Review on “Exploiting the entire near-infrared spectral range to improve the detection of methane plumes with high-resolution imaging spectrometers”.

The authors present an augmented version of the matched filter retrieval. The matched filter is employed in many studies for the identification and quantification of methane enhancements in hyperspectral data, commonly from airborne instruments in a top-down viewing geometry. Studies generally exploit methane absorption features in the shortwave infrared (SWIR) spectral range around the 2300 nm range for the algorithm. Yet, since the surface albedo may vary strongly in the SWIR, surface reflectance structures can align with the absorption features of methane spectrally and thus, cause artifacts in the produced image. These artifacts pose challenges to plume detection algorithms. Many state-of-the-art instruments observe wider spectral intervals in the SWIR. The authors present a technique to exploit the extended range between 1000 and 2500 nm to reduce image artifacts due to background heterogeneity. They evaluate the techniques performance on simulated and real data and compare it to similar retrieval adaptations from previous publications.

Monitoring anthropogenic methane emissions is an integral element of human efforts to mitigate climate change and has received particular attention in recent years. The matched filter retrieval is widely used in the scientific community due to its simplicity and efficiency. Retrieval artifacts have been identified as a major source of errors in the past, thus the findings of this paper are timely and scientifically significant. The performance of the adapted matched filter technique is impressive and achieves the goal of improving signal identification. Furthermore, it is applicable to many ongoing and upcoming satellite imaging instruments.
The applied methods are rigorous and valid, and the experiments support the findings. Yet, a quantification of some results is necessary to put them in context regarding applicability. Also, some open questions remain to be addressed in the review process.
The overall presentation quality is acceptable but should be improved during the review process. Some paragraphs are structured confusing, and findings are introduced before evidence of any kind was presented. Many paragraphs include colloquial language and need to be worded more concise.

Please find detailed comments below.

Major comments:

0. Abstract:
    a. Needs to be rephrased according to review changes.
1. Introduction
2. Materials and methods
    a. L111: The potential of the matched filter to use spectral information from distant bands to remove artifacts is (a) a conceptual difference to physical-based methods, which cannot do this, and (b) the main reason why the method presented in this paper works. Please dedicate more space to explain the idea and how information is communicated from the newly integrated bands via the covariance matrix of the scene.
    b. L122ff: The following paragraphs use results of the paper without stating this explicitly, such that they appear as claims without evidence. I believe the authors try to explain the behaviour of their adapted matched filter, but without the evidence/experience of the result section, these paragraphs are confusing. I try to go through them in detail.
    c. L122: “Although SWIR-MF reduces retrieval artifacts [..]” – has not been shown yet.

d. L129: "[…], which leads to a lower weight of the background variance direction." – I do not understand this, what is a weight in this context and how does it affect the retrieval performance? Are you referencing to the $C^{-1}t$ – vector?
e. L130: "This difficults the discrimination between plume pixels and background pixels, which can lead to enhanced background clutter and plume pixel attenuation." – I don't understand in detail why a smaller correlation between distant bands removes information from the filter. The information of the closer bands is still included, and also figure 6 shows that the SWIR-MF has a smaller variability.
f. Figure 4: Please add axis labels.
g. L140: How do you arrive at the conclusion that co-emitted carbon dioxide (a) affects the retrieval and (b) decreases the methane columns? Might this be mitigated by an exclusion of the carbon dioxide feature bands from the matched filter?
h. L142ff: I found the following paragraph very hard to understand.
As I understand it, the SWIR-MF produces overall smaller enhancements than the 2300-MF. This is especially true for 2300-MF artifacts and clutter regions. There are exceptions, in which the SWIR-MF will produce enhancements greater than the 2300-MF, which can be attributed to spurious correlations in the additional 1700 nm methane band. These exceptions are dealt with by simply replacing them with the 2300-MF clutter retrieval values, which results in the COMBO-MF. In order to keep the variances of the retrievals comparable, you introduce the scaling factor *f*. This procedure removes artifact from both retrievals and large clutter values of the SWIR-MF, thus increasing the contrast between background and plume pixels.
If my understanding is correct, consider rewording the paragraph along these lines. Also, I strongly suggest writing down a formula to make the procedure clear ($\alpha = \Delta$CH4):

$$\alpha_{combo} = \begin{matrix} \alpha_{swir} * f & if\ \alpha_{swir} < \alpha_{2300} \\ \alpha_{2300} & if\ \alpha_{swir} \geq \alpha_{2300} \end{matrix}$$

i. L153: What is a standard deviation here? Please give a definition.
j. L151: If you use the SWIR-MF as the more trustworthy retrieval and it has a smaller variance, why don't you scale the 2300-MF values before including them?
k. Figure 5: I believe the conditional comparisons are exchanged. Please include a formula for the technique, then I would suggest removing this figure.
l. The condition of the COMBO-MF will remove plume pixels which are detected in the 2300-MF but remain undetected in the SWIR-MF. Can you comment on if you could observe this behaviour, and if it might cause systematic biases for emissions estimates?
m. L206: The masking criteria is a bit strange. Why do you include your knowledge about the synthetic plume here? This removes partly the effect of artifact suppression on the plume mask, right?

3. Results
a. L256: As I understand it, the GF5-02 dataset includes a real CH4 plume. If that is the case, move it to section 3.2.
b. L56ff: Is the whole argument in section 2 about the matched filter underestimating CH4 in the case of co-emitted CO2 based on this result? If yes, consider supporting it further with e.g., the retrieval of synthetic co-emitted plumes. It feels you highlight this finding a lot, therefore it needs more evidence.
c. L288ff: I agree that the noise suppression works fine in this example, but the visual plume detection is an insufficient argument here. It would be much more robust if you used a plume detection algorithm and attributed the identified plumes to coal mine

ventilation shafts on the ground. Do you have the locations of the shafts, or how do you know that the identified plumes are indeed methane emissions?

4. Summary and conclusions
    a. Needs to be rephrased according to review changes.

Minor comments:

0. Abstract:
    a. You employ a lot of passive speech, consider wording your sentences in an active way. This way, the reader knows who/what performs a given task.
1. Introduction
    a. l50ff: This paragraph motivates the whole publication, please add some references which highlight the necessity, e.g., https://doi.org/10.1016/j.rse.2018.06.018.
    b. L36ff: Please make clear to which instruments you compare the hyperspectral imagers – they do not have a high spectral resolution objectively.
    c. L52: “raw” matched filter is not defined. Description required.
    d. Figure 1: “[…] resampled to 2 nm spectral *resolution*”. Also put “wavelength” on x-axis as long as lambda has not been introduced.
    e. Figure 2: What does “detected” mean in the context of this figure? How can you be sure all other detections are artifacts?
2. Materials and methods
    a. L65ff: Unit absorption spectra generation is not well described. Clearly state that you collect absorption cross-sections from HITRAN and use them in a simple lambert-beer transmission model, which accounts for observation and solar angles.
    b. L71: As far as I know, HITRAN does not calculate radiative transfer, therefore the scattering-sentence is confusing. Furthermore, you cannot state that scattering is negligible in the SWIR. Aerosol impact on retrievals is still ongoing research, and clouds most definitely have an impact on SWIR RT. Please formulate your assumption of a pristine atmosphere to back this simplified radiative transfer model.
    c. L90: This sentence belongs to the unit absorption spectrum paragraph.
    d. L65-L98: These paragraphs give a broader overview of the matched filter, bus are not sufficient to explain the method for an untrained reader. It is out of the scope of this paper to explain the matched filter in detail but add an introductory sentence in the beginning which redirects to explanatory literature would be appreciated.
    e. L111: “should be” is overused in this paper. It undermines the reliability of the findings, consider wording your findings more confidant if you are, or remove the sentences if it is speculation.
    f. Figure 3: Please mention in the caption that this is a 2-band matched filter.
    g. L183: Add a sentence on how the plumes are implemented in the dataset.
    h. L196: U_eff depends logarithmic on U_10
    i. L203: What is a standard deviation in this context?
    j. L215ff: Please mention which of the datasets include real methane enhancements and which of the are free from real sources, such that they may be used for the synthetic studies.
    k. L220: “Moreover, [..]” – add references to figures 3 and 4.
3. Results
    a. Figure 7: A zoom to the plume would help to show the effect on plume shape. Optional comment.

b. Figure 8: The distributions of the data points seem to be highly asymmetrical for smaller fluxes, since the uncertainty ranges reach to the negative. Could you make a boxplot for each flux?
c. L244: “lower” – please provide numbers
d. L266: You already commented on the “upper limit” in the text and in figure 10. If this is only a discussion about the colorbar-ranges, please limit it to the figure caption. If you have introduced an upper limit to one of the methods, this needs much more explanation.
e. L275: The last sentence seems to be connected to the unit absorption spectrum generation, please clarify this.
f. L284: “somewhat attenuated” → quantify, might this be due to my comment 2.l ?

4. Summary and conclusions
   a. L299: explicitly name the instrument from which the data was taken.
   b. L300: “different masking methods” sounds like you did a performance analysis. Make clear that results are either based on a thresholding approach or visual identification.
   c. L316: You could highlight that the improvement is especially prominent for small fluxes, which are typically hard to identify.

Technical corrections:

0. Abstract:
1. Introduction
   a. L31: “[..] come*s* from [..]” – missing ‘s’
   b. L37ff: “we find” is colloquial English, please use scientific language.
   c. L45ff: whole paragraph is colloquial English.
   d. L46: “[..] its interactions and the media *that travels* through.” – grammatical error.
   e. L58: “elaborated” – use present tense, e.g. “In this work, we present […]”
2. Materials and methods
   a. L64: “methods” not capitalized in heading
   b. L77: introduce vector notation from beginning, please write “spectral mean vector” at the first occurrence and type all vectorial quantities bold.
   c. L98: “in a per-column basis.” → “for each along-track column separately.”
   d. L101: citation order from oldest to youngest paper.
   e. L115: *a* scatter plot
   f. L116: a*n* EnMAP data set
   g. L128: scattered → wide-spread
   h. L130: difficults → complicates
   i. L133: not removed → additional
   j. L158: remove “as” before “COMBO-MF”
   k. L191: should not → will not
   l. L210: s*u*rroundings
   m. L279: we can observe → shows
   n. L280: “here there are” is colloquial, please rephrase
   o. L283: “the positive values” – unnecessary *s*
   p. L284: “migh” → might
3. Results
4. Summary and conclusions
   a. L304: not-simulated → real