

# Response to non-interactive referee comments on “A mid-latitude stratosphere dynamical index for attribution of stratospheric variability and improved ozone and temperature trend analysis” by William T. Ball et al

## General comments relevant to both referees:

We thank both reviewers for their continuing input. Please see our comments (blue) below in response to the reviewers (black). Any major changes to the text (see below) have been put in bold font in the updated manuscript.

### A. Karpechko (Referee)

The manuscript has improved compared to the previous version. I think the authors have made a good job. There are only few minor comments which I have on the revised version:

1. I believe the fact that your index does not correlate with the heat flux at 100 hPa (which represent the amount of wave activity entering the stratosphere in extratropics) is important and worth highlighting in conclusions and perhaps also in the abstract. The upper branch of BD is indeed driven by planetary waves and one would expect the 100hPa heat flux being an important proxy for it. But it is likely wave propagation and dissipation within the stratosphere that give raise the variability represented by your index, and this is not captured by the 100hPa heat flux.

Agreed. We think this is interesting and worth pursuing in the future.

2. P8L4 : ‘It seems surprising that these dynamical signals survive other processes, such as chemistry and radiative effects, on a monthly timescale and we suggest 5 this warrants further investigation, though this is beyond the scope of this paper.’

A dynamical event lasting just for a few days (or a sequence of such events) may influence the monthly mean statistics (temperatures, ozone, etc). So there is no need for an event to last for a month. Perhaps this is a sufficient explanation?

We have removed this sentence from the manuscript.

3. Figure 5 and Section 3.2:

How do you calculate TEMS and EPFD? References to the formulas would be helpful. Also do you use daily mean data or hourly (e.g. 6-hourly) data? Daily mean data may be not sufficient to represent forcing due to some tropical waves. This may be responsible for patchy patterns in the tropics, such as the negative anomalies surrounded by positive anomalies in Fig. 5a. Please give more details about your calculations

We have added the following sentence to the third paragraph of section 3.2: “We used six-hourly model output to calculate the monthly means and use equations 3.5.1 and 3.5.3 from (Andrews and Holton, 1987) to perform the calculations.”

*Reference: Andrews, D.G., Holton, J.R. and Leovy, C.B. (1987) Middle Atmosphere, Dynamics. Academic Press Inc., 491 p.*

4. Section 4:

I realize that I do not understand how you construct the index, specifically how you merge the hemispheres. Is it so that June-October values come from SH and November-May from NH? Then why there are several different periods mentioned in Section 4.1?

The DJF and JJA periods are used to find the peak correlation. Once found, we complete the remaining months missing from the index (i.e. MAM and SON) putting N-MA with DJF and M-AM with JJA. We think the confusion may come from the first sentence paragraph 3 in section 4: "For March--May and September--November months, ..." so we remove this sentence to reduce confusion.

5. Figure 6: I believe positive values of EP-flux divergence indicate decreased wave activity, which is consistent with weakened meridional circulation and high temperatures in the tropics (high-T composite).

We have corrected this.

6. I think 1-sigma confidence interval is more commonly used wording than 68% confidence interval.

While our confidence intervals are close to Gaussian, we use 68% because it is not necessarily Gaussian in every case.

7. P17L5: 'Nevertheless, use of AR0 or AR1, only appears to influence the uncertainties 5 on the trend estimates in a clear way, and only the mean values to a small degree (see Fig. 13).'

I do not understand this sentence. In any case I think it is enough to consider AR1 results and just skip the AR0 results.

We have removed discussion of AR results; due to an error in our previous analysis we have used AR2 instead since this is actually more appropriate (see response below to Lon Hood).

8. Figure 13 shows that the inclusion of the UBDC index makes the trends more negative, i.e. the trends are increased in the absolute values around 6-10 hPa, although within the uncertainty estimates, am I right? It is not absolutely clear from your text what do you mean by saying that UBDC index leads to a decrease of ozone trends: in terms of absolute values the trends increase.

This correct. We hope the new Figure makes this clearer. The absolute trends are generally positive, but they are almost always reduced in this region by ~0.5-1%; we made a minor adjustment to the text to make this point clear.

9. P18L12: 'temperature and ozone TREND estimates'

We have added this word.

### ***L. Hood (Referee)***

As stated in my first review, this is overall a valuable and significant effort to improve statistical estimation of stratospheric ozone and temperature trends and natural variability. The presentation is generally excellent and the changes made in the new version of the manuscript are very responsive to the criticisms provided by myself and the other reviewer. However, some minor but important revisions are still needed in my opinion.

First, Figure 13 of the original manuscript was the key figure of the paper because it showed the improvement in trend estimates for the four ozone datasets and the three temperature datasets when the MLS index was added to the regression model. In the new version of the manuscript, the old Figure 13 has been replaced with a new figure which shows the improvements for only one ozone dataset (SWOOSH) and one temperature dataset (SSU). The improvements are shown for cases with and without including a correction for autocorrelation of the residuals in the regression analysis. This is something of a step backwards because (a) there is no need to include the AR0 results because they do not correct for the autocorrelation; and (b) the improvements for the other three ozone datasets and the other two temperature datasets are not shown. Correction for autocorrelation is standard procedure in regression analyses so only the AR1 results are needed. One could put such a figure in an appendix or in supporting online information if it is desired to show the difference between the AR0 and AR1 results but it should not be in the main paper in my opinion. My suggestion is to correct the old Figure 13 using the AR1 results and put that in the final paper. Then readers can see the improvements for themselves for the various datasets. Adding the AR0 results only adds unnecessary complication to the figure in my opinion.

We have done this as requested, and the original figure has been replaced. We noticed a small error in our analysis routine, which means it was an error to choose AR1. We checked with a Durbin-Watson test and found that while AR1 was indeed at least necessary, AR2 was required to reach an optimal result. We thus have reprocessed all figures and state in the text "We consider the use of 'AR2' auto-regressive modelling through the procedure of (Cochrane and Orcutt, 1949) in all cases; see (Tiao et al., 1990) for a discussion of AR. The use of second-order auto-regression was determined after assessing the regression analysis using a Durbin-Watson test, which showed that AR1 was necessary, but not sufficient to account for auto-correlation in the residuals, and that AR2 was sufficient." Note that values (improvements and errors) changed only marginally.

Following this, we have followed Lon's advice in removing discussion on AR0 (or AR1), and replacing Fig 13 with the old version, but with a clearer plot given the multiple extra profiles included.

Second, in the abstract and conclusions, the improvements are stated as maximum improvements ("the index can account for up to 60% of the total variability ... the uncertainty on all multiple linear regression coefficients can be reduced by up to 30% and 25% in temperature and ozone, respectively ..."). While these statements are true, it would be more accurate to give the range of the improvements as a function of pressure level and dataset. Looking at the new Figure 13, the improvement in the

"SWOOSH ozone error bar is indeed 25% at 2.5 hPa but it decreases to about zero by 1 hPa and by 10 hPa. This is a rather strong altitude dependence and readers should be made aware of this."

We have added extra text to both the conclusion and the abstract to address Lon's concern of being clear about where the index is effective. The point regarding

"Also, how do these results depend on the different datasets? Are the improvements less for the other datasets? This should be shown in the final version."

This information is in Figure 10, where the error changes are included in the right panel of each subplot. Nevertheless we now briefly mention the different datasets in the main text, and the improved clarity of the revised Fig 13 makes it much easier to see the error bar improvement.