

## Anonymous Referee #4

Received and published: 11 February 2016

### Summary

To correctly investigate the impacts of solar variability on the climate using models, it is imperative that models correctly simulate the stratospheric ozone response to the solar cycle that may lead to an impact upon surface climate. The ozone response and feedback may be as important as the direct heating effect from solar variability. Thus, for models that do not calculate ozone variability online, a realistic representation is vital to correctly simulate this solar pathway to impact the climate. This requires using either modelled ozone responses, or those taken from multiple satellite sources. The difficulty is that different satellite data can tell different stories of how the stratospheric ozone has varied over time. Thus quantifying the true behavior, and extracting the solar component as an input into models is difficult. This paper, as the first of two parts, aims to investigate the behavior of several (SI<sup>2</sup>N) merged datasets that are relatively new and many of which are either based upon, or have a large component from, the long SAGE-II record. The authors present a comprehensive comparison of the behavior of the extracted solar signal in all seven SI<sup>2</sup>N ozone datasets. They also seek to quantify and understand why two versions of the same data in units of volume mixing ratio, v6.2 and v7.0, differ so much. Identifying the source of the difference due to the temperature data used to convert from number density, the authors apply their own conversion to investigate the differences in the two versions of SAGE and then expand their investigation and discussion to the other datasets based on SAGE. While this paper does not lead to a better understanding of why the solar signal extracted from the SI<sup>2</sup>N data differ so much with each other (except versions of SAGE II), or hint which one is likely the best to use in future studies, this is an important contribution to the field. The knowledge of how and where datasets differ will provide a step towards, not only, understanding the datasets, but potentially improving them in future work. The work done to understand why SAGE II v6.2 and 7.0 differ was an interesting, revealing and useful analysis. The results are generally clear and well communicated. In context of the two-part study this analysis aims to, and presumably will, inform in the production of an input ozone data set for the CMIP6 modelling runs. From the view of this reviewer, following a point that needs addressing, and some clarifications, the paper fulfills its aims and will be ready for publication.

*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

Page 25, lines 7-10: The point is made that for each of the sub-periods considered in the MLR in Figures 11a-f, both El Chichon and Mt Pinatubo are included. With both eruptions included, and yet the upper stratosphere spatial pattern changing with each sub-period, the authors suggest that this implies differences are unlikely to be volcanic. As stated by the authors, the eruption of El Chichon occurred in April 1982, so the effects of the eruption on the stratosphere would be expected to have gone by early 1985, and be less pronounced in 1984 than 1983 and 1982. However, while the authors are correct that both eruptions are included in 11a-b, and any effects likely present in

11c, El Chichon is not included in panels 11d-f, which begin in 1985, almost three years after the eruption. While temperature responses to volcanic eruptions are stronger in the lower stratosphere, and less, if any in the mid-stratosphere, there are hints that the mesosphere sees a response to volcanic eruptions (e.g. Beig et al., 2003), so there may indeed be an aliasing with volcanoes that decreases (as seen in 11c-f relative to 11a-b) when El Chichon's effect is removed by the sub-period chosen. Further to this, there is a very large anomaly present in both SBUVMOD and SBUVN8.0 around the time of the 1982 eruption (Fig 8) that lasts for 1-2 years; a similar event is not present following the Pinatubo eruption, and so is likely not of volcanic origin (unless it is related to a change in the atmospheric viewing of the instrument for a reason unique to El Chichon). Such a large anomaly may have an influence on the MLR leading to the change in the spatial patterns plotted in Figure 11, and then the authors may indeed be correct that it is not volcanic in origin. Perhaps it would be worth applying SBUV-Merged Cohesive to test this, as the anomaly is not visible in that time series in Fig. 8.

*Figure 11 has now been removed from the revised manuscript as a result of comments from another referee. Furthermore, the periods following the two major volcanic eruptions (El Chichon and Mt Pinatubo) are now excluded from the MLR analysis to reduce the potential for aliasing between the solar cycle and volcanic signals.*

**The following are suggestions for the authors to clarify or reword:**

Page 13; lines 20-26: Indeed, there also appears to be a larger positive anomaly in 1992-1994, a period of maximum and high activity, so it is possible this may also contribute to the enhanced signal seen in Fig. 3c.

*As noted above, the period June 1991-May 1993 has now been excluded from the analysis to remove the effects of the Mt Pinatubo eruption. This change has little impact on the results in Figure 3(c) (now Figure 4(c)).*

Page 15; line 13: While records indeed show there was a warming of  $\sim 0.5$ - $1.0$  K following the Pinatubo eruption, and perhaps there is a small increase in 1991/1992 in Fig. 4, it appears that at 30 hPa a warming began in 1989, followed by the  $\sim 2$  K decrease after 1992. The eruption does not appear to stand out in this time series, so perhaps the authors may wish to revise the focus of their comment here.

*We have amended the text to read: "At 30hPa, the evolution of the two temperature records is nearly identical during the period of overlap, with a long-term cooling trend of  $\sim 0.6$  K decade<sup>-1</sup> that is strongly connected to an apparent step-wise cooling of  $\sim 2$ K between 1992 and 1994."*

Page 15, lines 18-19: Two points of clarity here. It would better to reformulate to discuss NCEP first, as the MERRA data do not show the decline in the last three years, but in the last three years of NCEP. Note also that the solar cycle decline began in mid to late 2002, so this three year period is mainly during the maximum period. This of course does not change the point being made, and it is well worth highlighting also that this odd behavior in NCEP also occurs (though inversely) in this same, three-year period, at 5 and 10 hPa, hinting at an issue with NCEP.

*We have reformulated and expanded these sentences as suggested by the reviewer (see revised manuscript).*

Page 20, lines 19-22: The authors state that many of the SAGE-II based datasets have differences that are likely the result of merging procedures. Do the authors include in this comment also that SAGE-GOMOS 1&2 and SAGE-OSIRIS have less data (or more data gaps) in the equatorial region than SWOOSH and GOZCARDS? Or if not, might this additionally decrease the significance of the signal in the tropics and lead to the less 'smooth' appearance of the spatial patterns? The reviewer is also aware that Aura MLS v2.2 used in GOZCARDS, and v3.3 in SWOOSH have different short-term variability (larger in GOZCARDS). This might be worth checking/considering.

*In the revised manuscript, only 3 extended SAGE II datasets are analysed which provide ozone number densities rather than mixing ratios (SAGE-GOMOS 1/2 and SAGE-OSIRIS). Therefore SWOOSH and GOZCARDS no longer feature in the revised manuscript.*

*We now include figures in the Supplementary Information that show the number of data points used to diagnose the solar-ozone response as a function of latitude and height. This enables a comparison of the sampling by the original SAGE II data and the extended SAGE II datasets with the SBUV records. These differences in sampling between the datasets are briefly referred to in the main text.*

Page 22, line 29: The two SBUV records have almost the same datasets used, except the Merged Cohesive uses a little over a year from NOAA9 (see Tummon et al., 2015, Fig 1.). We have amended the text to read: *"The two SBUV VN8.6 datasets contain some differences in the data that is included from different instruments within a particular period (see Figure 1 in Tummon et al (2015)), and in the methods for averaging and merging these data. SBUV Merged Cohesive VN8.6 uses data from a single instrument in any time period; the individual records are then bias-corrected to produce a continuous record (Wild and Long, 2015). In contrast, SBUVMOD VN8.6 is constructed by averaging all available data within a particular time window (Frith et al., 2014). The SBUVMOD datasets extend back to 1970 by including data from the BUV instrument on Nimbus 4 from 1970-1976, whereas the SBUV Merged Cohesive dataset starts from 1978 with the first SBUV instrument on Nimbus 7."*

Figure 6 might benefit with a third column of difference plots, since the differences are well discussed, though specific altitudes and latitudes are usually not mentioned. Thus the difference plots might make it easier for the reader to locate what the authors are referring to. This is at the authors discretion. However, for the point made about the solar signal in Figs 6g and 6h, relative to 6a and 6b, that the signals are larger in NCEP than MERRA, while in absolute values this is the case, I wonder how significantly different, statistically these signals are? Figures 6g and h seem more similar than 6a and b do in the upper stratosphere; this may be helped with a difference plot with significance, as mentioned.

*A third column showing differences has been included in Figure 6 (now Figure 8 in revised manuscript).*

### **Technical corrections**

Page 19, line 4: "but the magnitudes [are] quite similar"

*Change has been made.*