

Weather Clim. Dynam. Discuss., referee comment RC2
<https://doi.org/10.5194/wcd-2021-32-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on wcd-2021-32

Daniela Domeisen (Referee)

Referee comment on "Flow dependence of wintertime subseasonal prediction skill over Europe" by Constantin Ardilouze et al., Weather Clim. Dynam. Discuss.,
<https://doi.org/10.5194/wcd-2021-32-RC2>, 2021

REVIEW of „Flow dependence of wintertime subseasonal prediction skill over Europe" by Ardilouze et al

Review by D. Domeisen

SUMMARY: This study investigates the dependence of winter forecast skill over Europe on the weather regime at initialization of the forecast. The authors find that strong NAO regimes at initialization increase subsequent skill during week 3 in two sub-seasonal prediction models.

ASSESSMENT: The paper is very well written and the analysis is very interesting and should be published. I have a range of minor comments about both the technical and the writing / interpretation aspects of the manuscript, which are detailed below and which I hope will be helpful for the manuscript.

My only major comment is that it didn't become fully clear if the skill following intense NAO regimes is mostly a factor of increased persistence due to the intensity of the event or if there is something intrinsically / dynamically different between these regimes. Even if this question cannot fully be answered, it can be addressed in more detail, see comments below.

COMMENTS AND RECOMMENDATIONS:

Line 17/18: "predictability well": interesting, I haven't heard this one. Just out of curiosity: I only heard the term "predictability desert" for S2S so far. "well" sounds much more positive, does this refer to a "source" of predictability? So the "windows of opportunity" don't seem like a contradiction, but rather a continuation of the "well"?

Lines 34 – 37: maybe further citations could help here, e.g. the MJO to NAO teleconnection has been described by Lin et al, 2009 (<https://doi.org/10.1175/2008JCLI2515.1>), and the resulting S2S predictability has been described in Vitart (2017), which is already cited elsewhere in this manuscript.

Line 98/99: "Because ERA5 and ECMWF reforecasts are derived from two versions of the same model": do you mean the use of ERAinterim as initialization for ECMWF as compared to ERA5 for CNRM (table 1)? Please clarify. (in this case, using ERAinterim for validating the results would probably be more suitable than JRA-55, but then again the differences in the results for using ERA5 versus ERAinterim would likely be even smaller than for comparing to JRA-55, so I don't suggest that the authors perform this comparison.)

Line 112: "the RMSE normalization method is arbitrary": not sure what you mean here, please clarify, which normalization did you use?

Section 2.3: do you use a minimum duration of each cluster, or can each consecutive day be assigned to a different weather regime? a persistence criterion may be useful when looking at S2S timescales.

Figure 1: it's surprising that all grid points and both models only show significant correlations, with the only exceptions being small areas in weeks 3 and 4. In particular, even correlation values below 0.2 still show significance – is this due to the large number of initializations used here?

Figure 4: I'm not sure the linear trendlines are very helpful here. Are you suggesting there is a linear trend in forecast skill? For figure 4, I would rather focus this figure on the connection to the NAO only. (I understand you've removed a linear trend, but that is the trend in T2m, not forecast skill or the NAO index, so it's not related to the trend shown in the figure. Good to know removing the trend does not make a difference in your results.)

Lines 216 – 219: (see also comment above about WR persistence above) I fully agree that a multi-day window should be used.

- Do you allow for several regimes in these 4 days? i.e. could these 4 days theoretically be assigned to 4 different WRs?
- Do you do this analysis separately for each ensemble member, or for the ensemble mean?
- Alternatively, you could introduce a persistence criterion or threshold value for WRs or average the WRs over a few days. We had to introduce such a criterion in this paper <http://doi.org/10.5194/wcd-1-373-2020>, but I'm sure there are others that do the same. I realize you did this to a certain extent by adding the zero regime in section 3.2, but I'm wondering if the results were more robust overall if you introduced a persistence criterion and a zero category throughout the manuscript.

Table 2: do I understand this correctly that each value represents the percentage among the skillful forecasts as opposed to the climatological frequency in brackets? If so, they should add up to 100 (they all do except the skillful forecasts for CNRM, please check).

Table 2 caption: "significantly different": I assume you mean that each value is significantly different from the value in brackets? Could you clarify?

Figure 5: it would be helpful to indicate the number of initializations in brackets next to each WR in the legend

Figure 5: in addition to the significance computed for difference from zero, it would be interesting to know if the ACC is significantly different for NAO+ as compared to other WRs, e.g. by showing error bars or shading (similar to Fig 9) showing the standard deviation to see if NAO+ overlaps with other WRs. I imagine it will not be clearly significantly different, which would not be a problem in my opinion, but it would be nice to get an estimate of the variability of the curves, e.g. to know if forecasts initialized in NAO+ also contain very poor predictions, or if most of them really show above average ACC. I think this would support the main message of the paper.

Figure 6 / lines 262-263: do you have the same plots for the other two regimes? these would be useful for comparison, as you here make statements about NAO versus non-NAO, but non-NAO initializations are not shown (at least a figure as supplementary material these would be useful). This would also allow for a better understanding if it's the higher persistence of the NAO regimes that makes their aftermath more predictable as compared to other WRs.

Line 275: "anthropic": do you mean "anthropogenic"?

Figure 7 / lines 278 - 279: did I get this correct that all 4 days have to have the same WR for the piControl simulation, while for the S2S data it only has to be the "regime with the greatest number of occurrence during the 4 initial days" (line 229)? Could you clarify? I

understand this will lead to a larger number of samples in piControl, but best to be consistent for comparison.

Figure 7: figure labels would be helpful in addition to the caption.

Line 287 – 288: “hemispheric positive AO pattern evoked earlier is a model artefact”: this is not clear – I’m pretty sure that all of these patterns will confidently project onto the positive AO pattern, despite their differences.

Line 300: “This agreement is much better for negative than positive NAO”: I’m wondering if this is due to the fact that NAO- is a much more pronounced North Atlantic regime than NAO+. In particular, if dividing up WRs into more than 4 regimes, NAO- remains a separate regime (equal to Greenland blocking), while NAO+ is sub-divided into separate regimes by the clustering algorithm. NAO+ is more of a mixture of several regimes that reflect the average state of the North Atlantic, while NAO- is a distinct regime. To paraphrase Brian Hoskins (I hope I’m doing this correctly), NAO+ is basically the “normal” state of the North Atlantic, while NAO- is a distinct anomalous state of the North Atlantic.

Line 311 – 315: this decorrelation timescale and behavior (e.g. the rebound) is consistent when looking at the decorrelation for a wide range of different NAO indices (Figure 3b in <http://doi.org/10.1175/JCLI-D-17-0226.1>, already cited elsewhere in this article).

Lines 331 – 332 / lines 370 onward: I don’t think that the regression analysis is proof that the NAO pattern at initialization influences the entire NH. There are many common remote drivers that will lead to both a NAO-type pattern over the North Atlantic and consistent anomalies elsewhere, e.g. precursors in the tropics, the North Pacific, or the stratosphere. If you want to include Figure 10 (it would be equally fine in supplementary), I think the text should be formulated more carefully. This might also explain your finding on lines 343-344: “However, no improvement of skill is detected over South East US and off the US Atlantic coast, as could have been expected from the teleconnection patterns.”

Line 335: earlier only the top 10% of strong NAO initializations were kept. What is the reason for using the quartile now? Increased sample size?

Line 354: “but performs reasonably well anyway”: can you be more specific / quantitative?

Section 3.4 / Figure 6: It didn’t become fully clear if the skill following intense NAO regimes is mostly a factor of increased persistence due to the initial intensity of the event or if there is something intrinsically / dynamically different between these regimes (see major comment above). I think it might help to repeat Figure 6 for initializations in intense

NAO regimes and to check if a clearer pattern emerges as compared to all "regular" NAO regimes and other WRs.

SOME TYPOS I FOUND:

Lines 9 and 367: conditioned -> conditioned

Line 11: others parts -> other parts

Line 18: traditionnal -> traditional

Line 19: reveals -> reveal

Line 68: of **the** new CNRM system

line 92: greenhouse gases emissions -> greenhouse gas emissions

line 175: tend -> tends

line 192: than other forecasts more spread out -> than other forecasts **that are** more spread out

line 193: figured in green and yellow shades -> **plotted** in green and yellow shades. **(missing period)**

Line 194: 2x counterparts

Line 206: forecast -> forecasts

Line 214: next section -> the next section

Table 2 caption: these frequency -> these frequencies

Line 287: over North Pacific -> over the North Pacific

Line 287: similitude -> similarity

Line 307: the figure 9 -> figure 9

Line 363: regime -> regimes

Figure A1 caption: "for week 1 to week for": do you mean "week 4"?