Review of "Remote sensing CO2, CH4 and CO emissions in a polluted urban environment" by O'Brien et al., ATMD

This paper describes a novel Observing System Simulation Experiment (OSSE) in which synthetic measurements from a proposed hyperspectral, short-wave infrared instrument called geoCARB are used to derive real-world sampling and error characteristics of the proposed instrument in an urban environment, which subsequently are used to drive a simplified, local-scale flux inversion. This is done in order to assess the ability of the proposed instrument to constrain local-scale fluxes in the face of real-world issues such as persistent cloudiness and aerosol contamination (diminishing the number of measurements) and real-world measurement errors as assessed by using a real retrieval (rather than assuming hypothetical errors as is typically the case). Rather than doing an actual flux inversion, an information-content-analysis type approach is taken in which the synthetic measurements, together with their error characteristics, are used to assess the uncertainty reduction in the urban fluxes afforded by geoCARB.

In general I found this paper well-written, and in fact enjoyable to read. It is refreshing to see papers moving away from old theoretical errors to actual retrieval errors with real retrievals, with the goal of moving towards full, end-to-end OSSE systems. I recommendation this article for publication in AMT after addressing the following comments.

High-Level Comments

The factor of 8 discrepancy between the posterior (i.e., noise-driven) CO uncertainties in comparison with the actual errors is curious. This is not explained by this paper and is potentially worrisome, especially because it is quite at odds with the results of the precursor geoCARB retrieval paper (Polonsky et al., AMT, 2014), which was closer to a factor of unity (typical errors ~3 ppb, vs. 10-15 ppb here). The paper should both state this discrepancy, and attempt to explain the source of the discrepancy between this work and the previous work.

Because a real flux inversion is not performed, only the statistics of the L2 retrieval errors are used, rather than the actual errors themselves. Thus, the outcome is still likely an optimistic assessment of the flux accuracy from using geoCARB, at least with the present-day trace gas retrieval system. Things could of course improve as improvements to the retrieval algorithm reduce systematic errors in the gas column measurements. This caveat is not mentioned in the discussion, and likely should be. (to be fair, it is mentioned in one sentence in the introduction).

Maps of the true and retrieved fields of the different gases (in addition to aerosol) would be highly useful to give a sense of the variability of these fields. For instance, the true field plus the

Specific Comments:

Section 5.2 – Prior Aerosol: The logic of the adjusted aerosol scheme is not entirely clear. Is the idea that places where the AOD is very low is dominated by background aerosol, and where it is high it is dominated by local-source aerosol? Why were the phase functions of both aerosol types set to that of Kahn 2B?

Line 16, page 9: Consider replacing the word "inversion" with "retrieval". Conventionally within this field, "inversion" implies a flux inversion.

Section 7.1: The 3 retrieval tests are never formally described. Please do this. A table may help. For instance, it was not clear to this reviewer what "cloud disabled" meant. Does that mean disabled in the simulation, in the retrieval, or both? And what is the motivation for disabling the cloud?

Section 7.2: Error vs. SNR section: Should make it clear that you could drive the OSSE with a fit to the posterior uncertainties, or the actual "retrieval-truth" uncertainty. Which is done, and why? Please make it clear in this section what the goal is, and which approach (or both?) is taken.

Section 8: Regarding CO problems and the water vapor prior constraint. If you are using an ACOS-style retrieval, this means that you retrieve an essentially unconstrained scale factor to the prior h2o profile. Your hypothesis implies that the vertical profile of h2o really matters, not just the scale factor (which should be almost completely unconstrained by the prior). This seems strange since most of the water vapor lives in the bottom few km of the atmosphere, where moving h2o around shouldn't have much of an effect.

Questions/comments about the CO error discrepancy:

- What were the values of the correlation coefficients in the posterior covariance matrix between CO and H2O, and CO and CH4? Ie, were they highly correlated, such that the algorithm **knows** that uncertainty in H2O (for example) is truly leading to high uncertainty in CO?
- It would be useful to constrain CO completely in the retrieval, and retrieve everything else, since it likely has only a modest effect on the other retrieved parameters of the algorithm (aerosol, surface pressure, co2 profile, etc). Then, a secondary 2.3-micron only retrieval could be run, fixing all other variables to the retrieved value in the first step. If this is done, the errors against the truth may get better. This would help identify the source of the error discrepancy. If this is simple enough to do, the authors may consider it in the revision phase of this work. An even simpler approach would be to remove all the clouds, aerosols, h2o, and CH4 from the true profiles, and re-run the simulations and retrievals (the latter not attempting to retrieve aerosols, clouds, h2o, or CH4). In this highly simplified case, the theoretical

and true errors **should** match. If they don't, it implies some kind of nonlinearity or even bug in the retrieval.