Review of "Validation of XCO_2 derived from SWIR spectra of GOSAT TANSO-FTS with aircraft measurement data" by M. Inoue, et al.

ACPD, 2013

Overview

This paper describes a reasonably thorough intercomparison of XCO_2 as calculated by aircraft versus that from the TANSO-FTS instrument aboard GOSAT. The aircraft data are a synthesis of profiles from the CONTRAIL, NOAA, and NIES measurement programs The GOSAT-derived XCO_2 are taken from the ver 02.00 NIES product. The authors find that aircraft and GOSAT XCO_2 values are in generally good agreement, with GOSAT exhibiting a low bias of about -1 to -2 ppm, and a 1-sigma scatter of "1-3 ppm".

General Comments

Overall, the paper presents a useful validation effort between GOSAT and aircraft measurements. However, the paper makes almost no mention of TCCON, which serves as the foundation of XCO_2 validation for GOSAT by almost all research groups. This paper needs to address the pros and cons of aircraft-based validation vs. TCCON validation. To me, the advantage seems primarily to be the increased number of locations, though many of the aircraft sites have few actual profiles against which to compare. The authors should also compare their results to those of Oshchepkov et al. (2012, hereafter Osh12), which compared NIES v02.00 to TCCON. For instance, Osh12 find a mean bias of -1.93 ppm of GOSAT-TCCON at Lamont (SGP), and a scatter of 1.13 ppm. In this work, you find an even lower bias of -2.6 ppm and a significantly higher scatter of ~ 1.7-1.9 ppm. This implies that the aircraft validation is inherently more uncertain than the TCCON-based validation.

This leads to my second point, that a detailed error analysis is required. The authors cover errors due to the stratospheric CO2 model, the model from the surface to the lowest aircraft measurement, due to the averaging kernel (CAK), and other sources. I suggest they do a real error analysis and attempt to combine all these error sources into a total error that contains all the individual uncertainties. Also, some uncertainty sources are not listed, such as due to the height of the boundary layer and tropopause. And many, many aircraft sites lack tower data. At the sites with tower data, it seems that there can be a large change of CO_2 concentration in the PBL. How large an uncertainty is introduced for sites that lack tower data? Finally, are the resulting total uncertainties on the aircraft XCO_2 very small (a few tenths of a ppm), or significantly larger (approaching 1 ppm)? Why are the 1-sigma standard deviations for the aircraft significantly larger than for TCCON?

The authors spend a lot of time discussion the averaging kernel effect only to find that it is small (sections 3.1.4 and 4.1). This was known from previous TCCON-based comparions (for example, Wunch et al. 2011). I suggest that section 4.1 be trimmed for brevity, and the details moved into the supporting materials. It is much too long as it stands just to report that the CAK-effect is small (we already knew this was true for TCCON; they've confirmed it for aircraft.)

Finally, I recommend that the authors try to be more quantitative in their conclusions. As written, we have learned very little from this aircraft validation effort that we didn't already know from TCCON (as expressed in Osh12). A low bias of "1-2 ppm" and scatter of "1-3" ppm is a huge range.

Specific Comments

1. In the Introduction, you need to state how this paper is related to Miyamoto et al. (2012). They same to make use of much of the same data.
2. Section 3.1.1 concerning the extrapolation of aircraft data to the surface. How did this work, and what kind of uncertainty does it introduce?
3. I suggest combining sections 3.1.2 and 4.2, as both are about the same thing: uncertainty due to the stratospheric & mesospheric CO2 profile. The same actually goes for 3.1.4 and 4.1 (CAK effect). This is more up to the authors but I think combing makes sense.
4. The paper is rife with statements about biases such as "In ocean regions, GOSAT data were underestimated by 1.64 ± 1.05 ppm". The second number (the standard deviation) implies an uncertainty on the first (the mean difference). This is not the case; in gaussian statistics, the uncertainty on the mean is the standard deviation divided by the sqrt of the number of observations. However, that is only the case in which the errors are uncorrelated, which is usually a bad assumption when it comes to XCO2. I suggest breaking these statements up into something like: "In ocean regions, the mean bias of GOSAT relative to aircraft was 1.64 ppm, with a standard deviation between the two datasets of 1.05 ppm." Also, some of these statements can be removed and the user directed to tables.
5. In the curve fitting method (section 4.3.2), I note that the equation cannot handle interannual differences in the seasonal uptake. It has been posited that throughout the northern hemisphere, seasonal land uptake was less in 2010 than in 2009; this is clearly visible in the author's SM plots. This will lead to additional errors in this approach. The authors need to state this. Further, it would be very helpful to see the standard devation of the curve fit – observation data for each of the plots in the SM – could the authors put this somewhere, and state in 4.3.2 what was the typical error in the fit itself (in terms of actually fitting the observations)?
6. Section 4.3.1, statement about difference between land & ocean biases does not actually hold for the curve fit results. I therefore recommend removing this, or saying that in the curve-fit results, the land-ocean differences were about the same (table 5).
7. Section 4.3.1, last sentence. What does "agreed" mean here? You need to be specific! Agreed to within 3ppm is not exactly a great success story when accuracies significantly better than 1ppm are required to improve carbon cycle science.