We would like to thank the reviewers for their feedback on the paper and their constructive comments and suggestions, which have all been answered or addressed and improved the paper. Below, we address each point in detail. The original review comments are shown in black; our responses are shown in blue.

Reviewer #1

Major comments:

1. The Introduction provides a great foundation, but at 2.5 pages, it feels a bit dense. To help the reader get to the core of your argument faster, I'd suggest condensing this section. For instance, the EU regulation details are excellent, but they might carry more weight if moved to the later sections where you discuss derived emissions, trends, and their relationships with the enforcement. Similarly, moving the specific species reporting details to Section 2 would keep the Intro focused on the 'big picture' while still preserving that valuable data for later.

   We agree that improving readability in the Introduction is important. However, we feel that the material currently included provides essential context for the study and is most appropriately retained in this section.

   To address the reviewer's concern, we have restructured the Introduction by introducing subsections. Specifically, we added Section 1.1 European Regulations to cover the EU regulatory framework and Section 1.2 Emission Monitoring to describe species-specific reporting.

2. In the results section, the discussion focuses primarily on (1) observed trends in relation to current regulations and (2) the discrepancies between UNFCCC-reported data and top-down estimates from various inversion systems. While these technical comparisons are essential, there is an opportunity to situate these findings within a broader context. Specifically, the paper would benefit from discussing how European emissions factor into the global total and identifying which countries are the primary contributors to NW European emissions. While percentage contribution of each compound is presented in one sentence at the end of each compound section, adding a dedicated section for this broader analysis would significantly enhance the paper's impact, elevating it from a regional estimate to a study with wider relevance.

   We thank the reviewer for this valuable suggestion to better contextualize our regional results in a global context.

   In response, we have added a new column to Table 2 (*NW / Global*), which quantifies the contribution of north-western European emissions to global totals for 2023 from [1], based on the average of the NAME- and FLEXPART-based emissions.

   In addition, we have included a short paragraph at the end of Section 4.6 Total HFC summarizing the contribution of total HFC emissions from north-western Europe in a global context and identifying the main contributing countries within the region.

   We considered introducing a dedicated section for this analysis; however, as the additional insight is primarily based on a limited set of summary metrics (now included in Table 2), we found that integrating this discussion into the existing structure provides a more concise and focused presentation.

   [1] Western et al., Global emissions and abundances of chemically and radiatively important trace gases from the AGAGE network, Earth System Science Data, 17, 6557–6582, https://doi.org/10.5194/essd-64517-6557-2025, 2025

Minor comments:

1. L14-15. Isn't this obvious – observations + modeling inversions is important? There is a community consensus on this. Is it really necessary to add this sentence at the end of the abstract?

   We understand the reviewer's point of view. However, we would like to emphasise the modelling effort behind the study, as multi-model evaluations remain relatively rare in the literature. We revised this sentence in the abstract accordingly.

2. Is the CBW station part of the AGAGE network or stand-alone observation sites? Please clarify. If it doesn't belong to AGAGE, you may consider move it out of the AGAGE network paragraph.

   The CBW station is not formally part of the AGAGE network; however, measurements conducted at the University of Bristol employ AGAGE calibration scales.

   We have retained CBW within the same paragraph because this section describes all measurements activities prior to 2023, whereas the subsequent paragraph focuses on newly introduced measurements.

3. Table 2. It would be helpful if you can add one column to show the fraction of NW Europe emissions out of the global emissions estimate for 2023. Providing the numbers will be useful to understand if there are variations of regional contribution among these compounds.

   We have implemented this suggestion by adding a new column ("NW / Global") to Table 2. This provides the requested quantitative comparison for 2023.

4. I see you use north-western most of the times, but north-west a few times. Please be consistent.

   We thank the reviewer for noting this inconsistency. The terminology has been standardized throughout the manuscript.

5. Why the less abundant species have larger relative uncertainties in their emissions estimates?

   We have clarified this point in the manuscript. The larger relative uncertainties are primarily driven by a poorer model performance for those minor compounds that are dominated by point sources. We now explicitly reference this in the text and direct the reader to Supplement S3 for further details on the analysis of inversion model performance.

6. Section 4.6. I found this section very informative. Besides stating the absolute emission numbers, could you also add trends in %/yr, instead of just describing the general characteristics, e.g. plateau, slow decline, substantial decline, etc?

   We appreciate this suggestion. We agree that quantitative trends are useful, particularly where the observational constraints are strongest. Accordingly, we have added trends expressed in % for north-western Europe and the UK, where the inversion results are most robust.

   For other regions, we have retained a more qualitative description, as the observational constraints are weaker; however, the corresponding figures and archived emissions provide necessary information for readers who wish to examine these trends in more details.

Reviewer #2

General comments:

1. The difference between EDGAR and the NIR estimates receives little discussion. Since EDGAR seems biased high, it makes me wonder if it wouldn't have been better to use the NIR country totals as priors. Would in that case that inversion have ended up systematically lower than the NIRs? If the goal is to evaluate the NIRs, using those as priors seems a more logical choice. I understand that NIRs do not provide gridded emission maps, but that would easily be solved using EDGAR for the emission distribution.

   We thank the reviewer for raising this important point. The choice of prior emissions is indeed a key aspect of inversion studies.

   We chose EDGAR as the prior primarily because it provides spatially distributed emissions across the full domain, which is required for the inversion framework. In addition, we apply relatively large prior uncertainties, allowing the observations to constrain the posterior estimates. For this reason, the posterior estimates of population-distributed HFCs are only weakly dependent on the choice of prior, and we do not expect that using NID-based priors would systematically shift our posterior emissions closer to, or further from, the NID values.

   We agree with the reviewer that using NID-based priors would also be a valid approach. However, if the objective is to evaluate national inventories, maintaining a degree of independence from those inventories is advantageous. In this context, using an alternative prior such as EDGAR allows for a more independent comparison.

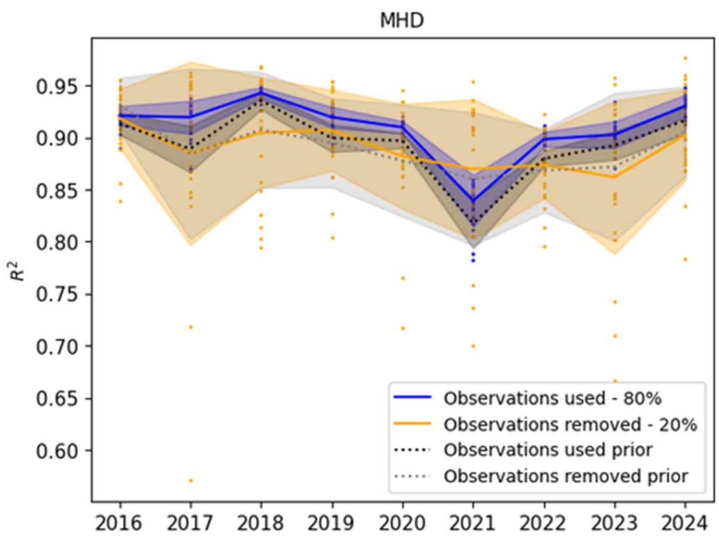
   A detailed investigation of the sensitivity of the posterior emissions to the choice of prior is currently underway and will be submitted for publication soon.

2. The benchmarking of inversion performance against measurements that are withheld from the inversion would provide an essential confirmation that the posterior emissions have indeed improved compared with the prior. I understand that the authors would like to keep the paper focused and concise, but such a test is nevertheless important to include in my opinion.

   We agree that such validation provides an important assessment of inversion performance, and we thank the reviewer for this suggestion.

   In response, we performed an additional evaluation using the inverse model InTEM. Twenty-four InTEM inversions were performed, each with 20% of data removed at random, in sixteen 5-day periods. Each produces site-specific modelled timeseries, $T_{posterior}$, for the seven sites and a posterior emission distribution. The latter from each of these inversions is used to calculate a simulated timeseries, $T_{removed}$, for the removed 20% of data. To test the impact of each inversion, the Pearson correlation coefficient value, R^2, is calculated for the two simulated timeseries, $T_{posterior}$ and $T_{removed}$, against the observed timeseries.

   The figure below shows the resulting R^2 values for HFC-134a at MHD, with the posterior $T_{posterior}$ shown in blue and $T_{removed}$ in yellow. The solid lines represent the mean R^2 across all 24 InTEM runs using the posterior timeseries, with shaded regions indicating the 1-sigma variability calculated from the 24 repeats for each time period. The dotted lines show the corresponding mean values obtained using the prior timeseries. The inversion posterior results show an increase in R^2 relative to the prior, along with a reduction in variability. In contrast, the non-inverted results show neither improvement. This indicates that the inversion improves the estimates.



We have added this analysis and the corresponding figure to the Supplement S4.

3. The geometry of the network change over time with new sites joining. For the goal of assessing the impact of new regulations on HFC emissions since 2017 it should be excluded that the change in network around the same time could have interfered. The experiment that has been performed to test the enhanced observational constraint offered by the new sites confirms the sensitivity of the emission estimates to those sites. However, the question how the inferred trends could have been affected doesn't receive the necessary attention.

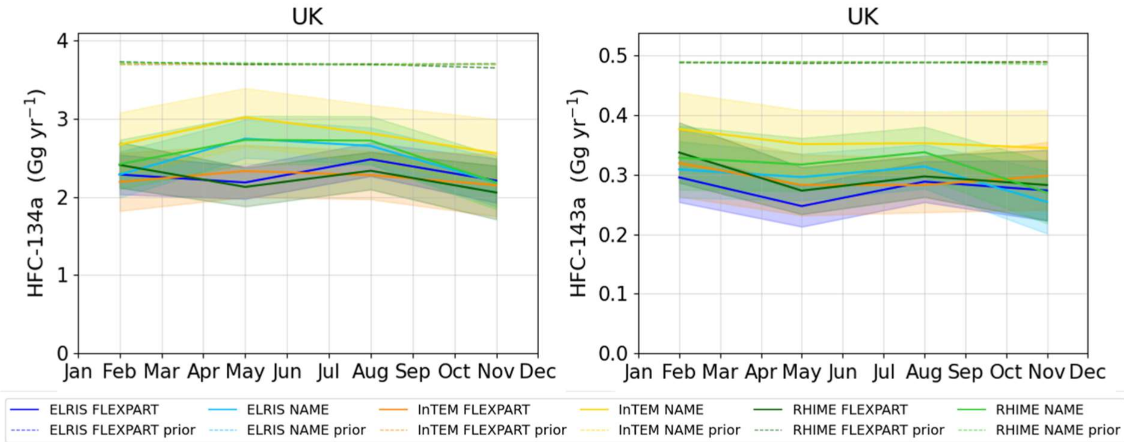
   The additional observation sites included in this study were introduced relatively recently (from 2023 onwards). As a result, the time period during which the expanded network is available is still too short to robustly assess its influence on long-term emission trends.

   While we have demonstrated that the new sites enhance the observational constraint, a quantitative assessment of their impact on inferred trends would require a longer time series.

4. If I understand correctly the a priori emissions are distributed uniformly within each year. While this seems a reasonable assumption, it would nevertheless have been useful to test if the posterior emissions show a significant seasonality. If so, this could provide important information about how well seasonal variations in atmospheric transport and vertical mixing are captured by the transport models.

   We thank the reviewer for this valuable suggestion.

   We believe that the UK is the only country with sufficient data density to generate results at a 3-monthly resolution. The figure below presents UK emissions of HFC-134a and HFC-143a estimated with a three-monthly resolution, for all six inversion systems, with values averaged by month over the period 2013—2024. No significant seasonality is identified. Therefore, we prefer not to include the results for only one country in the manuscript. We have, however, added a sentence in the manuscript mentioning that, with increased observational data density, higher temporal resolution inversions may become feasible, potentially allowing seasonal variations in emissions to be identified.



5. Some discussion is missing of why the FLEXPART estimated emissions are systematically below those inferred using NAME.

   We agree that this is an important topic. A detailed investigation of the systematic differences between FLEXPART- and NAME-based inversions is currently underway and will be submitted for publication soon.

Specific comments:

1. Line 110: Please specific which supplement.

   We have clarified this by explicitly referring to Supplement S1.

2. Table 1: The middle of the table seems to have an incomplete entry.

   The original formatting avoided repeating the TOB entry. We have now repeated the entry to improve the readability of the table.

3. Line 126: ‘The estimated total … ‘ This refers to the a priori emissions I assume? In that case a reference is needed.

   We confirm that this refers to the a priori emissions. Their magnitudes are an estimated guess and are intended to provide an initial estimate of the correct order of magnitude.

   In the inversion frameworks, relatively large prior uncertainties are assigned, meaning that the inversion is not strongly constrained by the prior and can adjust emissions based on observational information.

4. Line 134: Please specify which hourly intervals.

   We have clarified this by specifying “fixed 4-hour intervals” in the revised manuscript.

5. Line 138: Scaling factors to spatially and temporally constant boundary conditions?

   Yes, the scaling factors are applied to boundary conditions defined for each month and for the four cardinal directions of the domain.

6. Line 149: What is the a priori uncertainty on the station bias and what did the inferred biases look like?

The by-site bias has an a priori value of 0 and an a priori uncertainty equal to the baseline uncertainty. As mentioned in the text, this is estimated from the REBS method, using MHD observations. In summary, the REBS method estimates the baseline mole fraction by iteratively fitting a non-parametric local regression curve to observations. The uncertainty of the fit is what we use as a priori uncertainty of the by-site bias. A posteriori values of the bias vary from species to species but are usually in below 5 % of above-baseline pollution events. For some sites and species (e.g., TOB, HFC-134a) larger values of up to 15 % were observed. Please note that the by-site bias may absorb any site-specific transport bias. This may be related to the chosen model height above ground to represent smoothed topography but may also relate back to biases induced in the transported baseline mole fraction with respect to the relative sensitivity of a given site to the rather coarse baseline interfaces (11 segments). A more detailed discussion of baseline and bias impacts will be presented in a forthcoming publication.

7. Line 284: The posterior emissions for France seem to agree with EDGAR. Does the difference between EDGAR and the French NIR provide any clue why this is the case?

   This is an interesting observation. At present, we do not have sufficient detailed information on the methodologies underlying either EDGAR or the French NID to fully explain this agreement.

   As such information is not readily accessible, a thorough assessment is beyond the scope of the present study. We therefore leave this question for future investigation.