Response to RC3

Thank you for your careful review—please see our response in blue in the text below.

The paper describes analysis using synthetic data from a global model to evaluate the impact that broad averaging kernels have on deriving vertically resolved ozone trends from satellite observations. The topic is important because of society's need to determine the timing of ozone recovery as ODSs decline and to assess the ozone response to greenhouse gases. The paper is focused on one specific aspect of the problem but this aspect is generally relevant because many trend estimates rely on SBUV ozone profile timeseries stitched together from multiple platforms. The results will be of interest to those who compile and/or interpret trend estimates and to everyone who want to know the limitations of published trends.

The paper is clearly motivated and well written. The authors thoroughly describe the caveats to their analysis. I recommend some revisions before final publications, as itemized below.

General comments:

1. The units for the vertically resolved ozone from the model data (DU/layer) make it difficult to compare the simulated ozone with measurements or other models without knowledge of the model vertical grid. Even with knowledge of this, calculations would be needed. Can you show the ozone for these plots in the more conventional units such as ppm or number density? This applies to Figure 4 and Figure 5a. In this study, we are more focused on ozone variability and trends than we are with abundances or data set comparisons. The residuals in Figure 4 are only shown for illustrative purposes, to show the inner workings of the ToE method. In Figure 5, ozone abundances are shown as reference to interpret the other panels. Using vmr or ppm would be equally valid choices, but we initially chose DU/layer because they are the native units of the SBUV data set and the units used when applying the SBUV kernels, and because DU/layer profiles are easy to compare to total column results. However, to address your point, we converted DU/layer to DU/km to remove any effects of the vertical grid.

2. It was not clear what the advantage is of interpolating the model profiles to the SBUV grid before applying the averaging kernels (line 188-189). Doesn't this already remove some of the information about vertical structure that you are trying to identify in your study? The kernels are produced at the resolution of the SBUV data set (see Figure 2 for an example), so the model data *must* be interpolated to that resolution first in order to produce synthetic data. We now explain this in the text, and also specify in the caption of Figure 2 that the model profile shown is interpolated onto the SBUV vertical grid in order to isolate the effects of the averaging kernels.

3. Section 3.2.1 is hard to follow. Variables are defined (b, y, etc.) but the equation is not given. Since the final paragraph of this subsection appears to be key to the results that follow, it is important that it be clear. For example, do you compare the two emergence estimates y or y*? I could not tell which was identified as the time of emergence in Figure 5b-5d. Detailing the formalism of the ToE method here would require a significant amount of space, so we provide an overview + a Figure to illustrate, along with the reference. We reworked this section to include more detail and an example. We also specified in the caption of Figure 5 that the ideal time of emergence shown is y, to match

the title of the color bar, and in Figure 6 that the relative differences shown are calculated as $\dfrac{y^* - y}{y}$. We hope this will clarify that we compare y to y* (see Section 3.2.2) to quantify errors on the emergence of trends attributable to the averaging kernels. y is the 'omniscient' or 'ideal' ToE, and y* is the ToE corrected for the effects of the SBUV kernels, as stated at the end of Section 3.2.1.

Minor comments:

1. (line 98) By "optimistic" do you mean too low? Correct, we've replaced with 'underestimated'.
2. (line 287) The reminder that one should not over-interpret crossing an arbitrary threshold is appropriate; I'm glad to see it mentioned.
3. (line 320) Maybe I missed it but I think this is the first mention that the time to emergence depends on the magnitude of the trend. This is intuitively known but perhaps should be included in the introduction as one of the factors limiting trend detection. We've added a short sentence to this effect at the beginning of Section 3.2
4. At line 381, you state "Altogether, this analysis shows why trends should be analyzed as vertical profiles rather than at individual vertical levels." This is a good summary of the results shown in Figure 7 but is not quantitative. How do you decide whether that criterion has been met? For example, in Figure 8a the time to emergence is detectable over part of the profile but not all of it. However the text indicates that the trend in the upper levels is identified as detectable. RC2 had a similar comment. Our statement was meant to echo the general result that SBUV kernels can distort the trend profile; we've replaced it with a more general comment about the importance of accounting for averaging kernels when analyzing vertically resolved trends.