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.

It is stated that the method is described in the previous paper by Damadeo et al., 2014. It is not really the case since this paper deals with multiple instrument. The authors should add an Appendix or Supplement with the method description. In particular: Damadeo et al., 2014 analyzed data from just SAGE II. How exactly was the analysis of six satellite instruments (or just three? SAGEII, HALOE and ACE- FTS) done?

An appendix has been added that summarizes the technique from Damadeo et al. (2014) and adds some additional detail regarding how multiple data sets are incorporated simultaneously.

Was any instrument-specific weighting applied?

No. All instruments are treated equally. However, the regression is weighted and the instrument uncertainties do factor into that weighting so if one instrument is inherently less precise than another that will have an impact on the weighting.

The authors stated in many places that they try to use orthogonal functions. But the functions should be orthogonal on the dataset of available observations. Damadeo et al., 2014 mentioned that seven Legendre polynomials of latitude were used for the fit. However Legendre polynomials are orthogonal on 90S-90N, not on the 60S-60N interval where almost all measurements were taken. How was this handled?

Legendre polynomials in spherical harmonics are the logical choice to fit slowly varying data in latitude (as is done in many areas in physics) even if the data does not extend to the poles. Any polar gaps in a data set simply result in some degree of overfitting in those gaps, but ultimately it does not matter since we are not looking at those regions. Additionally, the data sets used here extend beyond 60S and 60N; we just show the results in this region because that is the primary region of focus in the community related to ozone trend studies.

Also, it seems that seven polynomials are too many. The authors should provide some justification.

Actually, this analysis uses 9 spatial terms instead of 7 (now added in the paper text) and it is very likely that more terms should be used. With a span of 180° in latitude, 9 terms yields a spatial resolution of 20°. However, some effects such as volcanic responses and the spatial extent of the peak of the QBO, for example, can be smaller than this. Choosing too many terms will create excessive overfitting through data gaps so 9 terms was chosen as a "middle ground" though no sensitivity study on this parameter was performed.

The authors compare their STS regression with the MZM method. I suggest the author reduce the part related to MZM and focus only on their STS results.

It is important to make the comparisons between the STS and the MZM as the MZM is currently the "de facto" methodology used by the community. Additionally, since we understand that it is very unlikely that the STS method will be whole-heartedly adopted, we decided that using the STS method to create "corrected" versions for use with the MZM method would be a reasonable compromise. As such, it is necessary to detail these "corrected" versions and how they compare as well as showing the relative impact of different sampling biases.

The MZM method is used in the paper in a very peculiar way: the authors just average all data within 10-deg latitudinal belts and assigned the value to the middle of the belt at the middle of the month.

We do not understand the reviewer's comment about the MZM being implemented in a "peculiar way" as this concept of creating monthly zonal means (i.e., averaging all of the data in a single month and latitude bin) is the default methodology that has been applied for almost all regression analyses of stratospheric ozone.

Most of the ozone variability is coming from the annual cycle. The annual cycle can be estimated, for example, by the same approach as discussed in the paper: by fitting all SAGE II data by a set of spherical (for latitude and, if necessary, longitude) and sin/cos functions (for time). Then the MZM method could be applied to the deviations from the annual cycle. The annual cycle is indeed orthogonal to the other proxies, so it should not affect their estimates.

The following description of fitting the seasonal cycle sounds like the common practice of deseasonalizing the data first. One could deseasonalize first, but doing so does not include information of any collinearity in the covariance matrix during the regression process and thus this information is not represented in the resulting coefficient uncertainties.

This step would largely remove most of the sampling problems and will likely produce results similar to STS.

Deseasonalizing cannot remove the sampling problems. If the sampling problems are constant from year to year, then not deseasonalizing only results in a biased seasonal cycle but does not impact the other terms (e.g., trends). However, in the case that the sampling patterns change from year to year (as has clearly been demonstrated), neither deseasonalizing nor fitting the seasonal cycle removes the problem. This is because the sampling biases create biases in the MZM values themselves and so any correction must be implemented on an event by event basis (e.g., the "DCorr" and "DSCorr" MZM data sets). Either that, or the data needs to be handled at a resolution much closer to the native resolution (e.g., the STS method).

P.4, L. 12. What data were used for this conversion? See box 2-1 from Ozone Assessment 2014 and comment on potential conversion errors.

As stated in the paper, the data used for the conversions were the pressures, temperatures, and altitudes given in the respective data sets. Box 2-1 from the 2014 Ozone Assessment cites McLinden and Fioletov (2011) and discusses both the expected differences in trends between number density and mixing ratio but also the potential uncertainties introduced during the conversion process.

First it is important to note that there is a distinction between these two phenomena. As cited, there is an expected difference in trend values depending upon choice of unit system due to underlying trends in stratospheric temperature. This just means it is important when comparing multiple analyses done with different unit representations to not expect the values to be the same. While this difference does not imply a unit conversion error, potential unit conversion errors can have an impact on the resulting data quality used for trend analyses. McLinden and Fioletov

(2011) show the potential impact using the SAGE II data. However, it is important to note that work made use of version 6.20 of the SAGE II data while more current works use version 7.00 with the important difference of the source of (and consistency of) the meteorological data in the middle to upper stratosphere. Damadeo et al. (2013) showed that the version 7.00 ozone product was much more robust for trend analyses than version 6.20 as a result of the change in meteorological data used for the retrieval. Additionally, Hubert et al. (2016) showed that the SAGE II v7.0, HALOE v19, and ACE-FTS v3.0 data products were relatively consistent in all unit representations when compared to the ground network.

All of that having been said, the possibility for unit conversions to introduce additional uncertainties is still present and an extensive study of this impact on resulting trends was not part of this work.

P.4, L. 22. ENSO is mentioned here, but no result was shown. Is it necessary to include it?

Not really. Damadeo et al. (2014) went into much greater detail on the results of the technique, to show that the results of all of the proxies were reasonable. However, even in that work, the impact of ENSO above ~20 km is fairly negligible and thus wasn't worth going into detail either. The same is true here. Rather, in this work, we only go into detail on the proxies that are impacted by the use of multiple data sets (e.g., solar and aerosol) or the sampling (e.g., diurnal) and, of course, the trends.

P.4, L. 22. The shape of the EESC function depends of latitude and altitude. What exactly was used? The authors used 2 "orthogonal" EESC functions and show the trend results. But how does the resulting EESC signal look like? What is the "phase"/delay? If the authors want to have an additional delay for EESC, it is more logical to introduce an unknown time lag.

The EESC proxy used here is detailed in Damadeo et al. (2014). To summarize, it derives from an empirical orthogonal function (EOF) analysis of EESC data with 6 different mean ages-of-air that subsequently have different turnaround times. The leading 2 EOFs account for 99% of the variance and can recreate the original 6 functions to within ~1%. In other words, using these two proxies allows for the regression to independently adjust the turnaround time and even allows for monotonic trend results. As such, there is no manually created phase delay.

P. 8. Solar cycle. It is difficult to get the 11-year solar cycle from SAGE data. Is the estimated solar signal statistically significant? Are the differences in the solar signal at different latitudes significant?

The solar cycle is, perhaps, one of the more difficult responses to derive from regression analyses owing to the fact that its duration of 11 years is often longer than most individual data sets. SAGE data, with a 21 year record, has quite often been used for this purpose. Unfortunately, within the SAGE II mission period (1984 to 2005), the volcanic eruptions of El Chichon and Mt. Pinatubo have coincidentally coincided with solar activity. This has greatly increased the collinearity of the two effects in regression analyses and made it much more difficult to separate the two. This was discussed in Damadeo et al. (2014). As discussed in this current work, using multiple data sets that span an even longer duration including the relative volcanically quiescent time period from 1998 to 2011 seems to have sufficient length and orthogonality to better extract the solar cycle response. Additionally, we have updated Figure 5 to include statistical significance.