Review of "Spectral Sizing of a Caorse Spectral Resolution Satellite Sensor for XCO_2" by Wilzewski et al, AMTD.

This paper aims to test the fidelity of single-band retrievals of XCO2 at low to moderate spectral resolution, based either on the weak CO_2 band only (at 1.6 microns, spectral resolution 1200) or the strong CO_2 absorption band (near 2 microns, spectral resolution 1600). The authors do this by applying the RemoTeC retrieval algorithm to both native-resolution, 3-band GOSAT observations, or GOSAT observations convolved with Gaussian ILSs (corresponding to the degraded spectral resolution) for the single-band retrievals. They find that the errors for the low-resolution, single-band retrievals are not terribly worse than those for native GOSAT, as compared to ground-based observations from the TCCON network. They use these results to argue that remote sensing of fossil fuel emissions (such as from power plants) may be possible from low spectral-resolution, single-band sensors with very high spatial resolution (on the order of 50x50 m^2).

Overall, the paper is very well written. However, I had some questions about their methodology and conclusions, and recommend publication only after addressing these concerns.

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.

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 the Gaussian ILS, in order to obtain an SNR in line with a more realistic value.

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.

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.

Specific Comments

P5L20: You assume 256 spectral channels in a single band. This seems like a high oversampling rate (~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).

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?

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.

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.

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).

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).

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.

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.

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.