The authors would like to again thank the referee for taking the time to review this paper. The referee's remaining original comments/concerns are listed below in red text, while the authors' responses to each comment are written below in black text. Further comments from the reviewer on the revised version of the paper are given in blue. Further responses from the authors are again written in black.

The revised version of the manuscript has fixed a few, but not all, of the concerns raised in my 1st review. I think the authors should still try to address these points, as discussed below. In many cases I feel other readers will have similar questions. The purpose of the revisions should be to make the paper clearer to the average reader, not to minimize the amount of work for the authors. When revising the authors should keep the following key questions in mind:

1.) Does the abstract clearly and concisely summarize the paper and state the main results? Can the abstract and main body of the paper each stand alone?
2.) Are the interpretations and conclusions adequately supported by the evidence presented? That is, are the assumptions valid, is the methodology sound, is the evidence adequate, and do the conclusions logically follow?
3.) Does this paper put the progress it reports in the context of existing published work? Is there adequate referencing and introductory discussion?
4.) Is the paper clearly and concisely written?

These questions are key review criteria (from the JGR website). They should also be fulfilled for ACP papers.

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.

It seems that the authors have chosen not to include these time-series, although they are shown in the online discussion. I am not sure why, but I have to admit that the time-series look complex, possibly reducing a readers confidence in the authors results. I still think that these proxy time-series need to be shown in the final paper. They are the basis for the regression results!!

The non-QBO EOFs are quite simple. The QBO EOFs are not overly complicated. The leading two terms look like what one would expect if performing EOF analysis of zonal wind data at different heights. The other two terms are simply the next leading two EOFs that describe the variation of the zonal wind. They are purely mathematical in nature and there is nothing "complex" about them. While the authors continue to believe that the figure provided in the interactive discussion does not contribute significantly to the content of the paper, we will include it.

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.

I still don't really understand how the regression is done, and what is iterated. I don't think the authors give a good description of what they are actually doing. To me, key criterion 4 (and to some degree 2) not fulfilled. But maybe the editor is happy with the description, and other readers can understand what is being done.

The reason the process of iteration is not detailed is because there is nothing in Appendix B that is novel or unique. The math and methodology are commonly found in any of a number of texts regarding

this subject. The reason the appendix is included is because the authors found details on how autocorrelation was corrected for lacking in many other papers as well as a total absence of the discussion of heteroscedasticity. However, to make things a little more clear, the process of applying the regression outlined in Appendix B is added as a final paragraph to that appendix. Section 4 remains reserved for the specifics of how certain aspects of the corrections are computed instead of the information commonly found in texts.

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.

Ok, looks like I was able to figure out what is shown in Figs. 3 and 4 (Figs. 2 and 3 of the old version). To help other readers as well, it would be good to add these references to Eq. B7 of the appendix to the description of Fig. 3 on page 11. Would make things easier for all readers. (Key criterion 4)
(Added)

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.

OK, I understand that the authors want to show the details. However, the process described in Section 4 is still not clear to me. One thing that would help greatly is the addition of a short description of the main steps. This should come in the 2nd paragraph on page 9 (the paragraph starting with "A generalized least squares"). It would be something like.

1.) The 1st step of the regression process is estimation of an initial set of bi,j and their uncertainties
2.) In order to get a better estimate of the uncertainty for the bi,j it is necessary to estimate and account for autocorrelation in the residuals remaining after regression. This requires a transformation of observed values and predictor values (e.g. Cochrane Orcutt method). Then step 1 is repeated until ...
3.) In addition, it is necessary to account for heteroscedasticity (i.e. a scatter of the residuals that varies over time, e.g. becomes larger as SAGE II sampling becomes sparser over the years). This is also done iteratively – following step 2 and then going back to step 1 (?) OR – after 1) + 2.) have been iterated and then iterating 1.) +

I hope the authors can see what is not clear to me (and probably many other readers). I think such a short outline in Section 4 would help a lot. The Appendix then gives the Equations and details, and then some more details are given in Section 4.

Please see the authors' new response to the referee's previous original comment.

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.

I don't see any changes in the revised paper, and I am still wondering about that.

Vertical granularity was added to the STS plots in Figs. 13 and 14 as requested. However, given the smaller latitudinal resolution of the STS plots, it may not be obvious at a cursory glance. Granularity was not added to the plots in Fig. 15 because contour lines do not plot well under those conditions.

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.

The authors probably refer to the 3rd paragraph of Section 7 (starting with "The results shown in Fig. 14). This discussion could still be made clearer. I think it should also refer to Fig. 10 and 11, NOT Fig. 11 and 12. The point is clearly that after 2001, SAGE II samples more sunsets (Fig. 10) – and thus higher ozone in the upper stratosphere (Fig. 11). So in the end, accounting for the different sampling in the STS method explains the smaller increase since 1997 in the SH. If this would be accounted for in the MZM method, the trend would also be smaller. To me, the presented evidence only suggests that the STS method does better accounting for SR / SS differences, but there is very little evidence that the STS regression itself is much better than the MZM method! The presented evidence does not support the authors' conclusions. (Key criterion 2 not fulfilled)

There is a reason why the third paragraph of section 7 refers to Figs. 11 and 12 (and not 10). Both Figs. 10 and 12 show a reduction in sampling in the later period, but Fig. 12 shows the latitudinal distribution of that disruption more clearly as a function of time than Fig. 10, which ties much more closely to the results of Fig. 14. The reason for the inclusion of Fig. 10 is less to show the reduction in sampling and more to show how precession has increased (due to orbit degradation) and how that specifically can affect the trend in the MZM method regardless of SR/SS differences.

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.

Key review criterion (#1) for a JGR paper is:

- Does the abstract clearly and concisely summarize the paper and state the main results? Can the abstract and main body of the paper each stand alone?

There are good reasons for this criterion, especially the possibility to just read the abstract and still get the main message of a paper. I think this criterion also applies to ACP papers (unless the editor wants to waive it). I think the current (unrevised) abstract does not fulfill this important criterion! I still urge the authors to rewrite the abstract and put the main messages and the main numerical values and uncertainties into the abstract.

Since the resulting trends are shown in many colored contour plots, the trends alone represent nearly 6,000 different trend values depending upon the latitude, altitude, method, and time-period of interest. This cannot be summarily written in the abstract. Only an investigation of the resulting figures can illustrate the results. Which region is of interest is up to each individual reader and so the authors cannot endeavor to choose which few numbers should be represented in the abstract. At most, the claim that "trends in the presumed recovery period" are different will be modified to reflect that the largest differences occur at mid-latitudes in the middle to upper stratosphere and state the maximum differences. A statement about how each monthly value can vary by upwards of 10% will also be included.

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.

Nothing has been done. I still think that a few older & comprehensive references are still needed. SAGE II results did not start with Damadeo et al. 2013. Key criterion 3 not fulfilled.

Various references have been used over the years to refer to the quality of the SAGE II ozone data for various older versions of the data. Now that a comprehensive paper both on the methodology of the SAGE II algorithm and the quality of the most recent version (v7.00) has been written (Damadeo et al., 2013), the SAGE II team would like it to be used as the standard reference. The SAGE II team also strongly advocates against the use of any older version of SAGE II data.

In a previous review, the reviewer suggested references to include at particular locations in the paper and why they should be included. If the reviewer could suggest some specific references to include in the same way they have before, the authors would be happy to include them. Otherwise, the authors do not currently see where or why any other references need be included.

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.

I think addition of the Figure with the time-series proxies (the one from the supplement) will help a lot with this problem. As stated above, I think this Figure should be added, and I don't see why the authors do not want to add it.

Please see the authors' new response to the reviewer's first response (first page).

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 practically 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.

OK. But then the authors should be fair and should differentiate more clearly between shortcomings of their MZM method due to not accounting for changes in diurnal sampling, and other shortcomings due to spatial sampling changes and simpler regression. Some numbers about the magnitude of trend-errors or solar-cycle errors due these effects would be good. From Figs. 13 and 14, I would say that the effect on trends is marginal nearly everywhere, except in the SH above 35 km after 1998. Key criterion 2 – questionable.

The purpose of this work was to show the difference between how time-series analysis is currently done (MZM) and a new method to properly account for non-uniform sampling (STS). It is true that the trends in Fig. 13 are not very different, but this is stated clearly in the paper. The differences in trends in Fig. 14 are well described and are the result of both the non-uniform diurnal sampling as well as the non-uniform temporal and spatial sampling. The authors do not see the need to account for only one cause (diurnal) and not another (temporal/spatial) when both can be easily accounted for. As such, the authors do not explore the intermediate step of performing the MZM method with a term to account for the SR/SS differences because it is still insufficient.

Furthermore, even if a separate scenario (MZM with SR/SS differences included) were computed, the results would not be useful. They would be specific only to SAGE II and not any other instrument or combination of instruments and would vary with latitude based on sampling and so could not be used to modify preexisting MZM-like calculations. If one wished, one could attempt to infer the direct effect of non-SR/SS differences on trends by comparing the top plots to the bottom plots in Figs. 13 and 14 in the regions of little to no SR/SS difference influence (Fig. 11) where it would be quite obvious that differences exist and (in some areas) seem comparable to some of the differences seen where SR/SS differences are prevalent.

Lastly, while it is beyond the scope of this paper, one of the biggest problems others have run into whilst attempting to derive trends from merged datasets is the seasonal cycle. Attempts to use multiple datasets via the MZM method have found that different instruments have different seasonal cycles and so, typically, each instrument has its own seasonal cycle removed before merging (part of the homogenization process). Since there is only one atmosphere, this has become necessary to attempt to account for the different sampling biases of different instruments. If the seasonal sampling were not at the center of each month and each bin but were consistent throughout the mission lifetime, the resulting seasonal cycle would be biased but it would not alias into derived trends. However, since the seasonal sampling drifts throughout the mission and the seasonal cycle is derived from means over the entire mission, both the seasonal cycle and the derived trends are biased. This source of bias would be apparent in almost any dataset, but more pronounced in sparsely sampled data (e.g., SAGE, HALOE, ACE, etc.). The authors believe the STS method can be used to account for and remove this source of bias.

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.

The point I was trying to make is that the comparison of piecewise trend (with no change of turnaround year with altitude because 1997 is used as fixed turnaround year) and two EESC terms (which allow a different turnaround year) is not fair. And it is even not clear that the varying turnaround year for the two EESC trends is correct: Its altitude dependence appears correct in the NH, but not in the SH, with reasons unknown. Again the authors make or imply a conclusion that is not fully supported by the presented evidence. Key criterion 2 not fulfilled, in my opinion.

Again, the purpose of this work was to explore the difference between what is currently done and one possible method to properly account for non-uniform sampling. Current time-series analyses (like most of those already referenced) use a piecewise linear trend with a single turnaround time applied everywhere. The fact that it is "not fair" to compare this with a variable turnaround method is because using a single turnaround time is likely inadequate and thus should not be done, as illustrated by the results in Fig. 15. The cause of the hemispheric asymmetry seen in the turnaround time is indeed unknown. The authors make no claim as to its origins and that is clearly stated in the paper. It is possible it derives from the lack of a good fit due to insufficient data to constrain the fit and this is also stated in the paper. To the authors' knowledge, no clearly defined claims are made in the paper that are not substantiated and the questions that arise from this work are also clearly stated and will be a matter of further study.