

Response to reviewer 1:

We would like to thank the reviewer for the valuable comments and suggestions, particularly for picking up on several inconsistencies and errors. The text and tables have been changed according to these comments and point-by-point responses can be found below. Reviewer comments are shown in italics, while the author responses are shown in grey and changes made to the manuscript in bold blue and red (changes highlighted with *latexdiff*).

1 Major Comments

Section 2.2 talks about how the monthly mean profiles of each of the seven data sets are all converted to a common grid. Is this a common pressure and latitude grid or just a common pressure grid for each? What is the resolution of the pressure grid that is used? If a spatial interpolation is performed, how exactly is that done?

This is a common pressure and latitude grid, however, no spatial interpolation has been carried out. For data sets that were natively at 5° horizontal resolution, we averaged them into 10° latitude bins to be consistent with the other (lower) resolution data sets. This has been more clearly described in the revised manuscript.

No spatial interpolation in the horizontal was applied; instead, data were averaged from their native resolution (5° or 10°) into latitude bands of interest (see Sect. 3). Furthermore no additional screening was applied to any of the seven data sets.

Equation 1 shows the regression model, where each coefficient A-H is actually some number of coefficients as part of Fourier pairs multiplied by some predictor parameter. The text on page 25072 states that the number in the subscript is the number of Fourier pairs. Does this mean that if NB=2 there are 2 pairs and thus 4 terms (12 month and 6 month, sine and cosine for each) or are there only 2 terms (12 month, sine and cosine) and thus 1 pair? Additionally, shouldn't each coefficient also have a constant term to represent the mean value (not seasonally varying) of each predictor?

If NB=2 there are indeed 4 terms (2 sine and 2 cosine functions), as well as one constant term representing the mean value. This has also been clarified in the revised manuscript.

Furthermore, each coefficient has a constant term, which represents the mean value (seasonally unvarying) of each predictor. The subscript of each term A-H indicates how many Fourier pairs the term was expanded into to account for ~~seasonality~~ the seasonal dependence of the basis functions on the ozone anomalies (Bodeker et al., 1998); for example NB=2 indicates two Fourier pairs (two sine, two cosine). Each coefficient also has a constant term.

At what vertical and spatial resolution is the regression model applied to each data set? Perhaps this ties in to a previous question about whether there is a standard spatial grid.

The regression model is applied to each level on the common pressure grid and for each latitude band region (35-60°N, 20°N-20°S, 35-60°S). This has been specified in the revised manuscript.

The regression was applied to each pressure level for each latitude band average and trends were only estimated if more than 50% of the data for a particular level were available (i.e. more than half of all months had data).

Section 3.1 discusses the annual cycles of each data set, computed simply as the mean of a particular month over the entire record and not from coefficient A in the regression. Do the results of coefficient A of the regression agree reasonably with the results of section 3.1? If not, their removal to derive trends from anomalies could introduce biases.

The annual cycles removed from the regression (coefficient A) agree well with the annual cycles shown in Figure 3 (compare with figure below). The choice was made to show rather the mean values of particular month to preserve the logical order of the paper, i.e. the results from the regression are presented later in the paper. Given that there were only very small differences between the computed

annual cycles and those derived from the regression model, we felt this was a justifiable choice.

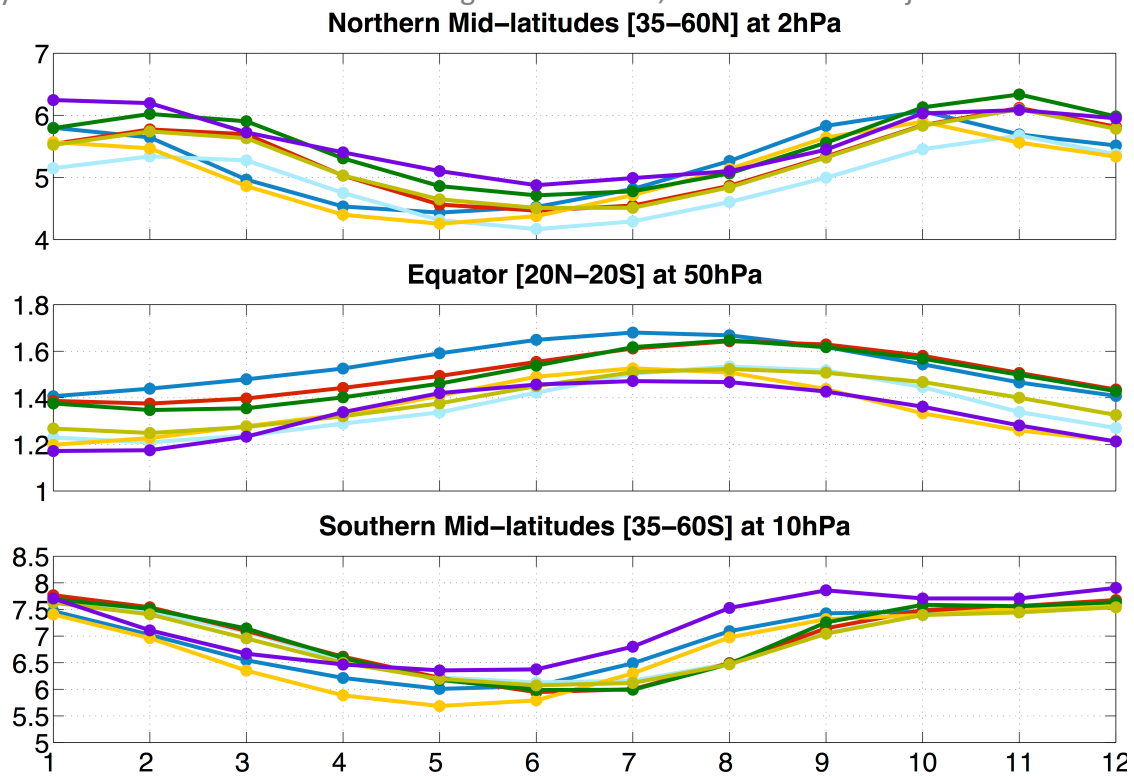


Figure 1: Annual cycle derived from the multiple-linear regression analysis.

The SAGE-OSIRIS anomalies from the MDM in Fig. 5b look odd around 2001. The data approach and then are identically zero in the time period in late 2001 when no data exist there. This behaviour is likely an anomaly of smoothing involving non-existent data and needs to be corrected. Possibly similar results are seen near 1986. I would think the code used to generate the data sets in each of these figures needs to be corrected to properly handle data gaps and rerun on each data set to ensure any potential anomalies are corrected. I cannot tell, given the overlapping data, if a similar problem exists for Fig. 6.

We would like to thank the reviewer for spotting this. There was indeed a bug in the code used to produce the figure, which falsely made averages where there was non-existent data. This has been corrected and the new figure 5 has considerably more missing data (as there should be, given the number of data missing as shown in figure 4).

2 Minor Comments

Table 1c states that SWOOSH uses a number of instrument data sets, including SAGE III v7.0. I am aware that there is a new version of SAGE II data, but was a new version of SAGE III data also released and used for that work?

Many thanks for picking up this error. Indeed, there is no SAGE-III v7.0. This has been changed to read SAGE-III v4.0. See table 1c (revised manuscript).

Table 1e: Just to clarify, the data screening mentioned refers to the method of screening used to create the merged data sets, correct? It does not refer to any kind of screening the authors used for this work in relation to the merged data sets provided to them.

No additional screening was applied to any of the data sets. This has been clarified in the revised manuscript.

Furthermore no additional screening was applied to any of the seven data sets.

Table 1c states that SWOOSH uses Aura MLS v2.2 but Table 1e states that data filtering uses guidelines from v3.3. The paragraph about SWOOSH on page 25699 also states that Aura MLS v3.3 was used.

Thank you for spotting this error. In all cases it should read Aura MLS v3.3. This has been corrected in Table 1c (revised manuscript).

Equation 2 shows the autoregressive model used to account for autocorrelation in the regression. I believe there is a typo here where ϵ_2 is not supposed to be multiplied by time but rather is a function of time (like ϵ_1). I just wanted to clarify the authors' intent.

Many thanks again for picking up on this error. The ϵ_2 should indeed be a function of time rather than multiplied by time. This has been corrected to read as follows:

$$R(t) = \epsilon_1(t) \times R(t - 1) + \epsilon_2(t) \times R(t - 2) + e_t$$

Corrected equation 2.

Pg. 25704, Line 1: "Averages and standard deviations were only calculated for months that had data for more than 20 of the 28 years available for analysis." Figure 3 states that data was only used if there was over half (14) of the 28 years. I was curious which criterion was used.

This is an error. In a previous version of the paper we used a stricter criterion (20 years) than that which was applied in the final version (14 years). The weaker criterion was chosen to ensure that a more complete annual cycle was available for some of the data sets in the lower tropical stratosphere (50hPa, Figure 3(b)). The text has been corrected to reflect that at least 14 of the 28 years were required (as is stated in the caption of Figure 3).

Averages, standard error, and standard deviations were only calculated for months that had data for more than 2014 of the 28 years available for analysis.