The document lists the comments (written here in small font slanted) and our responses to **Reviewer #1** .

We kindly appreciate the constructive review of our manuscript. The comments were great guidelines to improve the paper. Note that in the *Specific comments* some comments are only answered with DONE if the corrections were directly applied to the manuscript.

General comments:

This article applies several different CH4 retrieval schemes to HySpex imaging spectrometer data from an anthropogenic point source. It compares a nonlinear algorithm, involving spectrum fitting using a nonlinear radiative transfer model, to various linear schemes. It also compares several algorithm variants, including pre-clustering data with k means and accounting for the covariance of surface reflectance in the nonlinear model.

There are some clear and obvious achievements. The algorithm descriptions are incredibly comprehensive, with enough detail to serve as a reference for investigators coding their own implementations. The manuscript deals with an important and timely topic, adding to the growing literature on point source GHG detection from coarse-spectral-resolution imaging spectrometers. It independently validates these approaches, and adds some new data sets to the mix.

This said, I have some recommendations for how the manuscript might be improved.

My first major recommendation is to clarify the thesis statement. After reading the article, I'm still a bit unclear on the fundamental contribution. The manuscript focuses mostly on retrofitting the BIERRA algorithm for CH4 point source detection at coarse spectral resolution. However, by the end of the manuscript it is unclear what advantages this offers beyond the state of practice Matched Filter methods or other very similar nonlinear model-fitting methods in common use (like the Thorpe et al. IMAP-DOAS approach). The affect of albedo on nonlinear CH4 retrievals is great, but it has been investigated even more thoroughly before - see for example Ayasse et al. 2019 (https://doi.org/10.1016/j.rse.2018.06.018). I think the authors could do a better job of calling out what is new and significant about the BIERRA approach.

Firstly, we added a clearer statement of the research question (see before the Methodology section) to highlight the contribution of our work. Specifically, we have emphasized the retrofitting of the BIRRA algorithm for including covariance matrices in the fit which turned out to be a crucial aspect for fitting methane from hyperspectral datasets. In the revised manuscript the advantages of the weighted BIRRA solver over other methods is pointed out more clearly. Moreover, it is now also discussed in more detail how BIRRA compares to other methods (see Discussion section) and the Result section clearly highlights its superiority in terms of accuracy and precision compared to the other setups.

My second major recommendation is to have a quantitative performance comparison. The current assessment is fairly subjective, related to the quality of the plume image and the visual appearance of background interference. Couldn't the background variability outside the plume be used to quantify the detection noise for each method? And couldn't the strength of the plume enhancements then be used to create an SNR score or statistical confidence? As a part of this effort, it would be great to translate all of the plume maps into similar units. Currently maps appear variously as ppbv, alpha-CH4, and "enhancement factors," which makes it difficult to inter-compare. It should be possible to translate any one of the linear detection algorithm results into an equivalent CH4 mass enhancement, and compare the effective plume-to-background detection SNR of each of the algorithms.

To fulfill the request, we performed a more quantitative performance comparison by utilizing the background variability outside the plume to quantify the detection noise for each method. As part of this effort, we provide the same units for the CH4 enhancements (mole fractions) throught the results section, which makes it easier to compare the performance of different algorithms. Furthermore, we have created a score to evaluate the quality of the solvers (see Tables 2 and 3).

Minor suggestions:

1. Almost all of the prior literature cited on CH4 point source detection, and the vast majority of the imaging spectroscopy community working at these spectral resolutions, plot spectra in wavelength rather than wavenumber. Setting aside the question of which convention is more convenient or appropriate from a technical perspective, it would certainly be easier for the majority of the readership to quickly understand the figures if wavelengths were used. This would make the instrument sampling evenly-spaced in the horizontal direction.

We revised the corresponding figures to present the spectra in wavelength units in addition to wavenumber units. Wavenumbers are still around in the revised verison, e. g., for the naming of the retrieval windows but we think this does not compromise comprehension.

2. On line 61, are there citations for CarbonMapper or CO2Image missions? I think the claim that CarbonMapper operates at higher spectral resolution than average land surface imaging spectrometers (5-10nm sampling) could be incorrect.

Since we did not find reliable information on the spectral resolution of CarbonMapper we removed the instrument from the introduction.

3. On line 65, in the literature review of airborne CH4 point source campaigns, consider also the studies by Frankenberg et al. 2016 (https://doi.org/10.1073/pnas.1605617113) and Duren et al. 2019 (https://doi.org/10.1038/s41586-019-1720-3) which were earlier and larger.

Both source were added in the literature review of airborne CH4 point source campaigns.

4. Figure 2 (a) seems to be missing some lines. "Grass" is misspelled.

DONE

5. On line 183, can you provide any more specifics about the low order polynomial used? What was its degree and where was it centered? The details are significant because, as you note, the surface reflectance is often quite complex over these wide spectral ranges. Changes in the reflectance representation can have huge changes on albedo sensitivity.

The last paragraph in Sec. 2.3.2 answers this question (second order).

6. On line 195, the section comparing different least squares solvers seems ancillary. Least squares solvers are a commodity

DONE

7. Figure 10. I'm not sure what this is supposed to show. Could it be removed?

DONE

8. On Figure 12, the enhancement within the plume appears completely saturated, which makes it difficult to assess. Can you rescale the colormap to make it more similar to the other plume images? We adapted and rescaled the colormaps. They are now adjusted to a common scale to facilitate comparability. Most pixels should be within the colorbar range.

9. The discussion and conclusion is good, but would be further strengthened by quantitative claims about where and how the different algorithms outperform each other.

We modified and enhanced the discussion and conclusion sections. We added Tables 2 and 3 and a correlation plot to clearly present results more quantitatively and highlight areas where each algorithm outperforms the others.

With best regards,

Philipp Hochstaffl and co-authors