Response to Referee #1

This is another in a series of model simulation studies of the atmospheric response to 11-year solar forcing. It considers only one model (the UM-UKCA model) but does an extremely thorough and careful statistical analysis of the model simulations (three 45-year simulations) for comparison to a ∼ 38-year meteorological reanalysis data set (ERA-Interim). I wish I could say that this study sheds new light on how solar variability influences climate. Unfortunately, as discussed further below, despite the careful analysis of the model and observational data (combined with an excellent review and referencing of previous work), limitations of the model itself, which is representative of many existing chemistry-climate models, preclude any such progress. Revisions are requested in response to the following main comments and other more minor comments.

We thank the reviewer for their constructive comments. We reply to the specific comments below in blue.

We appreciate the reviewer's comments regarding the challenges of modelling the effects of the solar cycle variability on the atmosphere. However, we would like to stress that our primary motivation is to demonstrate the aspects of the solar cycle response that are robust and less robust to analysis methodology, thereby aiding interpretation of the solar cycle signals found in observational and modelling datasets. We use the new capability in UM-UKCA as a basis where careful analysis and testing is possible. One point to note in relation to some of the reviewer's comments is the large intrinsic uncertainty in the diagnosed signal of the 11-year solar cycle in observations/reanalysis, which can be understood from our study as being partly related to the large internal variability in the stratosphere, and partly related to issues associated with the datasets themselves (e.g. Mitchell et al., 2015b; Maycock et al., 2016). Hence, if the modelled best estimate response to solar forcing differs from the observations/reanalysis best estimate, this does not automatically mean the model is wrong as the reviewer suggests, as we must also consider the uncertainty associated with the two estimates.

Main comments:

(1) The abstract and summary (section 7) make little mention of the strong disagreements between the model-simulated responses of stratospheric ozone, temperature, and zonal wind to 11-year solar forcing and that derived from observations (here, SAGE II and ERA-Interim). These disagreements (e.g., Figures 3, 4, and 5 for the annual mean response for the 3 simulations combined together) are so strong that there is no chance that the model can simulate realistically the solar-induced climate response. Instead of drawing this obvious conclusion, the abstract and summary (and most of the last half of the manuscript) concentrate on investigating the roles of detection method and internal model variability in affecting (in relatively minor ways) the calculated model response to solar forcing. While the latter aspects are of academic interest, they pale in comparison to the overall failure of the model to simulate the stratospheric response and, hence, any derivative consequences for the troposphere. The reader is left with the false impression that the model does a reasonably good job of simulating the solar response and that any differences with observations can be attributed to uncertainties in reanalysis data sets and possible aliasing of observations by volcanic and ENSO forcing.

We first respond to the reviewer's assertion that there is a strong disagreement between the solar cycle signals diagnosed in UM-UKCA and observations/reanalysis. For example, in Fig. 2. the uncertainties in the best estimates of temperature and ozone responses in the tropics are overlapping throughout most of the stratosphere. The main exception to this is for temperature in the tropical upper stratosphere, where ERA-Interim shows a stronger peak temperature change. However, this is not found consistently across reanalysis datasets (Mitchell et al., 2015b), and it is known that uncertainties exists in reanalysis datasets in the upper stratosphere (and above) due to scarcity of long-term measurements (e.g. McLandress et al., 2014; Mitchell et al., 2015a; 2015b; Long et al., 2017). Thus, while we must of course make use of the available observational/reanalysis datasets, it should be recognised that these provide relatively weak constraints for models owing to the large uncertainty in the diagnosed solar cycle signals. This is the case for the ozone response (e.g. Dhomse et al., 2013; 2016; Maycock et al. 2016) and for the temperature and zonal wind response (Mitchell et al., 2015b). We therefore believe that our results in Sections 5-6 provide a valuable insight into the role of detection method and interannual variability for the detected solar cycle signal/response that has not been widely acknowledged in the previous literature.

We also wish to respond to the reviewer’s point that “investigating the roles of detection method and internal model variability in affecting (in relatively minor ways) the calculated model response to solar forcing.” While we show that the choice of analysis method has only a small effect on the annual mean ozone and temperature responses in the stratosphere, which is a useful result in itself, at high northern latitudes in winter the differences between the individual ensemble members are as large, or larger than, the differences between the ensemble mean response and ERAI-Interim. This finding alone emphasises the potential for internal atmospheric variability to affect solar signal detection in a system with relatively few degrees of freedom (i.e. covering only a few solar cycles).

Despite the above points, we accept that there are some differences between the solar signals in UM-UKCA and observational/reanalysis datasets which we did not try to hide. As suggested by the reviewer, we make it even clearer by adding the following paragraph to the abstract: “While many of the diagnosed signals of the solar cycle in UM-UKCA agree with those derived from observations/reanalysis within their respective uncertainties, there are some differences in magnitude, spatial structure and timing of the signals in ozone, temperature and zonal winds. This could be due to the large uncertainties in the observational estimates of the solar response and/or deficiencies in the model performance”.

(2) One fundamental problem with the model is its failure to adequately simulate the observationally estimated 11-year solar response of the upper stratosphere and lower mesosphere. Such a simulation is essential for initiating a strong zonal wind response near the time of winter solstice, which propagates downward and poleward later in the winter, ultimately leading to a tropospheric response (Kodera and Kuroda, 2002). The ozone response is too weak in the tropical upper stratosphere (dropping to 1 percent by 45 km; Figure 5a). Other models (see, e.g., Figure 1 of Hood et al., 2015) do better (about 2 percent at 45 km) but still fall short of that estimated from SAGE II data (about 3 percent). This shortfall of the simulated ozone response is at least partly responsible for the too-weak model temperature and zonal wind responses near the stratopause evident in Figures 3 and 4. At least some other climate models do a much better job of simulating the temperature response at these altitudes (see, e.g., Figures 4 and 5 of Mitchell et al., QJRMS, 2015) but I did not notice any mention of this in the text. Possible reasons for the poor ozone simulation are discussed only briefly in section 3.3.2 on p. 12. One factor that is not mentioned is the very coarse resolution of the 6 spectral bands evident in Table 1. The first spectral interval includes the entire UV region from 200 to 320 nm. As mentioned at the bottom of p. 9, the broad spectral band in the UV also has negative effects on the model's shortwave heating scheme and therefore its ability to simulate the full magnitude of the temperature response in the upper stratosphere and lower mesosphere. All of these problems combined together inevitably prevent the model from simulating realistically the downward propagating solar-induced dynamical signal. It is these deficiencies that should be emphasized in the abstract and summary sections for a fair assessment.

We first address the reviewer's comment: "The ozone response is too weak in the tropical upper stratosphere (dropping to 1 percent by 45 km; Figure 5a). Other models (see, e.g., Figure 1 of Hood et al., 2015) do better (about 2 percent at 45 km) but still fall short of that estimated from SAGE II data (about 3 percent)." We begin noting that while the reviewer's comment makes reference to the composite model results in Fig. 5a they do not mention that the corresponding MLR ozone response (Fig. 5b) at ~45 km is larger (~1.5%) and statistically significant than the composites. This indicates the MLR method is better suited for extracting the ozone response in this region. The apparent difference between the model and satellite observed best-estimate solar-ozone response is therefore smaller than suggested by the reviewer. We note that we have, however, already mentioned the differences between the model and SAGEII in the manuscript (see Section 3.3.2.). We would like to emphasise again here that there are large uncertainties in the observed solar-ozone response in the tropical upper stratosphere, and that the uncertainties in the model and satellite tropical ozone signals are overlapping (see Fig. 2c), which precludes the conclusion that the model ozone response is too weak.

Secondly we address the reviewer’s comment: “Possible reasons for the poor ozone simulation are discussed only briefly in section 3.3.2 on p. 12. One factor that is not mentioned is the very coarse resolution of the 6 spectral bands evident in Table 1.” We apologise for this confusion: the 6 spectral bands applies only for the model shortwave radiative transfer scheme (discussed further below). The Fast-JX photolysis scheme from UM-UKCA consists of a total of 18 bands over the spectral region 177-850 nm (see Section 2 of the manuscript). The ability of this 18-band version of Fast-JX to simulate the solar-ozone response was evaluated by Sukhodolov et al. (2016) against detailed line-by-line calculations. As noted in our manuscript, “Sukhodolov et al. (2016) found a small (~0.2-0.3 %) underestimation of the ozone response to the solar cycle in the upper stratosphere in the Fast-JX code used in UM-UKCA compared with line-by-line calculations; this gave rise to only a small temperature underestimation of ~0.05 K”. Hence the resolution of the photolysis scheme is unlikely to be a major contributor to any differences between the modelled and observed ozone response.

More minor comments:

(3) The model produces essentially no tropical lower stratospheric response of ozone and temperature, which contrasts with that derived from SAGE II and ERA-Interim data (Figures 3 and 5). This is probably mainly because of the weak modeled upper stratospheric response, which leads to, at most, a weak perturbation of the tropical upwelling rate via the Brewer-Dobson circulation. However, it may also be partly because the model does not have a coupled ocean; it only uses prescribed sea surface temperatures, etc. Recent work shows that there is strong coupling between the tropical lower stratosphere and troposphere, which will not be adequately simulated in a model with no coupled ocean. For example, observational analyses have recently shown that the stratospheric QBO influences tropical convection and the Madden-Julian oscillation (Yoo and Son, GRL, 2016; and references therein). If the same is true for solar forcing, positive feedbacks from the tropical tropospheric response may have the effect of amplifying the tropical lower stratospheric response in ways that would not be simulated in a model with no coupled ocean and no MJO. Please at least mention with references in the revised manuscript introduction this new evidence for coupling between the tropical lower stratosphere and tropical tropospheric convection.

We agree with the reviewer that feedbacks from the ocean can be important for the atmospheric response to the solar cycle. However, we note that our simulations are forced with prescribed SSTs from observations (HadISST, Rayner et al., 2003) and therefore implicitly include a solar cycle signal which may contribute to the simulated atmospheric response. This is explained in Section 3.2.3. To address the reviewer's request, we have added a sentence to the introduction regarding the possible role of atmosphere-ocean coupling for the tropical lower stratospheric response.

(4) With regard to the above noted evidence for an influence of the stratospheric QBO on the tropical troposphere, have the authors investigated whether such an influence can be simulated in the UM-UKCA model? Shouldn't such an investigation come first since many more QBO cycles are available for a robust statistical analysis? Presumably the model cannot simulate this since it has no coupled ocean and therefore, probably, no MJO. If not, then some discussion should be added to the manuscript about this deficiency of the model and whether it could affect the model's ability to simulate as well the 11-year solar forcing. Also, please mention in section 2.1.1 that the model is not able to simulate the MJO if this is the case.

See the response to (3).

(5) The font chosen for printing most of the manuscript text is difficult to read.

We are sorry to hear that the reviewer found this to be the case but, unfortunately, we had to use the required ACPD template for manuscript production. Therefore, the choice of font there was outside of our control.

(6) P. 5, section 2.1.2. Is the resolution of the model's SSI forcing daily or monthly? This is not clear from Figure 1 after 1979 where large fluctuations are present near solar maximum. In other words, does it only simulate the 11-year component of SSI variability or does it also simulate the 27-day component? The latter could have some non-negligible effects on the simulated 11-year atmospheric response since extrema of the 27-day cycle can differ significantly from the mean. While it is true that TSI and UV proxies such as F10.7 correlate well on the 11-year time scale as shown in Figure 1 (see text on p. 8, line 18), this is not at all true on the 27-day time scale. Also, F10.7 does not correlate adequately with the actual solar flux near 200 nm on the 27-day time scale. So, if daily resolution is used in the future, you would need to use the actual SSI near 200 nm as your solar variable.

The model is forced with yearly mean SSI (as stated in Section 2.1.2) and, therefore, only simulates the 11-year component of SSI variability.

(7) The monthly analyses for the Nov. to April season in section 4 are useful for showing that the model only simulates a strong zonal wind response in November centred near 60N latitude whereas observations indicate that the zonal wind response is initiated at much lower latitudes in the subtropics and continues on with downward and poleward propagation through the winter. The same high-latitude bias of the 11-year zonal wind response is seen in most or all climate models. No clear explanation for this bias has been advanced but it could be related to the ozone response, which is much larger in the SAGE II observations in the tropical upper stratosphere than it is in most models.

We first note that much of the downward and poleward propagation of anomalies found in reanalysis is not highly statistically significant (Mitchell et al., 2015b). Accordingly, as illustrated in Fig. R1 below, the errorbars associated with the NH winter UM-UKCA and ERAI zonal wind responses largely overlap:

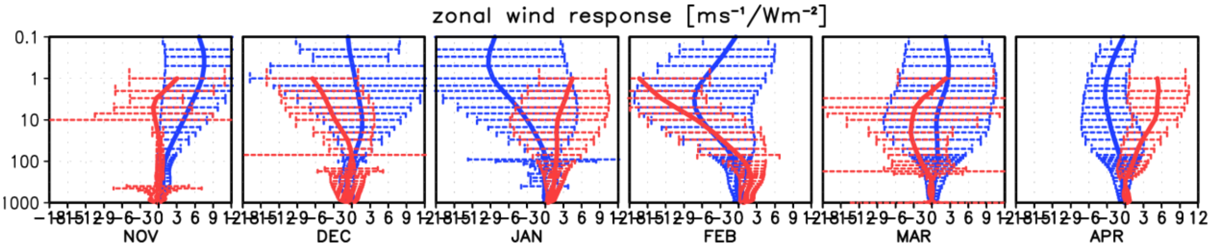

Fig. R1. November to April zonal mean zonal wind response [ms^{-1}/Wm^{-2}] at 60ºN diagnosed using MLR for UM-UKCA ensemble (blue) and ERA-I (red). The errorbars denote ±2 standard errors of the mean response.

We further note that the UM-UKCA simulations in Fig. 12 are 45 years long, which is longer than modern reanalyses covering the satellite era. Yet the individual ensemble members show diverse patterns in several winter months. It could be that the modelled signal is smaller than in the real world, as suggested by the reviewer, and/or that the modelled variability is too large, both of which would affect the signal-to-noise ratio and hence the detectability of the solar cycle response in the UM-UKCA ensemble. However, UM-UKCA captures the major components of stratospheric variability, including sudden stratospheric warmings and an internally-generated QBO, which makes the latter case less likely. It could be that the reanalysis record is too short to confidently diagnose the solar cycle signal in the high latitude stratospheric zonal winds, and that the reanalysis results in Section 6 are important for interpreting the solar cycle signal found in reanalysis datasets (which by definition represent only a single realisation of the real atmosphere currently for a few decades).

Nevertheless, we have mentioned in Section 4 the somewhat earlier timing of the NH dynamical response in UM-UKCA compared with the reanalysis best estimate and suggested this could potentially be related to the biases in the modelled climatological mean state.

(8) Most of the manuscript beginning with section 4.2 could be criticized as being either over-analysis of a deficient model or an investigation of issues that are mainly of academic interest. Can the authors find ways to delete or at least shorten some of this material?

As suggested by the reviewer, we have shortened Section 4.2.

Regarding Sections 5-6, we believe these sections provide a valuable insight into the role of analysis method and interannual variability for the detectability of the solar cycle within a 'perfect model' framework. Of course the model is not perfect and have some shortcomings as discussed above, but we still believe the results provide valuable information, e.g. about length of datasets required, ensemble sizes, that will be helpful for both future observational and model analysis studies. These issues have not yet been properly discussed in the literature, but are particularly relevant given the relatively short length of the satellite record compared with the quasi 11-year period of the solar cycle. The results also complement other important community efforts such as the SPARC SOLARIS-HEPPA working group on Methodological Analysis.

<u>References</u>:

See the references list in the manuscript, as well as:

Long, C. S., Fujiwara, M., Davis, S., Mitchell, D. M., and Wright, C. J.: Climatology and interannual variability of dynamical variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP), Atmos. Chem. Phys., 17, 14593-14629, https://doi.org/105194/acp-17-14593-2017, 2017.