The authors would first like to thank the referee for their time in reviewing this paper. There were many helpful comments that will be incorporated to make this a better paper. The referee's comments/concerns are listed below in red text, while the authors' responses to each comment are written below in black text. Some comments refer to similar topics and are thus grouped together.

Page 2, Line 7: account for any potential diurnal variation on what? Ozone presumably? or ozone trend?

Ozone. This will be added in the revised paper.

Page 2, Line 9: What is 'the recovery period'? First this presupposes that ozone is actually recovering from the effects of ODSs - something that your paper would need to prove first - which is also something that cannot be done only with regression analysis. I think that it would be clearer for the reader if you just stated the period here.

The term "ozone recovery" is used quite often in the current literature (e.g., Kyrola et al., Bourassa et al., and the most recent SPARC reports). In the case of Fig. 13, it is defined as after 1998. In addition, any paper that utilizes an EESC term for regression analysis is assuming both a decline period and a recovery period in their data set. Moreover, nearly every regression analysis shown reveals a period of negative trends followed by a period of positive trends whether piecewise linear trends or EESC terms were used. As such, the authors have no reservations about using the term "recovery period" (or perhaps "presumed recovery period"). That having been said, the authors will revisit the repeated use of the word "recovery" to see, for each case, if different terminology would be better suited.

Page 2, Line 9: Will readers unfamiliar with the field know what is meant by 'variable turnaround time'?

This will be added to the revised paper.

Page 2, Line 10: A hemispheric asymmetry in what? Ozone trends?

Correct. This will be added in the revised paper.

Page 2, Line 13: I find this difficult to believe. If you have a data set that runs e.g. from 1984 to 2000 and a second data set that is biased 20% low that runs from say 1998 to 2014, and you apply your 'global regression' to an uncorrected merge of those two data sets, I can't believe that the trends would not be biased negative? I think that you need to say more here about how this would work.

This would be true if one were to ignore any instrument dependencies. As mentioned in the latter half of the conclusion (perhaps "Conclusions and future work" would be more suitable), this could be achieved with instrument dependent conditional terms.

Page 2, Line 20: Is 1% the typical 1 sigma random uncertainty on each ozone measurement? If so, perhaps you should say so.

In most of the stratosphere, yes. This will be added in the revised paper.

Page 2, Line 21: Do you mean accurate measurements or precise measurements? Precision matters more for trend analysis than accuracy.

Precise. This will be added in the revised paper.

Page 3, line 15: It wasn't completely clear to me what you did here. I didn't understand what you meant by 'from multiple component data sets'. For example, consider the solar cycle basis function as 10.7 cm solar flux. How is this a 'multiple component data set'? Couldn't this be made much simpler by using Gram-Schmidt to orthogonalize the basis functions?

Page 4, line 1: Account for over 99% of the variance in what? It certainly can't be in ozone.

Page 4, line 17: Account for over 99% of the total variance in what?

As mentioned shortly thereafter, data sets like the QBO and EESC have multiple components (different pressure levels and mean ages of air respectively). Some data sets, like ENSO or the 10.7 cm flux, have only a single component. For data sets with multiple components, an EOF analysis allows for inclusion of terms that describe the greatest amount of variance in the data from which the EOFs derive. This will be more clearly stated in the revised paper.

Page 3, line 19: As done in Section 3 of Austin, J., K. Tourpali, E. Rozanov, H. Akiyoshi, S. Bekki, G. E. Bodeker, C. Brühl, N. Butchart, M. Chipperfield, M. Deushi, V. I. Fomichev, M. A. Giorgetta, L. Gray, K. Kodera, F. Lott, E. Manzini, D. Marsh, K. Matthes, T. Nagashima, K. Shibata, R. S. Stolarski, H. Struthers, and W. Tian (2008), Coupled chemistry climate model simulations of the solar cycle in ozone and temperature, J. Geophys. Res., 113, D11306, doi:11310.11029/12007JD009391 which you should therefore cite.

The authors were unaware of this particular instance of the use of this method. It is not novel or unique. It is a simple mathematical tool and thus a reference does not seem necessary.

Page 3, line 19: But this will not create a basis function that is orthogonal to all other basis functions. It will only create a basis function that is orthogonal to the one it is being shifted with respect to. While your approach does allow for phase shifts in any periodic basis function, it does not ensure that all basis functions being used in the regression are orthogonal. For that you need Gram-Schmidt orthogonalization.

The goal was not necessarily to create a set of predictor variables that were all orthogonal to each other, but rather to reduce the amount of multicollinearity between predictors of the same type (e.g., 4 QBO terms instead of 7, with each being orthogonal to the other). This will be more clearly stated in the revised paper.

The authors did explore the possibility of taking all of the predictor variables and creating a full set of orthogonal functions (e.g., via the Gram-Schmidt orthogonalization process). However, this does not inherently remove or reduce multicollinearity. For example, regressing a data set to 10 predictor variables will produce the same result as regressing to 10 orthogonal functions (OFs) derived from the 10 predictor variables (and then transforming back from orthogonal function space to predictor variable space). The only way to reduce multicollinearity, is to convert 10 predictors to 10 OFs and then remove some number of them; for example, say 7 of them explain 99% of the variance in the predictors. Regression can then be performed to the 7 OFs and the results transformed back to the 10 predictors. The pitfall of this technique is that the resulting coefficients can potentially be biased.

Ultimately, the authors decided to stick with the use of different sets of EOFs that may not be orthogonal to each other set. Exclusion of predictors is then done with iterative analysis of the uncertainty in the varying coefficients.

Page 4, line 4: This is novel. As far as I am aware, all previous regression studies have assumed that there is no delay in solar cycle effect on ozone and therefore did not consider any phase shift in the solar cycle basis function. Do you have any physical basis for expecting there to be a phase shift between change in solar output and ozone response?

Page 14, line 16: I would guess that it is the latter but that would be a guess. A key question to be addressed is: do you have any physically-based mechanism in mind that would cause ozone to respond two years later to a change in solar output? If no such mechanism can be envisaged, then it is probably safer to have only one solar cycle basis function in a regression.

Remsberg et al. (2010 and 2014 are similar analyses) make use of orthogonal pairs of solar cycle terms. In those particular papers, the solar cycle is represented by a pair of trigonometric functions with an 11 year period rather than the 10.7 cm flux, but the concept is the same. The authors chose to explore both possibilities (one or two solar terms) and compare. Ultimately, so long as a volcanic term was included, the total residuals and resulting trends were not different between the use of one or two solar terms. From this analysis, the authors cannot determine which is “right” or “wrong” and thus do not make any definitive conclusions about the use of 1 or 2 solar terms.

Page 4, lines 10-12: This is worded quite confusingly. Isn't it simply the case that you have one linear trend basis function that covers the whole period, and then a second linear trend basis function that is everywhere zero before 1997 which quantifies the change in trend after 1997?

It is as stated in the paper. One term is linear before 1997 and zero after while the other is the opposite. For the purpose of regression, this is mathematically equivalent to having two terms: one linear throughout the whole period and the other only linear after (or before) 1997 while being zero elsewhere.

Page 4, line 19: I think that you could avoid this 'pathological' behavior by orthogonalizing all of your basis functions. Assuming that the trend basis function comes before the EESC basis function in the Gram-Schmidt orthogonalization, this would effectively remove any trend from the EESC basis function.

If one were to create a full set of orthogonal functions (OFs) from all of the predictors, the resulting OFs would have no physical meaning separately. In order to derive any meaning from them, they would need to be transformed back into predictor variable space. If one were to take the 2 EESC functions and a linear term and create 3 OFs from that, one could look at the combination of the three OFs as the resulting long-term changes. However, looking at the three OFs together would be no different than looking at the 3 predictors together. As mentioned previously, multicollinearity is not reduced by orthogonalizing functions. Only excluding OFs can reduce multicollinearity, though it has the side effect of resulting in possible biases.

Page 4, lines 25-26: I disagree that 'A seasonal cycle in a predictor variable would interfere with the seasonal cycle in ozone'. In this case I suspect that you would find that e.g. expanding the volcanic aerosol regression model fit coefficient in Fourier pairs to account for seasonality would, if the mean of the signal is also subtracted, result in time series that are quite orthogonal to the mean annual cycle terms. You should at least do that test so that you can state this more categorically.

The concept of multiplying a term by a Fourier pair to account for a separately attributable seasonal cycle is a commonly used technique. Unfortunately, it can lead to spurious conclusions because it does not, in fact, compensate for multicollinearity. For example, if the seasonal cycle in ozone were a combination of temperature effects and response to seasonally varying aerosol, one would be inclined to use a seasonal term and a seasonal cross-term with aerosol. Unfortunately, what is then presented to the regression algorithm are two very similarly shaped terms. What ultimately ends up happening is that the algorithm cannot accurately differentiate between the two seasonal terms and assigns a spurious combination of weighting coefficients (perhaps 90/10, perhaps 50/50, perhaps 65/35, etc.). If the two terms to regress to are similarly shaped (i.e., in phase and of similar amplitudes), then it is impossible to use a regression approach to separate their influences.

Page 5, line 9: I think that you need to be a bit more specific in what you mean by 'cross-term'.

A cross-term is simply a product of terms. This will be added in the revised paper.

Page 5, line 15: No, the use of orthogonal basis functions does not allow for a change in amplitude with time. The net amplitude of two orthogonal functions is given by sqrt(sqr(A)+sqr(B)) where A and B are the fit coefficients for each of the basis functions. There is no time dependence in the amplitude. Really, including an orthogonal version of a periodic basis function has just one advantage i.e. it allows for the phase shift to be non-zero.

It is not stated that the amplitude can change with time. With the current implementation, the amplitude of various predictors can change with latitude (because of the two-dimensional regression) and altitude (because the regression is done separately for each altitude).

Page 5, line 20: Yes this is, in essence, a cross-term, but really this is nothing more than accounting for the QBO fit coefficient to be seasonally variable which can be accommodated by expending the QBO regression model fit coefficient in a Fourier expansion in season.

Yes, they are mathematically equivalent. However, the primary reason for the inclusion of only this cross-term was the apparent need for frequency modulation in the QBO response.

Page 6, line 11: I think that it would be good to have a Figure here to show an example of some data swaths.

A more detailed explanation of the sampling will be included in the revised paper. A figure may also be included.

Page 6, line 16: You need to explain what a 'non-dropped event' is.

It means dropped by processing in the SAGE II inversion algorithm. This will be included in the revised paper.

Page 7, equation 2: I have a big problem with the use of equation (2). This equation should only be used if you are measuring *the same quantity* N times. You are not measuring the same quantity N times. To make my point clear, consider the figure I have provided with this review (Example1.png). The upper and lower panels both show synthetic measurements of total column ozone on one day in one latitude zone. Both data sets have the same mean (300 DU) and both have the same sigma_Ybar (0.9081 DU) as calculated using your equations (1) and (2). And yet I would argue that I can have much higher confidence in my daily mean zonal mean for the upper panel since the noise level is lower - essentially I could fit a couple of Fourier pairs to remove the known structure i.e. the clear wave 1 pattern and then, using your equation (2), would derive a much smaller sigma_Ybar. Equation (2) is inappropriate for use in this case because it takes Ybar to be the best estimator for the true value at every longitude. For the upper panel Ybar is not the best estimator for the true value at every longitude. A fit of some sort (as a described above) would be better. This is an important issue that the authors need to deal with more comprehensively.

The referee makes a good point to reveal something the authors noticed shortly after publication to ACPD that requires a minor revision to part of section 5.1. The uncorrelated residuals reveal the extent of uncertainty in the data used by the regression (i.e., daily means) and thus represent the combination of instrument noise and geophysical variability within each daily mean that is not resolvable by the model applied in the regression. When the computation of heteroscedasticity is iterated, this measure of combined uncertainty is solved for. The attribution of the uncertainty to each of the instrument noise and unresolved geophysical variability cannot be separately determined. In this way, the regression technique cannot distinguish between the two scenarios outlined by the referee (i.e., upper and lower panels in the figure) and it doesn't have to since the regression model is not attempting to account for longitudinal variation.

The referee also points out a fact that the authors stated in the conclusion (again, perhaps "Conclusions and future work" would be a better section title), namely the fact that a transformation to a coordinate system with higher resolution would yield better results. In other words, applying a model to account for diurnal and longitudinal variation would allow for the reduction of the uncorrelated residuals. If the sampling used for regression is reduced down to the sampling of the instrument, then the uncorrelated residuals would represent the uncertainties in the measurements directly.

Page 8, line 20: This is very reminiscent of the Bodeker et al. (2013) approach.

Legendre polynomials have the properties of a zero derivative at the poles and are all mutually orthogonal. They are the logical choice to describe potential global variation of a phenomenon on a spherical body and are widely used in many sub-fields of physics. However, the similarities to the Bodeker et al. approach will be noted in the revised paper.

Page 9, line 17: For use in the correction of what? You haven't stated anything about a correction being required.

The reference is to the autocorrelation correction, which will be stated more clearly in the revised paper.

Page 9, lines 17-18: I didn't understand what you have done here. Specifically I didn't understand why any iteration is required to derive the autocorrelation coefficient. I think that you need to explain in more detail why autocorrelation matters and how you deal with autocorrelation in a data set that is multi-dimensional. It is not clear to me that it can be easily reduced to the equivalent of auto-correlation in a one dimensional data set. Or perhaps this is explained in detail in Appendix B?
Page 9, line 19: Again it is not clear to me why an iterative correction is required.
Page 11, line 8: What do you mean by 'the total residuals'? Do you mean the total of all of the residuals? A few lines later you say 'The total residuals are a combination of the correlated and uncorrelated residuals' but you don't say exactly how they are combined.

A similar question was posed by the first referee. The following was the authors' response to that inquiry:

"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 9 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. The application of GLS over OLS requires iteration, though practically very few iterations are required for convergence."

The main body of the paper describes the physical interpretation of the different residuals, while the appendix describes their mathematical computation. The SAGE II data can be described as two concurrent, serially correlated data sets: spacecraft sunrises and spacecraft sunsets. Each "separate data set" is made up of successive events that are correlated with each other so long as their spatio-temporal separation is sufficiently small. We make the assumption that these "two data sets" have the same correlation and thus compute the correlation coefficient using all of the data. This is why the correlation is made between equations B11 and B12. Each data set can be turned into "pairs of points" and combined into a single computation (B12) so long as large data gaps are ignored in the pair selection process.

Page 10, line 19: But isn't it possible that a seasonally dependent coefficient, such as that for the QBO, could be statistically significantly different from zero during one part of the year but not during another? What do you do in that case?

A coefficient is either statistically significant or it is not, there is not an in between. Groups of terms collectively, made up of statistically significant coefficients applied to predictors, can still create statistically insignificant fit values depending upon the time and/or latitude at which they are computed.

Page 12, line 20: But this would only be true of the aerosol basis function had non-zero values over the full period though right? If there was a period of 5 years when the aerosol basis function was identically zero, I believe that the regression model would assign all of the 'seasonality variance' to the annual cycle basis function.

Say for example, one had two functions that they wanted to regress to data. One function is sinusoidal with a period of 1 year and the other is the same except it has a span of 5 years where the function was zero. A regression of seasonal data to these two functions would attribute all of the variance to the first function. Now let's assume the second function has the same behavior as before, but now it also has a linear trend. Now assume the dependent data has both a seasonal component and a linear trend component. A new regression would yield a non-zero coefficient for the second function even though the second function has a pseudo-seasonal component in it, because the regression needs to account for the linear trend in the data. Since the second function now has a non-zero coefficient, the coefficient for the first function will be reduced, since the second function now accounts for at least some of the seasonal variation.

This example demonstrates the problem that the authors wanted to avoid. The volcanic proxy has no seasonality while a pure aerosol proxy does. Since ozone responds with volcanic activity, this attribution would detract from the purely seasonal term. The bright side is that this will be captured in the uncertainties in the coefficients, and would be easily recognizable by analyzing the residuals. Nonetheless, it can be avoided by using a purely volcanic term.

Page 13, line 19: Just to be clear I think that you should say 'at altitudes above where the proxy is available'.

This will be added in the revised paper.

Page 13, line 22: I think that you should say 'is larger around the time of the Pinatubo volcanic eruption' just to be clear.

This will be added to the revised paper.

Page 14, line 27: I don't understand what a 'regressive response' is.

A "regressive response" is an over-fitting in the realm of data gaps. This terminology was used to differentiate from physical responses or algorithmic responses (SAGE II inversion algorithm) and will be added to the revised paper.

Page 15, line 11: Given your Figure 8, I wonder if you would care to comment on the findings from previous studies that have reported on strong responses of ozone to Pinatubo over northern midlatitudes but only very weak responses over southern midlatitudes. See, for example, Shepherd et al. "Reconciliation of halogen induced ozone loss with the total-column ozone record" Nature Geoscience, DOI: 10.1038/NGEO2155.

It is true that Fig. 1 in Shepherd et al. reveals that the relative decline of total ozone is smaller in the Southern hemisphere than in the Northern hemisphere following the eruption of Mount Pinatubo. However, Fig. 3 in Shepherd et al. shows a larger absolute decrease in ozone attributable to Pinatubo in the Southern Hemisphere (~ -9 DU) than in the Northern hemisphere (~ -7 DU), and similar relative decreases in ozone attributable to Pinatubo relative to ODS-induced losses in each hemisphere. The following is from Shepherd et al. when discussing Fig. 1 (page 4, paragraph 2): "The lack of a decrease in Southern Hemisphere midlatitude ozone following the Mount Pinatubo volcanic eruption in 1991 is explained by the fact that chemical loss was masked by a dynamically driven increase, evident in the cODS simulation."

It is not trivial to tell simply by looking at Fig. 8 if the results of this study support or refute the results of Shepherd et al. regarding the amount of total ozone decline attributable to the Pinatubo eruption. Figure 8 of this study shows the peak response of vertically resolved ozone to Pinatubo as a percentage of the local mean, and Fig. 3 of Shepherd et al. shows the absolute response of total column ozone to the eruption (and ODSs). A completely separate analysis utilizing the results of this study would need to be done in order to compute a comparable metric.

Page 16, line 21: 'piecewise linear trend' means 'joined at some particular time' and so there is some redundancy in this sentence. I have concerns related to the EESC-derived trends plotted in Figure 13. Please conduct the following test: take your SAGE II data and replace all 1998 data with the data from 1997. Do the same for 1999, 2000, ... 2005 i.e. make it that ozone stays in 1998 to 2005 exactly as it was in 1997 by repeating 1997 for each year after 1997. The ozone trend from 1998 to 2005 is then, by design, zero. I contend that you will derive positive trends in ozone from your EESC-based analyses (but not from your piecewise trend analyses). I think that this is a very easy test to perform and I would like to encourage you to do this test. I believe that it will show that your EESC-based analysis of the ozone trends from 1998 to 2005 is not to be trusted.

The referee has a valid point, namely the use of a predictor term for regression against data that has no physical basis is a bad idea (and that it will not necessarily yield a coefficient of zero). While various other papers using regression use either a piecewise linear trend or EESC term to characterize the long-term trends in ozone, we opted to investigate both for this reason. However, the use of a long-term predictor that does not adequately describe the long-term trends should be revealed in an analysis of the residuals (i.e., a long-term trend in the residuals should be present).

Ultimately, the results of the piecewise linear trend terms reveal periods of decrease and increase in ozone (which the referee seems to be OK with) and the comparable metric of mean trends derived from the results of the EESC regression are very similar. There are some small differences between the two, but this may very well be due to the lack of a variable turnaround time in the piecewise linear trends. The variability of the turnaround time has been investigated before (e.g., in Kyrola et al.) and we believe that this needs to be included in the regression. Granted, the significances on the resulting recovery trends are small (as shown in Fig. 13) and this is why we suggest, in the conclusion, the need for the extension of this method towards the inclusion of additional data to further constrain the potential recovery of ozone.

Page 25, line 9: Always assumes a constant what?

… assumes a constant exists in the regression. This will be added to the revised paper.

Page 26, line 17: You either need to write proper acknowledgments or remove this section.

It was stated early in the appendix where these statistical methods come from, though perhaps it was not clear that it was almost entirely sourced from that text. This will be more clearly stated in the revised paper.

GRAMMAR AND TYPOGRAPHICAL ERRORS

Page 3, line 22: Replace 'variables are created' with 'variables is created'.
Page 3, line 27: Replace 'resulting seven' with 'resulting in seven'.
Page 4, line 13: Replace 'effective equivalent' with 'equivalent effective'.
Page 4, line 14: Replace 'amount of halogen compound loading on the stratosphere' with 'chlorine and bromine loading in the stratosphere'.
Page 5, line 6: Replace 'off of' with 'on'. And likewise elsewhere.
Page 6, line 14: Replace 'that data is' with 'those data are'.
Page 6, line 17: Replace 'are applied' with 'is applied'.
Page 6, line 20: Replace 'exclusion of of any' with 'exclusion of any'.
Page 7, line 3: Replace 'henceforth' with 'hereafter'. Likewise on line 8.
Page 7, line 4: Replace 'latitude bin' with 'latitude zone'.
Page 9, line 22: Replace 'this data' with 'these data'. The word 'data' is plural. It's singular form is datum.
Page 9, line 26: The word 'criteria' is plural so either 'filtering criteria' or 'a filtering criterion'.
Figure 3 caption: Replace 'insufficient data exists' with 'insufficient data exist'.
Page 12, line 8: Replace 'data itself' with 'data themselves'.
Page 12, line 17: Replace 'data was used' with 'data were used'.
Page 15, line 6: I think that the word 'extraneously' has been used incorrectly here. I would have gone with 'anomalously'.
Figure 11 caption: Replace 'data does not exist' with 'data do not exist'.
Page 16, line 14: Replace 'exists' with 'exist'.
Page 21, line 29: Replace 'An insufficient amount of data is' with 'Too few data are'.

These will be corrected in the revised paper.

Page 22, line 9: Replace 'regressed to, where' with 'regressed to where'.

This is actually grammatically correct as is.