We thank the reviewers for their thoughtful comments, which we have addressed below. All page and line numbers refer to those in the revised manuscript. Reviewer comments are in *italics*, our response is in plain text, and text in the revised manuscript is in blue.

Response to Comments from Anonymous Referee #1

1. *In section 4.1, one discusses the results presented in Figure 3. Although one can "see" the plume in the retrieved images (center and right) for the homogeneous scene when one knows it is there, I am not convinced that an uneducated guest would detect the plume without a significant number of false detection. It seems rather clear that, if the source was 100 kg/h (and not 500 and 900 kg/h) as in the simulated images, the signal would be hardly distinguishable for the noise. Thus, the claim that one would be able to detect and quantify plumes from 100 kg/h source is definitely not founded.*

We thank the reviewer for this helpful comment. We derive emission rates for each EeteS plume and describe the results in a new table, Table 2. We move the discussion of IME/emission rate derivation to Section 3, and give it its own section, Section 3.3.

We add the following analysis in the text:

Page 10, Line 343: "We examined the ability of the retrievals to quantify methane point source rates on the basis of the detected plumes, by applying the IME algorithm of Section 3.3 to the same ensemble of 5 WRF-LES plume realizations for each of the three different surfaces and for true source rates 100, 500, and 900 kg h^{-1}. Results are summarized in Table 2. We find that it is possible to quantify source rates as low as 100 kg h^{-1} for the Bright scene, and as low as 500 kg h^{-1} for the Grass scene, though the true source rates are underestimated by up to a factor of 2. There could be several factors behind this underestimate including (1) error correlation with surface reflectivity in the EnMAP retrieval that would cause some loss of the plume, and (2) use of the Varon et al. (2018a) U_{eff} -U_{10} relationship in equation (10) without customization for the EnMAP conditions. As pointed out by Varon et al. (2018a), the U_{eff}-U_{10} relationship should be customized to the plume mask definition and to the instrument pixel resolution and precision. This would require an ensemble of WRF-LES simulations specific to the EnMAP conditions and to the plume mask used here. The inability to quantify the 100 kg h^{-1} plume over the Grass scene is properly diagnosed in our retrieval by the failure of the plume mask to detect a plume. However, the surface artifacts in the Urban scene lead to spurious retrievals of source rates as the surface features are mistakenly attributed to plumes. This is due to the error correlation between X_{CH4} and surface reflectivity (explained in greater detail in Section 4.2) and can be diagnosed by inspection of the off-diagonal terms of $\hat{\mathbf{S}}$ (Equation 7)."

2. *Lines 229-230, it is said that the "8% precision [...] should enable EnMAP to successfully quantify 500 kg/h point sources in a single pass." There is no attempt at estimating sources in this section, so that there is no ground for this claim*

See response to comment #1.

3. *Line 235, it is said that, for a 900 kg/h source, the plume is "well defined against the background" which is an overstatement.*

We soften the language:

Page 9, Line 319: "The 900 kg h^{-1} plume is better captured over both surfaces, though major retrieval artifacts remain in the Urban scene."

4. *Line 284 "but a source rate can still be estimated successfully with EnMAP". There is no ground in the paper for that statement.*

See response to comment #1.

5. *Line 323 : "Nevertheless, the results do confirm that EnMAP should be able to detect plumes and quantify source rates down to ~ 100 kg /h". The analysis of the airborne data show overestimates by a factor up to 3 (mean 2). How can one see that as a confirmation that the source can be quantified?*

We clarify that the underestimate was confirmed by both assessments:

Page 12, Line 435: "The EnMAP underestimate is consistent with the results in Table 2 and may reflect the same sources of bias, in part correctable through an improved U_{10}-U_{eff} relationship. The results confirm that EnMAP should be able to detect plumes and estimate source rates down to ~100 kg h^{-1} when the scene is sufficiently bright."

6. *In the conclusion it is said that the space measurements can be used to "detect and quantify plumes of magnitude ~100 kg/h over relatively bright surfaces". Yet, the simulations have been performed with larger sources (factor 5 to 9). In addition, it is rather ambiguous whether the objective is to quantify the plume (and what that really means) or to quantify the source that generate it. This should be clarified.*

Thank the reviewer for this point and clarify in the text.

Page 13, Line 474: "We showed that these EnMAP-like images are able to detect actual plumes of magnitude ~100 kg h^{-1} over relatively bright surfaces. Source rates inferred from the plumes with a generic Integrated Mass Enhancement (IME) method are a factor of 1.2 to 3 lower for EnMAP than for AVIRIS-NG, which could be due in part to unaccounted dependence of the IME method on instrument pixel size and precision. This should be improved in further work by customizing the IME method to the EnMAP specifications."

7. *In addition, one major source of uncertainty for instrument with a "low" spectral resolution is the knowledge of the instrument response function. I understand that the authors have assumed that this response function is perfectly known. It would be nice to add a sensitivity test to analyze the impact of some uncertainty on this important parameter. To the very least, they should mention and discuss the potential impact.*

We clarify the importance of spectral calibration and include spectral shift in the retrieval:

Page 6, Line 182: "We also correct for uncertainty in the instrument's wavelength calibration with a spectral shift parameter (Thorpe et al., 2017; Frankenberg et al., 2005).

We give more information about EnMAP's spectral calibration:

Page 6, Line 230: "EnMAP has strict requirements of 1 nm spectral calibration accuracy and 0.5 nm spectral stability in the SWIR. Pre-flight calibration campaigns as well as onboard calibration means will be used to ensure the compliance with those requirements (Guanter et al., 2015)."

8. *Also, the paper uses a method for plume mask through "median and Gaussian filters" which is not described. Some sentences do describe the principle of the method would be useful.*

We clarify the purpose of the filters in the text:

Page 8, Line 289: "These filters help to remove spurious signals surrounding a plume and determine the spatial extent of the plume, which is needed for subsequent calculations"

9. The reviewer included many annotated comments directly on the manuscript. We update accordingly:

"livestock operations may not be point sources" "livestock operations may not be point sources"

Page 2, Line 42: "Anthropogenic emissions originate from a very large number of point sources (coal mine vents, oil/gas facilities, confined livestock operations, landfills, wastewater treatment plants) that are individually small, spatially clustered, temporally variable, and difficult to quantify (Allen et al., 2013; Frankenberg et al., 2016)"

"I assume "true" point sources, so that not like land fills for instance"

Page 5, Line 168. "This range is typical of large (but not unusually large) individual point sources (Jacob et al., 2016)."

"Not clear to me [reference to Page 5, Line 135 in original manuscript"

Page 5, Line 180. "We do not add noise or aerosol effects to the plume transmission spectra because the EeteS scene already accounts for those in the computation of back-scattered radiances, so that multiplying by the additional plume transmission already factors in the corresponding noise."

"The retrieval procedure assumes that the instrument spectral response is perfectly known ? Please state so and discuss the resulting uncertainty"

See response to comment #7.

"I do not see this parameter in the equations. Unit ? [in reference to Page 8, Line 203 in original draft]"

Since it the variance in a scaling factor, it is unitless. We clarify how it enters Equation 6:

Page 7, Line 277: $\mathbf{S_A}[1,1] = \sigma^2_{S_{CH4}} = 5$ (unitless)

"I would say these are rather optimistic comments with respect to the impression given by the figure. [in reference to Page 8, Line 217 in original draft]"

See response to comment #1

"??? There is realy no ground for this statement. One has no idea when "successfully quantify" means here. [in reference to Page 9, Line 229]"

See response to comment #1

“Rather optimistic to me [in reference to Page 9, Line 235]”

See response to comment #3.

“Not clear what the procedure is [in reference to Page 10, Line 279]”

See response to comment #8.

“how do I know that ? [in reference to Page 10, Line 284]”

See response to comment #1

“One finds source that are up to 3 times larger than the truth, and this is a confirmation that one can quantify source rates ?”

See response to comment #5

Response to Comments from Gerrit Kuhlmann

1. *The authors use the (relative) root mean square error (RMSE) for evaluating the precision of the methane retrieval. However, the RMSE is the sum of accuracy (mean bias) and precision (variance) RMSE = sqrt(MB^2 + Variance) and thus the analysis of the precision is potentially biased by the mean bias the retrieval. The mean bias might be caused by the strong dependency surface reflectance as discussed by the authors that apparently results in increased XCH4 as seen in Figure 3. Consequently, the author should not use the term "precision" as synonym for the RMSE as done in the text and in Figs. 4 and 5. The authors also need to check how much the computed RMSE is affected by a mean bias and variance and revise their results, discussions and conclusions accordingly. Using the variance will make the results better comparable with the a posteriori retrieval noise (second method), even if the latter is of course not affected by other (random) error terms in the retrieval.*

We thank the reviewer for this insightful comment. We switch to using just the relative residual standard deviation for precision estimates instead of RRMSE and theoretical precision.

Figures 4 & 5 updated

Page 9, Line 323. “Here we characterize the EnMAP instrument precision as the relative residual standard deviation (RRSD) between the true and retrieved column methane concentrations for individual 30 × 30 m^2 pixels in the scenes of Figure 2 including the WRF-LES plumes. Figure 4 summarizes the results for the four scenes of Figure 2. We find precisions of 3.5 ± 0.07% for Grass, 7.2 ± 0.1% for Urban, and 2.6 ± 0.08% for Bright scenes.”

We address how bias is not as important with a proper background definition:

Page 2, Line 57. “Bias may not be an issue if the plume enhancement is referenced to the local background.”

2. *The authors consider SNR of the instruments and other errors included in the EeteS simulator, but assume precise knowledge of wavelength positions. However, inaccurate spectral calibration is a potentially large error source not considered in the study. A further potential error source for the CH4 retrieval are radiometric calibration errors that can result in (systematic) high-frequency patterns in the spectra. The latter could in particular be a problem for instruments where the main application is not influenced by such high-frequency patterns. The authors should therefore discuss these limitations in their study and mention possible recommendation for the instrument developers, e.g. characterization in the lab, to make their instrument more suitable for measuring methane.*

See response to comment #7 from Anonymous Reviewer #1.

3. *P3, L61 and P10, L266: Please provide (rough) numbers of "most" and "majority of anthropogenic point sources".*

See response to comment #9 from Anonymous Reviewer #1.

4. *P6, L146f: Please specify what you did here. Applying a Gaussian filter with 10.0nm FWHM to AVIRIS-NG spectra with 5.0 nm FWHM would result in a spectral resolution of 11.2 nm FWM.*

We thank the reviewer for this point and clarify confusion in our workflow:

Page 6, Line 199: "…and further convolved these spectra with the appropriate Gaussian filter to match EnMAP spectral resolution and wavelength positions."

5. *P7, L183ff: Since this seems to be the first time that Legendre polynomials have been used in a DOAS analysis, it is probably worthwhile to provide some additional information here.*

Page 7, Line 249: "Orthogonal polynomials can potentially constrain surface reflectance with fewer terms, leading to better conditioning of the inverse solution"

6. *P7, L190f: Please explain why you are testing separated convolutions <*>. I assume this is due to the following inequality: <I0 * exp(-tau) > \ne <I0>*<exp(-tau)> (Frankenberg et al. 2005, Eq. 16).*

We add motivation for this analysis:

Page 7, Line 256: "Since the convolution operator is not linear (Frankenberg et al., 2005), …"

7. *P8, L223f: Please add parenthesizes, e.g.: (8.2 ±0.7)*

We keep as is because the reported numbers are objects of the preposition in the sentence.

8. *P11 L312: Varon et al., 2018 -> Varon et al., 2018a*

Fixed

9. *Table 1: It might be better to use the term "undefined" (or something else) instead of "TBD" which is quite ambiguous.*

We change the entry in Table 1 to read “Undefined”