Author response to review by Anonymous Referee 2

The representation of solar cycle signals in stratospheric ozone. Part II: Analysis of global models" by Amanda C. Maycock et al.

Overview of study

The authors look at how the ozone (mainly stratospheric) changes in response to solar cycle activity (not including many secondly solar effects, such as high energy particles). The use CMIP5 and 6 data, and compare with observations, mainly characterised in part 1 of these papers. I find the study comprehensive in its analysis, but not particularly novel in terms of the science, and certain not novel in terms of increased scientific understanding. The study essentially regresses out the solar signal from ozone in climate models, which has been done to death. I appreciate a lot of work has gone in to applying it to a new data set, but I can see little advancement in scientific knowledge in what is done. The authors' final summary seems testament to this, where their conclusion is essentially 'we need more data'. The scientific analysis is far from rigorous as well, with statistical significance very rarely performed, in unclear for the figures where it has been done.

We thank the reviewer for his/her detailed comments on the manuscript, which we address below. While the reviewer has raised some criticisms, which we have addressed and which have helped to improve the manuscript, we firmly believe that the manuscript contains relevant new results that will be of interest to the broad atmospheric and solar research communities, and therefore that the manuscript warrants publication in ACP.

<u>Concerns (major)</u>

1) The novelty to the study. Many studies have done very similar things to this study. Some of these studies are cited in the main text, but it is often not clear to a reader who is unfamiliar with the literature just how similar these studies really are. In many cases reproducing very similar figures. The authors need to be clear up front what is new here, and attribute all the repeated information to the correct papers.

a) Our study compares the representation of the solar-ozone response (SOR) in models with interactive chemistry (CCMs) against the prescribed SOR in GCMs. To our knowledge such a comparison has not been been performed before and therefore comprises an important advance to the field. It is particularly relevant for putting into context recent multi-model studies (e.g. Mitchell et al., 2015; Hood et al., 2015) that include CCMs and/or GCMs. Our results show that the representation of the SOR is crucial (arguably more important than changes to the SSI forcing dataset -- see Matthes et al (2017)) for determining differences in modeled solar cycle responses between CMIP5 and CMIP6. We deem these to be sufficiently interesting and important conclusions to justify publication in ACP.

b) To clarify that multiple regression methods have been widely employed to extract solar cycle variations in ozone datasets before, we have added the following text at the start of Section 2.2:

"Multiple linear regression models have been used to analyse drivers of secular trends and variability in stratospheric ozone for many decades (see e.g. Staehelin et al., 2001 and references therein). In the context of extracting solar cycle variability from ozone timeseries, there is a long history of similar methods being applied to both satellite observations (e.g., Soukharev and Hood, 2006; Remsberg 2008; Tourpali et al 2007; Remsberg and Lingenfelser, 2010; Dhomse et al 2016; Lee and Smith, 2003; Lean 2014; Randel and Wu, 2007; Merkel et al 2011) and chemistry-climate models (Austin et al., 2008; Sekiyama et al., 2006; Lee and Smith, 2003; Egorova et al., 2004; Dhomse et al., 2011; Dhomse et al., 2016; Hood et al., 2015; SPARC CCMVal, 2010). Here we follow the methodology described by Maycock et al (2016), which is very similar to the methods described in those earlier studies."

We have also edited the Introduction to state an explicit set of novel objectives for the study:

"The objectives of this study are therefore:

- to provide an update to previous CCM studies by analysing the SOR in CCMI-1 models.
- to evaluate the SOR in three pre-calculated ozone databases for climate models from CMIP5, CMIP6 and Bodeker et al (2013).
- to compare the CCMs and ozone databases with satellite observations from Part I (Maycock et al, 2016).
- to perform atmospheric model experiments to quantify the impact of differences in the SOR between CMIP5 and CMIP6 on the simulated atmospheric response to the 11 year solar cycle.

Collectively these objectives provide a comprehensive assessment of the represention of the SOR in current CCMs and global climate models."

We have also added at appropriate points in the results section statements connecting our results to figures in earlier multi-model studies such as Austin et al (2008) and Hood et al (2015). While the CCMI-1 model analysis is an update on these earlier studies, the explicit comparison with pre-calculated ozone fields is new.

We hope that the reviewer agrees these changes adequately acknowledge the earlier work that our study builds on.

2) The statistical significance in this study. This is very poor, and often non-existent. The authors need to be clear about what their significance test is, what it is showing, and most importantly, they actually need to do some significance testing for most of the plots. In some plots, different significance tests will be needed for each panel. i.e. Figure 3, a different test will be needed for the individual model, as for the MMM. I am not even sure if the MMM has significance in the current study?

The reviewer is correct that in the original manuscript the MMM result in Figure 3i did not show any estimate of statistical significance. This has been added in the revised manuscript based on regions where the MMM response is smaller than ±2 standard deviations of the intermodel spread derived from Figures 3(a-h).

Hatching denoting regions where the central estimate of the regression coefficients is not statistically significant at the 95% confidence level has also been added to Figures 5 and 7.

We have added shading to Figure 9 denoting ± 2 standard deviations of the interannual variations in temperature over the 50 year experiments as an estimate of the 2.5-97.5% confidence intervals for the ECHAM6.3 modelled responses.

All figures in the revised manuscript (with the exception of the raw timeseries) therefore now include appropriate estimates of the statistical significance of the results.

3) Regression methodology. The methodology the authors use has no measure of uncertainty in the basis functions. This is a fundamental problem, because some of the basis functions have a good deal of uncertainty associated with them. The authors should add this uncertainty in to better reflect the uncertainty in the final result. The regression method they use is cited as from Maycock et al, but in reality, it probably has roots in far earlier solar-regression studies (see my first point, of giving due where it is deserved, even if methods/results are slightly different). I expect the authors arguments to not including uncertainty in the basis functions will be 'it has been done multiple times before', and cite a number of studies. However, this does not mean it is correct, unfortunately. I feel at some point, one of these studies need to include these basis uncertainties, at the very least to show that it doesn't make a difference (although I expect that it does).

We have followed the reviewer's suggestion to add greater historical context for the regression methodology employed in the study in Section 2.2 (see response to major point 1). We agree with the reviewer that there are limitations of multiple regression analysis, which we emphasise are not limited to examination of the SOR. However, it is difficult to conceive of other current approaches that would have significantly less limitations. We have added the following text to the manuscript at the end of Section 2.2: "It is a challenge in geophysical science to develop statistical methods to extract forced signals from complex timeseries. The implementation of multiple regression analysis as described above has a number of limitations, including (but not limited to): assumption that the input basis functions have zero uncertainty; difficulties in separating a signal from noise in a relatively short or sparse record (Damadeo et al 2014); and potential issues with degeneracy between basis functions (Chiodo et al 2014). These limitations should be kept in mind when examining detailed aspects of the results."

We inform the reviewer that there is now a dedicated working group within the SPARC SOLARIS-HEPPA activity that will perform a detailed comparison of statistical methods for analysing solar-climate signals with the eventual aim of providing some recommendations for best practices.

4) Anomalies. Often the authors use anomalies of variables, rather than the absolute variables. It would be good to see who real values of the data. I realise this can not always be done, but in some figures, for instance Figure 4, this would be very informative. Anomalies often make things look better!

Since the focus of our study is on quasi-decadal variability in ozone, we believe it makes sense to show anomalies from the long-term annual cycle, so that the vertical scale on the timeseries in Figures 2 and 4 can be sufficiently narrow that interannual to quasi-decadal variations are visible. However, to respond to the reviewer's request we have added figures to the Supplementary Material showing timeseries of absolute tropical ozone mixing ratios in the CCMI models (Figure S1) and in the climate model ozone databases (Figure S3).

5) There is a lot of focus on the CMIP6 ozone data set. But seems to be absolutely no citation to documentation on this data set. I note that the creators of the data set are not authors on this paper, and perhaps some of that lack of knowledge is reflected in the text. Is there a CMIP6 ozone paper coming out? Should this current paper be kept out of publication till that exists? This should certainly be true if there is any overlap.

Throughout this work we have liaised closely with the creators of the CMIP6 ozone database, led by Michaela Hegglin. We have sent Michaela the draft manuscript for

comment and she has even posted a comment on the discussion of this article in ACPD. At no point has it been indicated to us that our study should not be published. To the best of our knowledge the forthcoming publication in GMD describing the CMIP6 ozone database will not focus on the representation of the SOR, and thus we do not anticipate any significant overlap between the studies.

Two co-authors of our study (Dan Marsh and David Plummer) are the principal investigators of the CCMs (CMAM and CESM1(WACCM)) used to produce the CMIP6 ozone database. These co-authors have provided detailed information about the CCM simulations used to create the CMIP6 ozone database. The parts of this information that are particularly relevant to simulation of the SOR are described in Section 2.1.3 of the manuscript.

We remind the reviewer that the CMIP6 ozone database is publicly available and CMIP6 modellers are already implementing the dataset in their models: <u>https://esgf-node.llnl.gov/projects/input4mips</u>.

Concerns (minor)

1. Line 1: This alone would not fully capture the response 'fully capture' changed to 'to aid in capturing'

2. The SOR seems a little strange in this context, because it is not obvious (until later) that the SOR is not a 'set thing', it is only known within uncertainty bounds (and so different CCM give different SORs).

We use SOR throughout the manuscript for consistency with Part I. Here we have changed the text to say 'comparison of the representation of the solar-ozone response (SOR)....' to make clearer that this is something with variable representation across models.

3. Line 9: . . . ozone databases' – at this point it is not clear if the ozone data basis is the prescribed ozone, or simulated ozone from a CCMI.

Throughout the manuscript we distinguish between analysis of output from chemistryclimate models (in this case CCMI models), analysis of pre-calculated ozone databases for models without chemistry (which can be constructed from observations and/or CCMs), and analysis of ozone datasets (i.e. satellite observations taken from Part I). The use of 'database' in the manuscript is therefore solely reserved for pre-calculated ozone fields used in models without chemistry. We feel that the preceding sentence makes clear the distinction between the analysis of CCM results and of the pre-calculated ozone databases for CMIP5/CMIP6.

4. Line 11 Make clear that you refer to historical period ozone Time period of analysis added.

5. Line 13: weak compared with what? This clause has been removed.

6. Line 76 – a citation is really needed here (see major concern). The dataset is publicly available at: https://esgf-node.llnl.gov/projects/input4mips. This link has been added to the text.

7. Line 89 – what time frequency is this data? 'monthly mean' added

8. Line 94 'transferring' – please revise this word.

Text changed to 'may play a role in driving the 'top-down' mechanism for the solar cycle influence on high latitude regional surface climate (see e.g. Gray et al. (2010)).'

9. L124. Why only 1 ensemble member? Please repeat with all of them. You need to capture the uncertainty.

We have updated the analysis to use all available ensemble members for the models. See Table 1.

10. Page 5 (top): This seems very similar to Hood et al, please state that. We are unsure of what the reviewer is referring to as being similar to Hood et al (2015) and have therefore not changed the text.

11. Line 37: 5 x 5degree. This is not normal, why has the interpolation taken place? This is the resolution at which the SPARC/AC&C CMIP5 ozone database is provided: see Cionni et al. (2011) doi:10.5194/acp-11-11267-2011. This is because the historical part of the CMIP5 database was derived from satellite observations (SAGE I and II) that are available on a 5 degree grid. We have not performed any further interpolation.

12. Equ 1: Please cite where this came from originally.

Additional references have been added at the start of Section 2.2 that make reference to earlier work using similar methods, including the review of Staehelin et al (2001) that discuss the history of multiple regression methods for ozone trends.

13. Equ 1: Do the authors have any views on the breakdown between the long terms solar response, and the 11-year response?

We tested the sensitivity of our results to removing >11 year variability from the F10.7cm solar flux timeseries and found that removing the lower frequency solar variability had virtually no effect on the results. For simplicity we therefore did not perform any pre-filtering to the timeseries of the solar basis function.

14. Figure 1: Surely these QBO signals are just from one model? These will change. Yes, the QBO indices in Figure 1 are just an example based on the observed winds. This is now stated in the caption. The QBO indices for the models are calculated from the individual model wind fields as described in Section 2.2.

15. Line 218-219: 'better proxy' not convincing to me. Please cite a paper that compares these.

Floyd et al (2004, doi:10.1016/j.jastp.2004.07.013) show that F10.7cm and Mg-ii are correlated at >.95 for daily timeseries and >.99 for variability on timescales longer than several months. This reference has been added to the manuscript.

16. Line 220-225: This section on the volcanic signal is vague. The data sets the authors use are not long, in fact some figure just use \sim 30 years of data. Very short for regression. Volcanic signals will cause issues in the regression, and it is not clear the authors have dealt with the properly (nor in Part 1).

Following the reviewer's comment and after further discussions, in the revised manuscript we now adopt the approach of Maycock et al (2016) by removing data in the periods immediately following the three major tropical volcanic eruptions since 1960: Mt Agung, El Chichon and Mt Pinatubo. This is because the ozone response to volcanic eruptions is a non-linear function of chlorine amount and thus it is not appropriate to include a basis function for volcanic effects in the MLR model. The description in the Methods section has been updated to reflect this change.

17. Line 240. I think the authors need to show the autocorrelation plots to the reviewers, so we can assess this evidence. I agree they probably do not need to go in the main text.

Figure R2 below shows the e-folding time in months of the autocorrelation function (ACF) of the monthly regression residuals in the CCMI models. Areas where the e-folding time of the ACF is greater than 2 months are evident in all of the models in the mid and lower stratosphere and hence our choice to adopt an AR(2) model.



Figure R2: e-folding time [in months] of the autocorrelation function of the MLR residuals for each CCMI-1 model.

In testing the effects of the AR model choice, as requested by reviewer 1, we identified some sensitivity of the estimated SOR in the polar lowermost stratosphere, which may be related to the longer timescales of the ACF in that region in several models. The sensitivity to the AR model choice across the remainder of the stratosphere was small, which the exception of the tropical lower stratosphere in SOCOL, which is discussed in the revised text. Therefore to avoid giving potentially misleading information about the SOR in the polar lowermost stratosphere we have restricted the plots in the revised manuscript to a maximum pressure of 100 hPa.

18. Section 2.3: please describe more how this model fits into the wider models of CMIP5. Then we can assess suitability.

A detailed description of the model is given in Section 2.3 of the manuscript. The model has a well-resolved stratosphere (model lid height above 50 km) and simulates the major aspects of the stratospheric circulation e.g. sudden warmings, the QBO (see e.g. Charlton-Perez et al., 2013; Schmidt et al., 2013). The response to 11 year solar forcing

in the CMIP5 version of ECHAM has been shown to be comparable to other high-top stratosphere resolving CMIP5 models (Mitchell et al., 2015).

Since the model does not include interactive chemistry, it provides a suitable test-bed for quantifying the effects of the pre-calculated ozone databases for CMIP5 and CMIP6.

19. Line 255-260: This is worrying that the lower signal might not be so well captured. We are unsure what the reviewer means by 'lower signal' in this context. The impact of the short wavelength absorption below 200 nm that is not captured in the ECHAM6.3 radiative code is quantified by Sudhokolov et al (2014) – see lower right panel of their Figure 2. The underestimation of the solar max-min shortwave heating anomaly in the stratosphere is ~15-20%, but is much larger above 60 km and thus we restrict our analysis to the stratosphere (<50 km) where the errors in the model radiation code are smaller. The simulation of the atmospheric response to the solar cycle in the CMIP5 version of ECHAM (MPI-ESM) was comparable and indeed compared better with reanalysis data than several other high-top CMIP5 models (see e.g. Mitchell et al. (2015)).

20. Figure 2: Please just plot the SAGE2 and SBUV observations on this plot. Done.

21. Line 291-295: power spectra would be useful here.

A plot showing the power spectra of tropical ozone anomalies at 3 hPa (roughly at the maximum of the SOR) for the CCMI models has been added to the Supplement as Figure S2. A peak around the decadal timescale is evident in the models.

22. Line 350-358: This is an important point the authors make. Are you saying this is a drawback of the CMIP6 ozone data set? Please expand on your recommendations here.

In our view it would be undesirable for a climate model to impose a QBO-ozone signal that is out of phase with its dynamical QBO. The counter-case is a model that simulates a dynamical-QBO, but that does not include realistic feedbacks from ozone. In practice, true consistency can only really be achieved in CCMs, but it seems potentially more problematic to impose an erroneous QBO-ozone signal than to neglect it altogether. Thus CMIP6 modelling groups may choose to post-process the CMIP6 ozone database in order to remove, or change the phase, of the QBO-ozone signal it contains to be consistent with their model. We are not making a specific recommendation, as we have not tested the impact of the QBO-ozone coupling on the simulation of the QBO; however, we feel this is an important feature of the CMIP6 ozone database to point out as it differs from the approach used in CMIP5.

23. Figure 3: What is the significance test here? Does the MMM have significance? See response to major comment 2. For individuals models the significance test criterion identifies whether the magnitude of the regression coefficients is distinguishable from zero based on the 2.5-97.5% confidence interval. Statistical significance for the MMM has been added based on where the MMM signal is larger than ± 2 standard deviations of the intermodel spread.

24. Figure 3: Why are tropospheric values masked out? The focus of the study (and of Part I) is on the stratospheric solar-ozone response. Hence tropospheric values are not shown.

25. Figure 4. Colors very similar

We have changed the colours of lines in Figure 4 so that they are hopefully more distinguishable.

26. Figure 5. Is there any significance on here? At this point (analysis of figure 5-9), I do not believe it constructive to have an in-depth review, because the significance is mainly missing, or hard to understand. You are interpreting potentially small signals compared to the noise.

See response to major comment 2. Significance testing has been added to Figure 5 and other Figures throughout.

27. Line 438-440: I think this is wrong, the SSTs do not constrain the upper tropospheric temperatures this much!

Figure R3 below shows the tropical mean (30°N-30°S) temperature response in a set of climate model experiments in which an idealised +2% increase in TSI has been imposed but the SSTs kept fixed at climatology. These experiments are not part of this study, but serve as a useful illustration to test the reviewer's hypothesis. Each model experiment is run for 30 years and the differences are taken with respect to a baseline experiment with the same fixed SSTs but without the TSI perturbation. With fixed SSTs, the tropospheric temperature change due to increased TSI mainly comes from increased shortwave absorption by water vapour and warming over land areas due to altered surface shortwave fluxes. The average tropospheric warming is ~ 0.2 K in the experiments. Note that the imposed solar perturbation of +2% is approximately 20 times larger than the solar max-min perturbation imposed in ECHAM6.3 in Figure 9. Therefore, from a simple scaling of the response in these fixed SST experiments, one could expect the tropospheric temperature response in ECHAM6.3 (which also uses fixed SSTs) to be around 0.01 K, which is consistent with the results in Figure 9. Therefore we conclude that the SSTs do impose a strong clamp on the upper tropospheric temperatures and we have therefore not changed the text.



Figure R3: Difference in tropical averaged (30°N-30°S) temperature [K] in a set of climate model experiments forced with an idealised +2% TSI perturbation with fixed SSTs. The perturbation imposed here is approximately 20 times larger than that used in the ECHAM6.3 model in Figure 9 of the manuscript. With fixed SSTs, the tropospheric temperature changes to the imposed solar perturbation are relatively small. Figure credit: Dr Chris Smith (University of Leeds).