Reply to reviewer 1:

Dear Wenjuan,

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

This manuscript presents a comprehensive evaluation of the 11-year solar cycle imprints from the middle atmosphere to the surface, especially the so-called "top-down" mechanism, based on 10-member CMIP5 historical simulations from the MiKlip (MPI-ESM-HR). Considering the diverse and debated conclusion on this topic in the existing studies, I think this paper is worth to be published after a minor review. The following are my comments on this manuscript:

Lines 17-18: "……on the North Atlantic Oscillation might lead to ……", I would suggest writing in a way like: "……on the sea level pressure in the North Atlantic might lead to a synchronization of the North Atlantic Oscillation via the 11-year solar cycle"

Rephrased: "More recently it has been proposed that the projection of possible top-down initiated decadal solar signals on surface climate might modulate and synchronize the North Atlantic Oscillation."

Lines 32-33: "……with opposite sign in response……" This also suggests the large internal variability could veil the possible solar imprints.

Indeed! We stress the potential impact of the internal variability in different atmospheric domains and ensemble members with respect to possible solar signals throughout the manuscript.

Added: "which might be a result of the large overall internal variability compensating rather small solar imprints."

Lines 93-104: all these studies are based on the same climate model, i.e., CESM-WACCM, of course, different versions, and different forcings. So even using the same climate model, the detected solar imprints or the interpreted results also could be controversial.

Thank you for pointing this out, since it makes our interpretation even stronger!

Added: "In the context of the most recent literature, it is difficult to understand why Chiodo et al. (2019) and Drews et al. (2022) arrive at a different assessment of the solar signal, even though the same model was used. This might point to the fact, that the complexity of the model is not the most relevant component in shaping potential surface solar signals, but rather the effects of internal variability in individual model runs and (to some degree) the applied analysis"

Lines 134-135: I guess the sea ice module is included in the MPIOM, right? Please add a few words here.

MPIOM includes sea ice dynamics, like any complex ocean model. The physical description of the sea ice model can be found in the cited literature. We think a more detailed description is not necessary or expedient.

Line 142: “We focus on the period 1880-1999” Why did not use all model years (1850-2005)?

We chose to analyze the 1880-1999 period, since the very early Hadley observations (both SSTs and MSLP) are only little reliable. Additionally, the MPI-ESM model was spun up using a constant solar forcing, which might influence the oceanic response prior to 1880.

Line 160: How did you perform the EOF analysis on the ensemble members and the ensemble mean? Did EOF on stitching data (i.e. 1200 model years) or did EOF on each member separately, but keeping the phases (sign in EOF1) the same as the observation? I ask this question because the EOFs could be opposite signs, if performing the EOF separately, need to double-check to make sure the patterns of all the members are in the same phase.

Thank you for pointing this out! We were aware of this problem and thus we stitched the input data and then performed the EOF analysis. This prevents our analysis from sporadic sign conversions.

Section 2 ：Significance level is shown in the lead-lag correlation and MLR analysis results, but the method description is missing. Considering after a band-pass filter, the effective degree of freedom of the time series will reduce a lot, so it’s necessary carefully calculate the EDF and describe the method in this section.

Correct! The description was not precise enough so far!

Added: “All correlation analyses have been performed by using the Python scipy.pearsonr function. Statistical significance of the correlations has been estimated by using a two-tailed Student’s t-test, as implemented in Python.”

By the way, how did you assess the robustness of the solar imprints?

We use the term robust in a non-statistical sense. We declare a result to be robust if, for instance, a comparable reaction in a meteorological variable (e.g., the stratopause temperature) is present in the majority of our ensemble members.

Line 174: “180 SSN”, here hard to understand this term for a reader not familiar with the MLR, please explain more.

We agree.

Added: “Based on this MLR analysis, we derived the model response to our chosen set of predictors, e.g., the temperature response per unit of the predictor (i.e., K per 1 SSN). To display the model response during solar maximum we scaled the coefficients to 180 SSN, which is a good approximation for a mean solar cycle amplitude between 1880 and 1999.”

Line 176: “the solar forcing lags the model data”, lags? for investigating the lagged response in climate, should solar forcing lead the model data, right? Did I misunderstand something?

It depends on the interpretation of leads and lags. As an example, if we would like to investigate the response of a meteorological variable one year after the solar maximum, we would analyze the model response in the year 1990 as a result of the solar forcing in the year 1989. Thus, the solar timeseries lags the model data by 1 year in this example. However, it is true that you could also say that we compare the model response to the solar forcing one year after the solar maximum. In this case we would analyze the lagged model response to the solar forcing.

The description, suggested by the reviewer is maybe more intuitive thus we changed the sentence.

Rephrased: "… the solar time series has been shifted in such a way that the model response lags the solar forcing by 1 to 4 years."

Lines 175-176: Are the other predictors keeping the same as the lag 0 when shifting the solar time series?

Yes. Otherwise, we would mix up signals.

Line 207: "period from 1850-1999" the period from 1850-2005?

Corrected. Replaced by "over the period 1880 – 1999"

Line 209: "pretty high minimum" do you mean compared to the SC19? I think the response of SWHR to the minimum of SC20 fits quite well, but a higher response to the maximum of SC20. So, please clarify your statement in this sentence.)

Thank you for pointing this out. The original sentence had been misplaced. It has been replaced by: "Only during SC20, the maximum SWHR response is higher than expected for that weak solar cycle."

Figure 1: a) and b) are missing in the figure. And I would expect a similar figure as Figure 1, but for temperature in the tropical mean (supplementary figure or panel into Figure 1), this figure will nicely show the sensitivity of the temperature response to the solar cycle amplitude. I also suggest plotting the original SSN time series and the scaled SSN time series used in MLR (180 SSN?), as a supplementary figure.

Unfortunately, only one EM was available with a diagnostic output for the swhr. Thus, producing a similar plot like Figure 2 is not possible. Saying this, we like to point out that one EM is enough to analyze the swhr, since they show the direct radiative response (and ozone is a prescribed variable in the MiKlip simulations) and the result will be identical among the different EMs. The sensitivity of the temperature response is shown and discussed in detail based on Figure 3.

Added: "a and b", and a Figure showing the original SSN time series (there is no scaled SSN time series available, since only the resulting coefficients are scaled not the SSN time series itself (please see comment above).

Line 219: "a robust" How did you define this?

Please see above.

Line 223: "a relative short time series of satellite…" maybe provide a specific time period? I think this period probably includes different solar cycles from the simulated time series. Repeating the model analysis in the same data period as the satellite would help to compare.

We like to show the model response over the complete period here, since it is important for the rest of the analysis. Since we don't know for sure the reason for these slightly higher values, we removed the statement.

Figure 2: A very brief caption. Please add more details. Like, is it a result of MLR? or a composite of annual mean zonal mean temperature in the solar maximum years? If it's the latter, how did you define the solar maximum years (maximum of SSN of each solar cycle I guess)? What does it look like in the solar minimum?

The result is based on MLR, as said in the caption, and thus the temperature response is scaled as described above. To derive the temperature profile during solar minimum, composite analysis could be an option. This, however, is not the scope of this manuscript.

It seems the second warming is absent in your simulations.

The long discussed secondary peak is, indeed, not present. However, we would like to point to most recent literature where the so-called secondary peak can no longer be found even in 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.

Section 4: a very long paragraph, that needs a break somewhere.

We created a new paragraph that starts with: "After having analyzed the variability of the TST, the PNJ and the 10 hPa zonal-mean zonal wind, we will now isolate potential solar signals by the aid of MLR."

Line 256: What does it imply, if the SSN is below the SC14 maximum? How strong the solar cycle could be if it's above the SC maximum? I think a figure of the original SSN time series will help to give the reader directly the impression.

Included. Please see comment above.

Line 266: please add a reference for this statement.

Added: "(Butchart, 2014).

Line 270: As the "dynamical response of the PNJ" is directly related to the meridional temperature gradient, I'm curious, can the solar-induced TST warming really increase the poleward temperature gradient? I think it's not the case for some ensemble members, at least for the EM4.

In our understanding the individual internal variability of each EM will determine if the (relatively weak) solar signals can shine through in the middle atmosphere or not. To clarify if this is the case, one would have to analyze the internal conditions of each individual ensemble member with respect to the polar vortex dynamics using very comprehensive analysis. In our paper, we would like to open the discussion and hope that future analysis will provide more insights on this.

Figure 4: same as Figure 2, the caption is too brief. Is it a result of MLR? MLR performed on ensemble mean or averaged regression coefficients of all the members?

Done.

Horizontal components of the E-P flux are hard to see in the bottom row of Figure 4, is it due to a very less reflection? Maybe scaling it to show more clearly.

The EP-flux vectors have already been scaled for a better visualization. The horizontal component is just much smaller than the vertical component. Scaling the horizontal component would lead to artificial results. Thus, we would like to refrain from doing this.

Line 339: "insulation" Do you mean "insolation"?

Corrected.

Lines 370-371: I think the temperature response in EM4 is very different from EM1, it's almost the opposite state in the Pole region.

Agreed and this, again, points to the importance of the internal variability in each model run.

About Figure 5, no significant temperature response in the tropical stratopause and no response of the lower mesospheric subtropical jet in EM4. Is the warming migration to the high latitude in EM1 due to solar insolation? If yes, why it disappears in EM4?

Agreed. If this signal would only be related to the march of the solar insolation, it should be present in all EMs. It is, however, possible that this signal is overprinted by the internal variability in this individual model run.

Line 384: “Drews et al., (2022)” => “Drews et al. (2022)”, they used a different climate model (CESM-WACCM)

Corrected. Yes, they used a different model but the statements in this publication were rather general in our opinion. And as pointed out by the reviewer above, even though using the same model it did not lead to comparable results.

Lines 413-415 and Figure 6: did a “top-down” mechanism show in observational data? An anomalous zonal mean zonal wind in the troposphere and surface in Feb?

It is not possible to answer this question since the results, depicted in Figure 6, include the analysis for the period 1880-1999 and middle atmosphere observations are not available over this complete period. This, however, is not only an obstacle in our manuscript but also in previous papers (e.g., Gray et al., 2013). Please also keep in mind that the observations only represent one ‘ensemble member’ which can lead to a misinterpretation (or overinterpretation) of potential solar signals.

Line 444-446: I guess that’s why we need some positive (slow) feedback from the ocean. But I’m not sure your model can simulate this.

We use one of the most widely used and complex general ocean circulation models, the MPIOM. So, we are rather confident about its general performance. Additionally, we don’t think that the ocean component is responsible but the fact that the solar signals are very small and (if at all) occur very irregularly, which can not lead to a regular phase relation. The ocean is just not sensitive enough to preserve these tiny signals over a longer period of time. Even very strong signals (e.g., anthropogenic or volcanic signals) are hard to detect in ocean dynamics. Furthermore, a detailed description of the underlying physics, forming the basis for the alleged lagged responses, is missing in literature.

By the way, if the “top-down” signals are not robust, how can we expect the surface response?

We can’t and this one of the main points of this paper. If the “solar signals” are already extremely small, irregular and hard to detect in the domain of the middle atmosphere (maybe due to the high internal variability there) the surface signals detected by MLR, composites or any other technique only capture the temporal state of the troposphere itself (sometimes NAO(+), sometimes NAO(-)) independent of the solar variability. In our opinion, future work should focus more on the inconsistencies in the middle atmosphere than directly discuss possible small surface signals or even a synchronization of hemispherical scale climate modes via the solar cycle. These points are given in the discussion section.

Figure 7: I suggest adding the surface wind or surface zonal wind in Figure 7. Same comments on the caption as above.

As the “problems” already start to appear in the middle atmosphere, including more tropospheric variables is not expedient in our opinion. The MSLP is pretty much the standard metric when discussing surface solar signals.

The spatial pattern of solar imprints in SLP may have different active centers from the EOF1 (NAO). Could you please compare them? Is the EOF1 (NAO) in your model the same as the observation?

Yes, we compared the spatial EOF1 pattern of the model and the observations. We include a figure below. The centers of action are comparable in both the model and the HadSLP2 data. In general, the pattern is more dominant in the model, however, please keep in mind that the observations only include 120 model years, while the model runs consist of 1200 years in total. A less dominant pattern might also be found in individual EMs.

Lines 458-463: this should be in the method section.

It is already included in the method section. We just remind the reader again here before discussing the results.

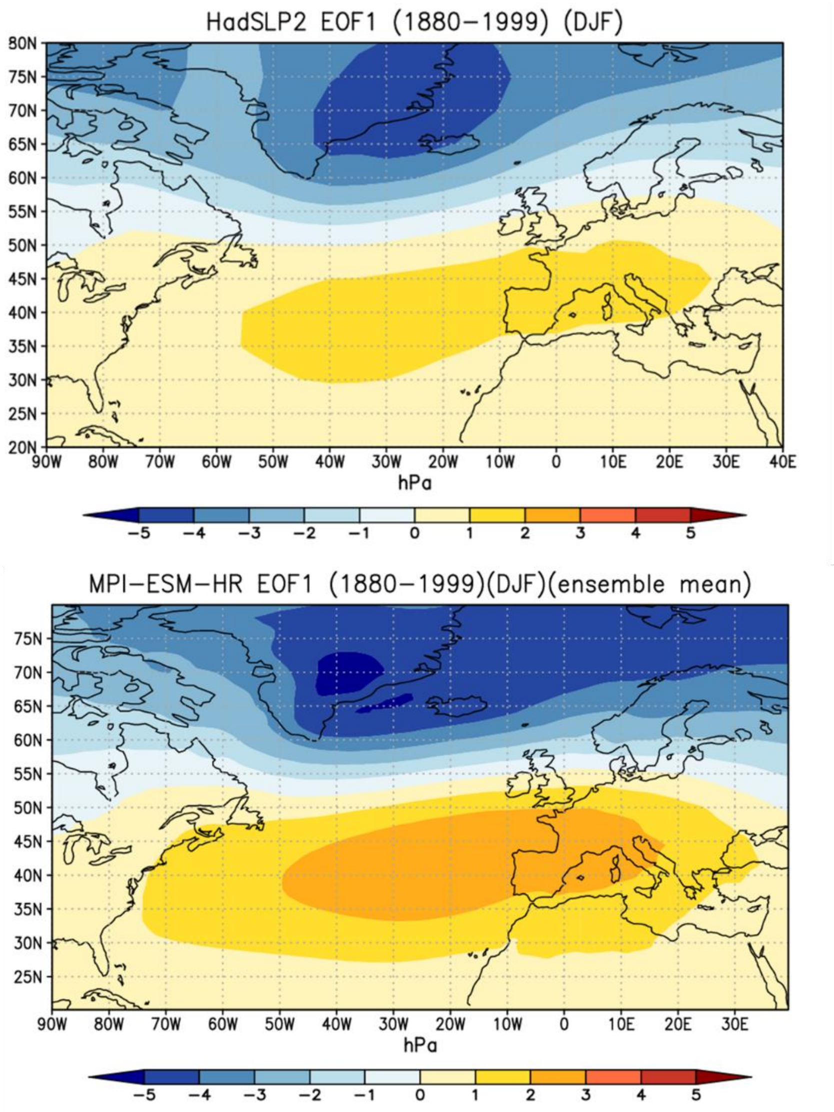

Figure for reviewer 1: EOF1 pattern (DJF) for HadSLP2 and MPI-ESM-HR data. Both calculated over the period 1880-1999

Literature

Dhomse, S. S., Chipperfield, M. P., Feng, W., Hossaini, R., Mann, G. W., Santee, M. L., & Weber, M. (2022). A single-peak-structured solar cycle signal in stratospheric ozone based on Microwave Limb Sounder observations and model simulations. *Atmospheric Chemistry and Physics*, *22*(2), 903-916.

Gray, L. J., Scaife, A. A., Mitchell, D. M., Osprey, S., Ineson, S., Hardiman, S., ... & Kodera, K. (2013). A lagged response to the 11 year solar cycle in observed winter Atlantic/European weather patterns. *Journal of Geophysical Research: Atmospheres*, *118*(24), 13-405.