Review of Lan et al. "Gradients of Column CO2 across North America from the NOAA Global Greenhouse Gas Reference Network"

This paper examines mean spatial patterns and seasonal cycles of partial and total column CO_2 (XCO_2) from NOAA aircraft locations in a number of regions across North America, and compares these to CarbonTracker estimates. They find that spatial variability of CO_2 is larger in certain regions as well is in summer vs. other seasons. They find that CarbonTracker generally agrees well with the aircraft-derived values for these average metrics (mesoscale and synoptic features are mostly averaged out). They find, as expected, that much (but not all) of the spatial variability is driven by remote fluxes and transport rather than local-scale fluxes. Finally, they pose a rebuttal to Reuter et al. (2014), which argues for a strong European sink based on satellite observations of column CO_2, arguing that the spatial patterns of XCO_2 derived in that work are unrealistic in comparison to CarbonTracker as well as their understanding of transport.

Overall Comments

In general, this paper is well-written and is a useful contribution to the literature. I have only a couple serious issues with the paper. First, since this paper is about columns, it would be useful to show direct comparisons of their aircraft-derived XCO_2 to TCCON, to get a sense of differences in remote sensing with respect to a similarly accurate measurement. Second, sometimes they seem to argue that nearly all spatial variability in XCO2 comes from large-scale + remote fluxes and transport, rather than regional-scale fluxes. Other times (such as in the main body), they argue that a significant contribution comes from large-scale + remote fluxes and transport, but a significant contribution also comes from more regional-scale fluxes (within North America in this study). The latter is more in agreement with their data and specifically their removal experiment of section 3.4, so we suggest they rewrite the paper to stress that both mechanisms affect the horizontal structure of XCO2.

Though the authors discuss at length their use of CT2015 to extend the aircraft profiles from 330 hPa to the top-of-atmosphere, they don't really discuss the potential errors from the fact that the aircraft measurements don't go all the way to the surface. While some sites do sample very close to the surface (e.g. 0.2 km at SGP), other sites don't even sample as low as 1 km AGL (e.g. CAR: 2.2 km agl, HIL: 1.1 km agl). Because the concentration can change quickly near the surface, the authors need to acknowledge this source of error, and ideally estimate its potential magnitude. They could do this by taking full CT profiles for their sites and comparing the column extended their way (lowest value held constant) vs. the CT value. They can also estimate it based on tower observations for a few sites.

Finally, their section on Reuter et al. (2014, hereafter R14) doesn't fully fit in with the rest of the paper. Specifically, they show that CarbonTracker (CT) agrees well with aircraft over North America. This is not surprising, given the excellent coverage of surface sites in North America, which CT assimilates. However, this does NOT guarantee that CT will be right everywhere, which they seem to imply throughout the paper. They need to stress that this work validates CT in North America specifically. Anywhere else, and they are merely speculating. Also, they focus on the "hot spots" in the R14 map, which are likely due to inhomogenous SCIAMACHY spatiotemporal sampling as much as anything, and may disappear in a regional inversion. Regarding the European sink, the stations that CT assimilates are all in Western Europe, and much of the controversy is really how much sink is portioned between the Europe and Eurasian Boreal Transcom regions, as discussed in both R14 and Reuter et al. (2016). Their central argument that the spatial patterns of XCO_2 from R14 are unphysical is qualitative at best, and, from this author's perspective, partially based on the unfortunate color scheme (rainbow) chosen by R14 which can accentuate very small spatial differences. They should tone down the language to something more like to say that the spatial patterns seem unphysical, but more work would be required to really rebut their physicality. They also need to cite Reuter et al. (2016), which gives a thorough overview of the "European sink controversy".

Specific & Technical Comments

L54: Suggest you add language such as "or else regional-scale biases can result" and cite Chevallier et al. (2014).

L65: While satellite retrievals of XCO_2 can certainly have regional biases, the Inoue et al. (2016) paper is hardly thorough. It discusses one particular retrieval (the NIES retrieval) of GOSAT data, and in fact the authors appear to cherry-pick the biases of 2.09 and 3.37 ppm they list, which come from stations with the fewest colocations (3 and 1, respectively). Most of the stations in that paper have biases less than 1ppm, consistent with their comparison to TCCON, and that of Kulawik et al (2016). The authors should give a fairer representation of Inoue et al, including their TCCON comparisons, and mention the mean biases relative to TCCON given in Kulawik et al. (2016).

L88: To the 2001 and 2004 papers they cite, they should add Chevallier et al. (2014), which more explicitly demonstrate this.

L119-125: Regarding the recently-discovered high humidity bias. To me, this seems like a pretty large error, especially if the mean / typical correction is 0.5 ppm! Since this is the first paper to discuss it, a plot seems warranted – authors should plot the estimated error based on humidity, and justify this 1.7% humidity cut-off value they give. Also, the authors aren't entirely clear on the units of "humidity" here. Presumably they mean volumetric mixing ratio = dry air mole fraction; please be more clear, or use a more typical unit (such as mass mixing ratio or specific humidity).

L158: I think the authors mean Figure 3, not 2.

L277 – Here is where the authors should discuss potential errors by extending the lowest measurement to the surface (see discussion above in *General Comments*).

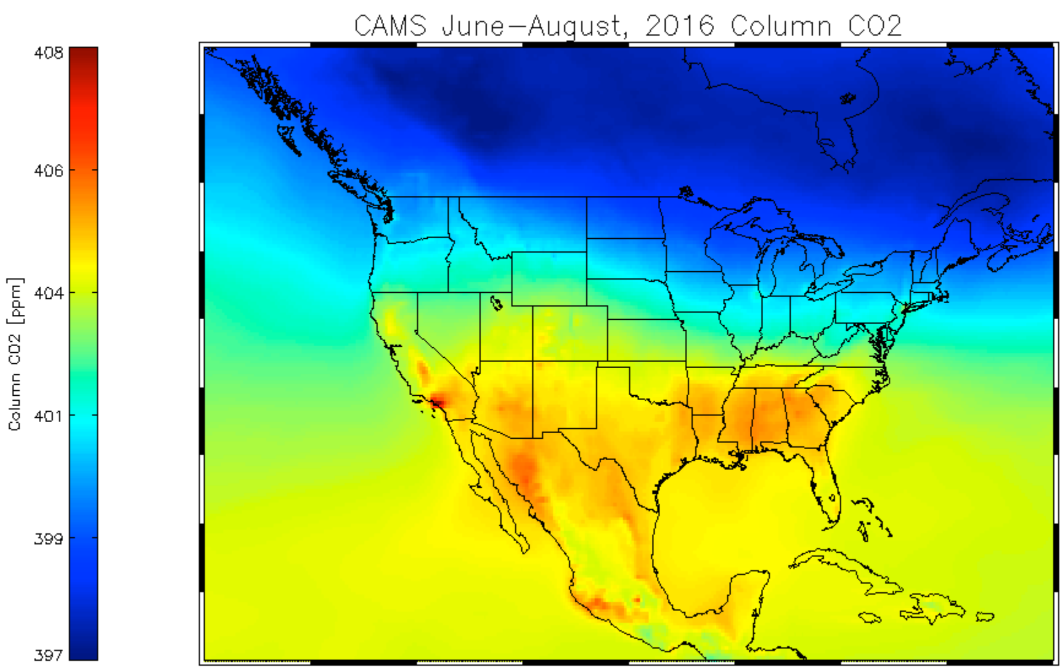
L308 – Please remind the reader here what "partial column" means. I think surface to 330 hPa, but it's not entirely clear.

L318 – Suggest you change the sentence to stress that CT captures these features of XCO2 *over North America* reasonably well. There is absolutely no guarantee that this will be true elsewhere, as discussed in *General Comments*. You can speculate that it will work in other locations, but you don't show it, so as written this statement is not justified.

L329 – Please define "drawdown". Does it mean peak to trough of CO2 in the seasonal cycle, or the minimum CO2 relative to MLO? The former is more typically accepted as the definition of drawdown. Then make Fig 8 consistent with your chosen definition.

L346 – Again emphasize that the agreement is only shown for North America.

L348-350 Regarding CT not showing hotspots. This is not necessarily surprising and could be due to CT's limited resolution. I've attached an unofficial plot of the CAMS nrt XCO2 over North America, for the June-August 2016 average (see e.g. Massart et al. 2016 for details), which has a resolution of roughly 0.25°, and hotspots over L.A. and Mexico city are clearly visible, consistent with the findings of Kort et al. (2012) and others.



L374-6 : This implies that local/regional fluxes account for ~50% of the horizontal gradient in this case. Authors should mention this fact, rather than argue that virtually all of the spatial gradients come from large scale transport, which is what they state in the abstract.

L392: Reuter et al (2014) doesn't only use SCIAMACHY; they use GOSAT as well, and in fact, they use a number of different GOSAT retrievals, not just one, and get consistent results from all of them. Statements in this paper should reflect this.

L409: The Feng et al. (2016) study is only applicable to global inversions. R14 uses a regional inversion with only observations in Europe, to make themselves insensitive to an overall regional bias, and only use the gradients in the satellite data to infer the European sink. This is somewhat at odds with the Feng et al. study, which shows that much of the apparent European sink comes from satellite observations outside of Europe.

References

Chevallier, F., Palmer, P. I., Feng, L., Boesch, H., O'Dell, C. W., & Bousquet, P. (2014). Toward robust and consistent regional CO2 flux estimates from in situ and spaceborne measurements of atmospheric CO2. *Geophysical Research Letters*, *41*(3), 1065-1070.

Massart, S., Agustí-Panareda, A., Heymann, J., Buchwitz, M., Chevallier, F., Reuter, M., ... & Hase, F. (2016). Ability of the 4-D-Var analysis of the GOSAT BESD XCO 2 retrievals to characterize atmospheric CO 2 at large and synoptic scales. *Atmospheric Chemistry and Physics*, *16*(3), 1653-1671.

Kort, E. A., Frankenberg, C., Miller, C. E., & Oda, T. (2012). Space‐based observations of megacity carbon dioxide. *Geophysical Research Letters*, *39*(17).

Kulawik, S., Wunch, D., O'Dell, C., Frankenberg, C., Reuter, M., Oda, T., ... & Miller, C. E. (2016). Consistent evaluation of ACOS-GOSAT, BESD-SCIAMACHY, CarbonTracker, and MACC through comparisons to TCCON. *Atmospheric Measurement Techniques*, *9*(2), 683-709.

Reuter, M., Buchwitz, M., Hilker, M., Heymann, J., Bovensmann, H., Burrows, J. P., ... & Ciais, P. (2016). How much CO2 is taken up by the European terrestrial biosphere?. *Bulletin of the American Meteorological Society*, (2016).