

Review by Judith Lean

Received and published: 12 February 2016

General Comments

As a result of the criticisms made by the reviewer we have made substantial modifications to the manuscript, as described below and in the 'General response to all reviewers'. While we accept some of the criticisms raised by the reviewer, and have modified the manuscript accordingly, we maintain that documenting and comparing solar-ozone signals in different satellite ozone datasets is a valuable contribution to the scientific literature. This is particularly motivated by the fact that there have been recent updates to the two main long-term satellite records, SBUV and SAGE II. Given the length and degree of detail of many of the reviewer's comments below, we have only attempted to respond to the main points raised. However, we emphasise that the revised manuscript has been substantially modified compared to the original, and we therefore encourage the reviewer to re-read the entire manuscript.

The task of this paper is, according to the title "The representation of solar cycle signals in ozone". The present paper is Part I of two papers, this first focusing on using observations to determine the solar cycle representation, the second using models. The overall goal of the work is to prescribe the ozone changes during the solar cycle for use as input to climate models that do not themselves calculate the ozone responses directly, thereby allowing these models to include indirect effects of solar forcing.

The manuscript states that the "goal is to synthesize current knowledge to inform a recommendation for including the solar-ozone signal in the prescribed ozone dataset being created for CMIP6". It does not accomplish this task. A representation of the solar cycle signals based on a synthesis of observations of stratospheric ozone is not recommended. Instead, the paper concludes that it is "unlikely that satellite ozone measurements alone can be applied to estimate the necessary solar cycle ozone component of the prescribed ozone database for future coupled model intercomparisons".

These aspects of the paper preclude publication in its present form. Unless, or until the authors can/do produce the product that they set out to produce then it would seem that a paper about not achieving their goal is unwarranted. Their conclusion that satellite data are not able to inform such a product is incorrect. Ozone has been measured in one form or another by multiple instruments for at least three solar cycles. Others have demonstrated that there is sufficient information available to quantify the solar cycle in ozone using these data, albeit with uncertainties (perhaps even large ones). Bodeker et al. (Earth System Sci Data 2013) produce just such a product. Their Tier 1.4 database is the natural component (solar plus volcanic) of vertical ozone profile variability extracted from observations by linear regression, from which the solar component can be further extracted. As well the manuscript cites various other such products reported previously e.g., by Randell and Wu, Hood and coworkers etc. So it's not that this task can't be done, it's that the present manuscript doesn't do it.

We have reframed the objectives of our study in the revised manuscript. We no longer state the aim of "synthesizing current knowledge to inform a recommendation for

including the solar-ozone signal in the prescribed ozone dataset being created for CMIP6". Instead we focus on comparing the solar-ozone responses in recently updated long-term satellite datasets (SAGE II, SBUV) with their predecessors. The study documents these differences and where possible gives insights into why these occur. These comparisons are important to document because many previous studies of the solar-ozone response have been published using the older datasets, but many fewer with the newer datasets, and in some cases substantial differences are found between them. This will provide a valuable point of reference for other researchers who are studying solar cycle effects on climate.

In its present form, the manuscript focuses on comparing multiple (nine plus) ozone datasets, each produced from a variety of data reduction techniques and assumptions, combined in various ways and covering different time periods and different lengths of time. For the most part, these datasets have already been described and compared in detail, including using multiple linear regression analysis. So the dataset comparisons themselves are not new material.

While multiple linear regression analysis has been applied to many of the datasets used in this study (e.g. Tummon et al., 2015), the main goal of this other work has been to document linear ozone trends, and most of them therefore do not discuss the solar-ozone response whatsoever. There has been no recent comparison of the latitude-height structures of solar-ozone responses across multiple satellite datasets (see e.g. the last major effort by Soukharev and Hood (2006)). We therefore disagree with the reviewer that these comparisons do not present new material. To give just one example from our study, the comparison of the solar-ozone responses in SAGE II v6.2 and v7.0 is new and provides useful and important insight.

Time Span of the Regression Analysis

For extracting the most reliable decadal solar signal the database should be as long as possible to minimize cross projection of the solar and other influences. As noted above, by limiting their analysis of ozone's response to solar variability to data mainly from 1984 to 2004 the authors are not utilizing the longest datasets available (by far!). Ozone data extend from 1979 to 2015 – during the additional 10 years from 2004 to 2014 the solar cycle has increased to another maximum, volcanic aerosols have been minimal, EESCs have continued to decline, and greenhouse gases to increase.

The authors investigate the effect of time spans on their results but their approach is not systematic and they conclude only that different time spans give different results. Their Figure 11 shows the patterns of ozone's solar cycle response derived from 21-year datasets over different epochs. It is probably not surprising that using different 21-year epochs gives different statistical models since, for one thing, the correlation among the various predictor times series likely differs for each 21-year period and this may affect the derive coefficients. This could occur even if the ozone time series were "perfect" so it is not necessarily true, as the authors conclude, that the differences in the ozone representations obtained from different 21-year epochs is due entirely to uncertainties in the database.

Examination of the stability of the statistical model coefficients is indeed important but the approach of using separate 21-year epochs is perhaps not the best way to establish

this. As a dataset lengthens, the magnitude of the coefficient of a given predictor should converge to a “stable” value. The usual way to evaluate the stability of the coefficients is to start with a core dataset of maximum length and then reevaluate the model coefficients using successively shorter lengths of the primary datasets. In revising their manuscript, the authors might consider developing a more quantitative metric (solar signal in total ozone derived by integrating the vertical profiles?) to establish the stability of the model coefficients as a function of length of the dataset. When accomplished using the longest possible dataset, such a metric will automatically provide feedback about limitations using shorter datasets (such as HALOE).

The authors proceed to extract solar signals from all nine plus of these datasets mainly over a common 21-year period 1984 to 2004; that there are notable differences leads them to conclude that the observations are not adequate to extract the solar cycle signal. But just because some datasets over this 21-year time period are not suitable for this task, does not mean that some longer datasets are also not; more importantly a few of the datasets extend over the period 1979 to 2015 – 37 years – and these (much) longer time series are more suitable and more likely to give meaningful results than those analyzed for just 21 years. A key characteristic of a database suitable for extracting a decadal solar cycle signal is that the record be as long as is possible. Even 37 years is just 3 solar cycles. This is crucial so that the regression analysis can properly separate the solar cycle from other influences on decadal time scales, namely EESC and volcanic activity, each of which has decadal scale variability. It is not surprising that datasets like HALOE that cover only the period 1991 to 2005, or other datasets analyzed over only 14 years (barely one solar cycle) do not yield statistically meaningful results or that results among them differ (e.g. Figure 10). To achieve the stated goal of the paper – namely the representation the solar signal in ozone using observations - the authors should use the longest available time series of ozone profiles that exists so as to minimize correlation among the predictors and decrease regression model coefficient uncertainties– this is from 1979 to the present 2015. A dataset of 37 years is far superior (even with other extenuating instrumental limitations) to one of 21 years for extracting the decadal solar cycle. Yet nowhere in the paper do the authors ever use the entire dataset available.

As requested by the reviewer, in the revised manuscript all ozone datasets are analysed for their entire lengths. We have removed the section discussing the stability of the regression coefficients. Furthermore, the revised manuscript is more selective about which datasets are included, and results from short records, such as HALOE, are no longer presented. Instead we focus on comparing SAGE II v6.2 and v7.0, a subset of SI2N extended SAGE records, and SBUV VN8.0 and VN8.6.

A truly observational representation of the solar cycle in ozone, which this paper seeks to achieve, is important for input to physical model simulations and as independent validation of those simulations. In its present form, the manuscript does not provide useful material for climate models to use. One assumes therefore, that they will recommend in Paper II that a representation of the solar cycle signal in ozone be based on model simulations. Since the present manuscript does not address such models - or assess their limitations - this conclusion cannot be justified here, especially given the many known uncertainties and limitations of models. Models require observations for validation and so in order to be suitable for publication, this paper needs to produce

this observational product - with as many caveats as needed and with realistic (possibly large) uncertainties– or be withdrawn.

In the revised manuscript, we no longer make reference to a recommendation for CMIP6. Furthermore, we make clearer recommendations for which SAGE II and SBUV datasets are likely to be most reliable for diagnosing the solar-ozone response.

Fortunately, from their exhaustive (albeit non-selective, somewhat unfocused and inconclusive) analysis of multiple datasets, the authors have the material available from which to extract the needed product and to avoid the obfuscation that results from reporting the details of multiple (less worthy) ozone datasets.

The revised manuscript has narrower aims and presents results from only a subset of the ozone datasets in the original manuscript. We believe that this results in a more selective, focused, and conclusive study.

One way for the authors to proceed to revise their manuscript and achieve their goal of producing an observational representation of ozone's response to the solar cycle (with uncertainties) might be as follows:

- 1) Select (on the basis of prior work and current analysis) the longest, most reliable SAGE-extended dataset.
- 2) Select (on the basis of prior work and their own analysis) the longest most reliable SBUV datasets (possibly NOAA Cohesive Merged Dataset – see below).
- 3) Develop a robust multiple linear regression methodology for extracting solar cycle signals separately from these two independent datasets on their native altitude and spatial grids (for this purpose the current MLR model needs to be expanded and tested – see below).
- 4) Merge, or otherwise combine/integrate- the solar representations from the two different approaches and propagate uncertainties from the statistical regression coefficients of the two representations.
- 5) Constrain as needed the ozone profile solar cycle changes to be consistent with that of total ozone (which the present analysis doesn't utilize at all).
- 6) Compare with Bodeker et al Tier 1.4 representation of vertical ozone profile responses to solar cycle or some other published representation (e.g. that used in CMIP5); discuss and quantify limitations and uncertainties in the final observational product.

Since the objectives of the manuscript have altered compared to the original submission, we have incorporated some but not all of the reviewer's recommendations (see below).

Specific Comments

Requirements of Physical Models

Nowhere does the manuscript define requirements for the task they are undertaking, namely the details of the inputs that physical models need for the ozone representation. What is the altitude grid (and resolution), latitude and longitude grid (and resolution) for which the models need these inputs? Lacking these stated definitions, the authors are without robust criteria upon which to build their product, or to assess how well the final product meets the requirements. Although the SBUV observations have poorer vertical resolution than do the SAGE observations, is this actually an issue for the climate models? What vertical (and spatial) resolution is actually needed? How does the magnitude of the uncertainties affect the usefulness of the observational representation; the authors show regions where significance exceeds 95% but it is quite likely that the patterns of the response, even including regions that are less statistically significant, can still provide useful model input and validation.

Is the ozone product to be represented by absolute values or anomalies – if the latter than is a reference distribution also needed?

The revised manuscript places much less emphasis on the goal of creating an ozone dataset for models, and thus some of the comments raised above are no longer applicable. Nevertheless, we have added a paragraph at the end of Section 1 that states “Given the potential application of the results described below for use in climate model simulations, it is prudent to briefly review the typical requirements of an ozone database for models by describing the CMIP5 dataset as a representative example (Cionni et al., 2011) (see also Bodeker et al. (2013)). The CMIP5 ozone database provided monthly mean ozone mixing ratios on a regular latitude/pressure grid at a horizontal resolution of $5^{\circ} \times 5^{\circ}$ (lon/lat) on 24 pressure levels covering 1000-1 hPa for the period 1850-2100. Data were provided on the following pressure levels: 1000, 850, 700, 600, 500, 400, 300, 250, 200, 150, 100, 80, 70, 50, 30, 20, 15, 10, 7, 5, 3, 2, 1.5, 1 hPa. Stratospheric ozone data (at $p \leq 300$ hPa) were given as zonal mean values. Therefore for any de- scription of the SOR must fulfil these (or similar) criteria to be viable for use in climate models (i.e. global coverage at monthly mean resolution and with sufficient vertical and horizontal resolution throughout the stratosphere).”

Dataset Comparison and Selection

As the authors explain, multiple ozone datasets now exist, in basically two categories – those based on the SAGE observations and those based on the SBUV observations. Much of the material in Section 2 – currently called “Methods” - is really about the datasets and might be renamed “Datasets” (with the MLR presented in its own separate and expanded section). A number of other papers already describe these various datasets in great detail - including their multiple linear regression analysis - and to some extent the current paper repeats more of these results than might be necessary, in that this paper is about the solar signal in the datasets, not the datasets per se.

We have renamed Section 2 “Ozone datasets”. As noted above, although MLR analysis has been applied to ozone datasets in the past, there has been no recent effort to analyse and compare the structures of the solar-ozone signals in different records (see e.g. Soukharev and Hood (2006)). This is certainly the case for SAGE II v6.2 and v7.0, as well as the SP²N datasets.

Most importantly, the authors should describe the selection of, and reasons for their selection, of the most suitable (longest) dataset in each category then extract the solar signals that will inform their product for climate model use from just these two datasets. In an unfocused and round about way, the present manuscript actually does this to some extent, but it doesn't use the longest available datasets and it fails to synthesize the many derived representations into a final product.

The SBUV record exemplifies this more streamlined approach. In the present manuscript the authors calculate and examine the solar signals in all three SBUV records when a judicious assessment of prior work would likely lead them to decide, right from the beginning of their study, that the best record for the present purpose is likely the NOAA Merged Cohesive Data from 1979 to 2015 - for a number of reasons (which the authors should lay out clearly and concisely). In a revision of the paper along these lines, the entire Section 3.3 about the SBUV record could appear much earlier in the paper, in a section about Datasets. The general (valid) conclusion presented already in this paper is that there are distinct instrumental differences between the NASA MOD V8.0 and MOD V8.6 datasets – a conclusion that Lean (JAS, 2014) also reached and discussed in detail, and which the authors might simply reference to justify why they probably shouldn't use MOD V8.6. So then, the authors need to obtain the solar signal in just one SBUV- based dataset. By elimination this would seem to be the NOAA Merged Cohesive Dataset which the authors show has a solar response like that of the (shorter) NASA MOD V8.0 dataset. They could then test and evaluate more comprehensively the MLR results for this one dataset (e.g., effects of lags, cross-correlation of predictors, addition of trend term, stability of model coefficients with length of database etc).

A similar selection could be made of the SAGE-based datasets – utilizing the material about these datasets already published and whatever additional tests the authors undertake to clarify their selection. A distinction among the SAGE datasets is that different versions use different temperature databases which the current manuscript expends much effort in analyzing and comparing. Another few pages of the manuscript (pages 11 and 12) compare various characteristics of the datasets such as their trends. But a comprehensive regression analysis should capture and extract these different trends without the trends necessarily affecting the separate (independent) solar component. And as with the SBUV dataset selection, this material about the SAGE-based datasets could be included in a section on Datasets – then in a subsequent section on Results, the authors need only extract and examine the solar signals from one, preferred, SAGE-based dataset - the longest one.

The revised manuscript presents results for a smaller number of ozone datasets (8 instead of 13) and analyses them all over their full lengths. The selection of datasets is justified by the revised goals of the study to compare solar-ozone signals in recently updated datasets (SAGE II v6.2 vs. v7.0 and SBUV VN8.0 vs. VN8.6), and to evaluate recent extended SAGE II datasets. Based on the comparisons amongst these versions, and consideration of the main sources of uncertainty, we now include recommendations for which are the most reliable datasets to use for diagnosing the solar-ozone response. We do not agree with the reviewer that the SBUV Merged Cohesive Data VN8.6 dataset is necessarily better than SBUVMOD VN8.6 for analyzing the solar-ozone response. Tummon et al. (2015) highlight a number of issues with the representation of long-term trends in the Merged Cohesive

Dataset, which may also affect the solar-ozone response. We therefore present both SBUVMOD VN8.6 and Merged Cohesive VN8.6 and discuss their respective pros and cons.

What about total ozone? Nowhere in the paper is use made of total ozone datasets, even though whatever altitude profiles are specified must be consistent with the total ozone amount. Total ozone datasets have, arguably, greater long term stability than the profile datasets, so that solar signals can - and have been - extracted somewhat reliably; differences among the different datasets can therefore guide the selection of the profile datasets, as suggested above for SBUV datasets. Analysis of the SBUV MOD V8.6 total ozone compared with TOMS MOD V8 strongly suggests instrument effects in SBUV MOD V8.6 around the 1996 time frame that cause a smaller solar response in MOD V8.6 than in MOD V8.0.

We now include a paragraph that discusses literature on the solar-ozone response in TCO. However, most of the results in our study are related to differences in the solar-ozone response in the upper stratosphere, which make only a small contribution to the TCO signal (e.g. Hood (1997)). Therefore to keep the study focused on the latitude-height structures of the solar-ozone response we do not include new analysis of total column ozone. As noted by the reviewer, the solar-ozone response in some of these column ozone datasets has already been analysed and described elsewhere (e.g. Lean 2014), and we now include text referencing these results.

The Regression Model Formulation and Results

Multiple linear regression (MLR) is the method that the authors use to extract the sought-for representation of the ozone's response to the solar cycle. Presenting the direct time series together with the solar index (Figure 2), and discussing (page 8) how a solar signal is not directly evident is not all that useful - it is not even evident in total ozone because of the various other influences.

The MLR analysis of the type that the authors use has been used extensively to model ozone variability statistically in terms of individual influences. But the model that the authors use doesn't take advantage of the understanding achieved from those prior studies. As a result it likely doesn't provide the best representation of the various predictors, especially EESC but also ENSO and trends.

EESC: As described for example in Bodeker et al (2013) and Lean (2014), among others, the EESC depends on the age of air and on bromine. It therefore has latitude - and possible altitude - dependencies. The peak of the EESC temporal structure shifts accordingly, which affects how the MLR apportions variance among EESC and other influences. The authors don't say which EESC profile they use but it seems that they use only one profile. However, there is no one EESC profile that is appropriate for all latitudes. Rather a suite of profiles is needed corresponding to different ages of air (and bromine); these can be obtained from GSFC's on-line capability.

The reviewer is correct that we have used a single EESC index for all latitudes/altitudes, which neglects the effects of age of air on EESC. We have tested the sensitivity of the MLR results to this assumption by using latitude-height dependent EESC timeseries taken from a chemistry-climate model (UM-UKCA) simulation (REF-C1 CCMI integration). This model

simulation implicitly includes effects of variations in stratospheric age of air on EESC. However, using this more sophisticated EESC index does not affect the diagnosed solar-ozone response in the datasets, and we therefore retain the use of a single EESC index for simplicity. A line has been added to the text that describes this sensitivity test to justify our choice.

ENSO: the ENSO signal in ozone may lag the MEI index by a few months – while this is not an issue for mid and high latitudes, where ENSO signal is small, it could affect tropical signal. Is zero lag of the MEI index the most appropriate lag?

We have tested the effect of lagging the Nino 3.4 index by 0-12 months and find that this does not affect the diagnosed solar-ozone response. We now state this in the revised manuscript. We also note that other recent MLR studies of long-term trends in satellite ozone datasets have also not used a lag for ENSO (e.g. Tummon et al., 2015; Harris et al., 2015).

Solar Irradiance: Why do the authors use the 10.7 cm flux and not modeled UV irradiance or the Mg index, which provide better representations of the true solar changes? While it is likely that for monthly means these differences won't be large, the authors nevertheless need to acknowledge possible limitations of the use of the 10.7 cm index in representing solar UV irradiance.

We have added a sentence “We adopt the widely used F10.7cm solar flux as a proxy for solar activity in the MLR model. This is a more appropriate measure for variations in the UV spectral region, the key driver of the stratospheric ozone response, than other indices such as total solar irradiance (Gray et al., 2010); however, it should be noted that the F10.7cm flux is not a direct measurement of UV variability, but rather is a proxy for variations at these wavelengths.”

Trend: The authors state that they include a trend in some instances but that it makes little difference to their “results” by which they presumably mean the MLR solar cycle coefficients (see Technical comment below about confusion due to the authors’ generic use of “results”). This is possibly because the 21-year time period of their analysis is too short. Using the longest available datasets, the authors would be able to – and would need to - separate the EESC and trend components. The manuscript discusses in quite a lot of detail the different trends among the different datasets with the inference that this provides a measure of the dataset quality. But as long as the trend and solar cycle terms are sufficiently orthogonal (low correlation coefficient) then a linear trend shouldn't influence the extraction of the solar cycle representation. The difficulty is that the long-term drifts in the ozone records aren't actually linear trends and that further more, the combined EESC and GHG influences aren't linear either. Other authors have used piece wise linear regression to account for the fact that the actual trend in ozone is some combination of that due to EESC and due to GHGs. Modelling the longest available ozone dataset to achieve the most reliable solar signal possible will very likely require the inclusion of a realistic trend term.

We now include a CO₂ term in the MLR model. However, it does not have a large effect on the diagnosed solar-ozone response.

Seasonality: The present paper cites the need for specifying the seasonality of the ozone response to the solar cycle, and the authors explore this (Figure 12) but without coming to concrete conclusions. Bodeker et al (2013) report that the seasonality is not pronounced for solar cycle variations in their regression model analysis. The authors could investigate this more robustly in a number of ways. One approach would be to add additional cosine and sine terms to modulate the solar index in the regression model. Another would be to perform the regression using 3 months to define the four primary seasons, which would likely be more representative (less noise?) than the figures shown for each sequential month in Figure 12. Either way, some statement is needed about the magnitude of this effect – is it important or not? (in a statistical sense). Will the solar signal representation still be useful as input to the physical model simulations without the seasonal modulation (see above comments about specifying requirements)?

Part of the motivation for including a section on “Seasonality in the solar-ozone response” in the manuscript is that this is a relatively unexplored area of research, and one that the authors believe warrants further investigation. Specifically, there is no scientific basis to answer the reviewer’s important question: “Will the solar signal representation still be useful as input to the physical model simulations without the seasonal modulation?” because a comparison of e.g. the climate response to solar forcing in a model with and without a seasonal modulation of the SOR has not been performed.

It is therefore difficult within the scope of this study (i.e. without performing the above model calculations) to quantify how important this effect is for climate simulations. Instead, we include a short section that explains why a seasonal component to the SOR is to be expected from photochemical theory, and possibly also from coupling between ozone transport and dynamics (see e.g. Hood et al. (2015)). In a similar manner to Hood et al (2015), we then attempt to extract this seasonal dependence from an observational dataset in spite of several challenges described in the manuscript.

There is motivation to discuss seasonality in the SOR on monthly timescales for a number of reasons:

- There may be intraseasonal variations in the SOR that would be smoothed out by taking a seasonal mean: e.g. Hood et al. (2015) emphasise the importance of solar-induced ozone anomalies at high latitudes in early winter.*
- Climate model ozone datasets are typically produced at monthly resolution.*

Upon performing the analysis, we find regions where the magnitude of the SOR on monthly timescales is considerably different from the annual mean. We then provide a discussion of the implications of these findings that concludes this is an under researched area and new studies are required to establish whether or not such seasonal fluctuations in the SOR are important for models.

Once a properly formulated MLR is applied to extract the solar cycle signal in each of the two basic (SAGE-based and SBUV-based) datasets, the extracted solar signals then provide baseline solar cycle representations, with associated statistical uncertainties. Post processing of these two solar representation could include (uncertainty weighted)

averaging and/or merging and error propagation to achieve a coherent synthesis and final product that the present manuscript lacks.

One of our main conclusions is that the SAGE data are most reliable for assessing the solar-ozone response in number density units. We have therefore not averaged the recommended SAGE and SBUV datasets to produce a final merged product because SBUV is provided as mixing ratios. Instead we suggest that chemistry-climate models be compared to both datasets in each of their native coordinates.

Derived Product –Quantitative Assessment, Uncertainties, Comparison with Similar Products.

The authors make numerous comparisons and describe differences among multiple ozone (and temperature) datasets but they do not synthesize their results or reach quantitative conclusions about the solar-ozone representation for input to climate model simulations. Having determined a best possible solar cycle representation of ozone and corresponding uncertainties, they could assess how this compares with independent representations and whether or not the uncertainties make the derived solar signal representation useful for input to the model calculations. The answer to this will depend on the magnitude of the uncertainties in conjunction with the requirements of the models (as noted above). Even if the uncertainty in the solar cycle signal is 50%, this may still be useful if, say, the models differ by 100% among themselves.

Independent validation is possible by comparison with Tier1.4 or other solar cycle representation of ozone published previously. This is different from the analysis that the authors have made to extract the solar signal, using their own regression model, from the Tier 0 database. Rather, comparing their solar cycle representation of ozone with that of the Bodeker et al. and others takes into account the different regression model formulation as well as the different datasets. Additionally, since it is likely that climate models will use the Bodeker et al. ozone profile dataset (it was developed expressly for that purpose) it would be good to compare the solar cycle representation in that dataset with the one developed from this work.

The solar-ozone response in the Bodeker et al dataset will be strongly affected by the version of SAGE II that it includes (i.e. SAGE II v6.2). Therefore the comparison of SAGE II v6.2 and v7.0 for number density and mixing ratios is extremely relevant for understanding the solar-ozone response in ozone datasets that are expressly produced for models such as Bodeker et al. and Cionni et al.

Since the scope of the manuscript has been revised, and the analysis is no longer orientated around CMIP6, we will instead include analysis of the solar-ozone response in ozone datasets for models (Cionni et al; Bodeker et al) in Part II of the study, which focuses on models. The revised manuscript includes more information about the statistical uncertainties in the solar regression coefficients extracted from the MLR which will enable the observed uncertainties to be more readily compared to uncertainties amongst models in Part II.

Technical Comments:

The text of this manuscript is not as precise as it might be and the authors should provide additional clarity when revising the manuscript. It can be quite difficult to discern the actual analyses being described. The discussion of Figure 3 is an example of this. The text states that “Figures 3(a) and (b) show latitude-altitude plots of the percentage differences in ozone number density between solar maximum and minimum for SAGE II v6.2 and v7.0, respectively”. What is not at all clear is how these percentage differences were obtained. On first reading they seemed to be differences of the direct time series themselves (shown in Figure 2) during solar maximum and minimum conditions. If so, then because of the other influences on ozone, such direct differences are not reliable indicators of the solar cycle signal. Are they, instead, derived using the MLR? The text doesn’t say.

Another example of this general lack of precision of the text is frequent generic reference to “results” (as noted above). An example is on page 16: “to test this, we add a linear trend term into the MLR; however, this does not strongly affect the results as compared to Figure 11 (not shown).” By results they presumably mean the coefficients of the solar predictor in the statistical model – so this is what they should say in the text. More generally, the paper presents many “results” about many things and each one should be properly articulated. Even if this lengthens the manuscript, it makes the message far more clear for the reader.

The acronym solar-ozone response (SOR) has been introduced throughout the text to clarify various references to “results”. We have also made textual changes throughout the manuscript to improve the overall clarity and precision.

Line 20 ff: the authors refer throughout to the SBUV VN8.0 and 8.6 datasets – do they mean SBUV MOD8.0 and 8.6 – where MOD is Merged Ozone Data?

Changed to SBUVMOD VN8.0 throughout.

Line 29-30: The authors make the mistake of confusing the absolute energy changes in the total and UV spectrum with their relative changes and suggesting that TSI changes are somehow much less than UV changes. The change in TSI from solar max to min is larger in absolute energy units than is the change in the UV spectrum. The manuscript needs to clarify this.

Changed to: “Whilst fractional changes in total solar irradiance (TSI) between the maximum and minimum phases of the approximately 11 year solar cycle are known to be small (<0.1%), there is enhanced fractional variability in the ultraviolet (UV) spectral region (>6%) (e.g. Ermolli et al. (2013)).”

Lines 454-477: These differences have been reported and interpreted previously in the corresponding total ozone datasets of the two versions. Analysis of the solar (and other signals) in the Ozone MOD V8.0 and MOD V8.6 total ozone datasets (Lean, JAS, 2014) shows the regional and global differences for the solar cycle signal in the two different datasets, finding a smaller solar cycle amplitude in the MOD V8.6 of 3DU versus 5DU on Mod V8. That paper discussed these differences, including the calibration issues.

We have added a citation to Lean (2014) and discuss their main findings.

Page 20, lines 670-673. The authors should remove their “encouragement” of instrument teams to better analyze their data! The instruments teams are already undertaking a very challenging and difficult task of space-based metrology and are, without doubt, (more than) fully aware of the need for properly specifying instrumental effects in their datasets as well as they can!

Removed.

A more helpful conclusion would be for the authors to generate an actual quantitative product – namely the solar representation of the solar cycle in ozone that is the stated goal of their paper - and assess the future needs of ozone observations in the context of the associated uncertainties of that product.

As noted above, the goals of the study have been reframed in the revised manuscript to focus on comparisons of the solar-ozone response between recently updated and previous versions of the main long-term satellite ozone datasets (SBUV and SAGE II). The conclusions have been amended accordingly.