Geophysical Research Letters

RESEARCH LETTER
10.1029/2022GL102071

Decadal Predictability of the North Atlantic Eddy-Driven Jet in Winter

Andrea Marcheggiani[1,2,3], Jon Robson[1], Paul-Arthur Monerie[1], Thomas J. Bracegirdle[4], and Doug Smith[5]

[1]Department of Meteorology, National Centre for Atmospheric Science, University of Reading, Reading, UK, [2]Geophysical Institute, University of Bergen, Bergen, Norway, [3]Bjerknes Centre for Climate Research, Bergen, Norway, [4]British Antarctic Survey, Cambridge, UK, [5]Met Office, Exeter, UK

Key Points:

- The winter North Atlantic eddy-driven jet is predictable on decadal time-scales with skill in jet speed comparable to that of the North Atlantic Oscillation (NAO)
- Anomalies in the jet are substantially smaller than expected from skill alone and so suffer from noise overestimation
- Skill drops significantly in the last decade, as hindcasts do not capture the return to positive NAO conditions post-2010

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Marcheggiani and J. Robson,
andrea.marcheggiani@reading.ac.uk;
j.i.robson@reading.ac.uk

Citation:

Marcheggiani, A., Robson, J., Monerie, P.-A., Bracegirdle, T. J., & Smith, D. (2023). Decadal predictability of the North Atlantic eddy-driven jet in winter. *Geophysical Research Letters*, *50*, e2022GL102071. https://doi.org/10.1029/2022GL102071

Received 9 NOV 2022
Accepted 22 MAR 2023

Author Contributions:

Conceptualization: Jon Robson
Data curation: Andrea Marcheggiani, Paul-Arthur Monerie
Formal analysis: Paul-Arthur Monerie
Funding acquisition: Jon Robson
Investigation: Andrea Marcheggiani, Jon Robson, Paul-Arthur Monerie, Thomas J. Bracegirdle, Doug Smith

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 focusing on decadal predictions made for years 2–9 from the sixth Coupled Model Intercomparison Project over 1960–2005 we find that there is significant skill in jet latitude and, especially, jet speed associated with the skill in the NAO. However, the skill in the NAO, jet latitude and speed indices appears to be sensitive to the period over which it is assessed. In particular, skill drops considerably when evaluating hindcasts up to the present day as models fail to capture the recent observed northern shift and strengthening of the winter eddy-driven jet, and more thus positive NAO. We suggest that the drop in atmospheric circulation skill is related to reduced skill in North Atlantic sea surface temperatures.

Plain Language Summary

Climate models have the capability to predict the evolution of mean atmospheric circulation over long timescales, 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. While climate models have demonstrated skill in predicting the North Atlantic Oscillation (NAO), this paper assesses model skill in predicting the evolution of the wintertime North Atlantic eddy-driven jet, with the aim to highlight how much of the predictive skill of the NAO is derived from accurate 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 (speed) over its location (latitude of 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). Atmospheric circulation variability also exerts a strong influence on the climate of the North Atlantic sector and Western Europe (Hall & Hanna, 2018; Sutton et al., 2018; Thompson & Wallace, 2001). Therefore, reliable decadal predictions of the NAO and jet's evolution are of prime societal importance as they would allow more time in the development of adaptation strategies.

Considerable evidence has now emerged showing that the NAO is predictable on seasonal (Scaife et al., 2014) to decadal timescales (Athanasiadis et al., 2020; Smith 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 fifth (CMIP5, Taylor et al., 2012) and sixth (CMIP6, Eyring et al., 2016) Coupled Model Intercomparison Project's hindcasts initialized between 1960 and 2005. Furthermore, S20 showed how NAO predictability 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 significantly underestimated—leading to the so-called signal-to-noise paradox, suggesting that models forecast observations

Methodology: Andrea Marcheggiani, Jon Robson, Paul-Arthur Monerie, Thomas J. Bracegirdle, Doug Smith
Project Administration: Jon Robson
Supervision: Jon Robson
Writing – original draft: Andrea Marcheggiani, Jon Robson
Writing – review & editing: Andrea Marcheggiani, Jon Robson, Paul-Arthur Monerie, Thomas J. Bracegirdle, Doug Smith

more skillfully than their own ensemble members. Large ensembles are required 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 the contribution of changes to the eddy-driven jet on the winter NAO skill in S20. Furthermore, we expect the relationship between the eddy-driven jet and winter NAO to change with the timescale, and thus relate to different processes (Baker et al., 2017; Woollings et al., 2015). 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 skillful prediction of shifts in the jet latitude (Parker et al., 2019). However, decadal timescale winter NAO variability has been linked more to changes in eddy-driven jet speed that, in turn, appear driven by sea surface temperatures in the subpolar North Atlantic (Woollings et al., 2015). The different aspects of jet variability (e.g., latitude and speed) are also known to lead to different impacts on sea ice, temperatures, and precipitation over the North Atlantic ocean basin and western Europe (Hall & Hanna, 2018; Ma et al., 2020). Therefore, evaluating separate features of jet behavior could provide insight on 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 not covered by CMIP5 hindcasts.

2. Data and Methods

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

As in S20, we define the NAO index as the difference in mean sea level pressure (MSLP) between two boxes located around the Azores (28°–20°W, 36°–40°N) and Iceland (25°–16°W, 63°–70°N).

We construct the indices of the eddy-driven jet's latitude (JLI) and speed (JSI) in accordance with Bracegirdle et al. (2018), which draws from Woollings et al. (2010) but uses monthly-rather than daily-averaged data. We calculate the zonal-mean zonal wind at 850 hPa in the North Atlantic sector (60°W−0°, 10°–75°N) and then identify the maximum and its location as the jet's speed and latitude, respectively.

Although the North Atlantic region is the main focus of this study, previous studies have highlighted that this region is particularly sensitive to signal-to-noise problems (Eade et al., 2014). Therefore, we also include an analysis of the Arctic Oscillation (AO) to put the analysis of NAO skill in a wider context. The AO index is calculated as the difference in MSLP between the midlatitudes (30°–60°N) and the high/polar latitudes (60°–90°N). This is an approximation to the canonical definition based on empirical orthogonal functions (Thompson & Wallace, 1998), but we find that the resulting indices are highly correlated (+0.98 in winter, not shown).

As in S20, we assess 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, thereby excluding their first winter season), our analysis only considers 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 overall starting dates and ensemble members, see Figure S1 in Supporting Information S1) from each ensemble member, 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 against 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 MSLP data from HadSLP2 (Allan & Ansell, 2006) to compute the observed NAO, which yields a lower variance than that from ERA5 (not shown). Although we expect skill scores to be different when using a different data set, we do not expect this difference in variance to be a major influence on overall skill as measured by anomaly correlation coefficient (ACC), which is sensitive to the phasing of the variability rather than its magnitude.

We measure the skill by evaluating the Pearson 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),

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

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"). For a perfect model, the RPC should be equal to one, while, in a scenario of low signal-to-noise ratio, models overestimate the unpredictable component of atmospheric variability, which corresponds to a RPC higher than one (Smith et al., 2019). As in S20, we test the statistical significance of the ACC estimates by using a block bootstrap approach.

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, 1962–2014) and a *long* period, which includes hindcasts initialized at the end of years 2006–2012 (thus covering 1962–2021).

3. Skill in the NAO and Jet Stream Indices

We first examine the 2–9 years prediction skill of DCPP-A for the NAO, 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 pronounced decadal and multidecadal variability (black curves in Figure 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 unable 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 Figure 1). Nonetheless, models demonstrate 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 ($P < 0.01$), comparable 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 improvement to NAO predictions by computing the lagged ensemble mean and by rescaling the variance to the observed. By pooling hindcasts in this way it is effectively assumed that the predictability of anomalies at years 2–9 varies more slowly than year-to-year changes in initial conditions. Hence, combining hindcasts reduces the unpredictable noise further. The resulting model predictions (thick red curves in right-hand panels of Figure 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 Figure 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 characteristic of climate models (Scaife & Smith, 2018), whereby a larger sample of the unpredictable, noisy component of climate variability is needed to improve the strength of the signal. This is particularly true for the North Atlantic, where skill is especially sensitive to the number of ensemble members (Eade et al., 2014). 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 Figures 1g and 1h.

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 Figures 1c and 1e). Models have higher ACC 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, Figure 1d) and the speed

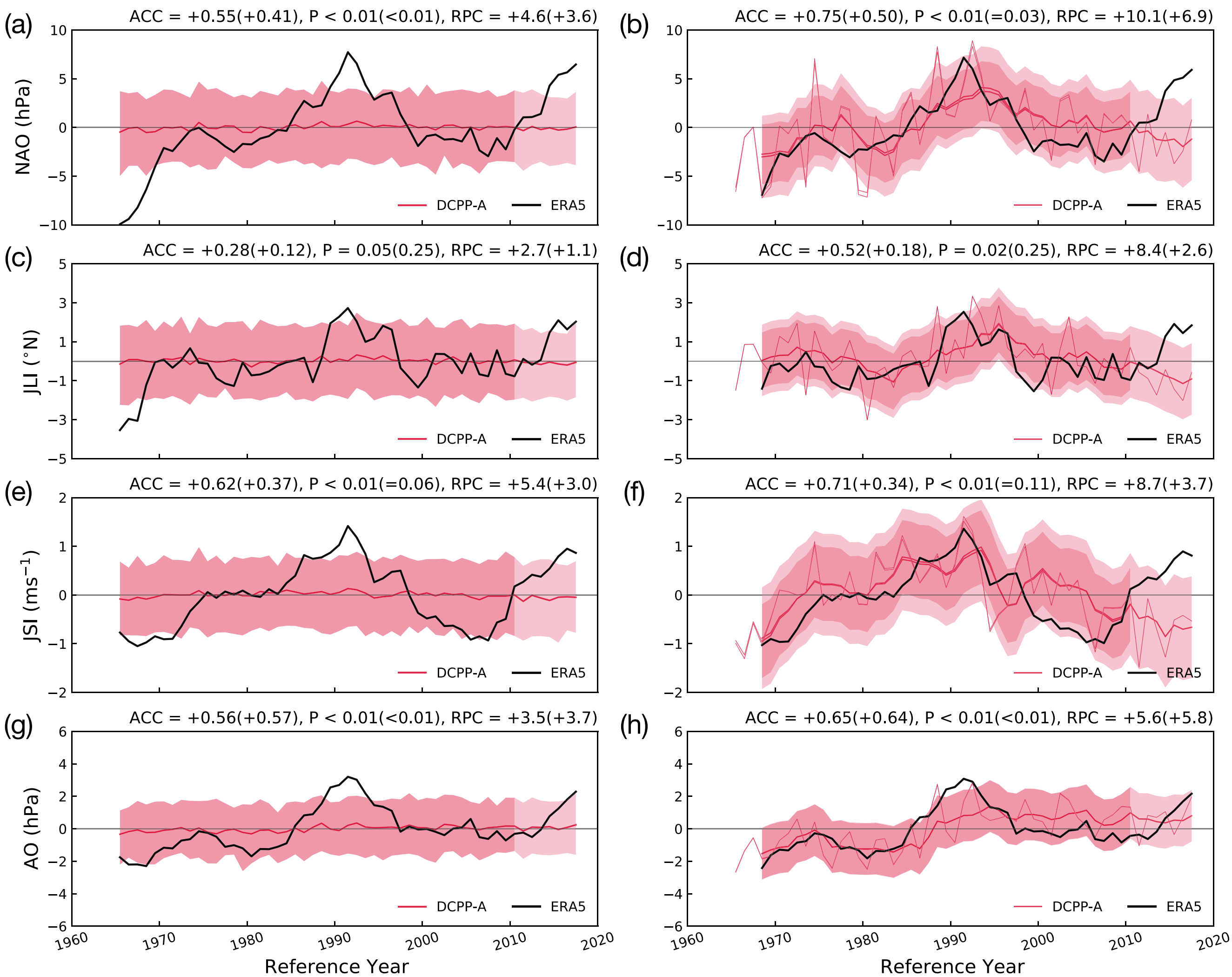

Figure 1. Evolution of 8-year running mean observed (black) and year 2–9 predictions from DCPP-A hindcasts (red) extended boreal winter (DJFM) NAO (a, b), Jet Latitude (c, d), Jet Speed (e, f) and AO (g, h) indices. Panels on the left show the raw ensemble-mean prediction (i.e., no rescaling of variance). Panels on the right are the same as those on the left, but showing the ensemble-mean forecast (thin red, comprised of 153 ensemble members) rescaled to have the same variance as the observations, and the lagged ensemble-mean forecast (thick red, comprised of 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).

(0.71, Figure 1f). The RPC also becomes larger, more than tripling for JLI (2.7–8.4), while the increase is more moderate for JSI (5.4–8.7). Therefore, the similar level 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 a wide range of lead times, as we observe comparable skill in NAO and JSI predictions (see Figure S2 in Supporting Information S1).

Figure 1 and previous work (e.g., Klavans et al., 2021; Scaife & Smith, 2018) 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 greatest skill (not shown). Therefore, we question 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 compute the skill for a random selection of different ensemble members that make up the lagged ensemble mean (612 members), and gradually increase the size of the selection.

10.1029/2022GL102071

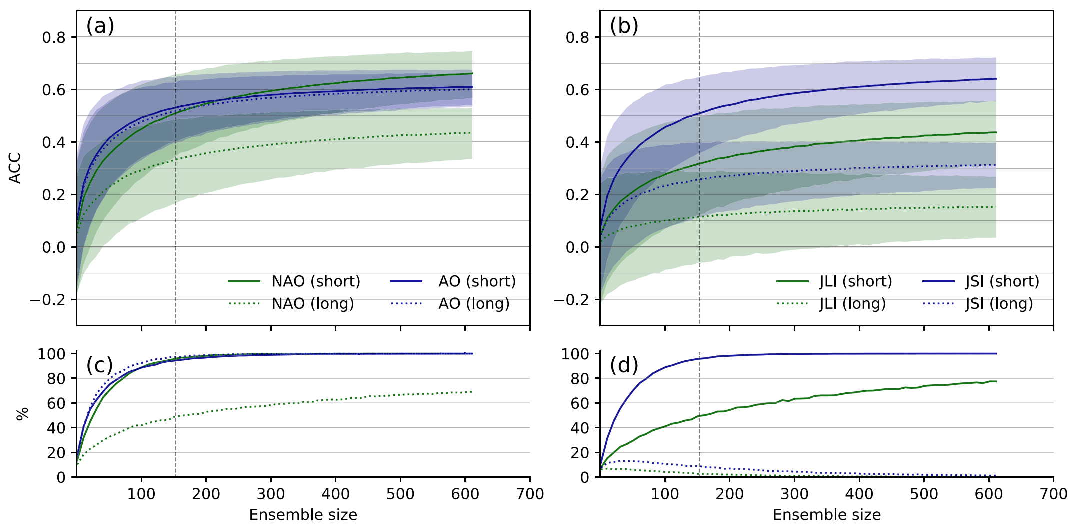

Figure 2. (a) Relationship between ensemble size and skill (ACC) at predicting the NAO and AO (green and blue, respectively) with lagged ensemble means, for the short (solid lines) and long (dotted lines) periods. Shading represents 5th–95th percentiles of the distribution of ACCs from 10,000 random combinations with replacement of a number of ensemble members; lines indicate the mean of such distributions. (b) As in (a), for JLI (green) and JSI (blue). Skill is estimated for the 2–9-year lead period in all ensemble members. Vertical lines denote the total multi-model ensemble size from DCPP-A of CMIP6, 153 members. (c) and (d) show the percentage of significant ACC scores among the 10,000 that make up the distributions for each ensemble size in panels (a) and (b).

Figure 2 shows the resulting skill at predicting the atmospheric indices considered herein as a function of ensemble size. Consistent with the evaluation of skill in Figure 1, the indices exhibit different levels of skill. However, it also shows that skill in the NAO, jet latitude, and jet speed appears to still be increasing when using the maximum number of ensemble members (612), suggesting that ACC skill could be expected to increase further with a larger number of ensemble members. We note that the shading in Figures 2a and 2b does not represent the uncertainty associated with the estimation of the correlation score, but rather indicates the spread in the distribution of the randomly selected combinations. Lower panels in Figures 2c and 2d show the percentage of the 10,000 combinations featuring a significant ACC for each ensemble size, which provides an indication of the significance of the spread in ACC values. In the short period, as an example, ensembles with more than 150 members are characterized by significant ACCs for all indices (apart from JLI, which reaches a maximum of less than 80% for the largest ensemble size here considered, Figure 2d), while significance in the long period is achieved almost exclusively for the AO index.

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 dependent on how basin-wide variability is measured (Bracegirdle, 2022).

4. Degradation of Skill in the Recent Period

The previous section, and results in Figure 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 with more observational data available for evaluation. Therefore, 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 substantial declines in the skill for the NAO and jet indices. For example, the lagged ensemble skill for the jet latitude and jet speed decreases from +0.52 to +0.71 respectively to statistically insignificant values of +0.18 and +0.34. Skill in the NAO index drops from +0.75 to a 0.50, lower yet still statistically significant. All differences in skill are deemed statistically

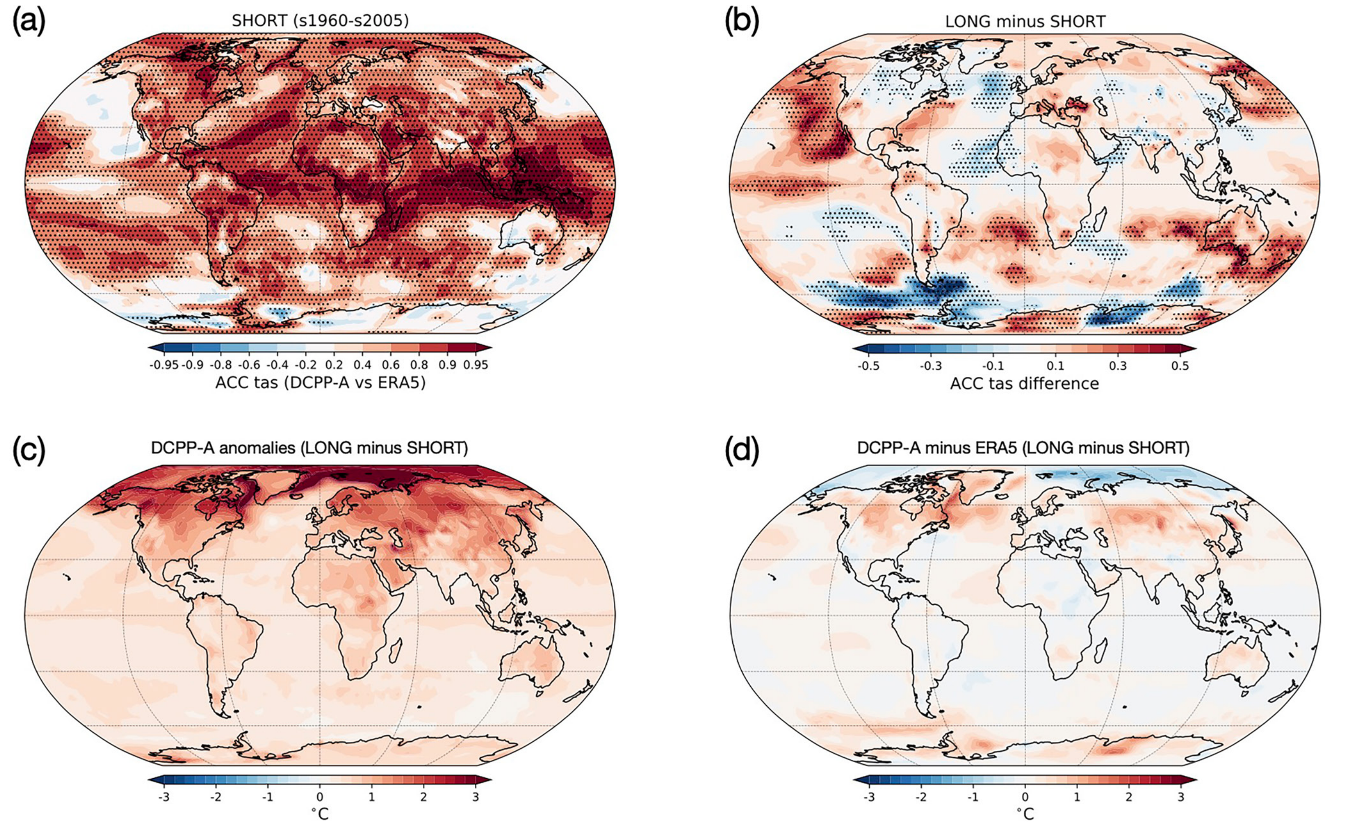

Figure 3. Surface temperature (TAS) skill (as measured by ACC) of year 2–9 hindcast from DCPP-A for the *short* period (a) and the difference, *long* minus *short* (b). Panels (c), (d) show changes in TAS over the latest decade with respect to the *long* period (i.e., *long* minus *short*) in DCPP-A 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).

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 (Figure 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 inspection of Figure 2. However, it is clear that model skill is not only lower over the *long* period, but that the increase in skill with ensemble size achieves saturation with fewer members (except for the AO index, whose skill remains virtually unaltered across the two periods, consistent with the lack of skill degradation seen in Figure 1).

In addition to atmospheric variables in the North Atlantic, surface temperature over the ocean exhibits a considerable decline in skill. Figures 3a and 3b shows the skill of DCPP-A hindcasts at predicting temperatures near the surface (TAS). For the short period (Figure 3a) there is significant skill over the majority of the globe, with particularly strong skill in the North Atlantic, tropical Atlantic Ocean, and 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 tropical North Atlantic (Figure 3b). This reduction of skill over the North Atlantic is associated with warm forecast errors in DCPP-A predictions over the subpolar North Atlantic, as shown by most recent changes (from the end of short period to the end of long period, i.e., 8-year periods spanning between 2008 and 2021) in TAS in DCPP-A models (Figure 3c) and the deviation of DCPP-A models from observations (Figure 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, warmer surface temperatures over the North Atlantic Ocean would also be expected due to the failure to predict the positive

NAO (e.g., positive NAO drives increased oceanic heat loss, Marshall et al., 2001; Grist et al., 2010), with other relevant factors possible. For example, previous work has highlighted that temperatures in the western tropical Pacific are a key driver of the NAO on decadal timescales (Kucharski et al., 2006; Latif, 2001). Nevertheless, we see no change in skill in this region between the short and long period (Figure 3b), suggesting that this is not the primary cause. External forcings have been linked to both NAO variability (Christiansen, 2008; Ortega et al., 2015; Sjolte et al., 2018) and skill in predictions of subpolar North Atlantic surface temperatures (Borchert et al., 2021). Furthermore, external forcings may explain the NAO skill in the short period (Klavans et al., 2021). However, different forcing factors (e.g., volcanic eruptions, anthropogenic aerosols, etc.) change through time and so it follows that the skill expected from external forcings is also likely to be sensitive to the time period used (Borchert et al., 2021; Sjolte et al., 2018). Additionally, state-dependent predictability of the NAO (Weisheimer et al., 2017) and North Atlantic SSTs (e.g., over the North Atlantic SPG, Brune et al., 2018), as well as state dependence on teleconnections (Fereday et al., 2020; López-Parages & Rodríguez-Fonseca, 2012; Weisheimer et al., 2020) may also play a role. Finally, we note that the drop in skill appears to be largely a North Atlantic atmospheric circulation's phenomenon, as there is no significant drop in skill in the predictability of the hemispheric-wide AO index (ACC values of +0.65 and 0.64 for the short and long period respectively, see Figures 1g and 1h), regardless of the size of the ensemble (Figure 2a). Further investigation is needed to unravel the causes of the reduction 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 and compare the skill of the eddy-driven jet latitude and speed in winter and the winter 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 ACC skill score for years 2–9 of the ensemble mean (after post-processing by considering a lagged-ensemble) is 0.52 and 0.71 for jet latitude and jet speed, respectively, 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 than observations (RPC of 8.4 and 8.7 for the jet latitude and speed, respectively), despite the high ACCs, indicating a low signal-to-noise ratio.
3. The skill for all indices drops substantially when evaluating hindcasts initialized between 1960 and 2012 (rather than 1960–2005). This drop in skill corresponds to a failure of hindcasts to capture both the return to positive NAO conditions post-2010 and the associated 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 and 2012 rather than 1960–2005.

This paper has demonstrated that, alongside the NAO, predictability in the winter North Atlantic eddy-driven jet exists on decadal timescales. In particular, skill at predicting the jet's speed is higher than that for its latitude. This improved skill for jet speed is consistent with Woollings et al. (2015), who showed that winter NAO interannual variability is dominated by jet latitude, while changes on longer time scales are dominated by jet speed. 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), rather than NAO predictions, could be more useful to certain sectors. In addition, efforts could be made to quantify the maximum level of achievable skill for the NAO and jet variability, and thus the optimal number of ensemble members needed to achieve it.

However, the skill in North Atlantic atmospheric circulation in winter is sensitive to the time period studied. Unfortunately, the reasons for this drop in skill remain unclear. Our results suggest that the drop in skill is primarily related to the physical mechanisms of the North Atlantic basin. In fact, the skill in predicting hemisphere-wide variability (e.g., the AO) did not exhibit similar degradation over the most recent period. One potential interpretation is that the drop in skill of the atmospheric variables is consistent with a reduction in skill at capturing

10.1029/2022GL102071

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, unpredictable noise (and its misrepresentation in models), 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, as well as the differences in NAO and AO predictability.

Data Availability Statement

The climate model hindcasts are available via the Earth System Grid Federation (ESGF) archive of the sixth Coupled Model Intercomparison Project (CMIP6) data (https://esgf-index1.ceda.ac.uk/projects/esgf-ceda/). ERA5 data is available from the ECMWF at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5.

Acknowledgments

We thank three anonymous reviewers for their insightful comments which helped improve the original manuscript. This work was funded by the Natural Environment Research Council (NERC) via the ACSIS project (NE/N018001/1 and NE/N018028/1). AM was additionally supported by the Research Council of Norway (NFR) via the BALMCAST project (NFR Project 324081). JR was additionally funded by NERC via the WISHBONE (NE/T013516/1) and CANARI (NE/W004984/1) projects, 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 UK 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. https://doi.org/10.1175/jcli3937.1

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. https://doi.org/10.1038/s41612-020-0120-6

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. https://doi.org/10.1175/JCLI-D-16-0864.1

Bell, B., Hersbach, H., Simmons, A., Berrisford, P., Dahlgren, P., Horányi, A., et al. (2021). The ERA5 global reanalysis: Preliminary extension to 1950. *Quarterly Journal of the Royal Meteorological Society*, *147*(741), 4186–4227. https://doi.org/10.1002/qj.4174

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., et al. (2016). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, *9*(10), 3751–3777. https://doi.org/10.5194/gmd-9-3751-2016

Borchert, L. F., Menary, M. B., Swingedouw, D., Sgubin, G., Hermanson, L., & Mignot, J. (2021). Improved decadal predictions of North Atlantic subpolar gyre SST in CMIP6. *Geophysical Research Letters*, *48*(3), e2020GL091307. https://doi.org/10.1029/2020GL091307

Bracegirdle, T. J. (2022). Early-to late-winter 20th century North Atlantic multidecadal atmospheric variability in observations, CMIP5 and CMIP6. *Geophysical Research Letters*, *49*(11), e2022GL098212. https://doi.org/10.1029/2022GL098212

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. https://doi.org/10.1029/2018gl078965

Brune, S., Düsterhus, A., Pohlmann, H., Müller, W. A., & Baehr, J. (2018). Time dependency of the prediction skill for the North Atlantic subpolar gyre in initialized decadal hindcasts. *Climate Dynamics*, *51*(5), 1947–1970. https://doi.org/10.1007/s00382-017-3991-4

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. https://doi.org/10.1175/2007jcli1657.1

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. https://doi.org/10.1002/2014gl061146

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. https://doi.org/10.5194/gmd-9-1937-2016

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. https://doi.org/10.1029/2020gl088664

Grist, J. P., Josey, S. A., Marsh, R., Good, S. A., Coward, A., De Cuevas, B. A., et al. (2010). The roles of surface heat flux and ocean heat transport convergence in determining Atlantic Ocean temperature variability. *Ocean Dynamics*, *60*(4), 771–790. https://doi.org/10.1007/s10236-010-0292-4

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. https://doi.org/10.1002/joc.5398

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (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. https://doi.org/10.1126/science.269.5224.676

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. https://doi.org/10.1038/s41612-021-00177-8

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. https://doi.org/10.1007/s00382-005-0085-5

Latif, M. (2001). Tropical Pacific/Atlantic Ocean interactions at multi-decadal time scales. *Geophysical Research Letters*, *28*(3), 539–542. https://doi.org/10.1029/2000gl011837

López-Parages, J., & Rodríguez-Fonseca, B. (2012). Multidecadal modulation of El Niño influence on the Euro-Mediterranean rainfall. *Geophysical Research Letters*, *39*(2). https://doi.org/10.1029/2011gl050049

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. https://doi.org/10.1175/jcli-d-19-0247.1

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. https://doi.org/10.1175/1520-0442(2001)014<1399:asotio>2.0.co;2

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. https://doi.org/10.1038/nature14518

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. https://doi.org/10.1029/2019gl084402

Robson, J., Ortega, P., & Sutton, R. (2016). A reversal of climatic trends in the North Atlantic since 2005. *Nature Geoscience*, *9*(7), 513–517. 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. https://doi.org/10.1038/18648

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., et al. (2014). Skillful long-range prediction of European and North American winters. *Geophysical Research Letters*, *41*(7), 2514–2519. 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. https://doi.org/10.1038/s41612-018-0038-4

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. https://doi.org/10.1038/s41561-019-0391-x

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. https://doi.org/10.5194/cp-14-1179-2018

Smith, D., Eade, R., Scaife, A., Caron, L.-P., Danabasoglu, G., DelSole, T., et al. (2019). Robust skill of decadal climate predictions. *NPJ Climate and Atmospheric Science*, *2*(1), 13. https://doi.org/10.1038/s41612-019-0071-y

Smith, D. M., Scaife, A. A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., et al. (2020). North Atlantic climate far more predictable than models imply. *Nature*, *583*(7818), 796–800. https://doi.org/10.1038/s41586-020-2525-0

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. https://doi.org/10.1175/bams-d-16-0266.1

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. https://doi.org/10.1175/bams-d-11-00094.1

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. (1998). The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical Research Letters*, *25*(9), 1297–1300. https://doi.org/10.1029/98gl00950

Thompson, D. W., & Wallace, J. M. (2001). Regional climate impacts of the northern hemisphere annular mode. *Science*, *293*(5527), 85–89. https://doi.org/10.1126/science.1058958

Weisheimer, A., Befort, D. J., MacLeod, D., Palmer, T., O'Reilly, C., & Strømmen, K. (2020). Seasonal forecasts of the twentieth century. *Bulletin of the American Meteorological Society*, *101*(8), E1413–E1426. https://doi.org/10.1175/bams-d-19-0019.1

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. https://doi.org/10.1002/qj.2976

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. https://doi.org/10.1007/s00382-014-2237-y

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. https://doi.org/10.1002/qj.625

References From the Supporting Information

Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., et al. (2021). NorCPM1 and its contribution to CMIP6 DCPP. *Geoscientific Model Development*, *14*(11), 7073–7116. https://doi.org/10.5194/gmd-14-7073-2021

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., et al. (2020). Presentation and evaluation of the IPSL-CM6A-LR climate model. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS002010. https://doi.org/10.1029/2019MS002010

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., et al. (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. https://doi.org/10.1029/2018MS001369

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D., DuVivier, A., Edwards, J., et al. (2020). The community Earth system model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*(2), e2019MS001916. https://doi.org/10.1029/2019ms001916

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., et al. (2021). The EC-EARTH3 Earth system model for the Climate Model Intercomparison Project 6. *Geoscientific Model Development Discussions*, 1–90.

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., et al. (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. https://doi.org/10.1029/2017ms001217

Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019). The Canadian Earth system model version 5 (CANESM5. 0.3). *Geoscientific Model Development*, *12*(11), 4823–4873. https://doi.org/10.5194/gmd-12-4823-2019

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., et al. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, *12*(7), 2727–2765. https://doi.org/10.5194/gmd-12-2727-2019

Williams, K., Copsey, D., Blockley, E., Bodas-Salcedo, A., Calvert, D., Comer, R., et al. (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. https://doi.org/10.1002/2017ms001115

Xiao-Ge, X., Tong-Wen, W., Jie Zhang, F. Z., Wei-Ping, L., Yan-Wu Zhang, Y.-X. L., Yong-Jie, F., et al. (2019). Introduction of BCC models and its participation in CMIP6. *Advances in Climate Change Research*, *15*(5), 533.