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.

We would like to thank Dr. D. Domeisen for the overall positive comments and also for the thorough and relevant review of our manuscript. We have made our best effort to address each point and hope the revised manuscript is now substantially improved.

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.

This point is addressed following your last comment 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”?

This is a very appealing interpretation of our choice to use the word “well” instead of “desert”. The truth is we just meant to pick a different word from “desert”, originally, without realizing the ambiguity. We thought more of a “Predictability sink” rather than a well but since you mention it, the image of a well fed by intermittent sources of predictability is nicer than a desert or a sink indeed. Identifying and detecting windows of opportunity could indicate when to draw predictability from the well. We wish to keep your interpretation full of imagery and have therefore rephrased as follows: (l.17-24)

‘The S2S horizon has been often considered as a "predictability desert" based on mean statistics and traditional methods and analyses inspired from seasonal-to-decadal climate prediction, but the most recent studies reveal instead so-called windows of opportunity based on the fact that under certain circumstances, and for specific events and regions, S2S predictability can be considerably increased (Mariotti et al., 2020). This conditional predictability is illustrated by a number of case studies showing the successful anticipation of extreme climate events by dynamical forecast systems beyond 15-day lead time (Domeisen et al. in revision).

Rather than a predictability desert, the S2S horizon appears more like a "predictability well' intermittently fed by these windows of opportunity. Timely drawings from the well, i.e. a priori identification of the windows of opportunity, are a major asset in an operational context for the development and uptake of climate services relying on subseasonal forecasts, but it remains a scientific challenge with some promising examples.'

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.

Thank you. These two references are now cited.

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.)

Actually this sentence was not related to how reforecasts are initialized but rather how they are evaluated. The idea was to avoid any suspicion about ECMWF prediction system being favoured by a reference dataset (ERA5) derived from a similar model. Assessing the skill with JRA-55, arguably independent from the IFS model, allows to discard this suspicion because we do not find any major difference.

We have modified the text as follows (see l. 103-106) to clarify our point:

"To verify if the assessment of the ECMWF predictions is favoured by the choice of this reanalysis, we have compared some of the results obtained with ERA5 to results using the JRA-55 reanalysis as a reference"

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

It has been normalized by the interquartile range of the observation, as stated in the previous sentence, but here, we just meant that there is no standard way of normalizing the RMSE. We could have normalized it by the observation mean for instance. However, we believe the normalization strategy is of minor importance in the context of this study and hence removed this comment from the revised manuscript. See line 119-120

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.

Both have been tested (3-day persistence criterion vs. no persistence criterion, see l. 249-259 of the original manuscript). We have added a few lines at the end of section 2.3 to clarify that both methods have been considered (l.148-150).

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?

Yes, we are positive that even weak correlations appear significant because of the very large sample of 320 initializations for each model. This is suggested by the comments relative to figure 11 a) and b) , where the subsampled initializations lead to patchier correlation patterns. However, we clearly state that significantly positive correlations do not always imply usefulness, hence our CRPSS analysis.

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.)

We thought it was worth assessing the impact of the T2m trend because it could have been that we obtained a larger number of skillful forecasts initialized during the most recent years as compared to forecasts initialized during the early years, for a "bad" reason. For example, we could have expected more skillful forecasts in 2016 than in 1997 just because the recent forecasts tend to predict warm temperature anomalies over Europe more frequently, in agreement with more frequently observed warm anomalies associated to the warming trend.

However, we agree that it is sufficient to indicate that removing the trend does not make a difference. We have thus removed the trend lines from figure 4 and also changed the text accordingly.

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?

  Yes, we allow for several regimes in these 4 days. More precisely, we count the occurrence of the 4 regimes for each ensemble member (this now specified in the manuscript l.233) and each forecast, so that percentages shown in table 2 are based on a pool of 2992 (for ECMWF : 11 members x 4 initial days x 68 skillful forecasts) or 2720 (for CNRM: 10 members x 4 initial days x 68 skillful forecasts) days. We proceed likewise to compute percentages shown in parenthesis, but based on the 252 (=320-68) other forecasts.

- 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.

  The threshold criterion has not been tested in this study but is mentioned at the very end of our conclusion, highlighting the potential limitations of our study stemming from the methodology. As for the persistence criterion we have only tested it for a limited number of analyses (see the dedicated appendix B) but found so little difference that we made the decision to carry out the study without such criterion. Another reason for this decision is that it does not require to discard forecasts characterized by a dominant 'zero' regime at initialization (concerning approx. 70 forecasts out of 320) at the expense of the sample size. Our approach benefits from more robustness due to a larger sample size which is critical for the subsampling approach performed in section 3.4.

  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).

Thank you for spotting this mistake. Indeed, for CNRM "good" forecasts, the frequency of blocking is 14.9% and not 20.5% (vs. 30.5% for the "wrong" forecasts). The number has been corrected and they now add up to 100%. Fortunately, it does not change the message.

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

Yes, this is correct. We have clarified the caption accordingly

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

Done

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.

We have reproduced figure 5 by adding ±1 standard deviation in dashed thin lines, and extending the y-axis down to -0.5 (fig. R1). As expected, no significant differences can be noticed between ACC curves. The standard deviation intervals are pretty similar even if, by the eye, the NAO- standard deviation interval (light green) seems slightly wider. We choose not to use this plot in the revised paper, but we have rephrased as follows, to also take into account a comment from Reviewer #1:

"For both systems, the mean ACC of the forecasts initialized in NAO+ conditions becomes higher than those initialized with other regimes by day 6 and more markedly from day 15 onwards, albeit not significantly (not shown) " (l.247)

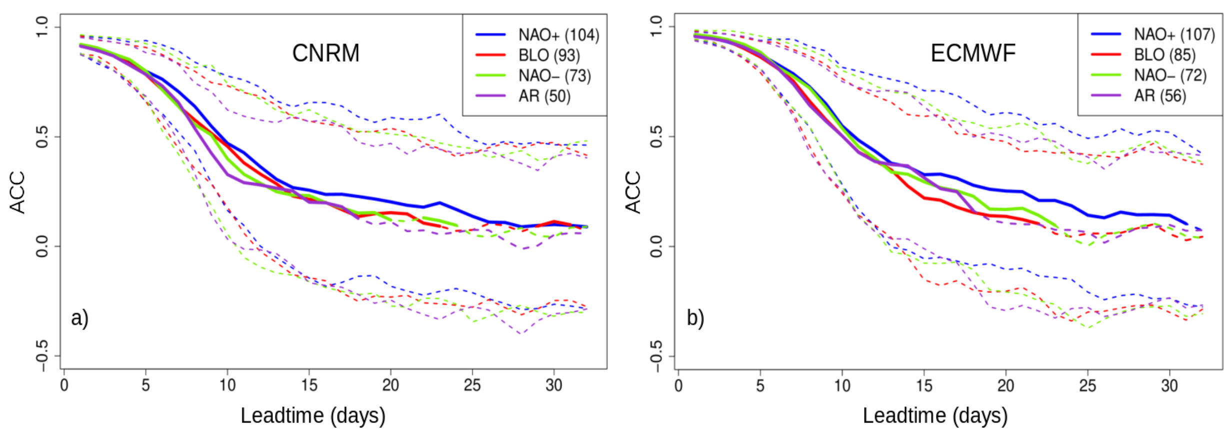


Figure R1: as Fig.5 with ±1 standard deviations in dashed lines

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.

The corresponding plots are shown below (fig. R2) and added in supplementary material. Forecasts initiated in BLO (AR) regimes do not show particular changes in BLO (AR) regime proportions at week 3 with respect to climatology. This result may suggest that the higher persistence of NAO regimes play a role in the predictability. But this conclusion remains highly uncertain, as discussed in reply to your last point below.

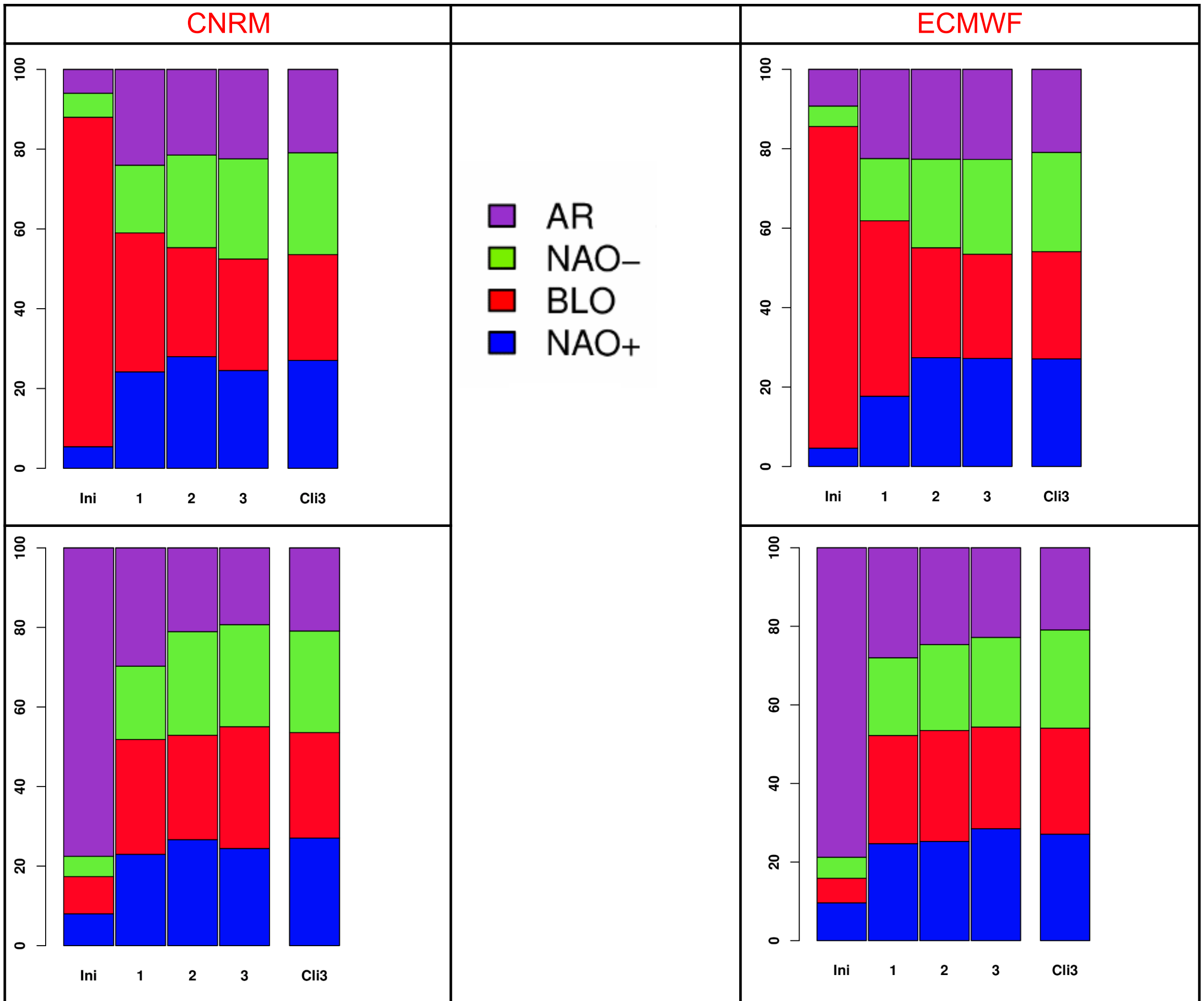


Figure R2: Weekly evolution of regime frequency among forecasts initialized in BLO (top row) or AR (bottom row) conditions for CNRM (left column) and ECMWF (right column). The leftmost bar corresponds to the 4 initial days. The rightmost bar corresponds to the climatological frequency for week 3.

Line 275: “anthropic”: do you mean “anthropogenic”? Yes. Corrected

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.

You understood correctly and we agree that our approach lacks consistency. We have therefore applied the "greatest number of occurrence" method to the piControl and ERA5 as well, and modified the text and composite maps in the revised manuscript accordingly (l. 292-294 and 303). In addition to the better consistency, this looser constraint in the piControl and ERA5 subsampling leads to a larger sample size, which is desirable for the comparison with the S2S data. Thank you for raising this issue.

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

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.

We have rephrased this assertion, to simply mention the divergence over the North Pacific with respect to the AO+ loading pattern (l. 305-306).

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.

In the study carried out by Falkena et al (2020) and cited in our conclusion, they suggest 6 WR instead of 4, with NAO+ still present, although with a slight shift in the centres of action. It is no longer the dominant WR though. However, their study meets your comment when they assert in their conclusion:

"the dominant occurrence of the NAO+ when there are only four clusters, which likely is due to it being the only regime with a low pressure area in the north, is reduced by the addition of two regimes that also have this feature. Therefore, six regimes allow for more variability in their representation of the circulation and prevent all data with a more zonal flow from projecting onto the NAO+"

Another interesting approach is that of Dorrington and Strommen (2020) . They show that removing the influence of the jet speed leads to an optimal clustering of k=3 or k=5 without NAO+, and with an increased stability of the remaining regimes

We have added a sentence suggesting that our clustering method may not be optimal, with NAO+ being a mere generic mode, and the other regimes are perturbations from this generic mode. Another interpretation is that NAO+ is not specific enough and potentially conceals a variety of WR. See line 319-320 and lines 410-414 in the conclusion.

Dorrington, J., & Strommen, K. J. (2020). Jet speed variability obscures Euro‐Atlantic regime structure. *Geophysical Research Letters*, *47*(15), e2020GL087907.

Falkena, S. K., de Wiljes, J., Weisheimer, A., & Shepherd, T. G. (2020). Revisiting the identification of wintertime atmospheric circulation regimes in the Euro‐Atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, *146*(731), 2801-2814.

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).

True. A comment has been added (lines 334-335)

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."

We agree that this is an overstatement. We rephrased the sentences at the beginning of section 3.4 as follows (lines 351-353):

"In the previous section, we have identified a statistical link between wintertime temperature anomalies over a number of regions of the northern hemisphere extratropics and the 3-week antecedent NAO index."

And later in the same section (lines 362-366) as:

"These regions match remarkably well with the regression patterns highlighted in observations (fig. 10) and they are relatively consistent between CNRM and ECMWF systems. Note that these regression patterns do not necessarily imply a causal relationship and may originate from a number of remote drivers (which are not investigated here). Indeed, no improvement of skill is detected over South East US and off the US Atlantic coast."

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?

The idea was to keep a sample large enough (40 forecasts out of 320) to draw robust conclusions, but it is a bit arbitrary. We have added a sentence to justify this choice (l. 357)

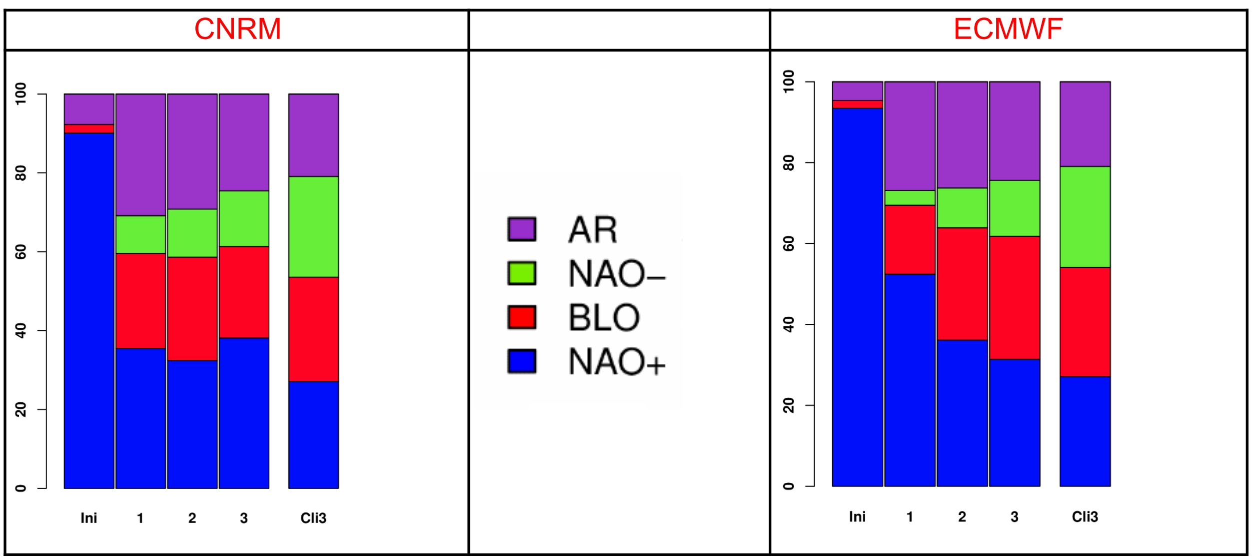
Line 354: “but performs reasonably well anyway”: can you be more specific / quantitative?

We have changed the text as follows:

“ but the spatial patterns compare relatively well”. Spatial correlations between CNRM and ECMWF skill is discussed in Section 3.1

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.

This is a key point indeed, and we have repeated figure 6 as suggested (see fig. R3 below). We do find that these forecasts show an increased proportion of NAO- (NAO+ to a lesser extent) at week 3, which may indicate that the persistence of these regimes is related to their initial intensity.



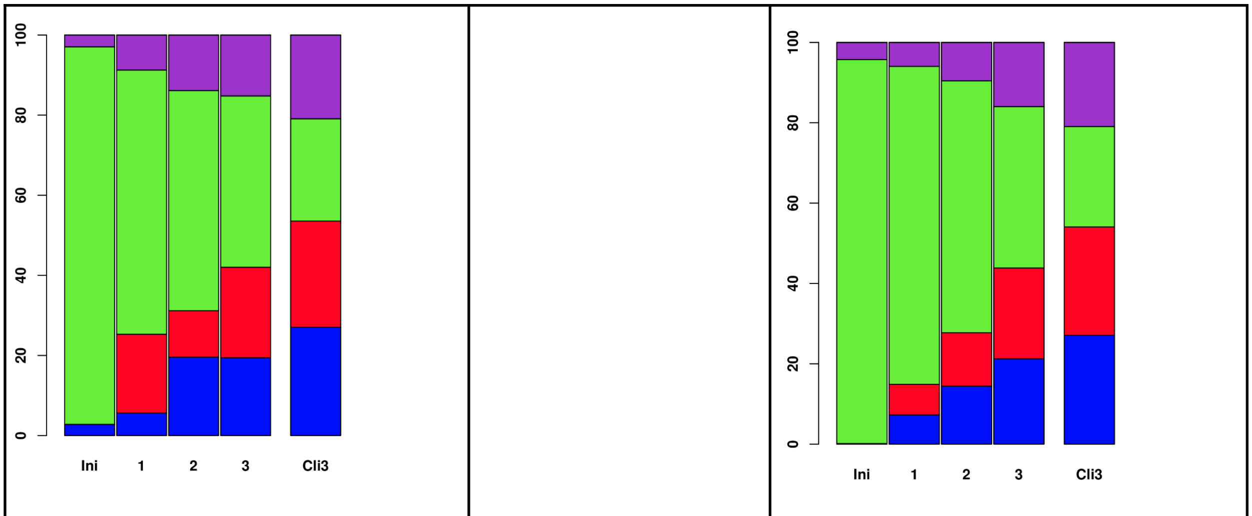


Figure R3: same as figure 6 after subsampling forecasts with the top 10% of strong NAO+ (top row) or NAO- (bottom row) initializations.

However, we wanted to go a bit further, and see if it is more the regimes persistence or their number of occurrences that depended on the initial intensity.

The scatterplots of figure 4 show for each forecast initiated in NAO- (NAO+) the number of occurrences of NAO- (NAO+) sequences of at least 3 consecutive days, as a function of the mean duration of these sequences. The red dots correspond to the forecasts with the highest initial NAO intensity.

For both models, the NAO- probability density function has a more elongated shape than the NAO+ counterpart, which is consistent with the well known higher persistence of the NAO- regime. It appears that forecasts with intense NAO- initial conditions (red dots) are overrepresented in the rightmost part of the PDF, thereby confirming to some extent the link between initial intensity and persistence of the NAO- WR. It is not the case for NAO+, although, for ECMWF, many red dots are located on the upper part of the probability density function, meaning that intense initial NAO+ could translate into more frequent occurrences of NAO+ sequences during subsequent weeks. There is no strong evidence for this conclusion given that it does not show in CNRM.

Finally, these additional analyses only provide a very incomplete answer to the main point you have raised about our study. An intense NAO- can lead to increased persistence of this weather regime, which probably depends on the interaction with other drivers (stratosphere ?). On the other hand, the NAO+ case is still unclear, potentially due to the lack of specificity of this weather regime when defined by k-means clustering with k=4 (see discussion of a previous point) .

We have decided to report these additional figures and this discussion in supplementary material. This is now indicated before the conclusion section (l. 371-376).

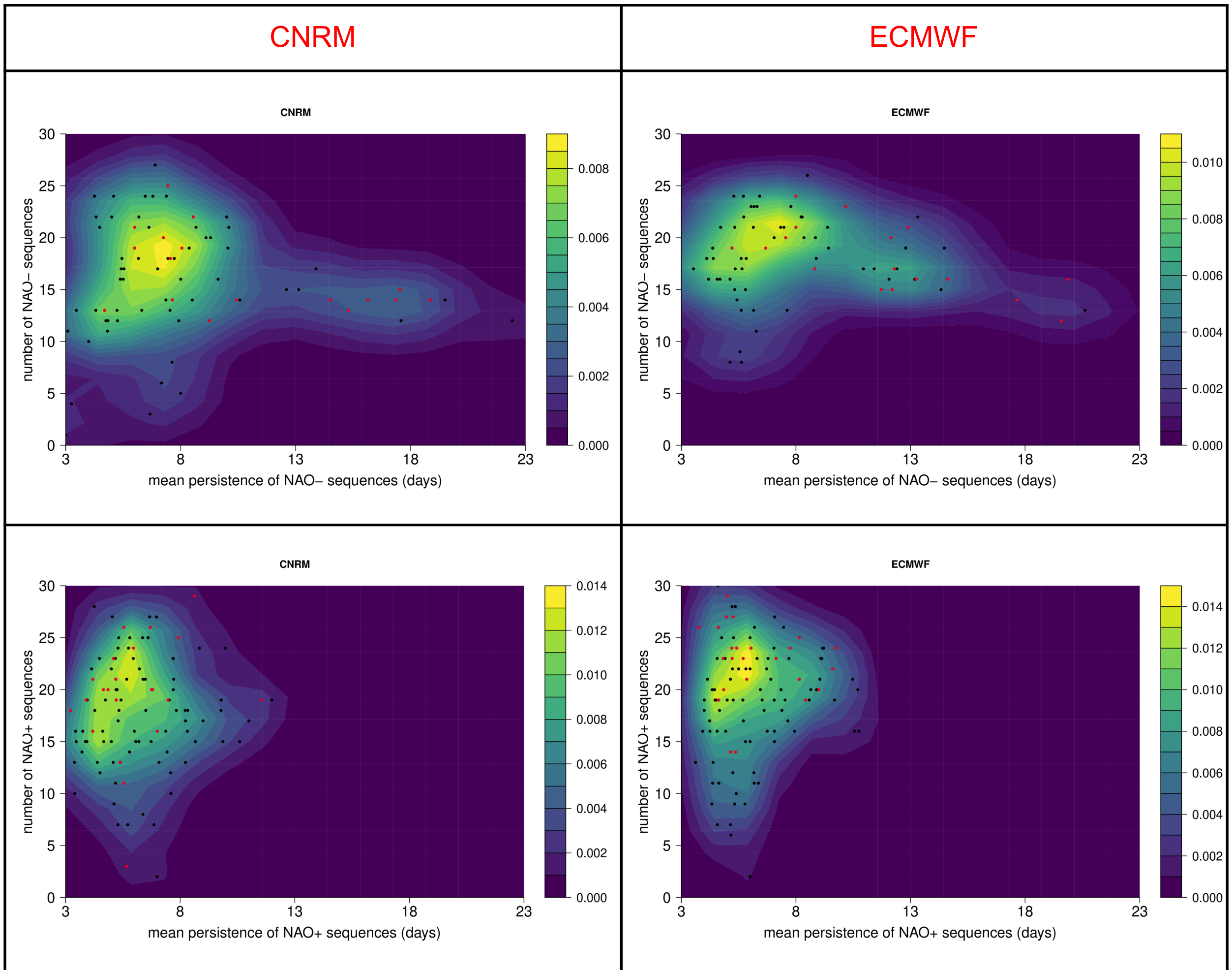


Figure R4: Scatter plot (dots) and associated probability density function (shading) of forecasts initiated in NAO- (top row) and NAO+ (bottom row). The y-axis indicates the number of 3-day or more NAO- (resp. NAO+) sequences in all the ensemble members and the x-axis the mean persistence of these sequences, in days. Red dots mark those forecasts initiated in strong NAO- (resp. NAO+) conditions.

SOME TYPOS I FOUND:

Thanks for reporting the typos. Corrected typos are marked with a red “check” symbol

Lines 9 and 367: conditionned -> 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”? Yes ☺ ✔

Flow dependence of wintertime subseasonal prediction skill over Europe

Constantin Ardilouze[1], Damien Specq[1], Lauriane Batté[1], and Christophe Cassou[2]

[1]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

[2]CECI, Université de Toulouse, CNRS, CERFACS, Toulouse, France

Correspondence: Constantin Ardilouze (constantin.ardilouze@meteo.fr)

Abstract. Issuing skillful forecasts beyond the typical horizon of weather predictability remains a challenge actively addressed by the scientific community. This study evaluates winter subseasonal reforecasts delivered by the CNRM and ECMWF dynamical systems and identifies that the level of skill for predicting temperature in Europe varies fairly consistently in both systems. In particular, forecasts initialized during positive NAO phases tend to be more skillful over Europe at week 3 in both systems. Composite analyses performed in an atmospheric reanalysis, a long-term climate simulation and both forecast systems unveil very similar temperature and sea-level pressure patterns three weeks after NAO conditions. Furthermore, regressing these fields onto the 3-week prior NAO index in a reanalysis shows consistent patterns over Europe but also other regions of the northern hemisphere extratropics, thereby suggesting a lagged teleconnection, either related to the persistence or recurrence of the postive and negative phases of the NAO. This teleconnection, ~~conditionned~~ conditioned to the intensity of the initial NAO phase, is well captured by forecast systems. As a result, it is a key mechanism for determining a priori confidence in the skill of wintertime subseasonal forecasts over Europe as well as ~~others~~ other parts of the northern hemisphere.

1 Introduction

Skillful weather and climate predictions for horizons beyond two weeks could benefit many users (White et al., 2017). Lately, so-called subseasonal-to-seasonal (S2S) forecasts have gained considerable attention and effort from the scientific community in multiple aspects, including the characterization of sources of predictability, such as atmospheric teleconnections, the initialization and generation of ensemble forecasts, the calibration and tailoring of the raw model outputs for enhanced usability and uptake by the application community (Merryfield et al., 2020). The S2S horizon has been often considered as a "predictability ~~well~~desert" based on mean statistics and ~~traditionnal~~ traditional methods and analyses inspired from seasonal-to-decadal climate prediction, but the most recent studies ~~reveals~~ reveal instead so-called windows of opportunity based on the fact that under certain circumstances, and for specific events and regions, S2S predictability can be considerably increased (Mariotti et al., 2020). This conditional predictability is illustrated by a number of case studies showing the successful anticipation of extreme climate events by dynamical forecast systems beyond 15-day lead time (Domeisen et al., In revision). ~~The a~~ Rather than a predictability desert, the S2S horizon appears more like a “predictability well" intermittently fed by these windows of opportunity. Timely drawings from the well, i.e. a priori identification of ~~these~~ the windows of opportunity~~is~~, are a major asset in an operational context for the development and uptake of climate services relying on subseasonal forecasts, but it remains a scientific challenge with some promising examples. For instance, Mayer and Barnes (2021) recently related the accuracy of North-Atlantic geopotential height forecasts issued by a neural network to their level of confidence. Their approach also allows to pinpoint the most relevant remote tropical regions leading to higher forecast skill.

At the subseasonal scale, European and Eastern North American climate are influenced by the phase of the North Atlantic Oscillation (NAO), the leading mode of climate variability over the North Atlantic sector (Cattiaux et al., 2010; Seager et al., 2010; Luo et al., 2020). The positive (NAO+) and negative (NAO-) phases of the NAO correspond to two well-identified weather regimes characterizing recurrent synoptic-scale atmospheric patterns in winter, along with the Atlantic Ridge (AR) and Scandinavian Blocking (BLO) (e.g. Vautard, 1990). The NAO is sometimes considered as the local manifestation of a hemispheric variability pattern called Northern Annular Mode or Arctic Oscillation (AO). AO and NAO are strongly correlated in present climate (Hamouda et al., 2021).

The tropics-extratropics teleconnection described by Cassou (2008) and Lin et al. (2009) illustrates the major role of the Madden-Julian Oscillation (MJO) phase in pre-conditioning North Atlantic weather regimes. Recently, Lee et al. (2019) found evidence of El-Niño Southern Oscillation (ENSO) modulating the strength of this teleconnection which largely contributes to the subseasonal predictability of the North Atlantic (Vitart, 2017). More generally, the tropical background state and variability are essential to induce subseasonal predictability of the northern hemisphere circulation, especially in winter, provided that the climate phenomena supporting the teleconnection, such as the atmospheric upper-level jet are adequately simulated (Yamagami and Matsueda, 2020). The stratosphere is another key precursor to the variability and predictability of the wintertime northern hemisphere circulation (Domeisen et al., 2020). A correct initialization (Zuo et al., 2016), together with a good representation of the stratosphere-troposphere coupling (Kolstad et al., 2020) accordingly contributes to skillfully forecast the NAO. The combination of ENSO evolution and stratospheric processes also drives the extended range NAO predictability (Sun et al., 2020).

Other studies have focused on the predictability conditioned by the wintertime weather regimes occurring at initialization time. Based on a specific set of weather regimes affecting North America, Vigaud et al. (2018) demonstrated the capacity of the ECMWF subseasonal forecast system to successfully predict some of them up to two weeks. Robertson et al. (2020) built on this study to emphasize the value of this weather regime approach for identifying forecasts of opportunity over North America, with high skill up to 30 days ahead for specific events or seasons. The flow-dependent variations of the subseasonal forecast skill over Europe was also evidenced (Ferranti et al., 2018), with a relatively good capacity of the ECMWF system to reproduce the preferred transitions between weather regimes. Ferranti et al. (2015) identified differences in medium-range weather forecast performances conditional to the regime flow in the initial conditions with initial NAO- states leading to more skillful forecasts. Beyond approaches based on weather regime prediction, Minami and Takaya (2020) recently found that Northern Hemisphere 500 hPa geopotential height was more predictable when following strong negative initial AO, due to an eddy-zonal flow feedback that contributes to persist this mode of atmospheric variability. This study emphasizes the role played by large-scale extratropical atmospheric dynamics in subseasonal predictability, on top of tropical and stratospheric precursors.

Our main goal here is to further explore the relationship between the circulation flow present in the forecast initial conditions, hereafter initial weather regimes, and subseasonal predictability of the 2m-temperature in winter over a broad North Atlantic European domain. In this study we analyze jointly the ECMWF forecast system, and the most recent CNRM (Météo-France) subseasonal forecast system, launched in October 2020. The next section presents these forecast systems, as well as reference data and methods adopted in this study. The main results are then developed in a dedicated section. The last section provides concluding remarks and prospects.

2 Data and methods

2.1 Forecast systems

Subseasonal forecasts delivered by CNRM have been routinely feeding the S2S database (Vitart et al., 2017) since 2015 with forecasts issued every Thursday. Lately, the CNRM forecast system version 1 (Ardilouze et al., 2017) has been superseded by a version 2 used in this study. Unlike the ECMWF extended range forecast system (which also feeds the S2S database), the CNRM upgraded system has been designed for research purposes and is not intended for operational aspects. Since the ECMWF system is often acknowledged as the most skillful system in several intercomparison comparison studies (e.g. Zheng et al., 2019; Specq et al., 2020), it will serve as a benchmark in the present work to assess the performance of the new CNRM system. The main characteristics of both forecast systems are described in table 1.

In this manuscript, 'reforecast' and 'forecast' indistinctly refer to retrospective forecasts, also named 'hindcast' in other studies. The comparison of ECMWF and CNRM prediction systems is facilitated by their comparable reforecast ensemble size and a common 20-year reforecast period. Here, we consider the December-to-March extended winters from 1997/1998 to 2016/2017.

However, because of different reforecast designs, initial dates do not exactly match between the two systems. This issue is addressed as follows. We first select for each winter 16 consecutive CNRM start dates (i.e. Thursdays) after November 13th, so that week 3 and 4 are always included within the December to March 4-month period. Then for each of these 320 (16x20) CNRM initial dates, we pick the closest date among the available ECMWF initial dates. Since ECMWF forecasts are issued twice a week, the resulting date from this selection either matches the CNRM counterpart or precedes/follows it by no more than two days, depending on the reforecast year. Note that each reforecast is evaluated against the corresponding reanalysis dates, to ensure a perfectly fair inter-model comparison.

Forecast and observed daily anomalies are considered rather than full fields, in order to remove the model bias : for the n^{th} 32-day forecast ($n \leq 16$) of a given winter, daily anomalies are computed by ~~substracting the daily climatology~~ subtracting the 32-day climatology as a function of leadtime, corresponding to the mean of the n^{th} forecasts of the 19 other winters.

Table 1. Characteristics of the CNRM and ECMWF subseasonal reforecasts

Characteristic	**CNRM**	**ECMWF**
Model	CNRM-CM6-1 HR (Voldoire et al., 2019)	ECMWF IFS CY43R3
Horizontal resolution	TL359 (~50 km)	Tco639 (~15 km) up to day 15, Tco319 (~31km) after day 15
Vertical resolution	91 levels up to 0.01 hPa	91 levels up to 0.01 hPa
Ocean resolution	0.25°, 75 levels	0.25°, 75 levels
Reforecast ensemble size	10	11
Reforecast frequency	Thursdays	Bi-weekly
Reforecast system	fix	on-the-fly
Atmospheric/Land initial conditions	ERA5 (Hersbach et al., 2020)	ERA-Interim (Dee et al., 2011)
Ocean/sea-ice initial conditions	Mercator Ocean International	ORAS5
Ensemble generation	Stochastic dynamics (Batté and Déqué, 2016)	Perturbed initial conditions (Singular vectors, Ensemble Data Assimilation) + Stochastic physics (SPPT and SKEB schemes)

In this study, we follow a frequently used convention in the S2S community to define weekly lead times (e.g. Vitart, 2004; Specq et al., 2020; de Andrade et al., 2021) : week 1 goes from day 5 to day 11, week 2 from day 12 to 18, week 3 from day 19 to 25 and week 4 from day 26 to 32.

For the composite analysis described in section 3.3.1, in addition to the forecast systems, we make use of a 300-year long pre-industrial simulation (known as piControl) of the same model used in the CNRM system, namely CNRM-CM6-1-HR, and performed in the framework of the Coupled Model Intercomparison Project Phase 6 (CMIP6, Eyring et al., 2016). This experiment is useful to assess the behavior of the model internal variability without any drift from initial conditions nor forcing interference stemming either from initialization or volcanic and anthropogenic aerosols as well as greenhouse ~~gases~~ gas emissions. Additionally, this simulation provides enough years to work on densely populated composite samples, thereby ensuring an enhanced robustness of the results.

2.2 Reference dataset and forecast skill metrics

The ERA5 reanalysis (Hersbach et al., 2020) serves as the reference for daily sea-level pressure and daily-mean 2-meter temperature. This reanalysis, although resulting from a model output, assimilates a wide array of observations, and will therefore be considered as our observational reference. For simplicity, we will use the term "observation" - albeit abusively - to refer to ERA5 in the rest of the manuscript. Because ERA5 and ECMWF reforecasts are derived from two versions of the same model, one may object that ERA5 is not a suitable reference for this study. ~~We have thus compared a few~~ To verify if the assessment of the ECMWF predictions is favoured by the choice of this reanalysis, we have compared some of the results obtained with ERA5 to results using the JRA-55 ~~reanalysis (Kobayashi et al., 2015) as reference instead of ERA5~~(Kobayashi et al., 2015) reanalysis as a reference. Given the very limited differences found (not shown), we have chosen to pursue with ERA5 only.

A common score to evaluate a subseasonal forecast system is the point-wise Pearson correlation between the ensemble mean forecasts and the corresponding observations over the entire reforecast period. Grid-point time correlation is a classic deterministic score, whose significance is here determined by a two-sided Student t-test at the 95% confidence level.

In order to evaluate the skill of an individual forecast, we also compute the anomaly correlation coefficient (ACC) which shows the level of spatial agreement between the forecast and observed patterns. This is performed over a domain covering Europe (hereafter EUR, 12°W,41°E,34°N,65°N). The domain boundaries are displayed on the map of figure 1. For ACCs, the significance is obtained by a bootstrapping method applied to the ensemble members of the forecasts: we compute the ACCs of 100 draws among the 10 (11) members of the CNRM (ECMWF) forecast and consider the forecast skillful if at least 95% of the 100 ACCs exceed zero. For the sake of convenience, our definition of skillful individual forecasts is arbitrary, and should be understood as “forecasts with the highest ACCs”. It does not imply that they systematically outperform climatological forecasts. This point is addressed by means of the probabilistic skill evaluation (see below). The root mean square error (RMSE) measuring the distance between the ensemble mean forecast and observation regardless of the sign of the anomaly has also been computed for individual forecasts, and normalized by the interquartile range of the observation. However~~the RMSE normalization method is arbitrary and~~ , this score has only been used to confirm a result found with the ACC in section 3.1.

In addition to deterministic scores, the ensemble forecasts can be evaluated by means of probabilistic skill metrics. The continuous ranked probability score (CRPS) is the quadratic difference between the cumulative distribution function (CDF) of an ensemble forecast and the empirical CDF of the observation. The smaller the CRPS, the more accurate the forecast. Let $F(x)$ be the forecast CDF for the variable x (e.g. weekly-mean 2-meter temperature), and y the corresponding observation, then the analytical expression of the CRPS is :

$$CRPS = \int_{\mathbb{R}} (F(x) - \mathbb{1}(x \geq y))^2 dx \qquad (1)$$

where $\mathbb{1}$ is the indicator function.

It is also insightful to compute a continuous ranked probability skill score (CRPSS) for a dynamical forecast system by comparing its CRPS ($CRPS_f$) with that of a climatological forecast ($CRPS_c$) so that :

$$CRPSS = 1 - \frac{CRPS_f}{CRPS_c} \qquad (2)$$

CRPSS ranges between -∞ and 1, 1 corresponding to a perfect forecast. Negative CRPSS values indicate that dynamical forecasts are less accurate than climatological forecasts. In this study, we consider 16 forecasts per winter over 20 years. Therefore, for the n^{th} ($n \leq 16$) forecast of a given year, the corresponding climatological forecast consists in a 19-member ensemble forecast grouping the n^{th} forecasts of the 19 other years. To take into account the differences in ensemble size between the forecasts and their corresponding climatological forecasts, a so-called 'fair' version of the CRPSS is computed, via an unbiased estimator for the score that would be obtained as the ensemble size increases to infinity (Ferro et al., 2008; Ferro, 2014).

2.3 Weather regimes and NAO index

The computation of weather regimes is performed on the ERA5 1979-2017 extended winter, i.e. the months of November to March (hereafter NDJFM). It consists in a k-means clustering of daily maps of sea-level pressure (SLP) anomalies of the North-Atlantic Europe (NAE) domain defined by the boundaries 90°W,30°E,20°N and 80°N. In order to facilitate this clustering, an Empirical Orthogonal Function (EOF) analysis is applied to the SLP anomaly maps, for which the 19 leading modes are retained, explaining more than 90% of the SLP variance. The four resulting clusters correspond to the typical North-Atlantic weather regimes widely described and used in the literature (e.g. Michelangeli et al., 1995). By decreasing order of frequency, these regimes are identified as positive phase of the North-Atlantic oscillation (NAO+), Scandinavian blocking (BLO), negative phase of the North-Atlantic oscillation (NAO-) and Atlantic ridge (AR). Each winter day of the reanalysis and the model simulations is then assigned to the weather regime for which the root mean square distance between the regime centroid and the map of SLP anomaly is minimal. Note that in this study, we have also tested a similar approach with a regime persistence criterion. More precisely, only sequences of 3 days or more corresponding to the same weather regime are effectively assigned to this regime. The impact of this persistence criterion is discussed at the end of section 3.2.

The assessment of teleconnections is facilitated by the use of a NAO index that quantifies this oscillation. Here, it is calculated as the normalized time series of the first principal component, resulting from the projection of the daily ERA5 SLP anomaly field on the leading EOF. For further robustness and because there are multiple ways to define the NAO (Pokorná and Huth, 2015), a comparison is made with another NAO index computed independently by the U.S. National Oceanic and Atmospheric Administration (NOAA) (NOAA Climate Prediction Center NAO index, 2020) on 500 hPa geopotential height fields from the NCEP/NCAR reanalysis and using a different method (Barnston and Livezey, 1987). Despite the many differences between the two daily NAO indices, their correlation for NDJFM 1979-2017 is as high as 0.77.

3 Results

In this section, we start with a general skill assessment to obtain a compared overview of the model ability to predict 2-meter temperature at the subseasonal horizon. The second and third subsections address the question of flow-dependence and the consequences on the forecast skill.

3.1 Skill of the subseasonal forecast systems

3.1.1 Northern hemisphere assessment

The pointwise Pearson correlation between forecasts and observation is shown for week 1 to week 4 forecast times in figure 1.

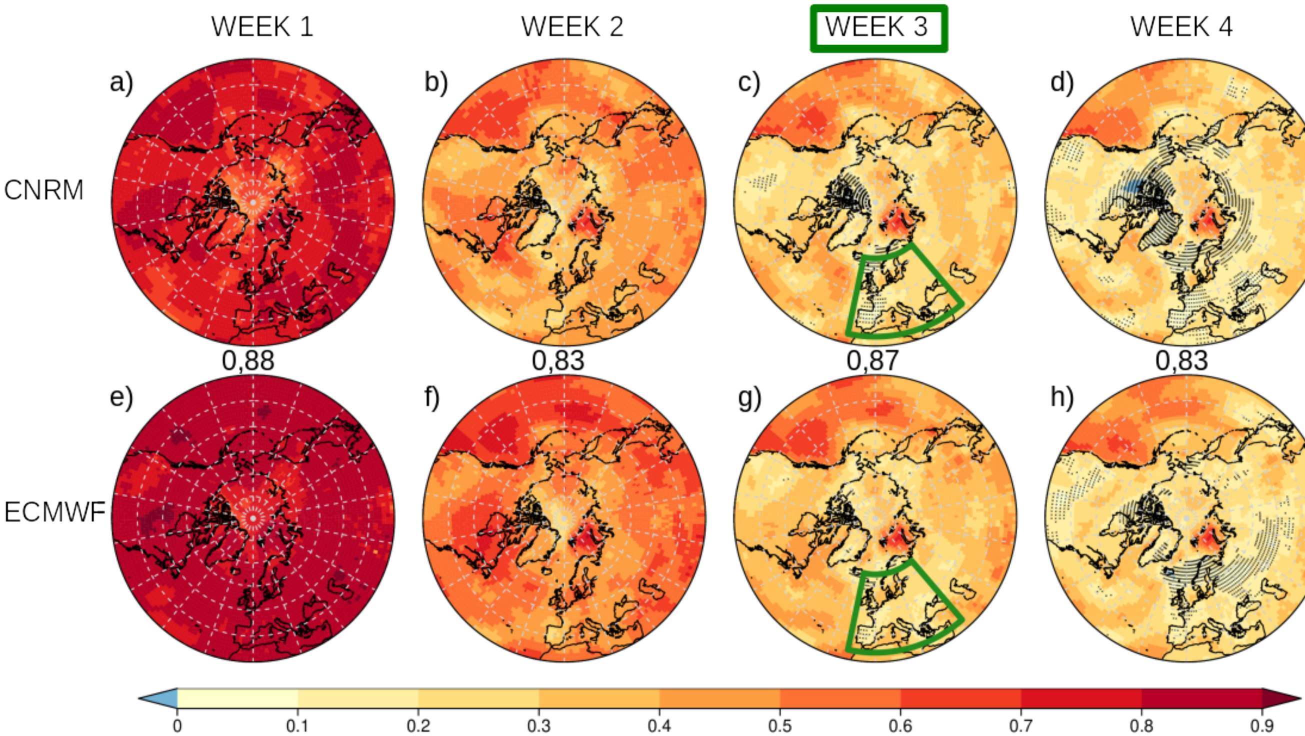


Figure 1. Correlation between week 1 to week 4 2-meter temperature forecasts and the corresponding observation for CNRM (a to d) and ECMWF (e to h) forecast systems. Stippling indicates grid points where correlation is **not** significantly positive at the 95 % confidence level. The numbers show the spatial correlation between CNRM and ECMWF maps for each week. Green boxes indicate the focus region (EUR) and forecast lead-time (week 3) targeted in section 3.1.2.

It clearly shows for both systems the sharp decrease of skill after week 1, and also the better performance of the ECMWF system for the 4 weeks. This result was somehow expected given the much finer spatial resolution of the ECMWF system (Vitart, 2017). The skill difference could also originate from the better fit between the ECMWF forecast system and the ERA-Interim initial conditions, derived from another version of the same IFS model, in particular for the land surface slow-evolving components such as snow cover, soil moisture and deep soil temperature. Nonetheless, discussing the impact on skill of ECMWF and CNRM modelling and forecasting strategies is out of the scope of this study.

For both models, the correlation at week 3 remains positive for large parts of the Northern hemisphere extratropics, albeit weakly over continents. At week 3 and 4, the ECMWF forecasts still show significant correlation over most of Europe, while this is only true over Eastern Europe for CNRM. Overall, while ECMWF exhibits higher skill than CNRM, the large-scale patterns of gridpoint correlation are strikingly similar between both models, as confirmed by the high values of spatial correlations reported on Figure 1.

However, positive correlations do not guarantee that these forecasts are more useful than a ~~naive~~ simple climatological forecast. To document this issue, we compare the CRPS probabilistic score with that of a climatological forecast, by means of the fair CRPSS (see section 2.2). On these maps (fig. 2), white and blue shadings indicate regions where the forecasts do not perform better than the climatology. This score highlights the much better performance of ECMWF over CNRM as early as week 2. The skill patterns look like those found in the correlation analysis, but they are more drastic. For example at week 3, over Europe, the CNRM system shows only remnant skill near the Baltic sea, and the ECMWF over the North of the continent as well as a limited portion of Central Europe. The contrast between the two systems is even more striking over North America. The comparatively poor CRPSS of CNRM could be the consequence of a lack of ensemble spread, resulting in a too narrow distribution of forecasts, which denotes overconfident predictions. The complementary analysis shown in Appendix A, which compares the intra-ensemble standard deviation of the two systems from week 1 to week 4, ~~tend~~ tends to confirm this hypothesis.

Interestingly, the systems remain relatively skillful over the Mediterranean sea but also the sea of Okhotsk, the Kara, Barents and Labrador seas, and, to a certain extent, the Baltic sea. This could be a consequence of persisting sea-surface temperature (for the Mediterranean) and sea-ice extent (for the Arctic and North Atlantic seas) anomalies leading to enhanced subseasonal predictability to the near-surface atmosphere (Chevallier et al., 2019; Bach et al., 2019), although indisputable evidence would require a dedicated study.

From now on, our work focuses on the predictability of week 3 only.

3.1.2 Focus on Europe

The forecast skill of EUR 2-meter temperature is assessed from the 320 reforecasts at week 3 for both systems, by means of the ACC. This score varies considerably between dates. Thus, in order to investigate the degree of consistency between models forecast skill, Figure 3 plots the distribution of ECMWF ACC against CNRM ACC over EUR, for each of the reforecast dates. Dots depict the 320 reforecasts and filled contours the corresponding probability density function. Red dots show the reforecasts where ACCs are significant at the 95 % level for both systems. This distribution is fairly symmetric, albeit slightly skewed towards higher values for ECMWF, which is consistent with results found in the previous section. This is also revealed by the mean and median points (black and grey triangles, respectively), located slightly above the $y = x$ identity line. The standard deviation of ACCs is similar (0.42 for CNRM vs. 0.40 for ECMWF). More interestingly, the correlation between CNRM and ECMWF ACCs reaches 0.52. The correlation is even higher (0.61) when considering the RMSE of the individual forecasts instead of the ACC (not shown). The scatter plot also reveals that the most skillful concurrent forecasts (red dots) are

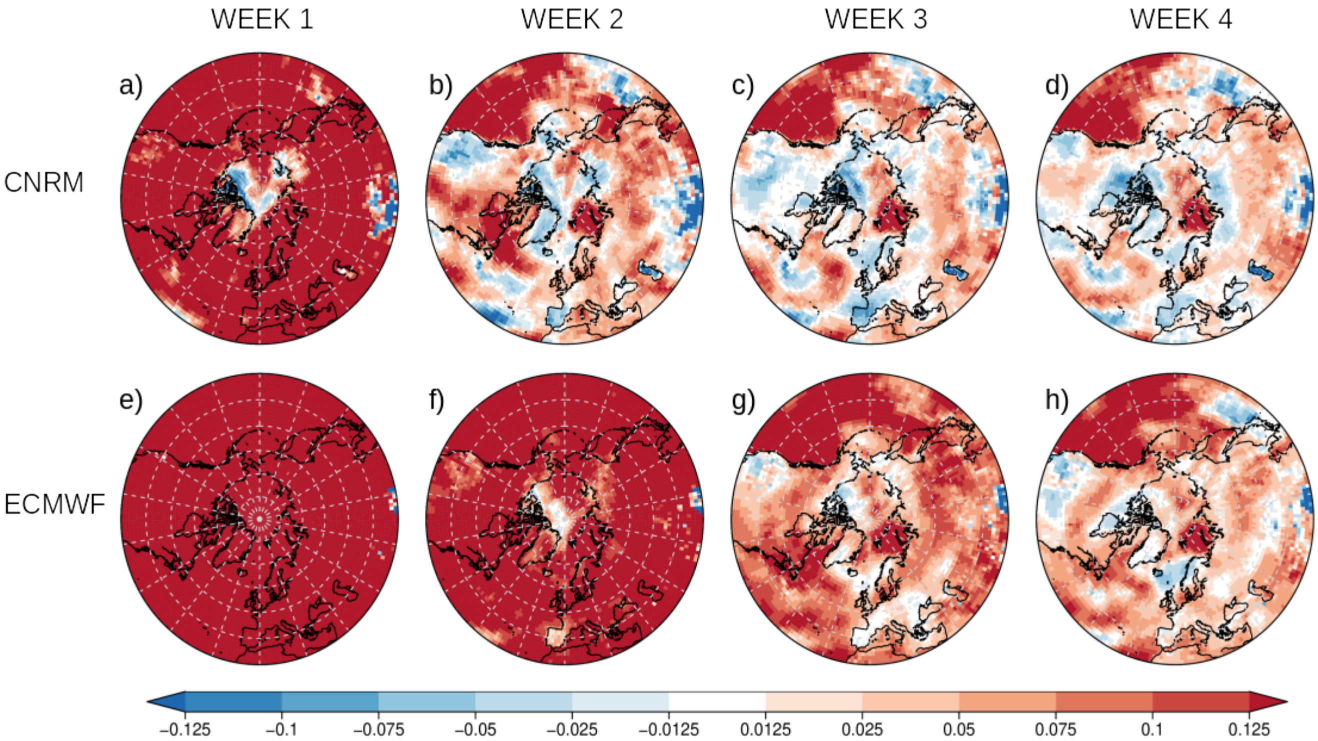


Figure 2. Fair CRPSS for week 1 to week 4 forecasts for CNRM (a to d) and ECMWF (e to h) forecast systems, against climatological forecasts (see text). Red shades indicate that the actual forecasts are more skillful than the climatological counterparts.

less scattered and more grouped along the $y = x$ identity line than other forecasts that are more spread out. They correspond to the maximum of the probability density function, ~~figured~~ plotted in green and yellow shades. This suggests that high skill forecasts contribute more to the correlation than low skill counterparts. In other words, CNRM and ECMWF systems are more prone to issue concurrently good forecasts than concurrently poor ones.

The synchronicity found in the level of skill between the CNRM and ECMWF week 3 forecasts therefore indicates the existence of a common source of predictability concerning the EUR region.

We now investigate the distribution of skillful forecasts along the 20 year period considered in this study. The barplots in figure 4 show a relatively consistent ~~trend and~~ interannual variability : the number of yearly skillful forecasts for ECMWF, in red, is significantly correlated to that of CNRM, in blue (r=0.61). ~~Linear trends have a similar slope, as reported on the figure.~~ We reprocessed figures 3 and 4 after removing a linear trend derived from the DJFM ERA5 2-meter temperature averaged over the Europe domain. We found no significant changes in the ACC distribution and correlation (0.519 instead of 0.521), nor in the interannual variability of skillful forecasts (not shown). Limited changes in the number of significantly positive ACCs per

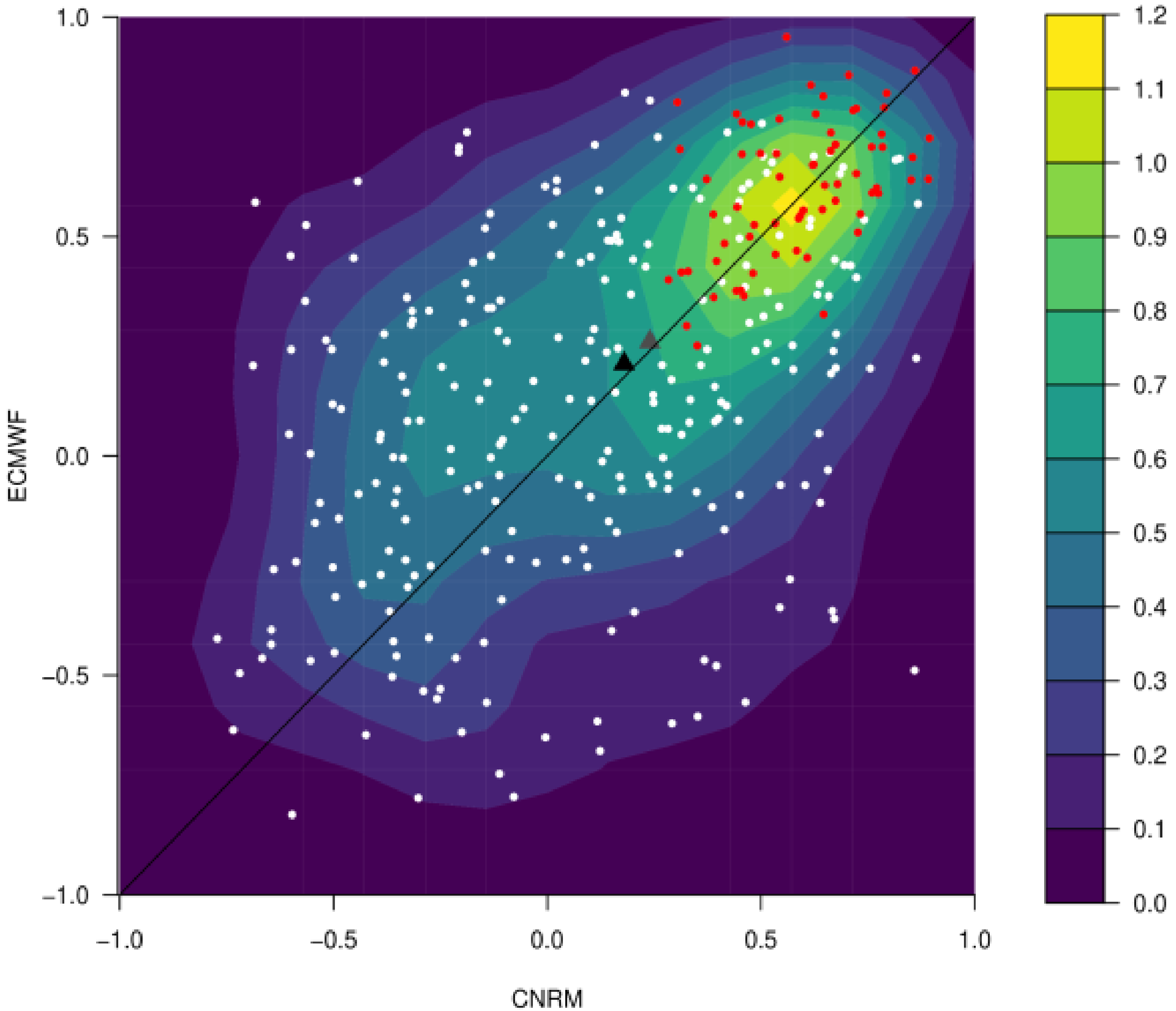


Figure 3. Scatter plot (dots) and probability density function (contours) of ECMWF ACC in function of the corresponding CNRM ACC for each of the 320 wintertime reforecasts of EUR 2-meter temperature at week 3. Red dots mark ACCs significant at the 95 % confidence level for both CNRM and ECMWF. The black and grey triangles correspond to the mean and the median point respectively, and the black solid line to the $y = x$ identity line

year and per forecast system before and after detrending confirm the minimal influence of the warming trend on the forecast skill.

In any case, 2009-2010 stands out as the winter with the maximum number of skillful ~~forecast~~ forecasts of the 20-year period for both CNRM and ECMWF systems either considered jointly (green bars) or separately (blue and red bars). Since that winter is characterized by a record-breaking negative NAO index (Cattiaux et al., 2010), we have computed the correlation between the yearly number of skillful forecasts and a winter mean NAO index (December-to-March) derived from our daily NAO index datasets. The correlation is not significant if the NAO index is computed by averaging daily NAO indices (not shown). However, when computed as the mean of daily absolute values of the ERA5 NAO index (brown broken line in fig. 4), the correlation found is significant. This is also true with the NOAA NAO index, with r ranging from 0.44 to 0.66. This result suggests that S2S EUR forecasts are more frequently skillful during winters characterized by a strong NAO index, either positive or negative.

Therefore, the next section focuses more specifically on the relationship between forecast skill and weather regimes.

3.2 Relationship between forecast skill and initial weather regime

We now consider the first 4 days after initialization as a relevant time window to discuss about initial weather regime. We argue that the choice of using 4 days instead of the single first day allows more robustness, since the latter may sometimes lie in between two different regimes. This 4-day window is also consistent with the S2S convention that defines the first forecast week as starting from day 5 onwards (see section 2.1).

We count separately for each member the occurrence of each weather regime assigned to the first 4 days of the forecast members, among the 68 EUR forecasts out of 320, that are concurrently skillful for CNRM and ECMWF.

In this sample of 68 skillful reforecasts, the frequency of initial NAO+ days is significantly higher, and that of initial BLO days lower than in the 252 other reforecasts, for both forecast systems (Table 2). The frequency of NAO- initial days is also higher in CNRM but not significantly for ECMWF.

Table 2. Initial weather regime frequency in % of skilful forecasts over EUR. Numbers in parentheses indicate the frequency for all the other forecasts~~and bold~~. Bold characters highlight where ~~these frequency~~ the regime frequencies of skilful forecasts are significantly different from those in parentheses at the 95 % confidence level as determined by bootstrap.

Weather regime	NAO+	BLO	NAO-	AR
CNRM	**39.2 (26.5)**	**~~20.4~~ 14.9 (30.5)**	**29.6 (22.6)**	16.3 (20.4)
ECMWF	**38.2 (27.8)**	**17.5 (26.8)**	27.8 (23.1)	16.5 (22.3)

If skillful forecasts tend to start more frequently with NAO conditions, we would like to verify the reciprocal, i.e. how skillful the forecasts starting with NAO conditions are. To this end, instead of subsampling the forecasts according to their level of skill, we now cluster the 320 forecasts in 4 groups determined by their initial weather regime and compare the mean

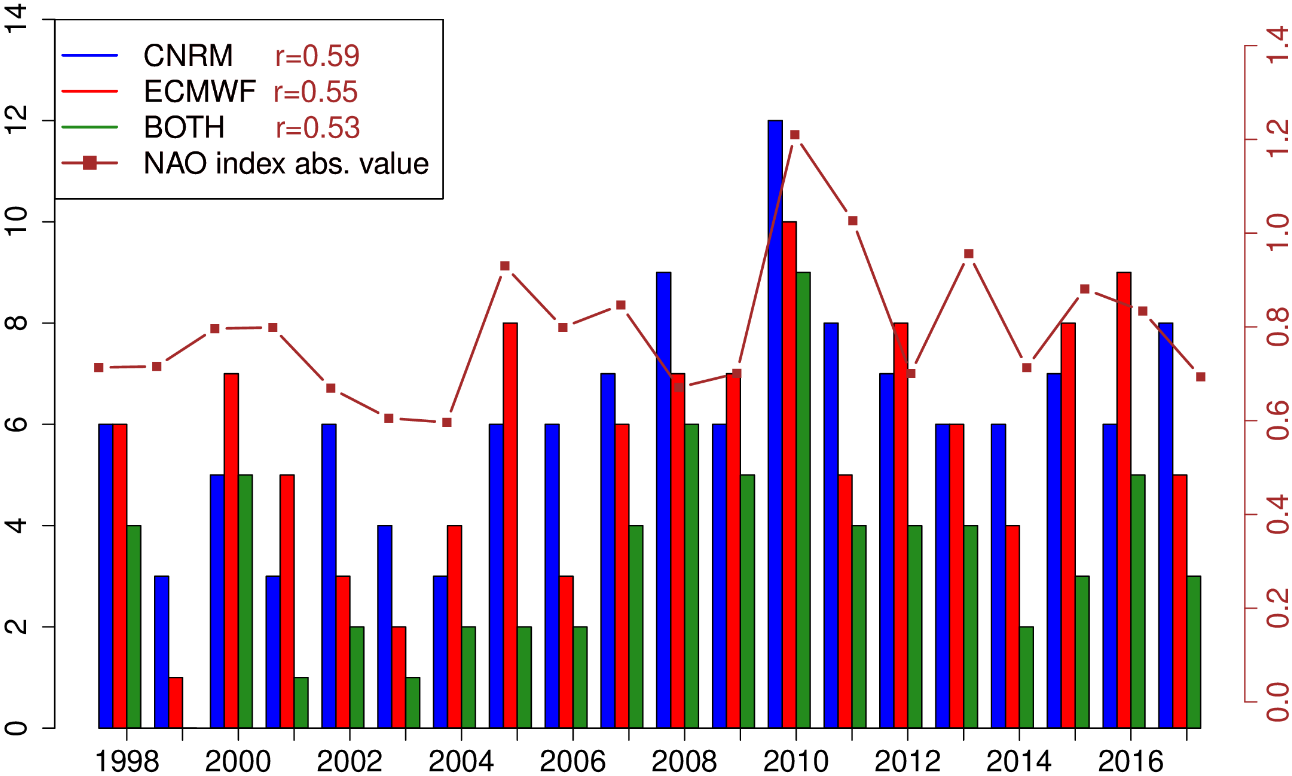


Figure 4. Yearly number of skillful forecasts for CNRM (blue), ECMWF (red) and both systems (green) computed on EUR week 3 temperature forecasts. The ~~dashed lines mark the respective linear trend whose slope value 'a' is reported in the legend. The~~ brown broken line shows the absolute value of the winter NAO index derived from ERA5 (see text). The 'r' values reported in the legend correspond to the correlation of this index with the yearly number of skillful forecasts.

skill evolution along the forecast time for each of these clusters (fig. 5). We define the initial weather regime of each forecast as the regime with the greatest number of occurrence during the 4 initial days.

Here the significance level is obtained by means of a bootstrapping method. More precisely, for a cluster of size N, a probability density function of mean ACC is built out of 1000 draws with replacement of N forecasts within this cluster. The forecast is then considered significant if the first percentile of the distribution is positive.

For both systems, the mean ACC of the forecasts initialized in NAO+ conditions becomes higher than those initialized with other regimes by day 6 and more markedly from day 15 onwards, albeit not significantly (not shown). The difference vanishes past day 25 for CNRM but not for ECMWF. Finally, the ACC remains significantly positive until the end of the forecast period

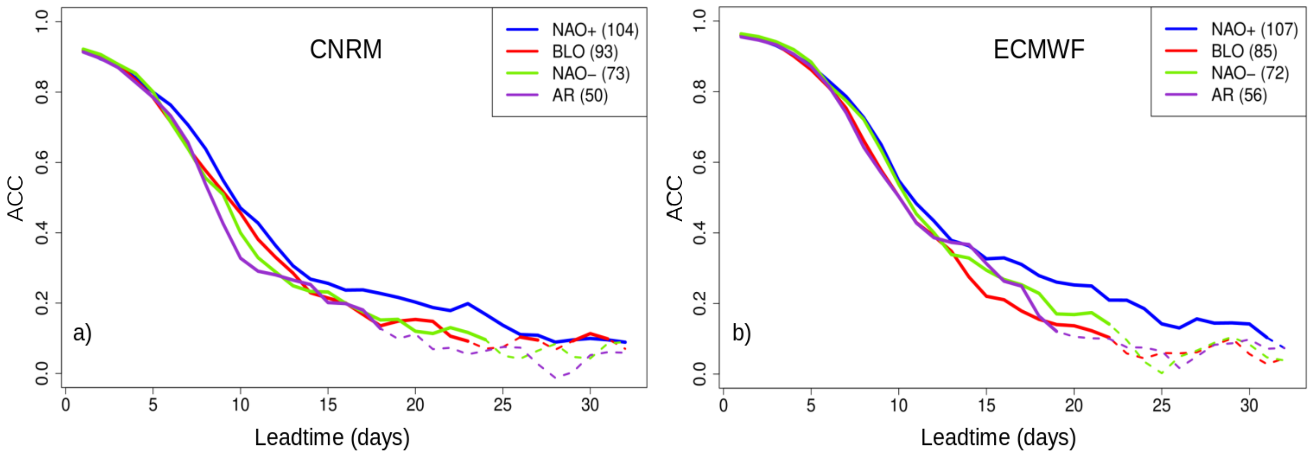


Figure 5. Mean ACC evolution with forecast time over Europe by initial weather regime for (a) CNRM and (b) ECMWF. Solid lines indicate values significantly positive at the 99 % confidence level. The number of forecasts for each initial regime is reported in parentheses.

in both systems although a positive ACC does not necessarily imply that the forecasts are useful, as discussed in section 3.1. It is also interesting to notice that NAO- conditions do not lead to particularly skillful forecasts at week 3 for CNRM, as could have been expected from table 2, and that models agree upon AR being the worst initial weather regime in terms of temperature subseasonal predictability over Europe, since the mean ACC of forecasts initialized thereby are no longer significant past day 18 or 19.

The next question arising from the previous results is the evolution in time of the regime frequency among these forecasts initiated under NAO+ or NAO- conditions. The stacked bar plots on figure 6 illustrate this evolution. Note that the residual non-NAO+ (resp. non-NAO-) regimes showing in the 4 initial days simply result from our clustering method based on the predominant regime counted within the 4 initial days of all ensemble members, thereby leaving some room for the occurrence of other weather regimes. We find that despite a rapid decrease of the NAO+ regime proportion with forecast time, it remains slightly larger than the climatological one at week 3. A similar but more pronounced result is found for NAO-. This suggests that the NAO regimes are persistent in the forecasts although this cannot be ascertained at this stage since no statistical significance test has been performed here, and furthermore, all the ensemble members are pooled together, which conceals the transitions between weather regimes.

Before exploring further the causes of the above results, we need to address in more detail the question of regime persistence. So far, every forecast or observed day has been assigned with one of the 4 weather regimes, regardless of the day-to-day variability of the regime sequence. Such variability occurs when the spatial distribution of high and low pressure systems of a given day does not match well any of the canonical weather regimes, or corresponds to a transition between two of them.

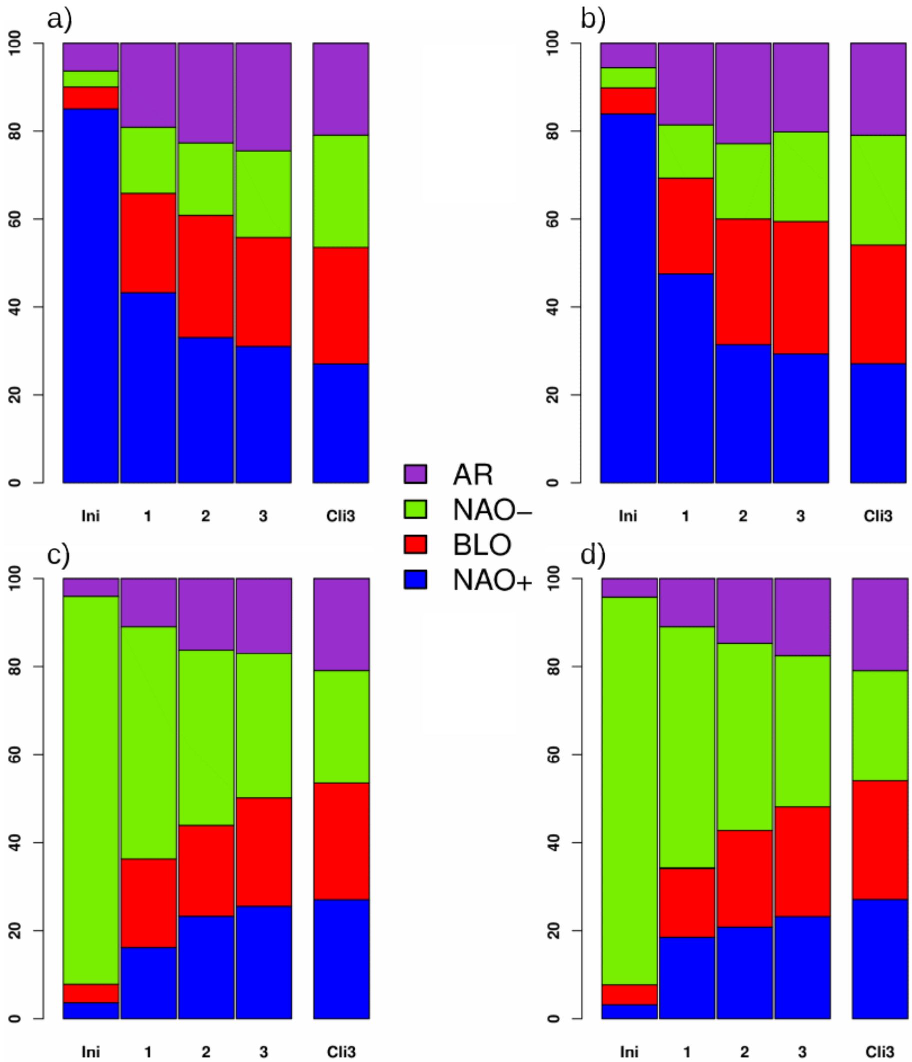


Figure 6. Weekly evolution of regime frequency among forecasts initialized in NAO+ (a) or NAO- (c) conditions for CNRM. (b) and (d) same as (a) and (c) for ECMWF. The leftmost bar corresponds to the 4 initial days. The rightmost bar corresponds to the climatological frequency for week 3.

To overcome this issue, we have defined a fifth category (called ’NONE’) assigned to days outside any sequence of 3 or more days belonging to the same weather regime. Excluding the forecasts initialized with the predominant ’NONE’ category results in 4 smaller clusters of forecasts. Nonetheless the mean ACC evolution is not dramatically changed, and the ACC dependence on the initial conditions lead to the same hierarchy of weather regimes as can be seen in Appendix B. Similar conclusions can be drawn regarding the weekly evolution of regime frequency after taking the ’NONE’ category into account. Given the limited impact of the screening based on regime persistence, the following sections rely on the original daily weather regime assignment, i.e. without the 'NONE' category.

3.3 Evidence of a lagged teleconnection

3.3.1 Composite analysis

The previous section has pinpointed a slight distort in the weather regimes distribution at week 3 for forecasts initialized in NAO conditions. For a broader comprehension, we thus compute the spatial composites of week 3 anomalies for this subset of forecasts for sea-level pressure (fig. 7a and b) and 2-meter temperature (fig. 7e and f). Given that 87 (93) forecasts out of 320 are concerned for CNRM (ECMWF) and that each of them comprises 10 (11) members, the composite maps result from the average of n=870 (n=1027) single realizations. The ECMWF and CNRM composites show some similarities over the Atlantic sector with a distinct low pressure anomaly over the Arctic and high pressure anomaly centers near the Azores archipelago. The temperature patterns are even closer to each other with a large scale warm anomaly stretching from Central Europe to Eastern Siberia and a cold anomaly over Canada, more pronounced near the Labrador Sea. The main difference between ECMWF and CNRM concerns the sea-level pressure anomaly over Europe, which barely reflects onto the temperature anomaly. If we consider the positive pressure anomaly over the North Pacific, it may remind of the Arctic Oscillation (AO) loading pattern (e.g. fig.1 in Thompson and Wallace, 1998), although this anomaly is not significant for CNRM, and more importantly not consistent with observations (see below).

The patterns found could be specific to the forecast systems, i.e. GCMs constrained by imposed initial conditions and external forcing affected by a strong ~~anthropic~~ anthropogenic imprint. To verify this hypothesis, we derive a set of single-member pseudo-forecasts from the CNRM-CM6-1-HR 300-year-long piControl simulation. For each simulated year, we extract sixteen 32-day time series starting every seven days from Nov. 13th to February 26th, so as to mimic successive S2S forecast start dates. Among the 4784 resulting pseudo-forecasts, those having ~~all 4~~ a majority of initial days assigned as NAO+ are sampled to compute sea-level pressure and 2-meter temperature anomaly composites (fig. 7c and g). In this case, it concerns n=~~579~~ 957 realizations. We proceed likewise for the 1950-2017 ERA5 reanalysis (fig. 7d and h), in order to compare the realism of this behaviour with respect to observation. This long ERA5 period is a trade-off between a sufficient sample size, requiring more than 20 years to be comparable with reforecast composites, and a stable structure of weather regimes given the decadal variations of the NAO (e.g. Jung et al., 2003; Woollings et al., 2015). However, despite long term shifts in the centre of actions occurring during the 1950-2017 period, the main features of NAO regimes, characterized by the Eurasia/Canada temperature dipole and a North Atlantic meridional pressure gradient, are preserved.

The piControl composite shows broadly consistent patterns over the mid-Atlantic notwithstanding differences in terms of relative amplitude and extent of pressure anomalies. For temperature, the warm anomaly over Southeastern US is somewhat stronger than in the forecast systems. The amplitude of the ERA5 composite patterns is generally larger, which is at least partly explained by the reduced size of the composite sample (n=~~148~~203). This observational composite shows a larger extent of the Atlantic high pressure belt also covering Southern Europe and central Asia, and conversely no high pressure anomaly over

North Pacific, which ~~tends to confirm that the similitude with a~~ diverges from the hemispheric positive AO loading pattern evoked earlier~~is a model artefact~~. In terms of temperature, the main difference is the greater extent of the warm anomaly over North America and the cold pattern near Bering strait, with respect to the forecast and piControl composites.

A similar composite analysis has been carried out with (pseudo-)forecasts initiated in NAO- (fig. 8). Here the pressure and temperature composites show even more similarities between forecasts and observation, in particular over the North-Atlantic-Europe region. Similar to NAO+, surface pressure patterns show more differences than temperature, in particular over East Siberia, West Pacific and North America. In the piControl composite, again, the patterns found are less intense but very consistent for temperatures, less so for surface pressure. One explanation for this reduced consistency could be that in the piControl time series, boundary conditions such as the ocean, sea-ice and stratosphere also influencing the atmospheric flow, have no reason to be coherent with observation, unlike the forecast composites initialized with reanalyzed atmospheric boundary conditions.

To summarize, this composite study reveals some significant agreement between forecast systems, unforced GCM and reanalysis as to prevailing atmospheric flow and near surface temperature anomalies during the third week following NAO conditions. This agreement is much better for negative than positive NAO, for temperature than pressure patterns, and for the North-Atlantic (pressure), Labrador, Europe and Siberia (temperature) regions. The lesser agreement found for NAO+ could relate to our clustering methodology, as discussed in the conclusion.

3.3.2 Observational NAO index

At this point, our study has only considered a weather regime assignment based on a root mean square distance criterion but this method may conceal a wide array of a atmospheric situations. Here we make use of the NAO index that quantifies the amplitude of the oscillation, and allows to identify periods of intense NAO+ or NAO- conditions. The composite analysis suggests that NAO initial conditions lead to NAO-like atmospheric flow. To verify this, we evaluate the extent to which the NAO index decorrelates with time in the observation. More precisely, ~~the~~ figure 9 depicts the correlation of the averaged day-1-to-day-4 NAO index with a sliding window of 7-day running mean NAO index. The grey line and envelope consider the 608 aforementioned time series (that is, 16 per winter of the 1979-2017 period) whereas the red counterparts only consider the 10 % characterized by the highest absolute value of the initial NAO index. Such screening selects the time series with an initial atmospheric flow characterized by intense NAO+ and NAO- conditions. Despite different datasets and methodology, the NAO decorrelation compares well to other studies when keeping the whole sample (Keeley et al., 2009) with a characteristic decorrelation time of 8 to 10 days. However, the decorrelation is much slower when considering only the subsample with intense NAO initial conditions. Keeley et al. (2009) also identified a similar "shoulder" or "rebound" in the NAO decorrelation function between 10 and 30 days and find it largely related to interannual variability, as opposed to intraseasonal. The decorrelation timescale and behavior are consistent when evaluated over a wide range of different NAO indices (fig. 3b in Domeisen et al., 2018). The overlap between confidence intervals indicates that the difference found is not significant beyond 10 days when the NAO index is derived from ERA5 sea-level pressure. However it remains largely significant for the NOAA NAO index based on 500 hPa

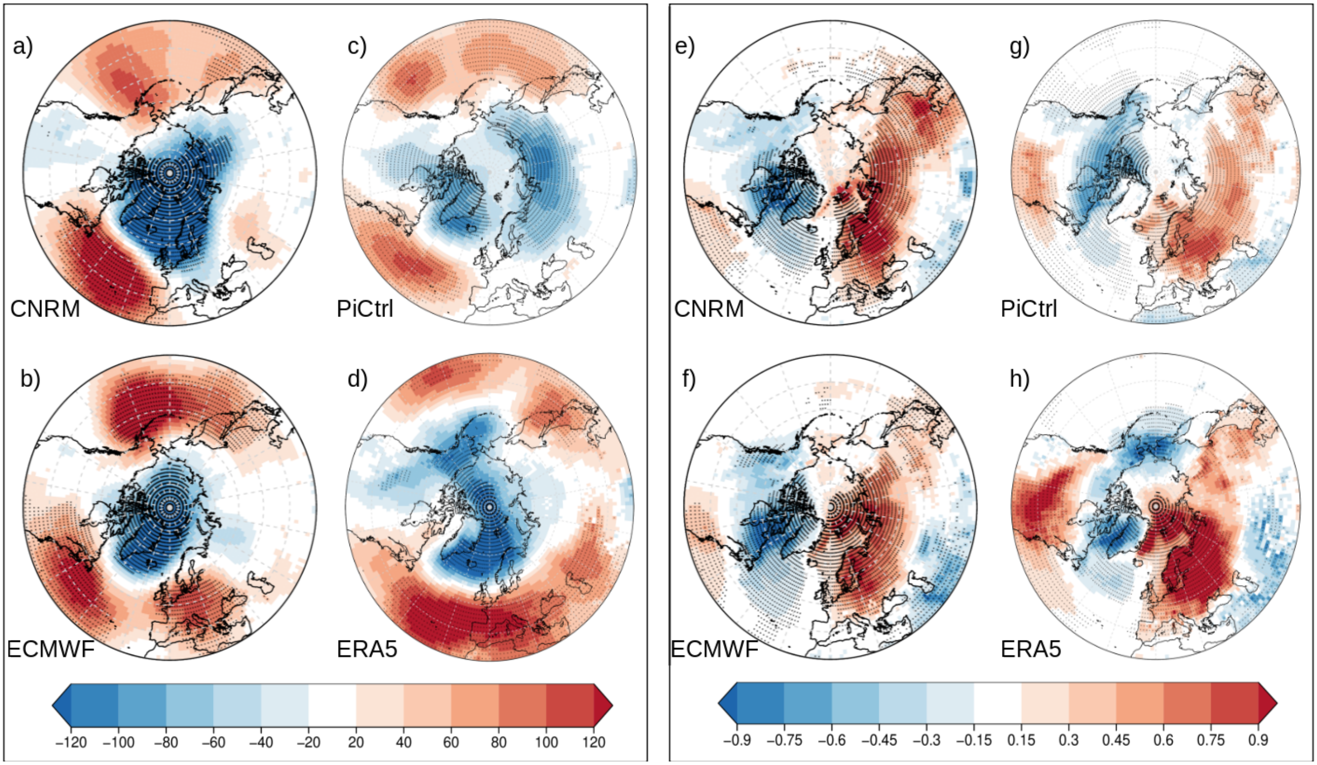


Figure 7. Composite anomaly of Week 3 sea-level pressure (in Pa) following NAO+ initial conditions in (a) CNRM forecasts (*n*=870) (b) ECMWF forecast (*n*=1027) (c) CNRM piControl pseudo-forecasts (*n*=~~579~~957) and (d) ERA5 reanalysis (*n*=~~148~~203). (e) to (h) like (a) to (d) for 2-meter temperature (in K) initial conditions. Anomalies statistically significant at the 95% level are stippled.

geopotential height, in particular three weeks after initialization where the correlation peaks up. For that matter, the sensitivity of NAO persistence to the NAO index definition is consistent with previous findings (Domeisen et al., 2018). Regardless of of the NAO index calculation method, our results provide observational evidence of a long-lasting persistence of NAO-like atmospheric flow in winter.

Finally, still with this observational subsample of ERA5 time series characterized by intense "NAO-like" initial conditions, we regress the week 3 pointwise 2-meter temperature onto the initial NAO index (fig. 10). Whether derived from ERA5 or NOAA, the patterns show similarities, with a stretch of positive correlation extending from South East US to Siberia with maximum values near the Baltic Sea, and two negative correlation lobes over Greenland / Labrador sea, and from the tropical North-Atlantic to North Africa and the Middle-East.

Given that the spatial extent of these correlation patterns encompasses large parts of the northern hemisphere, we will now evaluate if NAO initial conditions of winter subseasonal forecasts could translate into enhanced prediction skill beyond Europe, and how this relates to the regression patterns described above.

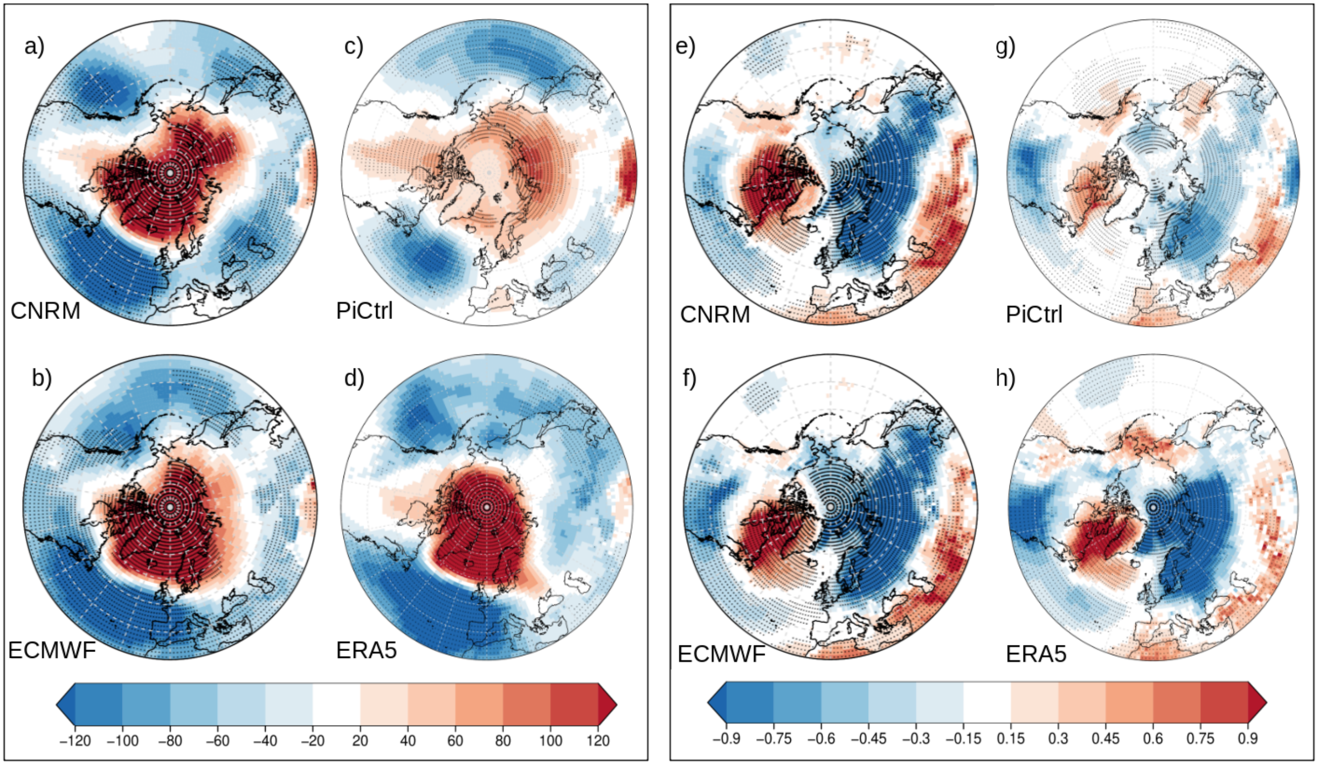


Figure 8. Same as fig. 7 for NAO- initial conditions.

3.4 Consequences on forecast skill outside Europe

In the previous section, we have identified a ~~lagged NAO teleconnection prone to impact temperatures over a large fraction of~~ statistical link between wintertime temperature anomalies over a number of regions of the northern hemisphere extratropics and the 3-week antecedent NAO index. We now return to the forecast skill evaluation, but this time, we proceed to a subsampling of the reforecasts based on two conditions: the initial weather regime and its intensity. More precisely, we select all the reforecasts initiated in NAO+ and NAO- and evaluate their initial NAO index from the NOAA dataset. We then retain only the "initial NAO+" ("initial NAO-") reforecasts for which the initial NAO index belongs to the upper (lower) quartile of the distribution. The choice of this percentile results from a trade-off between the strength of the initial NAO signal and a sufficient sample size. Figure 11 shows the week 3 2-meter temperature correlation after subsampling and the correlation difference with respect to the full sample of reforecasts (see fig.1c and g). The correlation patterns are patchier than in figure 1, because the sample size is considerably reduced, that is, 40 reforecasts instead of 320 for each system. Nonetheless, the correlation difference highlights a significantly increased skill over North-West Europe, and Central Siberia, as well as the Labrador seas and the South-East Mediterranean and Middle-East to a lesser extent. These regions match remarkably well with ~~those concerned by the NAO~~

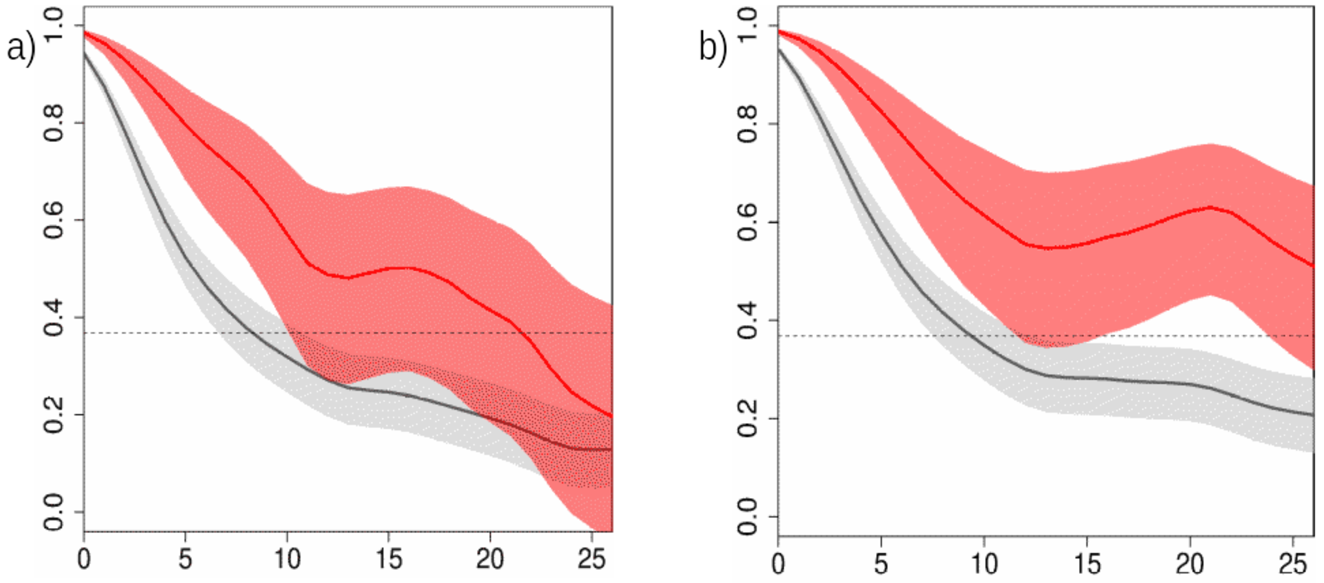


Figure 9. Correlation and 95 % confidence interval (solid line and envelope) of initial NAO index with weekly running mean NAO index derived from (a) ERA5 and (b) NOAA. Grey (red) shades consider the full sample (subsample with intense NAO initial conditions) of time series within the 1979-2017 wintertime period, as described in the text. The 1/e decorrelation threshold is marked with the dashed horizontal line.

~~lagged teleconnection~~ the regression patterns highlighted in observations (fig. 10) and they are relatively consistent between CNRM and ECMWF systems. ~~However~~Note that these regression patterns do not necessarily imply a causal relationship and may originate from a number of remote drivers (which are not investigated here). Indeed, 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~~.

Even if there is no one-to-one relationship between local increase or decrease of the prediction skill and the aforementioned regression patterns, our study reveals consistent evidence that the forecast systems are capable of capturing the lagged NAO ~~teleconnection~~ relationship to a certain extent. This provides additional sub-seasonal predictability at the continent scale, conditioned by the initial atmospheric flow.

In an attempt to better understand if the increased skill following intense NAO conditions was due to extended regime persistence or rather enhanced regime occurrence, we have performed additional analyses, reported and discussed in the supplement to this study. Although these analyses only marginally elucidate the question, they suggest that forecasts initiated in strong NAO- conditions tend do have more persistent NAO- patterns in both CNRM and ECMWF. There is no such evidence for the NAO+ case. It could be that strong NAO+ initial conditions are followed by an increased recurrence of the NAO+ regime along the forecast time, but this feature is only found - arguably not very distinctly - in ECMWF.

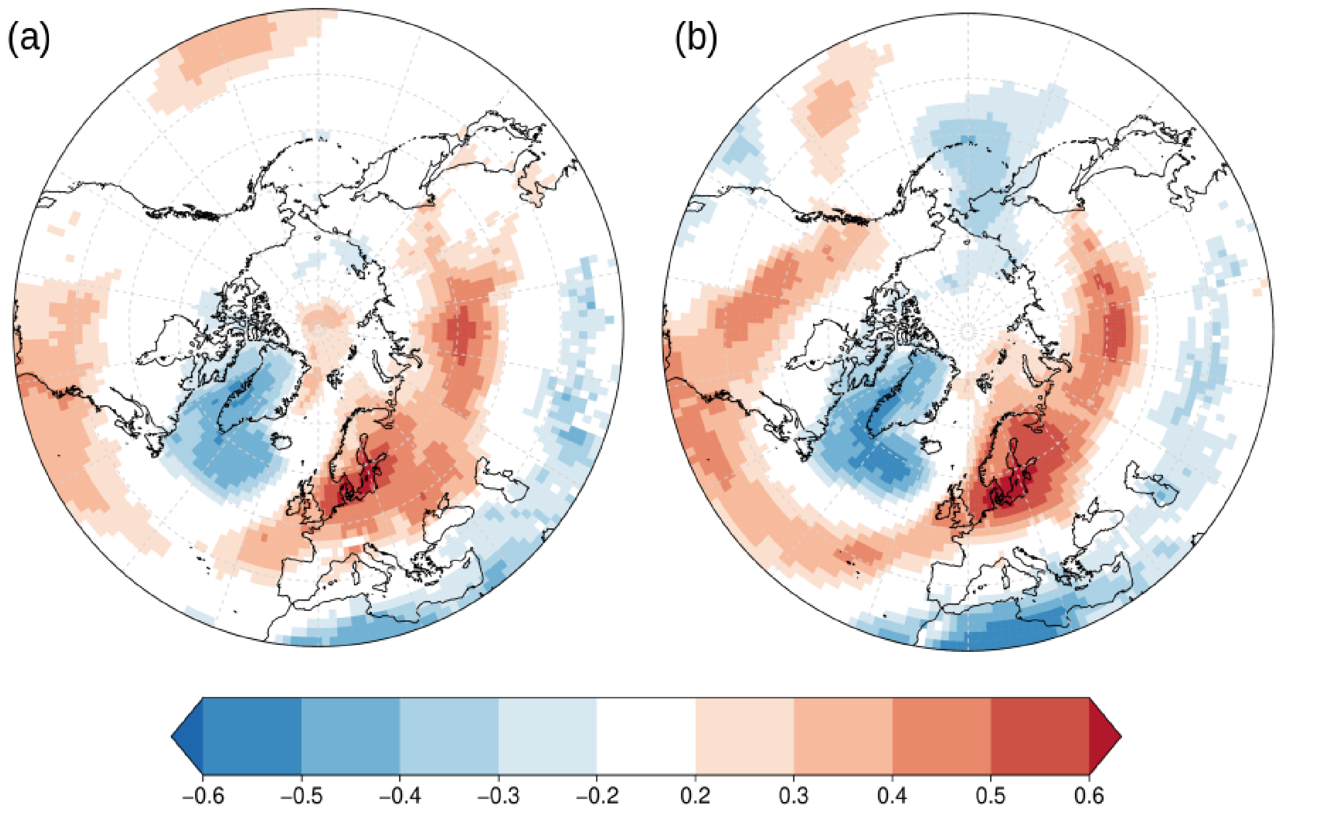


Figure 10. Correlation of ERA5 week 3 2-meter temperature with initial NAO index derived from (a) ERA5 or (b) NOAA. Only values significantly different from zero at the 99% confidence level are displayed.

4 Conclusions

The main objective of this study is to determine if the atmospheric circulation pattern in place at the time of initialization can impact the subseasonal predictive skill of forecasts delivered by state-of-the-art forecast systems. This study focuses on winter northern hemisphere extratropics near-surface temperature reforecasts issued by the new CNRM subseasonal forecast system as well as the ECMWF extended-range forecast system.

A first general skill assessment shows that the CNRM system proves less skillful than the ECMWF counterpart when considering the first 4 weeks after initialization but ~~performs reasonably wellanyway~~the spatial patterns compare relatively well. The ensemble spread of the CNRM forecasts is too weak over much of the Northern Hemisphere across all the prediction horizons, which likely penalizes this system in terms of probabilistic skill.

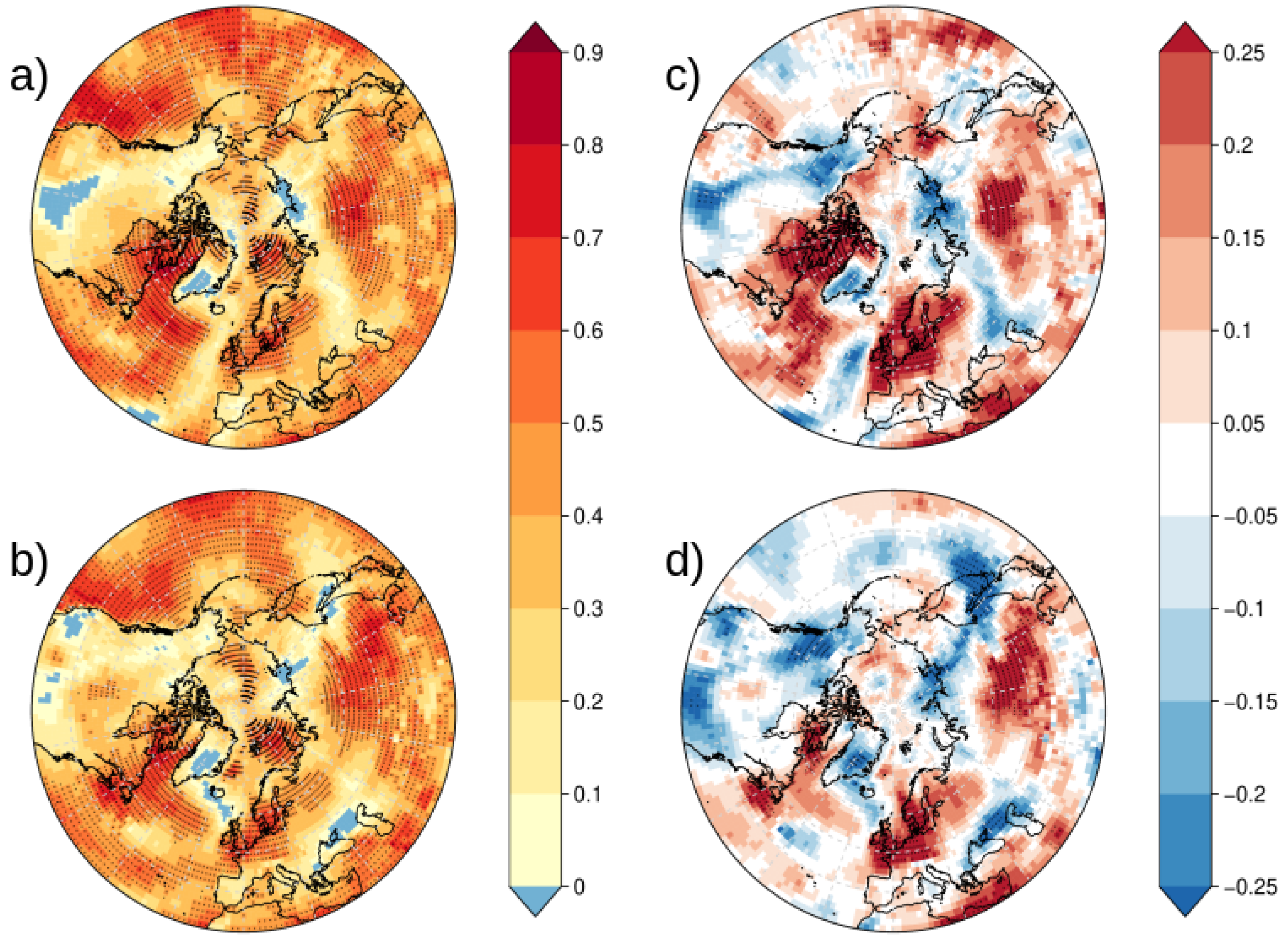


Figure 11. Correlation between week 3 2-meter temperature and the corresponding observation for the NAO initialized CNRM (a) and ECMWF (b) reforecasts. (c) and (d) depict the correlation difference (a) minus Figure 1(c) for CNRM and (b) minus Figure 1(g) for ECMWF. Stippling indicates significant values at the 95% confidence level.

When considering the performances of individual successive forecasts over Europe, the level of skill at week 3 tends to vary concurrently for both systems, thereby suggesting that they benefit from a common and intermittent source of subseasonal predictability. Since the European climate is known to be influenced by the North Atlantic Oscillation (NAO), a weather regime approach has provided evidence that forecasts initialized during positive NAO phases are slightly more skillful over Europe than those issued during the other 3 North-Atlantic weather regimes.

A composite analysis has shown that temperature and sea-level pressure anomalies typical of the positive (negative) NAO ~~regime~~ regimes tend to characterize the third week following the occurrence of such regime. This feature is well captured and comparable to a certain extent in forecasts, pre-industrial climate simulations and observations, particularly for temperature anomalies. The robustness of this time-lagged weather regime impact is further confirmed by the strong and persisting autocorrelation of the upper and lower tail of the NAO index distribution.

Ultimately, we show that the subseasonal predictive skill over Europe is more ~~pre-conditionned~~ pre-conditioned by intense NAO events, either positive or negative, than by the prevailing regime at initialization. We also find that this flow-dependent skill concerns mostly Northern Europe, but also central Siberia and regions surrounding the Labrador sea.

In a next study, it would be worth studying the atmospheric mechanisms involved in this NAO lagged teleconnection, and the extent to which they are properly captured by forecast systems. Such an approach could bring insight about the reasons why the NAO initiated forecasts do not show improved skill over most of Eastern North America, as could have been expected (Luo et al., 2020). At least for the coastal area, recent findings from Roberts et al. (2021) indicate that the skill could be improved by reducing the North-Atlantic sea surface temperature biases resulting from inadequate representation of mesoscale ocean eddies in coupled models. Factors influencing the persistence of NAO+ and NAO- phases should also be investigated to go a step further into the concept of flow-dependent "windows of opportunity" for subseasonal prediction. In particular the influence of sudden stratospheric warming events on the occurrence and persistence of the NAO- regime has been evidenced (Domeisen, 2019). Hence, subseasonal forecasts issued after the onset of such events and characterized by a strong initial NAO phase could be even more trustworthy, although this hypothesis would require a large reforecast dataset to be verified.

Another prospect for future works would be to evaluate the sensitivity of the results to the methodology. First, our strategy to identify wintertime weather regimes, although widely referenced in literature, may not be optimal ~~(Falkena et al., 2020).~~ (Falkena et al., 2020; Dorrington and Strommen, 2020). It could be that our clustering of the North Atlantic circulation into 4 weather regimes leads to NAO+ not being a mode of variability specific enough: it can be seen as a mere generic mode that potentially mixes a variety of distinct weather regimes. The robustness of our results would be worth assessing when considering a ~~different~~ larger or smaller set of weather regimes. Then, the reforecasts clustering strategy could also be questioned. In particular, a distance threshold between sea-level pressure patterns in reforecasts and the weather regime centroids could be applied in order to subsample only those reforecasts initiated in conditions very close the canonical modes of atmospheric variability.

Finally, including more forecast systems for a multi-model approach would bring considerable interest but also a great deal of additional complexity, given the many differences in the design of the S2S forecast systems.

Appendix A: Comparison of the CNRM and ECMWF ensemble spread

Figure A1 shows the weekly evolution with leadtime of the intra-ensemble standard deviation of the 2-meter temperature for the CNRM and the ECMWF subseasonal reforecasts. Since the CNRM ensemble size holds 10 members vs. 11 members for ECMWF, only 10 members of the latter have been used to guarantee a fair comparison of the two systems. The week-by-week differences (bottom row maps) help visualize that the ECMWF ensemble is more dispersive (red shades) than the CNRM counterpart over the vast majority of the Northern hemisphere whatever the prediction horizon. Only the North Pole and to a certain extent South Asia at longer lead times show more spread for CNRM. This lack of spread for CNRM is particularly pronounced over high latitude continents but considering the slow evolution of sea-surface temperature, the lack of spread over oceans is also meaningful and should not be overlooked.

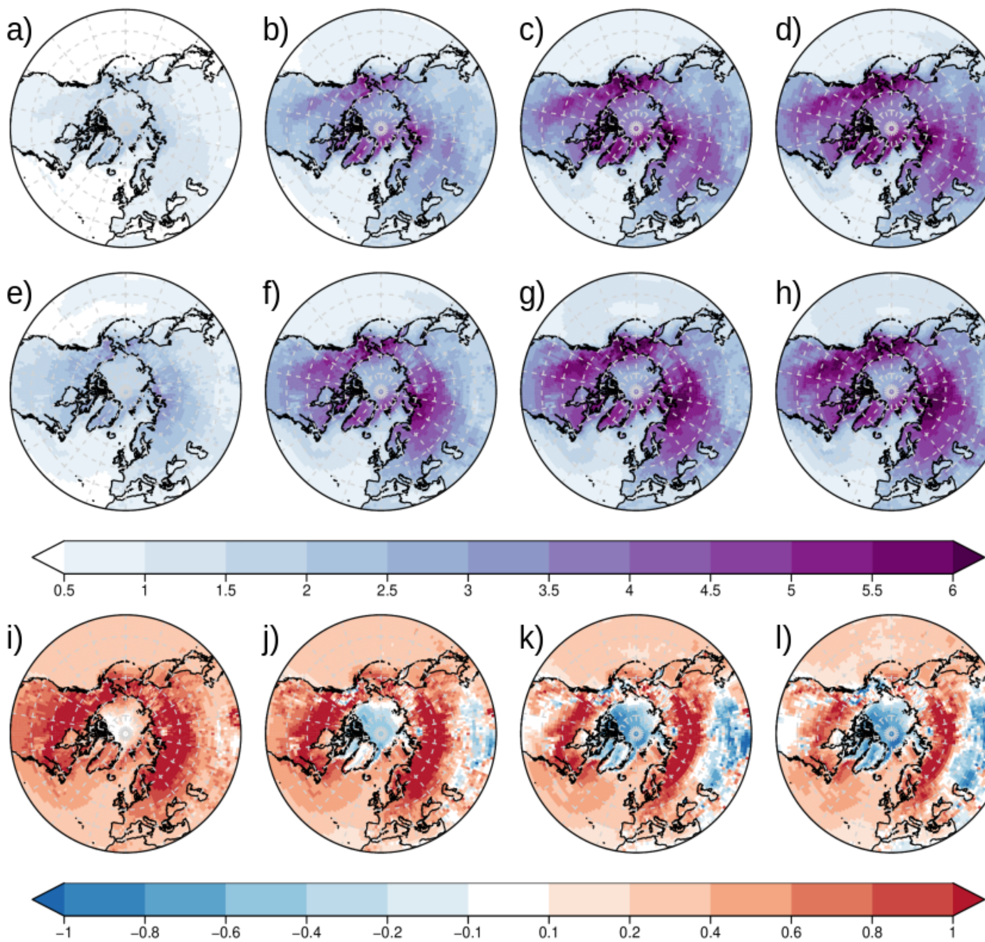


Figure A1. Ensemble standard deviation of 2-meter temperature for ~~week1~~ week 1 to week ~~for~~ 4 (a to d) CNRM and (e to h) ECMWF and week 1 to week 4 standard deviation differences 'ECMWF minus CNRM' (i to l). Differences not significant at the 95 % confidence level have been set to zero.

Appendix B: Approach based on persisting regimes

We here provide the results obtained after taking into account the persistence of the weather regimes, resulting in a new category "None" (see section 3.2 for details). As can be seen in comparing figure B1 with 5 and figure B2 with 6, the results found are very similar with or without this new category.

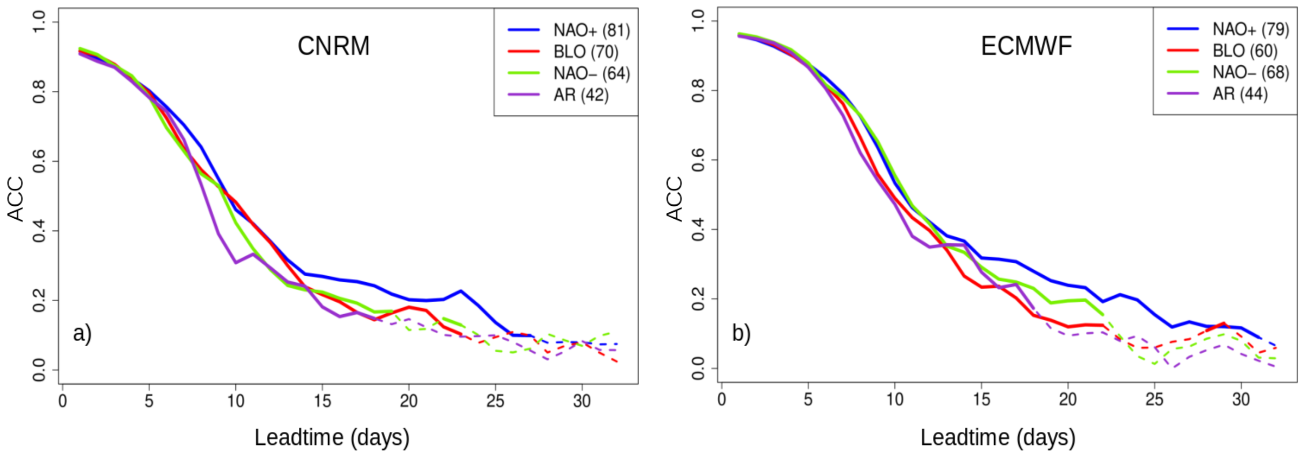


Figure B1. Like fig. 5 but with a fifth category ’None’ including days outside any persistent sequence of a canonical weather regime

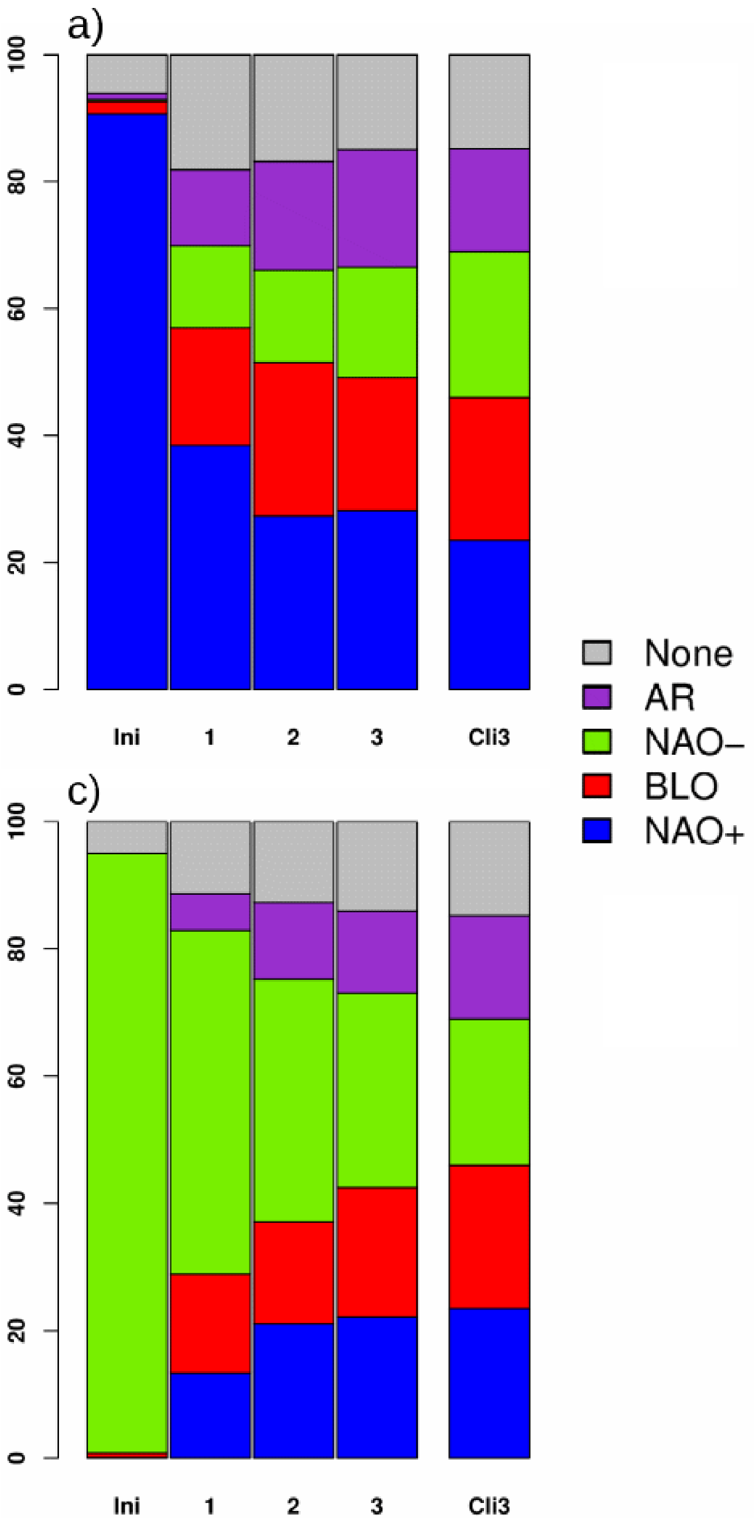


Figure B2. Like fig. 6 but with a fifth category 'None' including days outside any persistent sequence of a canonical weather regime

Code and data availability. The reforecast data used in this study is freely accessible from the S2S database (https://apps.ecmwf.int/datasets/data/s2s-realtime-instantaneous-accum-ecmf/levtype=sfc/type=cf/). The ERA5 reanalysis can be retrieved from the Climate Data Store (https://cds.climate.copernicus.eu/) and the CNRM-CM6-HR piControl simulation for CMIP6 from the Earth System Grid Federation platform (https://esgf-node.llnl.gov/search/cmip6/). The code for data analyses and plots is based on the free R software. Scripts are available upon request.

Author contributions. Constantin Ardilouze has collected and analysed the data, contributed to the design and drafted this article. Damien Specq, Lauriane Batté and Christophe Cassou have equally contributed to the design and the critical revision of this article.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work is part of the 2020-2022 project ICHEPS, supported by the French National programme LEFE (Les Enveloppes Fluides et l'Environnement).

References

Ardilouze, C., Batté, L., and Déqué, M.: Subseasonal-to-seasonal (S2S) forecasts with CNRM-CM: a case study on the July 2015 West-European heat wave, Advances in Science and Research, 14, 115–121, 2017.

Bach, E., Motesharrei, S., Kalnay, E., and Ruiz-Barradas, A.: Local Atmosphere–Ocean Predictability: Dynamical Origins, Lead Times, and Seasonality, Journal of Climate, 32, 7507–7519, 2019.

Barnston, A. G. and Livezey, R. E.: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns, Monthly weather review, 115, 1083–1126, 1987.

Batté, L. and Déqué, M.: Randomly correcting model errors in the ARPEGE-Climate v6. 1 component of CNRM-CM: applications for seasonal forecasts., Geoscientific Model Development, 9, 2016.

Cassou, C.: Intraseasonal interaction between the Madden–Julian oscillation and the North Atlantic Oscillation, Nature, 455, 523–527, 2008.

Cattiaux, J., Vautard, R., Cassou, C., Yiou, P., Masson-Delmotte, V., and Codron, F.: Winter 2010 in Europe: A cold extreme in a warming climate, Geophysical Research Letters, 37, 2010.

Chevallier, M., Massonnet, F., Goessling, H., Guémas, V., and Jung, T.: The role of sea ice in sub-seasonal predictability, in: Sub-Seasonal to Seasonal Prediction, pp. 201–221, Elsevier, 2019.

de Andrade, F. M., Young, M. P., MacLeod, D., Hirons, L. C., Woolnough, S. J., and Black, E.: Subseasonal Precipitation Prediction for Africa: Forecast Evaluation and Sources of Predictability, Weather and Forecasting, 36, 265–284, 2021.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, 2011.

Domeisen, D. I.: Estimating the frequency of sudden stratospheric warming events from surface observations of the North Atlantic Oscillation, Journal of Geophysical Research: Atmospheres, 124, 3180–3194, 2019.

Domeisen, D. I., Badin, G., and Koszalka, I. M.: How predictable are the Arctic and North Atlantic Oscillations? Exploring the variability and predictability of the Northern Hemisphere, Journal of Climate, 31, 997–1014, 2018.

Domeisen, D. I., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., Furtado, J. C., Garfinkel, C. I., Hitchcock, P., Karpechko, A. Y., et al.: The role of the stratosphere in subseasonal to seasonal prediction: 2. Predictability arising from stratosphere-troposphere coupling, Journal of Geophysical Research: Atmospheres, 125, e2019JD030 923, 2020.

Domeisen, D. I. et al.: Advances in the subseasonal prediction of extreme events, Bulletin of the American Meteorological Society, In revision.

Dorrington, J. and Strommen, K.: Jet speed variability obscures Euro-Atlantic regime structure, Geophysical Research Letters, 47, e2020GL087 907, 2020.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, 2016.

Falkena, S. K., de Wiljes, J., Weisheimer, A., and Shepherd, T. G.: Revisiting the identification of wintertime atmospheric circulation regimes in the Euro-Atlantic sector, Quarterly Journal of the Royal Meteorological Society, 146, 2801–2814, 2020.

Ferranti, L., Corti, S., and Janousek, M.: Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector, Quarterly Journal of the Royal Meteorological Society, 141, 916–924, 2015.

Ferranti, L., Magnusson, L., Vitart, F., and Richardson, D. S.: How far in advance can we predict changes in large-scale flow leading to severe cold conditions over Europe?, Quarterly Journal of the Royal Meteorological Society, 144, 1788–1802, 2018.

Ferro, C.: Fair scores for ensemble forecasts, Quarterly Journal of the Royal Meteorological Society, 140, 1917–1923, 2014.

Ferro, C. A., Richardson, D. S., and Weigel, A. P.: On the effect of ensemble size on the discrete and continuous ranked probability scores, Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling, 15, 19–24, 2008.

Hamouda, M. E., Pasquero, C., and Tziperman, E.: Decoupling of the Arctic Oscillation and North Atlantic Oscillation in a warmer climate, Nature Climate Change, 11, 137–142, 2021.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 2020.

Jung, T., Hilmer, M., Ruprecht, E., Kleppek, S., Gulev, S. K., and Zolina, O.: Characteristics of the recent eastward shift of interannual NAO variability, Journal of Climate, 16, 3371–3382, 2003.

Keeley, S. P. E., Sutton, R. T., and Shaffrey, L. C.: Does the North Atlantic Oscillation show unusual persistence on intraseasonal timescales?, Geophysical research letters, 36, 2009.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., et al.: The JRA-55 reanalysis: General specifications and basic characteristics, Journal of the Meteorological Society of Japan. Ser. II, 93, 5–48, 2015.

Kolstad, E. W., Wulff, C. O., Domeisen, D. I., and Woollings, T.: Tracing North Atlantic Oscillation forecast errors to stratospheric origins, Journal of Climate, 33, 9145–9157, 2020.

Lee, R. W., Woolnough, S. J., Charlton-Perez, A. J., and Vitart, F.: ENSO modulation of MJO teleconnections to the North Atlantic and Europe, Geophysical Research Letters, 46, 13 535–13 545, 2019.

Lin, H., Brunet, G., and Derome, J.: An observed connection between the North Atlantic Oscillation and the Madden–Julian oscillation, Journal of Climate, 22, 364–380, 2009.

Luo, B., Luo, D., Dai, A., Simmonds, I., and Wu, L.: Combined Influences on North American Winter Air Temperature Variability from North Pacific Blocking and the North Atlantic Oscillation: Subseasonal and Interannual Time Scales, Journal of Climate, 33, 7101–7123, 2020.

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., Dirmeyer, P. A., Ferranti, L., Johnson, N. C., Jones, J., et al.: Windows of Opportunity for Skillful Forecasts Subseasonal to Seasonal and Beyond, Bulletin of the American Meteorological Society, 101, E608–E625, 2020.

Mayer, K. J. and Barnes, E. A.: Subseasonal Forecasts of Opportunity Identified by an Explainable Neural Network, Geophysical Research Letters, n/a, e2020GL092 092, 2021.

Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho, C. A., Danabasoglu, G., Dirmeyer, P. A., Doblas-Reyes, F. J., Domeisen, D. I., et al.: Current and emerging developments in subseasonal to decadal prediction, Bulletin of the American Meteorological Society, 101, E869–E896, 2020.

Michelangeli, P.-A., Vautard, R., and Legras, B.: Weather regimes: Recurrence and quasi stationarity, Journal of the atmospheric sciences, 52, 1237–1256, 1995.

Minami, A. and Takaya, Y.: Enhanced Northern Hemisphere correlation skill of subseasonal predictions in the strong negative phase of the Arctic Oscillation, Journal of Geophysical Research: Atmospheres, 125, e2019JD031 268, 2020.

NOAA Climate Prediction Center NAO index: North Atlantic Oscillation (NAO), ftp://ftp.cpc.ncep.noaa.gov/cwlinks/norm.daily.na (Dec. 2020), 2020.

Pokorná, L. and Huth, R.: Climate impacts of the NAO are sensitive to how the NAO is defined, Theoretical and applied climatology, 119, 639–652, 2015.

Roberts, C. D., Vitart, F., and Balmaseda, M. A.: Hemispheric Impact of North Atlantic SSTs in Subseasonal Forecasts, Geophysical Research Letters, 48, e2020GL0911 446, 2021.

Robertson, A. W., Vigaud, N., Yuan, J., and Tippett, M. K.: Toward Identifying Subseasonal Forecasts of Opportunity Using North American Weather Regimes, Monthly Weather Review, 148, 1861–1875, 2020.

Seager, R., Kushnir, Y., Nakamura, J., Ting, M., and Naik, N.: Northern Hemisphere winter snow anomalies: ENSO, NAO and the winter of 2009/10, Geophysical research letters, 37, 2010.

Specq, D., Batté, L., Déqué, M., and Ardilouze, C.: Multimodel forecasting of precipitation at subseasonal timescales over the southwest tropical Pacific, Earth and Space Science, p. e2019EA001003, 2020.

Sun, L., Perlwitz, J., Richter, J. H., Hoerling, M. P., and Hurrell, J. W.: Attribution of NAO Predictive Skill Beyond 2 Weeks in Boreal Winter, Geophysical Research Letters, 47, e2020GL090 451, 2020.

Thompson, D. W. and Wallace, J. M.: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields, Geophysical research letters, 25, 1297–1300, 1998.

Vautard, R.: Multiple weather regimes over the North Atlantic: Analysis of precursors and successors, Monthly weather review, 118, 2056–2081, 1990.

Vigaud, N., Robertson, A. W., and Tippett, M. K.: Predictability of recurrent weather regimes over North America during winter from submonthly reforecasts, Monthly Weather Review, 146, 2559–2577, 2018.

Vitart, F.: Monthly forecasting at ECMWF, Monthly Weather Review, 132, 2761–2779, 2004.

Vitart, F.: Madden—Julian Oscillation prediction and teleconnections in the S2S database, Quarterly Journal of the Royal Meteorological Society, 143, 2210–2220, 2017.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., et al.: The subseasonal to seasonal (S2S) prediction project database, Bulletin of the American Meteorological Society, 98, 163–173, 2017.

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J.-F., Michou, M., Moine, M.-P., Nabat, P., Roehrig, R., Salas y Mélia, D., Séférian, R., Valcke, S., Beau, I., Belamari, S., Berthet, S., Cassou, C., Cattiaux, J., Deshayes, J., H. Douville, H., Franchisteguy, L., Ethé, C., Geoffroy, O., Lévy, C., Madec, G., Meurdesoif, Y., Msadek, R., Ribes, A., Sanchez, E., Terray, L., and Waldman, R.: Evaluation of CMIP6 DECK experiments with CNRM-CM6-1, Journal of Advances in Modeling Earth Systems, http://www.umr-cnrm.fr/cmip6/references, 2019.

White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J., Lazo, J. K., Kumar, A., Vitart, F., Coughlan de Perez, E., Ray, A. J., Murray, V., et al.: Potential applications of subseasonal-to-seasonal (S2S) predictions, Meteorological applications, 24, 315–325, 2017.

Woollings, T., Franzke, C., Hodson, D., Dong, B., Barnes, E. A., Raible, C., and Pinto, J.: Contrasting interannual and multidecadal NAO variability, Climate Dynamics, 45, 539–556, 2015.

Yamagami, A. and Matsueda, M.: Subseasonal Forecast Skill for Weekly Mean Atmospheric Variability Over the Northern Hemisphere in Winter and Its Relationship to Midlatitude Teleconnections, Geophysical Research Letters, 47, e2020GL088 508, 2020.

Zheng, C., Chang, E. K.-M., Kim, H., Zhang, M., and Wang, W.: Subseasonal to seasonal prediction of wintertime northern hemisphere extratropical cyclone activity by S2S and NMME models, Journal of Geophysical Research: Atmospheres, 124, 12 057–12 077, 2019.

Zuo, J., Ren, H.-L., Wu, J., Nie, Y., and Li, Q.: Subseasonal variability and predictability of the Arctic Oscillation/North Atlantic Oscillation in BCC_AGCM2. 2, Dynamics of Atmospheres and Oceans, 75, 33–45, 2016.