Atmospheric Chemistry and Physics Discussions

Atmos. Chem. Phys. Discuss., referee comment RC2
https://doi.org/10.5194/acp-2021-241-RC2, 2021


# Comment on acp-2021-241

Anonymous Referee #2

Referee comment on "The Sun's Role for Decadal Climate Predictability in the North Atlantic" by Annika Drews et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2021-241-RC2, 2021

The paper reports on the impact of the solar cycle on decadal predictability of the NAO in the WACCM chemistry climate model. The paper claims that the solar cycle "organizes" and "synchronizes" the decadal-scale component of the NAO. Based on these results, the paper concludes that the solar cycle substantially contributes to "potential predictability" of up to 25% in the North Atlantic, but that the solar influence is limited to decades with sufficiently strong solar cycle amplitudes, such as the second half of the 20th century. The subject of the paper is of relevance and interest for the readership at Atmospheric Chemistry and Physics, and the paper is well written. However, the evidence provided in support of the main two claims (1. the contribution of the solar cycle to predictability, and 2. the impact of the solar cycle amplitude) is not convincing, due to the convoluted and not sufficiently justified methods. Just naming a few examples: 1. the use of ppvf technique to quantify the "predictability", 2. the use of 90% significance level instead of the more standard 95% level, 3. the smoothing of the data, 4. the low and regionally limited significance of the results, etc... Hence, the implications of this paper concerning the potential prediction skill from the solar cycle are over-stated. The authors need to substantially revise some of these claims, perhaps toning them down and/or provide more convincing evidence in support of a robust and detectable solar signal. I also find the discussion rather biased at times. As examples: 1. the signals over the Atlantic are rather regionally very limited and yet the paper makes a big deal out of the small solar signals, 2. (1000) one THOUSAND years of model data are needed to get a "robust" solar signal (this amount of data means the signal can hardly be useful for decadal prediction...). Hence, I cannot recommend publication in the present form. Extensive revisions are needed, as detailed below.

MAJOR POINTS

1. The authors "disentangle" the solar-induced climate response from internal variability, using convoluted and not fully justified statistical methods. First, they show the "potential predictability" (Fig.1a) by using the ppvf technique on the **difference** between FULL and LOWFREQ, divided by the variance in FULL. Further, they smooth the data using an 8-year window, citing Goddard et al., to justify such choice (a paper which by the way does in no

way suggest one should use 8 years). This metric does not effectively show the impact of the solar cycle on the variance itself (which is what the authors are after) but rather the relative impact of the solar signal over that of the long-term trends in the solar forcing. A much easier (and more convincing) metric would be the ratio between the variance in FULL and LOWFREQ. Also, using an 8-year smoothing window to get an 11-year signal seems a very unwise way to filter the data. Why exactly are 8 years used? I suspect the results are sensitive to the window length used to smooth the data, but maybe the authors can prove me wrong.

2. Along the same lines, another line of evidence used in the paper to show a "solar signal" is the running mean correlation over time and against the solar cycle amplitude (Figs.5c-d). But instead of a canonical running mean correlation, we are seeing the correlation (against the solar cycle) of the **difference** between two (independent) ensembles! I find this an utterly confusing and strange metric. Why not simply looking at the running mean correlation itself, rather than the difference of two ensembles? I get it that there are other forcings at work too in the FULL ensemble and that the authors wish to extract the solar cycle component, but the solar signal should emerge from the (forced and unforced) noise... if it's of use for decadal prediction - in the real world, multiple forcings are at work and not only the solar cycle. Looking at the correlation of the ensemble mean itself would highlight how the NAO itself correlates with the solar cycle, rather than its 'solar derived' component (which is supposed to be represented by the FULL-LOWFREQ difference). Physically, it makes little sense to look at the correlation of the difference between two ensembles, if what we're after is quantifying how the solar cycle influences a specific variable.

3. The "emergence" of the solar signal in the strong solar cycle epoch is shown in Fig.5c, but the model vs observations comparison is intrisically flawed in this figure. First, a different lag is chosen for the model (lag 0) and the obs (lag 0 and 2). The model at 0 lag shows better agreement with the lag 2 in OBS data, so the lag 2 in the model should be compared with the lag 2 in the OBS. Moreover, the observational data (which is a "pure correlation") does not show the equivalent of the FULL-LOWFREQ difference but rather of equivalent of FULL alone, since it contains all the observed forcings... so this is not an apple vs apple comparison! This would be another argument in favor of using the actual running mean correlation in the FULL ensemble rather than of the FULL-LOW difference to show a detectable solar "impact". You could compare the running mean correlation of FULL and should be able to show that it's higher than in LOWFREQ to convincingly demonstrate that the solar signal "emerges" from internal variability (which is the claim of the paper).

4. The paper claims that the strong solar cycles organize and even synchronize the decadal-scale component of the NAO. This is an over-statement, which is not sufficiently supported by the evidence provided in the paper. Rather, the paper shows a small influence of the solar cycle over very limited portions of the North Atlantic (Fig.1a) and not over Europe. ALso, the solar influence is 2x smaller than the "forced" component (Fig.1b) and of the internal variability (Fig.1c). Can we deem such this a "useful" source of skill, considering that 1000 years of data are needed? Also, the method to extract the solar signal (see major comment 2. above) seems a bit ad-hoc rather than a robust and critical assessment.

5. The paper essentially rebuts another paper on this subject (Chiodo et al., 2019), but fails to discuss the reasons for the inconsistency. Chiodo et al. used the **same** climate model as this paper (WACCM) and a solar forcing which would qualify as "**strong solar cycle**" forcing **throughout 500 years** (i.e. repetition of the "strong" cycles 19-23). And yet, they found a signal which is time-dependent, much in the same way as Fig.5c of this paper. Hence, the solar cycle amplitude argument does not seem to hold in this study. I urge the authors to more explicitly state this inconsistency throughout the paper (in the

abstract AND conclusions) and discuss possible reasons for disagreement, rather than "brushing off" any evidence against their claim, as they do e.g. in L180.

6. The paper argues about enhanced "predictability", but does not provide actual metrics of enhanced skill scores, when the solar cycle is included in the model predictions. Could the authors show that the model's ACC score values increase during the strong solar cycle epochs? If not, I am afraid that any claims about predictability are not supported and remain pure conjectures.

SPECIFIC ISSUES

L37 this is an over-statement. Given the problems outlined above, I find it hard to believe this is "robust evidence of solar influence". The solar influence is still minimal, so at very least, change this to "albeit small compared to internal variability" or something similar.

L40 "relative to other external forcings" --> incorrect statement - other forcings have not been quantified, as you only compare solar against anything else.

L41 realistic solar forcing --> whether it's realistic is quite debatable, as solar forcing still a reconstruction using statistical models, so please remove the "realistic" word

L50 extracting an 11-yr signal using an 8-yr smoothing window is risky, as the window is close to the frequency you are interested in. Further, the paper by Goddard et al does not justify the use of 8-yr smoothing for the study of decadal signals. Are the results sensitive to this "smoothing"?

L54-55 where is the 25% number coming from? Fig.1 shows signals over quite remote sections of the N.Atlantic rather than Europe itself, and they are a small fraction of internal & forced variances. In any case, results do not seem to support this statement, as the authors haven't consistently shown that solar explains 25% of the decadal variance, but rather of a convoluted metric for the solar signal itself (i.e. the difference in FULL-LOW rather than FULL itself). I would urge authors to clarify how this number is obtained, or tone down this statement.

55-56 isn't this sentence in complete contradiction with what is stated in previous sentence? If this region is low in terms of predictability, then how can one say solar cycle influence on climate predictasbility ? How is the significance at all quantified?

Fig.1: how valid is it to disentangle the solar signal using the difference FULL - LOWFREQ? The authors are using a linear estimation for something which is intrinsically nonlinear, as they state. Can we also assume variances in FULL and LOWFREQ are really the same to allow this quantification, or do they change?

To show the impact of the solar cycle on the decadal variability, the authors should rather show the ratios of the variance in LOW vs FULL - this would be easier to interpret and also more convincing evidence for a solar impact rather than the convoluted metric used here var(FULL-LOW)/var(FULL). Further, estimating the variance of the difference FULL - LOW over time does not make much sense physically, since it's inconsistent with the physical state of either of the two ensembles.

Fig.1a shows that actually, the decadal predictability is quite limited regionally and does not extend to the European continent. But more generally, how can we get statistical significance on something which is 2x smaller than internal variance? How is the significance level effectively estimated in the ppvf technique?

General remark: The ppvf is applied on separate runs, but the technique is not really well known in the climate community... so I would please ask the authors to explain better how they use it. Otherwise, the results will not be reproducible.

Fig.2 - this is a nice schematic, but there is literally nothing new here over e.g. the schematic by Gray et al., 2010 and the ones by Kodera et al. - hence, I frankly do not see the value of this figure and would recommend removing it.

Fig.3b If the averaging over 10 ensemble members brings out the forced signal, then why is there not corresponding polar vortex strengthening around 1960, which is one of the strongest solar cycles on records? Further, why does the vortex at times even anticipate the solar cycle, such as e.g. at year 1980? Using multiple runs should bring out the signal even in individual cycles... so this should still work at all times!

Fig.5a What about the phase shifts in individual runs? Why does the sinus curve look shifted in some (2-3) of the runs? This rather hints at a sporadic & random process, rather than a "synchronization"... aLso, it does not make much sense physically to compare individual members in FULL against the ensemble mean of LOW, which is a separate ensemble!

Fig.5b What about the same calculation, for the individual LOW minus LOW-ensemble mean differences? This panel would be important to evaluate how much the unforced decadal variability itself can originate the apparent "synchronization"!

Fig.5c what does "FULL-LOWFREQ" mean? Is this the running mean correlation between the solar index and the difference in the NAO at each year of the simulation, or is this the difference in the running mean correlation in FULL vs LOWFREQ? If the latter is the case, then it's not really a running correlation. I find this method quite convoluted and not fully justified. Why not simply looking at the running mean correlation itself, rather than the difference? This would highlight how the NAO itself correlates with the solar cycle, rather than its 'solar derived' component (which is supposed to be represented by the FULL-LOWFREQ difference)

Fig.5c - why is only lag 0 shown for the model and not the lag 2, as done with the observations? The model shows better agreement with the lag 2 in OBS data... so the lag 2 in the model should be compared with the lag 2 in the OBS. Moreover, the observational data does not show the equivalent of the FULL-LOWFREQ difference but rather of FULL alone, so this is not an apple vs apple comparison! This would be another argument in favor of using the actual running mean correlation in the FULL ensemble rather than of the FULL-LOW difference to show a detectable solar "impact". You could compare the running mean correlation of FULL and should be able to show that it's higher than in LOWFREQ to convincingly demonstrate that the solar signal "emerges" from internal variability (which is the claim of the paper) - see major comment above.

Fig.5c - why is the dip in the model correlation around year 1900 not captured by the observations? Can the authors speculate?

Fig.5d - if the running mean correlation is calculated over 45-y windows, then the individual data points are not mutually independent, and this would reduce the degrees of freedom. Is this taken into account in the calculation of the 90% error bar?

Fig.5d If the scatter plot is for February, then the main conclusion about the "enhanced decadal-scale component of the NAO under a strong solar cycle forcing" only applies to this month, and not to, as previously argued in the literature, the whole boreal winter. This should be clarified in the abstract.

L148 "18% of the magnitude of internal variability" is again, a misleading statement, as the analysis using ppvf does not really quantify the magnitude of the internal variability on a specific time-scale.

L150 "small in magnitude but manifests itself as an organization and synchronization of internal variability as shown by the cross-correlations" --> the cross correlation is not really a cross correlation in the cleanest statistical sense, but rather an ad-hoc construct designed to isolate the "solar signal" (FULL-LOWFREQ) rather than the solar influence itself. This would be e.g. more convincingly shown by providing evidence that the decadal variance in FULL and LOWFREQ are significantly different.

L180 - actually, it's not inconsistent, as the signal is not really significant at the 95% level here either (only at 90% level), which indicates that there's a (non negligible) probability that the signal may be by chance. I think this should be stated here.

L195 "enforces the NAO phase if the solar forcing is strong enough" - this is really hard to believe, as the phase is really not constant over time. Moreover, Chiodo 2019 also used a strong epoch for the solar forcing, and got a time-dependent signal, too. Hence, the authors should at the very least comment on that, and elaborate possible reasons for the disagreement.

L215 interestingly, the disagreement between model and observations in terms of the lag is only noted here. Could it also be that part of the signal in the observations is by chance? Could this possibility at least be listed here?

L220 the pptf technique does not really convincingly demonstrate that the potential predictability is enhanced by this much (20-25%), as the signals over wide parts of Europe remain insignificant (Fig.1) and there is no convincing demonstration that the skill of the model is improved over the decades with a "strong solar cycle". Hence, this remains an unjustified claim rather than a science-based statement. Rather, this analysis shows that a small solar cycle signal may be present, but that an enormous amount of model data is needed to make it statistically detectable.

Also, if we need to run a model for 1000s of years to get a solar signal, then this would rather argue against an effective usability of the solar forcing for decadal prediction. Since this study does effectively not quantify the predictability (e.g. by using prediction skill scores metrics, or similar), I urge the authors to tone down any "predictability" statements.