Responses to Reviewer #2

We thank the reviewer for their comments and questions. Our responses are formatted as follows:

The reviewer's comment/question (numbered) is written in black italic text.

> Our responses are written in normal black text (indented).
>
> The revised text as it appears in the manuscript is written in normal blue text (indented), with relevant changes underlined.

Line numbers refer to the edited manuscript. We have also provided a tracked-changes document, but that has different line numbers.

1. P4, L92f: How would the mean optical depth of CO2 in both bands compare to that of CH4?

We have added the requested comparison to the text (L. 100).

The mean methane optical depth in band 12 is five times larger than that in band 11. <u>The mean CO_2 optical depth is about 5 times larger in band 11 and 24 times smaller in band 12 than that of methane.</u>

2. P4, L103: Concerning your assumption about the surface and aerosol reflectance properties in bands 11 and 12 – could you provide any references for your assumption that they are similar in both bands? According to a publication from Chen et al. (2006) using VIRS and MODIS observations (their Table 1), it seems that depending on surface type, the e.g., surface reflectance properties can differ at 1620 nm and 2130 nm.

The reviewer is right that similarity in spectral reflectance between the two bands is not guaranteed. Our multiband methods (MBSP and MBMP) will only work well when reflectances in the two bands are indeed similar. We explain this on lines 212-215:

> "One might expect the MBMP retrieval to be strictly superior to the SBMP and MBSP retrievals since it exploits both multi-band and multi-pass information to derive methane column concentrations. <u>However, this may not be the case for scenes with large differences in surface reflectance between bands 11 and 12 due to strong spectral dependence of the albedo (in which case SBMP might be superior)</u>, or if no good reference observation with consistent surface reflectance is available (in which case MBSP might be superior). We investigate the dependence of the methane retrieval precision on surface type and retrieval method in the following section."

3. Would you have any possibility using model simulations to investigate the effect of variable surface and aerosol reflectance properties on your retrieval (I assume the statement "The model accounts… not for … scattering" in L106f refers only to aerosol scattering)? I could also imagine that these effects are partly captured by your factor c in Eq. 1.

Indeed, scene-wide differences in surface and aerosol reflectances between spectral bands and satellite passes are captured by the factor c, but variability in reflectances across the scene leads to errors. We explain that in the text on lines 219-221:

> "Our Sentinel-2 methane column retrievals require close agreement between MSI surface reflectances measured on different satellite passes and/or in different SWIR spectral bands to isolate methane plumes in a scene. <u>The scaling factor *c* of Equations (1) and (3) is fitted to account for scene-averaged differences in reflectance between satellite passes or bands, but inter-pixel variability in surface reflectance remains and will be a source of error.</u>"

A detailed study of the impacts of surface/aerosol reflectance heterogeneity on Sentinel-2 methane retrievals over different scenes would be valuable, but is beyond the scope of the present work. We discuss generally the question of surface properties in Section 3.5, showing in equation (6) the relationship between reflectance variability and methane retrieval error.

4. P5, L115f: I agree with your statement that "Aerosol effects… (1) uniform across the scene, (2) … not co-emitted with methane…", however, what happens if the 'general' atmospheric aerosol load across the scene is relatively heavy on the day of the control observation but relatively clean on measurement days, or vice versa? Would this have an influence on the retrieved CH4 columns or also be correctly captured by c?

If the aerosol effect is uniform across the scene, then it will be captured by c. To clarify, we elaborate on the meaning of c in the SBMP method (L. 143-144). Please also see our response to comment #15 by reviewer #1.

… where c is a scaling factor that adjusts for scene-wide changes in brightness between satellite passes. Such changes could be due to variable observation zenith angles, atmospheric conditions, or surface conditions.

5. P5, L137: Any ideas why the fractional signals from the satellites are different? Is it related to the different windows and/or to slightly different instrument characteristics / calibration?

We clarify that the difference in signal change is due to slightly different band positions/widths (L. 153-154).

The difference between these two values is due to the instruments' slightly different spectral ranges (Figure 1).

6. P5, L141f: Similar tests as done for water vapour could potentially also be done for CO2 to quantify any induced bias on the CH4 column as indicated in L402.

CO_2 variability is much lower than that of H_2O. We performed the same test as for H_2O, but varying CO_2 between 400 and 410 ppm. This affects mSBMP by 0.01% or less (compared with 6% for the wide range of water vapor levels tested). We explain this on L. 162-163.

Water vapor columns over land may vary from 1 to 40 kg m^{-2}, as observed by the Orbiting Carbon Observatory (OCO-2; Nelson et al., 2016), but under nominal observing conditions m_{SBMP} varies by only 6% over this range (i.e., by roughly ± 0.002). Variability in background CO_2 is much less than for water vapour and has virtually no effect on m_{SBMP}; varying the CO_2 column between 400 and 410 ppm changes m_{SBMP} by ≤ 0.01%.

7. Additionally, would a relatively high atmospheric CO2 background concentration in the observed scene during e.g., the control observation, have a large influence on the retrieved CH4 column if the observation itself only exhibit low atmospheric CO2 background concentrations (in the case of the multi-band approaches)?

Please see our response to comment #6.

*8. P5, L142: "... this affects mSBMP by only 6%...": I do not quite understand what exactly the 6% refers to. Does it mean, a potential change in water vapour would cause a change of mSBMP by 6%, which would then be larger than the ~3% caused by a doubling of CH4 (L137), OR is the ~3% change caused by CH4 only modified by the stated 6% from water vapour so that, in this special case, the overall effect of water vapour on mSBMP would be only around 0.18% (0.06*0.03)?*

Thank you for raising this important question. We need to clarify that the 6% change is expressed relative to mSBMP, which is itself a relative quantity. So indeed, it is 6% of a roughly 3% fractional signal change, or a difference of about 0.2% fractional signal change, as the reviewer describes. We explain this in the text on L. 160.

Water vapor columns over land may vary from 1 to 40 kg m^{-2}, as observed by the Orbiting Carbon Observatory (OCO-2; Nelson et al., 2016), but under nominal observing conditions m_{SBMP} varies by only 6% over this range (i.e., by roughly ± 0.002).

9. P8, L199f: Would it be possible to also add figures of top-of-atmosphere reflectances for bands 11 and 12 for the scenes Korpezhe, Savannah, Urban and Farmland? That would be a good opportunity to visualize the variability in surface reflectivity for the scenes as mentioned in paragraph L223-228 (although I understand that the manuscript is already quite figure-heavy).

Thank you for this suggestion. The discussion of the additional scenes is intended only to illustrate the importance of surface type for the ability of Sentinel-2 to resolve methane plumes. Figure 3 and the discussion in Section 3.5 seem sufficient for this, and we agree that the manuscript already has a large number of figures. We hope that the geographic information we provide for the scenes will enable the interested reader to inspect them more closely online.

10. P12, L309: I would suggest also adding the median value for MBMP as done for SBMP and MBSP in L313.

We have added the median value (L. 339).

Precision fluctuates between about 0.1 and 0.25 mol m^{-2}, with a mean of 0.18 mol m^{-2} (27% of background, value reported in Figure 3), a median of 0.16 mol m^{-2}, and three outliers higher than 0.3 mol m^{-2}.

11. P13, L314: I think "less" would fit better here: "...because it is not subject to..." → ...because it is less subject to...

Good point. We clarify that the MBSP method is not subject to temporal variability (L. 345).

The MBSP shows the most consistent (but coarse) precision across the time series, presumably because it is not subject to <u>temporal</u> variability in surface conditions.

12. P13, L318f: As indicated at the beginning of Sect. 4.2, depending on the scene, one or the other method may work better. How would Fig. 6 and the mean emission look like if SBMP and MBSP were used instead of MBMP? Would one get a bias, enlarge the error bars or would the single emissions get 'more' variable?

The figure looks about the same, but with larger error bars for the MBSP and SBMP methods, and slightly different mean emission rates. We now state in the text the difference in mean values (L. 355).

Figure 6 shows the resulting time series of source rates, which range from 2.6 to 59.1 t h^{-1}, with a mean ± standard deviation of 9.3 ± 5.5 t h^{-1}. <u>Mean values estimated with the SBMP and MBSP methods are within 22% of this estimate.</u>

13. Additionally, how does a change of the control scene influence the retrieval precision (Fig. 5)?

We now compare the mean retrieval precision for the two multi-pass methods when using the second set of reference/control images (L. 345-347).

The SBMP and MBSP retrievals show much coarser precision, with mean (median) scene-wide standard deviations of 0.35 (0.28) and 0.38 (0.36) mol m^{-2}, respectively (omitting the extreme outlier in October 2019 from the SBMP calculation). The MBSP shows the most consistent (but coarse) precision across the time series, presumably because it is not subject to temporal variability in surface conditions. <u>We obtain similar multi-pass retrieval precisions using our second set of reference images, from September 2020; the mean precision is 0.39 mol m^{-2} for the SBMP method and 0.18 mol m^{-2} for the MBMP method.</u>

14. P13, L323f: Have you, in addition to the plume free scenes before and after the mission, also tested the 8 plume free scenes within the mission for the Hassi Messaoud oil field or are there any reasons why they cannot be used as control observations? Would their RMS value also be around 20%?

We have not tested this. A good reason not to use them is that they may contain active emissions that are simply below our detection threshold. We now explain this on L. 326.

Of the 20 non-detections, 12 were due to cloud cover and 8 showed no detectable plume, indicating a plume persistence rate of 93% for cloud-free observations. <u>Non-detections can be due source inactivity or to emission rates below the Sentinel-2 detection threshold.</u>

15. Fig. 1: I would suggest adding column to "slant optical depths" → slant column optical depths

Done.

16. Fig. 2: Just a comment: The SBMP retrieval appears to have some issues with surface elevation (or shadows caused by hills due to different illumination conditions on 2019-Oct-06 and 2019-Nov-20) if I interpret the red areas in the upper left and lower right corner in (d) correctly. Otherwise, the remaining part of that plot looks 'cleaner' than for the MBSP retrieval in (e).

Good point, we have taken note of this.

17. P4, L87: I would suggest capitalizing "high-resolution transmission molecular absorption (HITRAN)" → HIgh-resolution TRANSmission molecular absorption (HITRAN) database

Good suggestion.

18. P18, L444: "DJJ" → D.J.J.

Fixed. Thanks!

19. Fig. 4: Please add labels (a) and (b) to the two subfigures.

Done. Figure caption updated accordingly.