Atmos. Chem. Phys. Discuss., 11, C11298–C11304, 2011
www.atmos-chem-phys-discuss.net/11/C11298/2011/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Assessing filtering of mountaintop CO_2 mixing ratios for application to inverse models of biosphere-atmosphere carbon exchange" *by* B.-G. J. Brooks et al.

Anonymous Referee #2

Received and published: 2 November 2011

General Comments:

This paper is relevant (and practical) to individuals performing atmospheric inversions of biological trace gases such as CO2. It expands available CO2 observations for use in atmos inversions by providing filtering methodology (and comparisons) and does this in an area for which "regionally representative" CO2 measurements are very difficult to get. I would say that the most novel thing about this paper is the attempt to perform model-specific filtering techniques. This is a very interesting direction of research and I think there is a lot of research still to come on this topic. Unfortunately, I think CarbonTracker and TM5 might not be the best example to be used for this paper, or

alternatively, the right metrics of comparison between the model and the observations haven't been identified. Many coarse models have a very difficult to time with vertical transport, let alone in complex terrain, and therefore I think the authors should continue to look for more representative metrics of evaluation that can be used in their "model-specific" filter (lapse-rate in this paper). Additionally, for researchers attempting to employ the methods, the equations and notation need to be error-free and much easier to follow. I'm only listing "major revisions" because of this last point (math notation) and the fact that I think one figure needs to be redesigned, and I would like to make sure those two things are done. Otherwise, I have many comments, but mostly of "minor revision" nature. Overall, timely, stimulating, and most importantly, useful paper. Thank you.

Specific Comments:

(1) the derivation of a "common" data set for comparison to aircraft data was very hard to follow. Maybe some kind of timeline or spreadsheet comparison of times and associated data (lapse rates), in order to facilitate "seeing" the intersections between the data sets used for the different methods.

(2) since measurement error is likely small, the 'filtering' presented is really a way to subset observations so that model-data mismatch errors are lowered, i.e. so that obs are consistent w/ behavior that the models can reproduce. Hence future work would probably benefit from more model-obs comparisons, in addition to the aircraft comparisons which were very nice.

(3) Sections 3.4/3.5 need a review by the authors and/or a statistician. The notation is very difficult to follow and there appears to be a lack of definitions. This is a serious impediment (and serious comment) to anybody hoping to employ any of these methods. See assorted comments in Technical Corrections.

(4) Section 4.2 is very valuable part of this paper. Unfortunately, it seems to read a bit awkwardly as far as the subsetting is concerned. I'm very confused for how the different

"filtering" subsets relate to the 24 possible CARR vertical profiles. More details below.

(5) Section 4.2: A general comment is that the authors should emphasize that 1500 meters of well mixed air is a more stringent requirement than 50 meters (although the highest variability should be near surface) so that the fact that 218 of the 255 flights aren't "useable" doesn't NECESSARILY imply that the 218/255 % of the tower data isn't useable. This would depend on the reasons for the variability in the aircraft vertical profiles (aircraft hitting plumes/thermals of high/low co2 air, etc).

(6) Section 4.3: A general comment on this section. This represents a good opportunity for the authors to hypothesize on further research. For example, the ability of atmospheric models to accurately model storm fronts, topography, complex weather, etc largely a function of grid resolution. Although, fine resolution doesn't imply GOOD MODELING, coarse resolution certainly implies that many features can not be resolved. Therefore, it would seem that the Lapse Rate filter concept would benefit heavily from additional filtering/weighting based upon a model's (like CarbonTracker/TM5) ability to resolve wind speed, direction, vertical gradient etc, in addition to the "one-sided" lapse rate test alone. You are more likely to "accept" the observation into assimilation if the model shows it can reproduce a number of features in addition to lapse rate.

(7) Discussion: I'm still horribly confused by Figure 9. The authors need to reinvestigate how to display this data. Even if is correctly displayed (which I'm not sure of), if you can't figure it out in a few minutes, nobody is ever going to pay any attention to it. I really like "dense" images but in this case, I have to recommend to spend a little bit of time on a different/simpler visualization of this data. This is the figure I'm referring to in the "general" comments.

(8) Discussion: I like the idea of using model output to help subset the data used but the way in which the CT/TM5 data is used brings up some serious considerations. The CarbonTracker/TM5 CO2 has had some historical issues with surface CO2 being

VERY wrong, high CO2 I believe and not just at night under stable conditions. I'm not expert on TM5 but you mention NOAA folks in acknowledgements and thus you have access to this information. If you assume a model's lapse rate is indicative of it's ability to accurately model THAT lapse rate, and you have strange steep gradients of CO2 near the surface you really lose the ability to confidently say what you are trying to say. Furthermore, with increasingly high lapse rates coming out of models, one would have to assume that uncertainty in the modeled CO2 HAS to go up. In other words, if the model is WRONG w/ the lapse rate, you should throw the data out, but if the model is RIGHT about the lapse rate then you are trying to model VERY difficult conditions, stability, etc. The authors should comment first on the TM5/CarbonTracker surface CO2 issue (consult NOAA-ESRL if ?s) and possible effect on your lapse rates. I'm not sure of the answer here but I know that most people who read this, and have seen the CarbonTracker CO2 data, will think about these issues. Then the authors should certainly caveat the lapse rate filter by the fact that VERY incorrect transport in the model could essentially allow the data to come into the inversion essentially unfiltered, even when the model might be getting the dynamics very wrong. A recommendation would then be to speculate on additional model output that could be used to evaluate the model's ability to accurately model transport in complex terrain, in other words something to provide a "check" on simply using the lapse rate. I like the direction that the authors are going, but they have to be careful about specific claims and evaluation metrics.

Technical Corrections:

(1) abstract: change "..terrain are difficult to measure often due to..." to "terrain are often difficult to measure due to..."

(2) abstract: the phrase "standardized to common subset sizes" is too technical for abstract since I'm guessing the reader has no idea, a priori, to know why this is an issue, or what you are even referring to. Maybe the authors should drop that phrase and just talk about it in the text.

(3) Refs on page 4. I would put in some of Thomas Lauvaux regional/microscale papers as well since they belong in this "set" of inversion papers. Look up Lauvaux et al. 2009 and Lauvaux 2011 (ACPD) even though this was probably accepted after your submission. Also, possibly Gourdji, et al 2011 (Biogeosciences). I'm sure there are more, these are just what I'm immediately familiar with. Goeckede 2010 and Lauvaux 2009/2011 are probably most relevant at the scale that is being looked at here.

(4) Section 3.4:

X(n): is this a sub-daily time series?, daily time series? etc. It appears that "n" is a day but it is implied that X(n) contains subdaily values so please clarify EXACTLY what X(n) is. Same for x(n), the subsample.

Equation 1 has an "i" index which doesn't appear in the expression, I'm confused. Nothing in the summation notation, N, i, appears in the expression.

The definition of x(n) around line 264 is confusing also. x(n) appears to be defined in terms of itself?

I like the rigor with which the authors attempt, but please run by an objective non-informed statistician/mathematician in order to make sure the definitions and equations are clarified.

(5) Section 3.5

Same comments apply from 3.4 w/ respect to X(n) and x(n) notation.

It would also be helpful to officially define X "sub" i. It is indicated that X "sub" i. It is implied that X(n) "sub" i is somehow the hourly mix ratios belonging to day "n" and excluding some set of hours. Please clarify.

(6) Section 4.1

line 320: It would appear that the SI filter under all the different hourly subsamples is higher variability, not just under the "complete obs" as written.

line 321: Although technically correct, I would change the sentence "Also time-of-day sampling generally has little ..." to "Also time-of-day sampling *alone* generally has little ..." just to emphasize that the authors are only subsetting the times here and no other filtering is occurring. This is probably an important "benchmark" for readers to understand.

line 333: Change "on the order of -0.4ppm from the complete set." to "on the order of -0.4 ppm from the complete set, implying a slightly weaker seasonal amplitude than complete set" or something to that effect. I'm worried that "LARGER DIFFERENCE" implies to "LARGER SEASONAL CYCLE" to somebody reading this quickly.

line 337: Change "SVLG and SI subsets ..." to "Largely due to the way they are defined, SVLG and SI ..." or something to that effect. They are defined as mechanisms to filter based on stratifications, so this should be no surprise.

(7) Section 4.2

line 354: I'm assuming that the WM, SI were tuned (criteria) until the authors got "around" 20 or so common profiles? I'm not clear though how the hourly-stat filter subsets were handled. Were the filtered data sets FURTHER reduced to 20 common hourly profiles? based on individual CO2 gradients of each filtered data set? Might want to clarify this a bit.

line 352: I'm also a bit puzzled about the need for "standardizing" the comparisons. Although, I assume the authors have legitimate reasons for them, it would be nice to hear them. It would seem that the filters as set up (default) would not admit enough observations to be compared to aircraft? and so the filter had to be loosened in order to have ANY data to make the comparisons with. If this is correct, it would be better to explain the rationale for the filter adjustments in that fashion.

(8) Discussion

line 454: "were able to identify and retain CO2...". Sort of nit picky here but I don't like "identify", it seems to imply that the authors are identifying CORRECT observations or something similar. All these observations are presumably correct. The filter simply is trying to subset some that represent regional scale variability as opposed to local. Change "were able to identify and retain" to "retained", or something similar.

line 457: "diurnal"? Do these case studies represent DIURNAL or SYNOPTIC variability? I would assume synoptic might be more appropriate here?

line 470: where does the 86% come from?

line 479: What about night of 17th? big drop there too, was that not synoptic?

line 490; I don't recall, can the authors tell me whether 0-4LT filtering reproduces the seasonal cycle?

lines 512:513: Does this mean there was only 1 lapse rate from model for each time of day? So, a 12LT lapse rate from model for winter and summer were combined? It does seem that this needs to be considered more in the future as far as future lapse rate / transport model comparison filters are created. One would envision that there could be very strong differences seasonally. The authors might even elaborate on the actual differences in the lapse rates between winter and summer in the discussion, simply a couple averages of day/night over winter/summer would probably suffice.

lines 530:531: This comment is important. I have to say that most modelers will look at comparisons of NWR CO2 to 1x1 degree TM5 results in the mountains and say that there is no way you can compare these things. The authors do caveat with these comments, but they should not be taken lightly.

Interactive comment on Atmos. Chem. Phys. Discuss., 11, 25327, 2011.