Reply to RC1

The authors wish to thank the reviewer for their thorough review of the manuscript and their very helpful comments. The comments were carefully considered in the revised version of the manuscript. Below we respond to the different comments and suggestions.

- L18 - The waste and coal mining sectors are also important contributors of the anthropogenic emissions, which can also be controlled/reduced. In the waste case, it generally related to area source emissions. However, the coal mining sector is also related to point source emissions (Karacan, 2025), which is the main focus of this study.

  The mention of the coal sector was added in the text as well as the suggested reference.

- L41 - Please, you could also include here the following reference: Joyce (2023)

  The reference was added.

- L47 - The input are methane concentration maps or methane concentration enhancement maps? Please, clarify.

  The inputs are methane concentration enhancement maps. This was clarified in the text.

- L69-70 - Applying simulations of this kind to L1 data can lead to biases? In other words, can we trust the accuracy of the simulations? I am mostly concerned about how the simulations are convolved to the instrument spectral response function when integrating them into the L1 data. Not applying a correction as in Gorroño (2023) - Eq. 4 might lead to biases in reference to a real-like plume.

  There is indeed a small bias that occurs because we multiply two spectra already convolved with the instrument spectral response function. However this bias depends on the FWHM at the considered wavelength, the smaller the FWHM is, the smaller the bias is. In the case of EnMAP the FWHM is actually quite small with an average value around 12nm in the SWIR. We performed simulations to measure the magnitude of the bias for different methane concentration enhancements. We obtained an average relative error of 0.02% between the spectra convolved correctly and the spectra resulting from the product of two already convolved spectra. This relative error is lower than the SNR of EnMAP in the SWIR. Hence, we consider that the bias is negligible in the case of EnMAP.

- L71 - Later on in the text, it is mentioned that North America locations are more heterogeneous in comparison to the other two. Even with that comment, the location characteristics are vaguely defined. The brightness and heterogeneity of a scene is essential to assess the capability to detect and quantify. It would be very helpful to add further information about it. For instance, a Table listing the sites would be fine. IF the current list of sites are relatively bright and homogeneous, it is important to mention that the performance of the methodology could be worse in darker and more heterogeneous scenes. If so, testing the methods in this kind of scenes would be the ideal way to address this comment. However, mentioning that the location sampling does not consider this kind of scenes and that the results might be worse is also valid.

  We added in the text a list of the different sites as well as an analysis of some of the background samples. In this analysis we show that we have both heterogeneous and homogeneous backgrounds to generate our dataset.

- Section 2 (in general) - A diagram showing the methodology steps would be very useful for the reader. Please, consider to add one. One example of how to do it can be found in Gorroño (2023) - Figure 4.

  A diagram was added to the manuscript.

- L102-107 - Most part of this paragraph is redundant, since the previous paragraph already made this point. Please, consider to remove the redundant parts.

  We reformulated those parts of the text.

- Figure 1,5,6 - Please, add labels in the colorbar (i.e. ΔXCH4)

  Labels in the colorbar were added to all the figures.

- L163 - IME acronym was already defined and Frankenberg paper was already cited. Please, just use IME.

  This was changed in the text.

- L190 - It would be nice to show the Ueff fit. Since the number of points and the wind speed range is larger than in Guanter (2021) simulation dataset, there should be a higher robustness in the fit. Please, consider to add it. On the other hand, it is not explained how the Ueff is deduced. Please, a brief description of the process will be appreciated.

  The authors carefully considered this suggestion, however, we choose not to include the Ueff fit in the manuscript. We believe that it would suggest that there is a perfect or ideal Ueff to reach. However, the Ueff that is fitted depends on the masking procedure and the retrieval algorithm. A different masking procedure for example would lead to a different Ueff to fit and a different robustness of the fit.

  We added in the text that the effective wind model is obtained by linear regression and we described our masking procedure.

- L197 - Frankenberg paper was already cited in the text due to the IME method. Moreover, the IME acronym was already defined. Using the IME acronym would be more consistent than the 'Integrated Mass Enhancement'.

  This was corrected.

- L202 - Please, add a reference to the ConvNext models.

  The reference was added.

- L205 - The validation set should be independent from the training set. Please, clarify.

  The validation set is indeed independent from the training set. This was clarified in the manuscript.

- L211 - Why the shifts have not been done with higher pixel separation (more than 3 pixels)? Please, clarify.

  We used 3 pixels because it seemed empirically a good range for the source location error. In practice, the order of magnitude for the apparent plume source is 2x2 pixels. Hence, it is likely to have errors of this magnitude.

- L213-226: this part is hard to understand. Why is there uncertainty when applying plume rotation? Then, the rotation is only applied for the plumes to follow a x-axis direction? Then, the shift is only 0-3 pixels? Please, I recommend the authors to reword this part of the text to enhance clarity.

The authors thank the reviewer for this remark as this part of the text was not very clear. By "uncertainty" we meant that it is harder for the network to find and extract information from the plume pixels because any pixel can be a plume pixel. When pre-processing an image for inference the image needs to be rotated so that the plume is aligned with the x-axis and then the image needs to be cropped into a 100x100 image with the source in a fixed position. The 0-3 shifts are used in the data augmentation to be robust to imprecision in the image cropping. All of this was clarified in the text.

- L223 - Saying that the plume tail in the retrieval is very noise is not accurate. I would rather say that the authors meant that the plume tail enhancement is approximately at background level. Please, clarify.

  This was clarified in the text.

- Section 4.3 - What can be said about the level of uncertainty of the estimations? Since wind speed and plume masking (big error sources) were removed from the calculations, how does it benefit the flux rate uncertainty? Later on the text, when analyzing the controlled releases, it is mentioned that the precision is better. This is an important point because it improves the state of the art situation in which the uncertainty is huge. Please, emphasize this achievement in the text.

  A discussion on the level of uncertainties was added in the text.

- L332 - But the main problem was the appearance of retrieval artifacts. What happens if using MetFluxNet on retrievals with false positives? Wouldn't it be better to use MetFluxNet-sparse after cropping the area with the plume to not account for retrieval artifacts? Later on the text, it is shown that the artifacts in the controlled release retrievals do not have a big impact in the results. What would happen if (for instance) there is a facility with a high score in the retrieval? How would the prediction change? This is important because this is a relatively common case in real plumes. Please, discuss.

  We added a discussion on this issue in the text.

- L350 - In Sherwin 2023b, there are more PRISMA plumes. Since there are very few plumes from controlled releases to test the methodologies, it is important to leverage the whole extension of available data. Please, consider to extend the analysis to the rest of PRISMA plumes from Sherwin 2023b.

  We added to Section 4.5 the other PRISMA plumes, except for the plume released on the 27/10/2022, during a PRISMA overpass. This plume is not visible in our methane enhancement retrieval image, the order of magnitude of retrieved concentrations is lower than the background noise level. This was explained in the text.

- L384 - The IME method is based on the Ueff calibration. This calibration is made using a specific criteria for masking. Here, the masking process is not specified and the lack of accuracy in the IME method could be due to using a masking criteria different from the one used in the original calibration. Please, clarify.

  This was clarified in the manuscript. We added a description of our masking procedure. In particular, we ensured that the masking procedure used when reproducing the results of (Guanter et al.,2021) is the same procedure used in their paper.

- Conclusions - It would be great to talk about the implications of this work to other instruments such as EMIT or MethaneSAT. Can these insights be applied to other instruments?

  We added a discussion on how and when MetFluxNet could be generalized to other instruments. The main limitation is the spatial resolution which restrains the type and size of spatial features available for the network. Hence, an image from an other instrument needs to be at a 30 m resolution to guarantee a good result when using MetFluxNet.