Reply to
Review of “The Sun's Role for Decadal Climate Predictability in the North Atlantic” submitted by Drews et al. for publication in ACP

We thank the reviewer for the helpful comments.
Please find our replies and comments in italics.

The authors interpret two ten-member ensembles of CMIP-type “historical” simulations with respect to solar cycle signals in the boreal winter and in particular the North Atlantic surface climate. I think that simulations and their analysis are very appropriate for the purpose, writing and figures are in general very clear, such that I expect the manuscript to become publishable after consideration of a few issues that I will elaborate below. Overall I qualify the, in my view, necessary revisions as major although they may require relatively little effort. My issues with the current form of the manuscript are largely related not to the results themselves, but to their interpretation, discussion and framing.

The authors write that they intend “to partly rebut the conclusions” of a study by Chiodo et al. (2019) which speaks in its title of an “insignificant influence of the 11-year solar cycle on the North Atlantic Oscillation”. My understanding is that the authors don’t question the observational part of the analysis of that paper, but the analysis of observations which was done for DJF means, while in this manuscript, a surface signal significant at the 90% level is detected for February. The authors conclude “that it might be necessary to analyze monthly fields to capture a highly monthly varying signal”. Why speculate? I suggest to either analyze the Chiodo et al. model output for monthly signals and/or the new simulations for DJF signals.

Additional analysis for solar signals in DJF:
As suggested by the reviewer, we did additional analysis for zonal mean temperature, zonal wind (Fig. R1) and SLP (Fig. R2) as DJF means during the strong epoch. As shown in Fig. R1a, the solar signals in DJF mean zonal temperature are consistent with the monthly results in our paper (Fig. S4), primary warming in the tropical stratopause and secondary warming in the lower stratosphere (above the tropical tropopause). Similar westerly winds also can be found in the DJF mean (Fig. R1b) in the polar vortex region. However, only when using monthly data the top-down propagation and a significant surface signal can clearly be seen (Fig. S5).

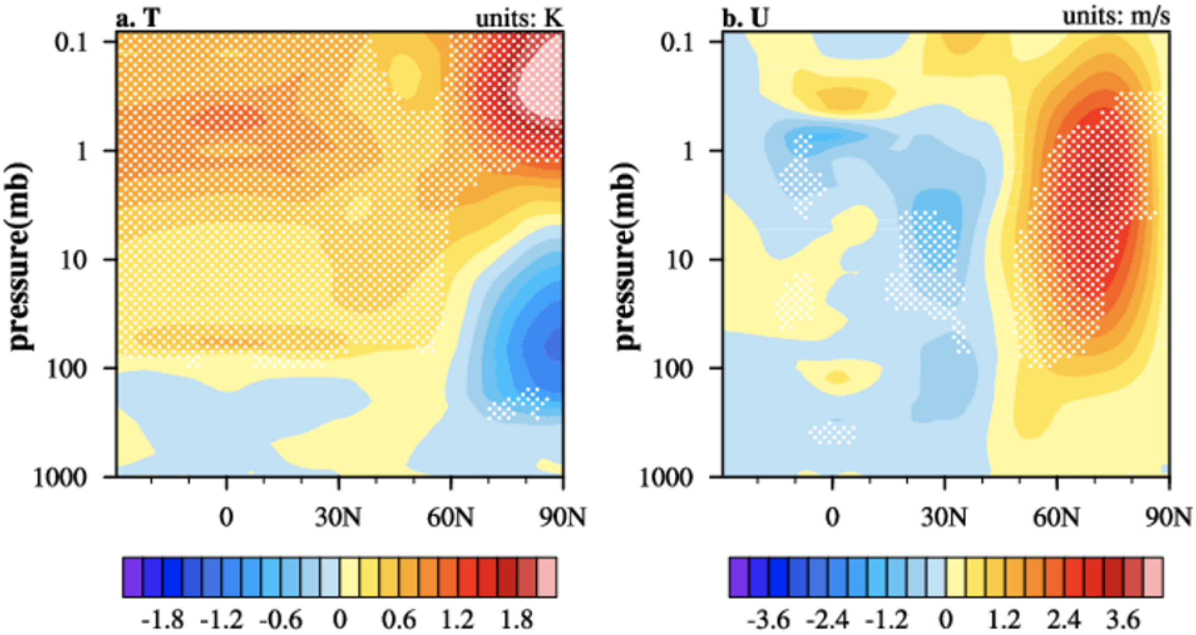


Fig. R1. *Composite differences between the solar maximum and minimum for (a) zonal mean temperatures (K) and (b) zonal mean zonal winds (m/s) in DJF mean during the strong epoch. Significance levels are indicated by white dots (95%) based on 1000-fold bootstrapping test.*

The North Atlantic surface signals for DJF mean during the strong epoch are shown in Fig. R2. There is a positive NAO pattern similar to that in February (Fig. 4a) but with smaller values. As shown in Fig. R1, the dipole zonal winds anomalies (easterly winds in the subtropics and westerly winds at high latitudes) in the troposphere are much stronger and significant in February, and smoothed out in the DJF mean. Accordingly, responses of the SLP and surface winds in the DJF mean are much weaker as compared to February.

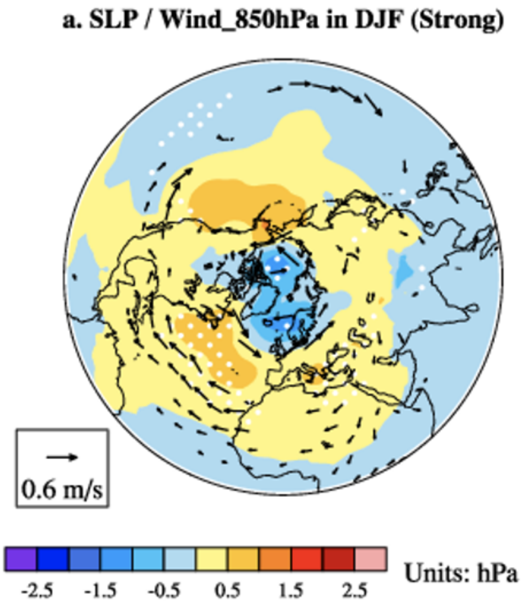


Fig. R2. *Composite differences between the solar maximum and minimum for SLP (contours) and wind at 850hPa (vectors) in DJF mean. Only those vectors where the zonal wind component is significant at the 90% level are shown. White dots indicate 90% statistically significant level based on 1000-fold bootstrapping test for SLP.*

We now added a sentence and the new figures referring to the new DJF mean analysis to the manuscript:

“Our results (see also Figs. S4-5) suggest that it might be necessary to analyze monthly fields to capture the top-down propagation of the solar-induced wind anomalies and surface signals. Analyzing our model using DJF means, we only find a very weak signal in SLP and no significant zonal mean zonal wind signal at the lower troposphere during the strong epoch either (Figs. S10-11).”

The authors mention further differences, as the missing low-frequency solar variation in the Chiodo et al. simulations. But why should this matter given that the authors make an effort to exclude this part of the solar signals from their analysis. More generally, it would be appropriate to discuss the agreement or disagreement of this study’s result with those of other papers more carefully. It seems the authors see their study in agreement of other papers they cite, e.g. in the Discussion “(Gray et al., 2010, 2013, 2016; Kodera, 2003; Kodera et al., 2016; Matthes et al., 2006)”, which they oppose to Chiodo et al.. However, several of those studies actually discuss North Atlantic surface signals only for DJF, so it seems to me that the actual results of this paper (no DJF signal) are rather similar to those of Chiodo et al., and it is mostly the framing where it differs.

We thank the reviewer for pointing out that it might be less the pure results, but rather the interpretation that has led to apparently different study results in the past. We now added a sentence to the end of the first paragraph in the Discussion section:

“However, as we show here, these seemingly discrepant results could be due to the analysis of DJF means in most studies, which likely are not sensitive enough to capture the signal reliably (see below).”

My second major point is that I think the authors do not adequately compare their model results to the available observation-based datasets. In the Discussion they mention that they “do not find a lagged NAO response... while the largest response in the observations appear at a lag of two years.” However, in the results section, when discussing Fig. 5c this is much less clearly presented. They write, e.g., that “running correlation ... begins to rise in the 1920’s both for the model and observations.” But this is not at all the case for the observations analyzed for lag zero. The following sentence is probably unintentionally unclear about “lag of two years” relating to model or observations or both. Why not, to compare apples and apples, also include model results for lag +2 in Fig. 5c? Moreover, it would be useful to identify also for other analyses if the observations are in the range of results provided by the individual ensembles, even if, unfortunately, one can neither conclude with certainty from such an agreement that the model is correct, nor that the observations provide a typical signal.

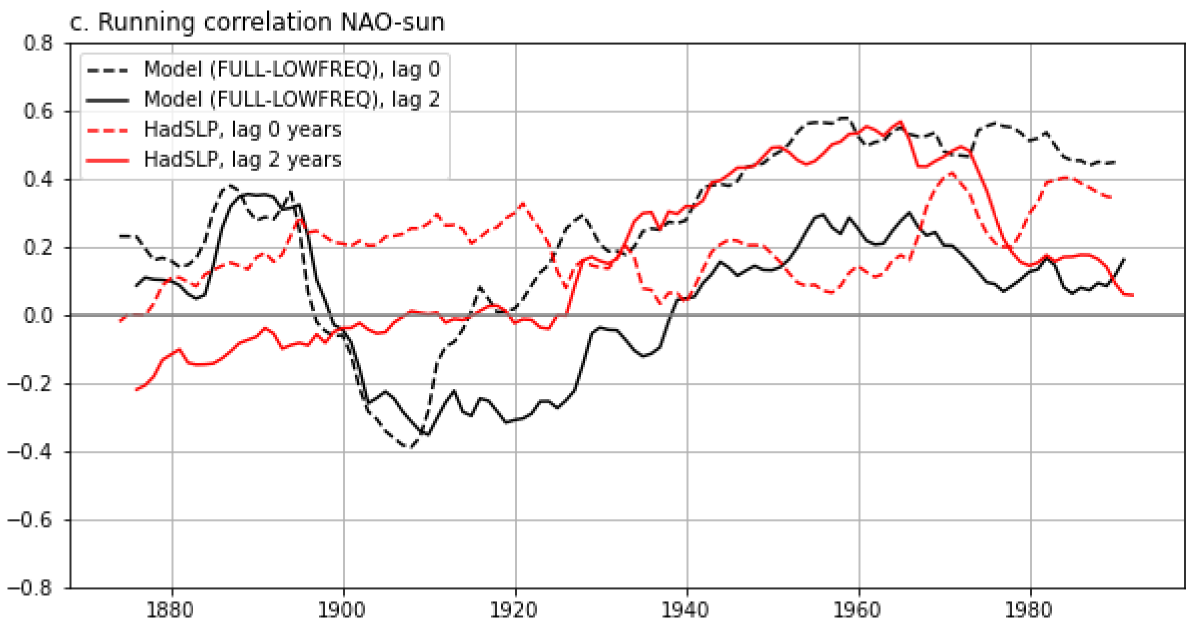
We agree that that sentence was a bit misleading. We now adjusted it to:

“Their running correlation for all overlapping 45-year windows is fluctuating in the earlier years but begins to rise in the 1920’s both for the model (at 0 lag) and observations (with a lag of 2 years) (Fig. 5c).”

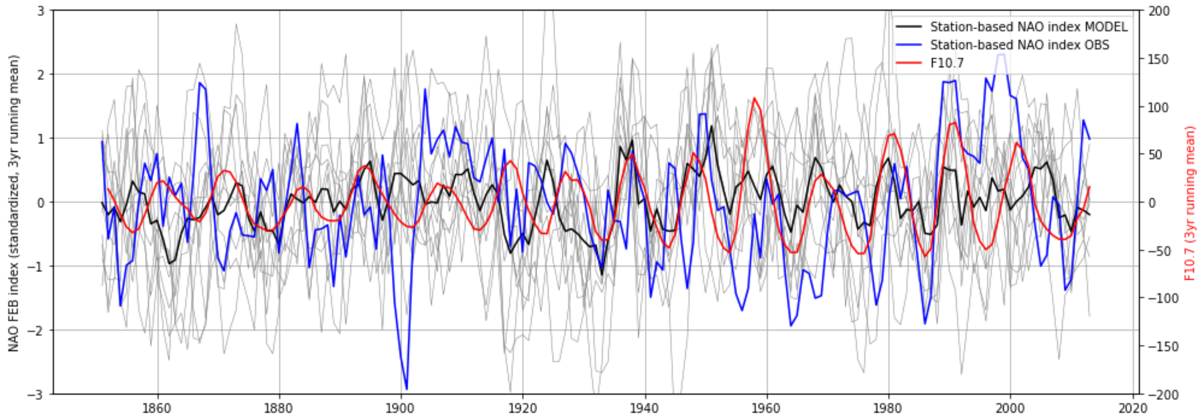
Furthermore, we would like to note that other studies did not find a lag in their model simulations either, and we therefore added a reference when addressing the lag differences in the Discussion:

“We do not find a lagged NAO response in our simulations (cf. Gray et al., 2013) [...]”

We attach here the figure of the running correlation of the model at lag 2, however, we would leave the figure in the manuscript as it is since it appears very full, and we clearly state that the model shows the strongest response at 0 lag.



Regarding the comparison of our analyses with observations, we add here the figure of the February NAO station-based index for all ensemble members and *observations, which shows that the observed index is well within the range of the model members.*



Furthermore, we use ERA5 to show zonal mean temperature and zonal mean zonal wind for comparison, and provide these figures below as well as in the new version of the supplement. Please note that the available data for zonal wind and temperature 1) are not as high as we plotted in our study, 2) it only covers three solar cycles (1979-2015), and 3) the solar forcing in ERA5 is a constant value.

The solar maximum years for these composite are: 1980, 1981, 1982, 1989, 1990, 1991, 2000, 2001, 2002.

The solar minimum years are: 1985, 1986, 1987, 1995, 1996, 1997, 2013, 2014, 2015.

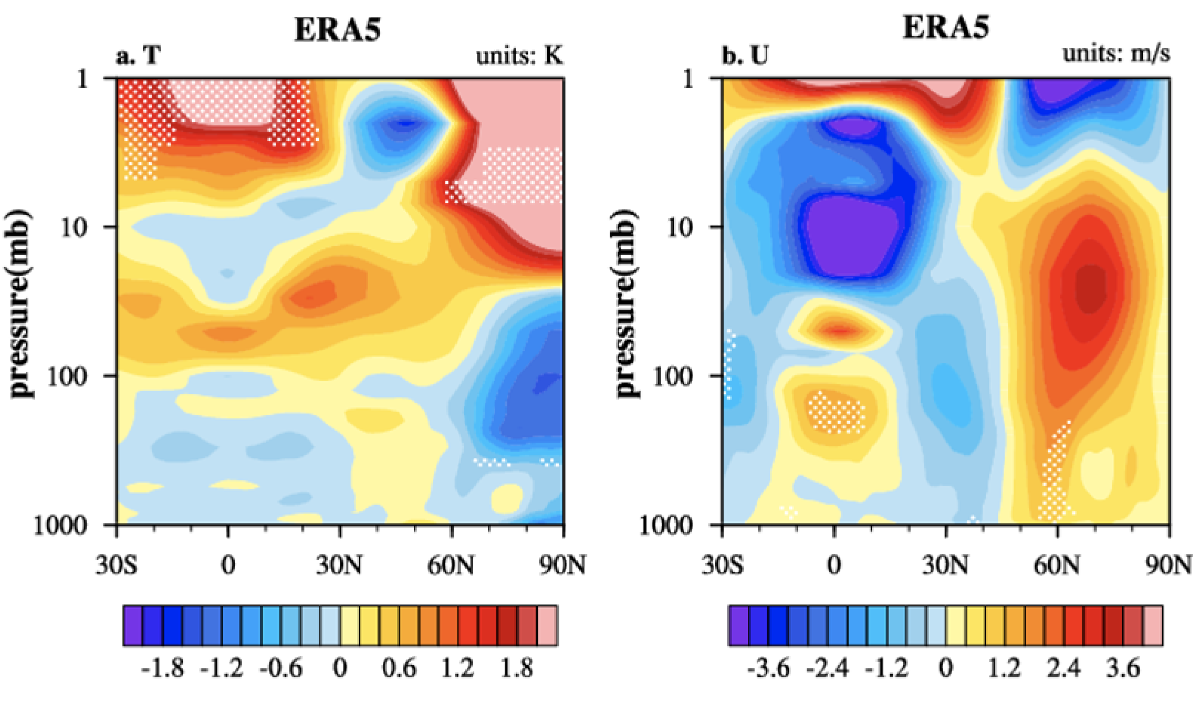


Fig. R3. *Same as Fig. R1, but for observational DJF mean ERA5 (1979-2015).*

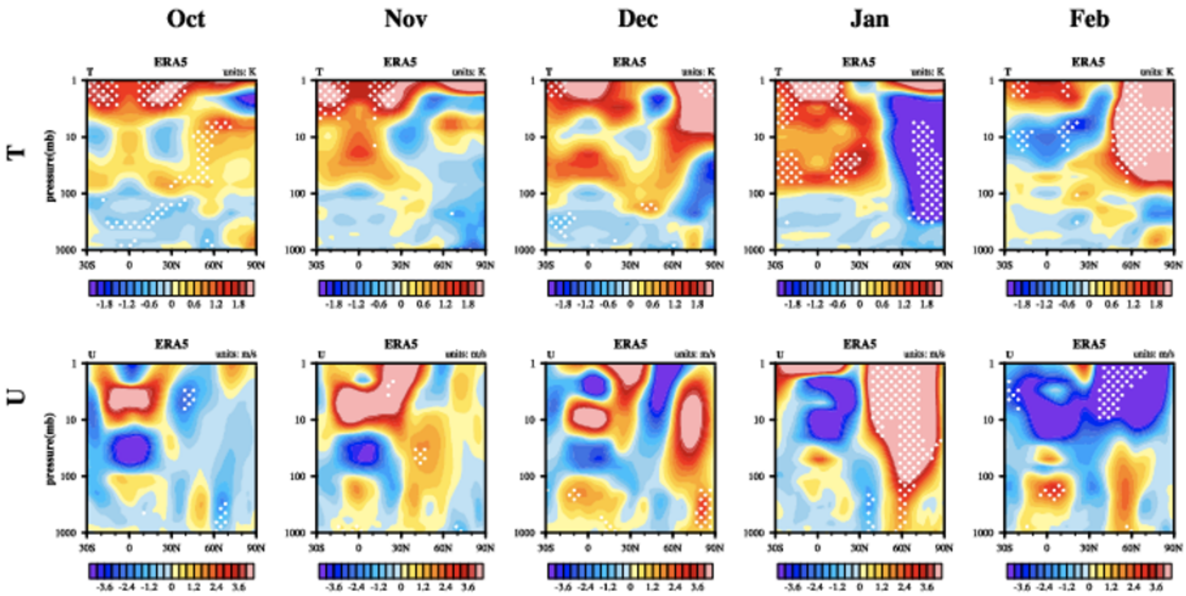


Fig. R4. *Same as Fig. R3, but for monthly ERA5 from Oct to Feb (1979-2015) (similar to Figures S4 and S5).*

Already in the abstract, the authors claim to “show that a strong solar cycle forcing organizes and synchronizes the decadal-scale component of the North Atlantic Oscillation”. They use this formulation a few more times. I claim they only show that this is true in their model reality. Of course, this is very useful, and the same mechanisms could also act in reality, but we can’t be sure. This in particular the case because of the involved non-linearities in the system, mentioned several times by the authors, and the apparently very different response to only slightly different forcings (compare weak and strong forcing epochs).

We agree with the reviewer as far as we consider our study alone. However, we would like to make the point here that our study’s result in this respect is not isolated but fits into a large collection of studies making use of different models and observational/reanalysis data. The understanding of the so-called top-down mechanism inducing a surface signal over the North Atlantic projecting onto the North Atlantic Oscillation is widely accepted in the research community dealing with this topic. We have to state that our study does not add much to the process understanding in this respect but we provide new insights as to why (when) this solar-related surface signal temporarily is hard to detect and what relevance this may have for (decadal) climate prediction efforts.

However, the reviewer is of course right, that we did not employ other models and no sophisticated analysis of observational/reanalysis data, hence we modified the respective statements in our manuscript to:

“The extratropical North Atlantic is a hotspot of solar cycle influence on climate predictability (Fig. 1a) where up to 25% of the decadal variability of winter surface air temperatures are explained by the solar cycle in our model.”

“We here show that in our model this “organization” depends on the solar cycle amplitude and it is large enough for a potential predictability variance fraction (ppvf) of up to 25% in the North Atlantic region.”

“We demonstrate that in our model the strength of the solar surface signal depends on the amplitude of the solar cycle.”

My last general point concerns the interpretation of the results with respect to decadal predictability. If this is supposed to be the main point, as the title suggests, I think this needs more careful and enhanced discussion. For instance, the authors claim that they use 8-year averaging because this is “a typical target of actual decadal prediction efforts”. However, in the reference they mention for this (Goddard et al., 2013) it is said that their choice of 1, 4, and 8 years may seem arbitrary, but was chosen to illustrate the effects of different temporal averaging. Many actual decadal prediction efforts show very weak skill beyond one or two years and certainly don’t concentrate on decadal (or 8-year) averages. So if the title should be kept, why not include a discussion of effects of different time averaging. Furthermore, in large parts of the analysis already different time-averaging is used and it is not mentioned how this relates to the main point of decadal predictability. It should also be mentioned that many of the hindcast systems used to evaluate decadal-scale forecast include observed solar irradiance. Moreover, forecasts of the strength of a solar cycle needed for actually deciding if a strong or weak solar forcing can be expected are far from being mature.

Thank you for this comment. It tackles a number of issues and for reasons of clarity we will sub-divide our response here into bullet points:

- *First, we would like to stress that the reviewer’s statement about the lack of skill in decadal prediction systems beyond the first two forecast years is only true when individual years (or even seasons) are considered. When multi-annual averages are considered - something that is recommended by Goddard et al. (2013) and usual practice in decadal prediction verification - the skill (as measured e.g. by the correlation) for near-surface temperature is typically the higher the longer the temporal averaging interval. That means typically that the skill for a year 2-9 prediction is higher than for the respective year 2-5 prediction. Of course this is not a result of higher predictive accuracy in the later forecast years but just an effect of eliminating more “noise” by averaging over a longer time. And this relates to the fact that the source of this skill (for temperature) is predominantly found in external forcing. That means that the predictions are highly skillful, it is just that the benefit from initialization is comparably limited. However, there are a few regions where a* *number of studies suggest predictability beyond 5 or even 10 years despite the absence of strong externally driven signals. The most prominent example for this is the North Atlantic (see, e.g., Christensen et al., 2020).*

- *Second, the reviewer is correct that there are quite different temporal averaging intervals used in decadal prediction studies and in recent years, most studies rather focus on lead times up to five years. However, there are still studies being published that deal with longer lead times and averaging periods, a very prominent example being the probably most-cited decadal prediction paper in recent years, that is the study of Smith et al. 2019 (analyzing year 2-9 predictions for the boreal summer season). Further examples are, e.g., Meehl et al., 2014; Kadow et al., 2016; Borchert et al., 2021; Hu & Zhou, 2021; Tian et al., 2021.*

- *Third, the reviewer correctly mentions that we use different time averaging in our analysis. We have to admit that this is not really done on purpose but rather the effect of bringing together skill-oriented analysis meant to match practices in climate prediction studies (as in Fig. 1, using the 8y running mean) and process-oriented analysis (as in Fig. 3 & 5) which were inspired by other studies such as Thieblemont et al. (2015). Our response to Reviewer 2 includes a sensitivity study to prove that the ppvf results presented in Fig. 1 are not sensitive to the averaging window. We therefore suggest adding the following sentence to the Methods section, subsection "Decadal potential predictability": "Other window lengths between 7 and 10 years were tested to exclude sensitivity of the results depending on the averaging period; the results were very similar for all window lengths." We just show this in a band around the currently used 8 year smoothing. If we would use a considerably shorter filtering window like 3 years, it is clear that some results would change. We would expect especially a much higher ratio of internal variability in the tropical Pacific, as a 3y running mean would not filter out ENSO-variability anymore. However, we consider a detailed discussion of the effect of different averaging windows beyond the scope of this paper.*

- *Fourth, the reviewer is correct that the observed solar forcing is included in decadal prediction systems. However, the story is quite complex. From our point of view it is highly questionable how much of the predictability indicated in our study can be exploited by today's decadal prediction systems. In order to model a proper representation of the top-down mechanism being associated with the surface climate signals seen in our study (and many others) over the North Atlantic requires interactive chemistry modeling (or at least an ozone forcing incorporating effects of the 11y solar cycle), a model top well above the stratopause (approx. 1 hPa), a sophisticated short-wave radiation scheme in the model, and the usage of spectral solar irradiance (SSI) as forcing dataset to account for the higher variability in the UV part of*

the solar spectrum compared to the visible and near-infrared part. The CESM1(WACCM) model used by us for this study fulfills these requirements, plus incorporating a parametrization of energetic particle precipitation and respective forcing. Most of today's decadal prediction systems fulfill only parts of these requirements. However, in this respect we notice a large development from CMIP5 - containing the first coordinated decadal prediction exercise - to CMIP6. The recommended forcing datasets (solar and ozone) for CMIP6 provide everything required above. This is why we speculate that some prediction systems in use for CMIP6-DCPP (and their counterparts used for CMIP6-historical) may be able to model some representation of the top-down mechanism, particularly those with a model top beyond 1 hPa, a state-of-the-art short-wave radiation scheme, and actually using the transient SSI-forcing as well as the ozone forcing for CMIP6. The study of Borchert et al. (2021), which we cite in the discussion of our manuscript, points out that the multi-model ensemble of CMIP6 simulations exhibits a response to external forcings that better matches the observed temperature evolution particularly in the North Atlantic and that this can be attributed primarily to the response to natural external forcings such as volcanic aerosols and solar forcing. A further distinction was not possible with the experiments analysed by Borchert et al.
However, there are still a number of models used in CMIP6 and DCPP that are far from fulfilling the requirements to represent the top-down mechanism. Some still use the total solar irradiance (TSI) only as solar forcing, scaling temporal TSI variability homogeneously over all parts of the spectrum which leads to unrealistically high variability in the visible and near-infrared and too low variability in the UV part of the spectrum. The result is a dampening of any potential top-down mechanism in these models.
We may inform the reviewer about the fact that the authors participate in the ongoing research project "SOLCHECK - Solar contribution to climate change on decadal to centennial timescales" - funded by the German Federal Ministry for Education and Research - with one of the main aims to particularly quantify the solar contribution to prediction skill in a state-of-the-art decadal climate prediction system (based on MPI-ESM).

- *Fifth, the reviewer is correct that the prediction of the 11y solar cycle itself is still a scientific challenge but positive prospects are given (see e.g. Petrovay, 2020). Furthermore, predictions of the progression of a cycle that already started are feasible within reasonable error margins and are actually produced by, e.g., the Solar Cycle Prediction Panel representing NOAA, NASA and the International Space Environmental Services (ISES), see https://www.swpc.noaa.gov/products/solar-cycle-progression. Therefore, we argue that the incorporation of predicted solar variability in quasi-operational decadal climate prediction may potentially be useful. Additionally, we consider the knowledge about a solar contribution to climate predictability valuable* *despite limitations of solar cycle prediction skill. This knowledge improves our overall understanding of climate predictability and shows ways to potentially improve decadal climate prediction, depending on the actual skill of (future) solar cycle prediction.*

Christensen, H. M., Berner, J., & Yeager, S. (2020): The Value of Initialization on Decadal Timescales: State-Dependent Predictability in the CESM Decadal Prediction Large Ensemble, Journal of Climate, 33(17), 7353-7370, DOI: https://doi.org/10.1175/JCLI-D-19-0571.1

Petrovay, K. (2020): Solar cycle prediction. Living Rev Sol Phys 17, 2, DOI: https://doi.org/10.1007/s41116-020-0022-z

Borchert et al.: Skillful decadal prediction of unforced southern European summer temperature variations. Environ. Res. Lett., 16, 104017 (2021)

Hu, S. & Zhou, T.: Skillful prediction of summer rainfall in the Tibetan Plateau on multiyear time scales. Science Advances, 7(24), eabf9395 (2021)

Kadow, C. et al.: Evaluation of forecasts by accuracy and spread in the MiKlip decadal climate prediction system. Meteorol. Z., 25, 631–643 (2016)

Meehl, G.A. et al: Decadal Climate Prediction: An Update from the Trenches, Bulletin of the American Meteorological Society, 95(2), 243-267 (2014)

Smith D.M. et al.: Robust skill of decadal climate predictions. npj Climate and Atmospheric Science, 2, 13 (2019)

Tian, T., et al.: Benefits of sea ice initialization for the interannual-to-decadal climate prediction skill in the Arctic in EC-Earth3, Geosci. Model Dev., 14, 4283–4305 (2021)

I will list a few more small issues in the following:

L11: “a systematic detection of solar-induced signals at the surface and the Sun's contribution to decadal climate predictability is still missing” Not clear what the authors want to say, here. Do they want to announce such a systematic detection in this paper? Certainly not, because they only do simulations. What would be a systematic detection? And is it at all possible with the available data?

We agree that this sentence starts a little unclear. We suggest changing it and writing the following: “Despite several studies on decadal-scale solar influence on climate, a systematic analysis of the Sun's contribution to decadal surface climate predictability is still missing.” We would like to stress in that context (and refer additionally to our response to major comment 6 of Reviewer 2) that “predictability” is *just a theoretical concept and commonly estimated based on model simulations alone.*

L29: "forecast skill for several years (...) beyond the externally forced climate response (Smith et al., 2019)" I don't think this is an appropriate interpretation of the reference. Smith et al. are actually much more careful in the interpretation of their results.

We are not entirely sure in how far the reviewer considers our reference here not being appropriate. Smith et al. (2019) claim that several other studies found a rather small benefit from initialization (i.e., beyond externally forced climate responses) mainly because their analysis approach and significance testing was not optimal. They further present a new approach assessing the impact of initialization in the residual space after subtracting the ensemble mean of uninitialized simulations from the initialized hindcasts as well as the observations. By doing so they find statistically significant benefits from initialization for a number of larger regions, e.g., over large parts of the Atlantic Ocean, Europe and parts of Africa, the Indian Ocean, Eastern Asia and parts of the Pacific, close to the Kuroshio (Extension). One issue that additionally needs to be considered though is that Smith et al. (2019) use a large multi-model ensemble of unprecedented size in this context. We suggest to rephrase the sentence in our manuscript to: "These prediction systems show forecast skill for several years (Bellucci et al., 2015; Yeager and Robson, 2017) and at least for large multi-model ensemble systems a robust benefit from initialization that goes beyond the externally forced climate response can be shown for number of regions globally (Smith et al., 2019)."

L53: The sentence starting here is one of the examples where the remark that this is a result from a simulation is crucially missing.

Agreed. We added "in our model" to this sentence.

S2: Information is missing on which simulations for which ensemble size are analyzed.

Agreed. We added the table with the models and ensemble sizes to the Supplement.

L94ff I guess correlation coefficients given here are only for a specific month. They seem to support a strong epoch-high correlation story, but numbers for wind in December, e.g. would look very different.

We now mention that it is December zonal mean zonal wind: "The ensemble mean zonal wind (here: December) gets more organized and in phase with the solar forcing during the strong epoch [...]"

To be specific, these are all correlation coefficients for DJF mean and December. "Var." is the variance fraction of the solar-induced changes compared to the magnitude of internal variability. "R" is the correlation coefficient with the solar forcing index.

	Weak epoch	***Strong epoch***		***Weak epoch***	***Strong epoch***
T_DJF	*Var. = 31%*	*Var. = 69%*	***U_DJF***	*Var. = 22%*	*Var. = 23%*
	R = 0.55	*R = 0.72*		*R = -0.12*	*R = 0.36*
T_Dec	*Var. = 31%*	*Var. = 37%*	***U_Dec***	*Var. = 18%*	*Var. = 24%*
	R = 0.21	*R = 0.58*		*R = -0.14*	*R = 0.33*

L107 "Synchronization" of what?

We now added "of the decadal NAO phase" to the previous sentence and hope this makes it clearer: "We find the "typical" downward propagation of zonal wind anomalies in later winter (Fig. S5) and a synchronization of the decadal NAO phase of the ensemble members (Fig. 5)."

L151 "We here show ..." Another case where the authors should mention that this refers to model reality.

Agreed and "in our model" added.

L188 I think that good studies don't necessitate such "first time" claims but results speak for themselves. Moreover, with model simulations this problem in observations can't be overcome.

We exchanged this by "With this unique dataset, ...".

L195 “The solar cycle enforces the NAO phase.” Even in these simulations, solar cycle forcing just changes the probability of occurrence of some phase.

We modified this sentence and it now reads: “This means the solar cycle enhances the probability of a specific decadal NAO phase if the solar forcing is strong enough.”