

## Reply to the editor:

Dear Thomas Birner,

we would like to thank you for taking the time to handle and carefully review our manuscript. Below you will find our response to your comments in blue indicating the respective changes/improvements where necessary.

Dear Tobias Spiegl et al,

thank you for submitting detailed responses to all reviewer comments and for preparing a revised manuscript. My assessment is that your responses and revisions mostly address the reviewer comments well and that the paper is essentially acceptable for publication. I do have a small number of minor follow-up revisions that I think would improve the presentation of your material further. Once you've addressed these further revisions I expect to be able to accept your manuscript for publication in WCD.

One particular suggestion concerns clarifications and misunderstandings by the reviewers that in some cases were only answered in the responses to reviewers without any changes to the manuscript. I suggest to treat those reviewer remarks as representative of other potential readers from the community and to reconsider adding short clarification remarks in the text. Chances are that they will help other readers.

*Thank you for this important hint! As suggested, we include more information now for more clarity (please see below).*

Specifically, I have the following reviewer comments in mind:

Reviewer 1

- usage of "robust" - certainly helpful for other readers to state how you use this descriptor  
*As suggested, we added a comment to section 2.1.*

*Added: "In this manuscript we use the term "robust" if a signal of the same sign (e.g., the temperature response at the tropical stratopause) appears in the majority of our ensemble members."*

- remark on line 142 ("Why did not use all model years (1850-2005)?")  
*Included now.*

*Added: "Since especially the very early years are little reliable in observations and the model has been spun-up with a constant solar forcing, we focus on the period 1880 – 1999."*

- remark on Fig. 2 ("It seems the second warming is absent in your simulations.")  
*Added more discussion and Dhomse et al. (2022).*

*Added: "In our simulations we can't find the sometimes observed secondary peak in the temperature profile in the lower stratosphere. This secondary peak, however, can no longer be found even in most recent analysis of satellite data. Dhomse et al. (2022) suggest that the secondary peak (found in earlier studies) emerged most likely due to aliasing effects related to the Mount Pinatubo eruption in 1991 and probably was not a result of solar variability."*

- remark on Fig. 5 why there's no significant response in EM4

*The individual responses to the solar cycle (including statistical significance) is a result of different internal variability in the middle atmosphere in individual model runs. A more detailed paper on this will follow, where SOLCHECK simulations (including even more model years and different models) will be exploited (personal communication: Wenjuan Huo).*

*Reformulated: " However, even though exactly the same solar forcing has been applied in EM4 as in EM1, the initial temperature signal is not significant (most likely due to the individual internal variability in this ensemble member) and the dynamical response of EM4 in the extratropical regions looks very different."*

- question about "top-down" mechanism in observational data

*Included the relevant information.*

*Added: " Furthermore, and due to the lack of data covering the whole atmospheric domain over the complete historical period, it is not possible to connect the potential surface solar signals to the seasonality in the middle atmosphere. This applies to our and the original studies (e.g., Gray et al., 2013)."*

- question about EOF1 (NAO) in model (related to Fig. 7)

*Added a respective sentence.*

*Added: " Before continuing, we compared the spatial pattern of the EOF1 between the modelled and observed data and found good agreement with respect to the centers of action and overall characteristics (not shown)."*

Reviewer 2

- "vertical bands in Fig. 3"

*For more clarity we added "Please note, that the same SSN time series has been used for the correlation for all individual ensemble members, leading to a "vertical arrangement" of the data in the scatter plots shown in Figure 3."*

Some further comments:

Reviewer 1 question about effective degrees of freedom in the MLR analysis (due to band-pass etc) was not really addressed - please include relevant information in the text.

There is some misunderstanding! The data used for the MLR have not been bandpassed or comparable. We use the raw data here as in earlier papers. We only bandpass the NAO and solar timeseries to derive the decadal component before correlation and applying the ttest (as mentioned in the manuscript). This is the same strategy as in the original paper of Thieblemont et al. (2015).

*For more clarity we added "We like to note, that we use the raw (unfiltered) model output as input for our MLR analysis."*

The added text in response to Reviewer 2 remark about signal-to-noise paradox seems confusing to me. Specifically, isn't your statement "that relatively small changes in the external forcing will not lead to detectable changes in the variability spectrum in both the real climate system and model simulations" in conflict with the reviewer remark that "models might underestimate any externally forced signals in dynamical aspects of the climate system"? Please reformulate.

*Thank you for spotting this.*

*Reformulated: " We would finally like to note that the detection of a significant decadal solar impact on the NAO in winter in the MPI-ESM-HR climate model, as in other climate models, might to some degree suffer from the 'signal-to-noise paradox', i.e., a low strength of predictable signals vs. a relatively high level of agreement between modelled and observed variability of the atmospheric circulation, which is particularly evident in the climate variability of the Atlantic sector (Scaife and Smith, 2018)."*

In some Fig. captions you state the meaning of hatching to "mark the 95% level of significance" -- please use the more precise phrase "statistical significance" to avoid misunderstanding.

*Thanks and changed accordingly.*