Airborne Measurements of CO_2 Column Concentrations made with a Pulsed IPDA Lidar using a Multiple-Wavelength-Locked Laser and HgCdTe APD Detector

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2017-360

This paper successfully details the results of the XCO2 dataset produced by the CO2 Sounder lidar during its 2014 and 2016 ASCENDS flights. It does a nice job of detailing the motivation and science goals of the ASCENDS mission, the CO2 Sounder lidar's role in that mission, the current state of lidar technology, and the CO2 Sounder's performance in the 2014 and 2016 flights. The paper is readable, and flows well, even in more technical sections such as the instrument setup and retrieval algorithm description. The included figures are, overall, very informative and helpful.

However, I would argue that the paper currently lacks a scientific punch. Providing more context for the instrument performance would help better communicate the scientific and technological impact of the results. Placing stronger emphasis on past instrument performance, and thus on the improvements in data quality since 2011, would help readers appreciate the instrument's achievements. For example, specific comparisons of 2014/16 noise or in situ bias to 2011 flight legs of similar altitudes; to 2011 spirals. Repetition of specific precision and accuracy improvements throughout the text can only improve readers' impression of the instrument progress.

General comments:

While the paper does a good job of giving the broad science background for the instrument, a background on past instrument performance would lend important context to the quality of the results. One way to achieve this would be to include a paragraph on past campaign data in the technical background section (section 2). This could also be its own section between sections 2 and 3.

There are several places throughout the text where comparing the 2014 and 2016 data to earlier data would help emphasize the how effective the instrument improvements have been. The abstract, for example, should mention the "five-fold improvement in precision over measurements made in 2011." This is likely the most important take-away message in the entire paper, but is not even given in the abstract. (And, it should be noted, perhaps this number would have be 7.5x better, instead of 5, if not for the "noisy reference voltage"?) This statistic packs a punch, and would draw readers in. Remarking *on* the feasibility of the CO2 Sounder lidar as a future space-based instrument in the abstract, as well, rather than just saying that this study serves as a demonstration of its feasibility, would be a good lead-in for potential readers. How close are you, in some kind of phase space, to technical readiness for a space-based version? Are you there on detectors, or are you a factor of 10 away? Are you there on laser power, or are you a factor of ten away? Throughout the rest of the text, I might recommend pulling numbers from earlier flights' similar-altitude flight legs to compare to those given here for the 2014 and 2016 flights. In the specific comments I list a few places where this can be done – including the summary.

One other area which might benefit from further detail is the 2016 flight that used 15 wavelengths instead of 30 – the authors could explain that a space-based instrument will likely sample fewer wavelengths, so it is important to test the capabilities of the instrument with that change implemented. This would tie this data, once again, to the larger science goal. Speculation on why that sampling change may or may not make a difference in the results, and whether a change was expected, could also be given. The results of other flights using fewer than 30 wavelengths could also be mentioned – did they show any significant difference? It says later that a space-based instrument would only use 8. Why then are you testing with 15 or 30, and not 8?

Finally, more discussion of your ability to determine your biases is needed. A low bias is perhaps the most important potential advantage of the lidar over passive measurements. Can you get your biases down to 0.3 ppm? 0.1 ppm? Therefore, doing a good job to quantify the level at which you can determine your biases is *a critical aspect* to these test flights. It can make or break a decision to do ASCENDS or not. From this paper, it appears that you cannot see any statistically significant biases in your retrieval (as compared to the in-situ). But what is the current level at which you can determine your biases? I note that while you are often within the standard deviation of the different lidar column measurements at a given altitude, you are virtually never within 1 *standard deviation of the mean*. If your errors are truly random, the errors should integrate down like the square root of N. Since you often have ~100 measurements per altitude bin, your noise-driven error is on the order 0.05-0.1 ppm on the mean XCO2 in an altitude bin, while your difference to the in-situ is often more like 0.5 ppm. That is a 5-to-10 sigma discrepancy (ie, it looks like a bias!). However, it could be that the in-situ columns are only good to 0.5 ppm, considering the change in the column that determines over the course of the spiral, the native in-situ errors, etc. Therefore, more commentary on and quantification of your biases and bias knowledge is sorely needed.

Specific comments:

Abstract: As stated above, please mention the factor of five improvement in precision in the 2016 flights compared to the 2011 version of the instrument.

Section 1. Intro - When describing currently operational (passive) space-based instruments and their role in carbon cycle science, it might be informative to comment on their abilities to meet the sub-ppm measurement requirement. It would further tie the statistical results of this study to the larger science goal, which is in part to improve upon passive datasets.

Section 2. Instrument background - Paragraph 3 of this section is the only place where the writing feels a bit disjointed and repetitive, and could be reorganized slightly to convey its message more clearly. It hops back and forth from spectroscopic information to environmental variables and Doppler-shift information a few times; I think that a simple rearrangement would help.

Section 3. 2014/16 Lidar setup - it is not clear in the text why the receiver optical transmission (in Table 1, which is referenced on page 3, line 45) was so low in 2014 (9.2%, as compared to ~50% and 60% in 2011, 2016). The authors should comment on this in this section. What specifically was done to improve it so much from 2014 to 2016?

Section 4. Data processing & retrieval – The retrieval summary mentions that "HDO absorption spectrum can bias the retrieved XCO_2 value if not taken into account," but it does not give the average value of that bias. The authors would should include either that number or a reference to further information on the topic. Also, there seems to be no reference provided for a more detailed description of the retrieval process. Finally, what is the posterior uncertainty on your measured HDO column? This could be a side benefit of your technique – knowledge of HDO can be used to study water cycle processes (see e.g. Frankenberg et al, 2013, "Water vapor isotopologue retrievals from high-resolution GOSAT shortwave infrared spectra", AMT).

Also, do you fit for a scale factor to a prior CO2 profile, or do you assume a uniform CO2 profile and solve for that single concentration?

Finally, and perhaps most importantly, is the short paragraph on the "clumped fitting" approach. This appears to be a novel aspect to this work. Has this approach of yours been described elsewhere in the literature (if so, please give the reference) or is it introduced here? How much does it reduce the errors (and in what way? Ie – is it simply a lowered scatter at a given altitude level?) between L2a and L2b? And how physically does it reduce the errors – what are the mechanisms? This could be an important aspect to this work, but it is downplayed – not mentioned at all in either the abstract or summary. Is is something that none of the passive teams (or other active teams, to my knowledge) currently do – you figured it out, so please make some more noise about it if it is really useful!

Section 5. Campaigns overview – There are a few opportunities here to reference the precision and accuracy of previous campaign data. For example, on page 6, line 35, the authors could state that this <1ppm agreement between lidar and in situ is improved from 2011 comparisons (Abshire et al. 2013b), when this agreement was more along the lines of <1.4ppm. Also, more should be stated about the "noisy reference voltage". Does this imply that if this were fixed, the precision would be another 33% better? That's significant.

Section 6. I might contest, on the return leg (bottom of Figure 18), the statement "There is good agreement between E-W gradients measured by lidar and those computed from model", especially in the 5.6km segment. The model predicts something like a 1ppm E-W gradient, whereas the observational gradient might be closer to ~4ppm. Please soften this statement.

SECTION 8. Discussion - Line 2: "Changing from 30 to 15 lines did not significantly change the retrievals, per the 2016 snow flight." It wouldn't hurt to include the numbers on this – how much of a difference did it make? Or was it truly negligible?

SECTION 9. Summary – to restate what I said above, more quantification of your biases, and at what level you can even determine them with comparisons to in-situ, is needed.

On line 28 would be meaningful to state specifically how much smaller the lidar vs. in situ biases are than in 2011, and how this was

Typos/grammar:

- Page 2, Line 4 – Cleaner wording – remove "so can" from "[…], and so can cause large retrieval errors."

- Page 2, Line 22 – Define IPDA acronym once before using it in the rest of the text.
- Page 2, Line 42 – Missing a period after "et al" in "[…] Abshire et al, 2013b)."
- Page 5, Line 25 – Typo – "from based on" – remove either "from" or "based on".
- Figure 7 – This figure appears blurry; update resolution.
- Figure 9 – Change RMS labels to include "initial" and "after fit" for better at-a-glance clarity. Also would recommend, in the caption, describing figures in the order in which they appear - top then bottom.
- Figure 14 & similar figures – If range and ground elevation are on the same axis (right-hand), perhaps clarify this by making them the same color but making one dotted/dashed.
- Figure 18
    - The bottom plot doesn't show any 10.8km data, even though it's included in the legend. Is this a mistake? If so, correct plot; if not, removed 10.8km data from legend.
    - A secondary legend for observations vs. model would be better than using arrows.
    - The statistics in the legends don't seem to be described anywhere; describe in caption. Is the correlation between observations and model?
- Page 8, Line 7 – Reference in wrong format? ("[54]")