Reviewer 2

This paper presents a synthetic study on the retrieval of methane plumes from satellites with high spatial resolution. This is a quickly developping area and a number of satellite (and aircraft) instrument have emerged that have succesfully demonstrated methane retrievals on a scale of tens of meters. Such observations will be important to detect and mitigate methane emisisons from localised emission sources. However, it is critical to put such methane satellite retrievals on a solid footing. This study addresses the question how well surface features and methane absorption can be seperated which is a key issue for instrument with lower spectral resolution. This is relevant for ongoing work with existing satellites but more importantly it provides guidance for the development of of future mission. The manuscripit is suitable for Atmos. Meas. Tech. and I recommend publishing it after addressing my comments below.

We thank the reviewer for the constructive comments and appreciate the thoughtful review.

Please find a point by point response below:

Figures: many figures in the manuscript are corrupted. This is probably simply an issue of the pdf conversation.

We apologize that this issue occurred. We did not hear of this problem from other reviewers, but will make sure that the figures are clearly shown in the publication, using only Vector graphics.

Instrument assumptions: The study provides a realistic model for the instrument and the measuerement noise calculation. The model makes uses of a number of instrument parameters given in Table 1. Can you please provide a justifcation of these assumptions. How does this compare to currently available systems and existing detectors. Is the assumption valid that the same parameters can be used for the two spectral range (1.6 and 2.3 micron): will detector quantuum efficienty, grating efficiency, spectral transmissivity/reflectivity of optical components not change between both ranges? Also, at 2.3 micron, I would assume that thermal emission of the optical bench will be a contributor to noise. Can you give some example values for dark current to support your assumption that this can be ignored. Finally, can you please clarify if noise has been added to the simulated spectra (Figure 8B suggest otherwise).

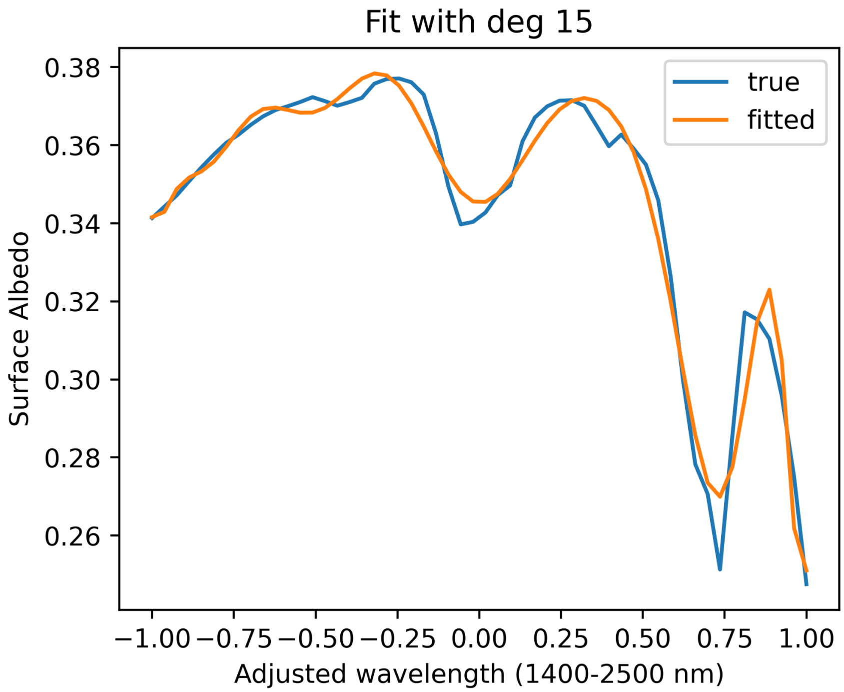
Most detector characteristics are in line with state-of-the-art detectors such as the Teledyne Chroma series (http://www.teledyne-si.com/products/Documents/CHROMA%20Brochure%20-%20rev%201%20v5%20-%20OSR.pdf). With cryo-cooling, dark current is very low and its impact on noise can mostly be neglected (also the thermal emissions from the optical bench). QE of the detector can indeed be wavelength dependent but the impact is small. Some publicly available can be found in https://www.teledyne-e2v.com/content/uploads/2018/10/ICSO_2018_Teledyne_IR_Sensors_PJerram_JBeletic.pdf (also consistent with assumptions made in the Strandgren paper).

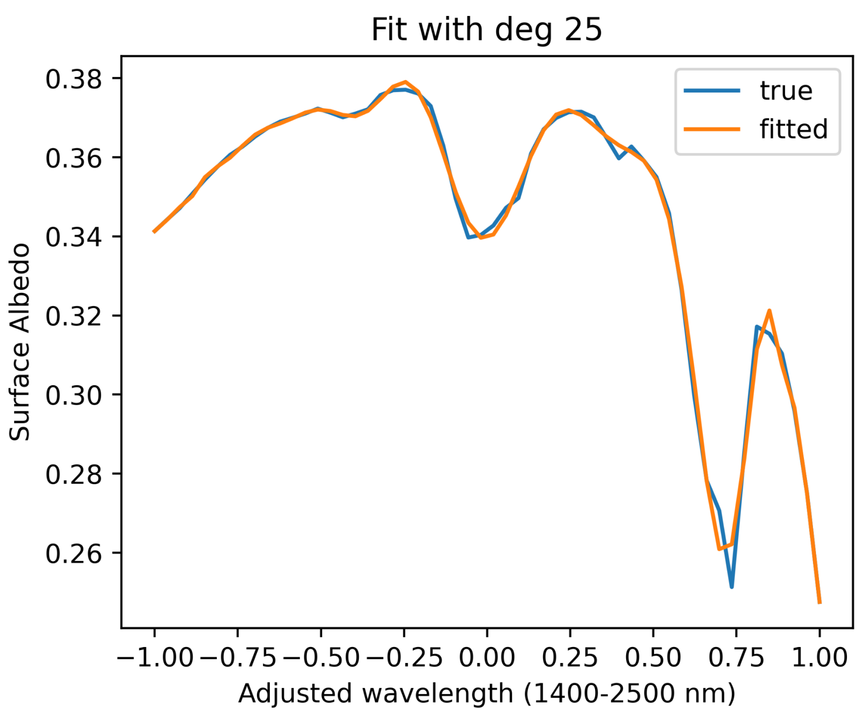
Noise has indeed been added to the spectra.

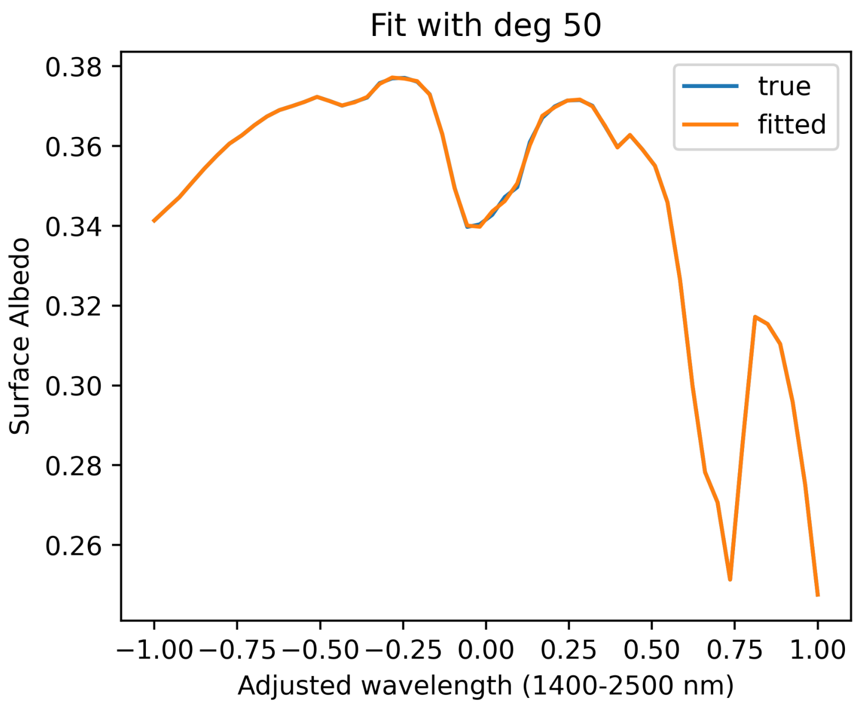
Surface features and polynomial degree: A key outcome of the study is the need for a very high degree of a polynomial to sufficiently accuratly describe surface features. However, the use of a polynomial of degree 50 makes me uneasy. Can you show with a direct polynomial fit to the underlying surface albedo data of the ECOSTRESS spectral library before using it in your forward model and without any spline interpolation that such a polynomial degree is needed? I would also expect that a high polynomial degree will lead to an increased number of non-converging retrievals when not carefully choosing their a priori value and a priori covariance; can you please elaborate on your choice. As you show in the paper, a high polynomial degree will increase the retrieval uncertainty for methane. At the same time the correlation with methane will increase so that you risk that the methane absorption will be taken out by the polynomial. Did you have a look at the correlation coefficients ?

We applied a polynomial degree of 50 as an illustration of how it could change bias and precision error. Over most surfaces, a polynomial degree of 25 seems adequate to significantly reduce the retrieval error (both bias and precision error combined).

The polynomial fits for an example surface (construction concrete) are shown in the following figures. The surface albedo in this demonstration is without any spline interpolation. Legendre polynomials were used and the range of wavelength is 1400-2500 nm. Even when the x-axis is adjusted to be in between -1 to +1 to make the fit easier, we clearly observed that high degrees such as 25 or 50 were needed to capture the surface albedo variations.


Fit with deg 15
0.38
0.36
0.34
0.32
0.30
0.28
0.26
Surface Albedo
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
Adjusted wavelength (1400-2500 nm)
true
fitted


Fit with deg 25
0.38
0.36
0.34
0.32
0.30
0.28
0.26
Surface Albedo
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
Adjusted wavelength (1400-2500 nm)
true
fitted



In our experiments, we set a prior value in our polynomial degree coefficients as 0.5, -0.1 for the first two degrees and 0.0 for the remaining degrees, and we use a loose prior covariance (1e20) to have no significant impact on the retrieved total columns or posterior errors. Based on this setup, the non-converging retrievals did not arise in our experiments.

Minor comments and typos: I have included them directly in the supplementary pdf.

We make edits throughout the text in response to the minor comments accordingly.

Fast detector: Enabling exposure times <50ms of a large focal plane array (to allow high spatial resolution, unlike say TROPOMI, which can integrate for 1s, which makes the readout easier as well)

And the unit of readout noise of 100 electrons.