assigned larger uncertainties to these observations). For this, we utilized the kurtosis of the a posteriori error frequency distribution and iteratively excluded observations causing the largest absolute errors until the kurtosis of the remaining error values fell below 5, approximating a Gaussian distribution)."

Otherwise, we did no filtering. See comment to reviewer 1.

- How was the baseline uncertainty of 0.15 ppt, and correlation length scales, arrived at? Why 70% for the prior uncertainty?

These values are based on sensitivity tests and values previously used in the literature, but are of course debatable. The emission data, unfortunately, do not contain uncertainty information, such that any value used is ambiguous and requires subjective judgement. The results are not very sensitive to the choice of the baseline uncertainty.

- I don't understand why a 70% level of prior uncertainty on a per-grid cell basis doesn't lead to a vanishingly small prior global uncertainty. Can you clarify? If you have ~5000 grid cells, wouldn't the global uncertainty be ~70% / sqrt(5000), which is ~1% (notwithstanding spatial correlations and minimum values).

The reviewer is right. We revised the uncertainties, which for global emissions indeed become very small. We therefore base our uncertainty estimates also on the differences obtained when using different a priori inventories.

- Surely the temporal correlation of 90 days plays very little role, given that you are solving for annual emissions in the main results? Is this term needed?

Yes, you are right. The term plays little role and is probably not needed in this case, even though it helps to regularize the problem.

L253 – L256: I would remove these statements (or at least the sentence on L256), as it suggests the inversion has more capacity to focus on "incorrect" parts of the model than it really has. It is of course better if the prior model baseline is better, but the optimization is of the whole system. Even if the prior model simulated a perfect baseline, errors in sensitivities to boundary conditions or footprints could still lead to an adjustment away from that perfect baseline.

We have removed the sentence: "This is important, as the optimization can focus on improving the emissions rather than correcting a wrong baseline". We left the rest of the text, since we think it is relevant to point towards the importance of a well-fitting baseline.

L261 – 265: I think these lines should be removed. I don't doubt that a 50-day simulation period is more "accurate" than a 10-day period. But it's not shown here.

We removed:

"Figure 5b further illustrates the advantage of choosing a rather long 50-day backward simulation period. With this long simulation period, we can see that this remote station is also directly influenced by emissions (i.e., enhancements over the baseline) that can be directly optimized. With shorter simulation times (e.g., 5-10 days), no emission contributions above the baseline could be seen, thus rendering this station useless for emission optimization. For a detailed discussion about the LPDM backward simulation period see Vojta et al. 2022."

L289 and throughout this section. Please provide an uncertainty to these quantities.

Yes -> done

L298 – 299: Remove the sentence about it being "reassuring". This is subjective and not needed.

Yes -> done

Section 3.3.4: My reading of all of these subsections is basically that there is, not surprisingly, very sensitivity to these regions. I suggest moving this content to the Supplement and summarizing this message in a paragraph or two in the main paper.

Thank you for the suggestion, but in this case, we would like to keep it as it is. Firstly, we would like to illustrate the big differences of the different a priori inventories in those regions, which are rarely discussed elsewhere. Also, (without overestimating the inversion's capability in those regions), we still think it is interesting that the results in many cases at least indicate a positive trend (even if very uncertain) and that the inversion derives smaller posterior emissions for the UNFCCC-ELE inventory compared to the prior which we suspect to overestimate emission in those regions (due to the overestimation of the global a posteriori emissions).

L434: Should this be "is larger than, and inconsistent with, the global atmospheric SF6 growth…". Furthermore, I wouldn't use "postulated" in this sentence (use "derived" or similar).

Yes -> done

L461: "could be brought relatively close to these previous estimates", rather than "known values" (there are no "known values").

Yes -> done

L462: Suggest deleting “which has rarely been achieved before”, as it’s too broad here. There are many studies using global Eulerian models that do this (although only one for SF6 that I’m aware of; i.e., Rigby et al., 2011).

Yes -> done

L462 – 468: I don’t agree with the framing of these sentences. The novelty of this work is that it attempts to create a global picture using a backward running Lagrangian model. This is very nice in itself. But we shouldn’t get carried away that 50-day back trajectories can really give us a full global picture, given the sparse measurement network. As this work shows, there is negligible sensitivity to large parts of the world, irrespective of the integration time. Without additional measurements, emissions derived from these regions will always be subject to biases from the prior and the accumulation of transport errors. Furthermore, the last part of these sentences is conjecture, that there is a “clear direction” in the adjustments to these unsampled regions. This seems to be subjective to me. I suggest cutting these sentences. The work is impressive in itself. You don’t need to over-sell it.

Agreed.

-> We deleted: We attribute this capability of simultaneously constraining both regional as well as global emissions mostly to our long backward calculation period of 50 days (Vojta et al. 2022) and our extensive observation data set.

-> and : Nevertheless, in most cases, the regional results at least indicate a clear direction in which \textit{a priori} emissions need to be corrected even for these poorly monitored regions.

Section 3.3.6: Note that seasonal emissions were also briefly noted for north-east Europe in Reddington et al. (2019). Similarly to Hu et al, these maximized in the winter.

Thank you for this reference, which we have added to the paper. Indeed, they mention a winter maximum and their Figure 111 seems to suggest it; however, without showing a very systematic seasonal cycle.

We rewrote:
While there is no clear seasonal cycle in the EU emissions, the Chinese seasonality is similar to the one in the Northern Hemisphere (Fig. 13b) ->

For EU emissions no clear seasonal cycle can be seen. Notice at this point that SF$_6$ emissions from North-West Europe were found to maximize in the winter (Redington et al., 2019) however without showing a very systematic seasonal cycle. For Chinese SF$_6$ emissions, the seasonality is similar to the one in the Northern Hemisphere (Fig. 13b).

L499: I suggest “boundary conditions”, rather than “initial conditions”

Yes -> done

L502 – 503: I suggest deleting the final sentence for the reasons outlined above (comment to L253)

Yes -> done

L504 – 505: I also suggest deleting the final sentence of this bullet for the reasons outlined above (comment on L462)

Yes -> done

L509 and throughout this section: provide uncertainties

Yes -> done

L517: Delete the final sentence, as this is conjecture.

We changed the sentence to:

This might suggest that the EU's new F-gas regulation was almost immediately successful in reducing SF6 emissions.

L527: Delete the two final sentences, as I don't see how you could know this. It's not supported by your investigation.

Our results showed that, when using the EDGAR prior emissions, the global a posteriori emissions showed the best agreement to the total global reference values. We therefore think that the aggregated prior estimated emission in poorly covered regions (residual between global emissions and emissions in well-monitored areas), should be a relatively good estimate, as otherwise we would expect larger biases in the global emissions.

We changed the two sentences to: The EDGAR bottom-up inventory seems to provide a relatively good estimate for the total emissions aggregated over all the poorly monitored regions (residual between global emissions and emissions in well-monitored areas), as otherwise, the global a posteriori emissions would be more strongly biased against the relatively well known global emissions based on atmospheric growth rates. Nevertheless, more observations are needed to investigate if also regional emission patterns in those areas are accurate.

L529 – 530: I think this bullet should be deleted, as there's so little sensitivity to this region.

We rephrased the bullet point:

Our inversions suggest globally significant and strongly increasing emissions in India since 2005. However, the results for this region are very uncertain because of a weak observational constraint. Adding monitoring capacity in this region should be a high future priority.

L542: Delete the final bullet, as it's well outside the scope of your work.

Yes -> done

Supplement:

L14: full stop needed.

Yes -> done

L44: Please confirm that the following is correct and has been checked in your analysis: The cited paper (Guillevic et al., 2018) quotes the ratio NOAA-2014 / SIO-05 = 1.002 ± 0.002. However, the wording on this line suggests that conversion from NOAA-14 to SIO-05 is by multiplication by 1.002. The cited reference suggests that division by 1.002 would be required.

Yes, we divided by the factor 1.002 and rephrased:

We used the factor -> we divided by the factor

References

Redington, A. L., Manning, A. J., O'Doherty, S. J., Say, D., Rigby, M., Hoare, D., Wisher, A., Rennick, C., Arnold, T., Young, D., and Simmonds, P. G.: Long-Term Atmospheric Measurement and Interpretation of Radiatively Active Trace Gases, Annual Report, Sept 2018 – Sept 2019, Department for Business, Energy and Industrial Strategy, London, UK, https://assets.publishing.service.gov.uk/media/5eddf868d3bf7f4601e57730/verification-uk-greenhouse-gas-emissions-atmospheric-observations-annual-report-2018.pdf, 2019.