Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC2
https://doi.org/10.5194/amt-2022-261-RC2, 2022

Comment on amt-2022-261

Anonymous Referee #2

Referee comment on "Understanding the potential of Sentinel-2 for monitoring methane point emissions" by Javier Gorroño et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2022-261-RC2, 2022

In the past few years, satellite imaging has been shown to be a very important tool to track methane emissions. This has proved to be an important step toward reducing unwanted emissions and fight global warming.
The goal of this work is to propose a novel methane emission detection and quantification method for Sentinel-2, an open-access satellite constellation from the ESA. On top of this method, the authors show how the proposed framework can also be used, alongside simulated plume shape, to simulate realistic synthetic data and perform an in-depth analysis of quantification errors as well as detection limits.
While such work had already been done for other satellites, such as PRISMA, this is the first time, as far as I know, that a quantitative study is performed for Sentinel-2 using realistic synthetic data. This allows for a more extensive study than real on-site experiments, such as Sherwin et al. (2022), that are both costly and polluting (since they require potentially large release of methane).

This work is mostly clear and very impactful. It is a very important block in the design of new methods for detecting and monitoring methane emissions but also helps understanding better the optimal conditions and limits of the current methods.

I list here more specific comments:

- While the conclusions are indubitably important, one can regret the low amount of testing. Indeed, only 5 plume shapes and three locations (at a specific date) were used. I think it is too little to derive an accurate conclusion, especially given the variance of the results:
  - I think that more plume shapes is necessary. Given the high variability of the plume shapes used I feel like that it is quite difficult to draw good conclusions. Indeed, the authors mention that a small but highly concentrated plume interacts differently that a widespread plume but this is derived using only one to two examples. This is even

more important since results with similar shapes (such as plumes 1 and 5) are not very similar (see Fig.9 and Fig.10).

- All dates selected are during the summer for the respected countries. This correspond to optimal sensing conditions (in the sense that the illumination of the scene is the best). It would be interesting to see similar results for other dates to measure the variability of the method and the impact of seasonality for example
- More example of sites in these environments would have also been beneficial for similar reasons as the one mentioned previously.
- With only three sites use, one can regret the lack of more diverse locations. In particular, no site from the Southern hemisphere were selected. At least a site with different weather conditions such as Canada or Russia (with snow for example) would have also been beneficiary. Even though these sites can be difficult to monitor in practice, the potential (and limits) can be interesting nonetheless. This is even more important since the SNR is signal dependent and as such depend on the type of albedo of the location. While probably not necessary for this paper, I think that it can be a very interesting follow up work for the future.

- How stable is the simulation process regarding to the spectral characterization? The ESA regularly updates the spectral characterization of both S2A and S2B so I assume that each time the analysis needs to done again but one can wonder how much would be the error is this wasn't done (and as such what is the error that is done when using an imprecise spectral characterization).
- The authors mentioned the spectral sensitivity difference between S2A and S2B. It is not entirely clear the satellites is used for the different experiences (it seems that the observation is from S2A and the reference from S2B based on l.106-112, and thus the synthetic experiment is using purely S2B). I think it could be interesting to study in that case the difference in detection limit between the two satellites (I assume that S2A limit is smaller than S2B since it is more sensitive to methane) and if there is a gain in using the reference image from the same satellite as well.
- l.60: The difference with the approach from Cusworth et al. needs to be more explicit. From the text, I notice at least two differences (simulated vs real images and the proposed correction term), are they other differences? I also think that it comparing the proposed retrieval approach to the one described in Cusworth et al. can be a valuable additional experiment. As it stands it is difficult to figure out the added value of the proposed quantification methods compared to the one previously presented since no explicit comparison is done.
- Some of the analysis is missing (or should be cited appropriately if it comes from another source), such as l.170 "it was found that ..." or l.178 "". I think that the experiments that lead to these assumptions should be presented as well to prove that these are indeed valid assumption.
- In section 3, the authors go from a 3D plume simulation to a 2D mass field thus neglecting the impact of the altitude of the plume. This should be clarified in the text (i.e clearly stating the assumptions + justification). Clearly specifying all hypothesis for the study is very important.
- Given that the correction term is scene dependent, It would have been interesting to have Fig.5 for the other sites as well to see if there are noticeable differences between sites.
- It is not entirely clear when the temporal normalization is the same as the observation image. I think this should be clarified for the different experiments. Another interesting experiment is the measure of the variability of the estimation as a function of the reference date. The method recommend using the closest reference but this allows to answer important questions such as "What happens when the closest is a poor reference?" or "Is the closest always the best choice or could another criterium be

better?"

- l.~300: I think that a metric such as the SNR between the plume (delimited by the mask) and the rest of the scene used in Ehret et al. (2022) could be useful to describe the difference behavior in this section.

Additional form comments:

- I suggest that the authors cite the 2021 report of the Intergovernmental Panel on Climate Change that contains a very detailed analysis of the impact of methane emissions on the planet.
- l.44, "high-quality calibration and high temporal revisit": I think that this should be clarified. The calibration and temporal revisit is poor compared to other satellites such as Sentinel-5P so I think it would be better to explicit the characteristic of Sentinel-2 instead of just qualifying as "high".
- For me some information are difficult to find in the text. For example, l.198 the authors state that "methane plume quantification is obtained from isolation of the term ...". It is only two sentence later that it is explained how (I think since no clear link is made). These links must be clearly explicit so that these sections are easier to understand and follow.
- I suggest adding an equation that clearly explicit that $L_{B12} = \int_{B12} L_{TOA}(\lambda)d\lambda$ (l.~150)
- The paper is filled with small typos. I will list the one that stood out to me but I'm sure there are many others:
  - l.96: missing ref
  - "normalisation" (l.108) vs "normalization" (l.204)
  - l.235: "acquitions" -> "acquisitions"
  - l.390: 5Ì□0 -> ~50