

# Variable External Forcing Obscures the Weak Relationship between the NAO and North Atlantic Multidecadal SST Variability<sup>©</sup>

JEREMY M. KLAVANS AND AMY C. CLEMENT

*Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida*

MARK A. CANE

*Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York*

(Manuscript received 25 June 2018, in final form 19 March 2019)

## ABSTRACT

North Atlantic sea surface temperatures (SST) exhibit a lagged response to the North Atlantic Oscillation (NAO) in both models and observations, which has previously been attributed to changes in ocean heat transport. Here we examine the lagged relationship between the NAO and Atlantic multidecadal variability (AMV) in the context of the two other major components of the AMV: atmospheric noise and external forcing. In preindustrial control runs, we generally find that after accounting for spurious signals introduced by filtering, the SST response to the NAO is only statistically significant in the subpolar gyre. Further, the lagged SST response to the NAO is small in magnitude and offers a limited contribution to the AMV pattern, statistics, or predictability. When climate models include variable external forcing, the relationship between the NAO and AMV is obscured and becomes inconsistent. In these historically forced runs, knowledge of the prior NAO offers reduced predictability. The differences between the preindustrial and the historically forced ensembles suggest that we do not yet have enough observational data to surmise the true NAO–AMV relationship and add evidence that external forcing plays a substantial role in producing the AMV.

## 1. Introduction

There is an ongoing debate as to how active a role the ocean plays in generating multidecadal climate variability in the Atlantic. Observational and proxy analyses identify a low-frequency, basin-wide warming and cooling of SST in the North Atlantic, which is generally referred to as the AMV (Deser and Blackmon 1993; Kushnir 1994; Schlesinger and Ramankutty 1994; Kerr 2000; Delworth and Mann 2000). Identifying a physical mechanism for the AMV is difficult given the short observational record relative to the time scale of the variability; long runs of climate models are the best available tools to better understand this oscillation (Latif and Keenlyside 2011; Buckley and Marshall 2016 and references therein). Research suggests that the three major

components of North Atlantic SST variability are natural variability in the Atlantic meridional overturning circulation (AMOC) (e.g., Delworth et al. 2017), the passive ocean response to atmospheric noise (e.g., Clement et al. 2015), and changes in external forcing (e.g., Bellucci et al. 2017). There is disagreement as to the relative contribution of each of these terms to overall SST variability in the North Atlantic (Vecchi et al. 2017; Buckley and Marshall 2016).

The first major component of the AMV is variable ocean heat transport, potentially related to variability in AMOC. Mechanisms that invoke changes in ocean heat transport build on the hypothesis that the quickly moving, low heat capacity atmosphere drives SST variability on short time scales, while the slowly moving, high heat capacity ocean drives SST variability on long time scales (Bjerknes 1964; Kushnir 1994). Recent studies of both observational products and climate models show a change of sign in the correlation of turbulent heat fluxes with SST as a function of time scale; this is interpreted as evidence in support of Bjerknes' (1964) hypothesis (Gulev et al. 2013; O'Reilly et al. 2016). There is an ongoing debate about whether this metric can be used to

<sup>©</sup> Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-18-0409.s1>.

Corresponding author: Jeremy M. Klavans, jklavans@rsmas.miami.edu

identify a causal relationship between the ocean circulation and SST (Cane et al. 2017; Zhang 2017). Many modeling studies have noted coherence (including lagged correlations) between the AMV and meridional ocean heat transport, which includes metrics that describe the AMOC (e.g., Delworth et al. 1993; Timmermann et al. 1998; Delworth and Greatbatch 2000; Latif et al. 2004; Jungclaus et al. 2005; Knight et al. 2005; Danabasoglu et al. 2012; Ba et al. 2014; Delworth et al. 2017; Zhang 2017; Wills et al. 2018). Generally, these studies hypothesize that buoyancy fluxes associated with atmospheric variability, often characterized by the North Atlantic Oscillation (NAO), drive changes in the strength of the AMOC; variations in AMOC then advect anomalous heat into the northern portion of the Atlantic, which produces SST anomalies. These changes in the AMOC are invoked as a source of predictability for subpolar gyre heat anomalies (e.g., Yeager et al. 2012). Despite the apparent connection between the AMV and ocean heat transport in some models, the literature has not established a conclusive observational link (Lozier 2010; Buckley and Marshall 2016).

The second major component of the AMV is the influence of atmospheric variability on a passive ocean. The NAO index describes a plurality of atmospheric noise over the North Atlantic; it is often calculated by taking the first empirical orthogonal function (EOF) of sea level pressure over the North Atlantic (Thompson and Wallace 2001). This index exhibits variability at all time scales in both models and observations (Deser et al. 2010; Woollings et al. 2015). (Climate models typically underestimate the level of low-frequency variance in the NAO, relative to observations; Kim et al. 2018). Interannual and higher frequencies of this oscillation are contemporaneously associated with an SST tripole, which is primarily forced by air-sea flux anomalies (Deser et al. 2010). The relationship between the NAO and SST is succinctly described by simple stochastic climate models, wherein SST anomalies are a function of air-sea fluxes and upper-ocean memory (Hasselmann 1976; Frankignoul and Hasselmann 1977; Wunsch 1999). This serves as the null hypothesis for SST variability (Deser et al. 2010). This relationship also holds true at longer time scales except in those regions where ocean heat transport variability is important such as the North Atlantic subpolar gyre (Buckley et al. 2015; Karspeck et al. 2015).

The third major component of the AMV is changes in external forcing. Recent modeling studies show that the AMV does respond to experimental changes in external forcing (e.g., volcanoes, CO<sub>2</sub>, aerosols). For example, Murphy et al. (2017) find that the variance of the AMV is larger in runs of climate models with variable external

forcing, which is consistent with work showing that models that include variable external forcing, particularly volcanoes and anthropogenic aerosols, can explain upward of 2/3 of AMV variability (Otterå et al. 2010; Booth et al. 2012; Bellucci et al. 2017; Bellomo et al. 2018). Zhang et al. (2013) challenge this view; they note that at least one model responds too strongly to external forcing. Similarly, climate models may undersimulate internal variability in both the atmosphere and ocean (Kim et al. 2018).

All three mechanisms described above contribute to overall climate variability in the North Atlantic; however, the relative contributions of each to the AMV and how those contributions may vary with time and time scale are uncertain. Clement et al. (2015) found that climate models with and without variable ocean heat transport can produce AMV indices with similar associated patterns and spectral characteristics. They show this using a climate model coupled to a mixed-layer depth ocean, which would explicitly exclude the ocean heat transport variability required by the first potential driver (Clement et al. 2015). However, many fully coupled climate models simulate a lagged relationship between the NAO and subpolar gyre SSTs at decadal time scales, whereas mixed-layer ocean models do not (Delworth et al. 2017; Peings et al. 2016; Zhang 2017). This discrepancy across model configuration is interpreted to mean that ocean dynamics, including the AMOC, exert influence on North Atlantic SST at multidecadal time scales.

This paper focuses on the lagged relationship between the NAO and AMV, which we presume to be mediated by some ocean dynamical mechanism (Delworth et al. 2017; Peings et al. 2016; Zhang 2017). We examine this relationship in the context of the other two major components of the AMV, atmospheric noise and external forcing. The relative importance of the NAO-AMV mechanism is evaluated statistically in preindustrial control runs of CMIP5 climate models. The relative importance of external forcing is evaluated experimentally by comparing historically forced runs to preindustrial control runs. Unlike many previous studies, we evaluate this relationship against the null hypothesis that accounts for the influence of filtering. We find that North Atlantic subpolar gyre SSTs exhibit a lagged response to the NAO in many fully coupled climate models with constant forcing, which is consistent with previous findings. However, we find that the lagged response to the NAO's contribution to the AMV is limited, offering only marginal explanatory power for North Atlantic SSTs. Further, we find that the relationship between the NAO and AMV is small enough that it is easily overwhelmed or obscured by the signal

associated with variable external forcing. In the presence of variable external forcing, knowledge of the prior state of the NAO may not be as useful a predictor of the AMV as it is in PI runs.

## 2. Data and methods

### a. Observations and observational products

Our analysis considers the relationship between atmospheric and SST variability in both models and observations. For SST, we consider two observational products. The first is the National Oceanic and Atmospheric Administration's global monthly observational product, Extended Reconstructed SST (ERSST) dataset, version 3 (Smith et al. 2008). The second is the Hadley Centre Sea Surface Temperature dataset (HadSST), version 3.1.1.0 (Kennedy et al. 2011). For ERSST, we linearly detrend each grid point before conducting analysis. For HadSST, we consider their median values of SST anomalies, which have also been detrended.

In this work, we characterize atmospheric variability using the NAO index. For our observational analysis, we choose the instrumental, normalized, station-based NAO index for December–March (DJFM) (see below for more detail; Hurrell and National Center for Atmospheric Research Staff 2018; Jones et al. 2003). For both SST and the NAO index, we utilize 153 years of observations from 1864 to 2016.

### b. Model output

Model output is selected from publicly available pre-industrial (PI) control runs and historically forced runs of CMIP5 climate models. We choose both PI control and historically forced runs from 27 individual models in the CMIP5 archive; additionally, we include two PI control runs from GFDL-CM2.1 and a configuration of CESM-CAM5 with a mixed-layer depth ocean to enable comparison to the existing literature. We report results from CMIP5 models where both PI and historically forced runs are readily available. All PI control runs have constant external forcing; the runs vary in length and resolution. A full list of climate models utilized appears in Table 1. Finally, we select historically forced runs from the CESM-CAM5 large ensemble, which we denote as CESM-LE. This 42-member ensemble covers the time period 1920–2005 (Kay et al. 2015); we consider the 41 of the 42 ensemble members that were branched from the first model run. Each member starts at slightly different initial conditions in air temperature (Kay et al. 2015). All model components are identical versions to those selected for our CESM-CAM5 PI control run.

### c. Definitions

We consider two indices of SST: the AMV index and the AMV<sub>MID</sub> index. The former is defined as the low-pass-filtered, linearly detrended time series of the cosine weighted average surface temperature over the ocean from 0°–60°N to 0°–80°W. The latter is defined as the low-pass-filtered, linearly detrended time series of the cosine weighted average surface temperature over the ocean in the region 40°–60°N and 20°–60°W. To examine multidecadal variability, we use Matlab's implementation of a fourth-order Butterworth filter with a half-power frequency of 1/10 yr. Other formulations of the AMV index eschew linear detrending, so as to more fully remove the forced signal (e.g., Trenberth and Shea 2006; Frankignoul et al. 2017). We believe nonlinear detrending may be inappropriate for our particular study; for example, removing global average SST from the AMV index (Trenberth and Shea 2006) may also remove any forcing that jointly affects the AMV and global mean temperature (Murphy et al. 2017). Our results are qualitatively similar whether the trend is removed linearly or nonlinearly (not shown).

To characterize atmospheric variability, we consider two forms of the NAO index. The first mimics the station-based or two-point index referenced above. This index is defined as the difference in normalized sea level pressure anomalies between the grid cells containing Stykkisholmur, Iceland, and the Azores. The second index is defined by the first principal component of sea level pressure in the region 20°–80°N, 90°W–40°E. The images below have been calculated using the EOF-based index; results are similar for both indices unless otherwise noted. To examine multidecadal variability, we again use Matlab's implementation of a fourth-order Butterworth filter with a half-power frequency of 1/10 yr.

### d. Methods

To study relationships between variables, we use lagged correlations and regressions between annual average SST and the wintertime (DJFM) average NAO index. Our choice of method only considers the linear portion of the relationship between the NAO and SST, which is consistent with the literature and allows for direct comparison to Delworth et al. (2017).

We establish statistical confidence by testing against the null hypothesis that low-pass-filtered SST variability is primarily driven by low-pass-filtered atmospheric noise. Filtering introduces spurious signals that are distinguishable from a random-walk process (Cane et al. 2017). Further, filtering complicates causal interpretation of statistical relationships. Cane et al. (2017) find

TABLE 1. This table explores the magnitude of the relationship between the low-pass-filtered NAO and the AMV in CMIP5 models, for both PI control runs and historically forced runs. The lag of maximum correlation is calculated as the maximum value of the cross-correlation function between the NAO index and AMV index. The NAO leads at positive lags; the AMV leads at negative lags. The percent variance explained by the contemporaneous NAO is calculated as the square of the NAO's correlation with the AMV or AMV<sub>MID</sub> index. The percent of AMV index variance explained by the lagged NAO is determined by linearly removing the lagged SST response to the NAO. We then find the correlation between the original AMV index and the AMV index of the temperature field without the signal from the prior NAO. A more detailed description is included in the main text.

		Observations				Models				PI control				% AMV <sub>MID</sub>		% AMV	
		Years	Years	Lag of maximum correlation	Lag of maximum correlation	Years	Years	% AMV <sub>MID</sub> variance explained by NAO <sub>t=0</sub>	% AMV variance explained by NAO <sub>t=0</sub>	Lag of maximum correlation	Lag of maximum correlation	% AMV <sub>MID</sub> variance explained by NAO <sub>t=0</sub>	% AMV variance explained by NAO <sub>t=0</sub>	% AMV <sub>MID</sub> variance explained by lagged response	% AMV variance explained by lagged response	% AMV <sub>MID</sub> variance explained by lagged response	% AMV variance explained by lagged response
ERSSV/ NAO	153	15	20%	7%	31%	19%											
HadSST/ NAO	153	14	5%	2%	69%	27%											
		Historical				CMIP5 models				CMIP5 models				7%		4%	
CESM1- CAM5						155	7	0%	2%	41%	35%	499	6	0%	3%	7%	7%
CESM1- SOM						155	25	12%	6%	15%	15%	499	6	0%	0%	5%	3%
GFDL- CM2.1						155	30	5%	1%	0%	2%	995	12	8%	2%	2%	3%
ACCESS1.0						155	18	3%	21%	8%	10%	1050	12	3%	2%	4%	3%
ACCESS1.3						155	8	7%	1%	1%	10%	276	13	24%	16%	4%	8%
CanESM2						155	7	15%	10%	5%	19%	329	9	13%	5%	0%	2%
CCSM4						155	-11	5%	4%	11%	9%	499	7	1%	1%	8%	12%
CMCC- CESM						155	8	0%	1%	6%	51%	499	11	3%	1%	1%	3%
CMCC-CM						145	-29	4%	6%	4%	16%	499	8	0%	0%	8%	9%
CMCC-CMS						144	15	2%	2%	7%	23%	499	8	2%	0%	0%	1%
CSIRO- MR3.6.0						144	4	11%	13%	1%	6%	499	6	3%	3%	5%	5%
GFDL-CM3						160	22	1%	3%	9%	12%	250	6	1%	0%	0%	1%
GFDL- ESM2G						155	-18	7%	2%	1%	6%	779	6	18%	15%	4%	3%
GFDL- ESM2M																	
GISS-E2-H- CC																	
GISS-E2-H																	

TABLE 1. (Continued)

	Years	Lag of maximum correlation	% AMV <sub>MID</sub> variance explained by NAO <sub>t=0</sub>	% AMV <sub>MID</sub> variance explained by NAO <sub>t=0</sub>	% AMV <sub>MID</sub> variance explained by lagged response	Lag of maximum correlation	Years	% AMV <sub>MID</sub> variance explained by NAO <sub>t=0</sub>	% AMV <sub>MID</sub> variance explained by NAO <sub>t=0</sub>	% AMV <sub>MID</sub> variance explained by lagged response	% AMV <sub>MID</sub> variance explained by lagged response
GISS-E2-R-CC	160	28	11%	3%	2%	10%	250	7	8%	11%	11%
HadGEM2-AO	145	-8	2%	5%	29%	18%	699	8	1%	0%	7%
HadGEM2-CC	145	30	0%	1%	29%	26%	239	9	36%	32%	3%
INM-CM4	155	8	2%	0%	15%	13%	499	10	7%	3%	1%
IPSL-CM5A-LR	155	-30	5%	7%	18%	2%	999	9	6%	1%	3%
IPSL-CM5A-MR	155	-21	14%	0%	20%	19%	299	7	12%	1%	1%
IPSL-CM5B-LR	155	27	0%	4%	10%	26%	299	22	4%	0%	3%
MIROC4h	55	11	19%	13%	37%	30%	99	13	11%	23%	1%
MIROC5	162	14	2%	6%	7%	10%	869	1	0%	0%	0%
MIROC-ESM	155	7	3%	1%	4%	13%	629	23	22%	8%	0%
MPI-ESM-MR	155	30	7%	0%	4%	21%	999	7	3%	0%	2%
MRI-CGCM3	155	13	24%	33%	7%	8%	499	14	5%	0%	3%
NorESM1-ME	155	7	0%	2%	6%	10%	251	8	6%	2%	2%
NorESM1-MCMIP5 average	155	10	14%	10%	16%	36%	500	7	10%	0%	1%
CMIP5 std dev	17.42	0.06	0.07	0.11	0.11	4.69	0.09	0.08	0.03	0.03	0.03

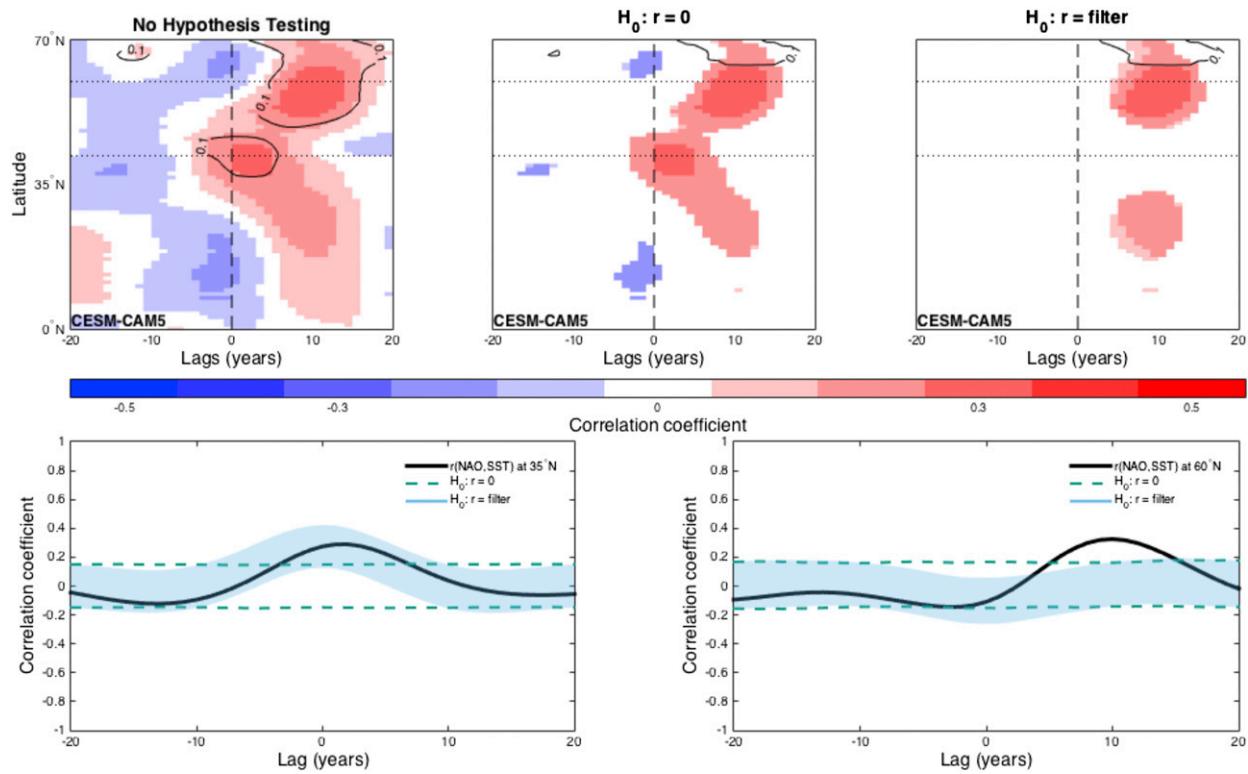


FIG. 1. We illustrate our approach to a hypothesis test that accounts for spurious signals introduced by filtering. (top) Zonal mean lagged correlation (colors) and regression (contours;  $^{\circ}\text{C per std dev NAO index}$ ) between the low-pass-filtered NAO index and low-pass-filtered annual average SST for a preindustrial control run of CESM1 at lags  $-20$  through  $20$ . The NAO leads at positive lags; the AMV leads at negative lags. (left) Zonal mean lagged correlations with no significance testing. (middle) Zonal mean lagged correlations tested against the null hypothesis that the correlation coefficient is equal to zero at all lags;  $95\%$  confidence intervals are constructed with a moving-block bootstrap. (right) Zonal mean lagged correlations tested against the null hypothesis that the correlation at a given lag will be equal to the autocorrelation of the filter at that lag; confidence intervals are again constructed via moving-block bootstrap. (bottom) Examples of the statistical test applied to the cross-correlation function at  $35^{\circ}$  and  $60^{\circ}\text{N}$  (black dashed lines in top plots). Green dashes represent the confidence intervals around zero (corresponding to top middle plot); blue shading represents the confidence intervals around the autocorrelation function of the filter (corresponding to top right plot). Differences between the blue-shaded and green-dashed areas primarily reflect the lag 0 correlation, as spread to other lags via filtering. Note that, without testing the filter-based null hypothesis, we may be tempted to interpret the positive lagged correlations at  $35^{\circ}\text{N}$  as an intergyre advective signal.

that the expected relationship between the low-pass-filtered time series of atmospheric noise and SST is approximately equivalent to the autocorrelation function of the filter, weighted by the portion of variance attributed to the atmosphere [see their Eq. (11e)]. Rather than testing the null hypothesis that the true correlation between the low-pass-filtered NAO index and low-pass-filtered SST is zero [as in, e.g., [Delworth et al. \(2017\)](#)], we test the null hypothesis that the true correlation between these two indices is equal to the autocorrelation function of the filter itself. We accomplish this by first calculating empirical  $95\%$  confidence intervals via a moving-block bootstrap, where the block size is equal to the length of the original time series. This choice of block-size allows us to preserve as much low-frequency covariance as possible ([Delworth et al. 2017](#)). We then recenter the confidence intervals around the autocorrelation function of the filter, scaled by the lag 0

correlation. One can validate this method in the one-dimensional noise forced model by comparing the autocorrelation function of the filter to the cross correlation of the low-pass-filtered atmospheric noise term and the low-pass-filtered SST term ([Cane et al. 2017](#); [Hasselmann 1976](#); [Frankignoul and Hasselmann 1977](#)). Note that this statistical test only accounts for the magnitude of the correlation at a given lag but does not account for uncertainty in the timing of the lag of maximum correlation. This statistical test is illustrated in [Fig. 1](#) and applied in [Fig. 2](#) (and Figs. S3 and S5–S7 in the online supplemental material).

In [Fig. 1](#), we consider lagged correlations and regressions between the low-pass-filtered NAO and low-pass-filtered, zonal mean SST from a PI control run of CESM1. First, we simply report the correlations with no additional processing; interpreted directly, this would suggest that negative SST anomalies ( $\sim -0.1^{\circ}\text{C}$ ) precede

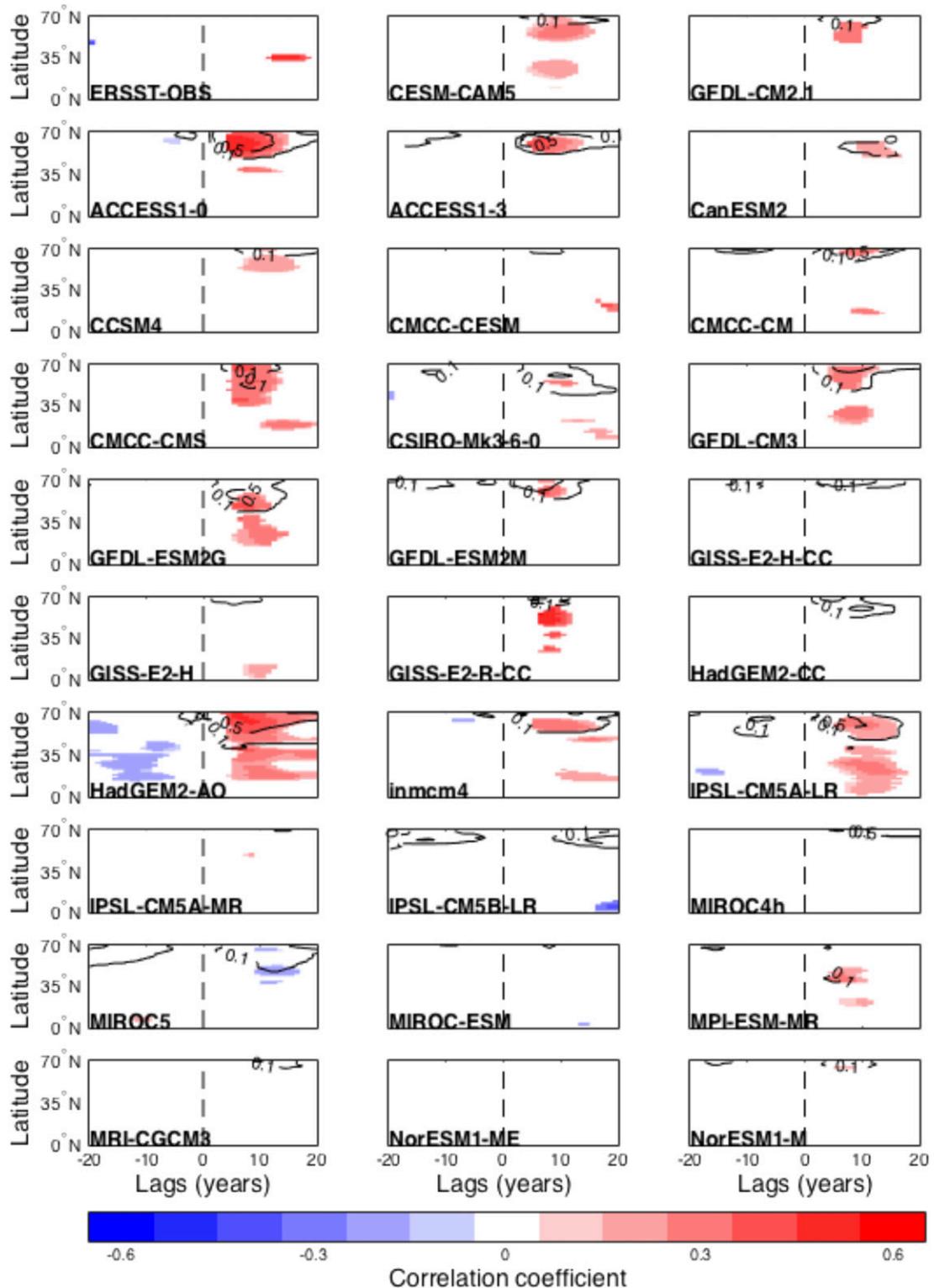


FIG. 2. The zonal mean correlation (colors) and regression (contours;  $^{\circ}\text{C}$  per std dev NAO index) between the low-pass-filtered NAO index and low-pass-filtered annual average SST for a selection of CMIP5 preindustrial model runs at lags  $-20$  through  $20$ . The NAO leads at positive lags; the AMV leads at negative lags. Pixels that are not statistically significantly different from the spurious signal associated with filtering at the  $95\%$  level are colored white. Note the variability in the lag of maximum correlation across models as well as the discrepancy between models and observations.

the positive phase of the NAO by about 10 years and positive SST anomalies ( $\sim 0.1^\circ\text{C}$ ) follow the positive phase of the NAO by about 10 years (Fig. 1, top left). Next, we consider the same regressions and correlations, but we test the null hypothesis that the correlation or regression coefficient at each grid point is equal to zero (Fig. 1, top center). This test is commonly used in the literature, notably in figures recently presented in Delworth et al. (2017). Even at the 95% confidence interval, one may be tempted to interpret this result as an advective signal, with warm subtropical SST anomalies associated with the NAO moving northward over time into the subpolar gyre. Finally, when we apply the null hypothesis which accounts for the imprint of the filter at the 95% confidence level, we isolate the lagged SST response to the NAO+ (Fig. 1, top right). Interpreted directly, this suggests that a positive SST anomaly ( $\sim 0.1^\circ\text{C}$ ) follows the NAO+ by about 10 yr in CESM1 but it does not indicate meridional advection. In isolation, this picture may lend support for a subpolar gyre-based mechanism for the AMV (e.g., Häkkinen et al. 2011; Piecuch et al. 2017); our results below remain agnostic as to the precise nature of the ocean dynamical mechanism behind the NAO–AMV relationship. This is in part because relaxing the confidence interval to the 80% may allow for a potential role for intergyre advection. Of course, these results are only for one model; a more detailed and generalizable analysis is presented below.

Temporal filtering may be a necessary evil in studying low-frequency climate variability, but we must account for its influence on our analyses. We believe the aforementioned method is a transparent attempt to account for the spurious signals associated with filtering; however, other researchers have proposed different methodologies. For example, recent work by Wills et al. (2018) attempts to deal with the issues associated with analyzing filtered data through their use of low-frequency component analysis. We establish no preference between the two methods but note that we find similar results to Wills et al. (2019) in PI control runs (although we interpret them in the context of historically forced runs).

### 3. Results

#### a. The small, lagged SST response to the NAO adds little predictive skill

We find that there is a statistically significant lagged warm SST response to a prolonged positive phase of the NAO in most fully coupled models (Fig. 2). This is consistent with several other studies that employ different statistical tests (e.g., Visbeck et al. 1998; Delworth

et al. 2017; Peings et al. 2016). Of the PI control runs we consider, we find that 17 out of 29 produce lagged positive SST anomalies that meet our threshold for statistical significance. In models, the warm SST response to the low-pass-filtered NAO index is generally limited to the subpolar gyre though with variable spatial extent and magnitude. For example, CCSM4, CESM1-CAM5, GFDL-CM2.1, and GFDL-CM3 (among others) tend to simulate a warm response that includes a signal confined to the northern portion of the basin, whereas the warming in HadGEM2-AO extends continuously into the subtropics. 29% (8 of 29) of PI runs produce a lagged warm signal in the subtropics so we cannot conclude that the tropical signal is robust. The connection between the tropical and extratropical portions of the AMV is considered an open research question both in models and observations (Delworth et al. 2017).

Previous work has identified an ocean dynamical mechanism as the lagged link between the NAO and SST on decadal time scales. The spatial scales (gyre-to basin-scale) and dominant forcing (e.g., wind- or buoyancy-forcing) required for this mechanism remain areas of active research (e.g., Delworth et al. 2017; Peings et al. 2016; Piecuch et al. 2017; O'Reilly and Zanna 2018). To evaluate the dynamical response of the ocean to a prolonged phase of the NAO, we consider volume transport and ocean heat transport (OHT) at  $56.5^\circ\text{N}$ . Foukal and Lozier (2018) find that a majority of interannual ocean heat content variability in the eastern portion of the subpolar gyre can be explained by heat advection that passes through the plane,  $S$ ,  $56.5^\circ\text{N}$ ,  $35^\circ$ – $9^\circ\text{W}$ , from the surface to the bottom of the ocean. When we consider volume transport and OHT integrated across this plane in a PI run of CESM, we find a lagged circulation response to the NAO at near-decadal lags (Figs. S1 and S2), which is consistent with increased import of warm water from the south as in Foukal and Lozier (2018).

Regression coefficients associated with the NAO–SST signal are small, generally  $0.1^\circ\text{C}$  per standard deviation of the NAO index or less, in both models and observations (Fig. 2). Similarly, correlation coefficients are small, typically less than 0.4. Significant correlations (at the 95% level) are largely confined to the higher latitudes and near-decadal lags. We cannot reject the null hypothesis that subtropical warming at short lags (0–4 yr) is a spurious signal introduced by the filter [cf. Fig. 2 herein to Fig. 7 of Delworth et al. (2017)]. Our statistical test does not allow us to conclude that heat anomalies are meridionally advected from the subtropical gyre to the subpolar gyre. Thus, accounting for the influence of the filter, we agree with the literature (e.g., Delworth et al. 2017; Piecuch et al. 2017) that the NAO is associated with a

small, lagged, warm response in the subpolar gyre but provides little help in understanding the subtropical or tropical portions of the AMV.

While this lagged signal appears in some PI control runs, there are large discrepancies between models and observations (Table 1). Observed SSTs warm about 14–15 yr after a NAO+, whereas simulated SSTs warm about 7–10 yr after a NAO+. The zonal mean spatial structure of the response in observations appears to be very different from most CMIP5 models. Without significance testing, observations are more similar to the broad response in HadGEM2-AO than the more common SPG response, as in CESM1 or GFDL-CM2.1 [see Fig. 1 in Delworth et al. (2017)]. However, when we apply our significance test, the observed subpolar response is no longer significant at the 95% level (Fig. 2). The only statistically significant correlation we find in observations is a lagged warming at about 35°N (see Fig. S6 for HadSST).

Small signals are valuable if they can explain a large portion of variance or if they can be used to predict a system. The lagged SST response to the NAO has a limited impact on the AMV or  $\text{AMV}_{\text{MID}}$  index and pattern. We show this by linearly removing the lagged SST response to the NAO from the overall SST field. In each model, we first determine the lag of maximum correlation from the cross-correlation function of the AMV and NAO (i.e., those listed in Table 1). Then, at each pixel, we regress the low-pass-filtered SST field on the lagged, low-pass-filtered NAO index, where the lag is the lag of maximum correlation. Next, we multiply the map of regression coefficients by the NAO index, which creates a time series that isolates the lagged, linear temperature response to the NAO. Finally, we subtract our constructed SST field from the original SST field. This yields a temperature field that should include all variability except for that associated with the lagged SST response to the NAO. It will include the contemporaneous SST response to NAO winds. We then consider the AMV and  $\text{AMV}_{\text{MID}}$  indices and patterns from this adjusted SST dataset.

The linear lagged SST response to the NAO only accounts for between 0% and 12% of the variance in the AMV index in PI control runs of CMIP5 models (4% on average) and 19% (27%) in ERSST (HadSST) (see Table 1). Similarly, the lagged response to the NAO explains less than 11% of  $\text{AMV}_{\text{MID}}$  variance in all models, while it explains 31% (69%) in ERSST (HadSST). For all models, the spatial pattern of the AMV pattern remains virtually unchanged by linearly removing the lagged response to the NAO (not shown). Conversely, the spatial pattern of the lagged SST response to the NAO varies across models (see Fig. S3).

Note that in observations this signal explains more variance in the AMV and  $\text{AMV}_{\text{MID}}$  indices; however, this method does not account for the reduced degrees of freedom available in observations or potential exogenous regressors (e.g., external forcing, atmospheric variability not captured by the NAO index). The relationship between observations and historically forced runs will be reported in the next section.

The decadal portion of the *contemporaneous* NAO explains little variance in the AMV index. On average, the unlagged, low-pass-filtered NAO index explains only ~5% and ~7% (2%) of the variability in the AMV index in control runs of CMIP5 models and ERSST (HadSST), respectively (see Table 1). The spatial structure of the low-frequency SST response to the low-frequency NAO appears to be somewhat model dependent. Areas near the Norwegian and Irminger Seas, where some of the largest responses to the NAO are found, appear to be particularly sensitive to model configuration (Fig. S4); the sign of the response to the NAO is not uniform across CMIP5 models.

In PI control runs, knowledge of the prior state of the NAO offers statistically significant, but small, improvements in predictability. We test the null hypothesis that the positive phase of the AMV follows the positive phase of the NAO by chance. For this null hypothesis, the AMV is a random variable with two potential outcomes: the AMV was warm (>1 standard deviation) or it was not warm. The expected probability that the AMV+ will follow the NAO+ (>1 standard deviation) a given number of times is defined by the binomial distribution,

$$f(k, n, p) = \binom{n}{k} p^k (1-p)^{n-k},$$

where  $k$  is the number of times the AMV+ follows the NAO+ by the lag of maximum correlation,  $n$  is the number of NAO+ years, and  $p$  is the climatological probability of the AMV+. The probability that the AMV+ will follow the NAO+ *at least*  $k$  times is defined by the cumulative distribution function,  $F(k, n, p)$ . These statistics are calculated for each model individually. For 20 out of 27 CMIP5 PI control runs, we reject the null hypothesis that the AMV+ followed the NAO+ as a result of chance at the 90% level. On average, there is an 11.2% chance that the AMV+ followed the NAO+ by chance in CMIP5 PI runs. For comparison to the CESM-LE, we randomly subsection our run of CESM-CAM5 (PI) into 41 85-yr segments and find that in 26 out of 41 segments, we reject the null hypothesis that the AMV+ followed the NAO+ as a result of chance at the 90% level. On average, there is a 13.3% chance that the AMV+ followed the NAO+ by chance in this synthetic

PI ensemble. (If we increase our segment length to 153 yr, to match observations, we will reject our null hypothesis for all members of the synthetic ensemble.) Overall, in 46 out of 68 CMIP5 (PI) ensemble and synthetic ensemble members we can reject the null hypothesis that the AMV+ followed the NAO+ by chance. Thus in these idealized model runs with prescribed constant forcing, knowledge of the prior state of the NAO does improve our ability to predict the AMV index.

*b. NAO–AMV relationship is obscured in model runs with variable external forcing*

In historically forced runs, the lagged SST response to the NAO is inconsistent and rarely statistically significant. Only 5 of 27 CMIP5 historically forced runs produce a statistically significant lagged warm response to the NAO in the subpolar gyre (Fig. S5). Only 7 of 41 CESM-LE members produce positive lagged warm responses to the NAO in the subpolar gyre (Figs. S6 and S7). Both the magnitude and latitude of the warm response varies across both historically forced ensembles; correlation coefficients are generally near zero but can be as large as about 0.5. Further, the lag of maximum correlation between the NAO and AMV are inconsistent in historically forced runs; the largest correlations can be found throughout the range of lags that we consider (−30 to 30; see [Tables 1](#) and [2](#)). Finally, there is a wider distribution of variance explained by the lagged response to the NAO in historically forced runs than PI runs ([Tables 1](#) and [2](#)).

We find that in historically forced runs, knowledge of the prior state of the NAO does not typically offer statistically significant improvements in predictability for SST. Note that when discussing predictability, we only include historically forced runs where we find a lagged warm response to the NAO (i.e., the lag of maximum correlation is positive in [Table 1](#) or [Table 2](#)). Using the same test as for the PI, for 9 out of 18 CMIP5 historically forced runs we consider, we reject the null hypothesis at the 90% level that the AMV+ followed the NAO+ as a result of chance. On average, there is a 42.2% chance that the AMV+ followed the NAO+ by luck in this ensemble. Likewise, for 8 out of 29 CESM-LE runs we consider, we can reject the null hypothesis that the AMV+ followed the NAO+ as a result of chance. On average, there is a 40.0% chance that the AMV+ followed the NAO+ by luck in this ensemble. Compared to [section 3a](#), we find that the NAO offers more predictability in PI control runs than in historically forced runs. The probabilities reported above are generous, given we ignore the 23 model runs (combined) where the lag of maximum correlation is negative. Overall, for only

17 out of 68 historically forced model runs we can reject the null hypothesis that the AMV+ followed the NAO+ by chance. This is in contrast to the 46 out of 68 model runs for the PI where we can reject the hypothesis, which demonstrates the role for external forcing in this measure of predictability.

The inconsistency in the NAO–AMV relationship from historically forced runs becomes more apparent when compared to CMIP5 PI control runs. The PI control runs discussed in earlier sections exhibit a reasonably robust set of cross-correlation functions; most models produce a small, statistically significant, lagged warm response to the NAO that is attributed to variable ocean heat transport or ocean noise ([Fig. 3](#), top) ([Delworth et al. 2017](#); [Zhang 2017](#)). Conversely, there does not appear to be any systematic lead/lag relationship between the NAO and AMV in historically forced CMIP5 models; both warm and cold events precede and follow NAO+ events ([Fig. 3](#), bottom). For example, between lags of 6–13 yr, all CMIP5 PI control runs show a warm or neutral response to the NAO (correlation coefficients greater than zero), whereas in the CMIP5 historical ensemble six models have a cold lagged response to the NAO. While the NAO–AMV relationship appears to be consistent within PI runs, different model physics will simulate different relationships so there is more spread in the CMIP5 PI ensemble ([Fig. 3](#), top) than in the CESM-CAM5 PI synthetic ensemble ([Fig. 4](#), top), created by randomly subsampling the 899-yr PI control run of CESM-CAM5 into 85-yr sections.

CESM-CAM5 is one of the models that produces a lagged relationship between the NAO and AMV in PI control runs ([Fig. 4](#), top). This synthetic ensemble shows that the relationship between the AMV and NAO is consistent throughout the length of the run and is well-defined in short time series. In contrast, members of the historically forced CESM-LE produce a diverse set of cross-correlation functions ([Fig. 4](#), bottom). In both CMIP5 and CESM-LE model runs that include variable external forcing, the AMV-NAO relationship is interrupted, obscured, or overwhelmed. Further, the ocean circulation response to the NAO in CESM-LE is also inconsistent with the PI, as shown by the cross-correlation functions of the NAO with volume transport and OHT (Figs. S1 and S2). The comparison of CMIP5 PI runs to CMIP5 historical runs shows that this result is independent of a specific model's physics parameterizations ([Fig. 3](#)). Combined with the results from CESM-LE, it appears that the large influence of variable external forcing on internal variability sharply diminishes the AMV-NAO relationship seen in the PI runs.

TABLE 2. As in Table 1, but for 41 members of the CESM-LE.

CESM-LE	Lag of maximum correlation	% AMV <sub>MID</sub> variance explained by NAO <sub>t=0</sub>	% AMV variance explained by NAO <sub>t=0</sub>	% AMV <sub>MID</sub> variance explained by lagged response	% AMV variance explained by lagged response
002	19	8%	9%	1%	10%
003	25	0%	26%	0%	4%
004	30	17%	42%	21%	13%
005	6	16%	24%	56%	52%
006	11	3%	5%	22%	15%
007	30	35%	1%	2%	50%
008	30	0%	2%	31%	47%
009	12	9%	15%	23%	61%
010	3	40%	43%	0%	22%
011	-25	15%	19%	15%	19%
012	25	1%	8%	39%	72%
013	30	1%	24%	56%	24%
014	-8	6%	0%	15%	40%
015	-30	19%	25%	0%	5%
016	14	6%	11%	57%	43%
017	30	0%	0%	62%	43%
018	13	7%	21%	16%	18%
019	-23	0%	20%	1%	23%
020	30	1%	4%	26%	17%
021	-12	0%	15%	23%	10%
022	23	10%	22%	19%	29%
023	30	16%	8%	20%	63%
024	12	4%	15%	70%	42%
025	10	0%	2%	8%	30%
026	19	9%	0%	30%	74%
027	21	1%	4%	64%	50%
028	19	0%	30%	1%	14%
029	8	11%	1%	20%	17%
030	-10	16%	1%	22%	38%
031	2	0%	3%	12%	26%
032	14	1%	3%	50%	8%
033	-30	20%	68%	0%	7%
034	-11	2%	0%	36%	19%
035	28	21%	2%	7%	42%
101	16	5%	0%	67%	39%
102	4	39%	22%	14%	35%
103	-16	1%	9%	1%	0%
104	-28	77%	74%	46%	46%
105	29	25%	61%	42%	49%
106	12	17%	41%	12%	10%
107	19	6%	16%	63%	48%
CESM-LE average	9.29	11%	17%	26%	31%
CESM-LE std dev	18.76	0.15	0.19	0.22	0.19

We can demonstrate how a strong forced response can interrupt a weak NAO–SST relationship with a simple illustrative statistical model in the form:

$$\text{AMV}_{\text{LP}} = \beta_1 \text{NAO}_{\text{LP},t=0} + \beta_2 \text{NAO}_{\text{LP},t=10} + \beta_3 \cos \omega t,$$

where the regression coefficients are derived empirically from relationships in climate models, not theoretically from first principals. In this model,  $\beta_1$  is the contemporaneous cool SST response to the positive phase of the

NAO. We calculate this term via a linear regression of the low-pass-filtered AMV on the low-pass-filtered *standardized* NAO in the PI control run of CESM-CAM5 (yielding a value of  $-0.05^{\circ}\text{C}$  per standard deviation of the NAO index). The term  $\beta_2$  captures the lagged warm SST response to the NAO. We calculate this term via a separate linear regression of the low-pass-filtered AMV on the low-pass-filtered NAO, lagged 10 years (yielding a value of  $0.10^{\circ}\text{C}$  per standard deviation of the NAO index). These regression coefficients are

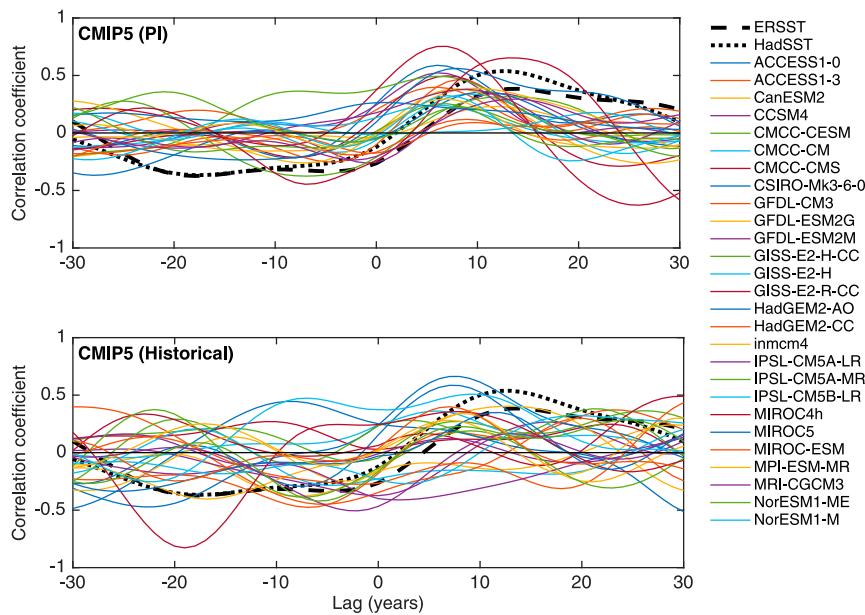


FIG. 3. Cross-correlation functions between the low-pass-filtered NAO index and the low-pass-filtered AMV index. The NAO leads at positive lags; the AMV leads at negative lags. The dashed black line is observations from ERSST, the dotted black line is observations from HadSST; both are the same in each panel. (top) CMIP5 preindustrial control runs. For each run, the full available length of the index time series is used in order to utilize the maximum amount of information. (bottom) CMIP5 historically forced experimental runs. For each run, the full available length of the index time series is used.

small, which mirrors the results presented earlier in this paper. Finally,  $\beta_3$  reflects the role of forcing in producing the AMV. This term is calculated by regressing the ensemble mean low-pass-filtered AMV on the ensemble mean low-pass-filtered NAO in CESM-LE (yielding a value of  $-0.68^{\circ}\text{C}$  per standard deviation of the NAO index). As noted above, this model is illustrative only and not meant to imply that the relationship between the NAO and AMV is strictly linear.

Given these coefficients, we use a series of low-pass-filtered normally distributed random values in place of the NAO. This is a reasonable approximation because the slopes of NAO spectra in climate models are typically flat and are unaffected by ocean model configuration (Peings et al. 2016; Kim et al. 2018). Additionally, as mentioned above, external forcing has little influence on the NAO in CESM-CAM5 (i.e., the ensemble mean NAO is near zero at all times). Low-pass filtering the random series is required to mimic both the memory within the climate system and (more importantly) the autocorrelation induced by filtering in the metrics used above. In the third term of our model, we choose to represent variable external forcing with a cosine function. Our qualitative results are not sensitive to the functional form (e.g., cosine “forcing,” linear trend, ensemble mean forcing) or periodicity of the forcing.

The amplitude of the forcing term is equal to the amplitude implied by our random “NAO” time series. Although we set the forcing term to be equal in magnitude to the forced response in the models, within the context of the simple model, this term could represent any (or a combination of) signals outside of the NAO–AMV system, including ocean heat transport that is unrelated to either the NAO or external forcing (O'Reilly et al. 2016).

A 42-member ensemble of this simple model shows that external forcing can obscure the linear portion of the NAO–AMV relationship (Fig. 5). We exclude the influence of external forcing by setting  $\beta_3$  to zero. In this configuration, we are able to produce a caricature of the cross-correlation function between the NAO and AMV exhibited by PI control runs of climate model (cf. orange lines to Fig. 4, top). If anything, the correlation coefficients produced within our simple model are too large, with values of nearly 1 at the lag of maximum correlation compared to a maximum of about 0.6 in CESM-LE. Including the  $\beta_3$  term obscures the NAO–AMV relationship just like in externally forced climate models (cf. the blue lines in Fig. 5 to Figs. 3 and 4, bottom). If external forcing is a large contributor to the AMV, the small relationship between the AMV and NAO can be overwhelmed.

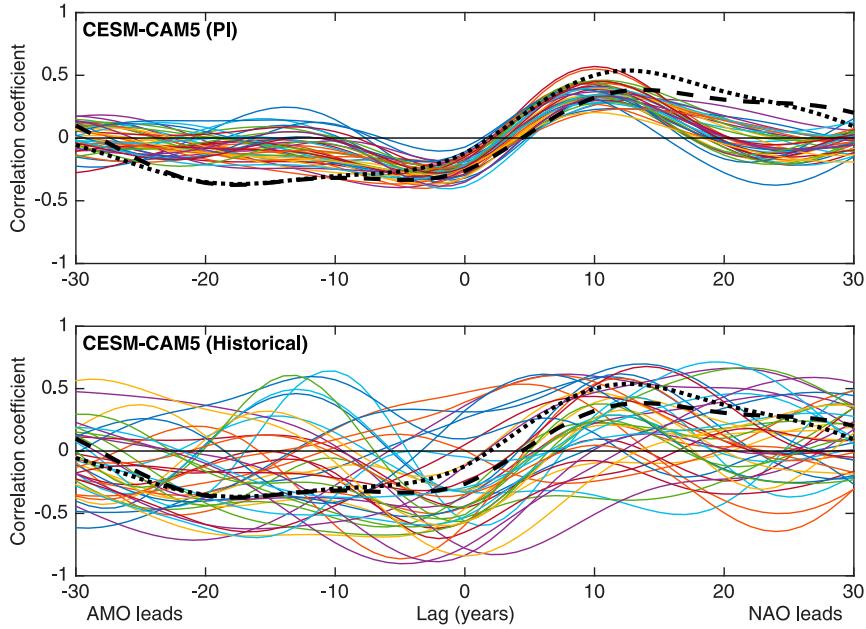


FIG. 4. Cross-correlation functions between the low-pass-filtered NAO index and the low-pass-filtered AMV index. The NAO leads at positive lags; the AMV leads at negative lags. The dashed black line is observations from ERSST; the dotted black line is observations from HadSST; both are the same in each panel. (top) A synthetic, 41-member, preindustrial “ensemble” created by randomly subsampling 85-yr segments of the long CESM-CAM5 PI run. (bottom) 41 members of the CESM-LE. Note that, each time series was linearly detrended; the ensemble average was not removed. Removing the ensemble average does not restore the NAO–AMV relationship from PI control runs.

In this statistical model, the magnitude of expected correlations between the NAO and AMV varies with sample size. That is, the width of the envelope of cross-correlation function is inversely proportional to the length of the time series; shorter time series produce larger correlations while longer time series produce correlations closer to zero (cf. Fig. 5 top to Fig. 5 bottom).

The NAO–AMV relationship in observations appears to be more like that found in historically forced runs than in PI control runs. The range of potential lagged SST responses to the NAO is larger in historically forced runs than in PI runs in terms of standard deviations and the range of spatial responses (see Tables 1 and 2, Fig. 2, and Figs. S5–S7 in the supplemental material). As noted in the literature, there are discrepancies between observations and PI control runs (e.g., Delworth et al. 2017). However, observations fit comfortably within the range of expected relationships from historically forced model runs. For example, both the lag of maximum correlation and the variance explained by the lag response fit this description (Fig. 6). Additionally, the observed cross-correlation between the NAO and AMV appears to be an outlier when compared to the PI

control runs but fit well within the range of expected relationships produced by our historical ensembles. Note that observations from the ERSST dataset yield a very similar response to the observed NAO as those from the HadSST dataset.

#### 4. Discussion

In many PI control runs, a prolonged positive phase of the NAO is associated with statistically significant warm anomalies in the subpolar gyre at about a 10-yr lag. These anomalies appear to rely on an import of warm water from the south, a mechanism consistent with prior work (O'Reilly et al. 2016; Delworth et al. 2017; Kim et al. 2018; Foukal and Lozier 2018). Knowledge of the prior state of the NAO can offer statistically significant increases in predictability of the AMV relative to climatology. However, the lagged correlation between the NAO and AMV index is small; the amount of variance explained by lagged low-pass-filtered NAO is reasonably consistent across PI control runs but is always less than 12% (Table 1). This is because the lagged SST response to the NAO is one of multiple significant components of multidecadal North Atlantic climate

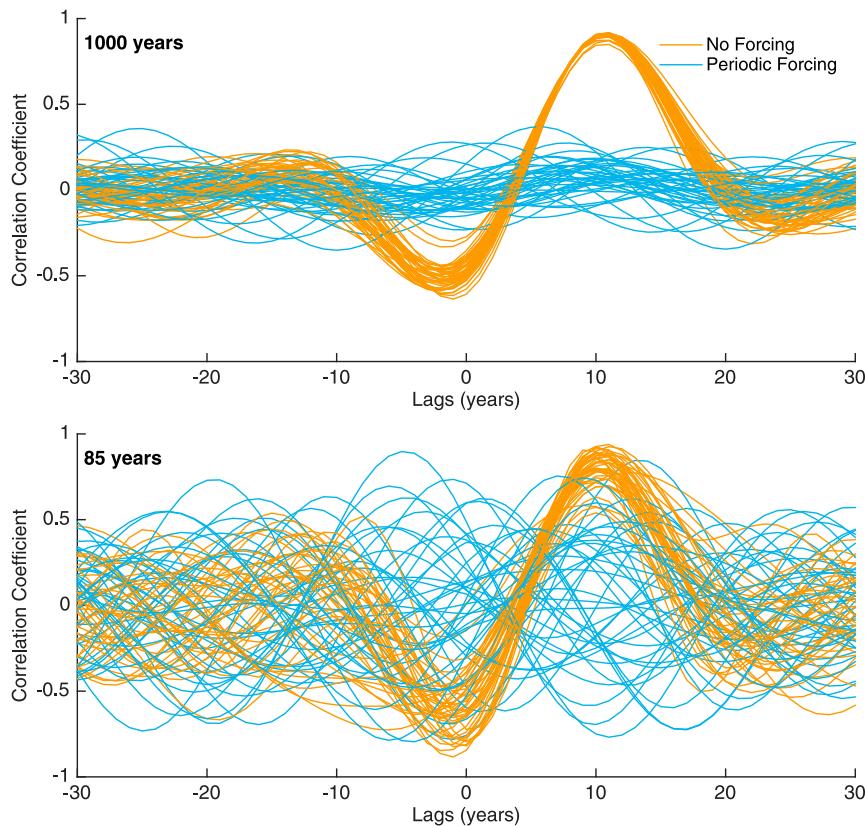


FIG. 5. Ensembles of a statistical model designed to illustrate the influence of external forcing on the lagged relationship between the NAO and AMV. The forcing in this simple model is a cosine function with a 30-yr period. (top) Ensemble of 1000-yr runs of the statistical model with and without the  $\beta_3$  term. (bottom) Ensemble of 85-yr runs of the statistical model with and without the  $\beta_3$  term.

variability. Our results are comparable to other studies that examine this relationship in PI control runs (e.g., [Delworth and Greatbatch 2000](#); [Ba et al. 2014](#); [Tandon and Kushner 2015](#); [Delworth et al. 2017](#)). Our interpretation differs when we compare to historically forced runs.

The relationship between the NAO and AMV is complicated by the inclusion of variable external forcing in climate models. The spatial pattern, timing, and magnitude of the lagged response become inconsistent when we account for external forcing. Because of this inconsistency, we generally find that in historically forced runs, the lagged, low-pass-filtered NAO does not offer predictability that is statistically significantly greater than climatology. In a simple model, we show how variable external forcing can obscure a prescribed relationship between the NAO and AMV. Our results can be compared to [Tandon and Kushner \(2015\)](#), who find a consistent relationship between AMOC and the AMV in PI control runs but an inconsistent relationship between AMOC and AMV in historically forced runs.

There are some notable discrepancies between the NAO–AMV relationship in PI control runs and observations. First, the warm SST response to the NAO appears to occur at a longer lag in observations (14 or 15 yr) than in CMIP5 PI control runs ( $9 \pm 5$  yr), as is also noted by [Delworth et al. \(2017\)](#). The observed lag of maximum correlation is statistically significantly different from that in our PI run of CESM1 at the 80% level (10–13 yr) when accounting for difference in the length of the two time series via a moving-block bootstrap. Next, the lagged NAO appears to explain a much larger amount of AMV variance in observations (19% or 27%) than in CMIP5 PI control runs (0%–12%). The observed percentage of AMV variance explained by the lagged NAO is statistically significantly different from that in our PI run of CESM1 at the 99% level (1.6%–4.5%) when again accounting for the difference in the length of the two time series via a moving-block bootstrap. Finally, unlike PI control runs, in observations there appear to be a cool anomaly preceding the NAO.

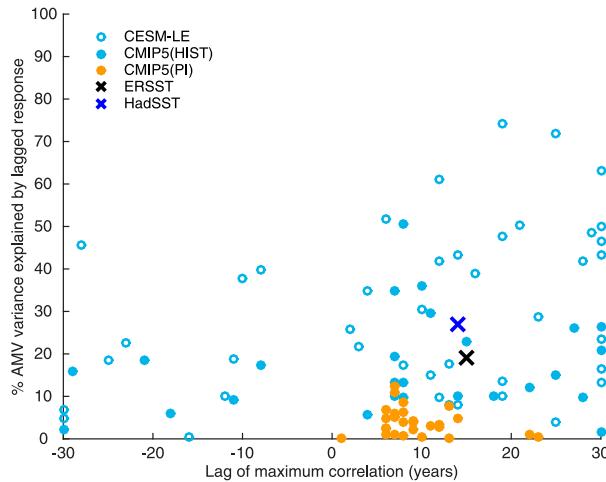


FIG. 6. Scatterplot of the lag of maximum correlation between the low-pass-filtered NAO index and the low-pass-filtered AMV index vs the percentage of AMV variance explained by the lagged response to the NAO. Blue circles represent the responses from historically forced models in both the CMIP5 ensemble (filled) and the CESM-LE (unfilled). Orange points represent responses from the CMIP5 PI ensemble described in text. Crosses represent the two observational datasets considered.

The NAO–AMV relationship in observations is more difficult to statistically distinguish from that in our historically forced ensembles. The observed lag of maximum correlation between the NAO and AMV fits within the wide distribution calculated from CMIP5 historically forced runs ( $8 \pm 17$  yr) or CESM-LE ( $9 \pm 19$  yr). Similarly, the amount of AMV variance explained by the lagged NAO fits within the large range calculated from CMIP5 historically forced runs ( $17\% \pm 11\%$ ) and CESM-LE ( $31\% \pm 19\%$ ). The widening of the distribution of the NAO–AMV relationship from PI runs to historically forced runs is well illustrated by the cross-correlation functions presented in Figs. 3 and 4. The cross-correlation function of the observed NAO–AMV relationship is outside of the range of the cross correlations simulated by PI control runs but fits within the wide envelope of NAO–AMV relationships simulated by historically forced runs. If observations are more like historically forced runs than PI runs, our interpretation of the observed record may change. For example, the aforementioned, observed cold anomalies that precede the NAO+ may not be the linear response to a prior NAO–, but rather could have been the result of other influences on the AMV (cf. Fig. 2 and Figs. S5–S7).

There is mounting evidence to suggest that variable external forcing is vital to understanding the AMV; Bellomo et al. (2018) estimate external forcing to be responsible for about 2/3 of total AMV variability in

CESM-CAM5. Our illustrative linear model suggests that the prior state of the NAO will only offer predictability for the AMV when  $<16\%$  of AMV variance is forced ( $\beta_3 < 0.4$ ). A large influence of external forcing on the AMV may suggest a pessimistic view of this mechanism for predictability, but these results do not rule out the possibility that the NAO can provide predictability on an episodic basis. For example, numerous studies have shown that knowledge of the initial ocean state improves the predictability of North Atlantic SSTs and upper-ocean heat content, particularly for the mid-1990s warming (Latif et al. 2004; Yeager et al. 2012; Robson et al. 2012; Msadek et al. 2014; Karspeck et al. 2015). Our results, however, show from a statistical point of view that the lagged response to the NAO does not always cause changes in the AMV.

We note that this analysis relies primarily on climate models. While these models are an excellent tool for research, they remain imperfect. Individual models may respond to external forcing with too much vigor (e.g., Zhang et al. 2013). On the other hand, most models appear to undersimulate internal variability (e.g., the NAO), as compared to external forcing (e.g., Kim et al. 2018). The ocean circulation may respond to different modes of atmospheric forcing depending on each model's primary site of deep-water formation; for example, the East Atlantic Pattern is found to be the primary driver of AMOC in IPSL-CM4 (Msadek and Frankignoul 2009). Put generally, the NAO is a convenient and useful index but other modes of atmospheric variability may more efficiently produce lagged SST anomalies (Branstator and Gritsun 2017). Finally, both the strength and the variability of AMOC may not be realistically simulated in all models (Zhang and Wang 2013; Buckley and Marshall 2016; Heuzé 2017; Yan et al. 2018). All of these caveats likely affect our comparison between models and observations.

Real-world external forcing complicates our ability to interpret the observed NAO–AMV relationship. The statistical relationship between the NAO and AMV in a historically forced ensemble appears to be different from that found in PI control runs. When we account for variable external forcing in climate models, there is no consistent or obvious connection between the two climate indices. If we consider historically forced climate models to provide plausible counterfactuals for our observed climate, we can infer that many other NAO–AMV relationships could have been possible given the length of the observed time series. While there are large discrepancies in the timing and structure of the lagged SST response to the NAO between observations and PI control runs, observations fall well within the wide distribution of relationships produced by the CMIP5

historical ensemble, CESM-LE, and a simple statistical model that includes variable external forcing.

**Acknowledgments.** This research was supported by the National Science Foundation Climate and Large-Scale Dynamics program (A. C. Clement) and the University of Miami Graduate School (J. M. Klavans). This research was conducted in collaboration with and using the resources of the University of Miami Center for Computational Science. We acknowledge the World Climate Research Program's Working Group on Coupled Modeling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their model output. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. NOAA ERSSTv3 data are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, and are publicly available from their Web site at <https://www.esrl.noaa.gov/psd/>. We also acknowledge the CESM Large Ensemble Community Project and supercomputing resources provided by NSF/CISL/Yellowstone. Finally, we thank the three anonymous reviewers for their valuable contributions; they have assuredly improved the quality and clarity of this manuscript.

## REFERENCES

Ba, J., and Coauthors, 2014: A multi-model comparison of Atlantic multidecadal variability. *Climate Dyn.*, **43**, 2333–2348, <https://doi.org/10.1007/s00382-014-2056-1>.

Bellomo, K., L. N. Murphy, M. A. Cane, A. C. Clement, and L. M. Polvani, 2018: Historical forcings as main drivers of the Atlantic multidecadal variability in the CESM large ensemble. *Climate Dyn.*, **50**, 3687–3698, <https://doi.org/10.1007/s00382-017-3834-3>.

Bellucci, A., A. Mariotti, and S. Gualdi, 2017: The role of forcings in the twentieth-century North Atlantic multidecadal variability: The 1940–75 North Atlantic cooling case study. *J. Climate*, **30**, 7317–7337, <https://doi.org/10.1175/JCLI-D-16-0301.1>.

Bjerknes, J., 1964: Atlantic air-sea interaction. *Advances in Geophysics*, Vol. 10, Academic Press, 1–82, [https://doi.org/10.1016/S0065-2687\(08\)60005-9](https://doi.org/10.1016/S0065-2687(08)60005-9).

Booth, B. B. B., N. J. Dunstone, P. R. Halloran, T. Andrews, and N. Bellouin, 2012: Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, **484**, 228–232, <https://doi.org/10.1038/nature10946>.

Branstator, G., and A. Gritsun, 2017: Alternatives to NAO driving of AMOC variability in CCSM4. CESM Ocean Modeling Working Group Winter Working Group Meeting, Boulder, CO, NCAR, <http://www.cesm.ucar.edu/events/wg-meetings/2017/presentations/omwg/branstator.pdf>.

Buckley, M. W., and J. Marshall, 2016: Observations, inferences, and mechanisms of the Atlantic Meridional Overturning Circulation: A review. *Rev. Geophys.*, **54**, 5–63, <https://doi.org/10.1002/2015RG000493>.

—, R. M. Ponte, G. Forget, and P. Heimbach, 2015: Determining the origins of advective heat transport convergence variability in the North Atlantic. *J. Climate*, **28**, 3943–3956, <https://doi.org/10.1175/JCLI-D-14-00579.1>.

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>.

Danabasoglu, G., S. G. Yeager, Y.-O. Kwon, J. J. Tribbia, A. S. Phillips, and J. W. Hurrell, 2012: Variability of the Atlantic meridional overturning circulation in CCSM4. *J. Climate*, **25**, 5153–5172, <https://doi.org/10.1175/JCLI-D-11-00463.1>.

Delworth, T., S. Manabe, and R. J. Stouffer, 1993: Interdecadal variations of the thermohaline circulation in a coupled ocean–atmosphere model. *J. Climate*, **6**, 1993–2011, [https://doi.org/10.1175/1520-0442\(1993\)006<1993:IVOTTC>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<1993:IVOTTC>2.0.CO;2).

—, and R. J. Greatbatch, 2000: Multidecadal thermohaline circulation variability driven by atmospheric surface flux forcing. *J. Climate*, **13**, 1481–1495, [https://doi.org/10.1175/1520-0442\(2000\)013<1481:MTCVDB>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1481:MTCVDB>2.0.CO;2).

—, and M. E. Mann, 2000: Observed and simulated multidecadal variability in the Northern Hemisphere. *Climate Dyn.*, **16**, 661–676, <https://doi.org/10.1007/s003820000075>.

—, F. Zeng, L. Zhang, R. Zhang, G. A. 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](https://doi.org/10.1175/1520-0442(1993)006<1743:SCVOTN>2.0.CO;2).

—, M. A. Alexander, 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>.

Foukal, N. P., and M. S. Lozier, 2018: Examining the origins of ocean heat content variability in the eastern North Atlantic subpolar gyre. *Geophys. Res. Lett.*, **45**, 11 275–11 283, <https://doi.org/10.1029/2018GL079122>.

Frankignoul, C., and K. Hasselmann, 1977: Stochastic climate models, Part II Application to sea-surface temperature anomalies and thermocline variability. *Tellus*, **29**, 289–305, <https://doi.org/10.3402/tellusa.v29i4.11362>.

—, 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>.

Gulev, S. K., M. Latif, N. Keenlyside, W. Park, and K. P. Koltermann, 2013: North Atlantic Ocean control on surface heat flux on multidecadal timescales. *Nature*, **499**, 464–467, <https://doi.org/10.1038/nature12268>.

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>.

Hasselmann, K., 1976: Stochastic climate models Part I. Theory. *Tellus*, **28**, 473–485, <https://doi.org/10.3402/tellusa.v28i6.11316>.

Heuzé, C., 2017: North Atlantic Deep Water formation and AMOC in CMIP5 models. *Ocean Sci.*, **13**, 609–622, <https://doi.org/10.5194/os-13-609-2017>.

Hurrell, J., and National Center for Atmospheric Research Staff, 2018: The Climate Data Guide: Hurrell North Atlantic Oscillation (NAO) Index (station-based). NCAR, accessed 15 December 2017, <https://climatedataguide.ucar.edu/climate-data/hurrell-north-atlantic-oscillation-nao-index-station-based>.

Jones, P. D., T. J. Osborn, and K. R. Briffa, 2003: Pressure-based measures of the North Atlantic Oscillation (NAO): A comparison and an assessment of changes in the strength of the NAO and in its influence on surface climate parameters. *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, Geophys. Monogr., Vol. 134, Amer. Geophys. Union, 51–62, <https://doi.org/10.1029/134GM03>.

Jungclaus, J. H., H. Haak, M. Latif, and U. Mikolajewicz, 2005: Arctic–North Atlantic interactions and multidecadal variability of the meridional overturning circulation. *J. Climate*, **18**, 4013–4031, <https://doi.org/10.1175/JCLI3462.1>.

Karspeck, A., S. Yeager, G. Danabasoglu, and H. Teng, 2015: An evaluation of experimental decadal predictions using CCSM4. *Climate Dyn.*, **44**, 907–923, <https://doi.org/10.1007/s00382-014-2212-7>.

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>.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *J. Geophys. Res.*, **116**, D14104, <https://doi.org/10.1029/2010JD015220>.

Kerr, R. A., 2000: A North Atlantic climate pacemaker for the centuries. *Science*, **288**, 1984–1985, <https://doi.org/10.1126/science.288.5473.1984>.

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>.

Knight, J. R., R. J. Allan, C. K. Folland, M. Vellinga, and M. E. Mann, 2005: A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophys. Res. Lett.*, **32**, L20708, <https://doi.org/10.1029/2005GL024233>.

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](https://doi.org/10.1175/1520-0442(1994)007<0141:IVINAS>2.0.CO;2).

Latif, M., and N. S. Keenlyside, 2011: A perspective on decadal climate variability and predictability. *Deep-Sea Res. II*, **58**, 1880–1894, <https://doi.org/10.1016/j.dsr2.2010.10.066>.

—, and Coauthors, 2004: Reconstructing, monitoring, and predicting multidecadal-scale changes in the North Atlantic thermohaline circulation with sea surface temperature. *J. Climate*, **17**, 1605–1614, [https://doi.org/10.1175/1520-0442\(2004\)017<1605:RMAPMC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<1605:RMAPMC>2.0.CO;2).

Lozier, M. S., 2010: Deconstructing the conveyor belt. *Science*, **328**, 1507–1511, <https://doi.org/10.1126/science.1189250>.

Msadek, R., and C. Frankignoul, 2009: Atlantic multidecadal oceanic variability and its influence on the atmosphere in a climate model. *Climate Dyn.*, **33**, 45–62, <https://doi.org/10.1007/s00382-008-0452-0>.

—, and Coauthors, 2014: Predicting a decadal shift in North Atlantic climate variability using the GFDL Forecast System. *J. Climate*, **27**, 6472–6496, <https://doi.org/10.1175/JCLI-D-13-00476.1>.

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>.

O'Reilly, C. H., and L. Zanna, 2018: The signature of oceanic processes in decadal extratropical SST anomalies. *Geophys. Res. Lett.*, **45**, 7719–7730, <https://doi.org/10.1029/2018GL079077>.

—, M. Huber, T. Woollings, and L. Zanna, 2016: The signature of low-frequency oceanic forcing in the Atlantic Multidecadal Oscillation. *Geophys. Res. Lett.*, **43**, 2810–2818, <https://doi.org/10.1002/2016GL067925>.

Otterå, O. H., M. Bentsen, H. Drange, and L. Suo, 2010: External forcing as a metronome for Atlantic multidecadal variability. *Nat. Geosci.*, **3**, 688–694, <https://doi.org/10.1038/ngeo955>.

Peings, Y., 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>.

Piecuch, C. G., R. M. Ponte, C. M. Little, M. W. Buckley, and I. Fukumori, 2017: Mechanisms underlying recent decadal changes in subpolar North Atlantic Ocean heat content. *J. Geophys. Res. Oceans*, **122**, 7181–7197, <https://doi.org/10.1002/2017JC012845>.

Robson, J. I., R. T. Sutton, and D. M. Smith, 2012: Initialized decadal predictions of the rapid warming of the North Atlantic Ocean in the mid 1990s. *Geophys. Res. Lett.*, **39**, L19713, <https://doi.org/10.1029/2012GL053370>.

Schlesinger, M. E., and N. Ramankutty, 1994: An oscillation in the global climate system of period 65–70 years. *Nature*, **367**, 723–726, <https://doi.org/10.1038/367723a0>.

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>.

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

Thompson, D. W. J., and J. M. Wallace, 2001: Regional climate impacts of the Northern Hemisphere annular mode. *Science*, **293**, 85–89, <https://doi.org/10.1126/science.1058958>.

Timmermann, A., M. Latif, R. Voss, and A. Grötzner, 1998: Northern Hemispheric interdecadal variability: A coupled air–sea mode. *J. Climate*, **11**, 1906–1931, <https://doi.org/10.1175/1520-0442-11.8.1906>.

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>.

Vecchi, G. A., T. L. Delworth, and B. Booth, 2017: Climate science: Origins of Atlantic decadal swings. *Nature*, **548**, 284–285, <https://doi.org/10.1038/nature23538>.

Visbeck, M., H. Cullen, G. Krahmann, and N. Naik, 1998: An ocean model's response to North Atlantic Oscillation-like wind forcing. *Geophys. Res. Lett.*, **25**, 4521–4524, <https://doi.org/10.1029/1998GL900162>.

Wills, R. C. J., K. C. Armour, D. S. Battisti, and D. L. Hartmann, 2019: Ocean–atmosphere dynamical coupling fundamental to the Atlantic multidecadal oscillation. *J. Climate*, **32**, 251–272, <https://doi.org/10.1175/JCLI-D-18-0269.1>.

Woollings, T., 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](https://doi.org/10.1175/1520-0477(1999)080<0245:TIOSCR>2.0.CO;2).

Yan, X., R. Zhang, and T. R. Knutson, 2018: Underestimated AMOC variability and implications for AMV and predictability in CMIP models. *Geophys. Res. Lett.*, **45**, 4319–4328, <https://doi.org/10.1029/2018GL077378>.

Yeager, S., 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>.

Zhang, L., and C. Wang, 2013: Multidecadal North Atlantic sea surface temperature and Atlantic meridional overturning circulation variability in CMIP5 historical simulations. *J. Geophys. Res. Oceans*, **118**, 5772–5791, <https://doi.org/10.1002/jgrc.20390>.

Zhang, R., 2017: On the persistence and coherence of subpolar sea surface temperature and salinity anomalies associated with the Atlantic multidecadal variability. *Geophys. Res. Lett.*, **44**, 7865–7875, <https://doi.org/10.1002/2017GL074342>.

—, and Coauthors, 2013: Have aerosols caused the observed Atlantic multidecadal variability? *J. Atmos. Sci.*, **70**, 1135–1144, <https://doi.org/10.1175/JAS-D-12-0331.1>.