First, I must apologise to the authors and the editors for the lateness of this review. I did read the paper a while back and formed my (positive overall) opinion, but was distracted from completing the review for various reasons. In the intervening time, I see two other reviews have come in and I find myself having little extra to add. My feelings resonate with those expressed in the excellent discussion from reviewer #2, and I am pleased to see that the response from the authors recognizes the validity of the points made.

I do think this is a nice study, well worth publishing in ACP. As reviewer #2 states, the Aura mission was never designed to fully capture and quantify all strato-
sphere/troposphere exchange events (although were it not for the cruel blow dealt to the HIRDLS instrument, we would be a lot closer), and recasting the paper as a 'confirmation' rather than a 'discovery' of this is a good thing to do.

My only somewhat major concern (also picked up by reviewer #2) is the discussion associated with the choice to include raw TES-Model comparisons in addition to those using the TES observation operator. I have no problem with this choice itself (particularly for the arguably somewhat qualitative discussion of figures 3 and 4). The scatter plots in figures 5-8 do show both comparisons. However, I found the discussion (26905 line 20 through 26906 line 5) rather oddly framed. If I’ve understood what is happening correctly, the model has a high bias in lower stratospheric ozone. If one therefore applies the TES observation operator to the model profile, this lower stratospheric bias propagates down to the upper troposphere. Is this a correct assessment? If so, then to say ‘The artificial high bias in the UT introduced by the TES operator’ (26905, line 20) seems to put all the blame on TES. I would rephrase this to make it clear that the TES operator propagates existing model biases to different altitudes.

My only other general comment is that I do find the figures a little cluttered and hard to read. Part of this is because of the transition from line-art to bitmaps that someone did (either the authors or the journal, why do people feel this compulsion!). Are figures 3 and 4 intended to be single column (i.e., pretty much at the size shown?). If so, they could be much more efficient - drop the y-axes in the right hand column, reduce the amount of white space, make the labels much bigger. The symbols in (b) are too small to see (though keep them the same size for OMI given the density). For figure 5 and its friends, I suggest the color bars should be wider. Alternatively, I wonder if you could chose a different normalization that would allow you to usefully have one color bar for all plots, that would increase the space that could be given over to data.

— Minor comments

Abstract - line 1: What does ‘instantaneous’ mean in this context? These measure-
ments are effectively coincident to within an hour or less. Does that count as 'instantaneous'? What does this word add?

26899, line 28: The model sees a whole elephant, but is it the same elephant that the instruments observe? It might be a mammoth! You might want to rephrase your analogy to capture that nuance.

26900, line 20: 'full' chemistry by whose definition? Not mother nature’s presumably.

26900, line 28: Could you expand on why the other definitions are more awkward for your model.

26901, line 20: Actually, not that many studies use L3 data (in fact there are no official L3 data for MLS for example).

26902, line 13: 'nearest and above the' is odd wording.

26903, line 8: Does OMI even have 3-6 layers in the troposphere? Actually, the discussion here and that for TES in the following paragraph feel a little disjoint. Would it be possible to make them more parallel? You discuss averaging kernels for TES but not for OMI. Are they available?

26904, line 9/10: Linear interpolation in log pressure I think should be the way is done (and probably the way you do do it).

26906, line 24: 'rest OF THE analysis'

26908, line 25: 'underestimates' is unclear here, what is underestimated, the O3 abundance? The O3 amplitude of the fold? The vertical depth to which the fold penetrates?

26909, line 20, 'on contrast' -> 'by contrast'

26910, line 6: The sentence that starts here feels like it should go before the one about tropopause height.

26901, line 28: It is my understanding that such MLS anomalies likely reflect contami-
nation from thick clouds, might be good to mention that.

26911, line 24: 'mismatches' is rather unclear, and then is not expanded upon. What aspects are mismatched (size, location, ozone amplitude etc.)

26915, line 9: 'normally DISTRIBUTED error'? (or Gaussian?)

Figure 2. Numbers on contours are hard to read, perhaps use discrete rather than continuous colors.

--- Follow up on response to reviewer #2

C12653, last paragraph: It is my understanding that there are circumstances when TES can resolve boundary layer ozone. These are typified by high thermal contrast between the surface and the lower troposphere. Perhaps this is one such case?

Interactive comment on Atmos. Chem. Phys. Discuss., 11, 26897, 2011.