Anonymous Referee #3
Received and published: 10 February 2016

General Comments
The paper seeks to explore the response of stratospheric ozone to solar cycle changes as observed in several global ozone data sets through a regression analysis. The observed response is presented, discussed, and contrasted between the various data sets. There appears to be as many differences as there are similarities in the observed responses. This paper is part one of a two-part series, with the second part said to focus more on atmospheric modeling.

While there are major, but correctable, flaws in the analysis of the data sets, the general qualitative conclusions of the paper are likely robust. The quantitative results, however, will need correction for issues listed below. Overall this is an important and timely study that will demonstrate the limitations of the information content in the existing historical stratospheric ozone record, as well as, the dependence of these records on the quality of ancillary data. The title of the paper is appropriate and the abstract is a complete summary of the paper's current content. References included in the paper strike a nice balance in both quality and quantity.

We thank the reviewer for reading the manuscript and providing their helpful comments and suggestions. We address their specific issues in turn below.

Specific Comments

The primary concern with the analysis methodology used in this paper has to do with the regression model - specifically, the lack of a diurnal term. The sparse spatial and temporal sampling provided by an occultation measurement system presents unique challenges. As has been done for decades now, the data are often reduced into monthly zonal mean time series with each mean treated as though it is representative of both the latitude and month of the center of the monthly zonal bin. While this is not usually too problematic, it is also common practice to assume that both the local sunrise and sunset sampling is unbiased in each mean so that the diurnal variability can be ignored.
While many published papers have already ignored the diurnal sampling issue, and undoubtedly many more will be submitted, it is now time to address this problem and develop a suitable approach for mitigating its impact. In order to demonstrate the presence of the issue in this work, look at figure 2. After the interruption of the SAGE II data record in late 2000 due to an instrument problem, the measurements resumed at a 50% duty cycle - alternating between sunset only and sunrise only periods of approximately 1 month duration each. This is seen in the deseasonalized monthly anomalies plotted in figure 2 as an abrupt increase in variance after 2000. The authors simply call this increased noise. Closer inspection of the figure shows that this "noise" in the equatorial-zone anomaly increases markedly with altitude. This variance is not simply noise, but rather the direct effect of a biased sampling of the known diurnal variability in stratospheric ozone. More subtle variances in diurnal sampling occur throughout the record and other latitudes all of which contributes to the apparent "noise". It is not immediately obvious how this additional unmodeled variance affects the regression results, but it is likely that it will correlate with some of the regression terms and produce biased results. Many, but not all (SAGE-GOMOS being an exception),

of the extended time series that add other data sets on to the SAGE II record have ignored the diurnal issue in the normalization process and are therefore highly suspect. More on these extended data sets momentarily. Computing monthly zonal means is not difficult and it should be possible to create time series where the local sunrise and local sunset measurements are kept separate. The regression would have to include a new "diurnal comb" term, but the number of data points in the regression would nearly double. The AR analysis would have to make sensible assumptions before application. Alternatively, it may be possible to keep the existing time series and add a new term that accounts for the relative sunrise to sunset proportions of events contributing to the mean. While also adding a term to the regression, this would reduce the degrees of freedom since no new data points are added to the regression. In this case, the AR analysis would probably apply as it is currently done. Either approach may require additional terms to concurrently deseasonalize the time series, since the current approach of subtracting the mean monthly mean would no longer work. Exactly which results from this Part I are being carried forward into the Part II paper or will emerge as relevant in the combined whole is not readily apparent. If, however, the quantitative results presented here are important, this regression analysis will need to be completely redone.

At the suggestion of the reviewer, we have added in a new term to the MLR model for the SAGE II datasets that quantifies the fraction of sunrise to total (sunrise + sunset) retrievals that are used to produce each monthly and zonally averaged profile. An example of this term for SAGE II v7.0 at 1hPa over the tropics is shown in Figure 2 of the revised manuscript. As the reviewer highlights, this index shows strong time dependency that is likely linked to the increasing variance of the SAGE II ozone anomalies with altitude in Figure 3. This variance should project strongly onto the SR/(SR+SS) term. The results for the SAGE II solar-ozone response in Figures 4 and 5 of the revised manuscript now include the effects of the SR/(SR+SS) term in the MLR. The inclusion of this term does not fundamentally change the diagnosed solar-ozone response, but it does slightly reduce the peak magnitude in the tropical upper stratosphere.

Because the extended SAGE II datasets take varying approaches for dealing with diurnal sampling, as pointed out by the reviewer, we have not included a similar diurnal sampling term in the MLR model applied to those datasets. However, the results for the SAGE II data suggest that this should not have a large effect on the estimation of the solar-ozone response in those SI^{2}N datasets.

Another concern, which is already discussed to some extent in the paper, has to do with the impact of the relative drift between data sets in time series comprised of data from more than one measurement system. It is reasonably clear that, given the timing of the end of the SAGE II data set, any such relative drift will bias primarily into the EESC term. The solar cycle term will also be biased, however, as there is some correlation between the EESC and F10.7 terms. What is not discussed is that the amplitude assigned to the volcanic term will change in response to varying degrees of drift between measurement systems via its correlation with the F10.7 term. This would seem to be an important diagnostic (especially in the upper stratosphere where the expected volcanic influence may, arguably, be small), but in reality it only serves to reinforce the conclusion that the currently available time series are too short to provide sufficient orthogonality between terms with predominantly low frequency content. It would seem prudent to add a figure showing the lat-alt distribution of the amplitude of the volcanic term. Attribution of the actual response of ozone to solar cycle variations solely to the amplitude of the F10.7 term is highly problematic with these extended data sets. Drift corrected composite time series, to the extent that they can be created, would seem to be required. The associated correlations and resultant coupled uncertainties should be more thoroughly discussed. On a positive note, the analysis done to attribute the SAGE II v6.2-v7 differences to either algorithm changes or Met data source selection are enlightening.

We agree with the reviewer that for relatively short time periods it is difficult to separate possible effects of multiple external drivers that may be partly correlated with each other. In the revised manuscript we analyse all datasets over their entire lengths, rather than focusing on a shorter common analysis period. Although this by no means removes all the issues highlighted by the reviewer, it does help with increasing the degrees of freedom available to separate individual drivers.

As a result of comments from the reviewers, we treat volcanic effects differently in the revised manuscript. Rather than including a volcanic term in the MLR, we instead exclude data from the analysis in the 2 year periods following the El Chichon and Mt Pinatubo eruptions. We hope that this addresses the reviewer's concern about aliasing between the solar cycle and volcanic signals.

With respect to drifts between individual datasets, this is most relevant for the extended SAGE II datasets and the SBUV records. We now show timeseries of ozone anomalies for the extended SAGE II datasets, which makes it possible to see behaviours of the combined records. For example, we now point out that OSIRIS shows persistent positive tropical ozone anomalies during the solar cycle 23 minimum and this might contribute to the reduced magnitude of the solar-ozone response in the SAGE-OSIRIS dataset compared to the SAGE II data.

For SBUV records, the process of data selection, calibration and merging is important and this is stated in the manuscript. We show how this can affect the solar-ozone response by comparing two SBUV VN8.6 datasets and describe how the differences in methodologies of these datasets affect the diagnosed solar signal.

A few less critical Specific Comments

Page 7 lines 224-225: Would it be possible to illustrate the effect of the AR2 vs AR1?
As a result of the reviewer comments, the Methodology section has now been significantly expanded and we feel that including a detailed comparison of the use of an AR1/AR2 model for the residuals would make this section even more dense when the effect on the results is minor. We have therefore not included this in the revised manuscript.

In several places, the authors "blame" NMC/NCEP for the poor quality of the SAGE II mixing ratio conversion when, in reality, the method by which the SAGE II team extends the NMC/NCEP profile to high altitude may be the culprit. The details of this process are discussed in a paper this work already references.
Our intention was not to "blame" either the NMC/NCEP or MERRA temperature datasets, but rather to point out that there are uncertainties in the evolution of stratospheric temperatures over the recent past (both in reanalyses and satellite observations). We have

amended the text to read: "The NMC/NCEP data show exceptional behaviour between 2000-03. At 1hPa, there is a warming of more than 3K over this short period, which is coincident with a warming of ~1K at 2hPa. In contrast, at 5 and 10hPa there is a cooling of more than 4 and 2K, respectively, over this period. The magnitude and vertical structure of these changes in the NMC/NCEP record seems inexplicable as to be related to any physical process, particularly when compared to the variations found in the remainder of the record. Some of these issues may be related to the method used to construct the NMC/NCEP temperature record itself. NCEP reanalysis data were only available for pressures greater than 10hPa, requiring the addition of operational analyses to extend the data to the stratopause. Data from an atmospheric model was used to further extend the temperature data to the mesosphere, but these levels are not considered here (see e.g. Damadeo et al (2013) for more details). The NMC/NCEP temperature record used to convert SAGE II is therefore constructed from several component datasets. Regardless of the exact cause, it seems likely that some of the temperature variations in the NMC/NCEP record are spurious and this may impact on the diagnosed SOR in the SAGE II v6.2 mixing ratio data."

Lines 526-527: Is the amplitude of the volcanic response invariant as the time sub-period is changed?
As noted above, to avoid possible aliasing between the solar cycle and volcanic signals the periods following major tropical volcanic eruptions are now excluded from the MLR analysis and we therefore no longer diagnose a volcanic response.

Minor Technical Comments

The URL link on line 111 does not appear to work as expected
URL link has been updated.

The time series show in Figure 1 should be extended to match the longest time period to which they are applied.
Change has been made.

Why does panel (b) in figure 5 appear to contain much coarser latitudinal structure than seen in panel (a)?
MERRA data have now been regridded to the same resolution as NMC/NCEP in this Figure.

Line 500 and elsewhere, this reviewer had some difficulty determining whether the term "increase" referred to trends or the solar cycle response. The authors may wish to introduce and use an acronym (e.g., SCR - Solar Cycle Response) to help clarify the topic under discussion rather than use a generic term such as "increase".
We have attempted to clarify the language to avoid the use of general statements such as "increase" and have introduced an acronym for the Solar-Ozone Response (SOR), as suggested by the reviewer.

Line 548: "improve" should be "improved".
Change has been made.