Predictable Decadal Forcing of the North Atlantic Jet Speed by Sub-Polar North Atlantic Sea Surface Temperatures

Kristian Strommen[a], Tim Woollings[a], Paolo Davini[b], Paolo Ruggieri[c], and Isla R. Simpson[d]

[a]Department of Physics, University of Oxford
[b]Istituto di Scienze dell'Atmosfera e del Clima, Consiglio Nazionale delle Ricerche (CNR-ISAC), Torino
[c]Department of Physics and Astronomy, University of Bologna
[d]Climate and Global Dynamics Laboratory, National Centre for Atmospheric Research, Boulder CO

Correspondence: Kristian Strommen (kristian.strommen@physics.ox.ac.uk)

Abstract. It has been demonstrated that decadal variations in the North Atlantic Oscillation (NAO) can be predicted by current forecast models. While Atlantic Multidecadal Variability (AMV) in sea surface temperatures (SSTs) has been hypothesised as the source of this skill, the validity of this hypothesis and the pathways involved remain unclear. We show, using reanalysis and data from two forecast models, that the decadal predictability of the NAO can be entirely accounted for by the predictability of decadal variations in the speed of the North Atlantic eddy-driven jet, with no predictability of decadal variations in the jet latitude. The sub-polar North Atlantic (SPNA) is identified as the only obvious common source of an SST-based signal across the models and reanalysis, and the predictability of the jet speed is shown to be consistent with a forcing from the SPNA visible already within a single season. The pathway is argued to be tropospheric in nature, with the SPNA-associated heating extending up to the mid-troposphere, which alters the meridional temperature gradient around the climatological jet core. The link between SSTs and surface temperatures, which mediates this pathway, is shown to be underestimated in the forecast models by approximately a factor of two, with potential implications for the 'signal-to-noise paradox'. The relative roles of anthropogenic aerosol emissions and the AMOC at generating predictable SPNA variability are also discussed. The analysis is extensively supported by the novel use of a set of seasonal hindcasts spanning the 20th century and forced with prescribed SSTs.

1 Introduction

European winter weather is strongly influenced by the variability of the North Atlantic eddy-driven jet, and it is therefore of high societal value to predict this variability as far in advance as possible. Recent studies have shown that, remarkably, retrospective ensemble forecasts ('hindcasts') are now able to skillfully predict some components of the low-frequency variability of the winter jet at lead times of up to 10 years (Smith et al., 2019; Athanasiadis et al., 2020). However, the exact source of the predictable signal and mechanisms involved remain unclear, making it uncertain to what extent one can rely on this skill to remain for genuine decadal forecasts of the future. The aim of this paper is to try to clarify these points.

In Simpson et al. (2018), by considering a jet index based on zonal winds at 700hPa, it was shown that the decadal variability of the jet is much stronger in March than in the boreal winter months December, January, February (DJF). This late winter jet

variability was argued to arise from the internally generated component of Atlantic Multidecadal Variability (AMV) in North Atlantic sea surface temperatures (SSTs), though the mechanisms were not elucidated; they also showed that models failed to capture the connection. In alignment with the enhanced decadal variability, the observed connection between the AMV and the jet was shown to be far greater in March than during DJF. Nevertheless, Smith et al. (2019) and Athanasiadis et al. (2020) showed that skillful decadal forecasts of the DJF-averaged North Atlantic Oscillation (NAO) could be achieved, and the latter suggested that the skill appeared to be driven by the AMV. Both Smith et al. (2019) and Athanasiadis et al. (2020) emphasised an apparent 'signal-to-noise paradox' in these forecasts, mimicking the phenomenon observed for seasonal winter NAO forecasts (Scaife and Smith, 2018). This 'paradox' effectively says that the real world appears to be much more predictable than the forecast model thinks it is, with the model underestimating the response to forcing or the response to predictable boundary conditions on seasonal-to-decadal timescales relative to the unpredictable noise. A practical consequence of this behaviour is that models are likely underestimating the predictable component of decadal winter jet variability, suggesting that the results of Simpson et al. (2018) and Athanasiadis et al. (2020) are consistent with a hypothesis that the AMV is driving predictable decadal jet[1] variability from December through March. In fact, several studies have found that even the total decadal variability appears to be systematically underestimated in models (Kravtsov, 2017; Wang et al., 2017; Kim et al., 2018; Simpson et al., 2018; Bracegirdle, 2022).

Numerous studies have been conducted on the potential for air-sea coupling to generate links between the AMV and the jet, starting with the pioneering work of Bjerknes (1964). The decadal variability of the AMV itself has been hypothesised to be driven by a combination of the Atlantic Meridional Overturning Circulation (AMOC) (Bjerknes, 1964; Delworth et al., 1993; Kushnir, 1994), anthropogenic aerosols and other greenhouse gases (Booth et al., 2012; Bellomo et al., 2018; Robson et al., 2022), and stochastic atmospheric forcing (Hasselmann, 1976; Clement et al., 2015; O'Reilly et al., 2019). An excellent recent overview on these topics with more complete references can be found in Zhang et al. (2019). Different mechanisms have been put forward for how the AMV affects the jet, including the direct modulation of low-level baroclinicity and stationary waves by the North Atlantic SST anomalies (Kushnir, 1994; Msadek et al., 2011; Kushnir et al., 2002; Peings et al., 2016); forcing from the tropical Atlantic (Davini et al., 2015); and stratospheric pathways (Omrani et al., 2014). However, the response in climate models to imposed AMV anomalies appears inconsistent and model dependent (Ruggieri et al., 2021), and the period of highly reliable observational data is short, making it challenging to distinguish between different hypotheses.

One major source of uncertainty in and across many studies is that the decadal variability is an average over several different processes occurring on different timescales, due to (a) the continuous nature of air-sea coupling and (b) the fact that the AMV pattern itself evolves over time, with the anomalies in the sub-polar North Atlantic (SPNA) arising first before propagating towards the tropical Atlantic (Zhang et al., 2019; Wills et al., 2019). This makes it difficult to attribute causality between atmospheric versus oceanic forcing, and also to distinguish between the role played by particular regions in the Atlantic ocean, such as the sub-polar versus tropical North Atlantic. However, several studies have emphasised the importance of the SPNA in particular (Gastineau and Frankignoul, 2015; Woollings et al., 2015; Ortega et al., 2017; Wills et al., 2019), especially on longer timescales (Delworth et al., 2017).

[1]Since the NAO is largely describing the variability of the jet, we will conflate these without comment for the remainder of the paper.

In this paper, we make crucial use of two techniques to help address these challenges:

1. The separation of the eddy-driven jet into two components, corresponding to the speed and latitude of the jet.

2. The use of two seasonal hindcast ensembles, named ASF20C and CSF20C, spanning the period 1900-2010. ASF20C is forced with prescribed, observed boundary conditions (Weisheimer et al., 2017), while CSF20C is fully coupled (Weisheimer et al., 2020) (more details in Section 2.1).

The first point is motivated by the fact that the variability and sensitivity of the latitude and speed of the jet are very different. The jet latitude exhibits multimodality and considerable variability on seasonal timescales, but shows no significant variability on decadal timescales beyond white noise (Woollings et al., 2010, 2014). The jet speed, on the other hand, is approximately Gaussian across all timescales and exhibits robust decadal variability (Woollings et al., 2014). Furthermore, Baker et al. (2017) showed that the latitude and speed respond differently to thermal forcing, and Woollings et al. (2015) showed that the processes responsible for latitudinal shifts in the jet clearly differ from those responsible for changes to the jet speed. This means that analysis based on single indices which amalgamate the latitude and speed (such as the NAO index) may struggle to identify robust links between the jet and SST anomalies. This approach has also recently been taken in Marcheggiani et al. (2023) using a complementary set of decadal forecasts.

To motivate the second point, we note that existing decadal forecasts only go back to 1954 at the earliest, leaving them with a relatively small effective sample size once any low-pass filtering or decadal averaging has taken place. The period 1954 to present also does not adequately sample the variability associated with the AMV and the AMOC. There is therefore great value in extending the analysis back to 1900. While taking decadal averages of a seasonal hindcast obviously does not constitute an actual decadal *forecast*, it nevertheless turns out to be useful to think of it as being like a 'nudged' forecast, where both the atmospheric and oceanic state are being nudged back towards observations at the start of each winter (and moreover, for ASF20C, the SSTs are being forecasted perfectly). We will show that in fact the decadal variability reproduced by the seasonal hindcasts completely matches the predictable decadal variability of a genuine decadal forecast ensemble, justifying this perspective post hoc. This not only allows us to confidently use ASF20C/CSF20C to extend our analysis back to 1900, but also introduces two considerable benefits: the lack of coupling in ASF20C simplifies the question of causality between ocean and atmosphere, and the fact that the forecasts making up ASF20C/CSF20C only cover a single season simplifies the question of timescales.

We will show that on decadal timescales there is no predictability of the latitude of the jet, and that all the observed skill at predicting the winter NAO arises from the predictability of the *speed* of the jet. By comparing observations with the ASF20C/CSF20C seasonal hindcasts and the Decadal Prediction Large Ensemble (DPLE) decadal forecasts (Yeager et al., 2018), we argue that predictable forcing of the jet speed arises from SST anomalies in the SPNA. Furthermore, we argue that the predictable forcing occurring on decadal timescales arises as the accumulation of a smaller forcing taking place already on seasonal timescales, with no need to invoke mechanisms spanning multiple seasons as in several other studies (Peings and Magnusdottir, 2014b; Kwon et al., 2020). Finally, we argue that the response of the jet speed to SST anomalies in the SPNA

can be understood simply as the adjustment of the jet to changes in the tropospheric meridional temperature gradient across the climatological jet core.

The structure of the paper is as follows. The data and methods are described in Section 2. In Section 3 we assess the predictability of the NAO, jet speed and jet latitude in the models, while Section 4 is devoted to identifying potential sources of skill from SSTs. Section 5 addresses the pathways, mechanisms and the 'signal-to-noise paradox', while Section 6 considers the roles of sulphate aerosols and the AMOC at generating predictable decadal variability in the SPNA. The results are finally summarised and discussed in Sections 7 and 8.

2 Data and Methods

2.1 Data

2.1.1 ERA20C

We use the 20th century reanalysis dataset ERA20C, spanning 1890-2010, as our observational 'truth' (Poli et al., 2013). This reanalysis is constructed using a cycle of the Integrated Forecast System (IFS) forecast model, and is made up of consecutive 1-day forecasts with data assimilation. Due to differences in available observations between the beginning and end of the 20th century, the atmospheric component of ERA20C only assimilates surface pressure, in order to maintain coherence over the whole period. Both ocean and sea-ice boundary conditions come from the HadISST2.1.0.0 dataset (Rayner et al., 2003). Although it is known that the internal variability is underestimated in the early 20th century (Dell'Aquila et al., 2016), ERA20C constitutes a reasonable reference for the status of the North Atlantic climate.

2.1.2 ASF20C and CSF20C

The ASF20C model data considered comes from an atmosphere-only seasonal hindcast experiment covering the 20th century (Weisheimer et al., 2017). A 51 ensemble member seasonal forecast is initialised every 1st of November from 1901 to 2010 and allowed to run for 4 months, thereby producing a December-January-February (DJF) prediction for every year in this period. The model used is version CY41R1 of the IFS. Its spectral resolution is TL255, corresponding to roughly 80km grid spacing near the equator, with 91 levels in the vertical. The model is run in atmosphere-only mode with prescribed observed sea-surface temperatures (SSTs) with boundary conditions and initial conditions from ERA20C. Further details can be found in Weisheimer et al. (2017).

We will additionally make use of the CSF20C hindcast. This hindcast is identical to ASF20C except it is run with dynamic coupling between the atmosphere and ice/ocean, and is initialised using the coupled reanalysis CERA20C (Laloyaux et al., 2018). The configuration is described in Weisheimer et al. (2020) and is similar to the SEAS5 operational seasonal forecast at the European Centre for Medium-range Weather Forecasts (Johnson et al., 2019). The ocean model used is NEMO version 3.4 (Madec and the NEMO team, 2016), and the ice model is LIM2 (Fichefet and Maqueda, 1997): both are run at a 1 degree horizontal resolution.

2.1.3 Decadal Prediction Large Ensemble (DPLE)

DPLE is made up of a suite of 40-member ensemble forecasts, each initialised on November 1st and run for 10 years. Forecasts are initialised every year from 1954 to 2015. The forecasts are run using CESM version 1.1, using the same model and component configuration as that used in the CESM1 large ensemble (Kay et al., 2015). The atmosphere component is version 5 of the Community Atmosphere Model (CAM5: Hurrell et al. (2013)), with a horizontal resolution of around 1 degree and 30 levels in the vertical. The ocean component is version 2 of the Parallel Ocean Program (Danabasoglu et al., 2012) and the sea ice model is version 4 of the Los Alamos National Laboratory (LANL) Community Ice Code (Hunke et al., 2010). Both are run at a 1 degree spatial resolution. Further details can be found in Yeager et al. (2018).

To be consistent with CSF20C/ERA20C we restrict to the overlapping period 1954-2010. When assessing the decadal forecast skill of DPLE, we always take averages over the entire 10-year period. For example, suppose we have a timeseries J made up of DJF averages of some quantity in reanalysis, and we want to assess the capacity of DPLE to predict $\overline{J}$, where overline denotes the average across the 10 winters from November 1954 to February 1964. Then the DPLE forecast of $\overline{J}$ is taken to be the ensemble mean over $\overline{x_k}(k = 1, \ldots, 40)$, where the x_k are the 10-year forecasts initialised on November 1st 1954, with k referring to ensemble member. By doing this for consecutive 10-year periods we obtain estimates of the decadal variability predicted by DPLE which we can correlate with the observed decadal variability.

Importantly, we do not perform any drift-correction of any kind in our analysis. The main goal of this paper is to understand how the atmosphere responds to decadally varying SSTs, and this can be assessed in DPLE irrespective of any drift taking place. Furthermore, it is not customary to de-drift seasonal forecasts, so no drift correction is done for ASF20C/CSF20C: not de-drifting DPLE therefore makes the analysis more directly comparable between the forecast products. For DPLE we will only ever consider two timescales: the response taking place in the first season or the 10-year mean across the whole forecast period. The drift taking place in the former is small, and the imprint of the drift in the latter is smoothed out by the large averaging-window. Finally, we note that Athanasiadis et al. (2020) found that robust decadal NAO forecast skill can be diagnosed in DPLE irrespective of whether drift-correction is carried out or not.

2.1.4 CMIP data

In order to help assess the mechanisms, we will be examining intermodel spread in the mean state of the jet across a number of climate model simulations. We analyse data drawn from the 5th (CMIP5) and 6th (CMIP6) phases of the coupled model inter-comparison project. We analyse the CMIP6 historical experiments for 31 CMIP6 models (Eyring et al., 2016) detailed in Table A1 of Appendix A in the Supporting Information (SI): only one member was selected for each model. We also analyse a total of 71 ensemble members from 28 distinct CMIP5 models (Taylor et al., 2012) detailed in Table A2. These historical experiments consist of coupled uninitialised climate runs forced with historical greenhouse gas and aerosol forcings over the 20th century, with initial conditions sampled from a free-running pre-industrial control run.

To further boost the sample size, we also make use of model data produced as part of the PRIMAVERA project (Roberts et al., 2018b), detailed in Table A3. These coupled simulations all follow the HighResMIP protocol (Haarsma et al., 2016),

and are therefore initialised in 1950, following a short 50-year spin-up. The simulations span the 65 years between 1950 and 2015. The historical forcing data is the same as for CMIP6, with the exception of aerosols and land-use. Six underlying models were used, each run at a number of different resolutions: CMCC-CM2 (Cherchi et al., 2019), CNRM-CM6 (Voldoire et al., 2019), EC-Earth3 (Haarsma et al., 2020), ECMWF-IFS (Roberts et al., 2018a), HadGEM3-GC31 (Williams et al., 2018), MPI-ESM1-2 (Gutjahr et al., 2019), and AWI-CM-1.0 (Sein et al., 2017).

Note that the model data is the same as that used in Dorrington et al. (2022), and the tables of this paper have been reproduced from there with permissions. In total, the climate model data used consists of more than 8400 DJF seasons generated from 76 different climate models.

In addition to the historical CMIP6 data, we make use of two piControl CMIP6 (500-year and 603-year long respectively) integrations from EC-Earth3 (Döscher et al., 2022). These simulations are atmosphere-ocean coupled runs with pre-industrial forcings. Such simulations are specifically interesting owing to their large internal variability induced by an internally-driven centennial oscillation of the AMOC (Meccia et al., 2022).

2.2 Methods

2.2.1 Metrics

The NAO index is computed as the leading EOF of daily deseasonalized DJF 500hPa geopotential height anomalies in the Euro-Atlantic sector 30-90N, 80W-40E. A seasonal cycle is estimated by taking the average daily NAO value, for each day in DJF, across all years available: this cycle is then removed. When computing the NAO index for ensemble forecast data, all the ensemble member data is used to compute the EOF, after which each members geopotential height field is projected onto the EOF pattern to obtain the individual NAO indices.

To compute indices of the jet speed and latitude, we follow the simplified methodology of Parker et al. (2019). Wind fields are first interpolated to a regular 1 degree grid. Daily DJF 850hPa zonal winds are then restricted to the region 15-75N, 60-20W and smoothed with a 5-day running mean. For any given day, the jet is said to be located at the latitude where the magnitude of the zonally averaged winds in this region is maximum. The jet latitude on that day is precisely this latitude, while the jet speed is the magnitude of the maximum. As with the NAO, a seasonal cycle is removed.

An 'SPNA' index measuring SST variability in the sub-polar North Atlantic is also used extensively. This is defined as the DJF-averaged SSTs, averaged over the region 49-57N, 50-25W. The motivation for this precise choice is given in the main text. The results are not sensitive to small shifts in this region.

Note that we do not remove linear trends from any of these timeseries. The question of whether and how to isolate internal variability in the AMV and related timeseries is not entirely clear (Deser and Phillips, 2021) and trends in the SPNA and jet speed timeseries we primarily consider here are small (e.g. around -0.5% per year on average for the SPNA), with the decadal variability dominated by the oscillations characteristic of internal variability. Removing the trends is thus found to have no meaningful impact on the analysis (e.g. conclusions drawn concerning significance).

2.2.2 Statistics

Our default stance on significance testing is to explicitly specify a statistical model representing the null hypothesis, and then generate significance thresholds by making 10,000 random draws using the model. Because the motivation behind the choice of each statistical model depends on the situation at hand, we introduce each such model in the main text as and when it is required. However, we note that when modelling SSTs we generally make use of the 'Fourier Phase Shuffling' method (Ebisuzaki, 1997). This method can be loosely described as follows. First compute the Fourier transform of the timeseries of interest. Secondly, for each Fourier mode, replace the computed phase by a randomly chosen one. Third, convert the resulting Fourier series back to a timeseries. The resulting randomly generated timeseries is guaranteed to have the same autocorrelation (at all lags) and degrees of freedom as the original timeseries, which is important given the considerable autocorrelation present in the SST timeseries we will be examining. In particular, a null hypothesis modelling SSTs in this way will typically produce much stricter significance thresholds than those that model SSTs using an AR1, which only specifies the autocorrelation at lag 1.

Finally, in cases where ASF20C/CSF20C and DPLE disagree on whether an effect is statistically significant or not, we always choose to place higher credence in ASF20C/CSF20C, due to the fact that (a) their sample size is approximately twice as large ($N = 109$ vs $N = 56$); and (b) they have more ensemble members (51 vs 40). It is possible that such disagreements reflect genuine differences in physical mechanisms, but we do not consider this possibility here.

3 Predictable jet variability in ASF20C and DPLE

3.1 Predictability of the jet speed and not the jet latitude

We first examine what low-frequency jet variability is skillfully reproduced by ASF20C. Figure 1 shows 10 and 30-year running means of the NAO, jet latitude and jet speed for ERA20C and the ASF20C ensemble mean. One aspect of the 'signal-to-noise paradox' is that the ASF20C ensemble mean standard deviation is considerably smaller than that of ERA20C: this can be seen here by comparing the magnitudes of the two y-axes, which highlight that the standard deviation in ERA20C is around 4-5 times greater than that of the ensemble mean. It can be seen that ASF20C skillfully reproduces decadal NAO variability across the entire period 1900-2010, with a correlation coefficient of $\sim 0.4 - 0.7$ depending on the choice of smoothing: using the raw seasonal data gives a correlation of 0.21. The correlations obtained using 10-year smoothing closely match those reported in Smith et al. (2019) and Athanasiadis et al. (2020) using genuine decadal forecasts, suggesting that the decadal forecast skill they reported might extend all the way back to 1900. Figure 1 also clearly shows that ASF20C cannot skillfully reproduce decadal jet latitude variability, but *can* skillfully reproduce decadal jet speed variability. Furthermore, if we regress out the (decadally averaged) ensemble mean jet speed from the ensemble mean NAO and correlate the residual with the observed NAO, we obtain ≈ 0.1 using 10-year averages and ≈ -0.1 using 30-year averages. Therefore, the majority of the skill that ASF20C has at reproducing decadal NAO variability is related to the jet speed. Figure B1 in Appendix B of the SI shows that, similarly, the coupled hindcast CSF20C has significant skill at predicting decadal variations in the speed but not the latitude.

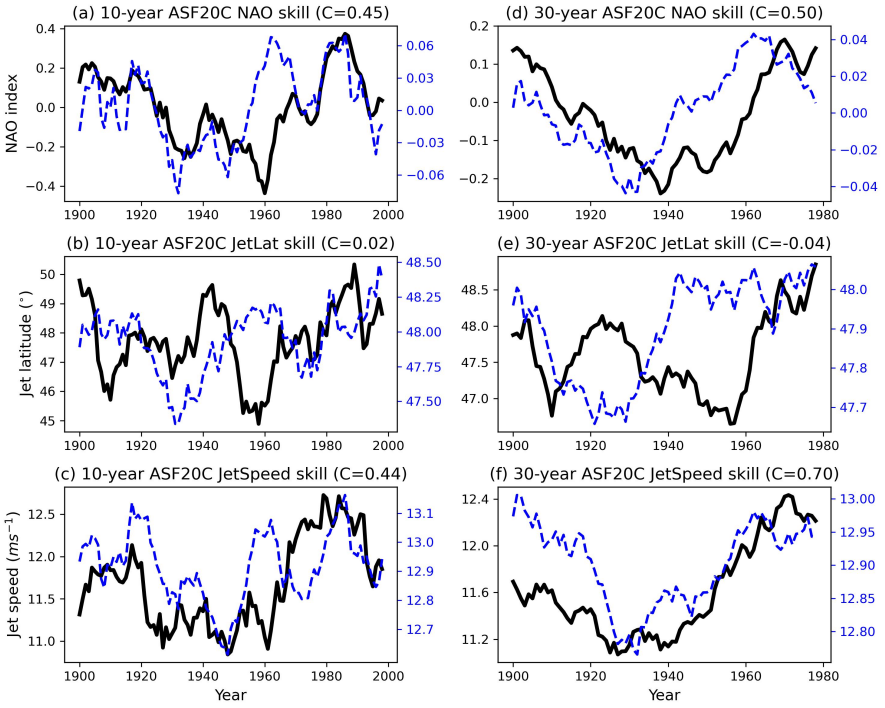

Figure 1. Timeseries of 10-year running DJF means of (a) the NAO, (b) the jet latitude, and (c) the jet speed. The same but with 30-year running means in (d), (e) and (f). The thick black curves are always ERA20C and the dashed blue curves are always the ASF20C ensemble mean. Note the different y-axes for the black and blue curves. The value C in each subplot is the correlation between the two timeseries.

JetSpeed correlations	10 year smoothing	30 year smoothing
Corr(ERA20C, ASF20C)	0.44	0.70
10% significance threshold	±0.42	±0.70
5% significance threshold	±0.48	±0.80
Corr(ERA20C, DPLE)	0.64	0.86
10% significance threshold	±0.62	±0.80
5% significance threshold	±0.71	±0.85

Table 1. DJF jet speed correlations between different data sets at different levels of smoothing. Significance thresholds assume a null hypothesis of the interannual jet speed timeseries being uncorrelated white noise.

Figure 2 shows that the same conclusion is true for the DPLE forecasts: there is no apparent predictability of decadal shifts in the jet latitude, but high skill at predicting shifts in the jet speed. Note that taking 30-year means is much less sensible for DPLE, given its shorter coverage of 56 years, but these are included anyway for direct comparison with Figure 1.

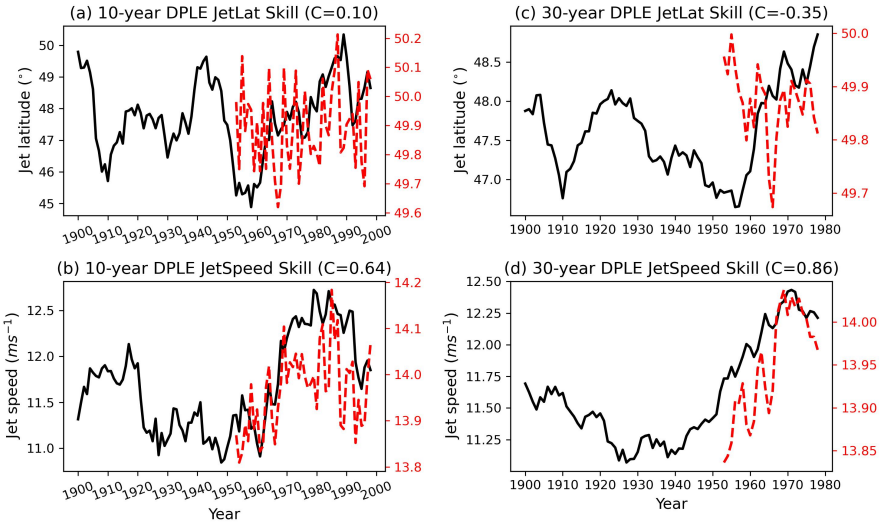

Figure 2. Timeseries of 10-year running DJF means of (a) the jet speed, and (b) the jet latitude. The same but with 30-year running means in (c) and (d). The thick black curves are always ERA20C and the dashed red curves are always the DPLE ensemble mean. Note the different y-axes for the black and red curves. The value C in each subplot is the correlation between the two timeseries.

How significant are the jet speed correlations reported in Figures 1 and 2? The 1-year autocorrelation is ≈ 0.0 for all three data sets, and we are therefore justified in assuming a null hypothesis of the DJF jet speed as being white noise. The approximate Gaussianity of the jet speed distribution has been previously noted (Woollings et al., 2010; Parker et al., 2019). By simulating 10,000 randomly generated DJF jet speed timeseries for each data set, taking 10 or 30-year running means and then computing correlations, we build up a distribution of correlations that can be obtained by chance. Note that taking running means will automatically introduce considerable autocorrelation to the resulting timeseries. Table 1 summarises the result of this significance testing. This shows that for ASF20C, whether considering 10 or 30 year smoothing, the decadal jet speed correlations are only significant to within $p < 0.1$ and not $p < 0.05$: the same is true for CSF20C (see Figure B1). For DPLE, the conclusion is the same except that the correlations using 30-year smoothing appear to be significant also with $p < 0.05$. However, as cautioned already, the small effective sample size here means that this level of confidence is probably not justified. It goes without saying that the jet latitude correlations are never significant. We also note that the relatively large negative correlation emerging for 30-year jet latitude variability in DPLE is not statistically significant when using a similar null hypothesis.

Our results here contrast somewhat with those of Smith et al. (2019) and Athanasiadis et al. (2020), which both report statistically significant decadal NAO forecasts with $p < 0.05$. However, their significance tests differ from ours, and the smoothing they use is also very close, but not identical, to the 10-year running means we use. Table 1 shows that the 10-year ASF20C and DPLE jet speed correlations sit neatly between the bounds of the 5 and 10% significance thresholds, and it is therefore

plausible that these differences can explain why they report a p-value less than 0.05 and we do not. For this article, we will assume that ASF20C and DPLE really do have significant skill at predicting decadal variations in the jet speed.

3.2 Signals driving decadal jet predictability are visible on interannual timescales

The similar behaviour of ASF20C and DPLE suggests that the extra information provided to ASF20C (the correct climate state every November 1st and the correct SSTs at all times) is not adding any extra jet speed skill beyond what is expected from an actual free-running, coupled decadal forecast. We interpret this as strong evidence for the assertion that the signals responsible for predictable decadal shifts in the jet speed are fully represented in ASF20C, and we will from now on assume that this is the case. This assumption has two important consequences.

Firstly, since each of the 109 ASF20C winter forecasts only knows about the initial conditions and boundary forcings of the season in question, the sources of predictable decadal jet speed forcing must be present in their entirety within a single winter season. Put differently, since the atmospheric variability in ASF20C is entirely generated by forecasts of a single season, any decadal variability ASF20C generates must arise due to atmospheric variability taking place within single seasons. In particular, the predictable atmospheric signals, and the mechanisms involved, do not seem to require multi-year lags as some studies have suggested they do (Peings and Magnusdottir, 2014b; Kwon et al., 2020). Secondly, since the ASF20C forecasts are uncoupled, the forcing being exerted on the jet does not essentially depend upon atmosphere-ocean coupling, in the sense that no within-season-feedbacks are necessary. To the extent that there is any interaction between the SSTs and the atmosphere in AS2F20C, it must purely be from the former to the latter. In the sections that follow we will make use of these two points repeatedly to simplify the analysis and reasoning.

We emphasise straight away that while surface coupling appears to not be required to reproduce the low frequency jet speed variability, missing or deficient coupling may play a role in generating the 'signal-to-noise paradox', as suggested in Scaife and Smith (2018).

4 Sources of predictability from sea surface temperatures

The midlatitude troposphere broadly speaking has a decorrelation timescale of about 2 weeks (Judt, 2020), and this is also true for the daily jet speed index (not shown). This strongly suggests that the skill in both ASF20C and DPLE cannot be explained by the persistence of atmospheric initial conditions, and is rather a result of forcing from some other slow timescale process. Due to the considerable body of work emphasising the importance of SSTs in generating forecast skill, we make the assumption here that decadal jet speed skill is a result of SST forcing. The validity of this assumption is discussed in Section 7.1.

4.1 Sub-polar North Atlantic SSTs as a common signal across observations and forecasts

In order to locate potential sources of skill in SSTs, we compute correlations of the winter jet speed against winter SSTs at every gridpoint. Because of our conclusion that these sources must be visible already on seasonal timescales (see Section 3.2), we do this using both the raw DJF timeseries as well as timeseries obtained by applying a 10-year running mean. Sources of

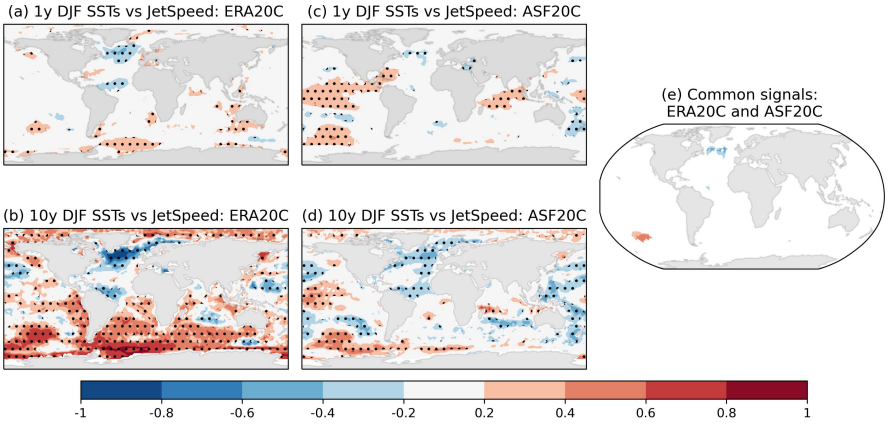

Figure 3. Correlations between the DJF JetSpeed timeseries and DJF SSTs at each gridpoint. In (a) for ERA20C using seasonal data; (b) for ERA20C with a 10-year running mean applied; (c) for ASF20C using seasonal data; (d) for the ASF20C ensemble mean with a 10-year running mean applied. In (e) are shown the correlations from (b) that are the same sign across all subplots and significant ($p < 0.05$) in each subplot by itself. Stipling indicates significance using a two-tailed t-test ($p < 0.05$). The period 1900-2010 is used.

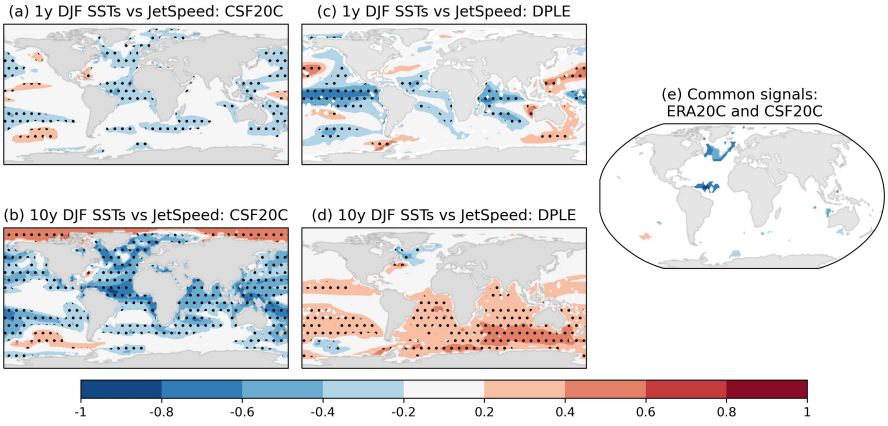

Figure 4. Correlations between the DJF JetSpeed timeseries and DJF SSTs at each gridpoint. In (a) for CSF20C using seasonal data; (b) for the CSF20C ensemble mean with a 10-year running mean applied; (c) for DPLE using seasonal data; (d) for the DPLE ensemble mean with a 10-year running mean applied. In (e) are shown the correlations from (b) that (i) are of the same sign in both (a) and (b) of this plot and subplots (a) and (b) of Figure 3, and (ii) significant ($p < 0.05$) in each of these four subplots by themselves. Stipling indicates significance using a two-tailed t-test ($p < 0.05$). The period 1900-2010 is used.

skill are then assumed to correspond to regions of significant correlations that are common to ERA20C, ASF20C, CSF20C and DPLE across both interannual and decadal timescales. While it is of course possible that the locations of signals differs

somewhat across the datasets (e.g. due to mean state biases in the forecast models), the simplest - and more physically sound - possibility is that the signals are identical across the datasets, so we consider this possibility first.

Figure 3 shows gridpoint correlations for ERA20C and ASF20C, at interannual (top row, (a) and (c)) and 10-year (bottom row, (b) and (d)) timescales. Similarly, Figure 4(a)-(d) shows interannual and 10-year gridpoint correlations but for CSF20C and DPLE. It is an instructive exercise to search for commonalities across these four subplots by eye, and doing so quickly highlights the SPNA. Already just comparing Figures 3(a) and (c) makes this clear. A comparison with CSF20C and DPLE corroborates this, and furthermore seems to rule out signals from both the tropical Pacific, tropical Atlantic and Indian ocean, since the correlations are opposite in sign between ASF20C/CSF20C and DPLE in these regions.

To justify this visual inspection further, we can filter the interannual-to-decadal correlations for common signals. In order to minimise false negative rates, i.e. accidentally discarding regions that are actually important, we deliberately use a relatively lenient filter here. Concretely, a gridpoint passes the filter if (i) the correlations at that gridpoint are of the same sign across the four subplots (a)-(d) of Figure 3, and (ii) statistically significant at $p = 5\%$ in each of the four subplots with respect to a two-tailed t-test. This is a lenient filter by virtue of the fact that the t-test does not take into account the long memory of SSTs and will therefore suggest significance far more often than reasonable. This leniency has the benefit that regions not passing the filter are unlikely to be important: stronger significance tests can then be applied to any regions that do pass the filter. The results are shown in Figure 3(e), where we have plotted the 10-year gridpoint correlations from Figure 3(d) that satisfy conditions (i) and (ii). This strongly suggests is that there are only two obvious signals common to ERA20C and ASF20C on interannual-to-decadal timescales: the SPNA, and an isolated patch in the south Pacific.

Similarly, Figure 4(e) shows the 10-year correlations of CSF20C common across ERA20C and CSF20C on both timescales. Note that the south Pacific patch has now nearly vanished. Comparing Figure 3(e) with Figure 4(e) therefore suggests that the SPNA is the only location where potentially robust correlations can be found on interannual-to-decadal timescales for ERA20C, ASF20C and CSF20C simultaneously. The same conclusion is drawn if one applies the same filter to search for common signals between ASF20C and CSF20C (not shown). Inspection of the signals in DPLE shows that the SPNA is also clearly highlighted on 10-year timescales (Figure 4(d)). On interannual timescales (Figure 4(c)), the signal from the SPNA is evidently weaker and mostly confined to the eastern side. However, because the sample size of ERA20C, ASF20C and CSF20C is almost twice that of DPLE, we consider it justified to weight these datasets higher than DPLE, and suggest that the weaker interannual DPLE signal is simply due to the reduced sample size.

To summarise the above discussion, the SPNA emerges as a common region of interannual-to-decadal correlations between SSTs and the jet speed across ERA20C, ASF20C, CSF20C and DPLE, and appears to be the only such common region. We interpret this as strong evidence for the importance of the SPNA in driving decadal jet predictability. Crucially, the coupled DPLE forecasts have skill at predicting the decadal SST variability in the SPNA region (Yeager et al., 2018; Yeager, 2020). We therefore define an SPNA timeseries as the DJF averaged SSTs across the domain 49-57N, 50-25W. This box was chosen because it encloses almost precisely the negative correlations highlighted in Figure 3(e). The remainder of the analysis we carry out is not sensitive to small shifts in the definition of this box. Other differences between the datasets seen in the two Figures, while interesting, do not seem relevant to the study at hand, and so are not discussed.

It is worth mentioning the phenomenon of false discovery rates, namely the fact that if one looks for correlations across sufficiently many predictors then some are bound to be significant by chance. This effect would a priori be expected to be high when considering correlations across thousands of gridpoints with high spatial autocorrelation. However, based on the results and discussion of previous sections, we are now assuming that (a) there is decadal forecast skill which needs to be explained and (b) that this skill comes from the SSTs. A rejection of any and all gridpoint correlations as 'false discoveries' would be directly counter to this assumption, and is therefore not done. The discovery of a region of correlations common across 4 data sets on the other hand can be seen as providing additional evidence towards our assumption.

4.2 Significance of the SPNA-JetSpeed link

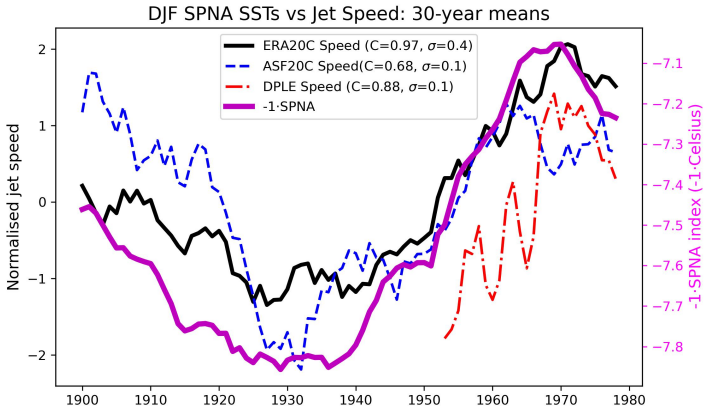

Figure 5. Timeseries of 30-year running means of ERA20C SPNA SSTs (thick purple: the sign has been flipped for visual convenience), ERA20C jet speeds (thick black), ASF20C ensemble mean jet speeds (dashed blue) and DPLE ensemble mean jet speeds (dashed red). The values of C in the legend indicate the correlation between the jet speed timeseries and the relevant SPNA timeseries, and the values of σ indicates the standard deviations of the jet speed timeseries. Note the different y-axes for the jet speed and SPNA timeseries. The jet speed indices have all been normalised to have mean 0 and standard deviation 1.

Table 2 summarises the correlations between SPNA SSTs and jet speed for the different datasets across different timescales. Figure 5 visualises the 30-year running mean timeseries: the sign of the SST index has been flipped for visual convenience. To generate the 5% significance threshold reported in the table, we assume a null hypothesis that the jet speed and SPNA are uncorrelated random variables. The interannual jet speed timeseries is modelled as a normal distribution with no memory across years, as before. To create synthetic SPNA timeseries, we apply the Fourier phase shuffle method to the interannual SPNA timeseries, in order to generate random draws that preserve the considerable autocorrelation. By correlating 10,000 such synthetic timeseries, along with those obtained by applying a 10 or 30-year smoothing, we can then numerically estimate the 5% significance threshold. Note that the confidence intervals for ERA20C, ASF20C and CSF20C were found to be almost

SPNA vs JetSpeed	Raw	10 year smoothing	30 year smoothing
ERA20C	**-0.45**	**-0.88**	**-0.97**
ASF20C	**-0.24**	-0.41	-0.68
CSF20C	**-0.27**	**-0.57**	-0.70
5% significance threshold	±0.19	±0.56	±0.84
DPLE	-0.13	-0.28	-0.88
5% significance threshold	±0.29	±0.81	±0.92

Table 2. Correlations between DJF averaged SPNA SSTs and DJF averaged jet speed at different levels of smoothing. The 5% significance thresholds shown using the null hypothesis described in the main text. Significant correlations are highlighted in bold.

identical, so we only report a single confidence interval encompassing all three, obtained by taking the average across the three individual intervals.

The correlation between SPNA and the jet speed is highly significant in ERA20C for all timescales considered here, suggesting a robust physical link between these quantities. Both ASF20C and CSF20C have significant correlations on the interannual timescale, and CSF20C also for 10-year timescales. The other correlations obtained do not pass the threshold. Of course, the lack of consistent statistical significance does not mean the physical link in ERA20C is not simulated by the forecast models, only that these correlations would not be sufficient to establish such a link when viewed in isolation. As emphasised in Shepherd (2021), it is crucial to include physical reasoning and prior knowledge when drawing conclusions about significance. We return to this, and the question of causality, in the Discussion section.

4.3 Decadal timescale forcing as the accumulation of seasonal timescale forcing

In Section 3.2, we argued that the presence of skill in ASF20C implies that predictable jet speed variability is a result of forcing from the boundary conditions taking place already within a single season. Having now argued that this forcing is coming from SPNA SSTs, it remains to test whether the forcing exerted by the SPNA on the jet taking place on seasonal timescales suffices to explain decadal timescale correlations seen in Table 2. That is, while the SPNA may be varying on decadal timescales, we want to now test whether the forcing it exerts on the atmosphere above can be understood as taking place on seasonal timescales without the need to invoke slower multi-annual-to-decadal timescale atmospheric processes.

To test this, we model the SPNA-JetSpeed system as follows. For a given dataset, we perform a linear regression of the interannual SPNA timeseries against the interannual JetSpeed timeseries to get

$$\text{JetSpeed} = a \cdot \text{SPNA} + b + \epsilon \tag{1}$$

for some constants a, b and a noise term ϵ which is normally distributed with mean 0. We then generate a random, synthetic timeseries of DJF SPNA using the Fourier phase shuffle method, as in the previous subsection. A corresponding JetSpeed timeseries is then obtained using the linear relationship just derived. We then take 10 and 30-year running means of these

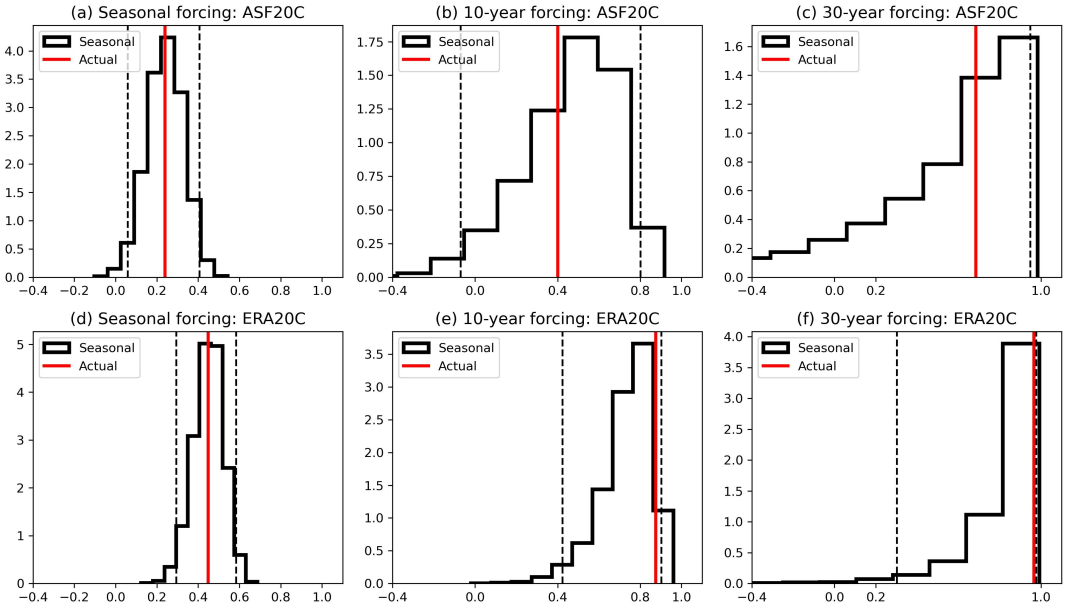

Figure 6. Histograms of the SPNA-JetSpeed correlations expected by chance from the seasonal timescale link, (a) and (d); taking 10-year averages of the seasonal link, (b) and (e); taking 30-year averages of the seasonal link, (c) and (f). See main text for details. The top row uses data fitted to ASF20C and the bottom to ERA20C. In each subplot the red line indicates the actually observed correlation in that dataset, and the dashed lines indicate the 95% confidence interval. The histograms are normalised so that the area underneath is 1.

synthetic DJF timeseries and compute the correlations between them. By repeating this procedure 10,000 times we can assess what decadal timescale correlations are expected from taking running means of the interannual SPNA-JetSpeed link.

The result of this for ASF20C is summarised in Figures 6(a), (b) and (c). The synthetic seasonal timescale correlations (Figure 6(a)) capture the observed seasonal correlation by construction. After taking running means, the distribution of synthetic correlations becomes notably skewed. For both 10 and 30-year running means (Figure 6(b) and (c)), the observed correlation is still comfortably within the 95% confidence interval of the synthetic correlations. The same analysis for ERA20C is shown in Figures 6(d), (e) and (f). The observed 10 and 30-year correlations are closer to the upper threshold here, but still fall within. This implies that the decadal timescale correlations of ASF20C and ERA20C can be completely explained by the interannual SPNA-JetSpeed link, as hypothesised.

We emphasise that we are not suggesting the SPNA variability is purely interannual, but that the connection between the SPNA and the atmosphere occurs within the span of a single season. The same conclusion can be seen to hold for CSF20C by comparing Table 2 with Figure 6; we also verified the analysis using the shorter DPLE (not shown).

5 Physical pathways and the signal-to-noise paradox

5.1 What affects the speed of the jet in climate models?

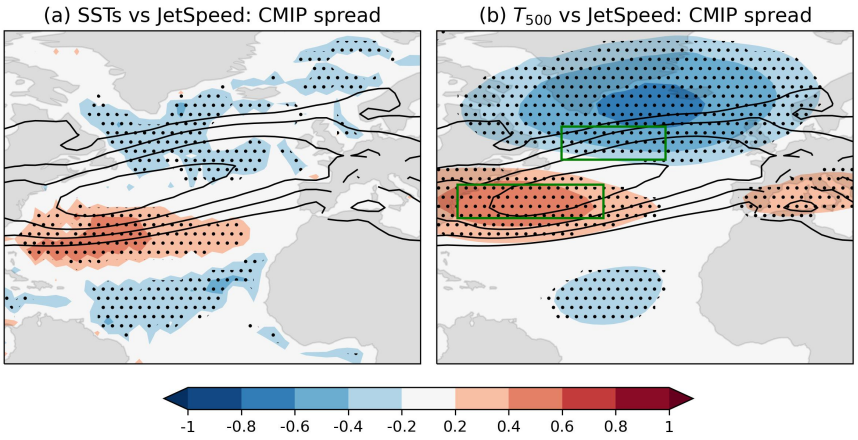

Figure 7. Filled contours: correlations between the spread in CMIP models' climatological DJF jet speeds and (a) the model spread in DJF SSTs at every gridpoint; (b) the model spread in DJF 500hPa air temperatures at every gridpoint. Line contours: the multimodel mean climatological zonal winds at 850hPa (contours drawn at 4, 6, 8 and 10 m/s). The climatologies are computed over the period 1980-2015. See main text for details of the models used. In (b), green boxes highlight the SPNA region and the jet-core region used in the computation of U_{jet} in Section 5.2. Stipling highlights statistical significance ($p = 5\%$) with respect to a two-tailed t-test.

In order to shed light on the pathways involved, it is helpful to first understand the related question of what affects the climatological speed of the eddy-driven jet in climate models. To this end, we show in Figure 7 correlations between the multimodel climatological DJF jet speeds and climatological DJF temperatures (SSTs and air temperatures at 500hPa) at multiple gridpoints for 76 coupled CMIP5 and CMIP6 climate models (see Section 2). In other words, for a given gridpoint x and model n, let $T_{x,n}$ be the climatological temperatures (at sea-surface or 500hPa) of model n at gridpoint x, and let $JetSpeed_n$ be the climatological jet speed for model n. Then the correlations plotted at gridpoint x is the correlation between the variables $T_{x,n}$ and $JetSpeed_n$ indexed by n. Thus a positive correlation at a gridpoint means that a model with warmer temperatures at that gridpoint would be expected to have a faster jet. The period used to compute the climatologies is 1980-2015. Stipling indicates significance ($p = 5\%$) with respect to a two-tailed t-test, where we simply assume the model means are all independent of each other and distributed normally.

Figure 7(a) shows the typical NAO-related tripole pattern in the North Atlantic, with the northern pole corresponding to the sub-polar North Atlantic and the southern pole corresponding to the tropical Atlantic. This is consistent with both a driving of the jet by the SSTs (Deser et al., 2007; Baker et al., 2019) and vice versa (Seager et al., 2000; Deser et al., 2010). Figure 7(b) shows that in the mid troposphere the dominant signal is the meridional temperature gradient between the jet core and the poleward flank of the jet. By computing an index based on the difference in 500hPa DJF temperatures across the two peaks in

Figure 7(b), we found that this meridional gradient accounts for more than 60% of the intermodel spread (correlation $C \approx 0.8$; not shown). Notably, at 500hPa the atmosphere above the tropical Atlantic has a relatively weak association with the model spread.

Changes to meridional temperature gradients in the midlatitudes necessarily alter the eddy-driven jet, both directly and immediately via thermal wind balance and more indirectly by changing the baroclinicity. We interpret Figure 7(b) as saying that the meridional gradient most relevant for the speed of the eddy-driven jet in climate models is the local one between the jet core and the atmosphere situated directly above the SPNA[2]. This offers independent evidence for the potential importance of the SPNA region, and will be made use of in the following section.

5.2 The tropospheric pathway from sub-polar North Atlantic SSTs to the jet speed

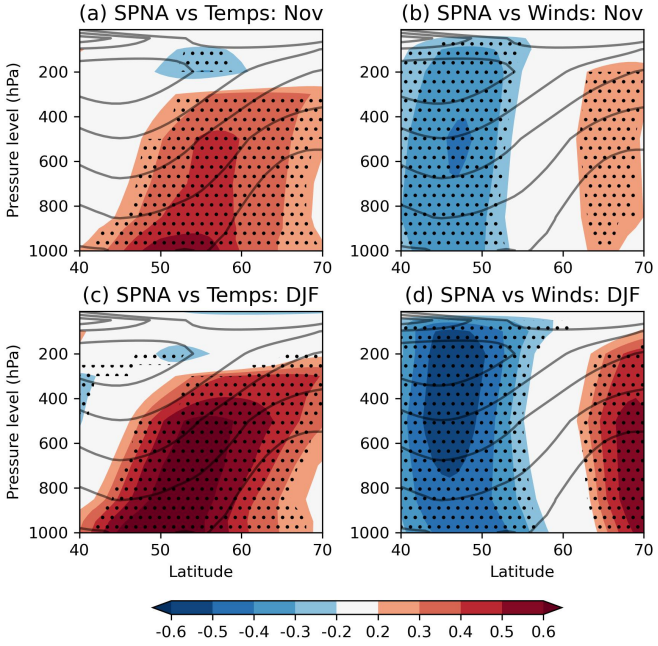

Figure 8. Correlations in ERA20C between SPNA SSTs and (left column) zonally averaged air temperatures at different pressure levels; (right column) zonally averaged zonal winds at different pressure levels. November averages are used in (a) and (b), while DJF averages are used in (c) and (d). The period 1900-2010 is used. The climatological zonal winds are shown in grey contours. Stipling indicates significance ($p = 5\%$): see main text for details.

[2]By contrast, in the southern hemisphere, Ceppi and Hartmann (2013) showed that the width of the Hadley Cell could explain intermodel spread in the southern hemisphere eddy-driven jet speed, suggesting a clear role for the tropics. They further reported that no such relationship could be found in the northern hemisphere, consistent with the lack of a strong tropical signal in Figure 7(b).

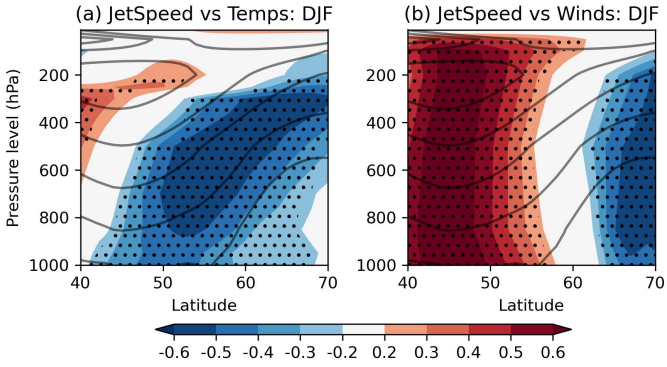

Figure 9. Correlations in ERA20C between jet speeds and (a) zonally averaged air temperatures at different pressure levels; (b) zonally averaged zonal winds at different pressure levels. DJF averages are used in both. The period 1900-2010 is used. The climatological zonal winds are shown in grey contours. Stipling indicates significance ($p = 5\%$) with respect to a two-tailed t-test.

Returning now to the question of how SPNA SSTs can modulate the jet speed, the results of the previous section suggest a simple explanation in keeping with the analysis of Woollings et al. (2015): anomalously cold SSTs in the SPNA cools the atmosphere aloft, thereby strengthening the local meridional temperature gradient around the jet, causing an intensification of the eddies and thereby an increased jet speed. Similarly for warm anomalies. To strengthen the case for this pathway, Figure 8 shows the vertical extent of the anomalies associated with SPNA SSTs in ERA20C, by correlating November SPNA SSTs with zonally averaged (a) temperatures and (b) zonal winds at different pressure levels: the averaging is done over longitudes 60W-20E. Significance at 5% (indicated by stipling) uses a null hypothesis which models the SPNA using the Fourier phase shuffle method and temperatures/winds as white noise. Figures (c) and (d) show the same but using DJF means. Figures 8(a) and (c) suggest that the heating does not remain confined to the surface, but extends up to around 300hPa, making the robust impact on the jet seen in Figure 8(b) plausible. For completeness, the reverse link from the jet speed to temperatures and winds is shown in Figure 9. Figure 9(a) shows that the heating pattern associated with the jet is maximal in the mid-troposphere. The heating associated with the SPNA (Figure 8(a) and (c)) projects well onto the jet pattern, but is maximal at the surface. Comparing Figures 8(d) and 9(b) confirms that the association between the SPNA and the zonal winds projects onto the jet speed signature.

Detailed analysis of the tropospheric response to an SST anomaly in the North Atlantic, such as in Deser et al. (2007), show that the response generally proceeds in two stages. To begin with, the induced anomaly is baroclinic in nature and localised to the heat source. In the second stage the baroclinic anomalies are removed by eddy heat and momentum transport, leading to a relatively barotropic structure with a more hemispheric scope. Figure 8 shows that in DJF, the structures are relatively barotropic in nature, consistent with this tropospheric pathway having taken place. Note that this approximately barotropic structure is also visible when plotting regression coefficients instead of correlations (not shown).

To further test the importance of the tropospheric heating anomalies over the SPNA, we will compute a rough estimate of what ERA20C jet speed anomalies are expected from SPNA variability using geostrophic wind balance. We define a region

405 labeled 'South' by 35-43N, 75-40W, and a region labeled 'North' using the same domain as the SPNA. Both regions are highlighted in Figure 7(b) with green boxes. We then assume a constant layer-thickness between 1000 and 300hPa, where 300hPa is the upper limit of the heating associated with the SPNA, as seen in Figure 8(c). Let $\hat{T}_N$ and $\hat{T}_S$ denote the layer-averaged DJF air temperature between 1000 and 300hPa in the North and South regions. By regressing our SPNA SST timeseries against $\hat{T}_N$, we obtain the SPNA-driven component which we denote by $\hat{T}_N(SPNA)$. We let $[\hat{T}_S]$ denote the average of $\hat{T}_S$ across
410 all years 1900-2010. The timeseries of the difference $\hat{T}_N(SPNA) - [\hat{T}_S]$ thereby roughly measures the layer-averaged meridional temperature gradient across the jet core purely associated with SPNA SST variability. The discussion of the previous section suggests this gradient should be highly relevant for determining the jet speed. Geostrophic balance with a constant layer-thickness relates this gradient to the zonal winds U_{jet} in the jet core according to

$$U_{jet} = -\frac{R}{f} \cdot \log(1000/300) \cdot \frac{1}{dy}(\hat{T}_N(SPNA) - [\hat{T}_S]), \tag{2}$$

415 where $R = 287.05$ is the gas constant, $f = 1 \times 10^{-4}$ is the approximate midlatitude Coriolis, and $dy = 2.553 \times 10^6$ metres. Here we have used that 1 degree latitude is roughly 111km. The U_{jet} timeseries is compared to the jet speed timeseries for ERA20C in Figure 10 using 30-year averages. They not only correlate extremely well ($C = 0.97$), but also share virtually identical magnitudes. In other words, the decadal jet speed variability of ERA20C can be fully accounted for by tropospheric temperature anomalies associated with the SPNA. It is worth emphasising that the key simplifying assumption made to compute
420 U_{jet} (that the only temperature gradient variability which matters is that associated with SPNA SSTs) is not part of standard geostrophic balance, and it is therefore not obvious, a priori, that such good agreement should be obtained.

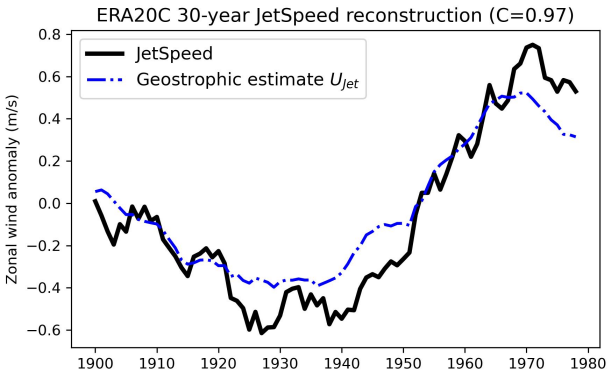

Figure 10. Black curve: 30-year averaged ERA20C DJF jet speed. Blue dashed curve: 30-year averages of the ERA20C U_{jet} timeseries, which roughly estimates the zonal wind variability associated with SPNA-induced changes to the meridional temperature gradient assuming geostrophic balance (see main text for details). The value C denotes the correlation coefficient between the two timeseries.

5.3 The signal-to-noise paradox and biases in heat transfer

As discussed in the Introduction, there is a 'signal-to-noise paradox' in the forecast models, which suggests that the models are responding too weakly to sources of predictable forcing. Since we have argued that the predictable forcing originates in the SPNA, it is natural to consider why the response to SST anomalies in this region may be too weak in the forecast models. As previously noted, the response of the atmosphere to an imposed SST anomaly proceeds in two stages, with the first stage dominated by the immediate localised response, and the second stage dominated by the eddies and eddy-meanflow interactions. There is already evidence to suggest that the second stage contributes to the 'paradox', due to a deficient eddy feedback in the forecast models (Hardiman et al., 2022). Here we would like to raise the possibility that biases in the first stage may also be contributing.

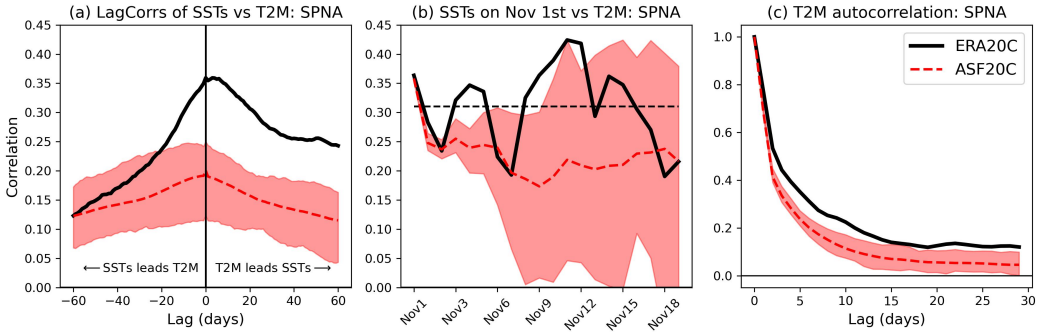

Figure 11. In (a): daily lag correlations between area-averaged SPNA SSTs and area-averaged SPNA two-meter temperatures (T2M). In (b): correlations between area-averaged SPNA SSTs on November 1st and between area-averaged SPNA T2M on subsequent days. In (c): daily autocorrelation of area-averaged SPNA T2M. In each case the thick black line is ERA20C, the dashed red line is the ASF20C ensemble mean, and the red shading shows the full range of the ASF20C ensemble spread. Only days in DJF are used, covering the period 1900-2010.

The fast atmospheric response to an SST anomaly is a result of forcing exerted by heatflux anomalies. There are several issues involved in comparing the heatflux variability across the forecast data we have considered here, which comprise a mix of coupled and uncoupled models whose heatflux variability will necessarily differ (Barsugli and Battisti, 1998). Furthermore, while ERA20C is created using daily forecasts of an uncoupled model, and thus at face value similar to ASF20C, the heatfluxes in ERA20C actually behave much more like those of a coupled model, like CSF20C or DPLE, and not like those in ASF20C (not shown). This is likely due to the constraints placed by the data assimilation, as well as the fact that the effects of air-sea coupling are small on daily timescales. Such subtleties would be important to account for carefully, since they may have a big impact on the interpretation of any analysis performed (O'Reilly et al., 2023). Finally, it is in general not straightforward to isolate the component of the heatflux variability associated with forcing by SSTs, as opposed to atmospheric forcing.

In an attempt to sidestep these complications, we will assess potential biases in the immediate localised response by analysing the link between daily SSTs and 2-metre air temperature (T2M). In Figure 11(a) we show the lagged correlations between these quantities averaged across the SPNA region for ERA20C and ASF20C. A bias in these correlations would be expected

in ASF20C when T2M leads the SSTs, due to the lack of coupling. However, a substantial bias is also evident when the SSTs lead T2M, even (and especially) at lag 0, with none of the 51 ensemble members able to replicate the behaviour of ERA20C. This implies that, in ASF20C, there is a substantial underestimation of how tightly linked SSTs and near-surface temperatures are.

The lag correlations in Figure 11(a) are computed using all days in DJF, so this underestimation could in principle be taking place exclusively in late winter. However, Figure 11(b) shows that the same bias is visible even when only correlating SSTs on November 1st (the initialisation day) with the T2M on subsequent days. We conclude that the daily SST-T2M link is likely too weak in ASF20C at all times. In particular, this suggests that the weak link is unlikely to be purely a result of deficiencies in the atmospheric adjustments due to the eddies, as these would be expected to manifest more clearly only after some time has passed. Instead, the weak link seems more obviously consistent with biases in the immediate surface temperature response. Such an inadequate surface temperature response from the slowly varying SSTs would also explain the significantly reduced T2M autocorrelation in ASF20C, shown in Figure 11(c).

We have argued that SPNA SSTs force the jet by altering the meridional temperature gradient around the jet core, not just at the surface but at depth (Figure 8(a)). In ASF20C the atmospheric temperature response to SSTs is already substantially weaker than observed just 2 metres above sea-level. We therefore cautiously propose that the 'signal-to-noise paradox' in decadal NAO forecasts may be related at least partially to deficiencies in the direct transfer of heat from the ocean to the atmosphere in models. Note that biases in extratropical surface coupling have previously been suggested as contributing to the 'paradox' (Scaife and Smith, 2018; Zhang et al., 2021). It is not clear if the bias we report here is due to the lack of coupling in ASF20C or due to a more fundamental problem with the one-way communication from the ocean to the atmosphere, though the fact that CSF20C and DPLE suffer just as badly from the 'paradox' as ASF20C is suggestive of the latter. We discuss this point more in Section 7.5.

6 Drivers of decadal SST variability and their link to the jet speed

We have highlighted SSTs in the SPNA as the main source of decadal predictability of the jet speed. Understanding what is driving the variability of these SSTs and their relation to jet speed variability is thus important to understand how reliable forecast skill is likely to be in the future. A comprehensive examination of this using the jet speed framework goes beyond the scope of the paper. Here we limit ourselves to only considering two key potential drivers: anthropogenic sulphate aerosols and the AMOC.

6.1 The role of sulphate aerosols

The question of how aerosols affect North Atlantic SSTs is still actively debated, and could include the role of both direct radiative adjustments local to the region, as well as more indirect impacts from upstream emissions (Booth et al., 2012; Robson et al., 2022). Here we will only consider the hypothesis put forth by Robson et al. (2022), due to it being based on a large multimodel ensemble, unlike most other studies. In their study, they argued that sulphate aerosol emissions over the eastern

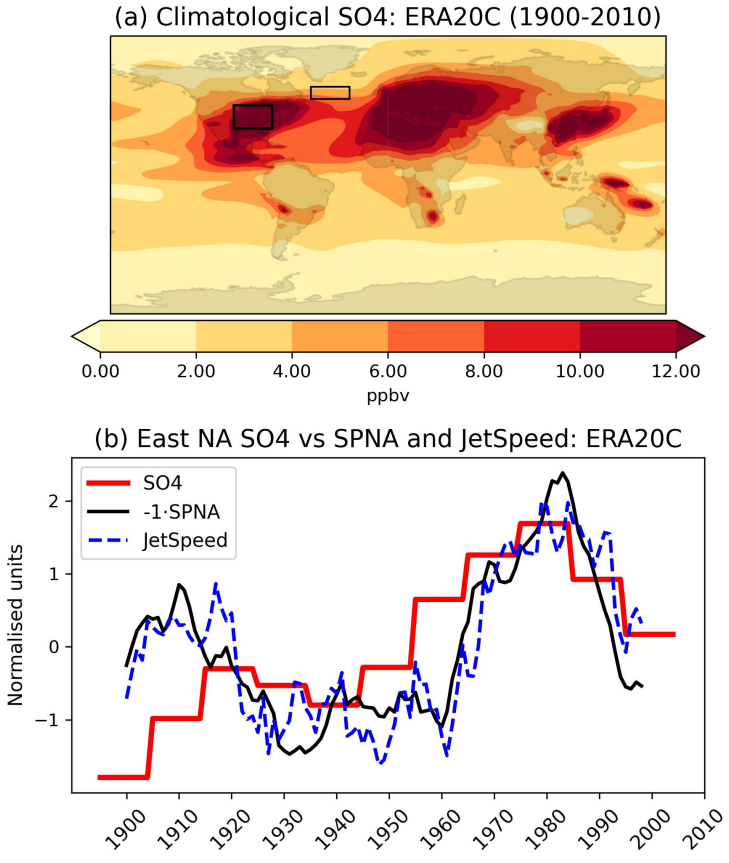

Figure 12. In (a): the climatological SO4 emissions of ERA20C (in parts per billion by volume), which is the same as those in ASF20C and CSF20C. In (b), timeseries of the 10-year averaged ERA20C SPNA SSTs, jet speed and area-averaged East North American SO4. The sign of the SPNA timeseries has been flipped for visual convenience. In (a), the domain of the East North American region has been marked with a thick black box and the SPNA domain with a thin black box.

part of North America led to a cooling of surface temperatures there, and that this cool air was then advected over the North Atlantic, thereby cooling the SSTs and ultimately modulating the AMOC. The possible direct impacts of emissions over the SPNA region itself will not be considered, since the emissions are generally very low in this region.

ERA20C, ASF20C and CSF20C all share the same prescribed aerosol emissions, namely those used in CMIP5 (Lamarque et al., 2010). The climatological SO4 emissions are shown in Figure 12(a), showing a clear concentration of emissions over the eastern North American region. To study the impact of these emissions on the SPNA and jet speed, we define an SO4 timeseries by averaging the emissions over the box 30-45N, 260-285E. This box is visualised in Figure 12(a) and approximately covers the eastern North American continent. This timeseries is compared to the 10-year averaged SPNA SSTs and ERA20C jet speed in Figures 12(b) and (c). The SO4 timeseries matches both the SSTs and jet speed in the period 1940-2010, with a correlation

of around 0.8. However, in the earlier period 1900-1940, the SO4 emissions are anticorrelated with the SSTs and jet speed, with a correlation of around -0.3. Note that the CMIP5 aerosol forcing data is constant across each 10-year period, explaining the step-like quality of the SO4 timeseries.

We conclude that while aerosol emissions over North America may have played a role in driving decadal SPNA variability, they are unlikely to explain the entire 20th century variability, especially in the early period when emissions are still very low. Similar conclusions were drawn in earlier work by Zhang et al. (2013). Note that the forecast models are able to capture the strong jet in this early period, despite the low aerosol emissions (Figure 1). If the aerosols are driving some of the SPNA variability, then this would be expected to lead to predictable shifts in the jet speed, as discussed in the present paper. It should also be noted that an aerosol-induced cooling of North America would strengthen the land-sea contrast, which would also be expected to drive a stronger jet (Portal et al., 2022). It is unclear if such a change in land-sea contrast alone can generate decadal timescale forecast skill.

6.2 The role of the AMOC

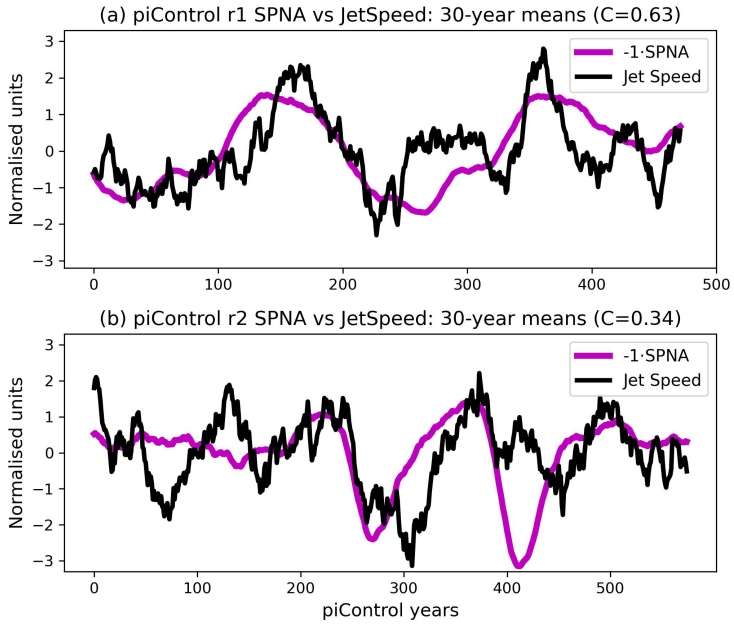

Figure 13. Timeseries of 30-year averages of the DJF SPNA SSTs (purple) and DJF jet speeds (black) in (a) the first EC-Earth3 pre-industrial control ensemble member, and (b) the same for the second ensemble member. The sign of the SPNA timeseries has been flipped for visual convenience. The value C in the titles indicates the correlation between the two timeseries. The timeseries have been normalised to have mean 0 and standard deviation 1.

To understand if the AMOC alone can drive SPNA-induced jet speed variability, we examine the two EC-Earth3 pre-industrial control simulations, which we refer to as piControl r1 and piControl r2. These simulations are particularly useful for

assessing AMOC forcing, since they are known to exhibit large and striking centennial timescale AMOC oscillations (Meccia et al., 2022). The lack of any anthropogenic emissions means the associated variability in the SPNA will be entirely due to the AMOC.

Figure 13 shows 30-year running DJF means of SPNA SSTs vs JetSpeed for the two ensemble members. The centennial AMOC oscillations are clearly visible in the SPNA timeseries. The correlations obtained between the SPNA and JetSpeed are -0.63 for the r1 member and -0.34 for the r2 member. A 5% significance threshold using the same null hypothesis as before (modelling JetSpeed as white noise, SPNA with the Fourier phase shuffle method) is approximately ± 0.5, implying that the r1 correlation is highly significant while the r2 is not.

We cannot directly compare the correlations from these free-running coupled simulations to the ensemble mean correlations of ASF20C. However, we note that if we concatenate back-to-back 5 ASF20C ensemble members, we get a timeseries of length $5 \cdot 109 = 545$, approximately the same length as the piControl members. By randomly drawing 5 ASF20C members, concatenating them, and computing the correlations between the resulting SPNA and JetSpeed timeseries, we can compare ASF20C and piControl in a like-for-like manner. If we carry out 1000 random such draws, we find that the expected ASF20C correlation is around -0.2 with a 95% confidence interval of approximately [-0.51, 0.18]. Almost identical results are obtained if using the CSF20C ensemble. This clearly suggests that the r2 correlation of -0.34 is perfectly consistent with the SPNA-JetSpeed link diagnosed in ASF20C/CSF20C. It also suggests that the r1 correlation of -0.63 is larger than expected, though not impossible, with a handful of concatenated ASF20C/CSF20C members producing correlations of around -0.6. Since the AMOC oscillations are particularly strong in the piControl runs, it is possible that this translates into SPNA-JetSpeed correlations which may be unrealistically large. The r1 member may also just be a random outlier.

Based on the above analysis we conclude that AMOC oscillations alone can in principle produce SPNA variations associated with an SPNA-JetSpeed link of the same strength as in the forecast models, even in the absence of any aerosol forcing. It is therefore natural to speculate that the AMOC is responsible for the coherent SPNA and jet speed variability seen in the early 20th century (Figure 12), and likely contributes to the variability later on as well.

7 Discussion

7.1 Other potential sources of skill

Our key argument for the importance of the SPNA can be understood as a conditional one: if one assumes that the skill comes from SSTs, then Figures 3 and 4 suggest that the SPNA is the responsible region. How reasonable is this assumption? In other words, could the skill originate from somewhere other than the ocean?

It is possible in principle that the skill comes from the stratosphere (Omrani et al., 2014), which has a decorrelation timescale of several months. However, the initial conditions of ASF20C (CSF20C) come from ERA20C (CERA20C), which only assimilates surface variables. Because these do not strongly constrain the stratosphere, the stratospheric variability in ERA20C is highly biased. Indeed, O'Reilly et al. (2019) found that the amplitude of the quasi-biennial oscillation (QBO) is substantially weaker in ERA20C compared to reanalysis products that assimilate the atmosphere. They also reported a greatly reduced down-

ward propagation in ERA20C. O'Reilly et al. (2019) also considered the same ASF20C hindcasts and found that the seasonal timescale association between the QBO and the NAO was essentially zero, unlike in forecasts assimilating the stratosphere. We interpret this as evidence that there is limited scope for skill from stratospheric initial conditions in ASF20C. Another argument towards the stratosphere not driving the predictability was given in Simpson et al. (2018), which showed using reanalysis that the magnitude of stratospheric anomalies do not match those of the jet on decadal timescales. Note that while they discussed only March explicitly, their Figure B1 suggest similar conclusions would likely be drawn for DJF.

Besides the stratosphere, the remaining obvious candidates are ice and anthropogenic emissions such as carbon dioxide and aerosols. However, the impact of ice on decadal jet variability has been found to be small compared to SSTs (Peings and Magnusdottir, 2016; Ogawa et al., 2018; Liang et al., 2021). Furthermore, observations of ice are sparse and unevenly distributed prior to 1979, with obvious implications for the potential quality of the ice data used in the ASF20C boundary forcing files. As for anthropogenic emissions, the analysis in Simpson et al. (2018) showed that globally averaged emissions cannot explain the decadal variability of a closely related jet index. This leaves open the possible influence of local emissions. The most obvious way for such local emissions to alter the jet is *indirectly*, by altering the radiative forcing and hence surface temperatures. Such indirect forcings on the jet would manifest themselves as an apparent forcing from surface temperatures. While this could potentially include forcing from temperatures over land, the decorrelation timescale over land is much faster than for SSTs, making a decadal timescale forcing seem less likely. On the other hand, any indirect forcing via the SSTs would already be visible in Figures 3 and 4, which highlight the SPNA.

It is hard to exclude the possibility of a more direct impact of aerosols on the jet, e.g. due to an altering of the local radiative fluxes in the troposphere (as opposed to at the surface). Figure 12(a) shows that sulphate aerosol concentrations are small on the north side of the jet core, making a direct contribution from there seem less likely. There are more emissions centred around the jet core, so these could be playing a role. However, as pointed out in Section 6.1, aerosol emissions drop considerably prior to 1940 and become anticorrelated with the jet speed timeseries in the period 1900-1940, even as forecast skill remains high (Figure 12(b)). Such direct effects are therefore unlikely to explain all the observed skill.

While all these other potential sources of skill cannot be conclusively ruled out without targeted forecast experiments, we nevertheless conclude by process of elimination that there is strong evidence towards our assumption that the skill comes from the SSTs, and hence from the SPNA in particular.

7.2 The question of causality between ocean and atmosphere

There is an obvious question concerning the causality of the SPNA-JetSpeed link we have proposed, especially since SSTs in the SPNA are known to be particularly sensitive to forcing from the jet (Visbeck et al., 2003; Barrier et al., 2014; Ma et al., 2020). Are the correlations in Table 2 in fact just indicative of stochastic atmospheric forcing on the ocean? Here we discuss several lines of evidence which suggest that at least a considerable portion of the correlations are due to a genuine forcing from the ocean to the atmosphere.

The first key point is the apparent existence of skillful decadal forecasts. We have shown that the SPNA appears to be the only clear, common SST-based signal between reanalysis and the forecasts we have analysed. If the SPNA-JetSpeed correlations

565 are mostly or entirely reflecting a causal influence of the atmosphere on the ocean then one seems forced to conclude that the decadal skill does not originate from the SSTs. This is problematic, given the discussion of Section 7.1.

Secondly, we remind the reader that ASF20C is uncoupled and uses prescribed boundary forcing, so correlations on interannual timescales in ASF20C cannot arise from a causal influence of the atmosphere on the SSTs in the model. While this is not true for CSF20C and DPLE, which are coupled, the fact that the correlations they attain are entirely comparable to those 570 of ASF20C, and all these correlations are explainable from the interannual link alone (Section 4.3), suggests that forcing from the SSTs to the atmosphere constitutes a pragmatic common explanation for the correlations found in all three forecast models. An alternative explanation would be the existence of a different source of skill, not related to SSTs, which influences both the jet and the SSTs (possibly via the jet). Since the correct SSTs are prescribed to ASF20C, such a common driver could result in correlations of the sort seen in Table 2. However, as before, the hypothetical existence of such a non-SST related driver is 575 problematic and currently unfounded.

Figure 8 showed correlations between DJF SPNA and DJF temperatures/winds. If we rather correlate November SPNA anomalies against DJF temperatures/winds, we get a similar picture, albeit weaker in magnitude, especially for the winds: see Figure B2 in the SI. This suggests that the vertical heating is not purely coming from the adjustment of the jet itself, but also from processes induced by the SST anomalies. The fact that the heating associated with the SPNA peaks at the surface (Figure 580 8(a) and (c)), is also suggestive of SST driving. In general, replacing DJF SPNA SSTs with November averages leads to smaller but non-zero correlations of the same sign.

Another important reason for having a high prior for a causal influence from the SPNA are that several studies using idealised models and GCMs have demonstrated that SST anomalies can causally affect the jet (Palmer and Zhaobo, 1985; Deser et al., 2007; Hassanzadeh and Kuang, 2016; Baker et al., 2017, 2019). Of particular relevance is the recent work of Drews et al. 585 (2023), who carried out pacemaker experiments which suggest that SSTs in the SPNA have a positive impact on decadal forecast skill. We also highlight Baker et al. (2017), who used a dry model to show that imposed heating anomalies modulate the speed of the eddy-driven jet by altering the meridional gradient around the jet core, making the causal pathway we propose here a plausible explanation for the observed decadal forecast skill. An important caveat here is the increasingly recognized role played by moist processes in determining the jet variability, which are not represented in such a dry model (Willison et al., 590 2013; Papritz and Spengler, 2015; Schemm, 2023; Fuchs et al., 2023). Note however that the results of Baker et al. (2017) are generally consistent with those in Baker et al. (2019) using a full moist GCM.

Finally, Simpson et al. (2018) showed that the DPLE forecasts can skillfully predict decadal SPNA SST variability, but fails to predict shorter timescale North Atlantic winds of an appreciable magnitude, strongly suggesting that decadal SPNA variability is not purely driven by stochastic atmospheric forcing. This was further corroborated by Yeager (2020), which found 595 evidence that the SST skill has an abyssal origin. We add to this by showing that quantitative estimates suggest the stochastic atmospheric forcing is weak on decadal timescales, as follows. Let x_k denote the 10-year averaged jet speed timeseries for each of the k CSF20C ensemble members ($k = 1, \ldots, 50$). We can decompose x_k as $x_k = s + \epsilon_k$, where s is the ensemble mean and ϵ is the residual for each member. The ϵ_k's can then be interpreted as different instances of stochastic atmospheric variability, with s representing a forced signal from the initial conditions and boundary forcings (e.g. SSTs). We then subtract

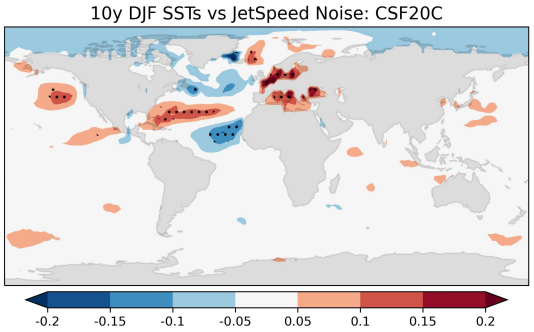

Figure 14. Correlations between the 'stochastic noise' component of the 10-year averaged DJF CSF20C jet speed and 10-year averaged DJF SSTs at each gridpoint. The period 1900-2010 is used. Stipling indicates statistical significance ($p = 5\%$) with respect to a Fourier phase shuffle test. Note the different colorscale compared to Figures 3 and 4.

the ensemble mean from each x_k and concatenate the resulting ϵ_k's back to back for all k. By correlating this timeseries with the CSF20C SSTs at each gridpoint (where ensemble members are again concatenated back to back in the same order), we get an estimate of the link between the SSTs and the atmospheric variability unrelated to initial or boundary conditions, shown in Figure 14. It can be seen that the stochastic components of the 10-year jet speed in CSF20C exhibit a tripole pattern consistent with atmospheric forcing associated with the NAO (Seager et al., 2000). The association in the SPNA region is weak and not significant almost everywhere, with correlations of around -0.06 (note the colorscale). It is unclear how such weak stochastic forcing could be responsible for the high correlations found in the ensemble mean. A different estimate of the stochastic forcing using the method of Ma et al. (2020) gave qualitatively similar results (not shown).

To conclude, we consider it likely that there is a sufficient causal influence of the SPNA on the jet to explain the observed decadal skill. Unambiguously confirming this would, however, require performing expensive forecast experiments, such as rerunning the full ASF20C hindcast using boundary forcing files for which the SPNA SST variability has been eliminated.

7.3 Relationship with Atlantic Multidecadal Variability

We have proposed that decadal timescale predictability of the NAO arises from a forced response to SSTs in the SPNA. It is natural to ask how this relates to previous studies suggesting a forced response from the AMV, with the NAO response typically lagging the AMV by several years (Zhang et al., 2019). We suggest that this apparent discrepancy is a result of two factors.

Firstly, while the AMV pattern on interannual timescales is the famous 'horseshoe' pattern, the pattern largely collapses to a signal confined to the SPNA after performing decadal timescale smoothing (Delworth et al., 2017; Simpson et al., 2018). This suggests that as far as decadal atmospheric variability is concerned, the sub-polar North Atlantic SSTs are more relevant than the 'horseshoe' pattern. The importance of this region has been highlighted previously (Gastineau and Frankignoul, 2015; Woollings et al., 2015; Ortega et al., 2017; Delworth et al., 2017; Wills et al., 2019).

Secondly, anomalies in the SPNA (the northern part of the 'horseshoe') are observed to lead anomalies in the tropical Atlantic (the southern part of the 'horseshoe') by around 2 years (Zhang, 2007). It is likely that this propagation of anomalies takes place both via subsurface ocean dynamics and coupling with the atmosphere (Zhang et al., 2019). This means that performing analysis based on the standard AMV 'horseshoe' pattern is likely mixing together several different processes happening on different timescales. Since the SPNA anomalies arise first, restricting attention to these might be expected to give a clearer picture and better highlight causal pathways from the ocean to the atmosphere. This argument has also recently been made by Wills et al. (2019). Note that this second point is presumably related to the first. The 'horseshoe' pattern includes the effects of the interannual timescale processes associated with the propagation of SST anomalies along the 'horseshoe': taking decadal timescale averages would be expected to remove these and highlight processes operating at much slower timescales, such as those associated with deep convection in the Labrador sea. This argument was also made in Delworth et al. (2017).

We therefore suggest that the apparent existence of an NAO response lagging the AMV by several years is, in fact, largely reflecting the existence of an essentially instantaneous atmospheric response to sub-polar North Atlantic SSTs, with the multi-year lag being an artefact of using an AMV index which averages across several different processes. A final point here is that the reverse impact of the NAO on the AMV has also been noted to be essentially lagged (Ma et al., 2020), additionally complicating analysis based on the AMV index.

7.4 Why is the jet latitude not predictable?

Several studies have argued that an AMV-driven NAO response involves a shift in both the strength and location of the storm track and/or baroclinic region (Msadek et al., 2011; Peings and Magnusdottir, 2014a; Frankignoul et al., 2015; Peings et al., 2016; Ortega et al., 2017). Such systematic latitudinal shifts in the storm track/baroclinicity, and hence the eddies, would be expected to manifest as predictable latitudinal shifts in the eddy-driven jet itself. However, as discussed in Section 3, there is no decadal predictability of the latitude of the jet in ASF20C, CSF20C or DPLE. This suggests that on decadal timescales, the eddies are randomly distributed around their mean position, and that the mechanism driving predictable jet variability does not essentially depend on latitudinal shifts in the storm track. Rather, such predictable changes appear to be more clearly related to the direct constraint of thermal wind balance along with an *intensification* of the storm track. This is fully consistent with Woollings et al. (2015), who highlighted the contrasting nature of forced shifts to the latitude and speed of the jet.

Of course, it cannot be ruled out that SST anomalies in the SPNA, or indeed elsewhere, force shifts in the latitude of the jet in the real world that are simply not captured by the forecast models we considered. However, until predictability has been established it does not seem possible to reject the null hypothesis that decadal variability in the jet latitude is chaotic and unpredictable. Indeed, the inconsistent latitudinal response in multimodel studies such as Ruggieri et al. (2021) lends concrete evidence for such inherent chaos. Further evidence for this is given in Figure B3 of the SI, which repeats the procedure of Section 4 to search for potential sources of jet latitude skill in ASF20C. This shows that there are no regions of significant SST-jet latitude correlations common to both ERA20C and ASF20C, consistent with the lack of jet latitude predictability.

It is also possible that biases in the climatological jet latitude (i.e., in the model mean state) of the forecast models are contributing to the apparent lack of jet latitude predictability. The sensitivity analysis of Baker et al. (2017) showed that the

response of the jet latitude to thermal forcing changes sign abruptly around the jet core, while the response of the jet speed is more uniform across the jet. The accurate response of a model jet to a given SST anomaly may therefore differ dramatically based on its climatological position. This could go some way to explaining the inconsistent jet latitude response in multimodel studies such as Ruggieri et al. (2021).

7.5 Some speculation concerning the 'signal-to-noise paradox' and turbulent heat transfer

In Section 5.3 we highlighted a bias in the daily timescale link between SSTs and T2M in ASF20C compared to ERA20C. As discussed there, the relationship of this bias with biases in the heatfluxes is non-trivial and requires its own investigation. However, we would like to briefly remark on a seemingly unexplored possibility, namely the potential impact of biases in the surface roughness lengths of forecast models. Indeed, the heatflux parameterisation in ASF20C uses the Charnock formula to compute surface roughness length (ECMWF, 2015). Greater roughness lengths are, broadly speaking, associated with increased friction and hence greater heat transfer, and GCM experiments have found that varying the roughness length in a localised ocean region can have substantial impacts on the fluxes and surface temperatures (Kirk-Davidoff and Keith, 2008). Over the ocean, roughness is strongly influenced by unresolved processes such as waves. In ASF20C these effects are represented by a globally fixed 'Charnock constant', which is undoubtedly inadequate, since the relationship between turbulent heat transfer and roughness length depends sensitively on the exact type of roughness and there is no general theory encapsulating all experiments: see e.g. the introduction to Wagner and Shishkina (2015). These biases would be expected to manifest immediately in the forecast, and it would be of interest to assess the potential role they play in driving the 'paradox'.

8 Conclusions

We briefly summarise the main results and arguments.

1. Decadal forecast skill of the winter NAO appears to be entirely due to the decadal predictability of the speed of the North Atlantic jet. There is no apparent predictability of decadal variations in the latitude of the jet, even in seasonal hindcasts with prescribed SSTs (Figures 1 and 2).

2. Initialised seasonal hindcasts can skillfully reproduce decadal variations in jet speed all the way back to 1900, and match the behaviour of a genuine decadal forecast in the period 1954-2010. This justifies using such seasonal hindcasts to diagnose decadal forecast signals using the full period 1900-2010, effectively doubling the available years compared to existing decadal forecasts. In addition, this strongly suggests that the decadal forecast signals are already fully present and visible on interannual timescales.

3. The only clear source of an interannual-to-decadal jet speed signal coming from SSTs, common to reanalysis and forecasts, is the SPNA region (Figures 3 and 4). SSTs in this region enjoy large and statistically significant correlations with the jet speed, which we argue are due at least in part to a causal link from the ocean to the atmosphere (Figure 5). We

also show that all of the decadal SPNA-JetSpeed link in the forecast models, along with the associated forecast skill, can be explained by a small but consistent seasonal timescale forcing from the SSTs (Figure 6).

4. The pathway from SPNA SSTs to the jet speed is argued to be tropospheric in nature: the surface heating anomaly extends relatively deeply into the troposphere (Figure 8) and is optimally situated to perturb the jet speed (Figures 7 and 10) both by direct adjustments consistent with thermal wind balance and subsequent reinforcements by the eddies.

5. The forecast models are found to significantly underestimate the link between SSTs and surface temperatures in the SPNA (Figure 11). This may lead to the forecast models underestimating the atmospheric response to SPNA anomalies, thereby contributing to the 'signal-to-noise paradox'.

6. Sulphate aerosol emissions in North America may explain part of the 20th century variability in the SPNA, but do not explain the early 20th century (Figure 12). AMOC oscillations alone are shown to be capable of inducing SST shifts of strength comparable to those connected with jet speed variability (Figure 13).

Note that the importance of the SPNA in driving predictable decadal jet variability was emphasised in Simpson et al. (2018) using a different metric of North Atlantic jet variability in March. Our work here is clearly closely related and was directly inspired by theirs.

There are several shortcomings to our analysis, chief of which is the fact that we have only studied two independent forecast models. However, Marcheggiani et al. (2023) have recently analysed the decadal forecasts used in Smith et al. (2019) and found that the jet speed is much more predictable than the jet latitude. They also linked the predictability to Noth Atlantic SSTs, thereby reinforcing two of our key conclusions using independent data. Our argument that the stratosphere is unlikely to be relevant due to the initialisation biases of the seasonal hindcasts may also be flawed, and it would be of interest to examine the role of the stratosphere more closely. We also interpreted the similar jet speed correlations in ASF20C/CSF20C and DPLE as evidence that the mechanisms involved are the same for these data sets. However, it may be that ASF20C/CSF20C obtain most of their 'skill' from the initialisation (lag 0), while DPLE obtains most of its skill from representing slower frequency coupled processes better (lags > 0). If so, conclusions drawn using ASF20C/CSF20C may not carry over to a genuine forecast context. Nevertheless, our findings have some important implications.

To begin with, we have reinforced the conclusion of earlier studies such as Baker et al. (2017) that the speed and latitude of the jet should be considered separately, and that the use of indices like the NAO (which amalgamate the two) may be misleading. Our finding that the decadally averaged jet speed is predictable while the latitude is not stands in amusing contrast to work on *seasonal* forecasts, which suggest the opposite picture of a predictable latitude and unpredictable speed (Parker et al., 2019). This reinforces the analysis of Woollings et al. (2015), who showed that the nature of forced jet variability differs depending on the timescale. On interannual timescales, the variability is dominated by the jet latitude, in turn associated with meridional shifts in the transient eddies and location of blocking. On decadal timescales, the variability is dominated by the jet speed, in turn associated with changes both to the *strength* of the eddy forcing and to the occurrence of transient Rossby wave breaking on both sides of the jet. Our analysis corroborates the speculation in Woollings et al. (2015) that this may result in

the nature and sources of predictability being different on the two timescales. In particular, our results are consistent with the following hypothetical view of the jet variability:

(a) On seasonal timescales the forcing from the SPNA is too small to be visible (resulting in the appearance of no jet speed skill), with jet latitude skill arising from accurately predicting meridional shifts in the eddies;

(b) On decadal timescales the eddies are varying essentially chaotically around their climatological position (resulting in no jet latitude predictability), but their intensity can vary predictably depending on the underlying SPNA SSTs, which strengthen or weaken the meridional temperature gradient.

A novelty of our work is the use of the initialised 20th century seasonal hindcast products ASF20C and CSF20C. While Parker et al. (2019) considered the fast and slow variability of the jet speed and latitude in ASF20C, the possibility of utilising these hindcasts to study the mechanisms underpinning decadal predictability seems to have been overlooked. The fact that ASF20C covers the entire 20th century, is reinitialised every year, and is uncoupled, are all highly attractive properties for simplifying analysis which we made extensive use of. We believe that ASF20C may be similarly beneficial for the study of decadal predictability in many other contexts.

Finally, we add to a growing body of literature suggesting that the SPNA is the main source of North Atlantic jet forcing on decadal timescales, and that the use of a larger AMV or AMO pattern might be counterproductive for questions of predictability by mixing together different mechanisms and timescales. We have shown that the SST variability in the SPNA driving predictable jet speed shifts can in principle be due to both the AMOC and aerosol emissions, and it seems likely that the observed 20th century variability is a result of both of these. We have also highlighted a model bias of an overly weak link between SSTs and surface temperatures, which may contribute the 'signal-to-noise paradox'. In future work we hope to examine this bias more carefully, and assess if the 'paradox' can be alleviated by improving the parameterisation of air-sea interactions in forecast models, e.g. through the use of stochastic physics and/or machine-learning methods.

Data availability. ASF20C and CSF20C data is freely available on CEDA (Weisheimer, 2020). ERA20C data is freely available via ECMWF at https://apps.ecmwf.int/datasets/data/era20c-daily/levtype=sfc/type=an/. DPLE data is available via the Earth System Grid Federation (Yeager, 2018). CMIP6 data is available for download via the Earth System Grid Federation.

Author contributions. KS led the writing of the manuscript and carried out the majority of the data analysis. PR and PD contributed data analysis related to idealised model results and the role of the AMOC in the pre-industrial CMIP6 simulations; aided with literature review; and helped interpret results. TW aided interpretation of results and provided expert guidance on physical mechanisms. IRS helped in the procurement and analysis of DPLE model data and interpretation of the results.

Competing interests. We declare that there are no competing interests.

Acknowledgements. KS was funded by a Thomas Philips and Jocelyn Keene Junior Research Fellowship at Jesus College, Oxford. PR acknowledges the use of computational resources from the parallel computing cluster of the Open Physics Hub (https://site.unibo.it/openphysicshub/en) at the Department of Physics and Astronomy of the University of Bologna. IRS was supported by the National Center for Atmospheric Research which is a major facility sponsored by the National Science Foundation under the Cooperative Agreement 1852977.

References

Athanasiadis, P. J., Yeager, S., Kwon, Y.-O., Bellucci, A., Smith, D. W., and Tibaldi, S.: Decadal predictability of North Atlantic blocking and the NAO, NPJ Climate and Atmospheric Science, 3, 1–10, 2020.

Baker, H. S., Woollings, T., and Mbengue, C.: Eddy-driven jet sensitivity to diabatic heating in an idealized GCM, Journal of Climate, 30, 6413–6431, 2017.

Baker, H. S., Woollings, T., Forest, C. E., and Allen, M. R.: The linear sensitivity of the North Atlantic Oscillation and eddy-driven jet to SSTs, Journal of Climate, 32, 6491–6511, 2019.

Barrier, N., Cassou, C., Deshayes, J., and Treguier, A.-M.: Response of North Atlantic Ocean circulation to atmospheric weather regimes, Journal of Physical Oceanography, 44, 179–201, 2014.

Barsugli, J. J. and Battisti, D. S.: The basic effects of atmosphere–ocean thermal coupling on midlatitude variability, Journal of the Atmospheric Sciences, 55, 477–493, 1998.

Bellomo, K., Murphy, L. N., Cane, M. A., Clement, A. C., and Polvani, L. M.: Historical forcings as main drivers of the Atlantic multidecadal variability in the CESM large ensemble, Climate Dynamics, 50, 3687–3698, 2018.

Bjerknes, J.: Atlantic air-sea interaction, in: Advances in geophysics, vol. 10, pp. 1–82, Elsevier, 1964.

Booth, B. B., Dunstone, N. J., Halloran, P. R., Andrews, T., and Bellouin, N.: Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability, Nature, 484, 228–232, 2012.

Bracegirdle, T. J.: Early-to-Late Winter 20th Century North Atlantic Multidecadal Atmospheric Variability in Observations, CMIP5 and CMIP6, Geophysical Research Letters, 49, e2022GL098 212, 2022.

Ceppi, P. and Hartmann, D. L.: On the Speed of the Eddy-Driven Jet and the Width of the Hadley Cell in the Southern Hemisphere, Journal of Climate, 26, 3450 – 3465, https://doi.org/10.1175/JCLI-D-12-00414.1, 2013.

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., Masina, S., Scoccimarro, E., Materia, S., Bellucci, A., and Navarra, A.: Global Mean Climate and Main Patterns of Variability in the CMCC-CM2 Coupled Model, Journal of Advances in Modeling Earth Systems, 11, 185–209, https://doi.org/10.1029/2018MS001369, 2019.

Clement, A., Bellomo, K., Murphy, L. N., Cane, M. A., Mauritsen, T., Rädel, G., and Stevens, B.: The Atlantic Multidecadal Oscillation without a role for ocean circulation, Science, 350, 320–324, 2015.

Danabasoglu, G., Bates, S. C., Briegleb, B. P., Jayne, S. R., Jochum, M., Large, W. G., Peacock, S., and Yeager, S. G.: The CCSM4 ocean component, Journal of Climate, 25, 1361–1389, 2012.

Davini, P., von Hardenberg, J., and Corti, S.: Tropical origin for the impacts of the Atlantic multidecadal variability on the Euro-Atlantic climate, Environmental Research Letters, 10, 094 010, 2015.

Dell'Aquila, A., Corti, S., Weisheimer, A., Hersbach, H., Peubey, C., Poli, P., Berrisford, P., Dee, D., and Simmons, A.: Benchmarking Northern Hemisphere midlatitude atmospheric synoptic variability in centennial reanalysis and numerical simulations, Geophysical Research Letters, 43, 5442–5449, 2016.

Delworth, T., Manabe, S., and Stouffer, R. J.: Interdecadal variations of the thermohaline circulation in a coupled ocean-atmosphere model, Journal of Climate, 6, 1993–2011, 1993.

Delworth, T. L., Zeng, F., Zhang, L., Zhang, R., Vecchi, G. A., and Yang, X.: The Central Role of Ocean Dynamics in Connecting the North Atlantic Oscillation to the Extratropical Component of the Atlantic Multidecadal Oscillation, Journal of Climate, 30, 3789 – 3805, https://doi.org/10.1175/JCLI-D-16-0358.1, 2017.

Deser, C. and Phillips, A. S.: Defining the internal component of Atlantic multidecadal variability in a changing climate, Geophysical Research Letters, 48, e2021GL095 023, 2021.

Deser, C., Tomas, R. A., and Peng, S.: The Transient Atmospheric Circulation Response to North Atlantic SST and Sea Ice Anomalies, Journal of Climate, 20, 4751 – 4767, https://doi.org/10.1175/JCLI4278.1, 2007.

Deser, C., Alexander, M. A., Xie, S.-P., Phillips, A. S., et al.: Sea surface temperature variability: Patterns and mechanisms, Annu. Rev. Mar. Sci, 2, 115–143, 2010.

Dorrington, J., Strommen, K., and Fabiano, F.: Quantifying climate model representation of the wintertime Euro-Atlantic circulation using geopotential-jet regimes, Weather and Climate Dynamics, 3, 505–533, https://doi.org/10.5194/wcd-3-505-2022, 2022.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., Bernardello, R., Boussetta, S., Caron, L.-P., Carver, G., et al.: The EC-Earth3 earth system model for the coupled model intercomparison project 6, Geoscientific Model Development, 15, 2973–3020, 2022.

Drews, A., Schmith, T., Yang, S., Olsen, S., Tian, T., Devilliers, M., Wang, Y., and Keenlyside, N.: Role of the subpolar North Atlantic region in skillful climate predictions for high northern latitudes: A pacemaker experiment, EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-13375, https://doi.org/https://doi.org/10.5194/egusphere-egu23-13375, 2023.

Ebisuzaki, W.: A method to estimate the statistical significance of a correlation when the data are serially correlated, Journal of climate, 10, 2147–2153, 1997.

ECMWF: IFS Documentation CY41R1 - Part IV: Physical Processes, no. 4 in IFS Documentation, ECMWF, https://doi.org/10.21957/p50qmwprw, 2015.

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, https://doi.org/10.5194/GMD-9-1937-2016, 2016.

Fichefet, T. and Maqueda, M. M.: Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics, Journal of Geophysical Research: Oceans, 102, 12 609–12 646, 1997.

Frankignoul, C., Gastineau, G., and Kwon, Y.-O.: Wintertime Atmospheric Response to North Atlantic Ocean Circulation Variability in a Climate Model, Journal of Climate, 28, 7659 – 7677, https://doi.org/10.1175/JCLI-D-15-0007.1, 2015.

Fuchs, D., Sherwood, S. C., Waugh, D., Dixit, V., England, M. H., Hwong, Y.-L., and Geoffroy, O.: Midlatitude jet position spread linked to atmospheric convective types, Journal of Climate, 36, 1247–1265, 2023.

Gastineau, G. and Frankignoul, C.: Influence of the North Atlantic SST Variability on the Atmospheric Circulation during the Twentieth Century, Journal of Climate, 28, 1396 – 1416, https://doi.org/10.1175/JCLI-D-14-00424.1, 2015.

Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J. H., von Storch, J.-S., Brüggemann, N., Haak, H., and Stössel, A.: Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP), Geoscientific Model Development, 12, 3241–3281, https://doi.org/10.5194/gmd-12-3241-2019, 2019.

Haarsma, R., Acosta, M., Bakhshi, R., Bretonnière, P.-A. B., Caron, L.-P., Castrillo, M., Corti, S., Davini, P., Exarchou, E., Fabiano, F., Fladrich, U., Fuentes Franco, R., García-Serrano, J., von Hardenberg, J., Koenigk, T., Levine, X., Meccia, V., van Noije, T., van den Oord, G., Palmeiro, F. M., Rodrigo, M., Ruprich-Robert, Y., Le Sager, P., Tourigny, E., Wang, S., van Weele, M., and Wyser, K.: HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR. Description, model performance, data handling and validation, Geoscientific Model Development Discussions, 2020, 1–37, https://doi.org/10.5194/gmd-2019-350, 2020.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geoscientific Model Development, 9, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016, 2016.

Hardiman, S. C., Dunstone, N. J., Scaife, A. A., Smith, D. M., Comer, R., Nie, Y., and Ren, H.-L.: Missing eddy feedback may explain weak signal-to-noise ratios in climate predictions, npj Climate and Atmospheric Science, 5, 57, 2022.

Hassanzadeh, P. and Kuang, Z.: The linear response function of an idealized atmosphere. Part I: Construction using Green's functions and applications, Journal of the Atmospheric Sciences, 73, 3423–3439, 2016.

Hasselmann, K.: Stochastic climate models part I. Theory, tellus, 28, 473–485, 1976.

Hunke, E. C., Lipscomb, W. H., Turner, A. K., Jeffery, N., and Elliott, S.: Cice: the los alamos sea ice model documentation and software user's manual version 4.1 la-cc-06-012, T-3 Fluid Dynamics Group, Los Alamos National Laboratory, 675, 500, 2010.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., et al.: The community earth system model: a framework for collaborative research, Bulletin of the American Meteorological Society, 94, 1339–1360, 2013.

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., et al.: SEAS5: the new ECMWF seasonal forecast system, Geoscientific Model Development, 12, 1087–1117, 2019.

Judt, F.: Atmospheric predictability of the tropics, middle latitudes, and polar regions explored through global storm-resolving simulations, Journal of the Atmospheric Sciences, 77, 257–276, 2020.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S., Danabasoglu, G., Edwards, J., et al.: The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability, Bulletin of the American Meteorological Society, 96, 1333–1349, 2015.

Kim, W. M., Yeager, S., Chang, P., and Danabasoglu, G.: Low-frequency North Atlantic climate variability in the Community Earth System Model large ensemble, Journal of Climate, 31, 787–813, 2018.

Kirk-Davidoff, D. B. and Keith, D. W.: On the climate impact of surface roughness anomalies, Journal of the Atmospheric Sciences, 65, 2215–2234, 2008.

Kravtsov, S.: Pronounced differences between observed and CMIP5-simulated multidecadal climate variability in the twentieth century, Geophysical Research Letters, 44, 5749–5757, 2017.

Kushnir, Y.: Interdecadal variations in North Atlantic sea surface temperature and associated atmospheric conditions, Journal of Climate, 7, 141–157, 1994.

Kushnir, Y., Robinson, W., Bladé, I., Hall, N., Peng, S., and Sutton, R.: Atmospheric GCM response to extratropical SST anomalies: Synthesis and evaluation, Journal of Climate, 15, 2233–2256, 2002.

Kwon, Y.-O., Seo, H., Ummenhofer, C. C., and Joyce, T. M.: Impact of multidecadal variability in Atlantic SST on winter atmospheric blocking, Journal of Climate, 33, 867–892, 2020.

Laloyaux, P., de Boisseson, E., Balmaseda, M., Bidlot, J.-R., Broennimann, S., Buizza, R., Dalhgren, P., Dee, D., Haimberger, L., Hersbach, H., et al.: CERA-20C: A coupled reanalysis of the twentieth century, Journal of Advances in Modeling Earth Systems, 10, 1172–1195, 2018.

Lamarque, J.-F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., Lee, D., Liousse, C., Mieville, A., Owen, B., et al.: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application, Atmospheric Chemistry and Physics, 10, 7017–7039, 2010.

Liang, Y.-C., Frankignoul, C., Kwon, Y.-O., Gastineau, G., Manzini, E., Danabasoglu, G., Suo, L., Yeager, S., Gao, Y., Attema, J. J., et al.: Impacts of Arctic sea ice on cold season atmospheric variability and trends estimated from observations and a multimodel large ensemble, Journal of Climate, 34, 8419–8443, 2021.

Ma, L., Woollings, T., Williams, R. G., Smith, D., and Dunstone, N.: How does the winter jet stream affect surface temperature, heat flux, and sea ice in the North Atlantic?, Journal of Climate, 33, 3711–3730, 2020.

Madec, G. and the NEMO team: NEMO ocean engine version 3.6 stable, Note du Pôle de modélisation de l'Institut Pierre-Simon Laplace, 27, 2016.

Marcheggiani, A., Robson, J., Monerie, P.-A., Bracegirdle, T. J., and Smith, D.: Decadal Predictability of the North Atlantic Eddy-Driven Jet in Winter, Geophysical Research Letters, 50, e2022GL102 071, 2023.

Meccia, V. L., Fuentes-Franco, R., Davini, P., Bellomo, K., Fabiano, F., Yang, S., and von Hardenberg, J.: Internal multi-centennial variability of the Atlantic Meridional Overturning Circulation simulated by EC-Earth3, Climate Dynamics, pp. 1–18, 2022.

Msadek, R., Frankignoul, C., and Li, L. Z.: Mechanisms of the atmospheric response to North Atlantic multidecadal variability: A model study, Climate dynamics, 36, 1255–1276, 2011.

Ogawa, F., Keenlyside, N., Gao, Y., Koenigk, T., Yang, S., Suo, L., Wang, T., Gastineau, G., Nakamura, T., Cheung, H. N., et al.: Evaluating impacts of recent Arctic sea ice loss on the northern hemisphere winter climate change, Geophysical Research Letters, 45, 3255–3263, 2018.

Omrani, N.-E., Keenlyside, N. S., Bader, J., and Manzini, E.: Stratosphere key for wintertime atmospheric response to warm Atlantic decadal conditions, Climate Dynamics, 42, 649–663, 2014.

O'Reilly, C. H., Weisheimer, A., Woollings, T., Gray, L. J., and MacLeod, D.: The importance of stratospheric initial conditions for winter North Atlantic Oscillation predictability and implications for the signal-to-noise paradox, Quarterly Journal of the Royal Meteorological Society, 145, 131–146, 2019.

Ortega, P., Robson, J., Sutton, R. T., and Andrews, M. B.: Mechanisms of decadal variability in the Labrador Sea and the wider North Atlantic in a high-resolution climate model, Climate Dynamics, 49, 2625–2647, 2017.

O'Reilly, C. H., Zanna, L., and Woollings, T.: Assessing external and internal sources of Atlantic multidecadal variability using models, proxy data, and early instrumental indices, Journal of Climate, 32, 7727–7745, 2019.

O'Reilly, C. H., Patterson, M., Robson, J., Monerie, P. A., Hodson, D., and Ruprich-Robert, Y.: Challenges with interpreting the impact of Atlantic Multidecadal Variability using SST-restoring experiments, npj Climate and Atmospheric Science, 6, 14, 2023.

Palmer, T. and Zhaobo, S.: A modelling and observational study of the relationship between sea surface temperature in the north-west Atlantic and the atmospheric general circulation, Quarterly Journal of the Royal Meteorological Society, 111, 947–975, 1985.

Papritz, L. and Spengler, T.: Analysis of the slope of isentropic surfaces and its tendencies over the North Atlantic, Quarterly Journal of the Royal Meteorological Society, 141, 3226–3238, 2015.

Parker, T., Woollings, T., Weisheimer, A., O'Reilly, C., Baker, L., and Shaffrey, L.: Seasonal Predictability of the Winter North Atlantic Oscillation From a Jet Stream Perspective, Geophysical Research Letters, 46, 10 159–10 167, https://doi.org/10.1029/2019GL084402, 2019.

Peings, Y. and Magnusdottir, G.: Response of the Wintertime Northern Hemisphere Atmospheric Circulation to Current and Projected Arctic Sea Ice Decline: A Numerical Study with CAM5, Journal of Climate, 27, 244 – 264, https://doi.org/10.1175/JCLI-D-13-00272.1, 2014a.

Peings, Y. and Magnusdottir, G.: Forcing of the wintertime atmospheric circulation by the multidecadal fluctuations of the North Atlantic Ocean, Environmental Research Letters, 9, 034 018, 2014b.

Peings, Y. and Magnusdottir, G.: Wintertime atmospheric response to Atlantic multidecadal variability: Effect of stratospheric representation and ocean–atmosphere coupling, Climate dynamics, 47, 1029–1047, 2016.

Peings, Y., Simpkins, G., and Magnusdottir, G.: Multidecadal fluctuations of the North Atlantic Ocean and feedback on the winter climate in CMIP5 control simulations, Journal of Geophysical Research: Atmospheres, 121, 2571–2592, 2016.

Poli, P., Hersbach, H., Tan, D., Dee, D., Thépaut, J.-N., Simmons, A., Peubey, C., Laloyaux, P., Komori, T., Berrisford, P., and Dragani, R.: The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C), ERA report series, 2013.

Portal, A., Pasquero, C., D'andrea, F., Davini, P., Hamouda, M. E., and Rivière, G.: Influence of Reduced Winter Land–Sea Contrast on the Midlatitude Atmospheric Circulation, Journal of Climate, 35, 2637–2651, 2022.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, Journal of Geophysical Research D: Atmospheres, 2003.

Roberts, C. D., Senan, R., Molteni, F., Boussetta, S., Mayer, M., and Keeley, S. P. E.: Climate model configurations of the ECMWF Integrated Forecasting System (ECMWF-IFS cycle 43r1) for HighResMIP, Geoscientific Model Development, 11, 3681–3712, https://doi.org/10.5194/gmd-11-3681-2018, 2018a.

Roberts, M. J., Vidale, P. L., Senior, C., Hewitt, H. T., Bates, C., Berthou, S., Chang, P., Christensen, H. M., Danilov, S., Demory, M.-E., Griffies, S. M., Haarsma, R., Jung, T., Martin, G., Minobe, S., Ringler, T., Satoh, M., Schiemann, R., Scoccimarro, E., Stephens, G., and Wehner, M. F.: The Benefits of Global High Resolution for Climate Simulation: Process Understanding and the Enabling of Stakeholder Decisions at the Regional Scale, Bulletin of the American Meteorological Society, 99, 2341 – 2359, https://doi.org/10.1175/BAMS-D-15-00320.1, 2018b.

Robson, J., Menary, M. B., Sutton, R. T., Mecking, J., Gregory, J. M., Jones, C., Sinha, B., Stevens, D. P., and Wilcox, L. J.: The role of anthropogenic aerosol forcing in the 1850–1985 strengthening of the AMOC in CMIP6 historical simulations, Journal of Climate, 35, 3243–3263, 2022.

Ruggieri, P., Bellucci, A., Nicolí, D., Athanasiadis, P. J., Gualdi, S., Cassou, C., Castruccio, F., Danabasoglu, G., Davini, P., Dunstone, N., et al.: Atlantic multidecadal variability and North Atlantic jet: a multimodel view from the decadal climate prediction project, Journal of Climate, 34, 347–360, 2021.

Scaife, A. A. and Smith, D.: A signal-to-noise paradox in climate science, npj Climate and Atmospheric Science, 1, 28, 2018.

Schemm, S.: Toward Eliminating the Decades-Old "Too Zonal and Too Equatorward" Storm-Track Bias in Climate Models, Journal of Advances in Modeling Earth Systems, 15, e2022MS003 482, 2023.

Seager, R., Kushnir, Y., Visbeck, M., Naik, N., Miller, J., Krahmann, G., and Cullen, H.: Causes of Atlantic Ocean climate variability between 1958 and 1998, Journal of Climate, 13, 2845–2862, 2000.

Sein, D. V., Koldunov, N. V., Danilov, S., Wang, Q., Sidorenko, D., Fast, I., Rackow, T., Cabos, W., and Jung, T.: Ocean Modeling on a Mesh With Resolution Following the Local Rossby Radius, Journal of Advances in Modeling Earth Systems, 9, 2601–2614, https://doi.org/10.1002/2017MS001099, 2017.

Shepherd, T. G.: Bringing physical reasoning into statistical practice in climate-change science, Climatic Change, 169, 2, 2021.

Simpson, I. R., Deser, C., McKinnon, K. A., and Barnes, E. A.: Modeled and Observed Multidecadal Variability in the North Atlantic Jet Stream and Its Connection to Sea Surface Temperatures, Journal of Climate, 31, 8313 – 8338, https://doi.org/10.1175/JCLI-D-18-0168.1, 2018.

Smith, D., Eade, R., Scaife, A. A., Caron, L.-P., Danabasoglu, G., DelSole, T., Delworth, T., Doblas-Reyes, F., Dunstone, N., Hermanson, L., et al.: Robust skill of decadal climate predictions, Npj Climate and Atmospheric Science, 2, 1–10, 2019.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological Society, 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.

Visbeck, M., Chassignet, E. P., Curry, R. G., Delworth, T. L., Dickson, R. R., and Krahmann, G.: The ocean's response to North Atlantic Oscillation variability, Geophysical Monograph-American Geophysical Union, 134, 113–146, 2003.

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., Douville, H., Ethé, C., Franchistéguy, L., Geoffroy, O., Lévy, C., Madec, G., Meurdesoif, Y., Msadek, R., Ribes, A., Sanchez-Gomez, E., Terray, L., and Waldman, R.: Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1, Journal of Advances in Modeling Earth Systems, 11, 2177–2213, https://doi.org/10.1029/2019MS001683, 2019.

Wagner, S. and Shishkina, O.: Heat flux enhancement by regular surface roughness in turbulent thermal convection, Journal of Fluid Mechanics, 763, 109–135, https://doi.org/10.1017/jfm.2014.665, 2015.

Wang, X., Li, J., Sun, C., and Liu, T.: NAO and its relationship with the Northern Hemisphere mean surface temperature in CMIP5 simulations, Journal of Geophysical Research: Atmospheres, 122, 4202–4227, 2017.

Weisheimer, A., Schaller, N., O'Reilly, C., MacLeod, D. A., and Palmer, T.: 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, https://doi.org/10.1002/qj.2976, 2017.

Weisheimer, A., Befort, D. J., MacLeod, D., Palmer, T., O'Reilly, C., and Strømmen, K.: Seasonal forecasts of the twentieth century, Bulletin of the American Meteorological Society, 101, E1413–E1426, 2020.

Weisheimer, A.; O'Reilly, C.: Initialised seasonal forecast of the 20th Century., https://catalogue.ceda.ac.uk/uuid/6e1c3df49f644a0f812818080bed5e45, 2020.

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H. T., Hill, R., Hyder, P., Ineson, S., Johns, T. C., Keen, A. B., Lee, R. W., Megann, A., Milton, S. F., Rae, J. G. L., Roberts, M. J., Scaife, A. A., Schiemann, R., Storkey, D., Thorpe, L., Watterson, I. G., Walters, D. N., West, A., Wood, R. A., Woollings, T., and Xavier, P. K.: 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, 357–380, https://doi.org/10.1002/2017MS001115, 2018.

Willison, J., Robinson, W. A., and Lackmann, G. M.: The importance of resolving mesoscale latent heating in the North Atlantic storm track, Journal of the Atmospheric Sciences, 70, 2234–2250, 2013.

Wills, R. C., Armour, K. C., Battisti, D. S., and Hartmann, D. L.: Ocean–atmosphere dynamical coupling fundamental to the Atlantic multidecadal oscillation, Journal of Climate, 32, 251–272, 2019.

Woollings, T., Hannachi, A., and Hoskins, B.: Variability of the North Atlantic eddy-driven jet stream, Quarterly Journal of the Royal Meteorological Society, 136, 856–868, https://doi.org/10.1002/qj.625, 2010.

Woollings, T., Czuchnicki, C., and Franzke, C.: Twentieth century North Atlantic jet variability, Quarterly Journal of the Royal Meteorological Society, 140, 783–791, https://doi.org/10.1002/qj.2197, 2014.

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.

Yeager, S.: Decadal Prediction Large Ensemble Project, https://doi.org/10.5065/D6DR2T8H, 2018.

Yeager, S.: The abyssal origins of North Atlantic decadal predictability, Climate Dynamics, 55, 2253–2271, 2020.

Yeager, S., Danabasoglu, G., Rosenbloom, N., Strand, W., Bates, S., Meehl, G., Karspeck, A., Lindsay, K., Long, M., Teng, H., et al.: Predicting near-term changes in the earth system: a large ensemble of initialized decadal prediction simulations using the community earth system model, Bulletin of the American Meteorological Society, 99, 1867–1886, 2018.

Zhang, R.: Anticorrelated multidecadal variations between surface and subsurface tropical North Atlantic, Geophysical Research Letters, 34, 2007.

Zhang, R., Delworth, T. L., Sutton, R., Hodson, D. L., Dixon, K. W., Held, I. M., Kushnir, Y., Marshall, J., Ming, Y., Msadek, R., et al.: Have aerosols caused the observed Atlantic multidecadal variability?, Journal of the Atmospheric Sciences, 70, 1135–1144, 2013.

Zhang, R., Sutton, R., Danabasoglu, G., Kwon, Y.-O., Marsh, R., Yeager, S. G., Amrhein, D. E., and Little, C. M.: A review of the role of the Atlantic meridional overturning circulation in Atlantic multidecadal variability and associated climate impacts, Reviews of Geophysics, 57, 316–375, 2019.

Zhang, W., Kirtman, B., Siqueira, L., Clement, A., and Xia, J.: Understanding the signal-to-noise paradox in decadal climate predictability from CMIP5 and an eddying global coupled model, Climate dynamics, 56, 2895–2913, 2021.