UCLA
UCLA Previously Published Works

Title
Modeled and Observed Multidecadal Variability in the North Atlantic Jet Stream and Its Connection to Sea Surface Temperatures Modeled and Observed Multidecadal Variability in the North Atlantic Jet Stream and Its Connection to Sea Surface Temperature...

Authors
Simpson, Isla R
Deser, Clara
McKinnon, Karen A
et al.

Modeled and Observed Multidecadal Variability in the North Atlantic Jet Stream and Its Connection to Sea Surface Temperatures

Isla R. Simpson and Clara Deser

Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, Colorado

Karen A. McKinnon

Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, Colorado, and Descartes Labs, Santa Fe, New Mexico

Elizabeth A. Barnes

Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado

(Manuscript received 22 March 2018, in final form 8 June 2018)

ABSTRACT

Multidecadal variability in the North Atlantic jet stream in general circulation models (GCMs) is compared with that in reanalysis products of the twentieth century. It is found that almost all models exhibit multi-decadal jet stream variability that is entirely consistent with the sampling of white noise year-to-year atmospheric fluctuations. In the observed record, the variability displays a pronounced seasonality within the winter months, with greatly enhanced variability toward the late winter. This late winter variability exceeds that found in any GCM and greatly exceeds expectations from the sampling of atmospheric noise, motivating the need for an underlying explanation. The potential roles of both external forcings and internal coupled ocean–atmosphere processes are considered. While the late winter variability is not found to be closely connected with external forcing, it is found to be strongly related to the internally generated component of Atlantic multidecadal variability (AMV) in sea surface temperatures (SSTs). In fact, consideration of the seasonality of the jet stream variability within the winter months reveals that the AMV is far more strongly connected to jet stream variability during March than the early winter months or the winter season as a whole. Reasoning is put forward for why this connection likely represents a driving of the jet stream variability by the SSTs, although the dynamics involved remain to be understood. This analysis reveals a fundamental mismatch between late winter jet stream variability in observations and GCMs and a potential source of long-term predictability of the late winter Atlantic atmospheric circulation.

1. Introduction

The North Atlantic is a region characterized by substantial variability in the circulation of both the atmosphere and ocean on a wide range of time scales. The highly populated surrounding continents of North America and Europe, as well as North Africa and Greenland, are impacted by this variability as it manifests in fluctuations in the characteristics of weather patterns, with impacts on storminess, temperature, and precipitation. This is particularly true during the wintertime when the North Atlantic storm track is at its most active and its variability is most pronounced (Hurrell et al. 2003). Accurate prediction of the climate in these regions using general circulation models (GCMs), and an appreciation of its uncertainties, whether it be on subseasonal-to-seasonal, decadal, or centennial time scales, requires an accurate representation of both forced and unforced variability in the atmosphere, the ocean, and the coupling between them. As a result of the wide range of complex processes that contribute to the climate of the North Atlantic and its variability, this continues to be a challenge (e.g., Woollings 2010).

One aspect where GCMs have recently been shown to exhibit a deficiency is in their decadal to multidecadal variability in the North Atlantic jet stream, as viewed through the North Atlantic Oscillation (NAO). The NAO is the dominant mode of variability in North

Corresponding author: Isla R. Simpson, islas@ucar.edu

Atlantic sea level pressure (SLP) (e.g., Walker and Bliss 1932; van Loon and Rogers 1978; Wallace and Gutzler 1981; Barnston and Livezey 1987; Hurrell and Deser 2010; Woollings et al. 2015); while it is not the only mode of variability in the North Atlantic jet stream (e.g., Athanasiadis et al. 2010), much of the research on North Atlantic atmospheric variability has centered around it. Hurrell (1995) demonstrated pronounced decadal to multidecadal time scale fluctuations in the NAO during the last century, with associated fluctuations in regional temperatures and precipitation over northern Europe, the Mediterranean, North America, and Greenland (Hurrell and van Loon 1997). This sparked debate as to whether such multidecadal fluctuations should be taken as evidence of multidecadal time scales in the atmosphere, with a particular underlying cause, or whether they might simply arise from the sampling of high-frequency white noise or weakly red noise processes (i.e., the averaging of atmospheric noise with no need to invoke an underlying low-frequency forcing) (Wunsch 1999; Stephenson et al. 2000; Feldstein 2000). With regard to the long-term positive NAO trend that was observed from the 1950s to 1990s (Hurrell 1995), while the sampling of high-frequency atmospheric noise almost certainly contributes, studies have argued for a role for forcing from increasing greenhouse gas concentrations (Gillett et al. 2003) and from sea surface temperature (SST) variability in the extratropics (Rodwell et al. 1999) or in the tropics (Hoerling et al. 2001; Hurrell et al. 2004). While some studies have shown that the magnitude of the long-term NAO trend seen over the latter half of the century is reproducible in models and consistent with internal ocean–atmosphere variability (Selten et al. 2004; Raible et al. 2005; Deser and Phillips 2009), models generally fail to capture the amplitude of the shorter-term trend from the mid-1960s to the mid-1990s even when they are run with prescribed historical SSTs and forced with time-evolving greenhouse gas and aerosol concentrations (Scaife et al. 2009; Kim et al. 2018).

While the debate over the underlying reasons for this multidecadal variability in the North Atlantic jet stream continues, it appears that the current generation of GCMs does not exhibit the same degree of multidecadal North Atlantic jet stream variability as has been observed over the twentieth century. Kravtsov (2017) and Wang et al. (2017) provided a comparison of winter and annual mean NAO variability on different time scales between observations/reanalysis products and models from phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012). They each found that interannual NAO variability was simulated with fidelity but almost all models were deficient in their

representation of NAO variability on longer time scales. Similar conclusions were reached by Kim et al. (2018) using the Community Earth System Model, version 1 (CESM1; Hurrell et al. 2013). They further argued that the deficient multidecadal NAO variability may be a cause of reduced variability in the ocean's Atlantic meridional overturning circulation (AMOC) and associated SSTs.

Given the importance of the North Atlantic jet stream for the climate of the surrounding continents as well as its potential global impact through the ocean circulation, it is important that models simulate its variability with fidelity. However, our understanding of the reasons behind the model deficiencies cited above, remains limited. One, or a combination, of the following three possibilities is likely responsible:

1) The observed multidecadal variability does indeed result from the averaging of atmospheric noise (e.g., Wunsch 1999; Feldstein 2000) and either the observed sequence of atmospheric noise was unlikely, or models are deficient in their representation of this atmospheric noise.

2) The observed multidecadal variability has been externally forced by factors such as changing greenhouse gas (GHG) or aerosol concentrations, volcanic eruptions, or solar variability, and either (i) the forcing datasets used to drive our models are in error or (ii) the models do not correctly simulate the response to these forcings. These forcings could either act directly on the North Atlantic jet stream or indirectly through their influence on the ocean or both.

3) The observed multidecadal variability arises from internal coupled ocean–atmosphere processes. That is, the long time scales inherent to the ocean are responsible for the long time scale jet variability seen in the observations, and models are either deficient in their representation of the important ocean processes, the ocean–atmosphere coupling, or both.

In this study, we elucidate some nuanced aspects of the observed multidecadal variability in the wintertime North Atlantic jet stream and its representation in models. We offer a slightly different view of this issue from previous studies by not limiting ourselves solely to the NAO and by considering the additional seasonality within the winter months. While, as will be discussed, our capacity for gaining a complete mechanistic understanding of the variability (either through the use of models or via the observed record before the satellite era) is limited, our analysis does lead us to conjecture that possibility 3 is likely playing an important role. That is, the observed multidecadal variability arises from internal coupled ocean–atmosphere processes and models are deficient

in their representation of these processes. This will, however, only be proven correct (or otherwise) as either our models improve or our observational record lengthens.

The observation- and model-based datasets and methods used are described in section 2. The variability in coupled models is then compared with that in observations in section 3, where it is shown that the observed variability is considerably greater than typically found in GCMs during the late winter. The nature of the variability that has been observed in terms of Atlantic jet structure is then briefly described in section 4 before its linkage with SST variability and external forcings is assessed in section 5. Discussions and conclusions are then provided in section 6.

2. Observation-based datasets, model simulations, and methods

a. Reanalyses

Our focus on decadal to multidecadal time scale variability necessitates the use of observation-based datasets that are of a sufficient length. We therefore use the reanalysis products that cover the entire twentieth century, namely the European Centre for Medium-Range Weather Forecasts (ECMWF) Twentieth Century Reanalysis (ERA20C; Poli et al. 2016), which is available from 1900 to 2010, and the National Oceanic and Atmospheric Administration's (NOAA's) Twentieth Century Reanalysis (20CR; Compo et al. 2011), which is available from 1850 to 2014. Only surface pressure observations are assimilated in 20CR while additional observations of marine surface winds are assimilated in ERA20C. Of these two datasets, our primary focus will be on ERA20C although the conclusions do not differ from 20CR. These will also be compared with the shorter reanalyses over the recent decades that assimilate a wider range of observations: ERA-Interim from 1979 to 2017 (Dee et al. 2011), MERRA2 from 1980 to 2017 (Gelaro et al. 2017), JRA-55 from 1958 to 2016 (Kobayashi et al. 2015), and ERA-40 from 1958 to 2001 (Uppala et al. 2005).

b. SST datasets and indices

Three SST datasets will be used to assess the linkage between Atlantic jet stream variability and SST variability: NOAA's Extended Reconstruction SSTs versions 3b and 5 [ERSSTv3b (Smith et al. 2008) and ERSSTv5 (Huang et al. 2017)], both of which are available from 1854 to 2017, along with the Hadley Centre Global Sea Ice and Sea Surface Temperature dataset (HadISST; Rayner et al. 2003) which is available from 1870 to 2016.

We make use of an index of area averaged SST from 80°W to 0° and from the equator to 60°N and refer to this as the Atlantic multidecadal variability (AMV) index. A variety of methods have been used to isolate the internal variability component of AMV from the externally forced component (Trenberth and Shea 2006; Ting et al. 2009; Frankcombe et al. 2015; Frankignoul et al. 2017; Murphy et al. 2017) and here we make use of three of them: the Trenberth and Shea (2006) method (the TS method), the linear inverse model (LIM) optimal method of Frankignoul et al. (2017) (the LIMoptimal method), and a simple subtraction of a linear trend following Murphy et al. (2017) (the Linear method). The first two methods, TS and LIMoptimal, are designed to remove the contribution of all external forcings and Frankignoul et al. (2017) demonstrated their success in doing so in models in the North Atlantic, particularly for the LIMoptimal method. On the other hand, the Linear method does not successfully remove the entire externally forced component (Tandon and Kushner 2015; Frankignoul et al. 2017) but we are motivated to consider it following Murphy et al. (2017), who argue that natural and anthropogenic aerosol forcing may have been an important driver of past AMV and that this index retains that component of forced variability.

For the TS method, the North Atlantic SST anomalies are taken relative to the average, from 60°S to 60°N over all ocean basins. For the LIMoptimal method the forced component of North Atlantic SST variability has been removed based on a LIM estimate with an optimal perturbation filter prior to taking the North Atlantic average [see Frankignoul et al. (2017) for more details]. This LIMoptimal estimate of internal SST variability has been provided in 3-month averages (January–March, etc.) for 1900–2015 for ERSSTv5 and HadISST (G. Gastineau 2017, personal communication). Finally, for the Linear method the linear trend over the length of the record for the given SST dataset was subtracted from each ocean grid point before calculating the North Atlantic average.

These indices are low-pass filtered by taking a running mean, typically of 20-yr length, but we emphasize that the pattern of SST variability that it represents does change with the time scale considered by showing the SST patterns that accompany the TS AMV index in Fig. 1. Interannually, it depicts a horseshoe pattern in North Atlantic SSTs with a positive anomaly in the subpolar gyre and an accompanying positive southern lobe (Fig. 1a). It is this horseshoe pattern that is typically referred to as the Atlantic multidecadal oscillation (AMO) (e.g., Kushnir 1994; Enfield et al. 2001; Trenberth and Shea 2006), and it has been argued that the direct forcing from North Atlantic wind variability plays an
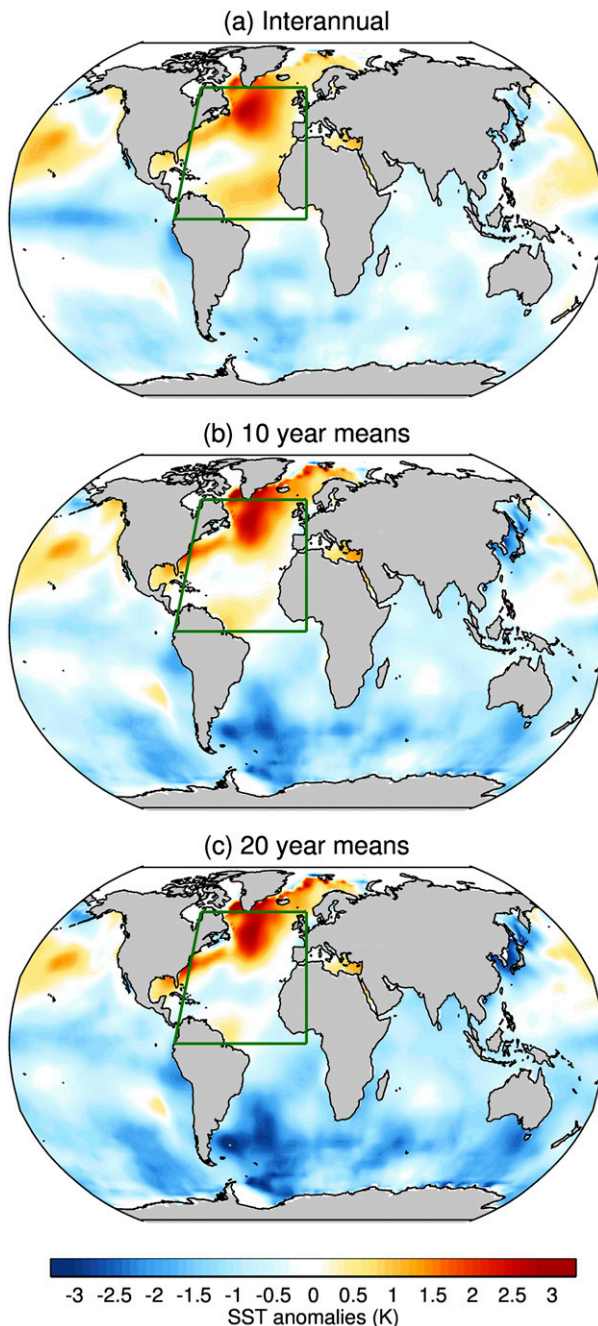
FIG. 1. ERSSTv5 DJFM SSTs regressed onto the DJFM Trenberth and Shea (2006) AMV index for (a) interannual variability, (b) 10-yr running means, and (c) 20-yr running means. The green box depicts the region used to define the AMV index.

negative anomalies over much of the low-latitude North Atlantic (Fig. 1c). The accompanying negative anomalies in the Southern Hemisphere (SH) also intensify as progressively longer time scales are considered. The more localized anomaly in the subpolar gyre region has been argued as evidence of an impact of anomalous heat transports by the ocean circulation on these longer time scales (Delworth et al. 2017). Since our focus will primarily be on 20-yr running means, it is the global pattern with a locally intense anomaly in SST in the subpolar gyre and oppositely signed anomalies in the SH (Fig. 1c) that our AMV index will be describing.

c. Model simulations

1) CESM SIMULATIONS

A 40-member ensemble of coupled simulations are available through the CESM Large Ensemble (LENS) project (Kay et al. 2015). These make use of the fully coupled CESM, version 1 with the Community Atmosphere Model, version 5 (CESM1-CAM5; Hurrell et al. 2013) at approximately 1° horizontal resolution in the ocean and atmosphere. Each member begins in 1920 and is forced with the CMIP5 historical forcings until 2005 and the representative concentration pathway 8.5 (RCP8.5) scenario thereafter. The first member was branched from a preindustrial control simulation, representative of conditions in the 1850s. The remaining 39 members were then branched from year 1920 of this initial simulation with an additional round-off level perturbation to the surface air temperature field to initiate a different evolution of the climate system due to internal variability. The accompanying 1200-yr-long coupled preindustrial control simulation (piControl) with forcings that are representative of year 1850 will also be used.

We also complement these free-running coupled simulations with initialized hindcasts. A 40-member ensemble of coupled decadal prediction simulations initialized from 1 November each year from 1954 to 2015 with the same version of CESM have been made available through the CESM decadal prediction large ensemble project (Yeager et al. 2018). These are run with the same forcings as LENS but with ocean and sea ice initial conditions obtained from a reanalysis-forced simulation with the ocean and sea ice models while the atmosphere initial conditions are those of the 40 members of the free-running LENS. These, therefore, represent ocean and sea ice initialized complements of the free-running LENS simulations.

To assess the importance of the historical evolution of SSTs, we make use of a 10-member "AMIP" ensemble with the same model version but with prescribed

important role in producing it (Clement et al. 2015; Cane et al. 2017). With 10-yr running means, the southern lobe of the horseshoe is weaker (Fig. 1b), and it has been argued that on these time scales, ocean dynamics plays a central role in AMV (Zhang et al. 2016). With 20-yr running means the southern lobe is actually replaced by

time-evolving observed SSTs and sea ice concentrations from ERSSTv4 (Huang et al. 2015). As with the coupled historical simulations these are run under the CMIP5 historical and RCP8.5 (after 2005) forcing scenarios and they extend from 1900 to 2009.

2) OTHER MODEL SIMULATIONS

The piControl simulation and each available historical member (from 1900 to 2005) for 35 models that participated in CMIP5 (listed in Table 1) will also be analyzed. Equivalents to the CESM AMIP simulations of sufficient length are not available through CMIP5, but we make use of an equivalent 10-member ensemble with the ECMWF model (the underlying model of ERA20C reanalysis), known as the ERA20CM simulations (Hersbach et al. 2015), and refer to them as ECMWF AMIP. These extend from 1900 to 2010 under CMIP5 forcings with prescribed SSTs and sea ice from HadISST2.

d. Methods

With our focus being the eddy-driven jet stream in the North Atlantic, the primary field of interest will be 700-hPa zonal wind (U700). Before comparing the variability between the reanalyses and the models, each dataset is first interpolated onto a $2° \times 2°$ longitude–latitude grid using a cubic spline interpolation and then isotropically smoothed in the spectral domain retaining only scales larger than total wavenumber 42 according to Sardeshmukh and Hoskins [1984, their Eq. (9) with $n_o = 42$ and $r = 1$]. This ensures the same spatial scales are represented in each dataset, thereby removing any potential grid dependency. The same procedure was followed prior to the analysis of SLP variability discussed in appendix A. The various datasets are all of different lengths. For any comparison, the maximum period of overlap for the datasets being considered will be used, with the minimum length being 91 years. The time period considered is, therefore, not entirely consistent throughout the analysis but results are unaffected by the exact time period used.

The primary quantity used to characterize multidecadal variability will be the standard deviation σ of running means (typically 20-yr means) of monthly or seasonal averaged U700 at grid points in the Atlantic. The extent to which this variability exceeds expectations from the sampling of white noise year-to-year variability will be assessed by comparison with the percentiles of the distribution of σ values obtained from the running means of 1000 synthetic time series of an equivalent length (and ensemble size in the case of the LENS), generated from Gaussian white noise with a standard deviation equal to the interannual standard deviation at a given grid point for the given month or season. A

TABLE 1. List of the number of historical members (from 1900–2005) and the number of years in the piControl simulation used for each model. (Expansions of acronyms are available online at http://www.ametsoc.org/PubsAcronymList.)

Model	Historical members	piControl length (yr)
ACCESS1.0	3	500
ACCESS1.3	3	500
BCC-CSM1.1	3	500
BCC-CSM1.1-M	3	400
BNU-ESM	1	559
CanESM2	5	996
CCSM4	6	1051
CESM1-CAM5	3	319
CESM1-WACCM	1	200
CMCC-CM	1	330
CMCC-CMS	1	500
CNRM-CM5	10	850
CSIRO-Mk3.6.0	10	500
FGOALS-g2	5	700
FIO-ESM	3	800
GFDL-CM3	5	500
GFDL-ESM2G	1	500
GFDL-ESM2M	1	500
GISS-E2-H	5	780
GISS-E2-R	6	850
HadGEM2-AO	1	700
HadGEM2-CC	1	240
HadGEM2-ES	5	576
INM-CM4	1	500
IPSL-CM5A-LR	6	1000
IPSL-CM5A-MR	3	300
IPSL-CM5B-LR	1	300
MIROC5	5	670
MIROC-ESM	3	630
MIROC-ESM-CHEM	1	255
MPI-ESM-LR	3	1000
MPI-ESM-MR	3	1000
MRI-CGCM3	3	500
NorESM1-M	3	501
NorESM1-ME	1	252

Kolmogorov–Smirnov test on the observed U700 confirms that a Gaussian distribution is a reasonable representation of U700 interannual variability (not shown).

When examining the link between U700 variability and SSTs/AMV, the Pearson's correlation coefficient will be evaluated. To assess the significance of a given correlation, the resampling methodology of Delworth et al. (2017) will be employed, which allows for a suitable assessment of significance with strongly autocorrelated time series. To assess the significance of the correlation between two time series TS1 and TS2 of length N, this procedure works as follows: TS1 is shuffled by obtaining a random number i between 1 and N and piecing together the segment from year $= i$ to year $= N$ with the segment from year $= 1$ to year $= i - 1$. This is repeated for TS2 with a different random value i, and the correlation between these two
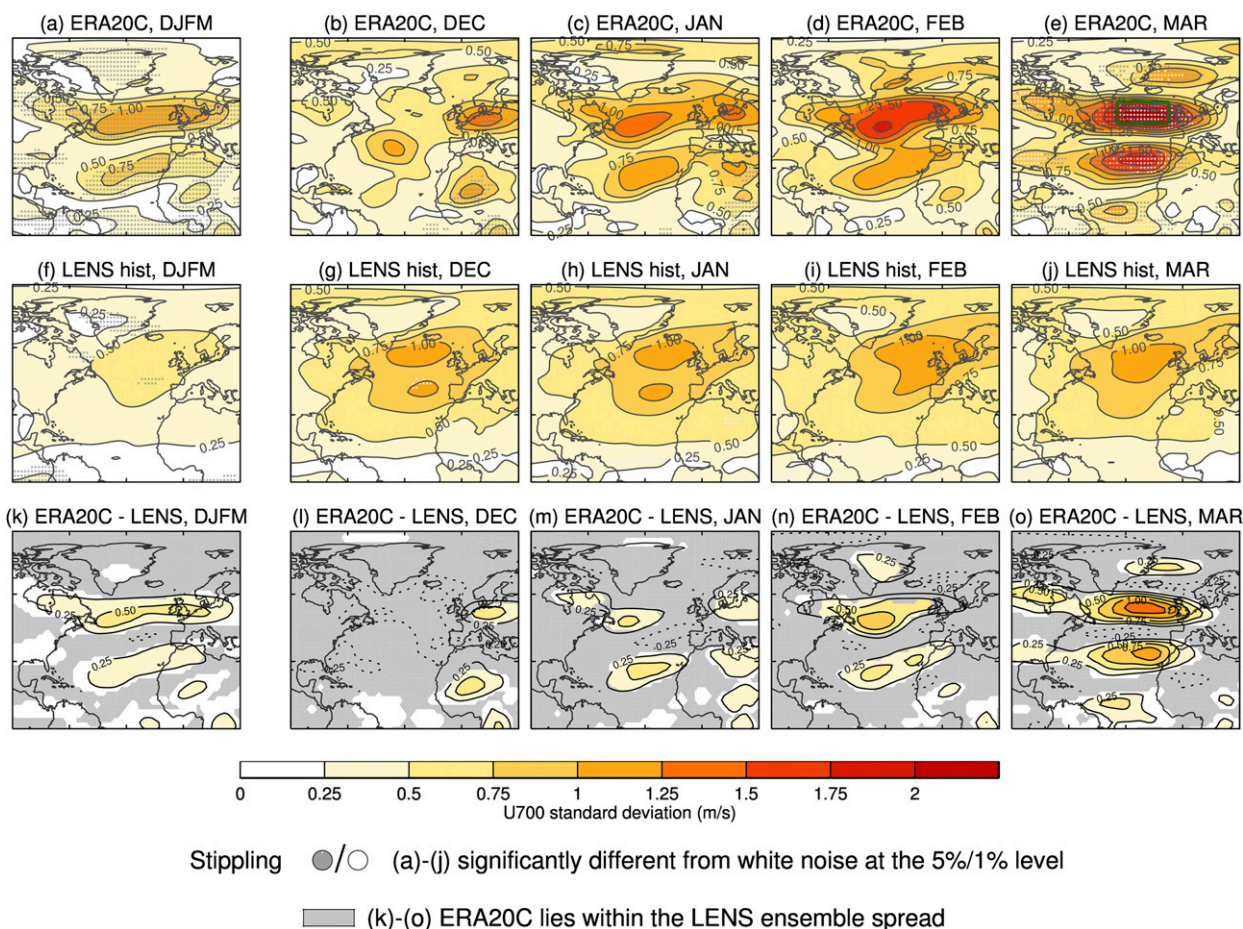
FIG. 2. U700 standard deviations σ of 20-yr running means for (a),(f),(k) the DJFM season, (b),(g),(l) December, (c),(h),(m) January, (d),(i),(n) February, and (e),(j),(o) March. (top) ERA20C reanalysis from 1920–2010, (middle) the CESM LENS from 1920–2010, and (bottom) the difference between ERA20C and LENS. The stippling in (a)–(j) indicates regions where the variability is significantly different from white noise at 5% (gray) and 1% (white) levels. Spatial correlation has been accounted for by increasing the threshold percentile value for significance by the false discovery rate method of Wilks [2016, their Eq. (3)] with a false discovery rate control value of 20%. Gray regions in (k)–(o) indicate where ERA20C lies within the LENS ensemble spread. The green box in (e) shows the region used for the U700NA index in subsequent figures. Only 1920–2010 is used for ERA20C so that equivalent record lengths are used in ERA20C and the LENS historical members for the shading in (k)–(o) but (a)–(e) are very similar when the full record is used.

shuffled time series is assessed. This is done 10 000 times with different values of i to build up a distribution of correlations that could occur between these time series by chance. Significance will be assessed by a two-sided test (i.e., for significance at the 5% level, we assess whether the correlation is greater than the 97.5 or less than the 2.5 percentile value of the 10 000 samples).

3. A comparison of multidecadal variability in North Atlantic zonal wind between models and reanalyses

Multidecadal variability in U700 is depicted on a grid point basis for ERA20C and CESM in Fig. 2 by showing the σ of 20-yr running means from 1920 to 2010 for the

December–March average and each of the individual winter months. For ERA20C the variability increases throughout the winter season, strongly maximizing in March (Fig. 2e), whereas for CESM LENS there is no apparent seasonality within the winter months (Figs. 2g–j). A comparison of the variability in the LENS with that in ERA20C (Figs. 2k–o) reveals that, in the later portion of the winter (March and to a lesser extent February), the variability in ERA20C is considerably greater than the mean variability calculated from the individual LENS members and is, in fact, greater than estimated from any individual LENS member. This March variability contributes to the December–March (DJFM) differences between CESM LENS and ERA20C, particularly to the west of the United Kingdom (Fig. 2k). However, it

should be noted that even if March were excluded and the DJF average considered instead, there are still significant differences between CESM LENS and ERA20C to the east of Newfoundland and in the South Atlantic because of the contributions from January and February. Note also that the lack of structure in the CESM variability is a result of the far greater sample size. Individual members can produce local centers of action in their variability similar to those seen in ERA20C but the magnitude is always considerably smaller (Fig. 3).

Where stippling is present in Figs. 2a–j the variability is significantly different from white noise at the 5% (gray) and 1% (white) levels. The lack of stippling in Figs. 2f–j indicates that the variability within LENS is, for the most part, not distinguishable from the sampling of white noise; that is, the variability is entirely consistent with the properties of a time series that exhibits no correlation from one year to the next and there is, therefore, no need to invoke an underlying causal mechanism acting on long time scales (Wunsch 1999; Feldstein 2000). Indeed, the variability in an atmosphere-only control simulation with prescribed climatological SSTs, where there is unlikely to be much of a source of persistence from one year to the next (with the possible exception of memory from the land), is found to be very similar to that in these forced coupled runs (not shown).

In ERA20C in the early winter, over the North Atlantic, the variability is also not greater than expected from sampling white noise. There are regions over Europe and North Africa for which this is not the case and, while not our focus, these will be discussed briefly in section 5. Over the North Atlantic, as the winter progresses, the variability becomes increasingly outside of the range expected from white noise, becoming significantly different from white noise at the 1% level in March (Fig. 2e). In fact, the variability in March of the area averaged U700 over the North Atlantic subpolar gyre (green boxed region in Fig. 1e, referred to as U700NA hereafter) lies around the 99.96th percentile of white noise samples; that is, there is less than a 1 in 2500 chance that the observed variability in 20-yr means has arisen from the sampling of white noise interannual variability, or given that four separate months have been sampled, less than a 1 in 625 chance of obtaining one month that exhibits this degree of variability. It should be noted that similar conclusions hold if instead synthetic time series were generated from the fit to an autoregressive red noise (AR1) process, since the autocorrelation from one year to the next is typically only around 0.2 and the associated added persistence in the synthetic time series does not substantially alter the stippled regions.

The U700NA index was chosen based on the region that exhibits the greatest multidecadal variability. To place this index into context, we show regression maps of
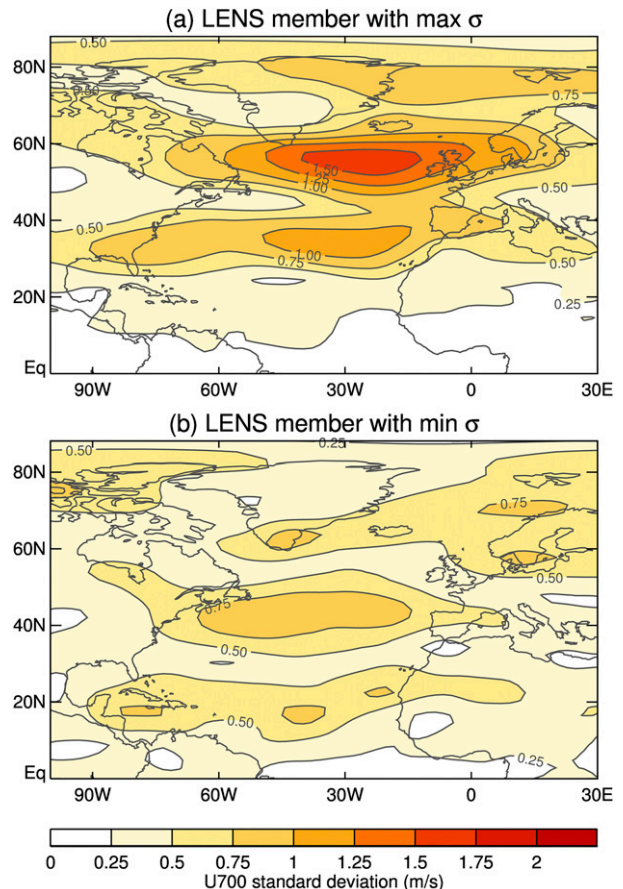

FIG. 3. As in Fig. 2j, but for the individual LENS historical member that exhibits the (a) maximum and (b) minimum March 20-yr running mean σ in U700NA (green box in Fig. 2e).

March U700 onto the March U700NA index for ERA20C (Fig. 4) and CESM LENS (Fig. 4b) on interannual time scales. Variability in U700 that accompanies this index is similar between ERA20C and CESM LENS and is characterized by zonally symmetric anomalies across the Atlantic sector, representing a poleward shifting of the Atlantic jet in the western part of the basin and a strengthening at the jet exit region. These zonal wind anomalies are very similar to those that accompany the dominant (shifting) mode of variability in the Atlantic jet stream (e.g., Eichelberger and Hartmann 2007, their Fig. 2b). While the structure of the anomalies is very similar between ERA20C and CESM LENS, due to differences in the climatological jet position, they represent a strengthening of the jet stream over a wider portion of the basin in CESM LENS.

That the variability in DJFM averaged U700 is significantly different from white noise in ERA20C (Fig. 2a) but not in CESM (Fig. 2f) is consistent with the NAO analysis of Kim et al. (2018). However, while
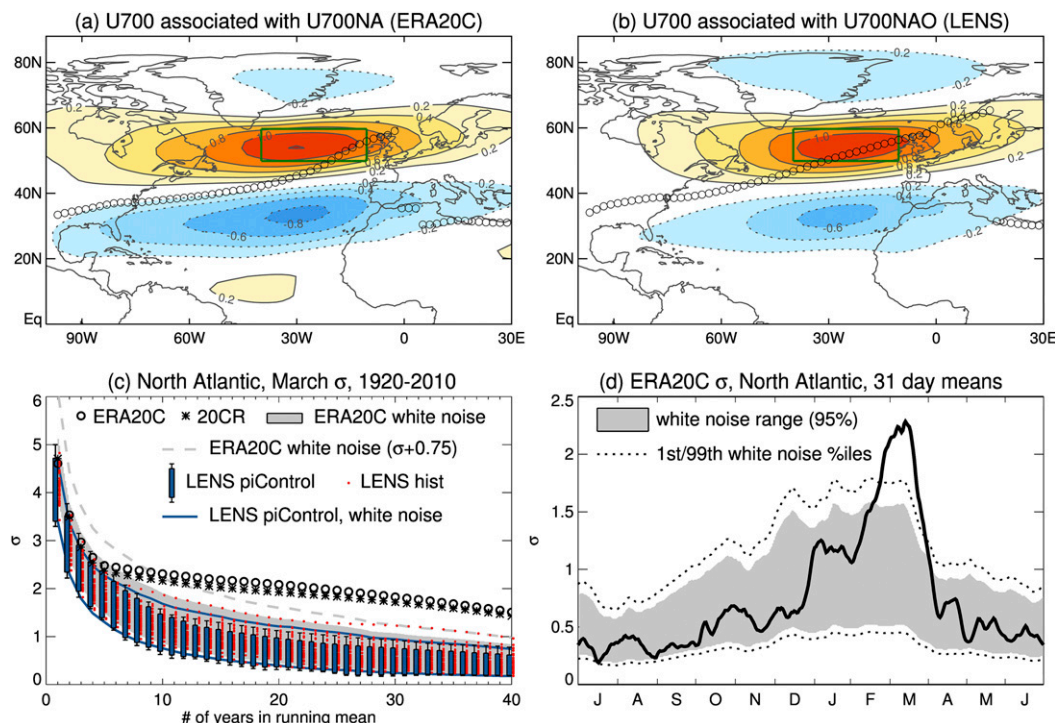
FIG. 4. (a) Interannual regression of March U700 onto March U700NA for ERA20C from 1900 to 2010. The green box shows the area used to define the U700NA region, and black points depict the local climatological jet maximum at each longitude determined by a quadratic fit to the maximum U700 on the 2° grid and the U700 at the two adjacent grid points. (b) As in (a), but for the ensemble mean of regressions calculated from the individual CESM LENS members. (c) The 20-yr running mean σ of U700NA as a function of the number of years in the running mean for March from 1920–2010. Data are shown for ERA20C (black circles) and the 20CR reanalysis (black asterisks); gray shading indicates the 95% confidence interval (2.5–97.5 percentile value range) on the 20-yr running mean σ for a white noise process with the ERA20C interannual standard deviation, the gray dashed line indicates the 97.5 percentile values for a white noise process with the ERA20C interannual standard deviation plus 0.75, the blue boxes and whiskers indicate the 2.5–97.5 percentile value range and minimum and maximum for overlapping 91-yr chunks of the LENS piControl, red dots indicate σ for each LENS historical member, and blue lines indicate the 2.5–97.5 percentile value range of a white noise process with the LENS piControl interannual standard deviation. (d) The σ of 20-yr running means of 31-day running means of U700NA for ERA20C from 1920 to 2010 along with the 95% confidence interval for the white noise distribution (gray shading) and the 1st and 99th percentiles of the white noise distribution (black dotted).

variability in U700NA is strongly correlated with SLP-based NAO indices, these NAO indices do not pick up the same degree of seasonality in the variability as this analysis of U700, nor do they pick up the same degree of discrepancy between the models and reanalyses since this U700 variability also depends on the finer-scale details in how the local SLP gradients change, not only the large-scale pressure patterns that characterize the NAO. This is discussed in more detail in appendix A.

The extent to which these conclusions are dependent on the length of the running mean considered can be assessed from Fig. 4c, which shows the standard deviations of March running means of a variety of lengths for U700NA. First, the CESM piControl and LENS simulations are very comparable in their behavior, as are ERA20C and 20CR. For anything longer than about an 8-yr running

mean, the reanalysis variability is greater than that seen in any of the 90-yr segments of CESM simulation considered. The interannual standard deviation (value of 1 on the x axis of Fig. 4c) of the reanalyses lies within the distribution of the estimates for CESM, but the range of CESM values suggest that the uncertainty on the observed estimate could be as much as ±0.75. Even if the true interannual standard deviation of the real world were equal to the observed standard deviation plus 0.75, beyond about a 10-yr running mean, the variability still exceeds the expectations from white noise sampling (gray dashed line in Fig. 4c).

The division into calendar months is somewhat arbitrary so the standard deviation of 20-yr running means of U700NA is shown for 31-day running means throughout the course of the year for ERA20C in Fig. 4d. This,
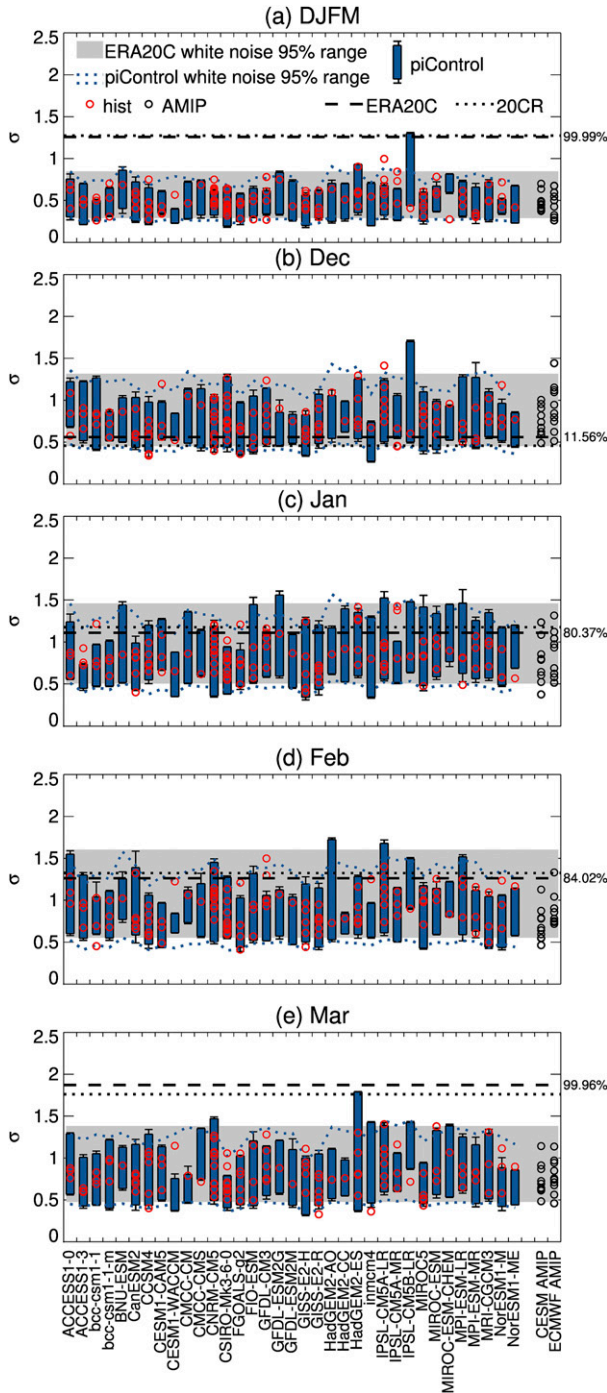
FIG. 5. Standard deviation of 20-yr running means of U700NA for 1900–2005 for (a) DJFM, (b) December, (c) January, (d) February, and (e) March. The blue boxes and whiskers indicate the 2.5–97.5 percentile value range and min and max of overlapping 106-yr chunks of the piControl simulations; red circles indicate historical members, the black dashed lines ERA20C, and black dotted lines 20CR. Gray shading indicates the 95% confidence interval for 106-yr chunks of white noise with the ERA20C interannual standard deviation. Blue dots are similar to gray shading, but use the interannual standard deviation for the piControl run of each
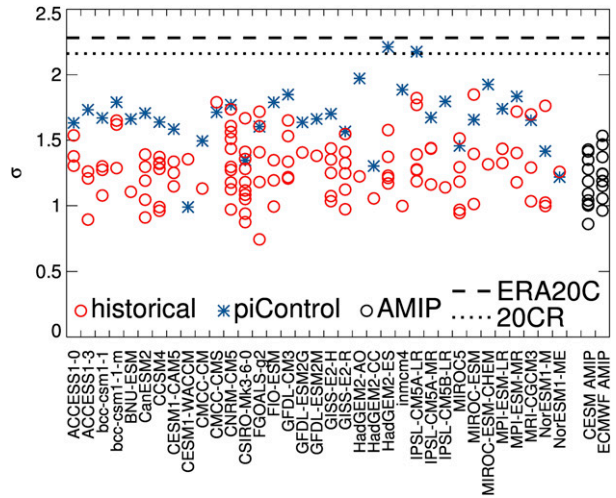

FIG. 6. Maximum standard deviation of 20-yr running means for March for 1900–2005, considering all grid points in the Atlantic domain (region plotted in Fig. 2). The black dashed (dotted) line indicates the maximum value for ERA20C (20CR). Blue points indicate the maximum value considering all overlapping 106-yr chunks of the preindustrial control simulations; red points indicate the maximum value for each historical member. Black circles in the final two columns indicate maximum values for simulations with prescribed historical SSTs.

again, demonstrates the greatly enhanced multidecadal variability in the late winter in the North Atlantic, reaching greater values than expected from white noise sampling only in late February and March. Since the chances of obtaining variability in 20-yr means as large as seen in March by sampling a white noise time series are less than 1 in 2500, then even if we account for the multiple (~12) tests that have been performed here, the chances of obtaining an extreme value like March are still less than 1 in 200.

So, CESM is very likely different from the reanalyses in terms of its representation of multidecadal variability in North Atlantic U700, but how do other models compare? The assessment of CMIP5 U700NA variability shown in Fig. 5 demonstrates that the CMIP5 models also sit within the expectations from white noise sampling (HadGEM2-ES in March and IPSL-CM5B-LR in the DJFM average are the only exceptions). This is also true of the CESM and ECMWF AMIP simulations (black points; i.e., those with prescribed historical

←

individual model. Black circles in last two columns indicate values for prescribed SST simulations. Percentages to the right indicate the percentile of the white noise distribution at which ERA20C lies.
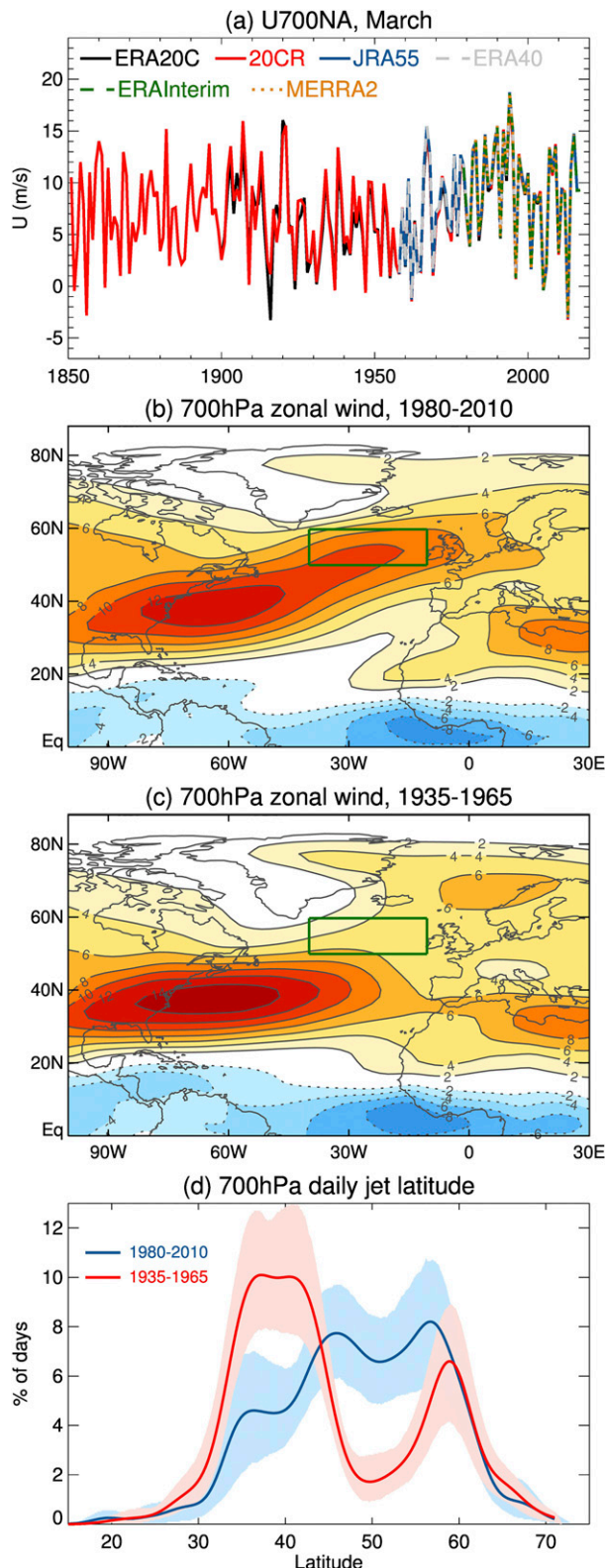
FIG. 7. (a) Time series of March U700NA (green box in Fig. 2e) for each reanalysis dataset. (b) March 700-hPa zonal wind for the 31-yr climatology from 1980 to 2010 using ERA20C. (c) As in (b), but

SSTs). The variability in ERA20C in March is roughly double the mean value for each model and 50% greater than the maximum value sampled taking all 106-yr segments for each model into consideration.

One remaining possibility is that perhaps the models are capable of simulating the observed magnitude of variability, but they just do not do it in exactly the same location. This possibility can be ruled out by an assessment of the maximum σ of 20-yr means of any grid point within the Atlantic domain (Fig. 6). For the vast majority of models, the maximum multidecadal variability seen in ERA20C is around 50% greater than the maximum variability that occurs anywhere in the North Atlantic sector. Only two models (HadGEM2-ES and IPSL-CM5A-LR) come close in a 106-yr segment of their piControl simulation.

To summarize, the multidecadal variability in U700 in the North Atlantic in CESM and in the vast majority of CMIP5 models is entirely consistent with the sampling of white noise interannual variability (i.e., there is no need to invoke a particular low-frequency cause behind the modeled multidecadal variability). The same is true for the early winter (December and January) of the observed record. In contrast, in the late winter (late February and March) the observed record displays multidecadal variability that is highly unlikely to have occurred as a result of the averaging of year-to-year variability indicating that possibility 1 outlined in the introduction for explaining the discrepancy between models and observations is extremely unlikely to be the only explanation.

4. The observed multidecadal variability in the North Atlantic in March

Before proceeding with an assessment of the likelihood of the remaining possibilities (2 and 3), it is worth briefly considering what this variability actually means for the climatology of the North Atlantic jet stream during March. The time series of March U700NA for the various different reanalyses are shown in Fig. 7a. First, even without the application of a running mean smoother,

←

from 1935 to 1965. (d) PDF of daily jet latitude based on 700-hPa zonal wind averaged from 60°W to 0° and applying a kernel smoothing to the PDF following Woollings et al. (2010). The PDFs are for March days after first performing the low-pass filtering so days prior to and after March are used for filtering. The shading depicts the 2.5–97.5th percentile value range determined by resampling the individual years within the time period considered, with replacement. The green boxes in (b) and (c) depict the area used to define the U700NA index for reference.

the multidecadal variability is clear with a notable minimum centered on the 1950s, rising to a maximum in the 1990s, consistent with the positive NAO trends observed in the wintertime average over this time period. Second, a comparison of the different reanalyses should allay any concerns over the fidelity of North Atlantic U700 in reanalyses that are only constrained by surface pressure observations. ERA20C and 20CR compare very well with each other over the period of overlap and they also compare well with the shorter, more constrained, reanalyses.

The structure of the North Atlantic jet stream at 700 hPa is shown for the 31-yr periods of 1980–2010 and 1935–65 in Figs. 7b and 7c, respectively. This makes clear that the structure of the observed March North Atlantic jet stream can vastly differ between 30-yr periods. Notably, the climatological flow over the United Kingdom in 1980–2010 is roughly double that of 1935–65 with related implications for storminess in that region.

Another metric that is commonly used to characterize the North Atlantic jet stream is the probability density function (PDF) of daily jet latitudes. Following a similar methodology to that of Woollings et al. (2010), this is obtained by first averaging the zonal wind from 0° to 60°W, then applying a 10-day low-pass Lanczos filter with 61 weights before identifying the latitude of maximum zonal wind between 15° and 75°N on a daily basis. The latitude of the maximum is obtained using a quadratic fit to the grid point with the maximum zonal wind and the two adjacent to it. The first three harmonics of the climatological seasonal cycle of jet latitude values are then subtracted before building the PDF of jet latitudes.

This PDF is shown, centered on the March climatological jet position for the two time periods (1980–2010 and 1935–65) in Fig. 7d. During 1980–2010, the PDF displays the three preferred jet latitude locations as discussed by Woollings et al. (2010) for DJF. However, for 1935–65 the PDF is drastically different with a distinct minimum in occupation around 50°N and a greatly increased preference for the occupation of latitudes between 35° and 45°N. Note that this is not at odds with the conclusions of Woollings et al. (2014), who found the trimodal structure of the PDF to be robust over the course of the twentieth century as they considered the DJF season (their Fig. 6). Figure 7d depicts March only and the DJF season for both periods exhibits the structure described by Woollings et al. (2014) (not shown). This variability in daily jet latitudes likely also corresponds to variability in the occurrence of blocking. As shown by Häkkinen et al. (2011) for the winter season as a whole, the 1950s and 1960s were characterized by more frequent North Atlantic blocking events compared to the later time period.

5. The relationship with SSTs and external forcings

The results of section 3 indicate that the chances that atmospheric noise has given rise to the observed variability in the late winter over the last century are extremely slim. We therefore proceed to search for evidence of connections with either multidecadal SST variability or external forcings to identify possible underlying causes of the excess variability.

a. The connection with SSTs

Figure 8a shows the correlation between 20-yr running means of SSTs globally and U700NA during March. Here ERA20C from 1900 to 2010 is combined with ERA-Interim from 2011 to 2017 and this is correlated with ERSSTv5, although results are essentially the same when using other combinations of datasets. Strengthened winds over the North Atlantic on these time scales are highly correlated with reduced SSTs in the subpolar North Atlantic and increased SSTs over the Southern Ocean. Even with the very few degrees of freedom that remain when taking a 20-yr running mean over a 118-yr period, locally the correlations in the subpolar North Atlantic and in regions of the Southern Ocean exceed the threshold for significance at the 5% level.

This global pattern of SST correlations is reminiscent of the SST variability that accompanies the TS AMV index on these time scales (Fig. 1c). Indeed, U700NA is strongly negatively correlated with the TS AMV index during March (Fig. 8b).[1] In fact, the correlation between North Atlantic zonal wind and the TS AMV is far greater in March than in any of the other winter months or the winter season as a whole (Fig. 8b) and this strong negative correlation holds if the period considered is extended back to 1854 using 20CR (Fig. 8c). So, the month of the winter that exhibits excess multidecadal variability in the North Atlantic winds (i.e., March) is the month for which the relationship between the AMV and the winds is by far the strongest.

This strong negative correlation with the TS AMV index is robust to the SST dataset used (Fig. 8d, left) and the LIMoptimal AMV index also produces correlations of a similar magnitude, although they narrowly fail to pass the significance thresholds (Fig. 8d, center). In contrast, when the linearly detrended AMV index is used, the correlation with North Atlantic winds is approximately halved (Fig. 8d, right). Given that the TS and LIMoptimal AMV indices are designed to isolate

---

[1] The AMV index here is calculated for the individual month or season, but the correlations are similar if the annual mean AMV index is used.
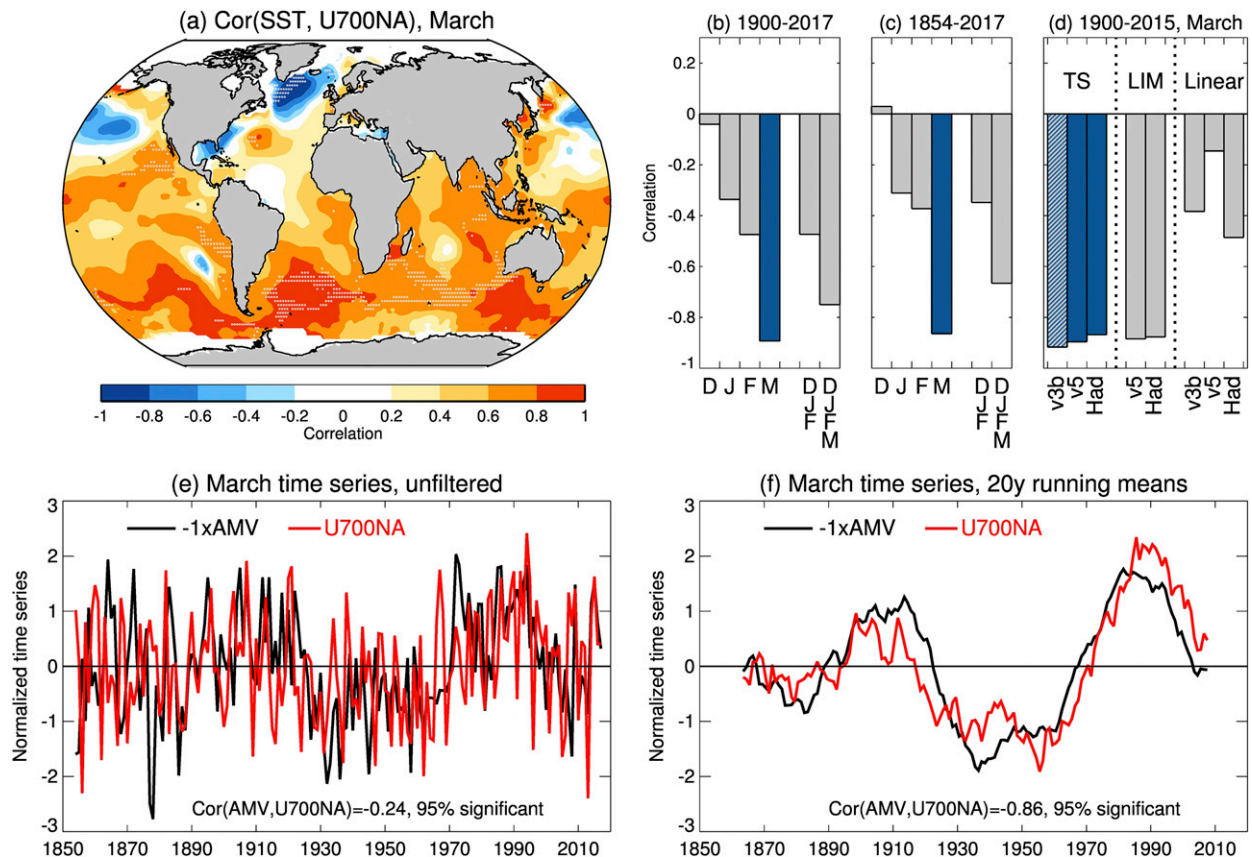
FIG. 8. The relationship between SSTs and U700NA (green box in Fig. 2e). (a) Correlation between SST (ERSSTv5) and U700NA (ERA20C + ERA-Interim, ERA-Interim after 2010) and 20-yr running means in March from 1900 to 2017. White stippled regions are significantly different from zero at the 5% level by a two-sided test although these regions do not pass a stricter test that accounts for spatial correlation by the false discovery rate method of Wilks (2016). Correlation between U700NA and AMV for 20-yr running means [blue hatched (solid) = significant at the 10% (5%) level by a two-sided test]: (b) correlations evaluated over 1900–2017 using the TS AMV index for ERSSTv5 and ERA20C + ERA-Interim (ERA-Interim after 2010); (c) as in (b), but for the full record from 1854 to 2017 using a combination of 20CR and ERA-Interim for U700NA (ERA-Interim after 2014); and (d) correlations for March evaluated from 1900 to 2015 using the ERA20C + ERA-Interim combination but for various different definitions of the AMV index and different SST datasets. (e) Time series of March zonal wind and −1 × AMV index for 20CR + ERA-Interim and the TS AMV index for ERSSTv5. (f) As in (e), but for 20-yr running means. All significance testing has been performed using the resampling strategy described in section 2d. In (b)–(d) the correlation has been performed using the AMV from that individual month or seasons except for the LIMoptimal index where the JFM averaged AMV has been used.

only the internal component, the multidecadal variability in March winds is, therefore, most strongly connected to the internal component of AMV.

Normalized time series of the unfiltered and 20-yr running mean TS AMV and North Atlantic winds are presented in Figs. 8e and 8f, extending back to 1854. There is a clear correspondence between them at low frequencies.

If we assume, for the moment, that this relationship represents an influence of SST variability on U700NA, then one possible reason why models fail to reproduce the degree of variability seen in the real world could be because they fail to simulate the magnitude of AMV that has occurred over the observational record, as shown by a

number of studies (Frankcombe et al. 2015; Murphy et al. 2017; Qasmi et al. 2017; Kim et al. 2018). But the CESM and ECMWF AMIP simulations similarly fail to reproduce the observed magnitude of U700NA variability (Figs. 5e and 6) and, much like the coupled models, are within the range expected from white noise sampling. Therefore, prescribing the historical evolution of SSTs does not solve the issue. Note that this was also the conclusion drawn by Kim et al. (2018) for NAO variability of the winter season as a whole in CESM.

Figure 9 shows the regression of zonal wind onto the TS AMV index for the observations; that is, we assume

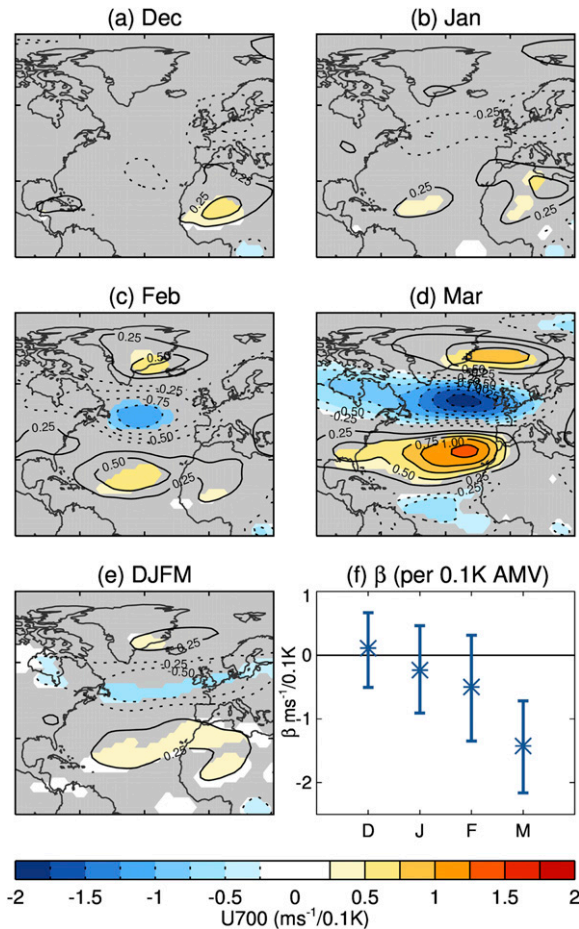$$\overline{U700NA} = \beta\overline{AMV} + \overline{\varepsilon}, \tag{1}$$

FIG. 9. (a)–(e) The regression of 20-yr running mean ERA20C U700 onto the TS ERSSTv5 AMV index (scaled by 0.1, which is approximately the standard deviation of the AMV in each month). Colored regions indicate where the regression coefficient is more than 50% larger than the maximum regression coefficient found across the 10 CESM AMIP members. The regions where the regression coefficient is large are also those where the correlation coefficient is high [most regions with a regression coefficient greater than $0.5\,\mathrm{m\,s}^{-1}\,(0.1\,\mathrm{K})^{-1}$ have a correlation coefficient greater than 0.8]. (f) The regression of 21-yr running mean U700NA onto AMV (scaled by 0.1 K) for each month. The 21-yr running means are used here such that the low-frequency time series is centered on a particular year. Error bars show the 2.5–97.5 percentile value range of bootstrapped estimates of the regression coefficients. These have been estimated by first dividing the U700NA and AMV time series into low-frequency (21-yr running means centered on a given year) and high-frequency (the actual value for that year minus the low-frequency time series) components. The high-frequency component is then resampled with replacement (preserving the relationship between U700NA and AMV) to produce a new high-frequency time series of equivalent length to the full record. The 21-yr running mean of this high-frequency time series is added to the low-frequency time series and the regression coefficient is recalculated. This is repeated 1000 times and the error bars show the 2.5–97.5 percentile value range of these bootstrapped regression coefficients. Note that (a)–(e) use the time series from 1900 to 2009 for comparison with the CESM AMIP experiments while (f) makes use of 1900–2017, appending ERA-Interim to ERA20C after 2010.

where $\overline{(\ )}$ represents a 20-yr mean, and ε represents noise that is unrelated to AMV. Figure 9 shows β, scaled by 0.1 K, which is approximately the standard deviation of $\overline{\mathrm{AMV}}$ in each month. Colored regions in Fig. 9 indicate where the magnitude of β is more than 50% greater than the maximum regression coefficient found across the 10 CESM AMIP members. In March, the observed β is more than 50% larger than the maximum found in the AMIP members over much of the North Atlantic (Fig. 9d). In fact, there is a striking correspondence across the winter season between the regions where U700 is strongly related to the AMV and those where the observations exhibit enhanced multidecadal variability, both relative to white noise expectations and relative to CESM. This begins with the regions over Europe and North Africa in the early winter (cf. Figs. 9a and 2l) and evolves to the regions over the North Atlantic Ocean in March, and to a lesser extent February (cf. Figs. 9c,d with Figs. 2n,o). This indicates a strong possibility that these regions are being influenced by SST variations that accompany the AMV index and that models (at least CESM and the ECMWF model) do not sufficiently respond to observed SSTs.

We can further assess whether the relationship between U700 and AMV is actually different in March from the other winter months by considering the uncertainty in β. The β values calculated from the regression of U700NA onto AMV (scaled by 0.1 K) are shown in Fig. 9f along with the 2.5–97.5 percentile value ranges. These uncertainty ranges are estimated using a bootstrapping with replacement methodology as described in the figure caption. The probability that the December, January, and February β values are equivalent to the March β is $p = 1.6 \times 10^{-5}$, 6.4×10^{-3}, and 0.141, respectively, where these p values represent the probability that the March value of β and the value of β for the other month fall within the region where their confidence intervals overlap. We can, therefore, be confident that the relationship between AMV and U700NA is different in December and January from that in March, but it is possible that a similar relationship exists in February and March with sampling of noise leading to the differences seen between them.

b. The connection with external forcings

SST variability has the potential to give rise to the multidecadal North Atlantic U700 variability in March, but is there also a potential role for external forcings that could either act directly on the North Atlantic winds or indirectly through their influence on the SSTs? The fact that U700 is only strongly connected with the internally generated component of AMV is already suggestive that
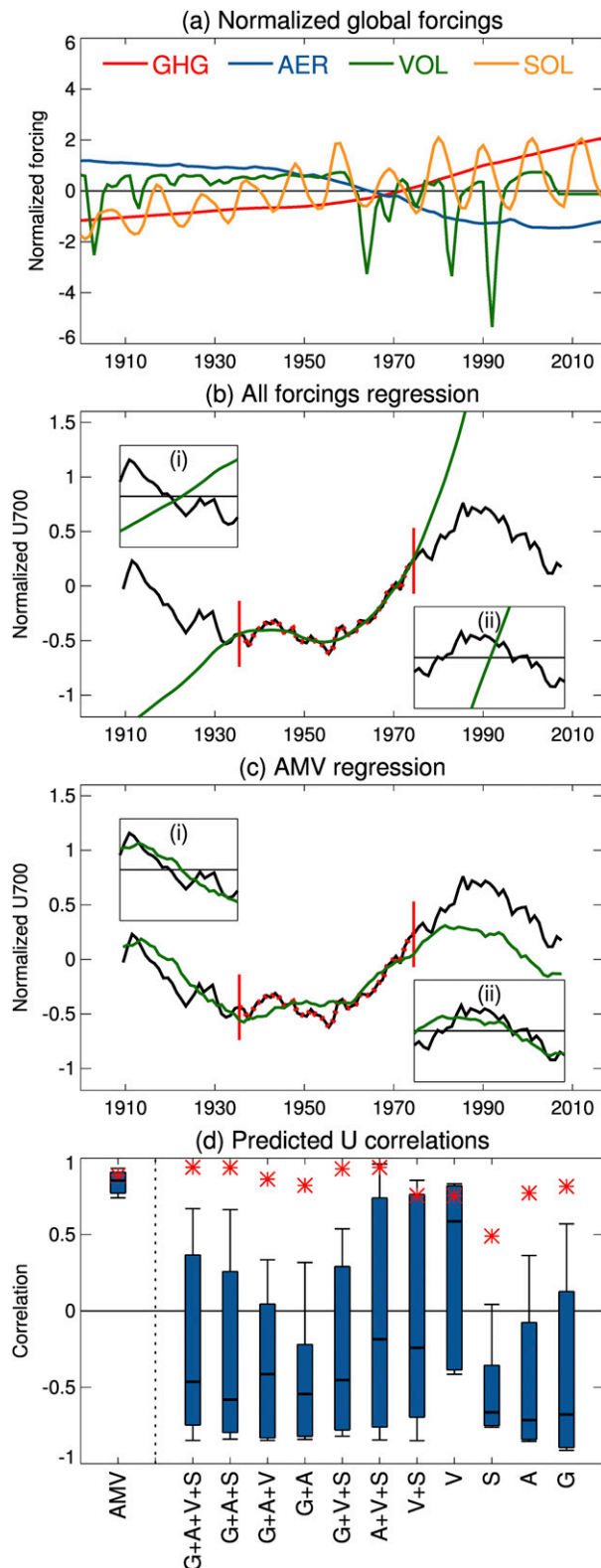
(a) Normalized global forcings

this is not the case, but here we also examine whether combinations of GHG, aerosol, volcanic, or solar forcings have the potential to be drivers of the U700 variability through a multiple linear regression approach. The forcing time series used are the global, annual average radiative forcings due to total greenhouse gas forcing (GHG), total direct aerosol forcing (AER), volcanic stratospheric aerosol forcing (VOL) and total solar irradiance (SOL) provided by Otto et al. (2015). Normalized versions of these time series with zero mean and unit standard deviation are shown in Fig. 10a.

We assess to what extent a linear regression model of 20-yr running mean U700NA fit to 20-yr running means of these forcing time series can explain the variability in March U700NA between 1900 and 2017. If all four predictors are included and the full record length is used to fit the regression model, then the U700NA predicted by this model is correlated at 0.94 with the actual U700NA (red asterisk in the second column of Fig. 10d). This actually slightly exceeds the correlation with U700NA predicted by a linear regression model fit to the AMV (0.89; red asterisk in the first column of Fig. 10d). However, the limited degrees of freedom available with 20-yr running means of a 118-yr time series mean that overfitting with the regression model when multiple predictors are used is very likely.

To avoid this, we take a cross-validation approach by assessing whether the regression model derived by fitting to a portion of the time series can explain the variations seen in the remainder. A demonstration of this approach is shown in Fig. 10b for the regression model fit to all four forcings. In this example, the regression model is fit only using U700NA between 20-yr means centered on 1935 and 1975 (red dotted portion) and below we summarize the results for similar fits over all overlapping 40-yr segments. The parameters obtained from the fit to this portion are then used to predict

←

taken from Otto et al. (2015). (b) A demonstration of the cross-validation procedure used to assess the effectiveness of a given regression model; in this case all forcings in (a) are included as predictors. Shown are the 20-yr running means of North Atlantic U700 (black) and the predicted U700 from the regression fit to the dotted red section (green). Insets i and ii depict the anomalies from the mean within the segments not used to fit the regression model. (c) As in (b), but using the ERSSTv5 TS AMV as the predictor. (d) Box and whiskers indicate the 10th–90th percentile range, min, max, and median of the correlation coefficients obtained from the cross-validation procedure using the ERSSTv5 TS AMV and various forcing combinations (G = GHG, A = AER, V = VOL, S = SOL). Red points show the correlation between predicted and actual U700 when the full time series are used to fit the regression model.

FIG. 10. An assessment of the ability of global forcings and the AMV to predict U700 in the North Atlantic. (a) Forcing time series, normalized to have zero mean and unit standard deviation,

U700NA in the remainder of the time series. To assess the ability of the regression model to do this, the mean from the time series within each segment is first subtracted (insets i and ii of Fig. 10b) to remove any built-in correlation that would arise from the long term variations over the time period used for the fit. The time series from segments i and ii are then combined, and the Pearson correlation coefficient between the predicted U700NA and actual U700NA for the combination of segments i and ii is calculated. In the example shown in Fig. 10b, the regression fit over the red dotted segment fails dramatically to predict the variations in the remainder of the time series giving a correlation of −0.32. In contrast, an equivalent example for the fit to the AMV does a reasonable job, giving a correlation of 0.8 (Fig. 10c).

Repeating this procedure for fits over all 59 overlapping segments of 40-yr length gives the 10th/90th percentiles and min/max of the correlation coefficients shown in Fig. 10d. The regression model is considered to fail when the 10th–90th percentile range is below, or encompasses, zero; that is, when more than 10% of the time, the predicted U700NA is negatively correlated with the actual U700NA. All regression models fit to combinations of external forcings fail this test. Even the median correlation typically lies below zero. In contrast, the regression onto the AMV passes convincingly.[2] This lends additional support to the conclusion that the multidecadal U700NA variability is most strongly related to the internally generated AMV.

c. Comments on ocean–atmosphere causality

While the above analysis leads us to conclude that the multidecadal variability in March U700 is most strongly connected with the AMV, some ambiguity remains over the causal nature of this connection. It is well known that variability in the North Atlantic atmospheric circulation and associated surface fluxes is, itself, a driving force for North Atlantic SST variability (Deser and Blackmon 1993; Deser et al. 2010), whether it be through the influence on the deep ocean circulation (Eden and Jung 2001; Danabasoglu et al. 2014; Yeager and Danabasoglu 2014; Delworth and Zeng 2016; Delworth et al. 2017) or direct influence on the ocean mixed layer (Seager et al. 2000; Clement et al. 2015; Cane et al. 2017). It is, therefore, plausible that these connections indicate a causal link in the opposite sense (i.e., the multidecadal

variability in the winds is driving the multidecadal variability in the SSTs). While the coupling between ocean and atmosphere likely goes both ways, we summarize here various lines of reasoning that support a directed causal link between the SST variability and March U700 in the sense of the SSTs driving March U700.

First, if SSTs do not explain the excess March U700 variability, then either an extremely unlikely occurrence by chance or external forcings would have to be invoked. External forcings are unlikely given that 1) the winds are most strongly connected to the AMV indices that are designed to isolate the internal component and 2) the regression analysis in the previous section fails to provide a link with external forcings.

Our remaining lines of reasoning are arguments for why the correlation between AMV and U700 does not represent a connection in the sense of U700 driving the AMV and, therefore, more likely represents a connection in the sense of the AMV influencing U700. First, consider the spatial structure of the SST anomalies that accompany the March U700 variability. On short time scales, NAO variability gives rise to a tripole pattern of SST anomalies through the direct influence of altered surface fluxes on the ocean mixed layer (Seager et al. 2000; Deser et al. 2010; Delworth et al. 2017). The high-frequency variability in U700NA, which is strongly connected to the NAO (see appendix A), is similarly associated with a tripole pattern (Fig. 11a). The near-surface wind anomalies that accompany a $1 \, \mathrm{m \, s^{-1}}$ increase in U700NA are very similar on low and high frequencies (cf. vectors in Figs. 11a and 11b). Therefore, if the low-frequency SST variations that accompany U700NA are a direct result of the influence of the wind variability on the ocean mixed layer, then we should expect the regression of low-frequency SST onto low-frequency U700NA to exhibit the same tripolar structure as seen at high frequencies. While this is the case for CESM (cf. Figs. 11c and 11d), it is not the case for the observations (cf. Figs. 11a and 11b). In observations, the SST anomalies that accompany the 20-yr running mean U700NA variability show a distinct pattern with one center in the subpolar gyre and oppositely signed anomalies elsewhere (except along the southeast United States). Delworth et al. (2017) recently argued that a structure, similar to this, is a signature of an influence of the ocean circulation. This would then imply that the March winds are not driving the multidecadal AMV through their direct influence on the ocean mixed layer alone.

The possibility remains that the March winds are a key driver of the accompanying SST anomalies through their influence on ocean heat transport. But, if that were the case, then there should be a lagged relationship between the wind variability and the SSTs given the long time

---

[2] The forcing regression also fails if the segment length used for the fit is increased while the AMV regression continues to pass, although the AMV regression does also start to fail if segments shorter than about 40 years are used.
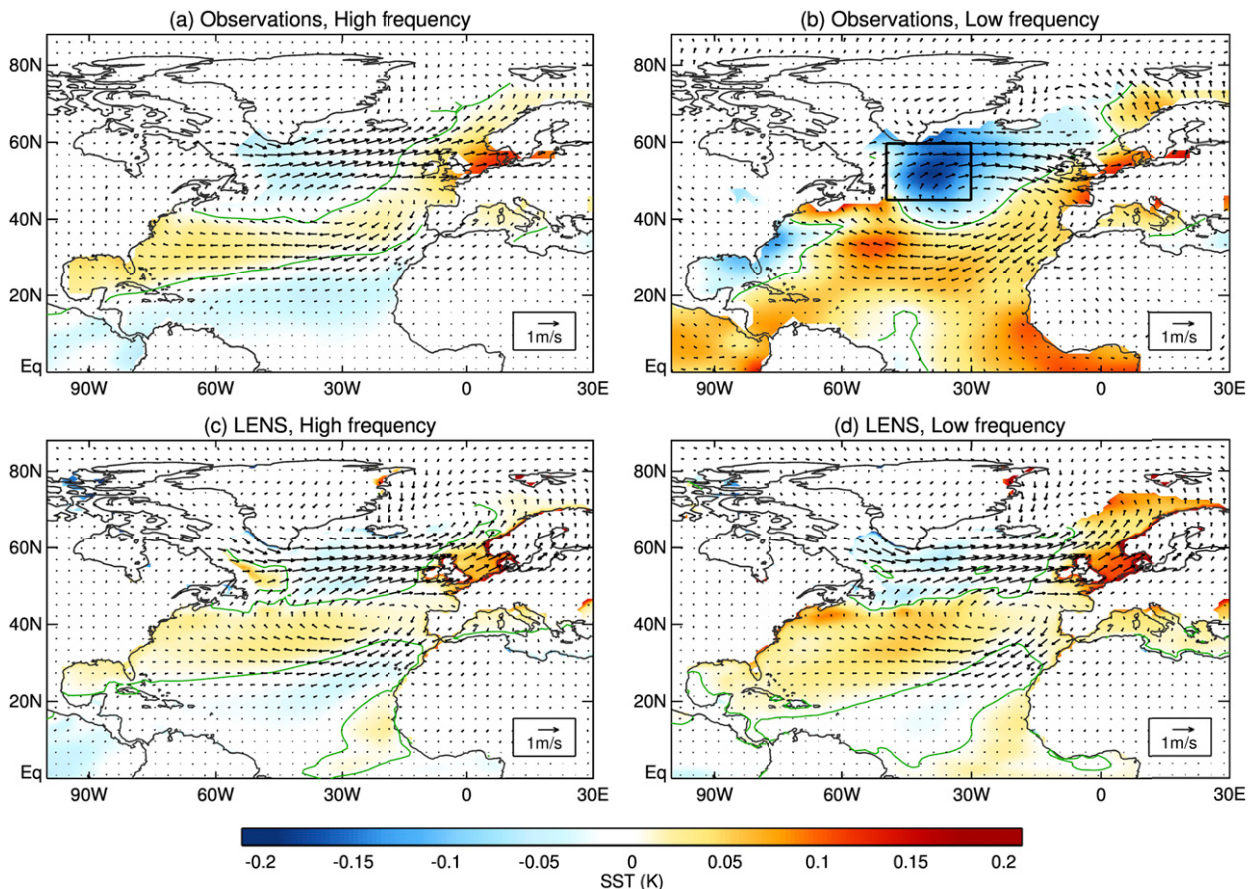
FIG. 11. The regression of SSTs (shading) and near-surface winds (vectors) onto U700NA (box in Fig. 2e) for March. (a),(b) Regressions for the observations (SSTs = ERSSTv5, winds = ERA20C + ERA-Interim 10-m winds, 1900–2017); (c),(d) the mean regression for the LENS historical ensemble (lowest model level winds are used). Note that (a) and (c) show the regressions on high frequencies and (b) and (d) show the regressions on low frequencies. Low frequencies are defined as 21-yr running means centered on a given year while high frequencies are the actual value minus the low-frequency value for that given year. The zero contour for SST is shown in green. The boxed region in (b) depicts the area used for the SST time series shown in Fig. 13.

scales inherent to the ocean circulation. Indeed, Delworth et al. (2017) showed that this SST pattern lags NAO variability in the observed record by at least 20 years (their Fig. 2h). This brings us to our third line of reasoning based on lead–lagged correlations between U700NA and AMV. Previous studies have inferred a driving of the AMV by the NAO through the use of lead–lagged correlations between low-frequency time series of the wintertime averaged indices (Peings et al. 2016; Delworth et al. 2017; Kim et al. 2018). Unsurprisingly, a similar lead–lagged relationship occurs between the wintertime averaged (DJFM) U700NA and the AMV (black lines in Fig. 12b). A positive AMV index (i.e., a warmer subpolar North Atlantic) is preceded by a positive anomaly in U700NA about 20–30 years earlier.

However, if there is an instantaneous connection between the AMV and U700NA, then given the strong autocorrelation of the AMV (Fig. 12a), a lagged correlation between U700NA and the AMV would be expected due to their instantaneous connection alone. We account for this possibility by removing this built-in correlation, that is, the lag-zero correlation between AMV and U700NA multiplied by the lagged autocorrelation of AMV (Fig. 12c), from the cross correlation between U700 and AMV, giving the red lines in Fig. 12b. For DJFM this shifts the lag of the maximum correlation slightly but the conclusion remains that a positive AMV index is preceded by a positive anomaly in U700NA. It should be noted that this same lead–lagged relationship is present, although noisier, for 10-yr running means (dotted lines in Fig. 12), and so it is not an artifact of the filter used (Cane et al. 2017). These correlations are not large enough to pass the 10% threshold for significance, but previous studies (e.g., Eden and Jung 2001; Delworth et al. 2017) have explicitly shown that the
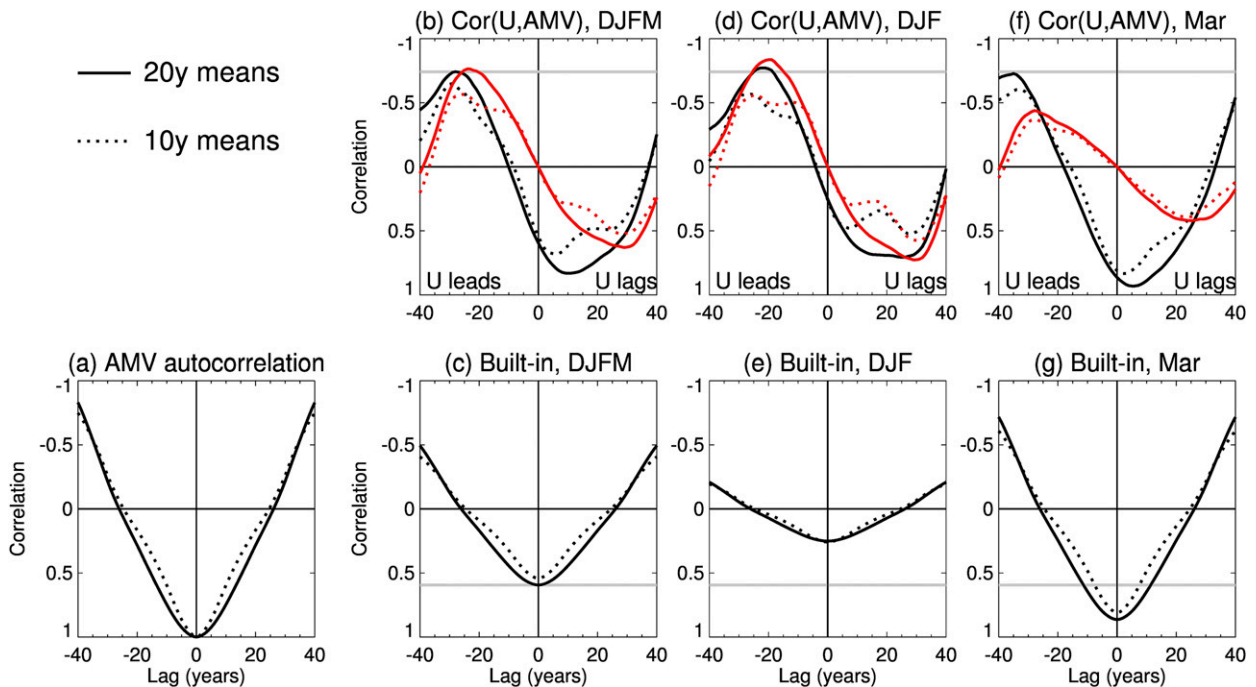
FIG. 12. The 20-yr running means (solid) and 10-yr running means (dotted). (a) The autocorrelation of the AMV index [note the flipped vertical axis for ease of comparison with (c), (e), and (g)]. (b) The lagged correlation between U700NA and the AMV index for DJFM (black) and the same but after subtraction of the built-in correlation shown in (c) (red). (c) The built-in correlation for DJFM obtained from the lag-0 correlation between U700NA and the AMV index multiplied by the AMV autocorrelation. (d),(e) And in (b) and (d), but for DJF. (f),(g) As in (b) and (d), but for March. In (b)–(g) the gray line shows a reference value from the DJFM panel, 20-yr running means, showing the maximum of the black solid line in (b), (d), and (f) and the minimum of the black solid line in (c), (e), and (g). The AMV used in all panels is the DJFM average TS index for ERSSTv5, and U700NA is from ERA20C + ERAInterim (1900–2017, ERA-Interim after 2010).

ocean circulation and the AMV respond to prescribed NAO surface fluxes with a similar lead–lagged relationship, suggesting they are physically meaningful.

Figures 12d and 12e are equivalent to Figs. 12b and 12c but using DJF U700NA while Figs. 12f and 12g use March U700NA only. These show that the March winds are not key to producing this lead–lagged relationship between the DJFM U700NA and AMV. In fact, a stronger correlation at negative lags is found when only the DJF winds are considered and, while a lead–lagged relationship with the March winds is present, it is relatively weak once the built-in correlation from their instantaneous connection is removed. The March winds, therefore, do not appear to be a critical component of the wind variability that precedes the AMV. Rather, based on the above reasoning, it is likely that the strong instantaneous negative correlation between March winds and the AMV (Fig. 12f) is evidence of an influence of AMV on the winds and that, while the March winds may contribute in return to the driving of the AMV, the other winter months likely dominate.

Our final, and perhaps most compelling, line of reasoning comes from the behavior of the CESM initialized decadal predictions. When initialized with observation-based ocean and sea ice conditions, CESM can, to a large extent, predict the behavior of SSTs in the region south of Greenland (i.e., the region that exhibits the strongest correlation with U700NA) a decade ahead (Fig. 13a). This prediction skill has already been discussed with an earlier model version by Yeager et al. (2012) and for this ensemble by Yeager et al. (2018). While the full twentieth century cannot be analyzed in these simulations, it is clear that over the latter half of the century, the initialized simulations can capture the low-frequency evolution of the SSTs south of Greenland, beginning with the decrease toward the late 1980s and the subsequent increase toward 2010. The free-running LENS simulations, on the other hand, do not capture this variability (Fig. 13b), so it not arising from the predictive power of the external forcings. Given that typical decay time scales of SST anomalies in this region are on the order of two years (Deser et al. 2003), if the winds are responsible for this low-frequency evolution of the SSTs, then the low-frequency behavior of the winds would have to also be predicted by the decadal prediction ensemble. This is not the case (Fig. 13c). Even
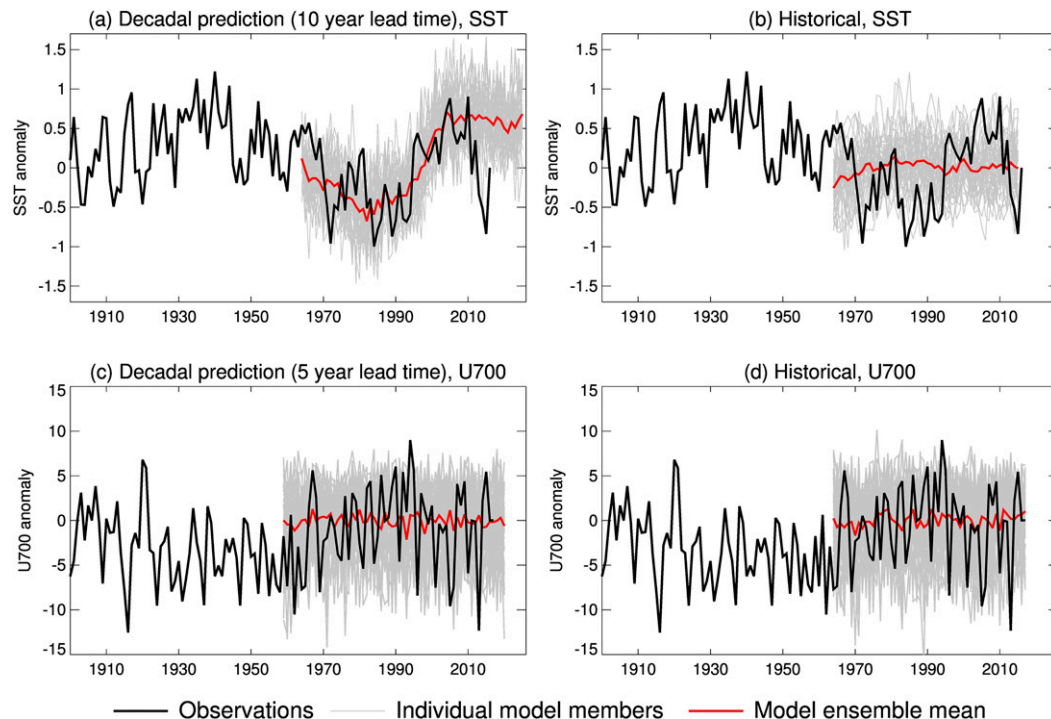
FIG. 13. (a) Time series of DJFM averaged SST anomalies in the North Atlantic (boxed region in Fig. 11b), black is ERSSTv5, gray is the prediction from each individual decadal prediction ensemble member that was initialized 10 years earlier, and red is the ensemble mean of the gray lines. (b) As in (a), but the model values are from the free-running CESM LENS historical simulations. (c),(d) As in (a) and (b), but for U700NA in the North Atlantic box using ERA20C + ERA-Interim (ERA-Interim after 2010) as the observational dataset and using only a 5-yr lead time for the decadal prediction ensemble. All time series are normalized by the mean over the 1964–2016 period. For the decadal prediction simulations in (a) [(c)] the values used are the anomalies of that 10th (5th) year prediction from the climatology of all 10th (5th) year predictions.

with a lead time of only 5 years, the decadal prediction ensemble fails to predict the low-frequency evolution of the winds (the same is true of longer lead times too). The predictability of the SSTs, instead, lies in the initialization of the ocean circulation and the associated skill in the prediction of ocean heat transports. This provides further evidence that the instantaneous connection between March U700NA and the SSTs south of Greenland does not represent a direct driving of the SSTs by the winds and that variations in ocean heat transport are an integral component of this low-frequency SST variability. Without a need to invoke the March U700 variability to explain the SST anomalies, we have no reason to expect them to be so highly correlated unless something about the SST anomalies is driving variability in U700.

6. Discussion and conclusions

The analysis presented in section 3 indicates that observed multidecadal variability in North Atlantic zonal wind displays a pronounced seasonality with strongly enhanced variability in the late winter, specifically late

February and March. Models fail to exhibit this degree of late winter variability, and they almost all exhibit variability that is entirely consistent with the sampling of uncorrelated year-to-year fluctuations of Gaussian white noise. The first of the three possible reasons for this deficiency laid out in the introduction was that the observed variability was the random chance sampling of internal year-to-year atmospheric variability, with no need to invoke an underlying low-frequency cause, and that either models are deficient in their year-to-year atmospheric variability or the observed sequence of variability was unlikely. While it is possible that the sampling of atmospheric noise has contributed to the variability that has been observed, the results demonstrate that the observed and modeled interannual variability are comparable and that the observed multidecadal variability was exceedingly unlikely to have occurred through the chance sampling of atmospheric noise alone, motivating the search for an underlying cause.

There are two main categories of low-frequency forcing that could explain the variability in question. One is forcing that is external to the ocean–atmosphere

system (e.g., anthropogenic forcings) and the other is internal ocean–atmosphere coupled variability, whereby the long time scales inherent to the ocean act as a low-frequency forcing on the atmospheric circulation. It is found that the variability in March zonal winds is strongly correlated with a global pattern of SST variability that is characterized by anomalies in the subpolar North Atlantic with oppositely signed anomalies in the Southern Ocean. This is the pattern that is represented by traditional AMV indices on these time scales (Fig. 1c) and indicates that low-frequency variability in the ocean does have the potential to explain the low-frequency variability in North Atlantic winds. If so, then given that simulations with prescribed historical SSTs were shown to also be deficient in their North Atlantic jet stream variability, it suggests that models may be deficient in their response to the SSTs.

The conventional view of North Atlantic ocean–atmosphere coupling, however, is that the primary direction of interaction is the opposite of this, with the winds influencing the SSTs. Evidence for a connection in the other direction, during winter, is somewhat patchy. Consistent with the results of the present study, Ting et al. (2014) and Gastineau and Frankignoul (2015) have both shown in reanalyses of the twentieth century that in the winter seasonal average, the positive phase of AMV is associated with a negative NAO signal; Gastineau and Frankignoul (2015) further argued based on lead–lagged relationships that this connection represents a driving of the NAO by the underlying SST variability. Other studies have investigated the impact of the decadal time scale AMV (Fig. 1b) on the North Atlantic atmospheric circulation using models with prescribed SST anomalies (Sutton and Hodson 2007; Hodson et al. 2010; Peings and Magnusdottir 2014; Omrani et al. 2014; Davini et al. 2015; Peings and Magnusdottir 2016) or, more recently, with nudging methodologies in a coupled framework (Ruprich-Robert et al. 2017). The multimodel study of Hodson et al. (2010) found no significant influence on the North Atlantic atmospheric circulation during winter but the studies since then have generally argued for a response in the sense of positive AMV being accompanied by a negative NAO. This response is, however, generally smaller than what would be inferred from the observational record over the central Atlantic [e.g., compare Peings and Magnusdottir's (2014) Fig. 2a with their supplemental Fig. 6]. In addition, the majority of studies focus on the winter season as a whole and the above analysis suggests these connections may have a rather strong seasonality within the winter months. Peings and Magnusdottir (2014) is one exception, where they did find a stronger response in the late winter, but their subsequent analysis with a different model (Peings

and Magnusdottir 2016) showed the opposite seasonality. If our hypothesis that models may be deficient in their response to SST variability in late winter is correct, then it is not surprising that the response found in modeling studies is generally weak or inconsistent.

However, given the strong evidence for atmospheric forcing of North Atlantic SSTs, one should rightly be cautious of ascribing a causal link in the sense of the SSTs influencing the North Atlantic winds from their strong correlation alone. We take a number of lines of reasoning to argue that this correlation does indeed represent an influence of the SST variability on the winds as opposed to the other way round. First, it is difficult to explain the wind variability through external forcings as demonstrated by the regression analysis in section 5b. Admittedly this approach was limited by making use of only a linear regression model and only the global forcing time series of greenhouse gases, aerosols, volcanic stratospheric aerosol, and solar variability. It is conceivable that there are nonlinearities or additional external forcings that could play a role, or more regional variations in external forcings that might have a different phasing through the century and these possibilities have not been considered here. But, a further line of evidence against a role for external forcings is that the zonal wind variability is most strongly connected with the internally generated component of SST variability, as determined by the LIMoptimal method of Frankignoul et al. (2017), suggesting external forcings do not dominate. This leaves only an influence of the SSTs or an extremely unlikely chance sampling of atmospheric variability as explanations.

The remaining lines of reasoning argue against the correlation with SSTs representing an influence of the winds on the SSTs and, therefore, there is no reason to expect such a correlation to occur unless the SSTs are influencing the winds. This includes the fact that the structure of the SST anomalies that accompany the variability is not what would be expected to result from the direct forcing by surface fluxes, which is suggestive of a role for heat transport by the deeper ocean circulation in the SST anomalies of importance. Then, if the March atmospheric variability were driving changes in the deeper ocean circulation, a lagged relationship would be expected rather than the instantaneous one that is found. Finally, a driving of the SSTs by the March winds is extremely difficult to reconcile with the CESM decadal predictions, which can predict the low-frequency evolution of SSTs in the region of relevance up to a decade in advance without predicting the low-frequency behavior of the winds themselves.

Each of these factors leads us to conjecture that the strong correlation between U700 and SSTs represents a driving of U700 by the SSTs and that possibility 3 laid out in the beginning of the paper is likely playing an important role. That is, the excess multidecadal variability in observed March

North Atlantic winds owes its existence to coupled ocean–atmosphere processes and the fact that models are deficient in their SST-forced wind variability. It is then possible that the March wind variability feeds back onto the North Atlantic SSTs, which could go some way to explaining why modeled AMV typically exhibits a shorter time scale (Peings et al. 2016; Ting et al. 2011) and weaker amplitude (Frankcombe et al. 2015; Murphy et al. 2017; Qasmi et al. 2017; Kim et al. 2018) than observed.

This argument would be made all the more convincing if a mechanistic understanding of the SST impacts on the winds could be obtained. This understanding currently eludes us given that it cannot be investigated in the current generation of models and achieving a mechanistic understanding through observations alone requires analysis dating back to the 1950s to capture a sufficient magnitude of the variability. This entails observations made before the satellite era when we may be concerned about the fidelity of aspects that may be important forcings of U700 variability (e.g., the divergent circulation or transient eddy fluxes), although this is a current topic of investigation.

It also remains to be understood why the seasonality in the relationship between AMV and U700NA exists. There are only small differences in the 1900–2017 climatological jet structure between the winter months, with the jet being slightly stronger and farther poleward in the early winter compared to the late winter. It is possible that these small differences somehow make the jet stream more responsive to AMV. Another possibility is that the AMV is forcing U700 anomalies throughout the winter but there is some time scale over which the anomalies grow, reaching a maximum in the later winter. There is some indication of the signal being present in January and growing in the subsequent months from Fig. 9, although without a more detailed understanding of the mechanisms involved we can only speculate that these could be reasons for the seasonality.

It has been argued that North Atlantic SST variability may influence the North Atlantic jet through a stratospheric pathway (Omrani et al. 2014) and we have investigated this possibility. In appendix B, evidence against this for an explanation of the March variability considered here is discussed. Other possible mechanisms for an AMV influence on the jet stream include forcing from the tropical Atlantic (Davini et al. 2015; Peings et al. 2016) or the local influence from the North Atlantic SSTs on, for example, the forcing of stationary waves or baroclinicity and eddy–mean flow interactions (Kushnir 1994; Msadek et al. 2011; Gastineau and Frankignoul 2012; Peings et al. 2016), or it could even arise from forcing from more remote regions such as the tropical Pacific, or changes in the tropical circulation that arise from the altered interhemispheric temperature gradients induced by the global SST pattern that accompanies the AMV (Fig. 1c).

In general, it is thought that the influence of local North Atlantic SST anomalies on the atmospheric circulation is small (Kushnir et al. 2002, and references therein). But this conclusion has mostly been drawn from model-based analyses and, given their apparent deficiencies, it is possible that the real world atmosphere exhibits a greater response to extratropical SST anomalies. Indeed, recent studies have suggested that improved resolution of ocean fronts and the overlying atmosphere could significantly alter the nature of ocean–atmosphere coupling (Smirnov et al. 2015; Parfitt et al. 2016; Siqueira and Kirtman 2016). A coupled simulation with CESM with horizontal resolution of 1/4° in the atmosphere and 1/10° in the ocean does not show improved multidecadal variability in the North Atlantic jet stream (not shown) but this simulation also exhibits a greatly reduced amplitude of AMV compared to observations. It has yet to be seen whether a high-resolution atmospheric model subject to realistic AMV SST variability exhibits an atmospheric response like that inferred from observations.

The coming decade or two of observational data may prove critical to fully understanding this issue. If the low-frequency behaviors of March winds and North Atlantic SSTs continue to evolve in step, our confidence in their connection will be reinforced. Furthermore, if the CESM decadal predictions prove correct (Fig. 13a), over the coming decade we should continue to see elevated SSTs in the North Atlantic and, from our analysis, we would predict that this would be accompanied by a more zonal jet stream in March with considerably reduced westerlies in the region west of the United Kingdom (as in Fig. 7c). This may give us a period of time that is analogous to the 1940s and 1950s but with sufficient data coverage to fully diagnose the mechanisms behind the SST–March winds connection in the observational record.

APPENDIX A

Comparison with SLP-Based NAO Indices

Here we clarify to what extent the U700 variability is connected with the SLP-based NAO: both the station-based (ST) and the empirical orthogonal function (EOF)-based indices (Hurrell 1995). The station-based NAO index is defined as the SLP difference between Lisbon, Portugal, and Reykjavik, Iceland. The EOF-based NAO index is the principal component (PC) time series of the first EOF of deseasonalized monthly SLP over 20°–80°N and 270°–40°E (cosine weighted). The EOF is calculated using all months in the DJFM season and the PC time series is normalized such that it has unit standard deviation for all months and years combined. These NAO indices have been a commonly used measure of Atlantic jet stream variability in previous studies (e.g., Kravtsov 2017; Wang et al. 2017; Kim et al. 2018).

The 20-yr running means of the NAO and the U700NA index used in the main body of the text are highly correlated during January, February, and March, but not during December (Fig. A1a). This is because the U700 variability that accompanies the NAO on these time scales differs from month to month (not shown, but can be inferred from the SLP regression in Figs. A1h–k). It follows that the NAO is also significantly correlated with the AMV during March (correlation = −0.91 for ST, −0.88 for EOF), but much less so during the other winter months (Fig. A1b).

While, during March, our U700NA index is correlated with the SLP-based NAO indices, a slightly different conclusion as to the seasonality of the variability and the extent to which models differ from observations would be obtained using the SLP-based NAO indices. Fig. A1c shows the σ of 20-yr running means of the ST NAO index for each month of the year, for ERA20C (black points) and the LENS historical members (red points) as well as the EOF-based NAO index for ERA20C only in December, January, February, and March (blue points). March does not stand out as so unusual when the NAO indices are considered. In fact, for the EOF-based index, January and February exhibit more variability. That

being said, it is only during March that the ST NAO variability exceeds the expectations from white noise sampling.

Insights into the reason behind this difference can be obtained by considering the variability in 20-yr running means of SLP over the Atlantic domain (Figs. A1d–g). The variability over Reykjavik and Lisbon (green points) does not differ too much among January, February, and March, consistent with their similar ST-based NAO variability (Fig. A1c). However, March differs from the other months by having greatly enhanced SLP variability more localized over the midlatitude North Atlantic Ocean in a region that would not be picked up by the ST-based NAO index. Similarly, on these low frequencies the regression of SLP anomalies onto the EOF-based NAO exhibits different structures across the season (Figs. A1h–k). While the NAO pattern is, by construction, the same for each month, the regression of SLP onto the NAO index on these 20-yr mean time scales exhibits enhanced SLP anomalies (and associated U700 anomalies; not shown) in the central Atlantic in March compared to the other months (Fig. A1k). The NAO indices, therefore, do not pick up on these detailed features and seasonal variations that are key to the enhanced U700 variability in March relative to the other months.

APPENDIX B

An Assessment of the Potential Role of a Stratospheric Pathway

It has been argued that the forcing of the North Atlantic jet by the AMV and the trends over the latter half of the twentieth century may have been driven via a stratospheric pathway (Scaife et al. 2005; Omrani et al. 2014). Through comparison of the response to a warming of North Atlantic SSTs in high-top and low-top model versions, Omrani et al. (2014) argued that a warming of the North Atlantic induces a strengthening of the polar vortex in midwinter that then propagates downward leading to a negative NAO anomaly in late winter. It is, however, difficult to reconcile the magnitude of the observed March zonal wind variability on 20-yr mean time scales with this argument.

While the DJF polar vortex strength of ERA20C compares reasonably well with more constrained reanalyses (Fig. B1a), one may be skeptical of the fidelity of ERA20C stratospheric winds further back in time. So, here we consider the change in the stratospheric polar vortex between two 20-yr periods that lie within the JRA-55 record: 1981–2000 and 1958–77. While the
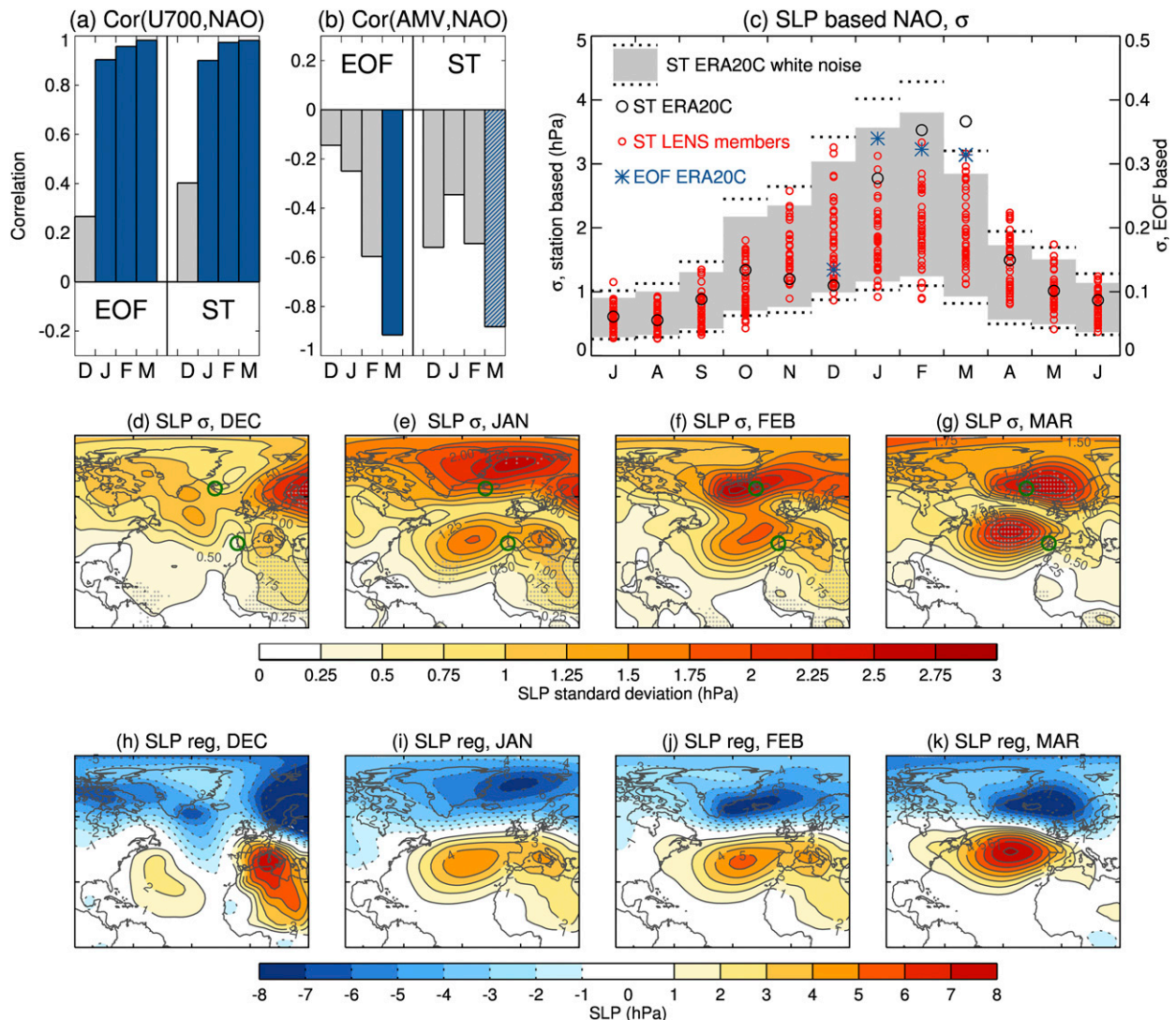
FIG. A1. (a) Correlation between 20-yr running mean U700NA and the EOF-based and ST-based SLP NAO indices for ERA20C + ERA-Interim, 1900–2017 (ERA-Interim after 2010). (b) Correlation between the TS ERSSTv5 AMV index and the EOF- and ST-based NAO indices, 1900–2017. (c) The 20-yr running mean σ of the NAO indices from 1920–2010. The left axis relates to the ST index and the right axis relates to the EOF index. Shown are the ERA20C ST index (black points), the ST index for the LENS historical members (red points), the 2.5–97.5 percentile value range of white noise samples of the ST index calculated using the ERA20C interannual σ (gray shaded region) and the same for the 1st and 99th percentiles (dotted), and the EOF index for ERA20C (blue asterisk). Note that the EOF points should not be compared with the model and white noise estimates for ST shown in the panel. (d)–(g) The σ of 20-yr running means of SLP from 1920 to 2010 for ERA20C. Gray stippling indicates where the variability is significantly different from white noise at the 5% level accounting for spatial correlation by the false discovery rate method of Wilks [2016, their Eq. (3)], with a false discovery control value of 20%. (h)–(k) The regression of 20-yr running mean SLP onto 20-yr running means of the EOF index. Note that differing lengths of record are used for each panel for consistency with comparable figures for U700 in the main text (Figs. 2, 4, and 8), but conclusions are not dependent on the exact record length used.

difference between these periods does not quite span the full range of variability seen in 20-yr means over the century, it spans a substantial portion of it with the later time period being characterized by greatly enhanced U700 in the North Atlantic compared to the earlier time period (Fig. B1c).

By the argument of Omrani et al. (2014), if the stratosphere were a key player in driving this response we should expect to see a strengthening of the polar vortex in midwinter in the later period, compared to the earlier period. Indeed, such a strengthening is found, but it is rather small compared to the interannual variability
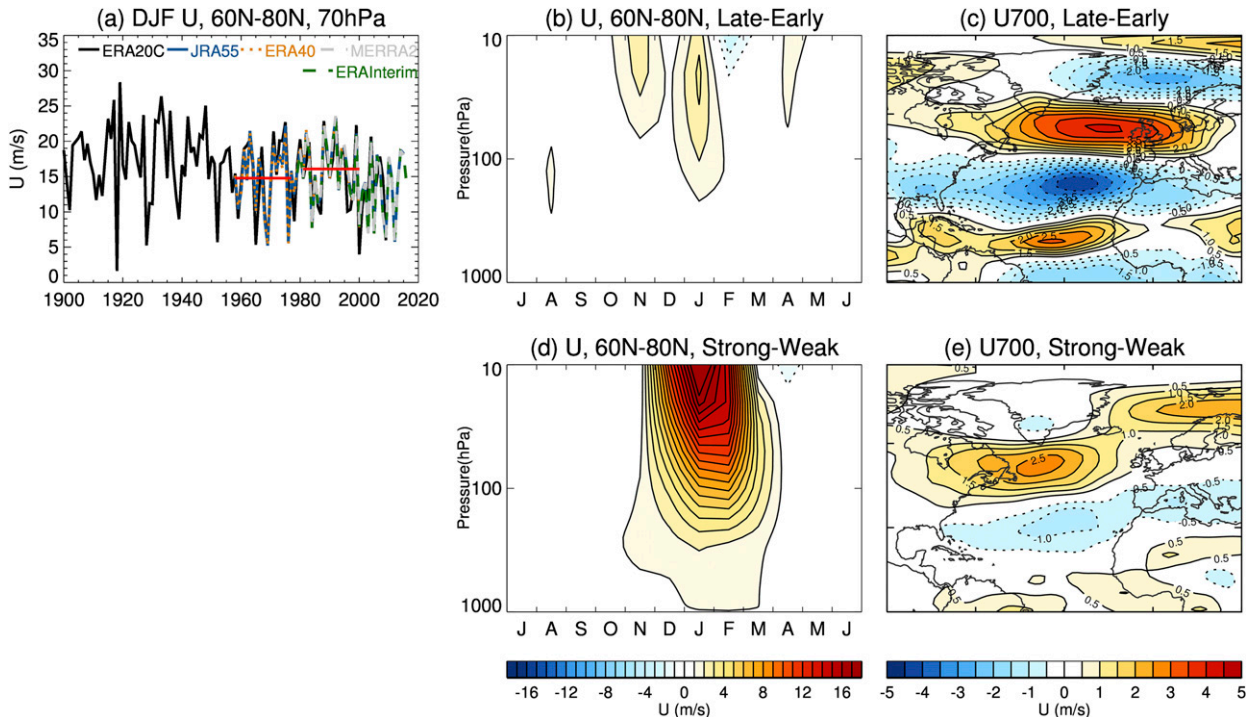
FIG. B1. (a) DJF averaged zonal mean zonal wind area averaged from 60° to 80°N at 70 hPa for all the reanalysis products. (b) The mean difference in 60°–80°N zonal mean zonal wind between the late and early periods depicted by the red lines in (a). Late is 1981–2000, and early is 1958–77. (c) The difference in March U700 between the late and early periods. (d) The composite mean difference in 60°–80°N zonal mean zonal wind between strong and weak vortex years. Strong (weak) vortex years are defined as those with DJF 60°–80°N, 70-hPa zonal mean zonal wind stronger (weaker) than average over the 1958–2016 period. (e) Composite mean difference in March U700 between strong and weak vortex years.

(Fig. B1b; also, compare the difference between the red lines in Fig. B1a with the interannual variability). To assess whether the magnitude of the March U700 anomalies is consistent with driving from the stratosphere, we compare with anomalies related to interannual variability in the polar vortex. Figure B1d shows the composite mean difference in polar vortex zonal winds between years when the DJF polar vortex is stronger than average with years when the DJF polar vortex is weaker than average. Here years have been composited from the full record length (1957–2016).

The interannual composite displays polar vortex anomalies at 70 hPa during DJF that are around 6 times stronger than the difference between the 1981–2000 and 1958–77 periods (cf. Figs. B1b and B1d) yet the March U700 anomalies in the North Atlantic are around half the magnitude (cf. Figs. B1c and B1e). In order for the changes in the stratospheric polar vortex to explain the difference between these two time periods, an explanation would be needed for why the stratospheric differences between the later and earlier periods are roughly 12 times more effective at producing March tropospheric zonal wind anomalies than interannual stratospheric

variability is. It therefore seems unlikely the March zonal wind anomalies are stratospherically driven.

REFERENCES

Athanasiadis, P. J., J. M. Wallace, and J. J. Wettstein, 2010: Patterns of wintertime jet stream variability and their relation to the storm tracks. *J. Atmos. Sci.*, **67**, 1361–1381, https://doi.org/10.1175/2009JAS3270.1.

Barnston, A. G., and R. E. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.*, **115**, 1083–1126, https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2.

Cane, M. A., A. C. Clement, L. N. Murphy, and K. Bellomo, 2017: Low-pass filtering, heat flux, and Atlantic multidecadal variability. *J. Climate*, **30**, 7529–7553, https://doi.org/10.1175/JCLI-D-16-0810.1.

Clement, A., K. Bellomo, L. N. Murphy, M. A. Cane, T. Mauritsen, G. Rädel, and B. Stevens, 2015: The Atlantic Multidecadal Oscillation without a role for ocean circulation. *Science*, **350**, 320–324, https://doi.org/10.1126/science.aab3980.

Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, https://doi.org/10.1002/qj.776.

Danabasoglu, G., and Coauthors, 2014: North Atlantic simulations in Coordinated Ocean-Ice Reference Experiments phase II

(CORE-II). Part I: Mean states. *Ocean Modell.*, **73**, 76–107, https://doi.org/10.1016/j.ocemod.2013.10.005.

Davini, P., J. von Hardenberg, and S. Corti, 2015: Tropical origin for the impacts of the Atlantic multidecadal variability on the Euro-Atlantic climate. *Environ. Res. Lett.*, **10**, 094010, https://doi.org/10.1088/1748-9326/10/9/094010.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

Delworth, T. L., and F. Zeng, 2016: The impact of the North Atlantic Oscillation on climate through its influence on the Atlantic meridional overturning circulation. *J. Climate*, **29**, 941–962, https://doi.org/10.1175/JCLI-D-15-0396.1.

——, ——, L. Zhang, R. Zhang, G. Vecchi, and X. Yang, 2017: The central role of ocean dynamics in connecting the North Atlantic Oscillation to the extratropical component of the Atlantic multidecadal oscillation. *J. Climate*, **30**, 3789–3805, https://doi.org/10.1175/JCLI-D-16-0358.1.

Deser, C., and M. L. Blackmon, 1993: Surface climate variations over the North Atlantic Ocean during winter: 1900–1989. *J. Climate*, **6**, 1743–1753, https://doi.org/10.1175/1520-0442(1993)006<1743:SCVOTN>2.0.CO;2.

——, and A. S. Phillips, 2009: Atmospheric circulation trends, 1950–2000: The relative roles of sea surface temperature forcing and direct atmospheric radiative forcing. *J. Climate*, **22**, 396–413, https://doi.org/10.1175/2008JCLI2453.1.

——, M. A. Alexander, and M. S. Timlin, 2003: Understanding the persistence of sea surface temperature anomalies in midlatitudes. *J. Climate*, **16**, 57–72, https://doi.org/10.1175/1520-0442(2003)016<0057:UTPOSS>2.0.CO;2.

——, ——, S.-P. Xie, and A. S. Phillips, 2010: Sea surface temperature variability: Patterns and mechanisms. *Annu. Rev. Mar. Sci.*, **2**, 115–143, https://doi.org/10.1146/annurev-marine-120408-151453.

Eden, C., and T. Jung, 2001: North Atlantic interdecadal variability: Oceanic response to the North Atlantic Oscillation (1865–1997). *J. Climate*, **14**, 676–691, https://doi.org/10.1175/1520-0442(2001)014<0676:NAIVOR>2.0.CO;2.

Eichelberger, S. J., and D. L. Hartmann, 2007: Zonal jet structure and the leading mode of variability. *J. Climate*, **20**, 5149–5163, https://doi.org/10.1175/JCLI4279.1.

Enfield, D. B., A. M. Mestas-Nuñez, and P. J. Trimble, 2001: The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S. *Geophys. Res. Lett.*, **28**, 2077–2080, https://doi.org/10.1029/2000GL012745.

Feldstein, S. B., 2000: The timescale, power spectra, and climate noise properties of teleconnection patterns. *J. Climate*, **13**, 4430–4440, https://doi.org/10.1175/1520-0442(2000)013<4430:TTPSAC>2.0.CO;2.

Frankcombe, L. M., M. H. England, M. E. Mann, and B. A. Steinman, 2015: Separating internal variability from the externally forced climate response. *J. Climate*, **28**, 8184–8201, https://doi.org/10.1175/JCLI-D-15-0069.1.

Frankignoul, C., G. Gastineau, and Y.-O. Kwon, 2017: Estimation of the SST response to anthropogenic and external forcing and its impact on the Atlantic multidecadal oscillation and the Pacific decadal oscillation. *J. Climate*, **30**, 9871–9895, https://doi.org/10.1175/JCLI-D-17-0009.1.

Gastineau, G., and C. Frankignoul, 2012: Cold-season atmospheric response to the natural variability of the Atlantic meridional overturning circulation. *Climate Dyn.*, **39**, 37–57, https://doi.org/10.1007/s00382-011-1109-y.

——, and ——, 2015: Influence of the North Atlantic SST variability on the atmospheric circulation during the twentieth century. *J. Climate*, **28**, 1396–1416, https://doi.org/10.1175/JCLI-D-14-00424.1.

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1.

Gillett, N. P., H. F. Graf, and T. J. Osborn, 2003: Climate change and the North Atlantic Oscillation. *The North Atlantic Oscillation: Climate Significance and Environmental Impact*, Geophys. Monogr., Vol. 134, Amer. Geophys. Union, 193–210.

Häkkinen, S., P. B. Rhines, and D. L. Worthen, 2011: Atmospheric blocking and Atlantic multidecadal ocean variability. *Science*, **334**, 655–659, https://doi.org/10.1126/science.1205683.

Hersbach, H., C. Peubey, A. Simmons, P. Berrisford, P. Poli, and D. Dee, 2015: ERA-20CM: A twentieth-century atmospheric model ensemble. *Quart. J. Roy. Meteor. Soc.*, **141**, 2350–2375, https://doi.org/10.1002/qj.2528.

Hodson, D. L. R., R. T. Sutton, C. Cassou, N. Keenlyside, Y. Okumura, and T. Zhou, 2010: Climate impacts of recent multidecadal changes in Atlantic Ocean sea surface temperature: A multimodel comparison. *Climate Dyn.*, **34**, 1041–1058, https://doi.org/10.1007/s00382-009-0571-2.

Hoerling, M. P., J. W. Hurrell, and T. Xu, 2001: Tropical origins for recent North Atlantic climate change. *Science*, **292**, 90–92, https://doi.org/10.1126/science.1058582.

Huang, B., and Coauthors, 2015: Extended Reconstructed Sea Surface Temperature version 4 (ERSSTv4). Part 1: Upgrades and intercomparisons. *J. Climate*, **28**, 911–930, https://doi.org/10.1175/JCLI-D-14-00006.1.

——, and Coauthors, 2017: Extended Reconstructed Sea Surface Temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *J. Climate*, **30**, 8179–8205, https://doi.org/10.1175/JCLI-D-16-0836.1.

Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, **269**, 676–679, https://doi.org/10.1126/science.269.5224.676.

——, and H. van Loon, 1997: Decadal variations in climate associated with the North Atlantic Oscillation. *Climatic Change*, **36**, 301–326, https://doi.org/10.1023/A:1005314315270.

——, and C. Deser, 2010: North Atlantic climate variability: The role of the North Atlantic Oscillation. *J. Mar. Syst.*, **79**, 231–244, https://doi.org/10.1016/j.jmarsys.2009.11.002.

——, Y. Kushnir, G. Ottersen, and M. Visbeck, 2003: An overview of the North Atlantic Oscillation. *The North Atlantic Oscillation: Climate Significance and Environmental Impact*, Geophys. Monogr., Vol. 134, Amer. Geophys. Union, 1–36.

——, M. P. Hoerling, A. S. Phillips, and T. Xu, 2004: Twentieth century North Atlantic climate change. Part 1: Assessing determinism. *Climate Dyn.*, **23**, 371–389, https://doi.org/10.1007/s00382-004-0432-y.

——, and Coauthors, 2013: The Community Earth System Model: A framework for collaborative research. *Bull. Amer. Meteor. Soc.*, **94**, 1339–1360, https://doi.org/10.1175/BAMS-D-12-00121.1.

Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) Large Ensemble Project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, https://doi.org/10.1175/BAMS-D-13-00255.1.

Kim, W. M., S. Yeager, P. Chang, and G. Danabasoglu, 2018: Low-frequency North Atlantic climate variability in the Community

Earth System Model Large Ensemble. *J. Climate*, **31**, 787–813, https://doi.org/10.1175/JCLI-D-17-0193.1.

Kobayashi, S., and Coauthors, 2015: The JRA-55 Reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, https://doi.org/10.2151/jmsj.2015-001.

Kravtsov, S., 2017: Pronounced differences between observed and CMIP5-simulated multidecadal climate variability in the twentieth century. *Geophys. Res. Lett.*, **44**, 5749–5757, https://doi.org/10.1002/2017GL074016.

Kushnir, Y., 1994: Interdecadal variations in North Atlantic sea surface temperature and associated atmospheric conditions. *J. Climate*, **7**, 141–157, https://doi.org/10.1175/1520-0442(1994)007<0141:IVINAS>2.0.CO;2.

——, W. A. Robinson, I. Bladé, N. M. J. Hall, S. Peng, and R. Sutton, 2002: Atmospheric GCM response to extratropical SST anomalies: Synthesis and evaluation. *J. Climate*, **15**, 2233–2256, https://doi.org/10.1175/1520-0442(2002)015<2233:AGRTES>2.0.CO;2.

Msadek, R., C. Frankignoul, and L. Z. X. Li, 2011: Mechanisms of the atmospheric response to North Atlantic multidecadal variability: A model study. *Climate Dyn.*, **36**, 1255–1276, https://doi.org/10.1007/s00382-010-0958-0.

Murphy, L. N., K. Bellomo, M. Cane, and A. Clement, 2017: The role of historical forcings in simulating the observed Atlantic multidecadal oscillation. *Geophys. Res. Lett.*, **44**, 2472–2480, https://doi.org/10.1002/2016GL071337.

Omrani, N.-E., N. S. Keenlyside, J. Bader, and E. Manzini, 2014: Stratosphere key for wintertime atmospheric response to warm Atlantic decadal conditions. *Climate Dyn.*, **42**, 649–663, https://doi.org/10.1007/s00382-013-1860-3.

Otto, F. E. L., D. J. Frame, A. Otto, and M. R. Allen, 2015: Embracing uncertainty in climate change policy. *Nat. Climate Change*, **5**, 917–921, https://doi.org/10.1038/nclimate2716.

Parfitt, R., A. Czaja, S. Minobe, and A. Kurano-Yoshida, 2016: The atmospheric frontal response to SST perturbations in the Gulf Stream region. *Geophys. Res. Lett.*, **43**, 2299–2306, https://doi.org/10.1002/2016GL067723.

Peings, Y., and G. Magnusdottir, 2014: Forcing of the wintertime atmospheric circulation by the multidecadal fluctuations of the North Atlantic Ocean. *Environ. Res. Lett.*, **9**, 034018, https://doi.org/10.1088/1748-9326/9/3/034018.

——, and ——, 2016: Wintertime atmospheric response to Atlantic multidecadal variability: Effect of stratospheric representation and ocean–atmosphere coupling. *Climate Dyn.*, **47**, 1029–1047, https://doi.org/10.1007/s00382-015-2887-4.

——, G. Simpkins, and G. Magnusdottir, 2016: Multidecadal fluctuations of the North Atlantic Ocean and feedback on the winter climate in CMIP5 control simulations. *J. Geophys. Res. Atmos.*, **121**, 2571–2592, https://doi.org/10.1002/2015JD024107.

Poli, P., and Coauthors, 2016: ERA-20c: An atmospheric reanalysis of the twentieth century. *J. Climate*, **29**, 4083–4097, https://doi.org/10.1175/JCLI-D-15-0556.1.

Qasmi, S., C. Cassou, and J. Boé, 2017: Teleconnection between Atlantic multidecadal variability and European temperature: Diversity and evaluation of the Coupled Model Intercomparison Project phase 5 models. *Geophys. Res. Lett.*, **44**, 11 140–11 149, https://doi.org/10.1002/2017GL074886.

Raible, C. C., T. F. Stocker, M. Yoshimori, M. Renold, U. Beyerle, C. Casty, and J. Luterbacher, 2005: Northern Hemisphere trends of pressure indices and atmospheric circulation patterns in observations, reconstructions, and coupled GCM simulations. *J. Climate*, **18**, 3968–3982, https://doi.org/10.1175/JCLI3511.1.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, https://doi.org/10.1029/2002JD002670.

Rodwell, M. J., D. P. Rowell, and C. K. Folland, 1999: Oceanic forcing of the wintertime North Atlantic Oscillation and European climate. *Nature*, **398**, 320–323, https://doi.org/10.1038/18648.

Ruprich-Robert, Y., R. Msadek, F. Castruccio, S. Yeager, T. Delworth, and G. Danabasoglu, 2017: Assessing the climate impacts of the observed Atlantic multidecadal variability using the GFDL CM2.1 and NCAR CESM1 global coupled models. *J. Climate*, **30**, 2785–2810, https://doi.org/10.1175/JCLI-D-16-0127.1.

Sardeshmukh, P. D., and B. J. Hoskins, 1984: Spatial smoothing on the sphere. *Mon. Wea. Rev.*, **112**, 2524–2529, https://doi.org/10.1175/1520-0493(1984)112<2524:SSOTS>2.0.CO;2.

Scaife, A. A., J. R. Knight, G. Vallis, and C. K. Folland, 2005: A stratospheric influence on the winter NAO and North Atlantic surface climate. *Geophys. Res. Lett.*, **32**, L18715, https://doi.org/10.1029/2005GL023226.

——, and Coauthors, 2009: The CLIVAR C20C project: Selected twentieth century climate events. *Climate Dyn.*, **33**, 603–614, https://doi.org/10.1007/s00382-008-0451-1.

Seager, R., Y. Kushnir, M. Visbeck, N. Naik, J. Miller, G. Krahmann, and H. Cullen, 2000: Causes of Atlantic Ocean climate variability between 1958 and 1998. *J. Climate*, **13**, 2845–2862, https://doi.org/10.1175/1520-0442(2000)013<2845:COAOCV>2.0.CO;2.

Selten, F. M., G. W. Branstator, H. A. Dijkstra, and M. Kliphuis, 2004: Tropical origins for recent and future Northern Hemisphere climate change. *Geophys. Res. Lett.*, **31**, L21205, https://doi.org/10.1029/2004GL020739.

Siqueira, L., and B. P. Kirtman, 2016: Atlantic near-term climate variability and the role of a resolved Gulf Stream. *Geophys. Res. Lett.*, **43**, 3964–3972, https://doi.org/10.1002/2016GL068694.

Smirnov, D., M. Newman, M. A. Alevander, Y.-O. Kwon, and C. Frankignoul, 2015: Investigating the local atmospheric response to a realistic shift in the Oyashio sea surface temperature front. *J. Climate*, **28**, 1126–1147, https://doi.org/10.1175/JCLI-D-14-00285.1.

Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006). *J. Climate*, **21**, 2283–2296, https://doi.org/10.1175/2007JCLI2100.1.

Stephenson, D. B., V. Pavan, and R. Bojariu, 2000: Is the North Atlantic Oscillation a random walk? *Int. J. Climatol.*, **20**, 1–18, https://doi.org/10.1002/(SICI)1097-0088(200001)20:1<1::AID-JOC456>3.0.CO;2-P.

Sutton, R., and D. Hodson, 2007: Climate response to a multidecadal warming and cooling of the North Atlantic Ocean. *J. Climate*, **20**, 891–907, https://doi.org/10.1175/JCLI4038.1.

Tandon, N. F., and P. J. Kushner, 2015: Does external forcing interfere with the AMOV's influence on North Atlantic sea surface temperature? *J. Climate*, **28**, 6309–6323, https://doi.org/10.1175/JCLI-D-14-00664.1.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Ting, M., Y. Kushnir, R. Seager, and C. Li, 2009: Forced and internal twentieth-century SST trends in the North

Atlantic. *J. Climate*, **22**, 1469–1481, https://doi.org/10.1175/2008JCLI2561.1.

——, ——, ——, and ——, 2011: Robust features of Atlantic multidecadal variability and its climate impacts. *Geophys. Res. Lett.*, **38**, L17705, https://doi.org/10.1029/2011GL048712.

——, ——, and C. Li, 2014: North Atlantic multidecadal SST oscillation: External forcing versus internal variability. *J. Mar. Syst.*, **133**, 27–38, https://doi.org/10.1016/j.jmarsys.2013.07.006.

Trenberth, K. E., and D. J. Shea, 2006: Atlantic hurricanes and natural variability in 2005. *Geophys. Res. Lett.*, **33**, L12704, https://doi.org/10.1029/2006GL026894.

Uppala, S. M., and Coauthors, 2005: The ERA-40 Reanalysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, https://doi.org/10.1256/qj.04.176.

van Loon, H., and J. C. Rogers, 1978: The seesaw in winter temperatures between Greenland and northern Europe. Part 1: General description. *Mon. Wea. Rev.*, **106**, 296–310, https://doi.org/10.1175/1520-0493(1978)106<0296:TSIWTB>2.0.CO;2.

Walker, G. T., and E. W. Bliss, 1932: World weather V. *Mem. Roy. Meteor. Soc.*, **4**, 53–84.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784–812, https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2.

Wang, X., J. Li, C. Sun, and T. Liu, 2017: NAO and its relationship with the Northern Hemisphere mean surface temperature in CMIP5 simulations. *J. Geophys. Res. Atmos.*, **122**, 4202–4227, https://doi.org/10.1002/2016JD025979.

Wilks, D. S., 2016: "The stippling shows statistically significant grid points": How research results are routinely overstated and overinterpreted, and what to do about it. *Bull. Amer. Meteor. Soc.*, **97**, 2263–2273, https://doi.org/10.1175/BAMS-D-15-00267.1.

Woollings, T., 2010: Dynamical influences on European climate: An uncertain future. *Philos. Trans. Roy. Soc.*, **368A**, 3733–3756, https://doi.org/10.1098/rsta.2010.0040.

——, A. Hannachi, and B. Hoskins, 2010: Variability of the North Atlantic eddy-driven jet stream. *Quart. J. Roy. Meteor. Soc.*, **136**, 856–868, https://doi.org/10.1002/qj.625.

——, C. Czuchnicki, and C. Franzke, 2014: Twentieth century North Atlantic jet variability. *Quart. J. Roy. Meteor. Soc.*, **140**, 783–791, https://doi.org/10.1002/qj.2197.

——, C. Franzke, D. L. R. Hodson, B. Dong, E. A. Barnes, C. C. Raible, and J. G. Pinto, 2015: Contrasting interannual and multidecadal NAO variability. *Climate Dyn.*, **45**, 539–556, https://doi.org/10.1007/s00382-014-2237-y.

Wunsch, C., 1999: The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillations. *Bull. Amer. Meteor. Soc.*, **80**, 245–255, https://doi.org/10.1175/1520-0477(1999)080<0245:TIOSCR>2.0.CO;2.

Yeager, S. G., and G. Danabasoglu, 2014: The origins of late-twentieth-century variations in the large-scale North Atlantic circulation. *J. Climate*, **27**, 3222–3247, https://doi.org/10.1175/JCLI-D-13-00125.1.

——, A. Karspeck, G. Danabasoglu, J. Tribbia, and H. Teng, 2012: A decadal prediction case study: Late twentieth-century North Atlantic Ocean heat content. *J. Climate*, **25**, 5173–5189, https://doi.org/10.1175/JCLI-D-11-00595.1.

——, and Coauthors, 2018: Predicting near-term changes in the Earth System: A large ensemble of initialized decadal prediction simulations using the Community Earth System Model. *Bull. Amer. Meteor. Soc.*, https://doi.org/10.1175/BAMS-D-17-0098.1, in press.

Zhang, R., R. Sutton, G. Danabasoglu, T. L. Delworth, W. H. Kim, J. Robson, and S. G. Yeager, 2016: Comment on "The Atlantic Multidecadal Oscillation without a role for ocean circulation." *Science*, **352**, 1527, https://doi.org/10.1126/science.aaf1660.