Author Reply to RC1, 10 Feb 2026

**RC1**: 'Comment on egusphere-2025-4542', Joseph Pitt, 01 Dec 2025

This study investigates sources of error in typical aircraft mass balance experiments and provides guidance for experiment design. It is well written and easy to follow, with clear conclusions: storage leads to random error in E_H/E_S, whereas extrapolation below the lowest transect can be an important source of systematic error. This is consistent with previous studies, but the thorough investigation presented here provides valuable insight for planning and evaluating future real-world data. I suggest that this paper is suitable for publication with only minor revisions. I think it would benefit from a slightly expanded discussion on the points below.

There is currently very little mention of the background concentrations. I understand that this is not a factor in the simulated data, as all the tracers released in the model come from sources within the domain. However, in real world examples variability in the background can be an important source of error, so at least some discussion of this is required. In particular, it impacts statements such as that in L368-370. Going further downwind may reduce the sources of random error addressed here, but there is a trade-off in terms of signal-to-noise above background.

Following the statement (end of 5th paragraph in Section 3.1.1, previously L370) we add the following text "*In a real-world scenario, there may be additional error due to noise in the concentration measurements, particularly for concentrations with a high background level. In these cases, moving further downwind (where the concentration enhancements above background due to the plume are much smaller), may result in a relative increase in error.*"

Further, at the end of the 3rd paragraph in the Conclusions, we add "*Generally, in real-life conditions, any reduced uncertainty due to flying further downwind from source must be balanced against increased relative uncertainty due to instrument measurement noise, especially for pollutants with high background concentrations (due to the weaker concentration signal as the plume disperses).*".

It is interesting that the kriging interpolation resulted in an overestimation of the instantaneous screen (L375). It would be great to see some more investigation of this. Was anisotropy in the variogram considered? I wonder if the variogram becomes more isotropic as you move further from the source? That would make intuitive sense to me. Were other functions (i.e. other than the spherical function mentioned) tested when fitting the variogram? It could also be interesting to see if this choice impacts the overestimation, although I appreciate that it is hard to draw general conclusions because the best function will always be specific to an individual flight. The same goes for the area source flights – L487 points to the kriging as a potentially significant error source so it would be good to see this case investigated too.

* We are currently working on a new analysis to be added as Sections 3.1.3 (for stack sources), and 3.2.2 (for area sources). The new sections are titled "Optimizing Screen Interpolation". Preliminary work shows that some flights do show anisotropy in the variogram, but the end results are not sensitive to the range of the variogram fits. For example, halving or doubling the model range value changes the average concentration in the screen by less than 2%.

As part of this new analysis, we compared results with spherical and exponential kriging, as well as Voronoi nearest-neighbour interpolation. Results (so far) are not sensitive to the model type, suggesting that the errors we are seeing are related to the sparse sampling and not the interpolation method. The results demonstrate that the differences between the instantaneous and non-instantaneous emission estimates (e.g. the red circles and black squares in Fig. 3) are primarily due to errors in the interpolation. These results will be discussed in detail in the revised manuscript.

The investigation of the vertical transect spacing is interesting but the results are hard to interpret. The hypothesis that plume movement could be responsible for the changes seen in the Sep 2 case seems plausible, but it would be nice to see this tested. Seeing as we are dealing with simulated flights, could a test be done where the order of the transects is changed?

We have done this analysis for both flight sets (Aug 20 and Sep 2). The figure now includes both cases in reverse, and the following text is added to the end of Section 3.1.2, "*To investigate this, we repeat the flight sets using the same flight paths with the directions reversed (i.e. top-to-bottom). In these cases, the flight begins at 16:20 (for the first flight in the sets) at the highest point and each flight samples at identical locations in the opposite direction, finishing at the lowest point at a height of 150 m. The reversed direction results in significant improvement in the estimation of* $E_H/E_S$ *for the Aug 20 flight sets (especially at smaller transect spacing), but increased error for the Sep 2 flight sets. In both cases, the variability between flights in each set is reduced for a transect spacing of* $\Delta Z$ *= 50 m, but is slightly higher than the variability of the bottom-to-top flight sets with* $\Delta Z \geq$ *100 m. As mentioned in Section 2.4, real flights are often flown in an upward direction so that the vertical extent of the plume can be determined while flying. While it is difficult to know the vertical extent of the plume beforehand (which would be required for a flight in the downward direction), these results demonstrate the potential advantage of flying once in the upward direction, followed by a subequent flight back in the downward direction.*"

L213/216 – refers to the known emissions as Es but I don't think this has been defined yet

In both cases we add "the *ratio of the storage term to the known emission rate* ($S/E_S$)."

L225 – in some cases even faster than 2 Hz. I know the UK FAAM aircraft has a CO2/CH4 LGR with a data acquisition rate of 10 Hz, although the cell turnover time means that the effective frequency of measurement is less than this (more like 7 Hz I believe).

This is modified to "…from 0.5 to *as high as 10* Hz *(e.g. France et al., 2021)*," and the citation for the France et al (2021) study is added.

L360 – it might be worth rephrasing this to clarify that it is E_H/E_S which is lower in the non-instantaneous cases (i.e. the underestimation is worse). A "lower underestimation" could perhaps be misinterpreted.

We have modified the text to "(*i.e. a lower* $E_H/E_S$ *value is estimated by the assumed below 150-m concentration relative to what would be determined with the actual below 150-m concentrations*)".

L413 – typo "sometimes"

Corrected.

Figure 9 – formatting error on some axes labels

This error occurred in the Word to PDF conversion. We have reformatted the figure so that this doesn't happen.

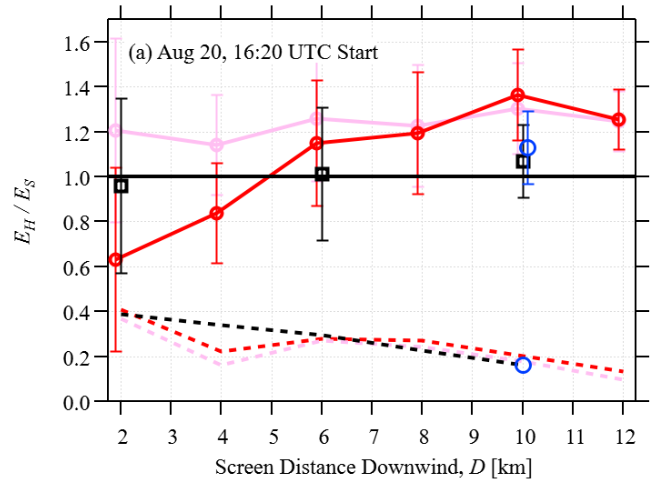
L576-577 – it makes qualitative sense that more information below the lowest transect would help. Could this be tested? At least for the case of a mobile vehicle you could presumably add an extra transect at z=0 with a typical vehicle speed and see what difference this makes

We have added a new section to investigate this, below is the added text at the end of Section 2.4 and the newly added section, which references a modified Figure 4.

"*We also investigate the potential improvement to the emissions estimate through ground-based mobile vehicle concentration sampling. For the case of the Aug 20 flight set (at 16:20) downwind of the stack sources, we simultaneously sample concentrations at the lowest model level beneath the flight path (Section 3.1.5). Although vehicle path locations are typically limited to roadways, we investigate here the highly idealized case where a car or truck can drive directly beneath the flight path of the aircraft for the duration of the flight. We assume a constant vehicle speed of 60 km/hr (16.7 m/s) and drive a single transect south-eastward from the most NW location. These values are then used in the interpolation of the screen without the need to assume a profile below a height of 150 m.*"

***"3.1.5 Adding Ground-based Vehicle Measurements***

*Figure 4 demonstrates that there can be considerable error associated with the assumed concentration profile below the lowest flight path of 150 m. As discussed in Section 2.4, in some situations, it may be possible to measure concentrations at ground level with a mobile measurement platform on a car or truck. Results for the Aug 20, 16:20 start flight set augmented by ground-based vehicle measurements are shown in Figure 3a. With these surface-level measurements, the assumption of a constant profile below the height of 150 m is no longer necessary, and the screen can be interpolated using these additional measurement values. The horizontal advective flux with ground-based measurements consistently overestimates the emission rate for all downwind distances, with values of $E_H/E_S$ ranging from 1.14 to 1.30 for $2 \leq D \leq 12$ km. This demonstrates that the underestimation of the horizontal advective flux close to the stacks ($D < 6$ km) with an assumed constant concentration below 150 m is predominantly due to a large amount of the plume being below the lowest flight path, as was shown in Figure 4. Further, the variability between flights is reduced by up to 6% (at $D = 4$ km). The results demonstrate significant value is obtaining surface-level measurements where possible, so that extrapolation below the lowest flight path is not required.*"




This is the modification to Figure 3a. The following text is added to the figure caption "*The pink lines in (a) demonstrate a case with additional measurements from a ground-based vehicle, discussed in Section 3.1.5.*"

---