

## August 2019 Review of acp-2018-1174

"Estimating background contributions and U.S. anthropogenic enhancements to maximum ozone concentrations in the northern U.S." by D. D. Parrish and C. A. Ennis

This manuscript is much improved! My concerns have been adequately addressed, and the readability is an order of magnitude better than the last version. This work is now presented in a way that others can understand and evaluate it. I look forward to the substantive and reasoned discussion which may follow in ACP and in the community at large.

Given the substantial amount of new text that the authors have generated, I have just a few comments that may improve the presentation.

A) The new section 2.3 describing the additional analyses was confusing to me.

- Should line 11 read "we also derive  $y_0$  through THREE somewhat different approaches that... " ?
- I recommend you remove the last sentence of the first paragraph, as it is repeated in the third paragraph.
- The third and fourth paragraphs seem to jump around a bit and are not presented in the same order as they are used later in the manuscript. Please see if this suggested re-organization describes what you intended:

Two additional approaches can approximately quantify the value of  $\tau$  in the northeastern states; both of these approaches assume that constant values of  $y_0$  and  $\tau$  are appropriate for all ODV time series included in each analysis. First, a linear fit to the initial period of decreasing ODVs provides direct information regarding the magnitude of  $\tau$  and  $y_0$ . The absolute value and the time derivative of Equation 1 when evaluated at year 2000 are  $y_0 + A$  and  $-A/\tau$ , respectively. Fits to two ODV time series provide four parameters ( $\tau$ ,  $y_0$ ,  $A_1$  and  $A_2$ ) if the  $\tau$  and  $y_0$  values are the same for the two time series. Algebraic manipulation gives  $\tau = -\Delta_{\text{intercept}} / \Delta_{\text{slope}}$ , where  $\Delta$  indicates the difference in the subscripted parameter between the two linear fits, and  $y_0 = (\Sigma_{\text{intercept}} + \tau * \Sigma_{\text{slope}})/2$ , where  $\Sigma$  indicates the sum of the subscripted parameter from the two fits. A complication with this approach is that the linear fits to time periods of significant length give biased measures of the derivative and year 2000 value of Equation 1; however, this bias can be corrected to first order through numerical comparison of a linear fit to the selected period of the exponential fit. The second approach is described in Section 2.4 of Parrish et al. (2017a) and is adapted here to the northeastern U.S. ODV time series. It uses an iterative, non-linear regression analysis that simultaneously derives values for  $\tau$  and  $y_0$ , plus the  $A$  parameter for each ODV time series included in the analysis. These two additional approaches help to constrain the uncertainty of the assumed value of  $\tau$  (21.9 years).

B) When the results of the method shown in Fig 13 are described in section 3.3.2, there is insufficient information to allow the reader to evaluate your "bias correction". The description of that "first order" correction "through numerical comparison of a linear fit" in section 2.3 does not shed much additional light. If it is possible to give a bit more information (or a reference?), the reader can have more confidence that your "corrections" reduce uncertainty, rather than increase it.

And for clarity of communication, I would suggest that "intercept" is not strictly the correct noun here, as the plots you show in Figure 13 would have fitting parameters with an intercept at Year 0. I recommend you either replot with the horizontal axis as Year Minus 2000 instead of Year, or stick to using the descriptor "value" or "absolute value at Year 2000" (e.g., revised manuscript, page 8, line 4). This will also remove any confusion about if the "intercept" for the Maine data set is to be chosen at 2000 or at 1991.

C) "Variance" is discussed often in this manuscript, and it seems that the meaning is slightly different in different contexts. On page 6, variance (in units of ppb) is described in terms of RMSD between a dataset and its fit. But then on page 9 and in Table 1, variance has units of ppb<sup>2</sup> and appears to be (but is never actually\*) defined as the square of the standard deviation of a dataset. Then both types are used back and forth through the remainder of the manuscript. It's especially tricky to parse at the top of page 19, where you are using both the ppb<sup>2</sup> values (like 251 and 13.4), but are also referring to Figure 9, which gives RMSDs of 3.5 and 5.6 ppb. I don't have a nice, tidy suggestion, but I would ask that the authors take a few minutes to think about how they might make their dual-use of this word a little bit easier for the reader.

\*It is given quite succinctly on page 6 of the Response document, but I couldn't find it in the manuscript itself. I recommend including it in the manuscript, especially since the rounding effects are just enough to make it questionable (e.g., pg 18: 3.7 \* 3.7 = 13.69 = 13.7, not 13.4).

And by the way, the Figure 3 legend says 252 ppb<sup>2</sup>, not 251 ppb<sup>2</sup>.

D) And a few small suggestions:

- Pg 9, line 14, add a comma: "Figures 3 and 4, and averages with standard deviations..."
- Pg 13, line 15 change to: Figure 10 plots the time series of these state maximum ODVs recorded in each year with respective fits over the 2010-2017 period.
- Is that really supposed to be 2010? In Fig 10, the solid fit lines start at 2000.
- Pg 13, line 19 change to: only the largest of the state's ODVs in a given year... {to match singular "across the state" later in the sentence"}
- Pg 13, line 27: is this "NYC urban maximum" a new subset? Or is it the same as the data which generated the red dashed line in Fig 7? This is not a big deal, but I got distracted for a while trying to figure it out.
- And a related question: In Fig S13, Does the fact that the 1:1 line goes through NJ data points at all values of "NYC urban max" indicate that the

"NYC urban max" data is actually entirely from NJ? (I understand the geography; that's not my point. If all the "reference" data is contained in the data plotted against it in Fig S13, that seems a bit circular.)

- Pg 13, line 28, remove comma after "is selected".
- Suggested clarification and nuance regarding spatial variability of ODVs on page 16: the derived US background ODV has significant variability on a continental scale. Within..... significantly smaller than in any of the western US regions, but shows no discernable spatial variability within this region. For context... NAAQS of 70 ppb. In contrast, in the northeastern US the A...
- Pg 18, top line has an extra space
- Pg 18, line18, strike "by" before (Fiore et al., 2015)
- Pg 18, line 21: change "limit" to "limited"
- Competing Interests section needs to be updated, as does the last line of the Acknowledgements.
- Fig 9 caption: The explanation of dashed, dotted, and y0 is awkward. Maybe try rewriting with dashed, then dotted, then y0 last.
- Fig 14, vertical axis label: Is this really US background, not NAB?
- Supplemental, pg 2, line 11: One...area
- Supplemental, line 21: some sites in the NY...
- Supplemental, line 32: neighboring states until 2013.
- Figs S3-S10: the gray circles are a bit too faint to really see. Could they be darker?
- Fig S10 legend, lower panel: green sites are called "interior" or "rural inland" elsewhere. I recommend choosing one of those in the lower legend for consistency.
- Fig S16: the blue dashed line is different from all the other blue dotted lines.