Atmos. Meas. Tech. Discuss., 7, C3156–C3161, 2014
www.atmos-meas-tech-discuss.net/7/C3156/2014/


Atmospheric Measurement Techniques Discussions

# *Interactive comment on* "Technical note: Detecting outliers in satellite-based atmospheric measurements" *by* P. E. Sheese et al.

**Anonymous Referee #2**

Received and published: 13 October 2014

From "Reviewer 2".

My apologies, somehow my earlier review got truncated. Here it is again, hopefully in full form.

Review of Technical note: Detecting outliers in satellite-based atmospheric measurements.

By Sheese et al.

This technical note describes the approach taken by the science team for the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS) instrument to identify "outliers" in their Level 2 dataset. The ultimate goal being to provide data



quality flags and information to eventual users of the ACE-FTS data. If this is the approach that the ACE-FTS team have chosen to adopt (as implied by this paper), then I fully recognize that it should be documented, and the process by which it was arrived at should be described to the user community. A technical note in AMT is a very suitable way in which to do that.

However, I have serious concerns about the efficacy and suitability of the approach described. In addition, in my opinion, the paper lacks important and valuable detail on the background to the method. The second of these issues can and should be remedied by the authors before publication. The first is a larger issue that may motivate more work, that is perhaps left to a follow up analysis and paper.

Fundamentally, this paper does not make it clear what phenomena give rise to these anomalous "outliers". Are they due to anomalous behavior within the ACE-FTS instrument? Are they due to poorly converged retrievals? Are they due to poorly modeled signal contamination from clouds or aerosols? One presumes that all three can contribute, individually or together, on a case by case basis to give such outliers. A discussion of these contributors should be included in the paper. Furthermore, for each of these contributors, surely the outliers that arise are better diagnosed in a phenomena-specific manner, rather than by resorting to some rather ad-hoc examination of Level 2 data. Instrument anomalies would, presumably, be detectable in raw telemetry, interferrograms or radiance spectra. Poorly converged retrievals could be identified through some radiance chi-square metric that the retrievals presumably (hopefully!) generate for each occultation scan (and/or even each spectrum). Cases of aerosol/cloud contamination would be identifiable as impacts to the "background" on observed spectra.

Only when all avenues to screening for each phenomena in its "native landscape" are exhausted would I consider a screening based purely on Level 2 data with no reference to underlying anomaly mechanisms. Perhaps the team have already exhausted such approaches. If so, the manuscript makes no reference to that effort. If not, I would urge that the authors return to their dataset and explore a more "mechanism-aware"

approach to the screening. Such approaches can really only be put together by a team with thorough end-to-end knowledge of the measurement system, and it is vital that this be done sooner rather than later, given the age of the ACE-FTS instrument, and the limited funding that can be expected for it in future years.

If the team is genuinely confined to approaches that rely purely on searching a dataset for outliers, without any reference to characterizable underlying mechanisms, then the approach described is probably the optimal one for the input dataset. However, all users need to be fully aware of the classes of potentially genuine atmospheric science phenomena that may be flagged as spurious by this approach. The authors give some examples of these. However, I can think of others, not least the unprecedented enhancements seen in lower stratospheric species in response to the "Black Saturday" fires in Australia in February 2009 (e.g., Siddaway 2011, Pumphrey 2011) and the episodic injections of moist air into the summer time northern mid-latitude lower stratosphere reported by Anderson et al. (Science 2012) and as seen in satellite observations (Schwartz et al., GRL 2013).

The approach described in this paper is, mercifully, not likely to reject large scale events out of hand (as happened, it seems, during the 1980s era of TOMS data processing). However, that episode has left many in the satellite atmospheric remote sensing community with understandable trepidation about all data screening methods of this type.

To summarize my "big picture" questions, the manuscript should be modified to describe the origins of these outliers, and why other approaches to performing the screening in a more "mechanism aware" manner are not viable (or at least are a "next step"). If they are viable, then my choice would have been to start with them rather than the approach documented here. However, having some screening is preferable to having none, so I do not necessarily want to hold up publication of this paper.

The manuscript is generally very well written, and the figure are clear (if a little small, but perhaps thats an AMTD vs. AMT issue).



==========

More minor comments.

—- Title

The title implies a far broader subject matter than is covered in the text. If the goal is to provide needed direction to the ACE-FTS data users, then it is critical that ACE-FTS be included in the title. Also, the "detecting outliers" title is unclear: are we supposed to think of these as spurious data to be rejected or geophysically real data of scientific interest (e.g., Black Saturday, H2O injection etc.). In point of fact, the abstract and the manuscript are also a bit vague on this in places also. It's only because the abstract talks about "quality flags" that one can draw the conclusion that it's the former.

For the title, I would suggest something like:

"Detecting [screening for?] spurious outliers in ACE-FTS atmospheric composition observations"

—- Abstract

Line 14: People might think "fitting error" means radiance fitting error, but that's not the topic of this paper (even though I feel that would be the preferred method). It would be good to clarify that this is a statistical fitting of the Level 2 data to some distribution.

—- Page 8395

Line 1: "This method is much less sensitive to extreme outliers" is not as clearly explained as it could be. Is that a good thing (it gets a better picture of what is "normal" because of that lack of sensitivity, so can more easily identify fliers), or is it a bad thing (it misses fliers).

Line 14: Add some more examples (e.g., those listed above).

Line 18: Add "spectral" between "high-resolution", just to be explicit?

—- Page 8396

Line 1: Perhaps "are" -> "have been" just to be completely clear.

Line 11: Is this the grid that the level 2 data are distributed on? If not, will users have to interpolate level 2 data to this new grid to use the flags? Also, if not, why not put the flags on the same grid as the data?

Line 13: "As well, it was..." -> "It was also..."

—- Page 8397

Line 1: "consistent" is unclear in this context. Consistent with what? Between the mean/median approaches? Between the different products?

Line 4: I would take issue with "clearly", again for the reasons discussed above (except in the case of strongly negative vmrs, I grant).

Equation 2: This seems strongly at odds with the traditional definition of both E(x) and of pdfs, for which one performs an integral to compute the probability that x is within a given range. I suggest you modify your nomenclature to avoid confusion and this break from tradition. If the references given below establish a new "tradition" then it might makes sense to stick with these unconventional definitions, but you should be more explicit about that.

Line 9: Just to be sure this is 10ˆ-4 on the scale where less than unity gave you pause before, yes?

—- Page 8399

Line 8: "... for in the fit. For each subset ..." is a bit unclear, is the discussion starting with the "For each subset ..." here the description of the "ad hoc" method introduced above, or is this a new topic?

—- Table 2



The meaning of the "percent error is not within 0.01-100%" cases was not made clear in the text. Is this the ratio of a Level 2 error bar to the retrieved value, or is it something else. In either case, more detail is needed. Apologies if it's there and I missed it.

—- Figure 2

Would be good to label the plots (or at least the x-axes) with O3, H2O.

—- Figures 3 and 4

These could be merged into one figure (granted the Figure 4 ones get followed up with further analyses, so perhaps it's OK as is). However, I also don't see why you don't include the green fitted histograms on figure 4.

---

Interactive comment on Atmos. Meas. Tech. Discuss., 7, 8393, 2014.