Dear Reviewer,

Thank you for your review of our paper. In order to respond we have kept your original comments in the black text below, our responses are in blue, and our changes to the paper are in underlined blue text.

This manuscript by Malina at al. examines whether S5P/TROPOMI and S5 can retrieve 13CH4 with sufficient precision to differentiate CH4 source contributions. Carbon isotope measurement in CH4 could provide useful information on the atmospheric CH4 budget and the changing CH4 sources/sinks that are not fully understood. Therefore, data availability of 13CH4 measurements are highly anticipated but challenging for carbon cycle science community.

The information Content (IC) analysis approach in the study is similar to the one presented by Malina et al. (ATM, 2018) for GOSAT-2. Both are feasibility tests to these new/future satellites. Both studies applied IC analysis to the synthetic measurements, assuming clear sky condition (non scattering atmosphere). One of the main differences between GOSAT-2 and S5P/S5 is resolution due the different spectrometers (resolutions), FTS (0.2 cm-1) and push broom spectrometer (0.45 cm-1).

The paper would be interesting for the readers of Atmospheric Measurement Techniques, in particular for those studying GHG from the space. This kind of feasibility test is essential before handling real data and contribute to exploring the possibilities in coming satellites. However, the paper needs to be improved. There are missing information and editorial errors (missing words, incomplete sentences, and so on), and inconsistencies with figures. Some figures are not clear.

The paper potentially contributes to the GHG community. However, at this stage, I recommend major modification for further consideration.

Major comments

1. Overall, the manuscript has a lack of information and consistency in figures. Please check the figures (including captions) and descriptions in the main text. Figures should be self-descriptive. Also, the figure legends should not overlap/distract the plots (see more in the specific comments).

Thank you for your comments, we have addressed your concerns in the specific comments below.

2. In Introduction, the authors said ‘the key questions becomes whether SNR or spectral resolutions is the key limiting factor in the retrieval of methane isotopologues”. At the end of Sect. 4.5, the authors conclude that "SNR is more important... than spectral resolution". However, this comes from the comparison between SWIR1 and SWIR3. SWIR1 has higher SNR than SWIR3, but they both have the same spectral resolution, which is lower than GOSAT-2. Before the conclusion, a discussion on spectral resolution vs SNR should be given.

Thank you for your point, however we respectfully disagree with the need for a discussion on spectral resolution vs SNR. This is a topic that has been extensively discussed elsewhere (e.g.

(Wunch *et al.*, 2011), and we do not feel that this is a controversial point. The key reasons for the difference between the SWIR1 and SWIR3 SNRs are the solar irradiance and surface albedo, both of which are lower in the SWIR3 as opposed to the SWIR1. This will lead to a lower SNR irrespective of spectral resolution. To emphasis this point, we have entered the following sentence into section 4.6.

The key reasons for the lower SNR in the SWIR3 band as opposed to SWIR1 is the lower solar irradiance and surface albedo at these wavelengths.

3. Definition of latitude bands (low-latitude, mid-latitude and, high-latitude) should be given before the results are presented (from Section 3 onward). Alternatively, specific regions should be referred to. Otherwise, readers might be lost. For example, in and mid-latitudes, the regional differences (between desert areas and non-temperate areas, inland and islands, etc) are more evident than the latitudinal characteristics.

We agree, and we have inserted the following sentences in to section 2.3 to provide a clear definition of high-latitude, etc.

Through this paper, we will refer to "latitudinal bands", which we split in three distinct areas. Tropical (~0-20°), mid-latitude (~20-60°) and high latitude (>60°). These are typically how model atmospheres are split (e.g. mid-latitude summer etc). Surface conditions will cause the results in this bands to vary, and we identify and significant regions that show significant deviation.

In the case of where regional differences being of high importance, we agree on this point. In our original submission, we typically highlighted specific areas that showed clear differences from other areas in a similar latitude band. For example highlighting the Amazon rain forest, or desert scenes. We therefore believe that we do not need to emphasis this point any further.

Specific comments

4. - Page 7, L4 'This paper builds on this study'. These two "this" must refer to different studies. Please rephrase this sentence.

Thank you, we have now re-written this sentence as follows:

"This paper builds on Malina et al. (2018)"

5. - Page 9, L7: "The spectral fit quality is good, with a chi-square value equal to 1". How was the chi-square value of 1 derived? Fig. 1 reads chi-square of 1.15, not 1.

We use the method outlined in (Galli *et al.*, 2012), and we have now included this reference in section 2.5, where the assessment statistics are described. We have also changed the chi-squared value reference in the paragraph below figure 1.

6. - Page 10, L1: "disagree with the 'truth', notably the methane lines at 1670 nm". However, no such disagreement is seen in the top panel in Fig. 1.

Agreed, we have looked further into this and found that the disagreement is more random in nature, rather than a persistent feature as originally thought. We have therefore modified this sentence to read.

However, there are some points which could be interpreted as not due to random noise, where the retrieval seems to disagree with the 'truth', notably the methane lines between 1645 and 1650 nm. However, upon further investigation we found that these features do not consistently appear in spectral residuals. Therefore the disagreements shown in Fig. 1 are random in nature.

7. - Page 10 L3: Please specify more of “complex behaviour” about the cause of difference?

This is now redundant given the changes outlined in point 6 above.

8. - Figure 1: Caption reads “-1.4S, -47.81W for a day in January 2015”. However the middle panel shows “Measured, 16/7/2015, 51.63 39.39”. They should be consistent. No right-hand scale in the bottom panel, which is mentioned in the caption.

This figure had been updated prior to submission, but unfortunately, the caption was not changed to reflect this. The figure has now been updated so that the legend and the caption match. References to left-hand scale and right hand scale have been removed as these are no longer relevant. Figure 7 has also been updated in this fashion.

9. - Figure 2: Overlapped legends are destructive. Please move them outside the panels (same for Figs. 8 and 12). Four colours indicate different regions, but it is hard to distinguish them. It would be more informative if the legends include the representing regions (not only latitude-longitude information).

Thank you, we have now moved the legends so they do not overlap with the plots. In addition we have highlighted the representative region for each example, e.g. Tropical or Mid-latitude etc. This has been performed for each of the Averaging Kernel plots.

10. - Figure 3: To see the overall performance, it would be helpful to have a total number of the measurement (before measurements with DFS (<1) have been filtered out.

In the original submission we state in section 2.3 that there roughly 10000 simulated synthetic spectra. We have now modified this to give the exact number (11041). This now gives context to the retrieval numbers stated in Figure 3, and other related figures.

11. - Page 11, L15: ‘almost 3000 additional valid retrievals, which is roughly 30% of the ensemble’? This statement is not clear. What do the authors mean by “3000 additional”?

We agree, and we have modified this sentence to read as follows:

The differences in measurement densities indicated in both Figs 3 and 9, show that the SWIR1 band has almost 3000 additional valid retrievals that pass the filtering criteria as opposed to SWIR3. These additional valid retrievals represent roughly 30% of the total ensemble, and is therefore a significant proportion.

We have also moved this sentence to section 4.3.

12. - Figure 4: 'lobal' should be 'Global'?

Thank you, updated.

13. - Page 12, L11: what 'show'? The subject of the sentence is missing.

Fixed.

14. - Page 12, L13: the last sentence seems incomplete.

We have modified the sentence to read as follows:

Meaning that a target uncertainty of 0.1 ppb is a more accurate requirement, as opposed to the original goal of 0.2 ppb.

15. - Table 3: The units should be given for 13CH4 and d13C intercept and sigma, separately (same for Table 4).

We have inserted the units "ppb" for the 13CH4 intercept and sigma columns for both Tables 3 and 4, and we have inserted the permil symbol for d13C.

16. - Figures 5, 6 and 11: Please give the units for x-axis and y-axis. Also, for Figs 5 and 6, please describe in the caption which parameters perturbed are.

The units for 13CH4 on these axes has been mode more prominent as compared to the original. The d13C plots have had 'permil' included on the axis. The captions of Figs 5 and 6 have been updated to included details on the parameter perturbation.

17. - Figure 6: Is the caption correct?

Thank you for identifying this. This caption has been updated.

18. - Page 16, L29: ILSF should be spelled out.

The acronym for ILSF had already been identified earlier in the document, page 7, line 6.

19. - Page 22, L4: SWIR3 should be SWIR1?

Thank you, updated.

20. - Page 23, L1: “Radiometric offset errors are not significant in the SWIR3 as opposed to the SWIR1 band’. More explanation and possible reasons for or discussion on this difference should be given.

We have replaced this sentence with the following paragraph to give more details.

Radiometric offset errors are less significant in the SWIR3 as opposed to the SWIR1 band. Radiometric offset errors typically lead to underestimation in the surface reflectance estimate (in the absence of aerosols/clouds) {Kuze et al., 2014). Given that the SWIR1 operates at a higher radiance magnitude than the SWIR3, any minor errors in the surface reflectance estimate will likely lead to larger errors in the calculated radiance, as opposed to the true radiance. Thus leading to larger errors in the SWIR1.

21. - Page 27, L34: What is "the L-curve method"? It should be introduced before being mentioned in the Discussion section.

We have expanded the description of the RemoTeC algorithm in section 2.2, this now includes an introduction to the L-curve method, and a background reference.

Please note that in addition to the changes indicated above in response to your comments, we have also made changes to additional editorials we spotted. In addition, in response to Thomas Roeckmann’s criticism, we have included a short section on the effects of scattering on the retrievals of $^{13}CH_4$. This now forms section 6 of the paper. Additional necessary details on the scattering elements of RemoTeC have been included in the RemoTeC section.

Please note that we spotted errors in Sections 3.5 and 4.5 of our original submission. The results shown in Tables 3, 4 and Figures 5, 6 and 11 were generated without the DFS > 1 filter that were included for the maps plots present in the rest of our submission. We have reprocessed this data, and have updated the relevant figures and tables, including the filtering criteria. We have updated the relevant portions of the text that reference the original results.

Galli, A. *et al.* (2012) ‘CH4, CO, and H2O spectroscopy for the Sentinel-5 Precursor mission: An assessment with the Total Carbon Column Observing Network measurements’, *Atmospheric Measurement Techniques*, 5(6), pp. 1387–1398. doi: 10.5194/amt-5-1387-2012.

Wunch, D. *et al.* (2011) ‘The Total Carbon Column Observing Network’, *Phil. Trans. R. Soc. A*, 369, pp. 2087–2112. doi: 10.1098/rsta.2010.0240.