The authors would like to thank the referee for taking the time to review this paper and for the many helpful comments that will be used to improve it. The referee's comments/concerns are listed below in red text, while the authors' responses to each comment are written below in black text.

Section 2 discusses used proxy variables and their orthogonalization at length. The resulting proxies are somewhat different from what is often used in regressions. I think it is absolutely necessary to explicitly show the time series for these resulting proxies. A corresponding Figure should be added.

A figure that shows the 4 QBO EOFs, 2 time-shifted ENSO orthogonal functions, 2 time-shift solar orthogonal functions, and 2 EESC EOFs will be added to the revised paper.

In Section 4, the regression methodology does not really become clear to me. Nowhere the paper clearly shows what is regressed against what. This really major point has somehow gotten lost. Instead, the discussion focuses very much (too much?) on residuals and statistical details.

It is the authors' opinion that the details regarding proper statistical analysis and investigation of the residuals are both vital to the success of the technique and a useful tool for gleaning additional important information regarding the quality of the technique itself and the data to which is it being applied. The authors believe that this is often skimmed over or ignored in many other analyses.

Eq. (5), in my opinion, is wrong. In the current form, the regressed temporal series T(t) would be the same everywhere, and a latitude dependence $\Theta(\theta)$ would "distribute" the regressed time series to different latitudes. That is clearly not what the authors did. Rather, the authors probably used:

$$\eta(\theta, t) = \sum_i \beta_i(\theta) X_i(t)$$

$$\beta_i(\theta) = \sum_j b_{i,j} L_j(\theta)$$

where $L(\theta)$ are Legendre polynomials, $X(t)$ are the time-series predictors, and b are the coefficients determined from the entire dataset(?) by the fitting procedure? This should be clarified/ corrected. The authors should clearly explain what is actually fitted.

Good point. This will be corrected in the revised paper.

In Sections 4 and 5.1, I am missing a few simple statements explaining what the total, correlated and uncorrelated residuals really are. I am assuming that they are from Equation B7 (pg. 17689, line 25). If I understood it right: The total residuals (top panels of Fig. 2) give the total residuals Ri in Eq. B7, and the uncorrelated residuals (bottom panels of Fig. 2) give the _i_i in Eq. B7. The total residuals are obtained in the 1st regression fit, the uncorrelated residuals are obtained only after iterative corrections / Cochrane-Orcutt transformations. Did I get that right? It would certainly help to add some clarification to Sections 4 and 5.1.

That is the correct description of the difference between the residual types, as written on page 17704, lines 15-17. All three residual types are updated after each iteration. Each iteration contains an autocorrelation correction. While the total residuals do not change after the autocorrelation correction, they do change with heteroscedasticity and residual filtering iterations.

So as not to overburden the reader (and the body of the paper) with excessive information about math, the assumption is made that the reader has a full understanding of generalized least squares (GLS) regressive techniques and already recognizes the terminology used (e.g., total/uncorrelated residuals, autocorrelation, and heteroscedasticity). If not, they are encouraged to read Appendix B as stated on page 17689 line 5.

The purpose of section 4 is not necessarily to explain GLS in detail, but rather to detail how some of the specifics (e.g., autocorrelation, heteroscedasticity, residual filtering) that can be unique to each regression are determined for this particular application.

Section 7: I feel that the authors tend to over-emphasize differences between monthly zonal mean (MZM) and simultaneous spatial and temporal (STS) regressions. To me, the different panels in Figs. 12 and 13 look very similar. I wonder if the minor differences are really significant. I realize that the MZM must lead to granularity in the latitude direction, different from STS which should be smooth in latitude direction. But why are the STS results not granular in altitude direction, like the MZM results?

This should be added in the revised paper.

Why are MZM results not stipled as insignificant, even when the trends are close to zero?

The MZM/STS piece-wise trend results are stippled where the results are statistically insignificant. It is possible for trends be near zero and still be statistically significant.

The MZM/STS EESC results are not stippled at all. The reason is because a statistical measure of the significance of the linear trend temporal coefficient can be computed for each part of the piece-wise linear trend, even when multiplied by multiple spatial terms. However, no similar mathematical calculation exists for two separate (but added) EESC temporal terms. A different measure of significance could be computed, but it would not be statistically robust and would not be comparable to the stippling shown for the linear trends.

I think the authors need to check that the MZM and STS results are really plotted in the same way. They should also be careful and not over-emphasize the differences. The authors may disagree, but my feeling from their results is that the old MZM method is actually doing quite a good job, and produces overall results that are comparable to results from the STS method. Of course the STS method is more advanced, does a better job in a few respects, and, particularly, gives better confidence that some possible sources of error are avoided, and results are more reliable.

The differences between the resulting trends of the two methods during the period of decline of ozone is small. However, the recovery period in the MZM method in the upper stratosphere at mid-southern latitudes is almost a factor of two larger than in the STS method. Granted, more data is needed to reduce the uncertainty of these results but the results are comparable to other studies and the reason for the bias is quite clear (as stated in the paper).

The MZM does a reasonable job, but it does have its limitations with regard to the non-uniform sampling and the resulting biases are shown in Figures 11, 12, and 13 and described in detail in the paper.

Abstract: The abstract should give numbers for the results: What are the SAGE uncertainties? How big are the sunrise-sunset differences? How big are the QBO, ENSO, solar cycle effects? How large are the pre-1997, post-1997 trends? How do these results compare to previous studies. This also applies to Section 8, which needs to put results into better perspective with respect to previous studies, e.g. summarized in the series of WMO-UNEP ozone assessments.

Perhaps it is a stylistic preference, but because the aim of this paper is to analyze all resulting trends, no single number can be stated to summarize the results. As such, the authors feel that only the resulting figures can summarily describe the results and that no numbers need be included in the abstract.

Again, since the aim of this paper is to analyze long term trends, detailed comparisons of the QBO, ENSO, and solar terms to other studies are beyond the scope of this paper. However, because the analysis of any single term necessitates the scrutiny of every term, a cursory analysis of each is discussed in the paper.

pg. 17682, line 21: An introduction should provide wider context from existing publications. Damadeo et al., 2013 is not a wide-context reference for SAGE II that has been providing good data and many papers since 1984. Some key papers from previous decades should be cited here. This criticism applies throughout much of the paper, where only papers from 2009 to 2013 are cited and the extensive body of work done with SAGE data in the 1980s, 1990s and 2000s is largely unreferenced / ignored. The authors should please make the effort and provide more scientific context, especially in Sections 1, 5.2.x, 7, and 8!!

This paper is not intended to be a review paper; rather it is an expansion of the current "state of the art." Many of the studies referenced are iterative works, repeated and refined over the years. As such, only the most recent of them are cited.

However, the authors will look into this to possibly include in the revised paper.

pg. 17683, line 6: What is meant by swath? Define / be more explicit.

This is simply referring to the ground-track pattern of events. This will be better described in the revised paper.

pg. 17683, line 8 (and many other places in the text): "based off of" ! "based on" ??

"based off of" is acceptable and occasionally common in informal, spoken American english whereas "based on" is more formal both in spoken UK english as well as academic writing. The typist apologizes for his colloquial slip up.

pg. 17685: It remains unclear what QBO proxies are used. Two orthogonal equatorial EOFs only? Sidebands with annual modulation? This could also be achieved by allowing for annually varying amplitude of the QBO fit. As suggested above, it would help to show and discuss the final proxy time series.

Page 17683, line 24 through page 17684, line 2 describe the creation of the QBO proxies and which are used (the leading 4 EOFs derived from equatorial zonal wind data at 7 pressure levels).

Page 17685, lines 19-21 state that the QBO can be modulated by the annual cycle (2 terms) and that these cross-terms are included. As such, there are 12 QBO related terms used in the temporal component of the fit (prior to the inclusion of spatial cross-terms).

While the 4 QBO EOFs will be plotted as previously stated, the inclusion of all of the cross-terms would be pointless and no information could be obtained from simply plotting them. However, the reason for the inclusion of the cross-terms (modulating the frequency at non-equatorial latitudes) can be seen in Fig 5, though it may take more than a cursory glance to see the non-dominant frequencies.

pg. 17686, lines 7-11: I think a plot showing a typical daily sampling pattern would help a lot here (or a reference to an existing plot).

A more detailed description of SAGE II sampling will be included in the revised paper. A figure may also be included if necessary.

pg. 17694, lines 6 to 10: This ambiguity between solar-cycle and volcanic aerosol signals (both peaking near 1983 and 1992) is a very old problem. There is an old Susan Solomon paper from 1996(?) that should be cited here!!

The authors will look into this and likely include it in the revised paper.

pg. 17695/6, Section 6: Here (and in a few other places, e.g. pg. 17698, lines 1-2), the comparison between MZM and STS is not always fair: MZM does not differentiate between sunrise and sunset events, STS does. However, MZM could also split between sunrise and sunset events. Then only the spatial/temporal biases between STS and MZM would show up! Sunrise / sunset differences would not alias into the MZM vs. STS comparison. I think this should be stated clearly here, and in several other places (e.g. pg. 17698, lines 1-2).

During this research, the authors considered looking at three methods (the other being an MZM that accounts for the different event types) but eventually decided against it. The reason being that, to the authors' knowledge, all other studies involving regression to SAGE data that are done in the MZM fashion (only one other study performs an STS-like method and it does not account for event type) do not account for the difference between event types. (All other MZM studies except Kyrola, 2013, a fact that will be mentioned in the revised paper.) Since the authors wanted to compare this new method with what is traditionally used, only two separate methodologies were employed.

That having been said, the differences between the MZM and STS methods are attributable to both the non-uniform diurnal sampling as well as the non-uniform spatial and temporal sampling. A look at Fig 11 reveals that, even where the diurnal variation is practially zero, differences between the two methods exist on the order of a few percent (larger than the measurement uncertainty) or larger (at high latitudes). Additionally, while the sampling is biased towards particular latitudes at particular times of year, and this bias is constant at the beginning of the mission (Fig 9), the bias drifts at the end of the mission. This drift in sampling patterns will alias into long-term trends in the MZM method and so a simple MZM method that also accounts for the different event types would still be insufficient.

pg. 17697, lines 14 to 25: Kyrölä, 2013 and many other studies consider data up to 2013. Here, however, only SAGE data up to 2005 are considered. This should be stated very clearly!!

This will be added to the revised paper.

It should be noted that these other studies make combined use of SAGE and other data. However, the other data generally starts after the early to mid 2000s. As such, the "anchor point" at the beginning of their recovery trends (as well as about half of the total time span) come squarely from SAGE data, and this is where the problem arises. The additional data at the end of the regression period is insufficient to compensate without additional data throughout the earlier part when sampling biases are not accounted for.

pg. 17698, lines 3 to 23: This "cherry picking" should be avoided: The two "orthogonal" EESC terms (I'd like to see them!) happen to pick a plausible altitude dependence of the turnaraound year in the Northern Hemisphere, but not in the Southern Hemisphere. Why? So are the orthogonal EECS terms good or not? The piecewise trend turnaraound date could also be changed /fitted. What would happen then? Statements based on spurious or unclear evidence should be avoided!

The authors are unsure of the "cherry picking" to which the referee is referring. If the reference is regarding the choice of latitudes for Fig 14, the reason for the choice of those latitudes to plot is simply because those latitudes (50 N/S) show the largest change in ozone over the mission lifetime. The equator is shown simply as another reference.

The reason for the hemispherical asymmetry is unknown. There is nothing wrong with the EESC terms, rather, the lack of significant turnaround in the southern hemisphere at some altitudes and latitudes is present in the data. Again, the reason is unknown and beyond the scope of this paper.

The piece-wise trend turn-around date could be changed. However, it would clearly be a function of latitude and altitude and the inclusion of a specific term to account for the turn-around time would make the regression non-linear (e.g., a*[t-b], where a and b both need to be solved for). Given the potential number of regressor terms, this would be problematic for convergence. Instead, the variable turn-around time would have to be estimated by guess and check. The purpose of the orthogonal EESC functions isn't necessarily to say that an EESC shape is strictly better than a piece-wise linear trend (though perhaps others may argue for that point), but rather because it allows for a linear regression to terms that account for the variable turn-around time that is present in the data.