Responses to Anonymous Referee #2

1. This paper by Boxe et al. presents a timely and much needed comparison of TES O_3 profiles with sondes at high latitudes with a high degree of spatial and temporal overlap. As such, the validation efforts presented here take significant steps forward over previous studies. The ability to quantify the empirical RMS error through multiple TES samplings of a single air mass is important. Overall, the content is well suited for ACP and while the paper could use some clarifications in a few areas, I believe it will be ready for publication in short order. We appreciate reviewer #2's comments.

2. The comparison between the empirical and theoretical expected errors comes across as overly optimistic. As this is one of the major findings of this work, it deserves to be treated a bit more carefully. For example, if the theoretical errors are 6% and the empirical errors are 9%, the authors are correct in saying that the difference between the two is small, only 3%. What they don't say, however, is that the theoretical errors are underestimates by 50%. Surely the latter framing of the comparison isn't as flattering, but it shouldn't be overlooked. Regarding the overall trends, I calculated a correlation of the mean theoretical errors and mean empirical errors in Table 4 to be only 0.52. So a regression of the empirical errors on the theoretical errors has an R_2 of only 0.27. It would be preferable to provide statistical quantification like this rather than qualitative statements such as "generally consistent". This type of analysis should be performed on the actual profile errors, not the mean profile errors as I did here. Statistical results should be presented in the discussion, abstract and conclusion. Statements therein that the empirical and theoretical errors are "consistent" might need to be reevaluated, or might be better supported by the new quantitative analysis. The random errors are with respect to the estimate and are, as the referee comments, not meaningful with respect to themselves. To address these concerns we have (1) Added language on why we expect the actual to be larger than the estimated errors.

(2) removed the word small and only state the values.

(3) Its not clear how adding correlations between the averaged actual and averaged calculated errors would add to the discussion because these values do not vary much from estimate to estimate. However, we can say that the altitude distributions for the errors are consistent which is our original intent. For example, the actually and estimated errors peak in the free troposphere and are lower in the stratosphere. We have added this language to the paper.

3. The discussion of the averaging kernels is bit sloppy. Some of the text in section 6.2 is repeated, e.g., "For instance, Fig. 4a and b show . . . " The discussion of aspects of Fig 4 following this doesn't clearly refer to 4.a vs 4b, which are different enough to warrant distinction. Overall, it's hard to follow the interpretation of the averaging kernel plots because the following was never clarified: the plots show @^x/@x . Does the y-axis or the orange color refer to the vertical coordinate of ^x or of x? The rows of the averaging kernels indicates how a given level is affected by all other ozone variations at all other pressure levels. This has been clarified in the text.

4. In Fig 1, the smoothed sonde profile nearly matches the a priori profile. This means that either x_{sonde} □ x_a = 0 or A_{xx} = 0. Since the TES O_3 profile doesn't

match A_{xx}, it looks like the instrument did have some sensitivity and A_{xx} is not zero. But from the figure, it also doesn't appear that $x_{sonde} = x_a$. Could this type of behavior be explained? That the *a priori* and sonde smoothed by the averaging kernel and *a priori* are nearly identitical is coincidental for this comparison. The reason that the TES estimate does not match the smoothed sonde profile is due to the bias in the TES estimate which is quantified in the bottom left panel in Figure 1. The other profiles tend to show stronger differences between the sonde and *a priori* in the upper troposphere and stratosphere.

5. 27269, 25: suggest "altitude" □ > "vertical distribution" This suggestion was incorporated in the manuscript.

6. 27277: was S_b defined? The parameter "b" is defined After Equation 1. To reduce confusion we have added language that these refer to un-retrieved parameters that can include temperature, water, or spectroscopic strengths. Sb is then the covariance for these terms. We have added this language to the manuscript to reduce confusion.

7. 27278: $S_{s\,\%}$? This error has been amended in the manuscript.

8. 27279, 10: is it standard practice to take a straight average of the averaging kernels? Or should it be a weighted average that accounts for varying measurement noise? For taking an average, where we can calculate the mean and variance about the mean then one averages the averaging kernels as noted in Equation (5). This effectively also weights the profiles accordingly.

9. 27282, 2: without repeating too much what is left to previous papers, could the "temperature difference criteria" be explained a bit more? We incorporated this suggestion in our manuscript.