**Atmospheric
Chemistry
and Physics
Discussions**

Interactive comment on "Five blind men and an elephant: can NASA Aura measurements quantify the stratosphere-troposphere exchange of ozone flux?" *by* Q. Tang and M. J. Prather

Anonymous Referee #2

Received and published: 8 November 2011

Review of "Five blind men and an elephant: can NASA Aura measurements quantify the stratosphere-troposphere exchange of ozone flux?" by Tang and Prather

Overall Comments

Lots of interesting comparisons in the paper – but it seems odd that the focus of the paper is trying to compute STE fluxes. If fully functional, HIRDLS data might have been useful for this purpose. But from planning stages onward, TES and MLS were never going to be the last word for this issue. OMI has some interesting possibilities but not for the standard profiles and the approach taken here. Even starting with the title – there is a piece of a science investigation whether using data or a model wherein there

is a choice of appropriate tools for the task. This step seems not to have been taken, and puts a negative tone to the paper where that is not necessary. The paper will be a lot more useful if the following comments are taken into account prior to publication.

The paper reads like the data have some inherent problem, when in fact the instruments were state-of-the-art for the era in which they were designed, and are providing all sorts of good information concerning the problems that they were designed to address. The poor resolution of TES is the best that can be obtained for nadir measurements with this remote sensing technique and space-based technology of our era, and also some 'features' of the TES instrument – but regardless, the technique has limited vertical information in the troposphere. The instrument was never designed for the study of STE – the 'science traceability matrix' would have had much different requirements for such a study, and if choosing an instrument for such a study, nobody would have ever chosen TES. The TES team chose to report many levels and all the information from the averaging kernels, thinking that this would give the user some freedom in the use of TES – but certainly the team has been consistent and patient in explaining the concept of degrees of freedom and best practices for use of TES data when comparing with global models. Even in Figure 1 – aren't we looking at a high bias in the CTM that the averaging kernel aliases into the troposphere? So if the CTM weren't biased, then the results would look better? Last – knowing that the CTM has a high bias and what the weighting functions look like before you ever start – isn't this a result that you could have anticipated? I don't think it serves any purpose for this to seem like a surprise. In your presentation, it would be much more rational to start from a statement – TES has only 1 DOF in the troposphere, so the appropriate comparison is with TCO (and explain) It is a poor practice to go through this whole production with TES profiles – and then conclude that since they only have 1 DOF in the troposphere, the appropriate comparison is with the TCO when you should have learned that by reading the data quality documents.

Something we have also given a lot of thought in recent years – if you start with an

excellent prior (which can formally be thought of as 'everything you know about the atmosphere before this measurement') and the retrieval does not change the prior – exactly what does that mean? In one (positive) sense, it means the measurement is 'consistent with all prior knowledge'. In a negative sense, it means that the measurement didn't reveal provide any new information.

I have strongly suggest that there be a change in tone for this paper. It will be understood better, have more impact and reflect better on the first author if the presentation tries to show how complementary information can be brought to bear on the same problem. The paper almost gets to this with the idea of using the model as the transfer standard and the platform that can see the whole elephant. The paper does this idea a disservice when it seems 'surprised' that Aura instruments don't resolve trop folds, or that the TES profiles have limited vertical information.

Specific comments

ABSTRACT: and the inconsistency amongst different instruments, such as from tropics to Northern Hemisphere mid-latitudes in July 2005 at 215 hPa and over tropics at 147 hPa for July 2005 and January 2006

None of the instruments is giving particular good information at 147 hPa in the tropics. MLS is best – so the inconsistencies should come as no surprise. It would be a lot more productive to think of the Aura instruments as providing complementary information and to acknowledge the importance of using the right instrument for the right problem – something you ultimately come to but do not state.

26899 l 5 I thought an important driver of differences was whether the effort was to get flux across the 380K surface or to get the flux actually entering the troposphere? Did all of these authors use the same definition? If no, then this is an overstatement b/c the comparison is apples/oranges.

L 17 The Aura instruments were never designed to resolve trop folds – at best, de-

signed to see their effects (in an integrated sense).

26901 7 'modeled profiles generally match sondes' – broad statement in Tang and Prather 2010 and also broad statement here. Really would love to see scatter plots, histograms, statistics.

L 14 – they aren't exactly 'ozone instruments' – they are four instruments that measure ozone (among other thing)

L19 – 'many studies' – there is no need to have a pejorative tone here. Often people use L3 because that is appropriate for their question and also for the instrument. For example – MLS single profile precision is often large so that for many studies an appropriate average makes more sense than relying on the swath profiles without averaging. This is also true for TES.

26902 HIRDLS – you probably should refer to the appropriate data quality document.

26902 MLS – mention single profile precision in the altitude range in question, and refer to data quality document.

26903 OMI – other TCO products need to be mentioned.

26903 TES – someplace you are going to say that there is only 1-2 pieces of independent information in the troposphere, right? And that the degrees of freedom depend on thermal contrast and thus have seasonal dependence? And maybe refer to the validation papers that show the noise in single profile comparisons?

26906 – at the end of the discussion about TES – perhaps you should consider what TES would tell you if the ozone in the lowermost stratosphere actually were very high compared to the prior. . . (instead of biased high). In that case, with the same sensitivity, TES would also place too much ozone in the UT (so the sensitivity, resolution and prior would need to be taken into account in interpretation) – as you correctly include (lines 24-27).

26908 – even when introducing the case studies – there are so few cases that meet your criteria – what is the effect of the 1 hour time coincidence? I think a lot of the launches were at middle latitudes and local noon, which would definitely miss . . . I think that our time window for validation was wider . . . Table 1 in Worden et al. (JGR, 2007), compares TES with ozonesondes, and shows that only about 10% of the sondes used in that study would have the possibility of meeting the coincidence criteria for TES . . .(not even thinking of the other instruments). Sondes were launched to be coincident only during campaigns and at some SHADOZ sites.

26909 – you also said that TES had only 1 DOF in the troposphere. Isn't the Pressure/Loc plot misleading?

26910 – does using the OMI operator make a difference?

26910 – these numbers for MLS either make sense with respect to single profile precision and data quality or they don't. Do they? For an appropriate average, does MLS give the correct mean? Is this a 2 or 3 sigma number?

26911 HIRDLS doesn't see below clouds – between this and the radiance correction for the Kapton, I don't think they claim to see anything much in the tropics below the tropopause.

26911 Pretty shocking to me to read that TES shows 'skill' in detecting boundary-layer O3, since the sensitivity is very low there. In fact there is a huge effort to combine TES/OMI information because by the UV-vis + ir has theoretical sensitivity to the troposphere.

Figures 3 and 4 You need more separation between the top group and the bottom (profile) group – or put the color bar on the right side of the (d) panel or something like that. Figure is very hard to understand b/c 'TCO (DU)' looks like a label. The visual comparison is pretty difficult I think in part because of the size and the model right panels showing full resolution and the data having all that white space.

C11626

26912 – not just noise, lack of sensitivity and vertical resolution. Only HIRDLS was conceptually designed to address this issue, and HIRDLS information on the issue is greatly compromised by the blockage.

26913 – 'cannot provide observations of all the individual events that could lead to a general, comprehensive integration' – people would have said this before launch – even with the best possible performance for all the instruments. Even before launch there was a role defined for CTMs and data assimilation systems to obtain such a global integral that would (hopefully) be consistent with Aura measurements.

26914 – 'suggesting low sensitivities and great noise' – the whole tone of discussion would change if here (and elsewhere) 'suggesting' were replaced with 'consistent with' This isn't a discovery.

those negative MLS values are telling you that single profile comparisons are not sensible because of single profile precision. You could look at an ancient paper (Schoeberl et al., GRL, 1993) that showed a reasonable way to use MLS ClO – prior to that paper people were interpreting the high ClO 'hits' as evidence of polar vortex leakage to middle latitudes.

26914 - 'The CTM-TES comparisons are almost always improved. . .' you really need to think about how you consider the TES averaging kernel and the prior. The TES data have limited information – the challenge to the user is to tease out the limited information and understand the problems for which the data are relevant. One way to look at some of these comparisons is that they are an example of an inappropriate comparison.

Poor syntax or grammar

26898 I 7-8 it is revealed in many cases that all four Aura datasets have some skill in catching the STE process, while missing many of them – this same statement appears in the conclusion (26916). For the 8 cases . . . despite missing many of them. Does

C11627

that mean many of the 8? Many of the others in the model?

26916 L11 – better CONTRAINT (not better 'constrains')

26909 l 14 – simulated swatch are resented

C11628