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

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. 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.

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.

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).

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.

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?

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.

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

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

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.

L107 "Synchronization" of what?

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

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.

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