Response to Reviewers for manuscript 'Maximizing Ozone Signals Among Chemical, Meteorological, and Climatological Variability'

Below we work through each of the reviewers' comments, with the comments in black and our responses in red. We also include any alterations to the text in red after our responses with the specific additions indicated with underlines.

Reviewer 1

The authors have addressed my main concerns satisfactorily with the improved Figure 2 and more careful caveating. I recommend the manuscript for publication with a few minor revisions/technical corrections, listed below.

Line 70: "Pawson" not "Dawson"

Corrected, Line 76.

Line 272-274: This sentence is rather convoluted. Please split in two or reword.

This sentence has been updated and clarified.

Lines 279-283: "*Figure 2 compares summertime (JJA) maximum daily 8-hour average ozone (MDA8 O_3) from the present-day model simulation (MOZ_2000, Figure 2a) to the year-2000 CASTNET observations (Figure 2b). Figures 2c and 2d plot the MDA8 O_3 standard deviation and variability for MOZ_2000, while Figures 2d and 2e compare the mean summertime MDA8 O_3 for the future simulations (MOZ_2050 and MOZ_2100).*"

Line 288: But which is the most relevant metric, relative standard deviation or absolute? You later discuss thresholds in units of ppb, which suggests an absolute metric.

Both are relevant metrics that give a different view of the ozone comparisons. We have clarified the sentence to make this clearer.

Lines 301-304: "*We consider both the standard deviation (ppb) and a mean-normalized standard deviation (as a percentage). The normalized standard deviation allows for a more direct comparison of the shape of the MDA8 O_3 distributions between the simulations and available observations, which accounts for the noted ozone biases (Figures 2b,c and Table 1).*"

Reviewer 2

General Comments

The revisions to this manuscript addressed many of the concerns expressed in my original review, and helped to improve significantly the quality of the manuscript. The framework of a "sliding-scale" view of ozone variability is a particularly innovative new contribution from this work. However, I have a few remaining concerns about the manuscript, which I list below. These concerns could be addressed through minor revisions and would not require new simulations or new substantive analyses.

Define better what is meant by "ozone signals". The intended meaning of this term is never defined. Throughout most of the paper, the emphasis is on the use of spatial and temporal averaging to decrease the meteorological variability. In this case, a robust estimate of the mean ozone mixing ratio (averaged appropriately) is the 'signal' that can be detected within a specified error tolerance. While this is one important type of ozone signal, the Abstract and Introduction also discuss ozone change 'signals', resulting for instance from emissions change or climate change, as opposed to meteorological variability. This type of ozone 'signal' is not discussed adequately in the paper. For instance, how can a forced trend in ozone concentrations be detected? Given the recommendations of 10-15 year averaging period, would this suggest the need for a 10-15 year average before an assumed instantaneous change in emissions, followed by a 10-15 year average afterwards? Presumably, the magnitude of the shift relative to the magnitude of the internal meteorological variability would be relevant to the required averaging period. Some discussion of such issues would greatly improve the usefulness and applicability of the methodologies in this paper.

We have added the following two paragraph to the end of the discussion section to address this point:

Lines 546-576: "*Smaller signals require longer temporal averaging periods to identify. Figure 4 shows that a 0.5 ppb MDA8 O_3 signal will emerge after 15 - 20 years of temporal averaging. The range here reflects different spatial averaging domains, with larger domains requiring shorter temporal averaging windows than smaller domains (i.e. 15 years for averaging over the Continental US and 20 years for averaging over the Northeastern US). This would mean that an average trend of 0.25 - 0.33 ppb/year would require a time series of at least 15 years to identify. Similarly, a 1.0 ppb MDA8 O_3 signal emerges after 7 - 15 years, which indicates an average trend of 0.14 -0.67 ppb/year would take at least 7 years to identify. Finally, a 5 ppb signal can be identified in less than 3 years, which indicates that an average trend of 1.67 ppb/year or greater would only require a 3-year time series. This presents particular difficulties if the ozone signal of interest is a trend spanning a time period on the same order. The 10 - 15 year averaging time scale we propose translates into a length of time beyond which you are likely to not see spurious trends above 0.5 ppb, but there are many cases in which the identification of a small trend is desired with less than 10 - 15 years of available data. For instance, Jiang et al. (2018) have found that NO_x emissions reductions since 2005 are not as strong as previously expected, showing a significant slowdown beginning in 2011. This has large implications for ozone and for short-term decisions for air quality managers within the United States, who have to promulgate policies on short-term scales without the luxury of postponing action until longer and more complete data sets become available. As we have shown, spatial and temporal variability due to meteorology is high, and the identification and quantification of trends over 5, 10, or 15 years is difficult, particularly at small spatial scales.*

However, as we have shown, a consideration of the impact on variability – and how variability changes over time – is often pivotal to understand the nature of the signals being examined. In *this paper, we have provided methods for quantifying the spatial and temporal variability and strategies for determining which types of signals are likely detectable at particular temporal and spatial scales. Some signals, especially small signals at small scales, are simply not large enough to emerge from the variability, and thus may not be detectable without additional data or expanding the temporal and spatial averaging scales used for analysis. Quantifying the signal-to-noise ratio at a variety of spatial scales, and determining an acceptable threshold of a particular signal, could be one accessible method for providing this context. The risk in neglecting the quantification and contextualization of the magnitude of the ozone signal relative to the magnitude of the variability induced by the internal meteorology – and the impact of temporal and spatial averaging – is primarily the risk of drawing conclusions that are more sensitive to a particular peculiarity in the underlying variability rather than the signal itself."*

We've also added the Jiang et al., 2018 reference to the references.

Minor Comments

Abstract, lines 43-46 -- Another possible reason for considering a shorter averaging period might be if a particularly large change (e.g., from rapid emissions reductions) were detected. If this change were large relative to the meteorological variability, a shorter averaging window could be used.

We have modified the abstract to make this point:

Lines 42-46: "*For signals that are large compared to the meteorological variability (e.g. strong emissions reductions), shorter averaging periods and smaller spatial averaging regions may be sufficient, but for many signals that are smaller than or comparable in magnitude to the underlying meteorological variability, we recommend temporal averaging of 10 – 15 years combine with some level of spatial averaging (up to several hundred kilometers).*"

We have also added a line to the conclusion:

Lines 620-622: "*For signals that are large compared to the underlying meteorological variability (e.g. strong emissions reductions), shorter averaging windows and smaller spatial regions may be used.*"

Section 2.1, lines 193-195 -- Sentence fragment. Revise.

This has been corrected:

Lines 200 – 202: "Offline forced meteorology *is taken* from the Modern-Era Retrospective analysis for Research and Applications (MERRA) reanalysis product…"

Section 4, line 447 -- "three" --> "four"

Corrected (Line 465).

Section 5, lines 538-540 -- Clarify what is meant by "the resultant magnitude of these changes" and "these changes and signals".

The sentence has been clarified:

Lines 591-592: "*However, these ozone signals (e.g. temporal trends or regional averages) are frequently small when* compared to the magnitude of the day-to-day ozone variability, and *thus* detecting these signals can be challenging."

Section 5, line 551 -- "out" --> "our"

Corrected (Line 604).

Section 5, line 558 -- Clarify the timescale over which the ozone variations of +/- 10-20 ppbv occur. Interannual variability of summer mean?
Clarified:

Lines 615-616: "…summertime *MDA8* O_3 variability is largest at the smallest *spatial and temporal scales*…"