Review AMT-2020-477 – High-frequency monitoring of anomalous methane point sources with multispectral Sentinel-2 satellite observations.

In this work the authors make novel use of the Sentinel-2 MultiSpectral satellite Instrument (MSI) to detect large methane emission sources, as opposed to the original intention of the instrument, a more general source of geo-information. The high spatial resolution of the Sentinel-2 MSI instruments (20 m) give MSI an advantage over the current satellite missions dedicated to methane retrieval (e.g. Sentinel-5P/TROPOMI), which typically have spatial resolutions >5 km. Several methods are proposed for detecting large methane sources and estimating emission rates as well as errors. These methods are applied to several test cases in well-known emissions locations, indicating that large methane well blow outs can be detected.

The paper is well written, the figures are high quality and the topic is novel and highly relevant, therefore the paper is a clear fit for AMT and I recommend publication. Before publication I recommend some expansion on the following points.

Major points:

There is no mention of instrument noise in the paper, for example equation (2) (and the related equations) rely on a fractional absorption model to infer column enhancements from reflectance changes in the measured radiance. The measured radiances will include instrument noise, with R and R' not showing the same noise due to capture under difference conditions. The fractional absorption model does not seem to account for instrument noise, meaning that the minimisation step between the R and m values may be underestimated. It is possible that the instrument noise is accounted for in the forward model calculations as identified in Jervis et al. (2020), if this is so, it should be explicitly stated, and the noise model identified. Further to this, do Sentinels-2A and B exhibit different behaviour due to differing instruments, and different noise levels?

The paper would benefit from some more validation, understandably this is challenging due to minimal measurements at this resolution (both spatially and temporally), and while there are some comparisons with GHGSat-D in this study (which show good results), I would have liked to have seen some TROPOMI retrievals over the test cases. These comparisons would distinctly show how TROPOMI cannot identify these plumes (or maybe they can), while Sentinel-2 can.

The authors briefly state at the beginning of page 10 that only unusually strong methane sources are detectable using the methods identified in this paper (~>3 t /hr). However, there is very little detail in the paper about how common such sources are and what they might be. The statement that most large sources fall between 1-10 t/hr is difficult to conceptualise without context. Therefore the overall utility of this method with Sentinel-2 remains unclear, and this should be improved. I recommend including a section placing the sources in the case studies in this paper in context with other global sources e.g. biomass burning events, or coal mining emissions, both significant methane sources (Saunois et al., 2020), therefore giving readers an idea of how the work in this paper could be applied globally.

Specific points:

P2 L 36 - (Schneising et al., 2020) should be identified here as well.

P2 L45 – It'd be useful to identify the main aims of Sentinel-2 in this section i.e. identifying land changes etc. This would help differentiate the scope of this study from the main aims of the Sentinel-2 mission.

P2 L53 – While 2300 nm does include significantly more methane spectral lines than 1600 nm, solar irradiance is several times higher at 1600 nm, and surface reflectance is typically higher at 1600 nm. Therefore, one has to be careful about interpreting the next statement '2300 nm being considerably more sensitive to methane than 1600 nm'. I recommend this section be qualified with some statement about solar irradiance.

P2 L54 – "Band 11 can therefore be used as a proxy for the continuum". It is not clear what is meant here? Surface reflectance values can still be significantly different between Band 11 and 12, especially in high albedo scenes. This combined with higher solar irradiance in band 11, indicate that there are significant differences between the bands.

P3 L67 – Are there any upcoming instruments that this method could be applied to?

P4 L93 – The risk of artefacts due to water vapour lines is identified here, but artefacts due to spectroscopic database uncertainties have not been identified. Since 2300 nm is very complex spectrally, it is likely retrieval artefacts will exist, especially over high reflectance environments where the lines can saturate. Therefore, I recommend to include a short discussion here (or elsewhere as appropriate) on the potential impact of spectroscopic parameter uncertainty.

P4 L100 – Here the concept of residual radiance analysis is introduced, the core of the analysis of this paper. The sentence reads a little bit as though this is a new technique, which it is not. I recommend that a discussion or identification of past uses of methane detection with residual radiance be included here e.g. (Leifer et al., 2006; Roberts et al., 2010).

P4 L105 – Here a radiative transfer model is briefly described, can access to this model be provided? Fundamentally, the results shown in the paper need to be reproducible, and the RTM is key in this regard.

Further, no mention of the model used for surface reflectance is identified which is surprising given the importance of reflectance in this study. Assuming the same model used in Jervis et al.

(2020) is employed, is a Lambertian model sufficient at such high spatial resolution? Would not a BRDF model yield improved results, possibly dealing with some of the heterogeneous scenes?

P4 L110 – Presumably the HITRAN2016 database?

P5 L115 – The assumption about aerosols is not well justified here, especially given the location of the case studies in this paper are in desert regions, well known to be affected by aerosols.

P5 Equation 2 – It is unclear as to how instrument noise in band 12 is accounted for in this calculation. The radiances used in equation 1 will include instrument noise, so there should be some accounting for this in equation 2?

P5 L136 – Are the differences between Sentinel-2A and 2B purely due to the spectral range differences?

P5 L141-142 – It is very unclear what the authors have undertaken here with assessing the impact of water vapour. It is stated that variations in background water vapour have a minor impact of 6% (although how does a 6% variation affect the precision). But what background is being used in this test? Are the water vapour variations the max that could be seen in the US standard atmosphere, or based off the max in a tropical scene?

P6 L149 – I'm not convinced that relying on similar radiance levels between the spectral bands works as suggested in this section. Even if surface reflectances are similar (which they may not be), solar irradiance and instrument noise will likely mean different SNRs between band-12 and band-11, yielding large magnitude differences. Can this all be accounted for in the 'c' factor in equation 3?

P8 L197 – It is stated that 'c' is used to account for scene-averaged differences in reflectances between satellite passes or bands. However earlier in the text (P6 L150), 'c' is identified as being used to account for calibration differences, implying minor variations. This line should be moved up to p6 to give more detail about the use of 'c'.

P9 L218 – The term 'plume detection limit' is not explicitly identified, please explain this term and how it is calculated. This term is also used in Figure 3, with no explanation as to what this is or how it is calculated.

P9 L227 – This MBMP precision is still significantly worse than the homogeneous scenes, is this still due to reflectance errors?

P10 L221 – This section should go in the conclusions.

P10 L253 – With regards to TROPOMI, it'd be useful to contrast the results of TROPOMI with those found by the methods shown in this paper. Therefore providing direct proof of the utility of this method with Sentinel 2, if such data is available at this time.

Technical points:

P2 L35 – TROPOMI now operates at a spatial resolution of 3.5x5.5 km^2.

P4 L88 – HITRAN -> HITRAN2016

References

Jervis, D., McKeever, J., Durak, B. O. A., Sloan, J. J., Gains, D., Varon, D. J., Ramier, A., Strupler, M., and Tarrant, E.: The GHGSat-D imaging spectrometer, Atmos. Meas. Tech. Discuss. [preprint], https://doi.org/10.5194/amt-2020-301, in review, 2020.

Leifer, I., Roberts, D., Margolis, J., & Kinnaman, F. (2006). In situ sensing of methane emissions from natural marine hydrocarbon seeps: A potential remote sensing technology. *Earth and Planetary Science Letters*, *245*(3–4), 509–522. https://doi.org/10.1016/j.epsl.2006.01.047

Roberts, D. A., Bradley, E. S., Cheung, R., Leifer, I., Dennison, P. E., & Margolis, J. S. (2010). Mapping methane emissions from a marine geological seep source using imaging spectrometry. *Remote Sensing of Environment*, *114*(3), 592–606. https://doi.org/10.1016/j.rse.2009.10.015

Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., Raymond, P. A., Dlugokencky, E. J., Houweling, S., Patra, P. K., Ciais, P., Arora, V. K., Bastviken, D., Bergamaschi, P., Blake, D. R., Brailsford, G., Bruhwiler, L., Carlson, K. M., Carrol, M., … Zhuang, Q. (2020). The Global Methane Budget 2000–2017. *Earth System Science Data*, *12*(3), 1561–1623. https://doi.org/10.5194/essd-12-1561-2020

Schneising, O., Buchwitz, M., Reuter, M., Vanselow, S., Bovensmann, H., & Burrows, J. P. (2020). Remote sensing of methane leakage from natural gas and petroleum systems revisited. *Atmospheric Chemistry and Physics*, *20*(15). https://doi.org/10.5194/acp-20-9169-2020