Manuscript: "Spectral Sizing of a Coarse Spectral Resolution Satellite Sensor for XCO2"
by JS Wilzewski et al.

Reply to interactive comment by anonymous reviewer #3

We thank the reviewer for the helpful comments to our manuscript. Below we repeat the reviewer's questions in **bold** font and subsequently provide our responses.

General Comments

Because these spectrometers will be for local-scale (power plant, urban scale) domains, the global-scale performance of individual GOSAT 10x10 km^2 really is only a starting point. It would be important to model the potential behavior of such a satellite using an OSSE (Observing System Simulation Experiment) over high-resolution, simulated local-scale domains. The authors should add a (potentially short) discussion of this limitation to the paper.

We agree that most of the analyses performed in this work are just a starting point towards evaluating a possible future CO_2 sensor. It will certainly be crucial to carry out detailed simulations of this proposed spectrometer with a thorough discussion of the actual instrument design and noise performance for representative local-scale domains. In the present manuscript we focus on investigating whether a coarse resolution, single band observation configuration could generally deliver sufficient information such that a meaningful retrieval of XCO_2 can be made. Details of the satellite sensor shall be studied in a forthcoming study, currently under preparation in our group.

We added "A forthcoming study addressing these aspects of the proposed sensor is currently under preparation." in the manuscript on page 18, line 8 – 9.

I have a methodological question as follows. In terms of taking real GOSAT data, and simply convolving it with a wider ILS, it seems like the SNR of the resulting measurement (with 256 channels per band) will be higher than one may actually be able to build in a realistic instrument. For instance, I performed a simulation of simple white noise for 1300 GOSAT channels spaced every 0.2 cm-1 (the approximate channel spacing for GOSAT) between 4740 and 5000 cm-1, and had a starting SNR of 700. In the simulation, when I convolved the spectrum (with realistic noise added) with a Gaussian ILS with FWHM=1.3 nm, the resulting SNR was ∼3400. This was due to the averaging effect of the

hi-resolution GOSAT data.

The authors do state (section 2) "Since we want to isolate the effects of spectral resolution and spectral band selection, we do not add extra noise to the convolved spectra." However, they are worried here about the effect of smaller ground pixels. BUT, it seems they are not taking into account this averaging effect 'beating down' the native GOSAT noise to unrealistically high SNR values. Here, the final SNR value of 3400 is NOT equal to the GOSAT value of 700, so I think they are not purely "isolating the effect of spectral resolution" since the SNR values are wildly different. Did the authors examine the resulting SNR of their low-resolution GOSAT measurements, and are they in line with what they would expect from their hypothetical instrument? I realize they somewhat avoid this question by not having a real instrument noise model proposed, but as written, the results may be misleading because they may assume unrealistically high SNR values for any possible instrument. The authors should discuss this point and make it clear. Also, this could be rectified by proposing a realistic instrument noise model, and then ADDING noise to the GOSAT spectrum after convolution with th Gaussian ILS, in order to obtain an SNR in line with a more realistic value.

It is evident that the effect of convolving the native GOSAT spectra with a wider ILS (sampled by 3 detector pixels) results in higher SNR per pixel for the setup with reduced spectral resolution then for the native configuration. And, indeed, we do not compensate for the "beating down" of the noise by adding extra noise (as mentioned by the manuscript).

Adding noise to the spectra would introduce an additional artificial element (besides the coarse ILS) to our analysis. We want to stick as close as possible to real measurements and, as stated by the manuscript, we want to isolate the one effect (i.e. coarse ILS).

Errors in native GOSAT retrievals are not dominated by noise. In fact, an SNR of 700 (at the radiance continuum?) as assumed by the reviewer is by far better than the observed spectral fitting residuals which are for the most part dominated by systematic patterns (unresolved scattering effects, spectroscopic errors, unaccounted instrument characteristics). Likewise, the noise errors on retrieved XCO2 are typically a factor 2-4 smaller than the standard deviations found when comparing to validation data (see also our figure 12). Thus, for a GOSAT-like setup, noise is a minor contributor to the errors. The noise for convolved and unconvolved spectra might be "wildly" different, but, for both, it is small compared to other sources of error. Accepting that the noise is small makes it straightforward to evaluate these other sources of error e.g. through the parameter correlations shown in the paper, which we chose to be the focus of the present paper.

Adding noise to the spectra to mimic a new sensor with fine ground resolution would result in a different paper. We would need to discuss the instrument optical and electronic setup and describe the noise model. Such a paper is in preparation including a noise evaluation with simulated data. The present paper aims at discussing whether it is reasonable at all to try out a coarse-spectral-resolution configuration.

Essentially, our results are representative under circumstances where the noise can be assumed small compared to other sources of error. The next paper will address how to build the instrument, for what scenes noise is indeed negligible, and what to expect if noise becomes large for dark surfaces. To make these aspects clear, we add the following paragraph to the manuscript:

"Our approach essentially relates to conditions under which the detector noise is negligible as typical for GOSAT. Under such conditions, other sources of error can be addressed e.g. through evaluating geophysical parameter correlations (section 3 and 4). A forthcoming study will discuss noise performance and retrieval simulations for a hypothetical instrument design." (page 5, line 20 – 23).

Another concern is the impact of not using the O2A band. The authors should discuss the feasibility of seeing power plant plumes in the face of realistic pointing errors, and if the pointing will be sufficiently good such that surface pressure estimates from meteorological reanalysis, hypsometrically adjusted to account for the local topography, will be a relatively small error or not.

As the proposed sensor will have imaging ability, the spectrometer shows promise to have a good pointing knowledge. Any errors in pointing may be 'recalibrated' when scenes with prominent surface reflectance features, such as shorelines, etc., are observed. We expect that even if pointing accuracy is low, one would be able to obtain a good correction in order to correctly calculate airmass for the XCO_2 retrieval.

We added "Errors in the calculation of the airmass can be caused by erroneous satellite pointing; these errors are part of the overall errors reported for the TCCON validation sites (section 3)." (page 7, line 26 – 27).

A critical concern is the ability to properly filter the data. For many XCO2 retrievals, cloud and aerosol filtering is a critical component of any retrieval system, yet this is completely left out of this analysis as the authors start with data pre-filtered using the native GOSAT 3-

band retrievals. It is therefore not clear how robust the conclusions would be if the sensor had to solely rely on filtering from a single, low-resolution SWIR band. While this study is a good start, results from a proper simulation-retrieval experiment including the effects of clouds & aerosols and the role of pre-filtering is of critical importance to realistically judge if such a simple sensor could truly determine power plant emissions.

We would have liked to analyze the impact on cloud-screening, however, due to computational costs, we could not. It should be pointed out that the SWIR-2 configuration, which is favored for the future instrument, has two CO_2 absorption bands with very different optical depths, which opens up an avenue to set-up a cloud filter using the SWIR-2 window alone. By retrieving XCO_2 from both CO_2 bands, one could filter for large discrepancies caused by the presence of clouds. This is a variant of the cloud filter currently used for the native GOSAT soundings. The actual implementation and verification of this approach must be postponed to a future study.

We now mention in the paper that we did not carry out a cloud filtering exercise on page 5, line 8: "Due to computational costs, we restrict our analysis to cloud-free, quality screened soundings over land as identified by the native GOSAT retrievals of the RemoTeC algorithm...".

Also, we added a discussion of the SWIR cloud filter option in the discussion (page 18, line 4 – 7): "Additionally, the SWIR-2 seems better suited for the construction of a cloud filter, because its CO_2 bands have very different optical depths. Similar to the cloud filter currently in use for GOSAT measurements, one could retrieve XCO_2 from the two SWIR-2 bands individually and filter for discrepancies. This scheme should be tested in the future."

Specific Comments

P5L20: You assume 256 spectral channels in a single band. This seems like a high oversampling rate ($\sim$3 for both SWIR-1 and SWIR-2), considering that there are roughly 86 fully independent spectral samples in each band, given your proposed resolving powers. This rate appears to have been carefully chosen. Please speak to any knowledge you have on the importance of the spectral oversampling, as it may be an important consideration (for SNR or retrieval accuracy/precision). I just noticed this is also discussed on page 9, but the factor of 3 oversampling is again assumed there, and not questioned or discussed as any kind of instrument parameter to be optimized (in the way that spectral resolution is, in this study).

We have assumed a spectral sampling ratio of three throughout this work. A sampling ratio of 2 would be the lower limit according to Nyquist's theorem. Generally, the higher the sampling ratio, the better. Detectors with a very high number of pixels (e.g 2000 pixels) could enable a significantly higher sampling ratio. Yet, previous space-based CO_2 missions have been successful by spectrally over-sampling the FWHM by a factor 2-3 (e.g. GOSAT, OCO-2, OCO-3, TanSat). Thus our choice of sampling ratio is based on what is currently in use for similar sensors.

P6L17. The improvement of your 3-aerosol-parameter retrieval vs. a non-scattering retrieval is curious, consider the extremely low DFS for aerosol you cite (0.38). It therefore seems possible that your results may be sensitive to the prior assumption on aerosols. How are the aerosol priors for the 3 parameters chosen, and did you test your sensitivity to the aerosol prior, given the low DFS?

Given that the retrievals estimate 3 aerosol parameters with little DFS, the retrievals, by definition of DFS, depend on the a priori. We have conducted a sensitivity study how various aerosol priors map into XCO2 errors. As prior aerosol we had selected reasonable numbers for scattering optical depth (τ=0.1), scattering layer height (z_{par}=3000 m) and size parameter (α_{par}=3.5) throughout the study. These values are routinely used as prior for GOSAT retrievals with RemoTeC. Table 1 shows the changes in scatter around TCCON as well as the changes in correlation coefficients for SWIR-2 retrievals at 1.29 nm resolution with changed aerosol priors. As we only have ~0.4 degrees of freedom to be distributed to the fit of three aerosol parameters, it is clear that the aerosol prior can have an impact on retrieval performance.

We find that our results are moderately sensitive to small changes in τ or z_{par}, while larger variations in the prior have a big impact on XCO_2 retrieval performance. For instance, changing τ by a factor 2 or $\frac{1}{2}$ leads to relatively small deviations from our benchmark SWIR-2 retrieval regarding scatter around TCCON and geophysical correlations on a global scale. We observe σ_{TCCON} =3.19 ppm for τ=0.05 and σ_{TCCON} =3.50 ppm for τ=0.2. Correlation coefficients to albedo and other geophysical parameters (as in Fig. 11 of the manuscript) are collected in Table 1. Changes in retrieval performance also occur for small changes in the initial scattering layer height. For prior layer heights of 1000 m and 5000 m, the standard deviation of SWIR-2 retrievals around colocated TCCON data amounts to 3.32 ppm and 3.71 ppm, respectively. This indicates a stronger dependence on scattering layer height priors than on optical depth priors. A significant change in retrieval performance occurs for a prior aerosol scenario, where rather large scattering particles are placed at the top of the troposphere (τ=0.07 z_{par}=11600 m α_{par}=3.67). In this case, σ_{TCCON} =4.14 ppm is higher than for all other aerosol prior options we studied here.

As a result, extreme prior aerosol values have to be avoided for our retrievals. This sensitivity study shows that the retrieval performance of the proposed sensor may be enhanced by a few tenths of a ppm by using a good aerosol prior. An additional aerosol sensor would help to inform and optimize the retrievals.

We added "An investigation of the impact of the aerosol priors on retrieval performance showed that SWIR-2 XCO_2 is only moderately sensitive to the aerosol priors. For instance, varying aerosol prior optical depth by a factor of two or one half resulted in small changes in standard deviations around TCCON (+0.22 ppm and −0.08 ppm, respectively). Changing scattering layer height priors to z_{par}=1000 m or z_{par}=5000 m increased scatter around TCCON by +0.04 ppm and +0.43 ppm, respectively. Similarly, scatter around TCCON changes by +0.22 ppm and −0.05 ppm if α_{par} is set to 3.0 and 5.0, respectively." in the manuscript (page 10, line 20 – page 11, line 2).

Aerosol prior	σ_{TCCON} / ppm	R(albedo)	R(SOT)	R(N_{par})	R(z_{par})	R(α_{par})
τ=0.1 z_{par}=3000 m α_{par}=3.5	3.28	-0.18	0.26	-0.5	0.21	0.01
τ=0.07 z_{par}=11600 m α_{par}=3.67	4.14	-0.48	0.17	-0.5	0.17	0.09
τ=0.05 z_{par}=3000 m α_{par}=3.5	3.19	0.04	0.31	-0.45	0.21	-0.07
τ=0.2 z_{par}=3000 m α_{par}=3.5	3.50	-0.30	0.21	-0.50	0.19	0.07
τ=0.1 z_{par}=1000 m α_{par}=3.5	3.32	0.20	0.31	-0.4	0.19	-0.11
τ=0.1 z_{par}=5000 m α_{par}=3.5	3.71	-0.36	0.21	-0.47	0.20	0.06
τ=0.1 z_{par}=3000 m α_{par}=3.0	3.42	-0.24	0.23	-0.49	0.20	0.07
τ=0.1 z_{par}=3000 m α_{par}=5.0	3.23	-0.05	0.3	-0.5	0.2	-0.09

Table 1: Comparison of the effect of different aerosol priors on standard deviation of retrieval results around TCCON ("σ_{TCCON}") and on the correlation coefficients ("R(X)") with respect to geophysical parameters (albedo at 2.1 nm, SOT, particle amount, scattering layer height and size parameter) as in Fig. 11 of the manuscript. The highlighted row shows the parameters for the prior with which we have carried out the calculations for the manuscript. The respective aerosol prior is shown in the first column.

Also, is this only for SWIR-2? I would be curious if you attempted scattering retrievals for SWIR-1, to prove that they are no better than non-scattering is right. If my hypothesis is correct, they may be better for the same reason as for SWIR-2? the the information is more from the prior, and not the measurement itself.

We did attempt to include scattering in the SWIR-1 retrievals as mentioned on page 7, line 1 – 2, but even at native GOSAT spectral resolution, a SWIR-1 single band retrieval accounting for scattering typically has an average of 0.24 degrees of freedom for three aerosol parameters. At coarse spectral resolution we encountered low information content and worse retrieval performance with respect to scatter around TCCON. Thus, neglecting aerosol particles in the retrievals seemed the better choice. We added "the SWIR-1 band suffers from low information content and results in worse XCO2 retrieval performance than under the non-scattering assumption" on page 7, line 2 – 3 in the manuscript.

P7L19: The 1.86% scaling factor is interesting. Which way does it go? e.g., do you require a +1.86% scaling of the gas absorption coefficients at 2.01 to match 2.06? Please state this explicitly, as spectroscopists might be interested.

This was indeed unclear. We have added "(i.e. cross sections of the 2.01 μm band need be scaled by 0.981)" in the manuscript (page 7, line 35) to explain this scaling.

P9: I think it is also important to examine the change in standard deviation (scatter) of GOSAT-TCCON at individual sites, to see if that increases more for some sites over others. The global numbers (3.0 and 3.28 ppm vs. 2.43), but it would be interesting to see what these are for individual sites. This information would be usefully presented in a table. In fact, I think a table is important, where the basic information per site is presented (N, mean bias, Stddev). Currently, you try to graphically represent only the per-site bias (in Figure 5).

This information is indeed useful and we have decided to expand Fig. 5 to also show scatter around TCCON at individual sites (we also changed the caption accordingly). Furthermore, the figure was updated to contain information about the number of colocated soundings at each station. In addition, we added Table 2 in the form of a supplementary material.

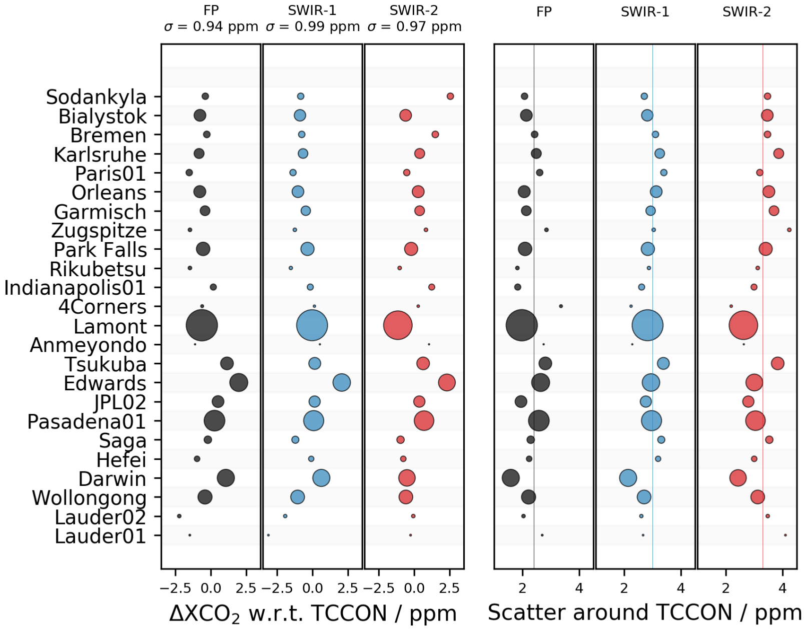

Figure 1: Comparison of retrieval performances at individual TCCON stations sorted north to south. Marker size indicates amount of colocated soundings at each station. Left: Station-by-station mean differences between TCCON and the native (black), SWIR-1 (red), and SWIR-2 (blue) retrievals from GOSAT. The standard deviation of mean differences among the stations, σ, amounts to 0.94 ppm (native), 0.99 ppm (SWIR-1) and 0.97 ppm (SWIR-2). Right: Scatter around TCCON per station for the native, SWIR-1, and SWIR-2 retrievals. Vertical lines mark the average standard deviations (native: 2.43 ppm, SWIR-1: 3.00 ppm, SWIR-2: 3.28 ppm).

We added “Figure 5 also shows XCO_2 retrieval standard deviations per TCCON station. The corresponding data for retrieval performance at individual sites can be found in the supplementary materials.” on page 10, line 1 – 3.

TCCON site	N			Bias / ppm			σ / ppm		
	FP	SWIR-1	SWIR-2	FP	SWIR-1	SWIR-2	FP	SWIR-1	SWIR-2
Sodankyla	217	211	217	-0.38	-0.83	2.55	2.08	2.71	3.47
Bialystok	714	673	708	-0.76	-0.88	-0.6	2.14	2.82	3.46
Bremen	229	218	229	-0.28	-0.75	1.49	2.43	3.11	3.47
Karlsruhe	512	478	512	-0.82	-0.66	0.39	2.49	3.26	3.86
Paris01	215	211	214	-1.52	-1.37	-0.52	2.61	3.4	3.2
Orleans	740	712	736	-0.77	-1.02	0.28	2.06	3.13	3.52
Garmisch	493	462	493	-0.4	-0.47	0.39	2.14	2.93	3.7
Zugspitze	69	66	69	-1.47	-1.24	0.83	2.85	3.04	4.24
Park Falls	940	905	896	-0.53	-0.35	-0.21	2.09	2.84	3.41
Rikubetsu	68	60	68	-1.47	-1.52	-1.02	1.82	2.87	3.12
Indianapolis01	195	193	188	0.18	-0.15	1.24	1.84	2.62	2.99
4Corners	45	30	34	-0.6	0.14	0.29	3.36	2.24	2.19
Lamont	5047	4939	4208	-0.62	-0.02	-1.15	1.98	2.83	2.62
Anmeyondo	9	9	9	-1.1	0.53	1.05	2.75	2.29	2.63
Tsukuba	837	731	830	1.15	0.16	0.63	2.81	3.38	3.83
Edwards	1666	1575	1462	1.98	2.07	2.31	2.64	2.95	3.0
JPL02	713	652	659	0.52	0.15	0.36	1.95	2.77	2.79
Pasadena01	2209	2084	1979	0.27	0.09	0.7	2.58	2.97	3.05
Saga	293	264	287	-0.2	-1.21	-0.96	2.29	3.31	3.53
Hefei	159	148	159	-0.96	-0.09	-0.77	2.24	3.2	3.0
Darwin	1521	1510	1404	1.07	0.63	-0.51	1.59	2.14	2.43
Wollongong	1029	975	974	-0.4	-1.05	-0.58	2.22	2.7	3.12
Lauder02	65	61	65	-2.22	-1.92	-0.06	2.04	2.61	3.48
Lauder01	17	15	17	-1.48	-3.11	-0.25	2.7	2.67	4.1

Table 2: Comparison of full-physics (FP), SWIR-1 and SWIR-2 retrievals for soundings colocated with individual TCCON stations. Stations are sorted north to south in the first column. Number of soundings (second column from left), mean differences between the present retrievals and TCCON (“bias”; third column) and standard deviation (“scatter”; last column).

P9L30: For the parameter correlations, I think you should also look at the retrieved aerosol parameters from SWIR-2 when looking at the XCO2 from SWIR-2. At least check it. I would be surprised if those correlations were not higher than they are for the parameters from the native retrieval, which is VERY different (3 bands, high spectral resolution, etc).

We analyzed correlations of ΔXCO_2 (SWIR-2 - TCCON) with aerosol parameters retrieved from the SWIR-2 configuration. We find that, in comparison to the correlations to the full physics aerosol parameters we used previously,

- correlation with N_{par} changes from -0.05 (FP aerosol parameters) to -0.21 (SWIR-2 aerosol parameters)
- correlation with z_{par} changes from -0.32 (FP aerosol parameters) to 0.05 (SWIR-2 aerosol parameters)
- correlation with α_{par} changes from 0.08 (FP aerosol parameters) to 0.29 (SWIR-2 aerosol parameters)

As the reviewer argued, it does make a difference which aerosol parameters are used here. Interestingly, the SWIR-2 XCO_2 error with respect to TCCON correlates more stronlgy with particle amount and size in case of the SWIR-2 aerosol retrieval than for the FP aerosol parameters. Scattering layer height, however, correlates less with retrieval errors when the SWIR-2 layer height is used.

Section 4: You should state the purpose of the extensive comparison of the modified SWIR-1 and SWIR-2 retrievals to the native GOSAT retrievals. You take the native GOSAT retrievals as the reference, but they are NOT truth. So the value of several of the Figures (7-11) is dubious. You could shorten the paper by removing some of these figures, since you honestly do not know, in many instances, whether the low-resolution, single band retrievals are actually less accurate than the high-resolution, 3-band retrievals.

The native GOSAT retrievals were shown to compare better to TCCON than the coarse resolution SWIR retrievals in section 3. Of course, the native-GOSAT XCO2 data are not perfect, but at least they have been shown to be useful in many studies of GOSAT measurements. For this reason, we illustrate retrieval errors with respect to the native GOSAT retrieval (e.g. Figs 7,9,10,11). We do believe it is helpful to show these plots as they give insight into SWIR retrieval errors caused by geophysical dependencies on a global scale. A comparison to TCCON is limited to the site locations of the network and does not reflect variations in geophysical parameters that are observed globally. These plots also help to demonstrate limitations of the proposed sensor. At the same time they help to make the point that our coarse resolution approach is generally comparable to the native RemoTeC GOSAT XCO_2 product.

P11/Fig 7: What are the R (or R^2) values for SWIR-1 and SWIR-2

vs. Native? These are useful to see as well. I suggest also including these numbers in Fig. 9, and perhaps the corresponding main text as well. Ie, is 90% of the variance explained, or 50%? Etc.

We have included Pearson's correlation coefficient in the plots. For both, SWIR-1 and SWIR-2, the value is 0.90.

P17/Fig 12: Per the discussion of the SNR, this relates to my general comment above, about whether the SNRs you actually ran tests on are even remotely achievable. In practice, most instrument builders will tell you that there is a trade off between SNR and spectral resolution. They are not independent, as this work seems to imply. This should be stated more clearly. As I said above, my preference would be to consult with instrument builders and find out what are reasonable noise models for the type of instrument you want to build, and actually run retrieval tests on those, rather than on the likely unrealistic SNR values within this work.

As we discuss in the introduction of the paper, several authors have proposed pursuing a coarse spectral resolution spectrometer for the detection of localized CO_2 and CH_4 emissions from space (e.g. Dennison et al. (2013), Thorpe et al. (2016)). In light of these previous studies, we investigate here whether a CO_2 satellite monitoring mission would be generally within the realms of possibility and which spectral resolutions are favorable. Instrument design is currently in progress and will be the subject of a forthcoming paper. From a technical point of view, the instrument will require a large telescope (e.g. 15 cm diameter) and a fast opitcs (f-number < 2.5).

References

Dennison, P. E., Thorpe, A. K., Pardyjak, E. R., Roberts, D. A., Qi, Y., Green, R. O., Bradley, E. S., and Funk, C. C.: High spatial resolution mapping of elevated atmospheric carbon dioxide using airborne imaging spectroscopy: Radiative transfer modeling and power plant plume detection, Remote Sensing of Environment, 139, 116 – 129, https://doi.org/https://doi.org/10.1016/j.rse.2013.08.001, 2013

Thorpe, A. K., Frankenberg, C., Green, R. O., Thompson, D. R., Aubrey, A. D., Mouroulis, P., Eastwood, M. L., and Matheou, G.: The Airborne Methane

Plume Spectrometer (AMPS): Quantitative imaging of methane plumes in real time, in: IEEE Aerospace Conference Proceedings, Vol. 2016 – June, https://doi.org/10.1109/AERO.2016.7500756, 2016