Review of Worden et al, “Evaluation and Attribution of OCO-2 XCO_2 Uncertainties”

In this manuscript, the authors investigate whether the variation in OCO2 XCO_2 is consistent with the error statistics reported by the ACOS retrieval algorithm. They consider the variation of retrieved XCO_2 within a “small” area of ~100 x 10.5 km^2 (~14 sec of ground track), and check whether the statistics of that variation are consistent with (a) random error, (b) correlated random error, or (c) a slowly varying bias from non-CO_2 elements of the retrieval state vector. Their approach is systematic and well elucidated, and yields the not surprising conclusion that a slowly varying bias from interference terms is a key component of XCO_2 variation over small areas. The numbers they derive for the effective precision and accuracy of land and ocean soundings are reasonable.

I have one major comment, and several minor comments. If the authors can respond to these satisfactorily (especially the major comment), I would recommend publication of the manuscript in AMT.

Major comment

A key question the authors need to answer when considering the variation of retrieved XCO_2 over a small area is “how much of this is real?”, since that variation must be “taken out” to quantify the contributing factors behind the remaining variation. To do this, the authors consider XCO_2 fields from CarbonTracker (CT), which is run at 1° x 1° over North America, which the authors call “high resolution”. The N-S gradient of CT XCO_2 has an RMS of ~0.3 ppm/100 km, which the authors consider a plausible measure of flux- and transport-driven variability.

I disagree with their approach and conclusion for two reasons. First, the N-S gradient of a 1° x 1° model is not expected to mirror gradients in the real atmosphere, especially when the objective is to explain gradients seen by an instrument which takes soundings every ~2 km going from S to N. In that context, 1° x 1° is hardly “high resolution”, despite what the authors claim, and CT XCO_2 is expected to be much smoother (and hence N-S gradients much smaller) compared to gradients at the scale observed by OCO-2. To illustrate my point, I’ve made plots analogous to Figure 1 of the manuscript, but taking XCO_2 from a NASA GMAO high resolution (~7m km globally) free running GEOS-5 CO_2 simulation available at https://gmao.gsfc.nasa.gov/global_mesoscale/7km-G5NR/data_access/. The modeled XCO_2 within a box enclosing the contiguous United States was sampled at 20:00 UTC to be close to the 13:30 local overpass time of OCO2 over the geographical center of the US. While the fluxes in this model are not optimized, they are realistic. More importantly, the transport-induced variability over short spatial scales is expected to be more realistic than in CT. As can be seen in Figure R1 below, the N-S gradient in the model is highly dependent on the resolution, and is in general much higher than those evaluated from CT by the authors.

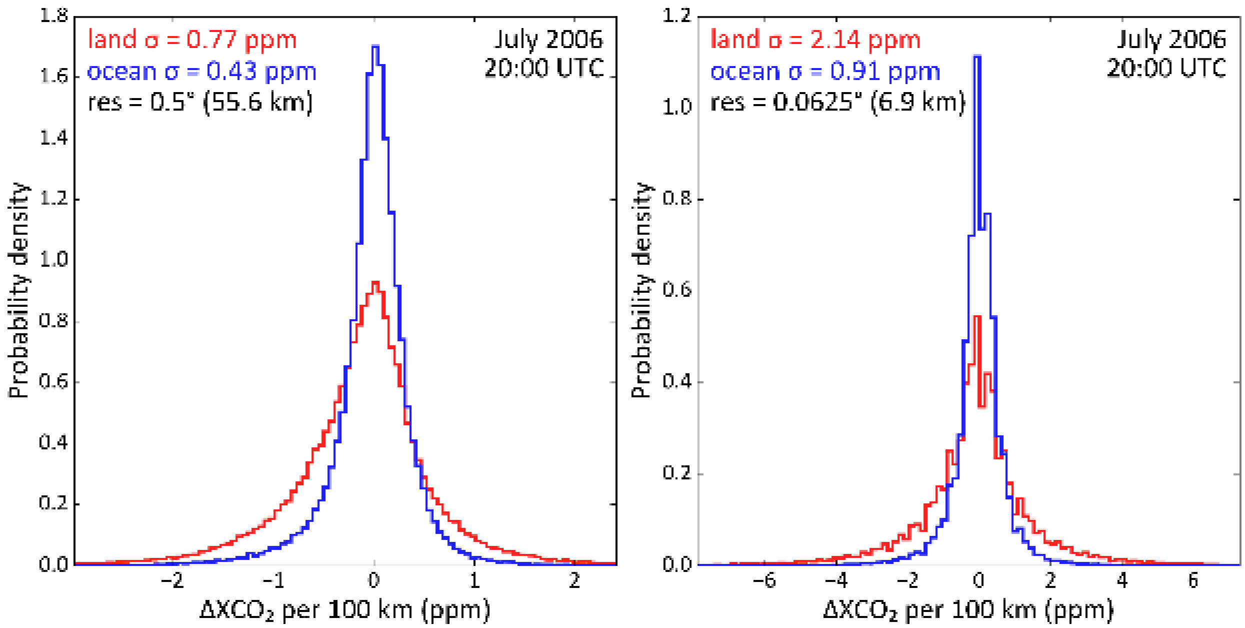

Figure R1: Histograms of north-south gradients of XCO_2 calculated from a ~7 x 7 km^2 model run at NASA GMAO. The model fields are available at 1/16 and 1/2 degree resolutions. In both cases, the fields were sampled at 20:00 UTC (corresponding to 13:30 local time in the center of the conterminous US), and the histograms represent all possible N-S gradients within a box (24°N to 50°N, 127°W to 64°W) covering the conterminous US. The modeled field at a higher resolution has steeper N-S gradients, much higher than the gradients from 1° x 1° CT.

Second, I would argue that for the authors' purpose what is important is not the N-S gradient but rather the variability of XCO_2 in the real atmosphere within a ~100 x 10.5 km^2 area. This is impossible to get from 1° x 1° CT fields, since all of that area is within a single grid cell. I evaluated that variability using the previously mentioned 0.5° and 0.0625° model fields in Figure R2. At 0.0625°, which is still coarser than the OCO2 footprint, a significant fraction of the small areas considered had variability larger than 0.4 ppm over land. Since any Eulerian model variability is limited by numerical diffusion, we can expect that the real atmosphere has even more variability at the ~2 km length scale commensurate with OCO2 pixels.

All this is to say that the variation of OCO2 XCO_2 seen by the authors within each small area could be entirely explained by variability in the real atmosphere, and perhaps the authors don't see that because they look at the N-S gradient (not the variability) of a fairly coarse resolution 1° x 1° model. Their assertion on page 3, lines 21-23 ("the expected variability in XCO_2 …

comparable or less than the calculated OCO-2 uncertainties") may not hold for the real atmosphere.

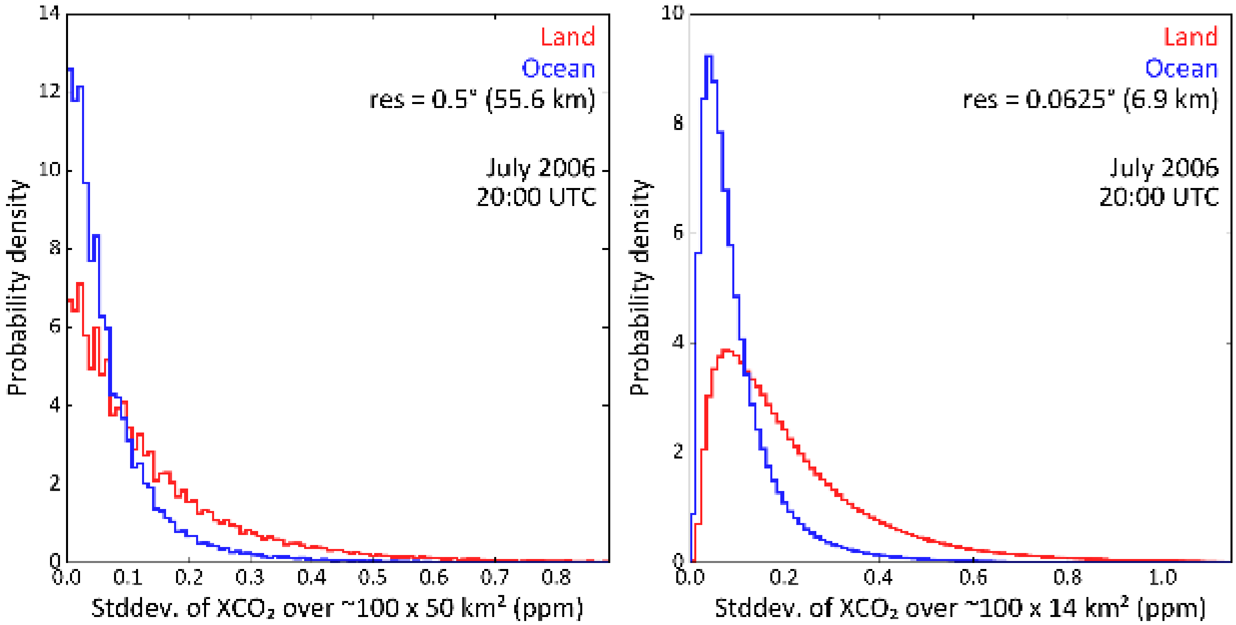

Figure R2: The standard deviation of modeled XCO_2 over a ~100 km N-S stretch for two model resolutions. The models were subsampled analogous to Figure R1 to adhere closely to what might be observed by OCO-2. As model resolution is increased, the modeled variability increases. It is expected that OCO-2, with a footprint of ~1.25 km x 2 km, will observe an even higher variability than a 7 km x 7 km model.

I would like the authors to respond to this argument, i.e., what would happen to their estimate of the different factors behind the variation of XCO_2 over small areas, if it turned out that their CT-derived estimate of the atmospheric variability was too low, and in fact the real atmospheric variability was high enough to explain all of the OCO-2 observed variability?

Minor comments

1. P4, L5: It is not correct to say that the OCO-2 instrument always observes the "glint spot" of specular reflection, since, as the very next sentence explains, there are both "glint" and "nadir" modes.
2. P4, L15: How do the authors know that the statistics of the target mode soundings are spurious? What makes them spurious?
3. P4, L24: The authors use bias-corrected XCO_2 for this exercise. In theory, bias correction should remove long-range correlations in the error in XCO_2 by reducing what the authors call interference error. However, if the bias correction parameters are not chosen correctly, the bias correction itself will introduce slowly varying biases. Can the authors verify that using non-bias corrected (or "raw") XCO_2 from ACOS leads to a larger estimate of the slowly varying error in H3?

4. P4, L25: As far as I know, the bias correction depends not just on TCCON XCO_2 but as well as on the so-called "southern hemisphere approximation" and a small area analysis where XCO_2 is assumed constant over < 100 km along track (see page 14 of http://disc.sci.gsfc.nasa.gov/OCO-2/documentation/oco-2-v7/OCO2_XCO2_Lite_Files_and_Bias_Correction.pdf).
5. P4, L28-29: There is emerging consensus in the OCO-2 flux inversion community that filtering by warn levels (WL) only lets in retrievals with significant bias from interference terms. Rather, filtering by **xco2_quality_flag**, which is WL < 15 plus some additional criteria on retrieved aerosol and CO_2 parameters, is a much better way of reducing the number of biased samples. Can the authors confirm this by showing that if they use soundings with **xco2_quality_flag** = 0 they get a smaller contribution from the slowly varying bias of H3?
6. P4, L29: The highest WL is 19, not 20.
7. P5, L13: Should be "XCO_2" instead of "$_{XCO2}$"
8. P5, L21: The CT-based variability in the N-S gradient of XCO_2 was estimated only over North America, yet it seems to have been used everywhere between 30°S and 30°N. How valid is this assumption?
9. P7, L11-13: In the statement of hypotheses, I think the authors mean "variations in XCO_2" and not "uncertainties". If I understand correctly, the entire point of the manuscript is to see whether variations in XCO_2 within a small area are consistent with XCO_2 errors being primarily from random noise, correlated noise, or a slowly varying bias. So the choice of words in L11-13 is important, and I'd like the authors to either confirm or refute my understanding that "uncertainties" should be replaced by "variations in XCO_2".
10. P9, L18-20: Can the onset of this strong inverse relationship between calculated and actual uncertainty below a certain threshold be used to filter out seemingly low noise (high SNR) soundings over the tropical oceans that might be biased?
11. P10, L1: I think the authors mean "Figure 2" (or 3, or 4) instead of "Figure 1".
12. P11, L3: Why the lag of 0.3 sec? Is it because OCO-2 cross-track "strips" are spaced 0.3 sec apart along track? If so, that should be mentioned.
13. P12, L18-20: Recent results shown at OCO-2 science team meetings and telecons suggest that over small areas, surface elevation has a strong impact on retrieved XCO_2. Is this included in $\mathbf{GK_y}$, i.e., is surface elevation in the vector $\mathbf{y}$?
14. P13, L1: Each of the distributions (Gaussian, Lorentz, Laplace) considered by the authors has a physical basis, i.e., there are reasons why a quantity might follow one of the three distributions. E.g., if two independent variables each follow an exponential distribution, then their difference follows a Laplace distribution. Can the authors speculate why the slopes in Figure 8 might behave like such a quantity?