Decadal predictability of the North Atlantic eddy-driven jet in winter

Andrea Marcheggiani[1,2], Jon Robson[1], Paul-Arthur Monerie[1], Thomas J. Bracegirdle[3], and Doug Smith[4]

[1]National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading, UK

[2]Geophysical Institute, University of Bergen, and Bjerknes Centre for Climate Research, Bergen, Norway

[3]British Antarctic Survey, Cambridge, UK

[4]Met Office, Exeter, UK

Key Points:

- The winter North Atlantic (NA) eddy-driven jet is predictable on decadal timescales with skill (ACC) comparable to that for the winter NAO
- Anomalies in the NA jet are substantially smaller than expected from the ACC skill alone and so suffer from the signal-to-noise issue
- Skill drops significantly over the most recent period, as hindcasts do not capture the return to positive NAO conditions post 2010

Corresponding author: Andrea Marcheggiani, `andrea.marcheggiani@reading.ac.uk`

Corresponding author: Jon Robson, `j.i.robson@reading.ac.uk`

Abstract

This paper expands on work showing that the winter North Atlantic Oscillation (NAO) is predictable on decadal timescales to quantify the skill in capturing the North Atlantic eddy-driven jet's location and speed. By focussing on decadal predictions made for years 2-9 from the 6th Coupled Model Intercomparison Project over 1960-2005 we find that there is significant skill in both jet latitude and speed associated with the skill in the NAO. However, the skill in all three metrics appears to be sensitive to the period over which it is assessed. In particular, the skill drops considerably when evaluating hindcasts up to the present day as models fail to capture the latest observed northern shift and strengthening of the winter eddy-driven jet and more positive NAO. We suggest the drop in atmospheric circulation skill is related to reduced skill in North Atlantic Sea surface temperature.

Plain Language Summary

Climate models have been shown to be capable of predicting the evolution of the mean atmospheric circulation over long time scales, from annual to decadal and longer. However, models are overestimating the chaotic, unpredictable component of the climate's variability and, although model predictions follow the observed oscillations of the climate, the strength of these oscillations is critically underestimated. Recently, it was shown that climate models have skill in predicting the North Atlantic Oscillation and in this paper we assess model skill in predicting the evolution of the North Atlantic eddy-driven jet in winter, with the aim to highlight how much of the skill at predicting the NAO derives from good predictions of the jet's state. We find levels of skill similar to that for the NAO, with slightly higher skill for the jet's strength (or speed) over its location (latitude of its maximum speed). We also notice a drop in skill over the last decade, as models fail to capture the latest trends in the NAO and jet's evolution, and suggest that it might be related to degradation in skill at predicting surface temperature variability.

1 Introduction

The North Atlantic climate system is characterized by significant atmosphere and ocean variability that occurs on a wide range of time scales. In particular, the North Atlantic Oscillation (NAO) represents the leading pattern of climate variability in the North

Atlantic region, with positive NAO typically associated with stormier and wetter conditions over Western Europe, while negative values correspond to drier, colder weather (Hurrell, 1995). The NAO is also closely linked to the intensity and position of the North Atlantic eddy-driven jet (Thompson et al., 2003; Woollings et al., 2010). Furthermore, the atmospheric circulation variability also exerts a strong influence on the climate of the North Atlantic basin and Western Europe (Thompson & Wallace, 2001; Sutton et al., 2018; Hall & Hanna, 2018). Therefore, reliable predictions of the NAO and jet's evolution are of prime societal importance for Northern and Western Europe.

Considerable evidence has now emerged showing that the NAO is predictable on seasonal (Scaife et al., 2014) to decadal timescales (Smith et al., 2020; Athanasiadis et al., 2020). In particular, Smith et al. (2020, henceforth, S20) revealed a high level of skill at predicting decadal variability of the winter NAO in the 5th (CMIP5, Taylor et al., 2012) and 6th (CMIP6, Eyring et al., 2016) Coupled Model Intercomparison Project's prediction systems for hindcasts initialized between 1960–2005. Furthermore, S20 showed how the predictability of the NAO can be used to improve decadal predictions of other climate variables (e.g., surface temperature, mean sea level pressure, precipitation). However, the magnitude of the predictable signals in seasonal and decadal predictions appears to be significantly underestimated – leading to the so-called signal-to-noise paradox – and large ensembles are needed to reveal the predictable signal (Scaife & Smith, 2018).

In contrast to the NAO, decadal predictions of the eddy-driven jet have not yet been assessed. Thus, we do not know what aspect of the eddy-driven jet changes are associated with the winter NAO skill in S20. Furthermore, we expect the relationship between the eddy-driven jet and the winter NAO to change with the timescale and may be related to different processes (Woollings et al., 2015; Baker et al., 2017). For example, jet latitude changes appear to dominate interannual variability of the winter NAO (Woollings et al., 2015) and skillful seasonal predictions of the winter NAO have been associated with a skillful prediction of shifts in the jet latitude (Parker et al., 2019). However, decadal time-scale winter NAO variability has been linked more to changes in eddy-driven jet speed that, in turn, appear to be driven by sea surface temperatures in the subpolar North Atlantic (Woollings et al., 2015). The different aspects of jet variability (e.g., latitude or speed) are also known to lead to different impacts on sea ice, temperatures and precipitation both over the North Atlantic ocean basin and over western Europe (Hall &

Hanna, 2018; Ma et al., 2020). Therefore, understanding the different aspects of skill could be useful in understanding what sectors would benefit most from improved predictions on these timescales.

In this paper, we build upon the analysis of S20 to evaluate the skill of the eddy-driven jet. In particular, we address how much of the winter NAO skill on decadal timescales is associated with skill in predicting the eddy-driven jet latitude and speed. We focus our analysis on the CMIP6 models, which were not all available at the time of S20, and extend the analysis over observations of the latest period that was not covered by CMIP5 hindcasts.

2 Data and Methods

In this study, we assess a multi-model ensemble of decadal predictions from prediction systems taking part in *component A* of the Decadal Climate Prediction Project (DCPP-A, Boer et al., 2016) as a contribution to CMIP6. A list of the models considered is provided in Table S1 in the Supporting Information. The multi-model ensemble consists of 10 models and 153 members in total (of which 120 were also considered in S20).

As in S20, we define the NAO index as the difference in mean sea-level pressure between two small boxes located around the Azores ($28°-20°$W, $36°-40°$N) and Iceland ($25°-16°$W, $63°-70°$N). The Arctic Oscillation (AO) index is calculated as the difference in mean sea-level pressure between the midlatitudes ($30°-60°$N) and the high/polar latitudes ($60°-90°$N). We construct the indices of the eddy-driven jet's latitude (JLI) and speed (JSI) by following their definition in Bracegirdle et al. (2018), which draws from Woollings et al. (2010) but uses monthly averaged data instead of daily: we first calculate the zonal mean of the zonal wind at 850hPa in the North Atlantic sector ($60°$W$-0°$, $10° - 75°$N) and then identify the maximum and its location as the jet's latitude and speed. As in S20, we focus on assessing skill for years 2–9 of the hindcasts, restricting our attention to the extended boreal winter (December, January, February and March, DJFM).

The different forecasting systems are initialized towards the end of each starting year. While the first winter of a hindcast is not necessarily complete (some models are initialized at the end of December, so their first winter season does not include it), it does not affect our analysis as we consider hindcast years 2–9 (winter of year 2 is complete for all models).

Multi-model ensemble mean anomalies are constructed by first subtracting the model mean state (i.e. the time average between hindcast years 2–9 over all starting dates and ensemble members, see Fig. S1 in Supporting Information) from each ensemble member and then taking the equally weighted average of all ensemble members. Finally, we consider the time mean of years 2–9 winters. Following S20, we construct a lagged ensemble by combining each hindcast with the previous three start dates, thus quadrupling the number of ensemble members from 153 to 612. We refer to the resulting multimodel ensemble mean as the "lagged" mean.

The skill of DCPP-A is assessed against reanalysis data from the ERA5 data set (Hersbach et al., 2020), between 1979 and 2021 and its back-extension for years 1960–1978 (Bell et al., 2021). Indices from reanalysis are computed in a similar way (removing the seasonal climatology across the time period considered) and then smoothed through an 8-year rolling average so that the observations and hindcasts cover the same time periods. Reanalysis and model data were interpolated to a 2.5°×2.5° grid before analysis. S20 used mean sea level pressure data from HadSLP2 (Allan & Ansell, 2006) to compute the observed NAO, which appears to have a lower variance in time than ERA5. However, we do not expect this difference in variance to affect the skill estimates, which are dependent on the phasing of the variability rather than its magnitude.

We measure the skill by evaluating the Pearson anomaly correlation coefficient (ACC) between the observations (ERA5) and the multi-model ensemble mean and estimate the Ratio of Predictable Components (RPC) as in Eade et al. (2014),

$$\mathrm{RPC} = \frac{\sigma^o_{sig}/\sigma^o_{tot}}{\sigma^f_{sig}/\sigma^f_{tot}} \approx \mathrm{ACC}\frac{\sigma^f_{tot}}{\sigma^f_{sig}}, \qquad (1)$$

where σ_{tot} and σ_{sig} are, respectively the expected total (signal plus noise) and signal standard deviations in the observations/reanalysis ('o') and forecast ('f'). We test the statistical significance of the ACC estimates by using a block bootstrap approach (as in S20).

We assess skill over different time periods: a *short* period consisting of years 2–9 of hindcasts initialized at the end of years 1960–2005 (corresponding to the time period studied in S20, that is 1962 to 2014) and a *long* period, which includes hindcasts initialized at the end of years 2006 to 2012 (thus covering the period 1962–2021).

3 Skill in the NAO and jet stream indices

We first examine the 2–9 year prediction skill of DCPP-A for the NAO and jet latitude and speed, initially focusing on the same start dates examined by S20 (i.e. the *short* period, from 1960 to 2005).

Figure 1a shows predictions of the NAO time series. The observed NAO features a pronounced decadal and multidecadal variability (black curves in Fig. 1), with a generally increasing trend between the 1960s and 1990s followed by a decrease persisting until the late 2000s. As noted in S20, the multi-model ensemble mean appears not to be able to capture the observed decadal variability, with the observed extremes in the 1960s and 1990s lying outside model uncertainties (red shading in the left panels of Fig. 1). Nonetheless, models do show skill at predicting the phasing of such decadal variability, as indicated by the significant positive ACCs. Over the *short* period, the ACC of the multi-model ensemble mean for the NAO is 0.55 ($P < 0.01$), which compares to 0.48 ($P = 0.03$) in S20 over the same period, and is also affected by a low signal-to-noise ratio (RPC of 4.6 here, 4.2 in S20).

S20 also showed NAO predictions can be improved by computing the lagged ensemble mean, which helps filter out the unpredictable noise, and by re-scaling the variance to the observed. The resulting model predictions (thick red curves in the right hand panels of Fig. 1) are visibly improved as the magnitude of the signal is closer to that of observations. We also obtain a higher level of ACC consistent with S20 (compare the ACC in left panels to those in right panels of Fig. 1). At the same time, the RPC also increases, almost doubling in magnitude compared to the raw ensemble mean. This is indicative of the low signal-to-noise ratio that is characteristic of climate models (Scaife & Smith, 2018). Models also show similar levels of skill for the AO index (+0.55 and +0.63 for the raw and lagged ensemble means, respectively), as shown in Fig. 1g,h.

We then examine the skill of DCPP-A models at predicting the eddy-driven jet's variability (latitude and speed), which also shows decadal timescale variability similar to the NAO (see Figure 1c and e). Models have higher skill in predicting the speed of the jet (0.62, Figure 1e) than its latitudinal location (0.28, Figure 1c). The RPC for the jet latitude (2.7) is lower than that for the jet speed (5.4), consistent with the lower skill in the former. Again, the skill improves when using the lagged ensemble mean for both the jet latitude (0.52, Fig. 1d) and the speed (0.71, Fig. 1f). The RPC also becomes larger,

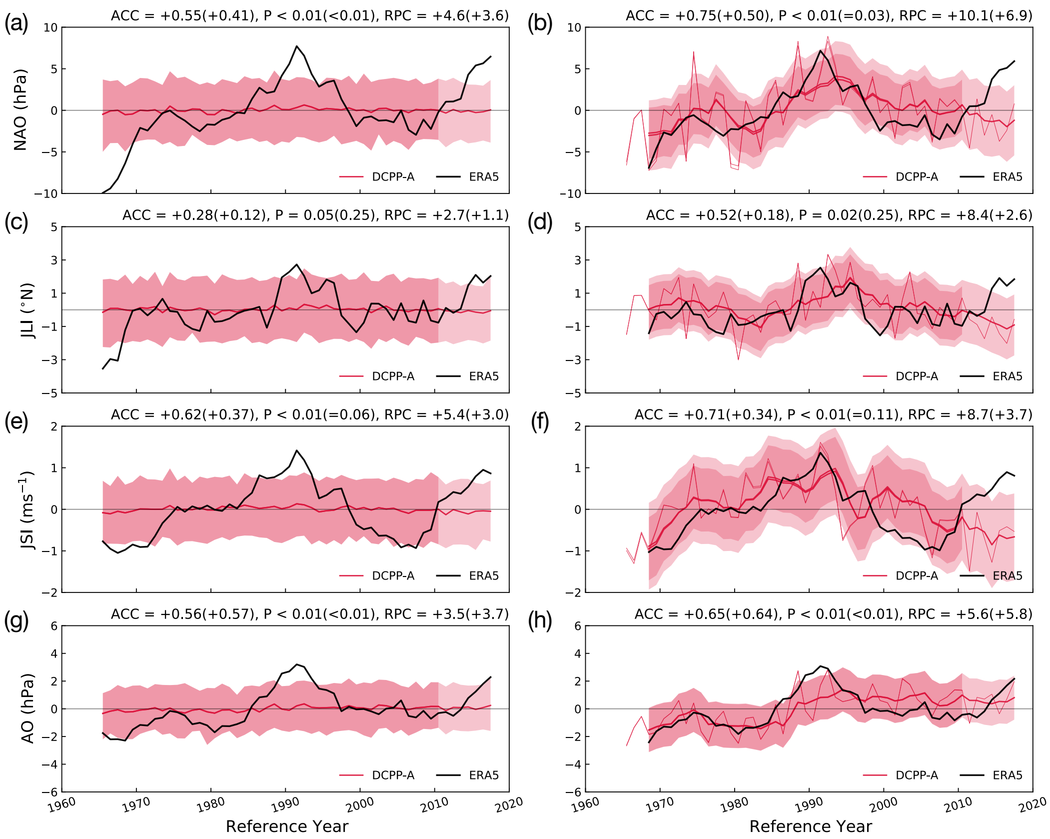

Figure 1. Evolution of 8-year running mean observed (black) and year 2–9 predictions from DCPP-A hindcasts (red) extended boreal winter (DJFM) NAO (a,b), Jet Latitude (c,d), Jet Speed (e,f) and AO (g,h) indices. Panels on the left show the raw ensemble-mean prediction (i.e., no re-scaling of variance). Panels on the right are the same as those on the left, but showing the ensemble-mean forecast (thin red, resulting from 153 ensemble members) rescaled to have the same variance as the observations and also the lagged ensemble-mean forecast (thick red, resulting from 612 ensemble members, rescaled by the same factor as for the non-lagged). The red shading in panels (a,c,e) represents the 5th-95th percentiles of all ensemble members (dark shading corresponds to *short* period; the additional years in the *long* period are shown in lighter shading) while in panels (b,d,f) it indicates the 5%-95% confidence interval estimated from the root-mean square error of the lagged ensemble with respect to the observations. At the top of each panel, we indicate the ACC with its significance (P) and the corresponding RPC for the *short* period (*long* period inside brackets).

more than trebling for JLI (2.7 to 8.4), while the increase is more moderate for JSI (5.4 to 8.7). Therefore, the similar levels of skill for the NAO and the jet speed suggests that the skill in the NAO on decadal timescales is associated with skill in the jet speed rather than its latitude. This also appears to be the case for quite a wide range of lead times, as we observe comparable skill in NAO and JSI predictions (see Fig. S2 in Supporting Information).

Figure 1 and previous work (e.g., Scaife & Smith, 2018; Klavans et al., 2021) have shown that prediction skill is sensitive to the number of ensemble members. Such a result is also underlined by the fact that, of the models that contributed to DCPP-A, the models with the biggest ensemble size also have the largest skill (not shown). Therefore, an obvious question is whether the skill scores computed here for DCPP-A represent the upper limit of skill, or whether more skill could be expected. To assess the upper limit of skill we plot how skill changes with the number of ensemble members. We do this by computing the skill for a random selection of different ensemble members that make up the lagged ensemble mean (612 members) and gradually increasing the size of the selection.

Figure 2 shows the resulting skill at predicting the atmospheric indices considered in this study as a function of ensemble size. Consistent with the evaluation of skill in Fig. 1, it highlights the different levels of skill for the different indices. However, it also shows that skill in the NAO, jet latitude and jet speed appear to still be increasing when using the maximum number of ensemble members (e.g. 612), suggesting that ACC skill could be expected to increase further with a larger number of ensemble members. We point out that the shading in Fig. 2 does not represent the uncertainty associated with the estimation of the correlation score, rather it indicates the spread in the distribution of the random selection of combinations.

As an aside, we find that the overall skill for the NAO and eddy-driven jet is sensitive to the inclusion of March in the winter season mean (e.g., DJFM compared to DJF). The increase in skill is especially clear for the jet latitude, which is associated with a significant drop in skill when assessing DJF rather than DJFM (not shown). This drop in skill appears to be consistent with the larger decadal and multidecadal variability observed in the North Atlantic eddy-driven jet in March (e.g., Simpson et al., 2019), although the larger variability on decadal timescales appears to be dependent on how basin-wide variability is measured (Bracegirdle, 2022).

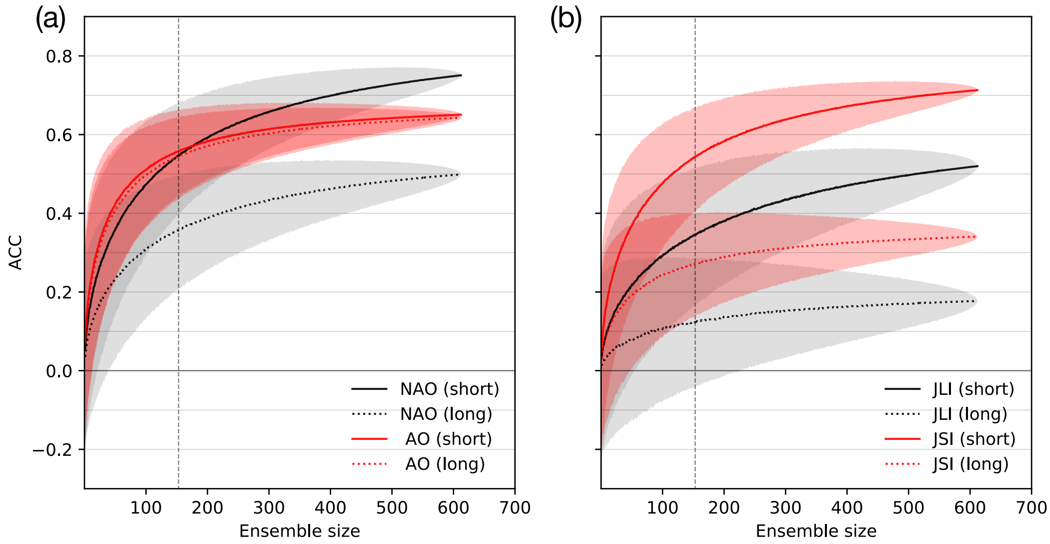

Figure 2. (a) Relationship between ensemble size and skill (ACC) at predicting the NAO and AO (black and red, respectively) with lagged ensemble means, for the short (solid lines) and long (dotted lines) periods. Shading represents 5th-95th percentiles of distribution of ACCs from 10,000 random combinations of a number of ensemble members; lines indicate the mean of such distributions. (b) As in (a), for JLI (black) and JSI (red).

4 Degradation of skill in the recent period

The previous section, and results in Fig. 1, focused on evaluating hindcasts initialized over 1960–2005 (i.e., the *short* period) to be consistent with results from S20. However, DCPP-A hindcasts from CMIP6 cover a longer time period and longer observational data is available to evaluate them. Therefore, here we extend our analysis to evaluate hindcasts initialized over 1960—2012, which we call the *long* period.

When evaluating DCPP-A hindcasts over the long period, we find the skill for the NAO and the jet indices drops substantially. For example, the lagged ensemble skill for the jet latitude and jet speed decreases from +0.52 and +0.71 respectively to statistically insignificant values of +0.18 and +0.34. Skill in the the NAO index drops from +0.75 to a 0.50, but the latter value is still statistically significant. The differences in skill are found to be statistically significant via block-bootstrapping. The drop in skill is also related to a drop in RPC values, which decreases to 6.9 for the NAO, and down to 2.6 and 3.4 for the jet latitude and speed respectively. The drop in skill appears to be related primarily to DCPP-A hindcasts failing to capture the observed positive trend in the indices over the 2010s (Fig. 1). In particular, this period corresponds to a return to positive NAO conditions associated with a stronger and more northerly jet. Such a drop in skill is also visible from the inspection of Fig. 2. However, it is clear that model skill is not only lower over the *long* period but the increase in skill with ensemble size also appears to reach saturation at smaller ensemble sizes (except for the AO index).

Alongside the drop in skill of the atmospheric variables, there is also a drop in skill of surface temperature over the North Atlantic Ocean. Figure 3a,b shows the skill of DCPP-A hindcasts at predicting temperatures near the surface (TAS). For the short period (Fig. 3a) there is significant skill over the majority of the globe, with particularly strong skill in the North Atlantic and across the tropical Atlantic Ocean, and also in the Indian and western Pacific Oceans. However, for the longer period we find a significant reduction in skill over the eastern subpolar North Atlantic and in the tropical North Atlantic (Fig. 3b). This reduction of skill over the North Atlantic is associated with DCPP-A predictions being too warm over the subpolar North Atlantic, as suggested in lower panels of Fig. 3 where we show the latest changes (from end of short period to end of long period, i.e. 2010–2017) in TAS in DCPP-A models (Fig. 3c) and the deviation of DCPP-A models from observations (Fig. 3d). In other words, the DCPP-A multimodel mean does not capture the recent cooling of the subpolar North Atlantic post-2005 (Robson et al., 2016). Anomalously cold temperatures over the subpolar and tropical North Atlantic Ocean have been suggested as drivers of positive NAO and a faster jet (Rodwell et al., 1999; Woollings et al., 2015). Therefore, one interpretation is that a drop in TAS predictability is the cause of the drop in NAO and jet indices.

However, it is important to note that warmer surface temperatures over the North Atlantic Ocean would also be expected due to the failure to predict the positive NAO (e.g., because positive NAO drives increased oceanic heat loss, Marshall et al., 2001; Grist et al., 2010) and there are other factors that may be relevant. For example, previous work has highlighted that temperatures in the western tropical Pacific are a key driver of the NAO on decadal timescales (Latif, 2001; Kucharski et al., 2006). Nevertheless, we see no change in skill in this region between the short and long period (Fig. 3b), suggesting that this is not the primary cause. External forcings have been linked to NAO variability (Christiansen, 2008; Ortega et al., 2015; Sjolte et al., 2018) and may explain the skill in the short period (Klavans et al., 2021). However, different forcing factors change though time and the skill expected from external forcing is also sensitive to the time pe-

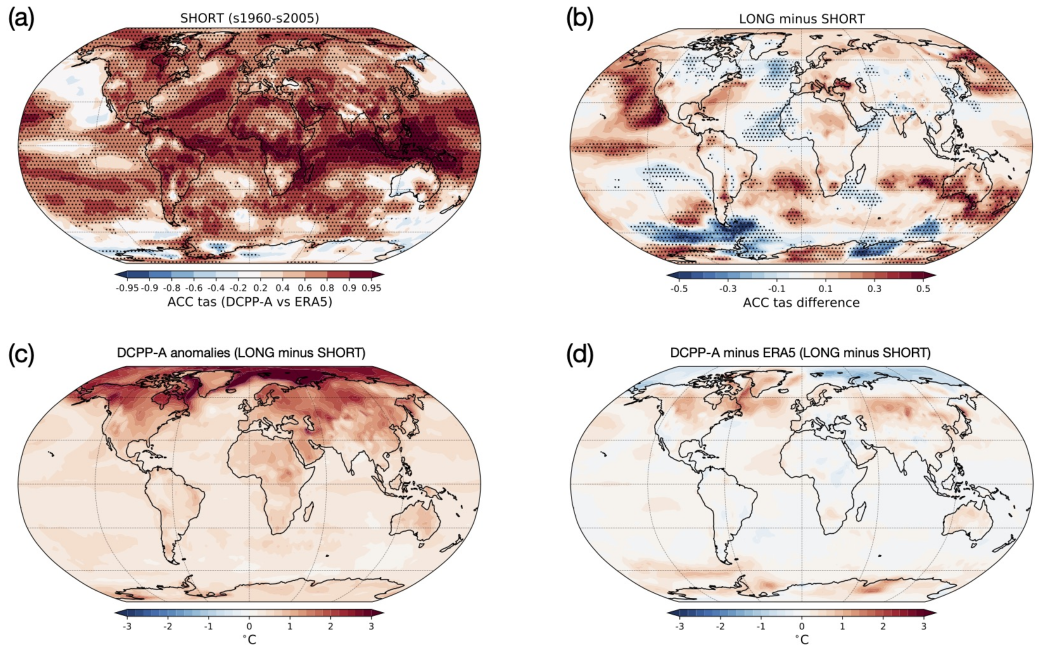

Figure 3. Surface temperature (TAS) skill (as measured by ACC) of year 2–9 hindcast from DCPP-A for the *short* period (a) and the difference *long* minus *short* (b). Panels c,d show changes in TAS over the latest decade (i.e. *long* minus *short*) in DCPP-A models and deviations of DCPP-A models from ERA5, respectively. Stippling indicate statistical significance ($P< 0.05$) of the ACC (a) and its difference across the two periods (b).

riod used (Sjolte et al., 2018). Additionally, state-dependent predictability of the NAO (Weisheimer et al., 2017) , as well state dependence in teleconnections (López-Parages & Rodríguez-Fonseca, 2012; Weisheimer et al., 2017; Fereday et al., 2020) may also play a role. Finally, we also note that the drop in skill appears largely an Atlantic phenomenon as there is no significant drop in skill in the predictability of the Arctic Oscillation index (ACC values of +0.65 and 0.64 for the short and long period respectively, see fig 1g and h). Therefore, further work is needed to unravel the causes of the drop in skill.

5 Conclusions

In this paper we expand upon the analysis presented in Smith et al. (2020) to assess the predictability of the North Atlantic eddy-driven jet (latitude and speed) in winter (December to March) in decadal predictions made for CMIP6. In particular, we evaluate the prediction skill of the eddy-driven jet latitude and speed in winter and we compare with skill in the winter North Atlantic Oscillation (NAO). Our key results are as follows:

1. The North Atlantic eddy-driven jet is predictable on decadal time-scales when evaluating hindcasts initialized over the period 1960–2005 (i.e., the same time-period as used in Smith et al., 2020). The Anomaly Correlation Coefficient skill score (ACC) for years 2–9 of the ensemble mean (after post-processing to reduce unpredictable noise, i.e. considering a lagged-ensemble) is 0.52 and 0.71 for jet latitude and jet speed, respectively, and is consistent with the ACC of 0.75 for the winter NAO.
2. As with the NAO, the amplitude of predicted anomalies in the North Atlantic eddy-driven jet is substantially smaller compared to observations (RPC of 8.4 and 8.7 for the jet latitude and speed, respectively), despite the high level of ACC, indicating that they also suffer from a low signal-to-noise ratio.
3. The skill for all indices drops substantially when evaluating hindcasts initialized between 1960–2012 (rather than 1960–2005). This drop in skill was due to hindcasts failing to capture both the return to positive NAO conditions post 2010 and the poleward extension and strengthening of the jet. As a result, the skill of the NAO drops to 0.50 and significant skill is no-longer present in the North Atlantic eddy-driven jet indices.
4. Alongside the drop in skill of the atmospheric circulation in the North Atlantic, there is also a significant drop in skill at capturing the surface air temperature over the subpolar and tropical North Atlantic when evaluating hindcasts initialized between 1960–2012 rather than 1960–2005.

This paper has demonstrated that, alongside the NAO, it is possible to predict the winter North Atlantic eddy-driven jet on decadal time-scales. However, as with the NAO, the predictable signal appears too weak. Future work could explore calibrations of the predictions as in Smith et al. (2020) in order to provide more relevant information to society, and to explore whether jet predictions (e.g., latitude or speed) could be more useful to some sectors than the NAO predictions.

However, it is also clear that the skill in North Atlantic Atmospheric circulation in winter is sensitive to the time period over which it is computed. Unfortunately, the reasons behind this drop in skill are still unclear. Our results suggest that the drop in skill is primarily related to the physical mechanisms that unfold in the North Atlantic basin. In fact, the skill in predicting hemisphere-wide variability (e.g., the Arctic Oscillation) was not found to be affected by a similar degradation over the most recent period. Furthermore, one potential interpretation is that the drop in skill of the atmospheric variables is consistent with a reduction in skill at capturing surface temperature anomalies over North Atlantic Ocean. However, the drop in North Atlantic atmospheric circulation skill could be related to other factors, such as external forcing changes, state-dependent predictability, or poorly related processes. Therefore, in order to have confidence in future predictions, it is important that future work explores the reasons behind changing skill.

Open Research

The climate model hindcasts are available via the Earth System Grid Federation (ESGF) archive of the 6th Coupled Model Intercomparison Project (CMIP6) data (`https://esgf-index1.ceda.ac.uk/projects/esgf-ceda/`).

Reanalysis data from ECMWF's ERA5 is available from `https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5`.

Acknowledgments

This work was funded by the Natural Environment Research Council (NERC) ACSIS project (NE/N018001/1 and NE/N018028/1). JR was additionally funded by NERC via the WISHBONE project (NE/T013516/1) and via NCAS. TJB was additionally supported through the NERC British Antarctic Survey research program Polar Science for Planet Earth. DS was supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. This work used JASMIN, the U.K. collaborative data analysis facility.

References

Allan, R., & Ansell, T. (2006). A new globally complete monthly historical gridded mean sea level pressure dataset (hadslp2): 1850–2004. *Journal of Climate*, *19*(22), 5816–5842.

Athanasiadis, P. J., Yeager, S., Kwon, Y.-O., Bellucci, A., Smith, D. W., & Tibaldi, S. (2020). Decadal predictability of north atlantic blocking and the nao. *NPJ Climate and Atmospheric Science*, *3*(1), 1–10.

Baker, H. S., Woollings, T., & Mbengue, C. (2017). Eddy-driven jet sensitivity to diabatic heating in an idealized gcm. *Journal of Climate*, *30*(16), 6413 - 6431. doi: 10.1175/JCLI-D-16-0864.1

Bell, B., Hersbach, H., Simmons, A., Berrisford, P., Dahlgren, P., Horányi, A., … others (2021). The era5 global reanalysis: Preliminary extension to 1950. *Quarterly Journal of the Royal Meteorological Society*, *147*(741), 4186–4227.

Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., … others (2021). Norcpm1 and its contribution to cmip6 dcpp. *Geoscientific Model Development*, *14*(11), 7073–7116.

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., … others (2016). The decadal climate prediction project (dcpp) contribution to cmip6. *Geoscientific Model Development*, *9*(10), 3751–3777.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., … others (2020). Presentation and evaluation of the ipsl-cm6a-lr climate model. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS002010.

Bracegirdle, T. J. (2022). Early-to late-winter 20th century north atlantic multidecadal atmospheric variability in observations, cmip5 and cmip6. *Geophysical Research Letters*, e2022GL098212.

Bracegirdle, T. J., Lu, H., Eade, R., & Woollings, T. (2018). Do cmip5 models reproduce observed low-frequency north atlantic jet variability? *Geophysical Research Letters*, *45*(14), 7204–7212.

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., … others (2019). Global mean climate and main patterns of variability in the cmcc-cm2 coupled model. *Journal of Advances in Modeling Earth Systems*, *11*(1), 185–209.

Christiansen, B. (2008). Volcanic eruptions, large-scale modes in the northern hemisphere, and the el niño–southern oscillation. *Journal of Climate*, *21*(5), 910–922.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D., DuVivier, A., Edwards, J., ... others (2020). The community earth system model version 2 (cesm2). *Journal of Advances in Modeling Earth Systems*, *12*(2), e2019MS001916.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., ... others (2021). The ec-earth3 earth system model for the climate model intercomparison project 6. *Geoscientific Model Development Discussions*, 1–90.

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., & Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical research letters*, *41*(15), 5620–5628.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958.

Fereday, D., Chadwick, R., Knight, J., & Scaife, A. (2020). Tropical rainfall linked to stronger future enso-nao teleconnection in cmip5 models. *Geophysical Research Letters*, *47*(22), e2020GL088664.

Grist, J. P., Josey, S. A., Marsh, R., Good, S. A., Coward, A., De Cuevas, B. A., ... others (2010). The roles of surface heat flux and ocean heat transport convergence in determining atlantic ocean temperature variability. *Ocean dynamics*, *60*(4), 771–790.

Hall, R. J., & Hanna, E. (2018). North atlantic circulation indices: links with summer and winter uk temperature and precipitation and implications for seasonal forecasting. *International Journal of Climatology*, *38*(S1), e660-e677. doi: https://doi.org/10.1002/joc.5398

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J. (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049.

Hurrell, J. W. (1995). Decadal trends in the north atlantic oscillation: Regional temperatures and precipitation. *Science*, *269*(5224), 676–679.

Klavans, J. M., Cane, M. A., Clement, A. C., & Murphy, L. N. (2021). Nao predictability from external forcing in the late 20th century. *npj Climate and Atmospheric Science*, *4*(1), 1–8.

Kucharski, F., Molteni, F., & Bracco, A. (2006). Decadal interactions between the western tropical pacific and the north atlantic oscillation. *Climate dynamics*, *26*(1), 79–91.

Latif, M. (2001). Tropical pacific/atlantic ocean interactions at multi-decadal time scales. *Geophysical Research Letters*, *28*(3), 539–542.

López-Parages, J., & Rodríguez-Fonseca, B. (2012). Multidecadal modulation of el niño influence on the euro-mediterranean rainfall. *Geophysical Research Letters*, *39*(2).

Ma, L., Woollings, T., Williams, R. G., Smith, D., & Dunstone, N. (2020). How does the winter jet stream affect surface temperature, heat flux, and sea ice in the north atlantic? *Journal of Climate*, *33*(9), 3711–3730.

Marshall, J., Johnson, H., & Goodman, J. (2001). A study of the interaction of the north atlantic oscillation with ocean circulation. *Journal of Climate*, *14*(7), 1399–1421.

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., ... others (2018). A higher-resolution version of the max planck institute earth system model (mpi-esm1. 2-hr). *Journal of Advances in Modeling Earth Systems*, *10*(7), 1383–1413.

Ortega, P., Lehner, F., Swingedouw, D., Masson-Delmotte, V., Raible, C. C., Casado, M., & Yiou, P. (2015). A model-tested north atlantic oscillation reconstruction for the past millennium. *Nature*, *523*(7558), 71–74.

Parker, T., Woollings, T., Weisheimer, A., O'Reilly, C., Baker, L., & Shaffrey, L. (2019). Seasonal predictability of the winter north atlantic oscillation from a jet stream perspective. *Geophysical Research Letters*, *46*(16), 10159–10167.

Robson, J., Ortega, P., & Sutton, R. (2016). A reversal of climatic trends in the North Atlantic since 2005. *Nature Geoscience*, *9*(7), 513. doi: https://doi.org/10.1038/ngeo2727

Rodwell, M. J., Rowell, D. P., & Folland, C. K. (1999). Oceanic forcing of the wintertime north atlantic oscillation and european climate. *Nature*, *398*(6725), 320–323. doi: https://doi.org/10.1038/18648

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., ... Williams, A. (2014). Skillful long-range prediction of european and north american winters. *Geophysical Research Letters*, *41*(7), 2514-2519. doi: https://doi.org/10.1002/2014GL059637

Scaife, A. A., & Smith, D. (2018). A signal-to-noise paradox in climate science. *npj Climate and Atmospheric Science*, *1*(1), 1–8.

Simpson, I. R., Yeager, S. G., McKinnon, K. A., & Deser, C. (2019). Decadal predictability of late winter precipitation in western europe through an ocean–jet stream connection. *Nature Geoscience*, *12*(8), 613–619.

Sjolte, J., Sturm, C., Adolphi, F., Vinther, B. M., Werner, M., Lohmann, G., & Muscheler, R. (2018). Solar and volcanic forcing of north atlantic climate inferred from a process-based reconstruction. *Climate of the Past*, *14*(8), 1179–1194.

Smith, D. M., Scaife, A. A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., ... others (2020). North atlantic climate far more predictable than models imply. *Nature*, *583*(7818), 796–800.

Sutton, R., McCarthy, G. D., Robson, J., Sinha, B., Archibald, A., & Gray, L. (2018). Atlantic multidecadal variability and the uk acsis program. *Bulletin of the American Meteorological Society*, *99*(2), 415–425.

Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., ... others (2019). The canadian earth system model version 5 (canesm5. 0.3). *Geoscientific Model Development*, *12*(11), 4823–4873.

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., ... others (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in miroc6. *Geoscientific Model Development*, *12*(7), 2727–2765.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of cmip5 and the experiment design. *Bulletin of the American meteorological Society*, *93*(4), 485–498.

Thompson, D. W., Lee, S., & Baldwin, M. P. (2003). Atmospheric processes governing the northern hemisphere annular mode/north atlantic oscillation. *Geophysical Monograph-American Geophysical Union*, *134*, 81–112.

Thompson, D. W., & Wallace, J. M. (2001). Regional climate impacts of the northern hemisphere annular mode. *Science*, *293*(5527), 85–89.

Weisheimer, A., Schaller, N., O'Reilly, C., MacLeod, D. A., & Palmer, T. (2017). Atmospheric seasonal forecasts of the twentieth century: multi-decadal variability in predictive skill of the winter north atlantic oscillation (nao) and their potential value for extreme event attribution. *Quarterly Journal of the Royal Meteorological Society*, *143*(703), 917–926.

Williams, K., Copsey, D., Blockley, E., Bodas-Salcedo, A., Calvert, D., Comer, R., ... others (2018). The met office global coupled model 3.0 and 3.1 (gc3.0 and gc3.1) configurations. *Journal of Advances in Modeling Earth Systems*, *10*(2), 357–380.

Woollings, T., Franzke, C., Hodson, D., Dong, B., Barnes, E. A., Raible, C., & Pinto, J. (2015). Contrasting interannual and multidecadal nao variability. *Climate Dynamics*, *45*(1), 539–556.

Woollings, T., Hannachi, A., & Hoskins, B. (2010). Variability of the north atlantic eddy-driven jet stream. *Quarterly Journal of the Royal Meteorological Society*, *136*(649), 856–868.

Xiao-Ge, X., Tong-Wen, W., Jie ZHANG, F. Z., Wei-Ping, L., Yan-Wu ZHANG, Y.-X. L., Yong-Jie, F., ... others (2019). Introduction of bcc models and its participation in cmip6. *Advances in Climate Change Research*, *15*(5), 533.