Assessing External and Internal Sources of Atlantic Multidecadal Variability Using Models, Proxy Data, and Early Instrumental Indices

CHRISTOPHER H. O'REILLY, LAURE ZANNA, AND TIM WOOLLINGS

Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, United Kingdom

ABSTRACT

Atlantic multidecadal variability (AMV) of sea surface temperature exhibits an important influence on the climate of surrounding continents. It remains unclear, however, the extent to which AMV is due to internal climate variability (e.g., ocean circulation variability) or changes in external forcing (e.g., volcanic/anthropogenic aerosols or greenhouse gases). Here, the sources of AMV are examined over a 340-yr period using proxy indices, instrumental data, and output from the Last Millennium Ensemble (LME) simulation. The proxy AMV closely follows the accumulated atmospheric forcing from the instrumental North Atlantic Oscillation (NAO) reconstruction ($r = 0.65$)—an "internal" source of AMV. This result provides strong observational evidence that much of the AMV is generated through the oceanic response to atmospheric circulation forcing, as previously demonstrated in targeted modeling studies. In the LME there is a substantial externally forced AMV component, which exhibits a modest but significant correlation with the proxy AMV (i.e., $r = 0.37$), implying that at least 13% of the AMV is externally forced. In the LME simulations, however, the AMV response to accumulated NAO forcing is weaker than in the proxy/observational datasets. This weak response is possibly related to the decadal NAO variability, which is substantially weaker in the LME than in observations. The externally forced component in the proxy AMV is also related to the accumulated NAO forcing, unlike in the LME. This indicates that the external forcing is likely influencing the AMV through different mechanistic pathways: via changes in radiative forcing in the LME and via changes in atmospheric circulation in the observational/proxy record.

1. Introduction

In situ observations of North Atlantic sea surface temperatures (SSTs) have been made from the midnineteenth century to the present day. In this ~150-yr observational period, North Atlantic SSTs exhibit significant variability on time scales of decades and longer; this is often referred to as the Atlantic multidecadal variability (AMV) or Atlantic multidecadal oscillation (e.g., Kushnir 1994; Enfield et al. 2001; Deser et al. 2010). The AMV has been shown to influence decadal climate in the surrounding continental regions, including North America (McCabe et al. 2004; Sutton and Hodson 2005; Knight et al. 2006; Ting et al. 2009; Hu and Feng 2012; Ruprich-Robert et al. 2018), Europe (Sutton and Hodson 2005; Knight et al. 2006; Sutton and Dong 2012; O'Reilly et al. 2017; Ruprich-Robert et al. 2017; Qasmi et al. 2017; Ghosh et al. 2017; Årthun et al. 2018), and the Sahel (Palmer 1986; Folland et al. 1986; Zhang and Delworth 2006; Martin et al. 2014; Martin and Thorncroft 2014). The AMV has also been linked to decadal variability in hurricane frequency (e.g., Zhang and Delworth 2006; Yan et al. 2017) and Arctic sea ice (e.g., Zhang 2015). The influence the AMV exhibits on the surrounding climate means it is an important source of potential climate predictability and future decadal variability.

The AMV has been linked to variability in ocean circulation and variability in external forcing, as well as other generation mechanisms. Many studies have shown that the AMV in general circulation models is associated with changes in ocean circulation—and more specifically, changes in ocean heat flux convergence (e.g., Delworth and Mann 2000; Knight et al. 2005). This has been associated with changes in the Atlantic meridional overturning circulation (AMOC; e.g., Danabasoglu 2008; Zhang and Wang 2013; Wills et al. 2019), but the AMV has also been attributed to changes in horizontal gyre circulations, particularly in the extratropical part of the AMV (Williams et al. 2014, 2015; Piecuch et al. 2017). The study by Clement et al. (2015) showed that similar North Atlantic SST variability is exhibited in climate models with a slab ocean model that does not feature ocean circulation variability. However, other studies have shown that ocean circulation is an important factor in reproducing the observed characteristics in the extratropics; the decadal SSTs in the extratropical North Atlantic are correlated with upward turbulent heat fluxes in observations and fully coupled models but not in slab ocean models (Gulev et al. 2013; O'Reilly et al. 2016; Zhang et al. 2016; Drews and Greatbatch 2016). Moreover, the observed influence of the AMV on surrounding continental regions, and particularly over Europe, is not reproduced in models without ocean circulation variability (O'Reilly et al. 2016; Sun et al. 2019).

Contrary to studies citing the importance of ocean circulation variability, several studies have demonstrated that changes in external forcing could be responsible for the observed AMV. Booth et al. (2012) demonstrated that, in a particular climate model, the indirect radiative effect of anthropogenic aerosols and their effect on downward shortwave radiation was responsible for much of the AMV over the twentieth century. However, Zhang et al. (2013) demonstrated that this particular simulation failed to capture some key features associated with the observed AMV, such as the upper-ocean heat content anomalies and the covariation of surface salinity anomalies (also see Zhang 2017). Nonetheless, models in phase 5 of the Coupled Model Intercomparison Project (CMIP5) historical ensemble, the same model generation as the single model study of Booth et al., do have a substantial externally forced component of the AMV (Murphy et al. 2017). In the CESM "Last Millennium Ensemble" simulations, Bellomo et al. (2018) showed that much of the externally forced AMV over the CMIP5 historical period is from anthropogenic aerosols, but with substantial contributions from decadal variations in greenhouse gas concentrations and volcanic aerosol emissions. However,

the influence of external forcing does not necessarily eliminate the role of ocean circulation in the AMV and indeed signatures of both external forcing and AMOC variability were found in the study of Tandon and Kushner (2015).

In contrast to studies demonstrating that external forcing has driven the observed AMV, there is substantial evidence that the AMV is an "internal" mode of the climate system. For example, in studies by Li et al. (2013) and McCarthy et al. (2015) it has been demonstrated that the observed AMV over the twentieth century closely follows the accumulated signal of the North Atlantic Oscillation (NAO). The ocean circulation changes as a result of the NAO forcing on decadal time scales, influencing heat content anomalies in the North Atlantic. In a recent modeling study, Delworth et al. (2017) showed that prescribing extratropical turbulent heat flux anomalies associated with decadal NAO variability was sufficient to generate the recent observed AMV changes, through changes in the overturning circulation. Kim et al. (2018b) also show evidence that ocean dynamics have played a key role in recent AMV behavior. Similarly, Robson et al. (2012) found that surface turbulent heat fluxes were responsible for the recent period of rapid North Atlantic warming in the mid-1990s. In a preindustrial control simulation, Sun et al. (2015) showed that the accumulated NAO leads the AMV and then there is a subsequent feedback that acts to generate substantial multidecadal variability. The NAO forcing of the AMV is present in most coupled models in the CMIP5 ensemble, in both historical and longer preindustrial control simulations, though the relationship is generally weaker than that seen in observations (O'Reilly and Zanna 2018; Peings et al. 2016). Since this mechanism is seen in preindustrial simulations (with constant external radiative forcing) the AMV driven by atmospheric forcing can be considered internal to the climate system. It is likely that intrinsic oceanic variability, which is not driven by accumulated NAO forcing, makes a substantial contribution to the oceanic circulation and thereby the AMV; however, it seems that this is lower in the extratropical North Atlantic where the atmospheric-forced variability dominates (Leroux et al. 2018). We refer to the AMV forced by anomalous atmospheric circulation as the "internal" AMV throughout this study.

A separation of the external forced AMV from the internal AMV is difficult to calculate using observations, primarily due to the relatively short observational period. To better understand the AMV and its behavior in the past, several studies have examined the presence of AMV using proxies. Gray et al. (2004) used a sparse tree-ring dataset to produce an AMV index from the

late 1500s to the present that exhibited distinct multi-decadal variability in the reconstructed period. More recently, Wang et al. (2017a) produced an annually resolved AMV proxy, which revealed that some of the multidecadal variability in the North Atlantic can be attributed to the response to volcanic eruptions, similar to the results of the modeling study of Otterå et al. (2010). However, a large amount of the AMV in the Wang et al. reconstruction cannot be attributed to external forcing.

In this study, we use proxy data to investigate the internally generated AMV and the externally forced AMV components. To do so, we analyze several proxy AMV reconstructions in combination with a reconstruction of sea level pressure (SLP) over the past ~350 years. The sea level pressure reconstruction is produced using early instrumental observations over Europe (Luterbacher et al. 2001). We use the proxy AMV datasets and reconstructed SLP to examine the internal AMV generated by anomalous atmospheric forcing. In addition, we analyze a long ensemble simulation over the same period to estimate the external forcing of the AMV in the proxy datasets. The datasets and methods are described in section 2. In section 3 we demonstrate that there is a large fraction of the proxy AMV that appears to be internally driven by atmospheric circulation anomalies; we also find evidence of an externally forced fraction of the proxy AMV. Conclusions follow in section 4.

2. Datasets and methods

a. Proxy AMV indices

To analyze the AMV over an extended period we utilize three available proxy-based indices. The AMV in all these indices is defined as the SST averaged over the entire North Atlantic Ocean (0°–70°N). The first, by Wang et al. (2017a), was produced using 46 annual-resolution terrestrial proxies including tree rings and ice cores that exhibit positive correlations with local temperatures and/or hydroclimate, from the Past Global Changes 2000 yr (PAGES 2k) Consortium database (Ahmed et al. 2013). Wang et al. used these proxy records to reconstruct the (nondetrended) AMV for the extended boreal summer season [May–September (MJJAS)], as many proxies are particularly sensitive to summer variability and the AMV exhibits a particularly large signal in summer temperatures (e.g., O'Reilly et al. 2017), though this is very similar to the annual AMV over the observational period. The AMV was reconstructed using a nested principal component regression method and was demonstrated to be well validated using a sliding window approach. It was also shown to be insensitive to

the various changes to the proxy composition and nesting method. The proxy reconstruction method was also validated using pseudoproxy climate experiments, where artificial proxies in a climate model are produced at the locations of the real proxies and are used to reconstruct the model AMV (e.g., Smerdon 2012). The Wang et al. proxy AMV index will be referred to as AMV-Wang herein.

The second proxy AMV index we analyze is from the study of Mann et al. (2009). In their study, Mann et al. used climate field reconstruction to produce a dataset of annual mean gridded temperatures. This temperature reconstruction is actually only resolved on decadal time scales, from which areal averages are taken to produce an AMV index. This dataset was produced using a set of proxies with global coverage, some of which were also used to produce AMV-Wang. However, the decadal resolution of this dataset reduces the degrees of freedom available over the instrumental period, which therefore reduces the confidence levels of the calibration and verification. In spite of these possible limitations, we will analyze the Mann et al. AMV index (AMV-Mann) since it provides an additional estimate over the proxy period.

The final proxy AMV index we will analyze is from the study of Gray et al. (2004). This index was produced using a sparse dataset of only 12 tree rings, with annual resolution, which are completely independent of the proxy datasets used to produce AMV-Wang and AMV-Mann. Gray et al. used a principal component reconstruction method to produce their AMV index, targeting detrended annual mean temperatures over the North Atlantic. However, the proxies were selected from a large tree-ring database based on the magnitude of their correlation (i.e., $|r| > 0.25$) with the observed AMV index over the instrumental period, rather than local temperatures as in the AMV-Wang and AMV-Mann reconstructions. Selecting the tree-ring proxies to include in the principle component reconstruction based on the observed AMV in this way leads to a selection bias that is likely to substantially overestimate the accuracy of the reconstruction (Miller 1984). For example, in the Gray et al. reconstruction, tree rings that are uncorrelated with local temperatures or precipitation but by chance have a reasonable correlation with the AMV would be included in the reconstruction. Despite the flaws in the proxy selection criteria, we nonetheless include the Gray et al. AMV index in our analysis (AMV-Gray). AMV-Gray is provided in terms of standard deviation, so is scaled by the instrumental AMV standard deviation ($\sigma = 0.19$ K) to be comparable with the other AMV indices.

We analyze the AMV-Wang and AMV-Mann indices over the period 1659–1999, the common period with the

NAO index (described in the following subsection), and the AMV-Gray index over the period 1659–1990 since AMV-Gray is not available after 1990. We will present analysis for all three indices but have the most confidence in AMV-Wang, for the reasons outlined above. In addition, we analyze the AMV index from observations, which is the nondetrended North Atlantic SST (0°–70°N) from the Kaplan SST dataset (downloaded from https://www.esrl.noaa.gov/psd/data/timeseries/AMO/), this is used over the period 1856–1999.

The three proxy AMV reconstructions disagree substantially in some parts or the analysis period. While effort is made as to the accuracy of the reconstructions in the source literature (Gray et al. 2004; Mann et al. 2009; Wang et al. 2017a), it is not clear that these uncertainties will be constant going back in time. The AMV reconstructions are considered to be more uncertain than the NAO index calculated from a sea level pressure reconstruction (see next section), since this is produced using instrumental data from barometers among other measurements.

b. NAO index from early instrumental data

To investigate the internally generated AMV in the proxy datasets we analyze the NAO, which is the dominant mode of large-scale atmospheric variability in the North Atlantic sector. To calculate the NAO index, we use the mean SLP reconstruction produced by Luterbacher et al. (2002). This dataset is a monthly $5° \times 5°$ gridded SLP dataset over Europe and the eastern North Atlantic (30°W–40°E; 30°–70°N) and is available from 1659 onward. The SLP reconstruction was produced primarily using early instrumental station time series of pressure, temperature, and precipitation. The station pressure time series are the most important input data for the SLP reconstructions owing to the large spatial scale of monthly SLP anomalies. The number of predictors in the reconstruction increases substantially from 1755 onward, such that the quality of the reconstruction is not constant throughout the dataset. Numerous other proxy NAO reconstructions are available that extend the NAO back for up to 1000 years, though there is a large uncertainty and spread across these proxy indices (Pinto and Raible 2012), so for this reason we only analyze the NAO derived from the instrumental-based SLP reconstruction.

The NAO time series is calculated as the SLP difference between two boxes, one over Iceland and one over the Azores (shown in Figs. 2a,b), with each box consisting of four grid points, following Luterbacher et al. (2001). We calculate the NAO index for all months, normalizing over the entire monthly time series, such that there is more NAO variability in the winter months

than in the summer months. This means that the winter months contribute more to the annual mean indices but the contribution from summer months is not neglected. The annual mean NAO index is computed from the monthly NAO index.

The NAO was calculated using this traditional gridpoint method rather than an empirical orthogonal function (EOF) method because the reconstructed SLP dataset is only provided over a limited spatial domain (30°W–40°E; 30°–70°N). We tested computing the NAO using the EOF method over this limited domain and compared this with the (EOF based) NAO index from Hurrell et al. (2003) from 1899 to 1999.[1] The gridpoint NAO exhibited a substantially higher correlation with the Hurrell NAO index than the NAO calculated from the EOF method ($r = 0.88$ and $r = 0.77$, respectively); therefore, the gridpoint-based NAO seems to give a better representation of the observed large-scale NAO.

c. LME simulations

To complement the proxy and instrumental data we also analyze data from the CESM Last Millennium Ensemble (LME; Otto-Bliesner et al. 2016). The LME was produced using the CESM climate model, running with 13 ensemble members from 850 to 2005. Each ensemble member is forced with the same transient evolution of solar intensity, volcanic emissions, greenhouse gases (GHG), ozone/aerosols, land-use conditions, and orbital parameters in an attempt to simulate the evolution of the climate system over the last millennium. The LME was run with a resolution of about $2° \times 2°$ resolution in the atmosphere and about $1° \times 1°$ resolution in the ocean. The 13-member ensemble means that the ensemble mean evolution of variables in the LME is expected to provide a good estimate of the externally forced mode response. In addition to the 13-member ensemble performed with all of the forcings, smaller ensembles (from 3 to 5 members) were performed with individual forcing components separately. Here we will analyze simulations with solar variability, volcanic emissions, GHGs, and ozone/aerosol forcings in addition to the full ensemble.

The AMV indices from the LME (AMV-LME) were calculated as the areal average of the annual mean SST anomalies over the North Atlantic (i.e., 0°–70°N), as for the observed indices. The NAO indices from the LME are calculated using the SLP difference over the same boxes as those used for the NAO index from the early

---

[1] The Hurrell principal component (PC)-based NAO index was downloaded from https://climatedataguide.ucar.edu/climate-data/hurrell-north-Atlantic-oscillation-NAO-index-pc-based.

instrumental data. The NAO indices are calculated for each ensemble member separately. The NAO calculated from the LME has the typical NAO pattern of anomalies with low pressure anomalies in the northern part of the North Atlantic basin and high pressure anomalies to the south, which is similar to that seen in the early instrumental NAO index (shown in Fig. S1 in the online supplemental material). The LME data is analyzed over 1659–1999 to compare with the proxy AMV indices and the instrumental NAO reconstruction.

d. CMIP5 historical simulations

In addition to the LME, we also analyze data from the historical simulations in the CMIP5 archive (Taylor et al. 2012). The CMIP5 historical ensemble is useful for putting the results from the LME into the context of other climate models. The first ensemble member for the 43 different models in which the historical simulations cover the period 1860–1999 is used. Similar to the LME, the CMIP5 historical simulations include the transient evolution of external forcing factors including solar intensity, volcanic emissions, GHGs, and ozone/aerosols. AMV indices and NAO indices were calculated in the same way as for the LME.

e. Significance tests

To test the significance of the correlation values we used a Monte Carlo phase randomization technique—following, for example, Ebisuzaki (1997). In this technique, 10 000 dummy time series are produced from one of the input time series that have the same spectral properties but are uncorrelated. These dummy time series are produced by taking the Fourier transform, then randomizing the phase of each component but retaining the amplitude—the inverse Fourier transform is then taken to give the dummy time series. These 10 000 dummy time series have a zero-mean correlation with the target time series. The significance is calculated as the percentage of the dummy time series that have a correlation with a lower amplitude than the original correlation between the input time series.

3. Internal AMV in proxy and early instrumental data

We begin our analysis by examining the presence of internal AMV (i.e., NAO-driven AMV) in the proxy indices through comparison with the evolution of the NAO. The annual mean NAO index, calculated from the instrumental data, is shown in Fig. 1a. The NAO reconstruction clearly has more interannual variance after about 1855—the period when a much larger number of station indices are included. However, the

NAO index exhibits substantial variability on longer time scales throughout the reconstruction (i.e., thick black line, Fig. 1a). Li et al. (2013) showed that over the observational period, low-frequency AMV changes are proportional to the low-pass-filtered NAO (i.e., $\overline{\text{NAO}}$):

$$\frac{d}{dt}\text{AMV}(t) \propto \overline{\text{NAO}}(t). \qquad (1)$$

This can equivalently be written in the integral form and several studies have shown that the AMV is proportional to the accumulated NAO (e.g., McCarthy et al. 2015; Delworth et al. 2017; Sun et al. 2019):

$$\text{AMV}(t) \propto \text{NAO}_{\text{int}}(t) = \int_{t_0}^{t} \text{NAO}(t') \, dt', \qquad (2)$$

where NAO_{int} is the accumulated NAO forcing and t_0 is the time at the start of the NAO record (i.e., 1659). Here we investigate the presence of internally generated AMV going further back in time using the proxy datasets.

The integrated NAO index (NAO_{int}) and the low-pass proxy AMV indices (calculated using a 7-yr moving average) are shown in Fig. 1b. From visual inspection it is clear that the AMV indices generally follow the integrated NAO, and this is particularly clear for the AMV-Wang index. It should be noted that the NAO accumulation precedes the AMV in this analysis, so covariability between the integrated NAO and the AMV does not occur due to the coincident influence of the NAO on North Atlantic temperatures and the associated proxies. The correlation between the integrated NAO and the AMV indices is shown explicitly in Fig. 1c, over 100-yr moving windows as well as for the full period (1659–1999), early period (1659–1855), and late period (1856–1999). This late period represents the instrumental period over which the proxy indices were calibrated. Over the full period, the integrated NAO is most highly correlated with AMV-Wang ($r = 0.65$) but also exhibits significantly positive correlations with AMV-Mann and AMV-Gray (both $r = 0.51$). Much of this positive correlation is from the late period, during which there is a significantly positive relationship between the observed AMV index and the integrated NAO ($r = 0.67$), which is spanned by the three AMV indices. However, there is also a clear relation between the integrated NAO and AMV-Wang in the early period ($r = 0.56$), which is entirely based on the proxy data. The other two proxy AMV indices have nonsignificant but positive correlations with the integrated NAO forcing.

The AMV-Wang index shows a particularly clear relationship with the integrated NAO forcing from around 1750 or so to the start of the instrumental period in 1856.
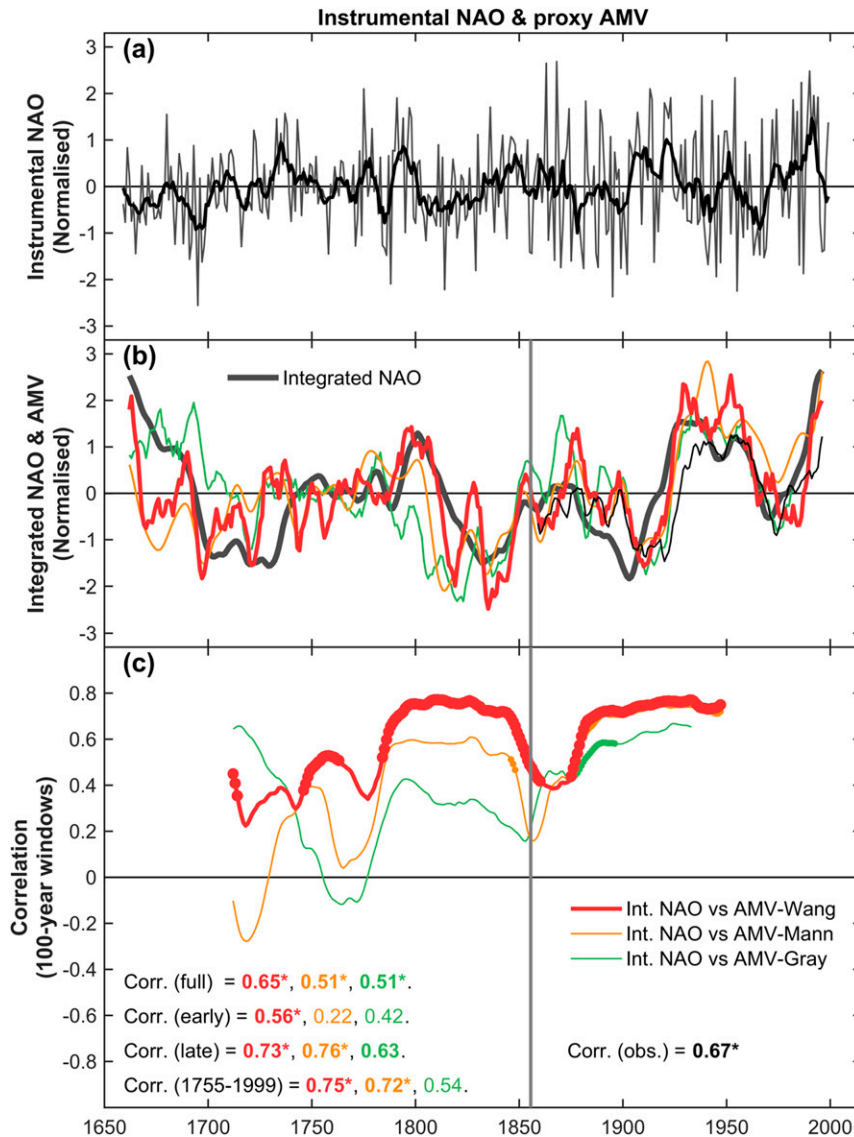
FIG. 1. (a) Annual reconstructed instrumental NAO index from Luterbacher et al. (2002) is shown in gray; the thick black line shows the 7-yr running mean NAO index. (b) Integrated NAO index (gray) along with the three proxy AMV indices from Wang et al. (red), Mann et al. (orange), and Gray et al. (green); the indices have been normalized after taking a 7-yr running mean. (c) Correlation between the integrated NAO and each proxy AMV index over running 100-yr windows (plotted at the center of the 100-yr window), where the filled circles denote correlation values that are significant at the 90% level. Also shown in (c) are the correlation values between the integrated NAO and the proxy AMV indices during the full period (1659–1999), the early period (1659–1855), the late period (1856–1999), and also the period 1755–1999. Correlation values in bold font are significant at the 90% level, and correlation values with an asterisk are significant at the 95% level. The significance values were calculated from a Monte Carlo phase randomization technique (see methods).

During this period the AMV-Wang index is entirely outside the instrumental calibration period and is only determined by the proxy data; remarkably, the relationship with the integrated NAO is even stronger than during the instrumental period (Fig. 1c). From 1755 onward, the significantly more predictors go into the instrumental SLP reconstruction Luterbacher et al. (2002) and this extended period (i.e., 1755–1999) exhibits the stronger correlation between the integrated NAO and the AMV that in the late period.
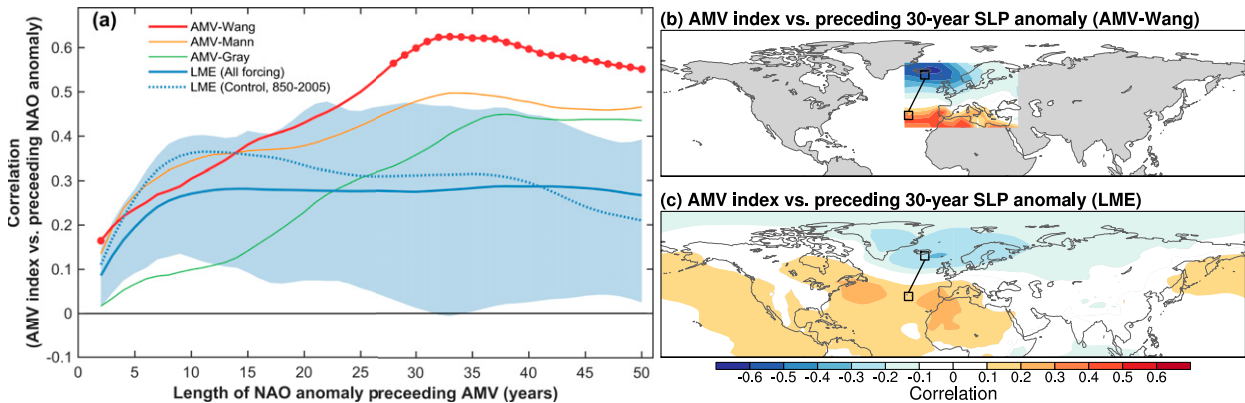
FIG. 2. (a) Correlation between preceding NAO anomaly (i.e., averaged from $t - N$ to $t - 1$ years) and the following 7-yr AMV anomaly (i.e., averaged from t to $t + 6$ years). The red, orange, and green lines in (a) show the correlation for the proxy AMV indices and the instrumental NAO index; the solid blue line shows the correlation averaged over all 13 ensemble members of the LME and the shading shows the interquartile range of the ensemble member correlations; the dotted blue line shows the correlation calculated from 1155 years of a single control simulation of LME model without external forcing. The filled circles indicate where the proxy AMV and instrumental NAO correlation is larger than at least 12 of the 13 LME ensemble members. (b) Map of the correlation of AMV-Wang with the preceding 30-yr SLP anomaly (from the Luterbacher et al. SLP reconstruction). (c) Map of the correlation between the AMV index in the LME with the preceding 30-yr SLP anomaly. The black boxes in (b) and (c) indicate the region used to calculate the NAO indices.

By comparing the integrated NAO with the AMV, we are implicitly testing the forcing of the AMV by the large-scale atmospheric circulation changes. To test this more simply we can compare the AMV with the preceding NAO anomaly over a given period—this is shown for varying periods in Fig. 2a. The preceding NAO anomaly is positively correlated with the AMV in all the proxy datasets, peaking at around 30–35 years, and is largest in AMV-Wang. Correlation of the AMV with the gridpoint SLP anomaly over the preceding 30 years is shown in Fig. 2b. The SLP pattern clearly corresponds to the positive phase of the NAO, with a particularly strong dependence on the strength of the Icelandic low, similar to that seen over the modern instrumental period and in CMIP5 models (O'Reilly and Zanna 2018; Peings et al. 2016).

The relationship between the early instrumental NAO index and the proxy AMV is consistent with there being a substantial component of the AMV that reflects the response of the North Atlantic to large-scale atmospheric forcing. This is particularly the case in AMV-Wang, the most modern and well verified of the proxy AMV indices, for which the accumulated NAO forcing can explain over 40% of the variance (i.e., $r^2 = 0.42$). We now turn our attention to the externally forced AMV component of the proxy AMV.

4. Externally forced AMV in the Last Millennium Ensemble and proxy indices

To assess the externally forced component of the AMV, we analyze the AMV from the LME simulations.

The AMV indices from each member of the LME, as well as the ensemble mean, are shown in Fig. 3a. The ensemble mean AMV can be interpreted as the externally forced component of the AMV in the LME. We compare the externally forced AMV in the LME with the proxy AMV indices to estimate the externally forced component. This estimate of the externally forced component in the proxy AMV indices can perhaps be considered a lower bound on the externally forced AMV due to uncertainties associated with the prescribed external forcings and errors in the model response to the external forcing. However, the component of the proxy AMV indices that covaries with the LME can be interpreted as an externally forced component of the proxy AMV with a reasonable degree of confidence.

The proxy AMV indices are shown alongside the AMV-LME indices in Fig. 3a, all plotted using a 7-yr moving average filter. While there is relatively little similarity between the proxy AMV indices and AMV-LME over the whole period, there are certain periods during which there is clear covariability, such as the early 1800s and periods of the twentieth century. The correlation between the AMV-LME ensemble mean and the proxy AMV indices are shown in Fig. 3b for all filter lengths from 1 to 20 years. The correlation is significant at the 10% level for all proxy AMV indices and is largest for filter lengths of around 7 years, peaking at $r = 0.37$ for AMV-Wang. This suggests that a significant portion of the AMV proxies, over 10% of the explained variance (i.e., $r^2 = 0.13$), is externally forced. Correlations between the AMV-LME and the proxy AMV
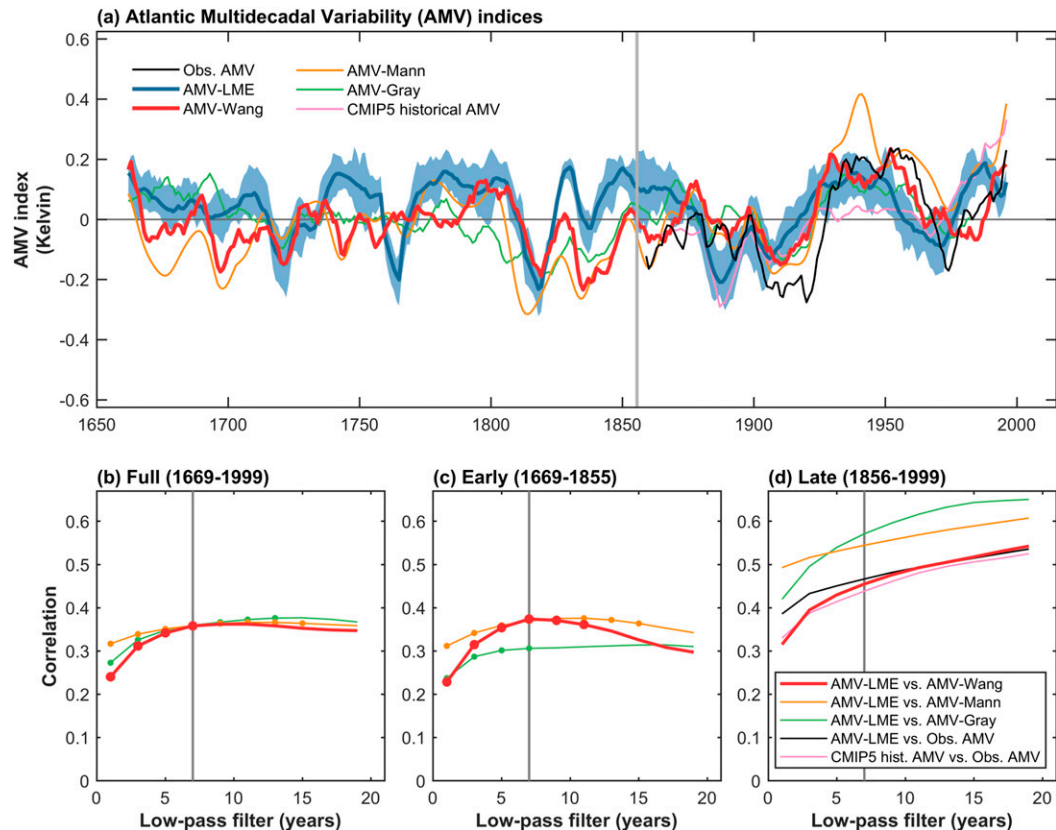
FIG. 3. (a) Proxy AMV indices and the AMV indices from the LME ensemble members (interquartile range shaded in blue) and ensemble mean (thick blue line); these indices have been filtered using a 7-yr moving average. Also shown in (a) are the observational AMV index (from 1856 to 1999) and the ensemble mean AMV index from the CMIP5 historical simulations (from 1860 to 1999). The correlations between the ensemble mean LME AMV index and the proxy AMV indices are shown for various low-pass filter lengths for the full analysis period, 1659–1999 in (a) and also for (b) the early period, 1669–1855, and (c) the late period, 1856–1999. The filled circles in (a)–(c) indicate where the correlation is significant at the 10% level according to a Monte Carlo significance test (see methods, section 2e).

indices in the early period (i.e., outside of the instrumental calibration period) are shown in Fig. 3c. Similar and significant correlation values are seen in this early period, peaking at $r = 0.38$ with a 7-yr moving average filter for AMV-Wang. This confirms that the externally forced component in the proxy AMV occurs in the preindustrial period and corroborates the similar conclusions in Wang et al. (2017a), which were made using a simple statistical model of the externally forced AMV component.

In the late period, during which instrumental data are available, the correlation between AMV-LME and the observed AMV is slightly higher but is not significant at the 10% level and the proxy AMV indices exhibit similar behavior (Fig. 3d). As the length of the low-pass filter increases, the correlation between the observed AMV and AMV-LME increases, with $r > 0.5$ for a 20-yr moving average, although it is not significant at the 10%

level. At first glance, the low level of significance of the late period may seem odd but this period is shorter and exhibits fewer degrees of freedom than in the early period, which is reflected in the low level of significance. Also shown in Fig. 3c is the correlation between the observed AMV and the ensemble mean AMV from the CMIP5 historical simulations (pink line in Fig. 3a). The correlation between the CMIP5 historical AMV and the observed AMV over the late period is very similar to the AMV-LME and increases with filter length. These correlation values are slightly lower values than those reported in the recent studies by Bellomo et al. (2018) for the LME and Murphy et al. (2017), which reported correlation values of around $r \approx 0.7$ for the corresponding ensemble means. This could be due to the slightly different periods used in the analysis, the different filtering methods used or the linear detrending that is performed in both of these previous studies,

which we do not apply to any of the indices in our analysis. It therefore seems likely that some fraction of the observed AMV over the later period is externally forced, but the exact amount is very uncertain. Moreover, as the correlation increases with the filter length, it could be that there is some aliasing between external forced temperature signals in the model and internal AMV variability in the observations over the late period.

5. Internal AMV in the LME

We now examine the internal AMV present in the LME to compare with the external component. Figure 4 shows the integrated NAO in the LME and the AMV-LME, analogous to the proxy AMV analysis in Fig. 1. The correlation between the integrated NAO and the AMV-LME is shown for each ensemble member over 100-yr windows in Fig. 4c, along with the mean correlation over all 13 ensemble members. It is immediately obvious that there is considerable spread across the ensemble members, though the LME typically demonstrates a small positive correlation throughout the analysis period. The light blue dots on the right-hand side of Fig. 4c show the correlation between the integrated NAO and AMV-LME, which is positive for all ensemble members with an ensemble mean of $r = 0.21$, substantially less than that found for the proxy AMV indices (shown as colored crosses in Fig. 4c). Over the late period, there is a marginally stronger relationship between the integrated NAO and the AMV-LME (i.e., brown dots in Fig. 4c), though in all ensemble members the relationship is weaker than seen in observations (i.e., black cross in Fig. 4c). Across the CMIP5 ensemble there is clearly a much larger spread in the relationship between the integrated NAO and AMV, shown by the pink dots in Fig. 4c; while the CMIP5 ensemble mean is positive it is very small, and many models even exhibit a negative correlation. Only 2 of the 38 models in the CMIP5 ensemble exhibit a correlation between the integrated NAO and AMV as strong as that seen in observations. The relationships between the integrated NAO and AMV in the LME ensemble members over the late period are mostly within the upper half of the CMIP5 ensemble, so is certainly comparable with most of the CMIP5 models.

To further investigate how the AMV is forced by the NAO in the LME we now compare the AMV with the preceding NAO anomaly over a given period, which is shown in Fig. 2a. At lead times of between 10 and 50 years the correlation between the preceding NAO anomaly and the AMV averaged over the 13 ensemble members is small but positive; however, there is substantial spread across the 13 ensemble members. At lead times of 25 years and longer, the positive correlation between the early instrumental NAO anomaly and AMV-Wang becomes substantially larger than in the LME. This indicates that at these long lead times, the LME struggles to capture the stronger relationship between the NAO and AMV, which is also clear in the gridpoint correlation map in Fig. 2c. The difference in the fraction of NAO-forced variability in the model and proxy data may be due to the difference in the magnitude of the NAO variability on decadal time scales. The standard deviation of the decadal early instrumental NAO reconstruction is $\sigma = 0.36$ (using a 10-yr moving average), whereas the standard deviation of the decadal NAO in the 13 members of the LME are all in the range $\sigma_{LME} = 0.13$–0.20. Therefore, the lack of NAO-driven AMV in the LME may be due to a lack of decadal NAO variability. Across the 13 ensemble members, those that have the highest decadal NAO variability tend to be those that have a stronger relationship between the NAO and AMV (Fig. S2), though the correlation is not clearly significant. The lack of decadal NAO variability in the LME is a typical feature of the current generation of coupled climate models, as demonstrated in the recent study of Bracegirdle et al. (2018).

To examine whether or not the presence of the external forcing influencing the link between the NAO and the AMV in the model, we performed the same analysis in the single-member 1155-yr LME control simulation, which has no external forcing (dashed line in Fig. 4a). The LME control simulation exhibits a very similar relationship between the NAO and the AMV as in the full LME, demonstrating that in the model this relationship is too weak on long time scales, even in the absence of external forcing. A similarly weak relationship is also found in the CMIP5 historical simulations over the recent observational period, as demonstrated in O'Reilly and Zanna (2018), so the LME is not unusual in this regard (see also Peings et al. 2016).

6. Assessing sources of externally forced AMV

Having established that there is a clear externally forced AMV component in the LME, which is to a lesser degree apparent in the proxy AMV indices, we now investigate the source of the externally forced AMV. To do so, we make use of the individual forcing experiments that were performed alongside the full LME: volcanic forcing (5 members), solar forcing (4 members), GHG forcing (3 members), and aerosol and ozone forcing (3 members). These experiments were all performed over the full analysis period except the aerosol and ozone forcing experiment, which was only run for the industrial period, from 1850 onward. Early and late periods are
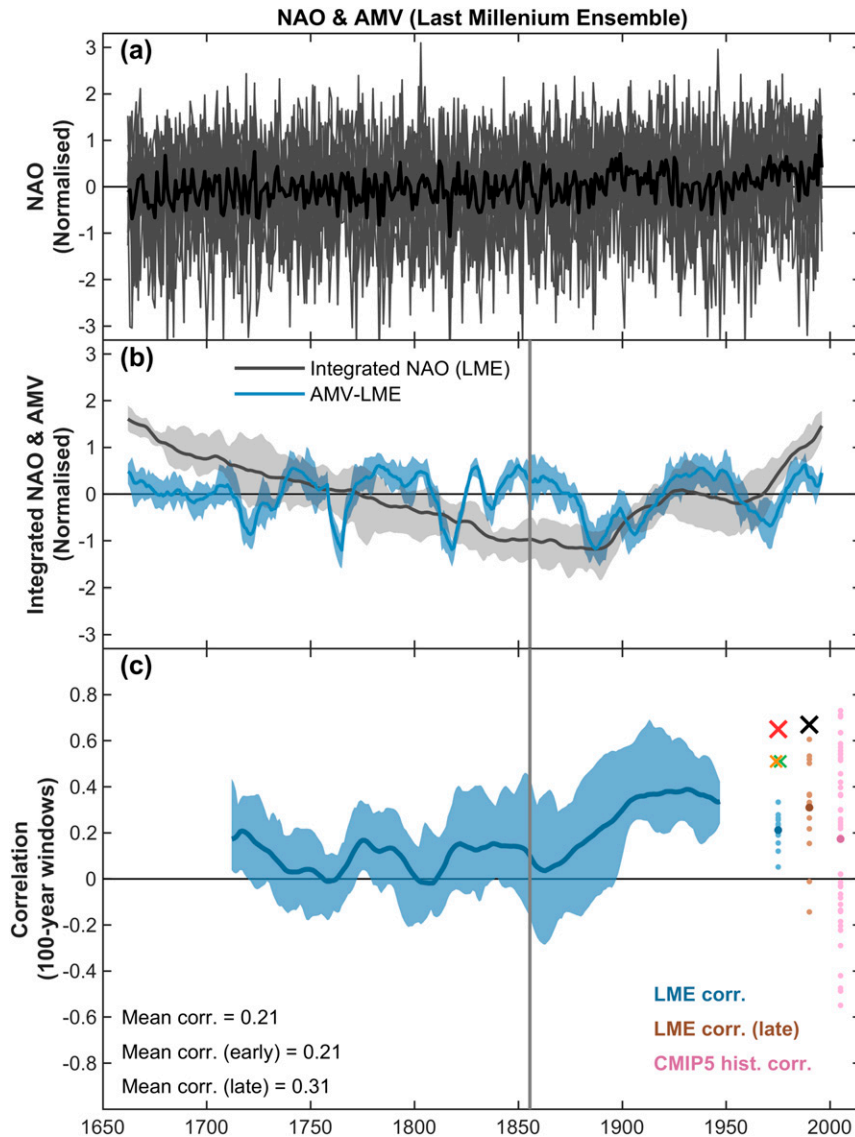
FIG. 4. (a) Annual NAO indices from the 13 LME members in gray; the ensemble mean is shown in the bold black line. (b) Integrated NAO indices from each LME member (gray) along with the AMV indices from each LME member (blue)—the interquartile range of the individual ensemble members is shaded and the mean shown by the line. (c) Correlation between the integrated NAO and AMV index for each LME member over running 100-yr windows (plotted at the center of the 100-yr window) with the ensemble mean shown in dark blue. Also shown in (c) are the correlation between the integrated NAO and AMV from the LME over the whole period (light blue dots; ensemble mean in dark blue), over the late period (light brown dots; ensemble mean in dark brown), and the correlation between the integrated NAO and AMV in each CMIP5 historical simulation (light pink dots; ensemble mean in dark pink). The crosses in (c) show the correlation of the integrated instrumental NAO (i.e., Fig. 1) with the AMV indices over the full period of observations (black), Wang et al. proxy (red), Mann et al. proxy (yellow), and Gray et al. proxy (green).

defined, as in the earlier analysis, as 1669–1855 and 1856–1999, respectively.

The ensemble mean AMV from each of the individual forcing experiments is shown in Fig. 5a, along with the

ensemble mean AMV-LME (i.e., the full forcing experiment). Much of the variability in the AMV-LME is captured by the volcanic forcing experiment, particularly in the early part of the analysis period, yet
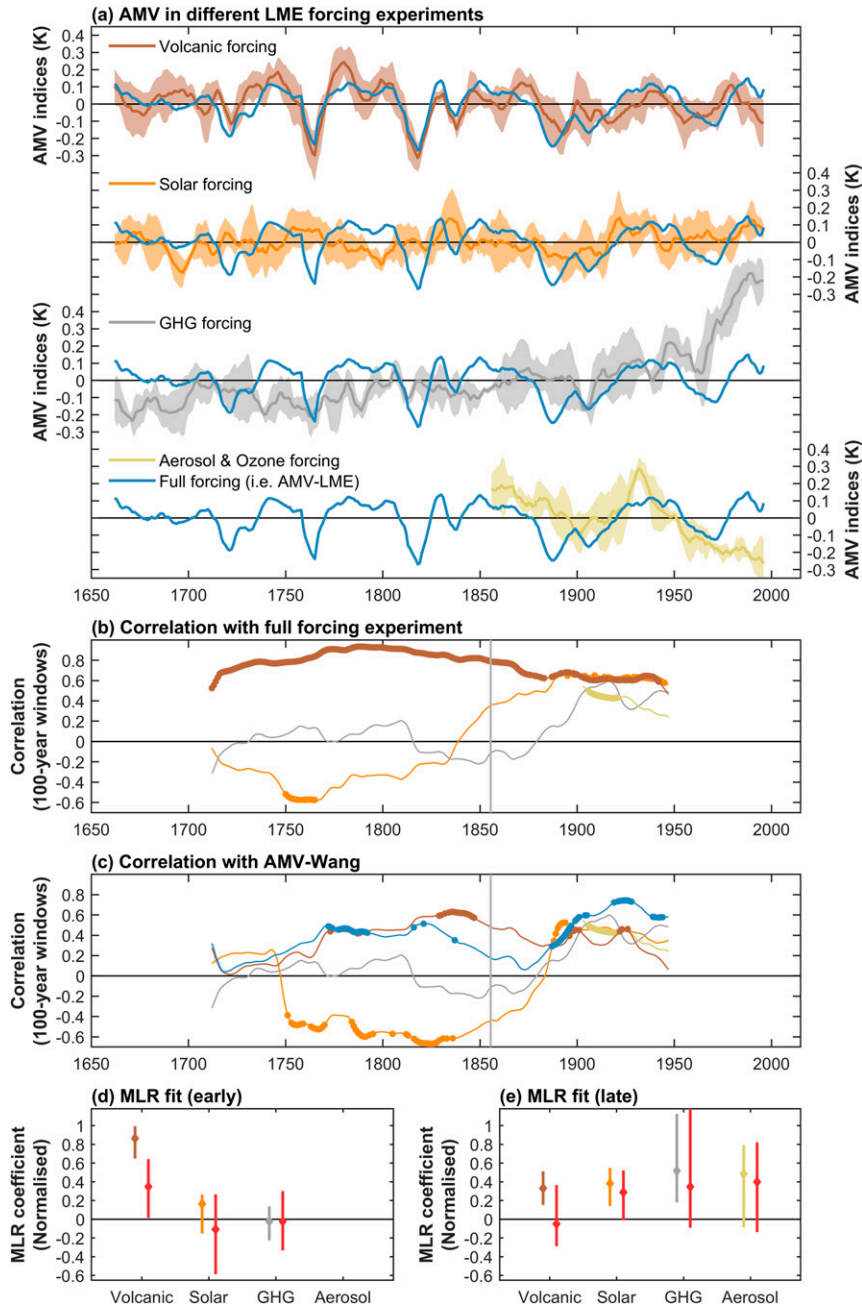
FIG. 5. (a) Ensemble mean AMV in the full forcing (i.e., AMV-LME, 13 members), volcanic forcing (5 members), solar forcing (4 members), GHG forcing (3 members), and aerosol and ozone forcing (3 members) experiments. The shading around each line in (a) indicates ±1 standard deviation of the ensemble. (b) Correlation of the ensemble mean AMV index from the full forcing experiment with the AMV indices from the other experiments, over moving 100-yr windows. (c) Correlation of the proxy AMV index from Wang et al. with the AMV indices from the different experiments, over moving 100-yr windows. Coefficients from the multiple linear regression fit of the different forcing experiments to the full forcing experiment (in different colors) and to the proxy AMV index (red). The bars in (d) and (e) show the 90% confidence interval of the regression coefficients calculated using a block bootstrap resampling (e.g., Wilks 1997), using blocks of 20 years.

similarity with the other forcing experiments is less clear. To quantify the similarity between the individual forcing experiments and the AMV-LME, we calculated the correlations between them over 100-yr moving windows, which are shown in Fig. 5b. Over the early period, the volcanic forcing AMV is highly correlated with the AMV-LME, whereas the GHG forcing AMV and the solar forcing AMV exhibit approximately no correlation and negative correlation, respectively. In the late period, when all four individual forcing experiments are available, there are positive correlations between all the experiments and the AMV-LME, indicating that all four of these forcing factors could be influencing the external AMV over the late period. The volcanic and solar forcing experiments exhibit the strongest correlations, the aerosol forcing also exhibits significantly positive correlation with the AMV-LMV but the correlation with the GHG forcing experiment is not significant.

Similarly, we can compare the AMV from the individual forcing experiments with the proxy AMV indices and correlations over moving 100-yr windows with AMV-Wang are shown in Fig. 5c. The volcanic forcing AMV exhibits the largest correlation with AMV-Wang in the early period and exhibits similar correlations with AMV-Wang as the AMV-LME (shown in blue). The GHG forcing AMV is uncorrelated with AMV-Wang in the early period and the solar forcing AMV is significantly negatively correlated with AMV-Wang. This suggests that either the solar variability was not important for past changes in the AMV or the model is incorrectly capturing these mechanisms through which solar variability influences the AMV, though it could be that the solar forcing acts to dampen the influence of the volcanic forcing on the AMV. Over the late period there are positive correlations with AMV-Wang in all of the individual forcing experiments; however, there are only significant correlations with the volcanic, solar, and aerosol forcing experiments. Moreover, it is noticeable that none of the individual forcing AMVs are as closely correlated with AMV-Wang as the ensemble mean AMV-LME, which suggests that several of the different forcings may be contributing to the externally forced component of the observed AMV.

To estimate the relative importance of each of the external forcing components, we performed a multiple linear regression (MLR) analysis by fitting the individual forcing experiment AMV indices to the AMV-LME and AMV-Wang. All ensemble mean indices were normalized prior to the analysis and the normalized regression coefficients are plotted for the early and late periods in Figs. 5d and 5e, respectively. Uncertainty estimates for the regression coefficients were calculated

using block bootstrapping, using a block length of 20 years (e.g., Wilks 1997). The predictors are mostly uncorrelated except for the GHG and aerosol forcing AMV indices in the late period ($r = -0.66$), where both exhibit substantial trends, so this may be expected to increase the uncertainty in the regression coefficients. The AMV-LME during the early period is dominated by the volcanic forcing component, whereas the solar forcing and GHG forcing make no significant contribution. The volcanic forcing also dominates the MLR fit to AMV-Wang, although the regression coefficient is much smaller than in AMV-LME. In the late period, the MLR suggests that all of the forcing components contribute somewhat to the AMV-LME, though there is large uncertainty in the relative contributions from each forcing. The MLR fit to AMV-Wang in the late period also has contributions from all the individual forcing components, with the exception of the volcanic forcing, which makes a very small contribution to the MLR fit.

7. Is the AMV-LME responding appropriately to external forcing?

Our analysis to this point has shown that there is a substantial externally forced AMV component in the LME (i.e., Fig. 3), which can explain a significant but modest fraction of the proxy AMV variance. Subsequent analysis of the internal AMV, specifically that driven by changes in the large-scale NAO forcing, showed that the LME tends to underrepresent this component of the AMV (i.e., Fig. 4). This suggests that the AMV in the LME may be responding too strongly to the external forcing. In this section we will attempt to assess how appropriately the AMV-LME responds to external forcing and use this analysis to interpret the externally forced AMV in the proxy indices.

To investigate whether the externally forced AMV is too strong in the LME, we assess the externally forced or predictable signal in comparison to internal variability in each ensemble member. To do this, we first calculate the perfect model correlation of AMV-LME, which is simply the correlation of each ensemble member of the LME with the ensemble mean. The perfect model correlation of the AMV-LME is shown for a varying number of ensemble members during the full, early, and late periods in Fig. 6 (solid blue lines). The perfect model correlation can be compared with the actual correlation between the AMV-LME with the proxy and observed AMV indices, shown in Fig. 6 for AMV-Wang. In the full, early, and late periods, the perfect model correlation is almost 0.7 with all 13 ensemble members and is significantly larger than the actual correlation with AMV-Wang and the observed AMV (during the
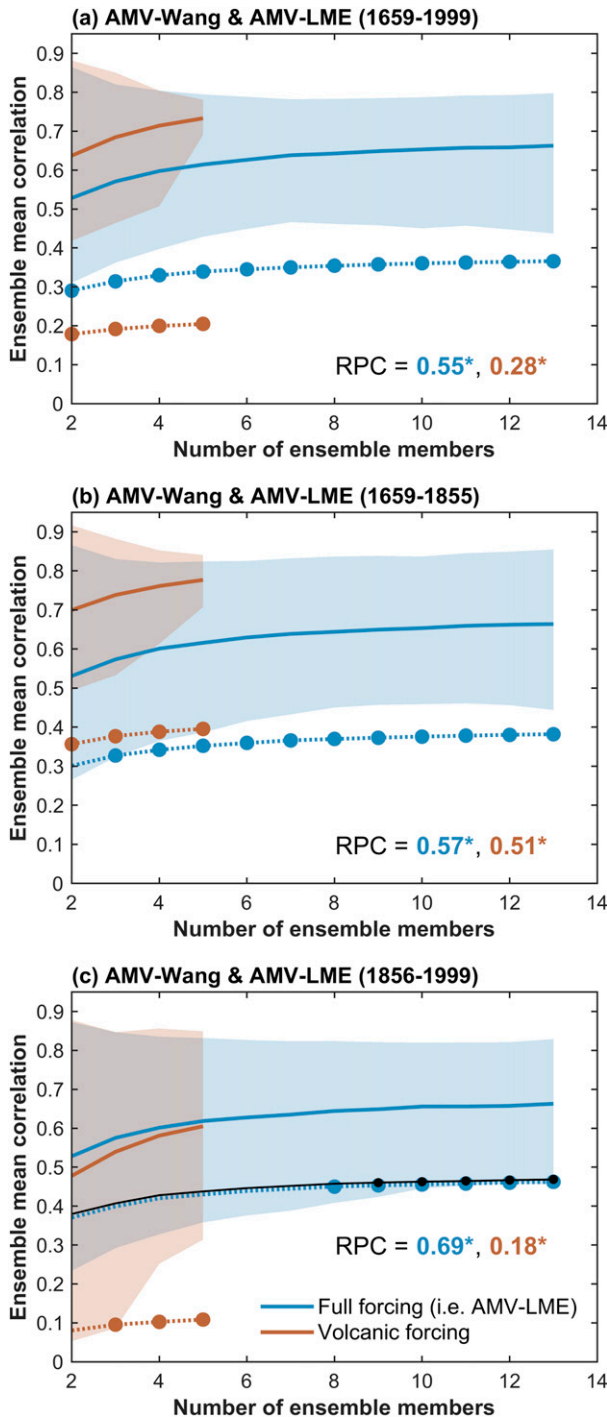
FIG. 6. Correlation of the ensemble mean AMV with the proxy AMV (from Wang et al.) depending on number of ensemble members is shown in dotted lines for (a) the full period, 1659–1999; (b) the early period, 1659–1855; and (c) the late period, 1856–1999. The solid lines show the average correlation of the ensemble mean AMV with individual ensemble members, the shading shows the 5%–95% confidence interval calculated using a random subsampling of the ensemble members. The filled circles show where the correlation of the ensemble mean with individual ensemble members is significantly different from the correlation with AMV-Wang (and observed AMV) at the 90% level, based on random subsampling of ensemble members.

late period). This suggests that the externally forced signal in the AMV-LME may be too large. This can be quantified by calculating the *ratio of predictable components* (RPC), which is the ratio of the predictable component in the real world to the predictable component (or signal to noise) in the model and has previously been used in studies of seasonal/decadal predictability (e.g., Eade et al. 2014; Scaife and Smith 2018; O'Reilly et al. 2019). Over the whole period, the RPC is significantly less than 1 (see Fig. 6), which indicates that the predictable signal in the LME is larger than the signal that is captured in the proxy/observed AMV.

If we were able to assume that the external forcings prescribed in the LME were a perfect replica of the real world forcing over the analysis period and that the proxy AMV indices are trustworthy, then we would be able to state with confidence that the AMV in the model is responding too strongly to external forcing. Of course, this assumption is not well justified, as many of the prescribed external forcings are estimates that are very uncertain. For example, while there is a general consensus on timings of large volcanic eruptions over the past 400 years, there is considerable uncertainty as to the amount and type of volcanic aerosols emitted (e.g., Crosweller et al. 2012). Nonetheless, over the recent analysis period—when the external forcing factors are expected to best constrained and we can compare with the observed AMV—the LME still appears to respond too strongly to the external forcing. This is certainly consistent with the AMV-LME not having a large enough internal AMV component, that is, responding too weakly to preceding NAO forcing (e.g., Figs. 2 and 4).

To this point, we have treated the externally forced AMV as distinct from the internal AMV that is associated with preceding large-scale circulation anomalies. However, previous studies have suggested that external forcing, particularly from volcanic aerosol emissions, can influence the phase of the NAO (e.g., Ortega et al. 2015). Therefore, it is possible that at least some of the NAO signal that leads the AMV in the observations and proxy records is externally forced. To examine this, we compare the externally forced (ensemble mean) AMV-LME with the integrated NAO from the early instrumental NAO index, which we previously found was able to explain over 40% of the variance in the AMV-Wang proxy index. These indices are plotted again in Fig. 7a, along with the ensemble mean integrated NAO index from the LME and the correlation coefficients calculated between the integrated NAO indices and the AMV-LME. As discussed above, the correlation between the integrated NAO in the LME and the AMV-LME is positive but very low ($r = 0.18$). However, the
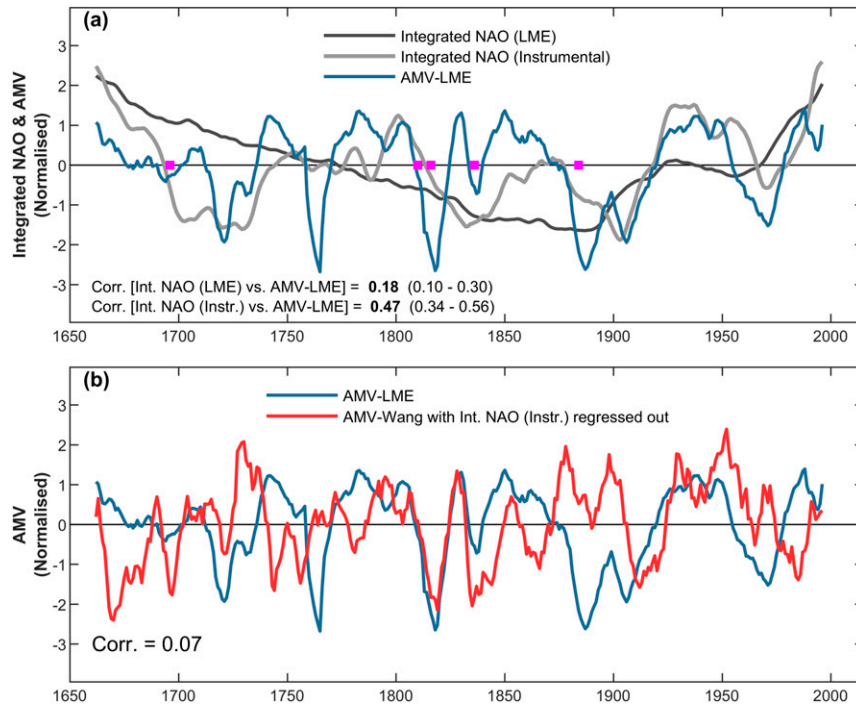
FIG. 7. (a) Integrated NAO from the Luterbacher et al. (2002) SLP reconstruction (light gray), as in Fig. 1b; ensemble mean of the integrated NAO indices from the LME (dark gray); ensemble mean AMV from the LME (in blue). The magenta squares in (a) indicate the five large volcanic eruptions between 1659 and 1999, following Swingedouw et al. (2017). The correlation between the integrated NAO indices and the AMV from the LME are also shown in (a), along with the 90% uncertainty range based on ensemble subsampling of the LME. (b) Proxy AMV (from Wang et al.) and the AMV index with the integrated NAO index from observations regressed out; these indices are essentially uncorrelated.

AMV-LME is much more closely related to the integrated NAO from the instrumental SLP reconstruction, with a correlation of $r = 0.47$ and a 90% confidence interval of $r = 0.34$–0.56 (calculated by randomly sampling across the 13 ensemble members). Since the ensemble mean AMV-LME is the externally forced component of the AMV in the model, the reasonably large correlation with the integrated NAO from the early instrumental data indicates that a significant fraction of this may be externally forced.

We now test whether the externally forced component identified in the proxy AMV occurs through the influence of the NAO. To test this, in Fig. 7b we again compare the AMV-LME with AMV-Wang, but here we have (linearly) regressed out the integrated NAO influence. The correlation between AMV-LME and the AMV-Wang was $r = 0.37$ prior to the regression (e.g., Fig. 3); however, after regressing out the integrated NAO from AMV-Wang there is practically zero correlation between the indices ($r = 0.07$). Therefore, the component of AMV-Wang that was identified as being externally forced in section 3b (i.e., Fig. 4) seems to be

wholly associated with the integrated NAO. It is therefore interesting that in the LME the externally forced AMV is not as strongly associated with the integrated NAO, suggesting that the external forcing is exerting a greater influence on the large-scale circulation in observations than in the LME. In the LME, the externally forced AMV is seemingly generated more directly through changes in radiative forcing, as argued in Bellomo et al. (2018). It appears, therefore, that while some of the externally forced AMV is correct, the pathway through which the external forcing influences the AMV is incorrect.

Detailed examination of the pathways through which external forcing is influencing the AMV is a difficult task given the relative paucity of observations. Nonetheless, we can try to examine the response of the model and observations/proxies to volcanic forcing, which was found to be the dominant individual forcing factor in section 3d. We perform a so-called superposed epoch analysis on the five explosive eruptions that occurred during our analysis period and are common to three alternative volcanic reconstructions, based on Swingedouw
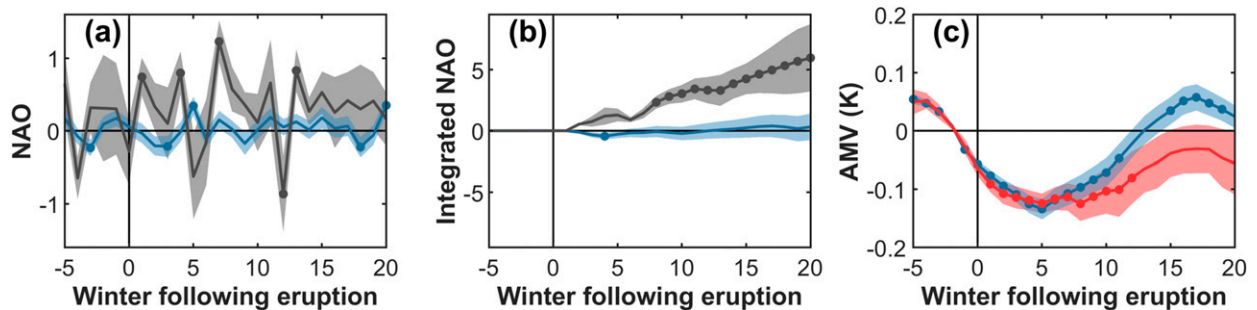
FIG. 8. Superposed epoch analysis of climate impact following the five largest volcanic eruptions for (a) NAO index, (b) integrated NAO, and (c) AMV index. Anomalies are shown relative to the 5-yr period prior to the eruption for the reconstructed NAO (in gray), the proxy AMV from Wang et al. (in red), and from the LME (in blue). In (a)–(c) the shading indicates the standard error of the mean; the dots indicate where this is significantly different from zero at the 90% confidence level.

et al. (2017), and these are indicated in Fig. 7a. We focus on the most explosive eruptions because the response to more minor eruptions can be difficult to detect as they do not necessarily inject substantial amounts of aerosol into the stratosphere. The LME uses the Gao et al. (2008) reconstruction, which was also used to identify the large eruptions, so it is reasonable to also perform this analysis for these same events in the model. Figure 8a shows the average response of the NAO in the winters following the large volcanic eruptions.

In the early instrumental NAO, there is a positive NAO response in all 4 years following the eruption. Of course, the response is fairly noisy because it is only averaged over five eruptions. Nonetheless, this is consistent to previous observational and modeling studies that have found that the NAO is typically positive in the years following an explosive volcanic eruption (e.g., Graft et al. 1993; Shindell et al. 2004; Ortega et al. 2015; Swingedouw et al. 2017). In the LME, however, there is no consistent NAO response to the major volcanic eruptions, which indicates that the LME is missing the dynamical response of the large-scale circulation to external volcanic forcing. In terms of integrated NAO response, the absence of a consistent response in the LME means there is no signal in the integrated NAO following the major volcanic eruptions (Fig. 8b); whereas in the early instrumental NAO index, the integrated response typically increases and becomes significant between 5 and 10 years following the eruption. Despite the differences in the NAO response to the eruptions, however, the impact of the radiative forcing in the LME is enough to give a similar AMV response to the proxy AMV, which consists of a cooling for the years following the eruption (Fig. 8c). This simple analysis demonstrates that, for these five major eruptions, the LME is apparently missing the dynamical response in the large-scale circulation, but the radiative effects are

able to produce a similar AMV response, in both sign and magnitude. It is therefore possible that the dynamical response to other external forcings, such as solar variability (e.g., Gray et al. 2010; Ineson et al. 2011), is not captured by the model but that some of the externally forced AMV variability is being realized in the model through radiative effects alone.

8. Further discussion and conclusions

In this study we have analyzed sources of AMV over a period of more than 300 years, using proxy indices and early instrumental data, along with a complementary ensemble climate model simulation over the same period. The proxy AMV is found to closely follow the accumulated NAO forcing over almost the entire analysis period, referred to as an "internal" source of AMV as it can be expected to occur to some extent in the absence of external forcing. The dependence of the AMV on the accumulated NAO forcing has previously been demonstrated over the observational period, the last 150 years or so, but here we show that this is also evident over a significantly longer period (i.e., Fig. 1). These results provide additional observational evidence that much of the variance in the AMV is driven by the oceanic response to the buildup of surface atmospheric forcing, as has been demonstrated in targeted modeling studies (e.g., Delworth et al. 2017).

The AMV response to the accumulated NAO forcing is found to be present in the LME simulations, though it is much weaker than that seen in the proxy/observational datasets. Similar behavior is seen in the majority of the CMIP5 historical simulations so, at least in this respect, the behavior of the LME is fairly representative of other models. The weak AMV response to the accumulated NAO forcing is particularly clear on time scales greater than a decade in the LME, which was also found

to be the case in the preindustrial control and historical CMIP5 simulations in O'Reilly and Zanna (2018). This could be due to a weaker-than-expected response of the ocean circulation to the accumulated NAO forcing (Xu et al. 2019). Alternatively, the weak response to the NAO could be due to the relatively weak NAO variability on multidecadal time scales. In the LME, the decadal NAO variability is significantly lower than in the NAO reconstruction, with $\sigma_{LME} = 0.13–0.20$ in the LME compared with $\sigma = 0.36$ for the early instrumental NAO reconstruction. Kim et al. (2018a) showed that the weak multidecadal NAO variability is likely responsible for the weak AMV amplitude in the CESM large ensemble, which employs a very similar model setup to the (CESM) LME analyzed here. Weak multidecadal NAO variability has been shown to be a robust feature of the North Atlantic jet stream across almost all the CMIP5 models in the recent studies by Simpson et al. (2018) and Bracegirdle et al. (2018); see also Woollings et al. (2015). Therefore, the lower magnitude of the decadal NAO forcing may explain the weak response of the AMV to the accumulated NAO in the CMIP5 ensemble (e.g., Fig. 4c). Moreover, the weak multidecadal NAO forcing may therefore explain why the amplitude of the AMV is too low on multidecadal time scales in the CMIP5 ensemble (Peings et al. 2016), as highlighted by the study of Wang et al. (2017b). The misrepresentation of the multidecadal NAO forcing on the AMV clearly has implications for modeling past changes in the North Atlantic but it also has important implications for modeling the contributions of internal multidecadal variability in transient climate projections. Models that are missing multidecadal NAO variability are likely to underestimate the amplitude of the AMV and the internal multidecadal variability in the continental regions surrounding the North Atlantic, including Europe and the Sahel.

It is possible that the weak decadal variability in the LME is related to a weak response of the NAO to the AMV. In observations, a negative NAO on decadal time scales is found to follow a positive AMV anomaly (e.g., Peings and Magnusdottir 2014). Lag correlations between the 7-yr low-pass NAO and AMV indices are shown in Fig. 9 for the proxy and LME indices. Here it is clear that the negative correlation between the AMV and the NAO leading by 10–15 years is not captured by the LME, which has weaker correlations when the AMV is leading. Sun et al. (2015) showed that in a preindustrial control simulation (using the CCSM4 model) the coupled feedback of the AMV onto the NAO at later times likely contributed to the pronounced multidecadal NAO variability in the simulation. It is likely, therefore, that the weak feedback of the AMV onto the
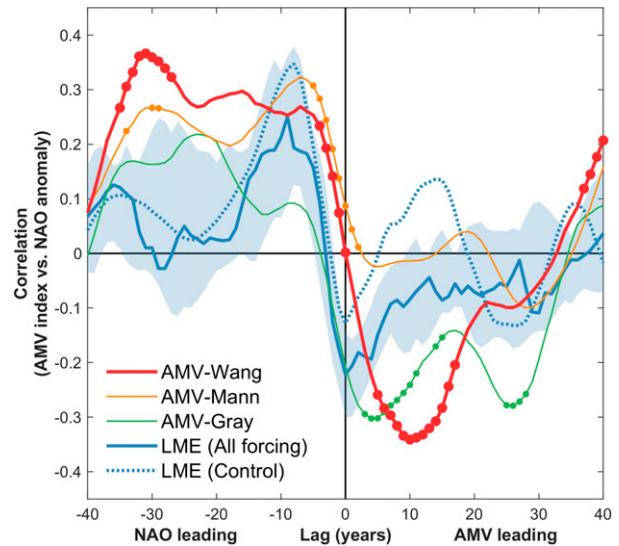

FIG. 9. The red, orange, and green lines show the lag correlation between the 7-yr mean proxy AMV indices and instrumental NAO index; the solid blue line shows the median correlation of the 13 ensemble members of the LME and the shading shows the interquartile range of the ensemble member correlations; the dotted blue line shows the correlation calculated from 1155 years of a single control simulation of LME model without external forcing. The filled circles indicate where the proxy AMV and instrumental NAO correlation is larger/smaller than at least 12 of the 13 LME ensemble members.

NAO is contributing to the weak NAO variability on decadal time scales in the LME. The weak feedback from the AMV to the NAO is not unique to the CESM model analyzed here and is found to be a general feature of coupled CMIP5-era models, as shown by Peings et al. (2016) (see also Simpson et al. 2018).

The analysis of the ensemble mean AMV in the LME reveals a substantial externally forced component. The externally forced AMV exhibits a modest but significant correlation with the proxy AMV (i.e., $r^2 = 0.13$), which implies that *at least* this fraction of the proxy AMV is estimated to be externally forced. The AMV in the individual forcing simulations show that the externally forced AMV is largely driven by volcanic forcing in both the AMV-LME and proxy AMV before the mid-nineteenth century. In the later period, however, GHG forcing, aerosol/ozone forcing, and to a lesser extent solar forcing can explain more of the AMV in both the LME and the observational/proxy records. While the LME is able to confirm an external influence on the observed AMV, the model appears to respond too strongly to this external forcing. Or, alternatively, the AMV in the LME does not exhibit enough internal variability. This may be related to the very weak multidecadal NAO influence on the AMV in the model,

compared with observations, which would in turn be expected to result in a weak internal AMV.

Finally, we showed that externally forced AMV identified in the proxy AMV is related to the accumulated NAO forcing. In the LME, however, very little of the externally forced AMV that arises in the model is through changes in the accumulated NAO forcing. This evidence demonstrates that the external forcing could be influencing the AMV through different mechanistic pathways: via changes in radiative forcing in the LME and via changes in atmospheric circulation in the observational/proxy record. We demonstrated this by analyzing the response to large volcanic eruptions in the model and early instrumental NAO indices. The NAO response seen in the early instrumental data is completely absent in the model, which is nonetheless able to capture the AMV response. If the model was able to better respond to external forcing then it is possible that more of the AMV could be explained by the external forcing, though this is largely speculative. The response of the large-scale atmospheric circulation to changes in external forcing is already an area of active research. However, our findings motivate further study of the impact of these circulation responses on climate variability on multidecadal time scales via interaction with the ocean; this could also be important in governing forced responses over the Pacific Ocean (i.e., the Pacific decadal oscillation) as well as over the North Atlantic as outlined in this study.

REFERENCES

Ahmed, M., and Coauthors, 2013: Continental-scale temperature variability during the past two millennia. *Nat. Geosci.*, **6**, 339–346, https://doi.org/10.1038/ngeo1797.

Årthun, M., E. W. Kolstad, T. Eldevik, and N. S. Keenlyside, 2018: Time scales and sources of European temperature variability. *Geophys. Res. Lett.*, **45**, 3597–3604, https://doi.org/10.1002/2018GL077401.

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.

Booth, 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.

Bracegirdle, T. J., H. Lu, R. Eade, and T. Woollings, 2018: Do CMIP5 models reproduce observed low-frequency North Atlantic jet variability? *Geophys. Res. Lett.*, **45**, 7204–7212, https://doi.org/10.1029/2018GL078965.

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.

Crosweller, H. S., and Coauthors, 2012: Global database on large magnitude explosive volcanic eruptions (LaMEVE). *J. Appl. Volcanol.*, **1**, 4, https://doi.org/10.1186/2191-5040-1-4.

Danabasoglu, G., 2008: On multidecadal variability of the Atlantic meridional overturning circulation in the Community Climate System Model version 3. *J. Climate*, **21**, 5524–5544, https://doi.org/10.1175/2008JCLI2019.1.

Delworth, T. L., 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., 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.

Drews, A., and R. J. Greatbatch, 2016: Atlantic multidecadal variability in a model with an improved North Atlantic current. *Geophys. Res. Lett.*, **43**, 8199–8206, https://doi.org/10.1002/2016GL069815.

Eade, R., D. Smith, A. Scaife, E. Wallace, N. Dunstone, L. Hermanson, and N. Robinson, 2014: Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophys. Res. Lett.*, **41**, 5620–5628, https://doi.org/10.1002/2014GL061146.

Ebisuzaki, W., 1997: A method to estimate the statistical significance of a correlation when the data are serially correlated. *J. Climate*, **10**, 2147–2153, https://doi.org/10.1175/1520-0442(1997)010<2147:AMTETS>2.0.CO;2.

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.

Folland, C. K., T. N. Palmer, and D. E. Parker, 1986: Sahel rainfall and worldwide sea temperatures, 1901–85. *Nature*, **320**, 602–607, https://doi.org/10.1038/320602a0.

Gao, C., A. Robock, and C. Ammann, 2008: Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models. *J. Geophys. Res.*, **113**, D23111, https://doi.org/10.1029/2008JD010239.

Ghosh, R., W. A. Müller, J. Baehr, and J. Bader, 2017: Impact of observed North Atlantic multidecadal variations to European summer climate: A linear baroclinic response to surface heating. *Climate Dyn.*, **48**, 3547–3563, https://doi.org/10.1007/s00382-016-3283-4.

Graft, H., I. Kirchner, A. Robock, and I. Schult, 1993: Pinatubo eruption winter climate effects: Model versus observations. *Climate Dyn.*, **9**, 81–93, https://doi.org/10.1007/BF00210011.

Gray, L. J., and Coauthors, 2010: Solar influences on climate. *Rev. Geophys.*, **48**, RG4001, https://doi.org/10.1029/2009RG000282.

Gray, S. T., L. J. Graumlich, J. L. Betancourt, and G. T. Pederson, 2004: A tree-ring based reconstruction of the Atlantic multidecadal oscillation since 1567 AD. *Geophys. Res. Lett.*, **31**, L12205, https://doi.org/10.1029/2004GL019932.

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.

Hu, Q., and S. Feng, 2012: AMO- and ENSO-driven summertime circulation and precipitation variations in North America. *J. Climate*, **25**, 6477–6495, https://doi.org/10.1175/JCLI-D-11-00520.1.

Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck, 2003: An overview of the North Atlantic Oscillation. *The North Atlantic Oscillation: Climatic Significance and Environmental Impact, Geophys. Monogr.*, Vol. 134, Amer. Geophys. Union, 1–36, https://doi.org/10.1029/134GM01.

Ineson, S., A. A. Scaife, J. R. Knight, J. C. Manners, N. J. Dunstone, L. J. Gray, and J. D. Haigh, 2011: Solar forcing of winter climate variability in the Northern Hemisphere. *Nat. Geosci.*, **4**, 753–757, https://doi.org/10.1038/ngeo1282.

Kim, W. M., S. Yeager, P. Chang, and G. Danabasoglu, 2018a: 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.

——, ——, and G. Danabasoglu, 2018b: Key role of internal ocean dynamics in Atlantic multidecadal variability during the last half century. *Geophys. Res. Lett.*, **45**, 13 449–13 457, https://doi.org/10.1029/2018GL080474.

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.

——, C. K. Folland, and A. A. Scaife, 2006: Climate impacts of the Atlantic multidecadal oscillation. *Geophys. Res. Lett.*, **33**, L17706, https://doi.org/10.1029/2006GL026242.

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.

Leroux, S., T. Penduff, L. Bessières, J.-M. Molines, J.-M. Brankart, G. Sérazin, B. Barnier, and L. Terray, 2018: Intrinsic and atmospherically forced variability of the AMOC: Insights from a large-ensemble ocean hindcast. *J. Climate*, **31**, 1183–1203, https://doi.org/10.1175/JCLI-D-17-0168.1.

Li, J., C. Sun, and F.-F. Jin, 2013: NAO implicated as a predictor of Northern Hemisphere mean temperature multidecadal variability. *Geophys. Res. Lett.*, **40**, 5497–5502, https://doi.org/10.1002/2013GL057877.

Luterbacher, J., and Coauthors, 2001: Extending North Atlantic Oscillation reconstructions back to 1500. *Atmos. Sci. Lett.*, **2**, 114–124, https://doi.org/10.1006/asle.2002.0047.

——, and Coauthors, 2002: Reconstruction of sea level pressure fields over the eastern North Atlantic and Europe back to 1500. *Climate Dyn.*, **18**, 545–561, https://doi.org/10.1007/s00382-001-0196-6.

Mann, M. E., and Coauthors, 2009: Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly. *Science*, **326**, 1256–1260, https://doi.org/10.1126/science.1177303.

Martin, E. R., and C. D. Thorncroft, 2014: The impact of the AMO on the West African monsoon annual cycle. *Quart. J. Roy. Meteor. Soc.*, **140**, 31–46, https://doi.org/10.1002/qj.2107.

——, ——, and B. B. Booth, 2014: The multidecadal Atlantic SST–Sahel rainfall teleconnection in CMIP5 simulations. *J. Climate*, **27**, 784–806, https://doi.org/10.1175/JCLI-D-13-00242.1.

McCabe, G. J., M. A. Palecki, and J. L. Betancourt, 2004: Pacific and Atlantic Ocean influences on multidecadal drought frequency in the united states. *Proc. Natl. Acad. Sci. USA*, **101**, 4136–4141, https://doi.org/10.1073/pnas.0306738101.

McCarthy, G. D., I. D. Haigh, J. J.-M. Hirschi, J. P. Grist, and D. A. Smeed, 2015: Ocean impact on decadal Atlantic climate variability revealed by sea-level observations. *Nature*, **521**, 508–510, https://doi.org/10.1038/nature14491.

Miller, A. J., 1984: Selection of subsets of regression variables. *J. Roy. Stat. Soc.*, **147A**, 389–425, https://doi.org/10.2307/2981576.

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.

——, T. Woollings, and L. Zanna, 2017: The dynamical influence of the Atlantic multidecadal oscillation on continental climate. *J. Climate*, **30**, 7213–7230, https://doi.org/10.1175/JCLI-D-16-0345.1.

——, A. Weisheimer, T. Woollings, L. Gray, and D. MacLeod, 2019: The importance of stratospheric initial conditions for winter North Atlantic oscillation predictability and implications for the signal-to-noise paradox. *Quart. J. Roy. Meteor. Soc.*, **145**, 131–146, https://doi.org/10.1002/QJ.3413.

Ortega, P., F. Lehner, D. Swingedouw, V. Masson-Delmotte, C. C. Raible, M. Casado, and P. Yiou, 2015: A model-tested North Atlantic oscillation reconstruction for the past millennium. *Nature*, **523**, 71–74, https://doi.org/10.1038/nature14518.

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.

Otto-Bliesner, B. L., and Coauthors, 2016: Climate variability and change since 850 CE: An ensemble approach with the Community Earth System Model. *Bull. Amer. Meteor. Soc.*, **97**, 735–754, https://doi.org/10.1175/BAMS-D-14-00233.1.

Palmer, T. N., 1986: Influence of the Atlantic, Pacific and Indian Oceans on Sahel rainfall. *Nature*, **322**, 251–253, https://doi.org/10.1038/322251a0.

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.

——, 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.

Pinto, J. G., and C. C. Raible, 2012: Past and recent changes in the North Atlantic Oscillation. *Wiley Interdiscip. Rev.: Climate Change*, **3**, 79–90, https://doi.org/10.1002/wcc.150.

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.

Robson, J., R. Sutton, K. Lohmann, D. Smith, and M. D. Palmer, 2012: Causes of the rapid warming of the North Atlantic Ocean in the mid-1990s. *J. Climate*, **25**, 4116–4134, https://doi.org/10.1175/JCLI-D-11-00443.1.

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.

——, T. Delworth, R. Msadek, F. Castruccio, S. Yeager, and G. Danabasoglu, 2018: Impacts of the Atlantic multidecadal variability on North American summer climate and heat waves. *J. Climate*, **31**, 3679–3700, https://doi.org/10.1175/JCLI-D-17-0270.1.

Scaife, A. A., and D. Smith, 2018: A signal-to-noise paradox in climate science. *npj Climate Atmos. Sci.*, **1**, 28, https://doi.org/10.1038/S41612-018-0038-4.

Shindell, D. T., G. A. Schmidt, M. E. Mann, and G. Faluvegi, 2004: Dynamic winter climate response to large tropical volcanic eruptions since 1600. *J. Geophys. Res.*, **109**, D05104, https://doi.org/10.1029/2003JD004151.

Simpson, I. R., C. Deser, K. A. McKinnon, and E. A. Barnes, 2018: Modeled and observed multidecadal variability in the North Atlantic jet stream and its connection to sea surface temperatures. *J. Climate*, **31**, 8313–8338, https://doi.org/10.1175/JCLI-D-18-0168.1.

Smerdon, J. E., 2012: Climate models as a test bed for climate reconstruction methods: Pseudoproxy experiments. *Wiley Interdiscip. Rev.: Climate Change*, **3**, 63–77, https://doi.org/10.1002/wcc.149.

Sun, C., J. Li, and F.-F. Jin, 2015: A delayed oscillator model for the quasi-periodic multidecadal variability of the NAO. *Climate Dyn.*, **45**, 2083–2099, https://doi.org/10.1007/s00382-014-2459-z.

——, F. Kucharski, J. Xue, and X. Li, 2019: Contrasting spatial structures of Atlantic multidecadal oscillation between observations and slab ocean model simulations. *Climate Dyn.*, **52**, 1395–1411, https://doi.org/10.1007/S00382-018-4201-8.

Sutton, R. T., and D. L. Hodson, 2005: Atlantic Ocean forcing of North American and European summer climate. *Science*, **309**, 115–118, https://doi.org/10.1126/SCIENCE.1109496.

——, and B. Dong, 2012: Atlantic Ocean influence on a shift in European climate in the 1990s. *Nat. Geosci.*, **5**, 788–792, https://doi.org/10.1038/ngeo1595.

Swingedouw, D., J. Mignot, P. Ortega, M. Khodri, M. Menegoz, C. Cassou, and V. Hanquiez, 2017: Impact of explosive volcanic eruptions on the main climate variability modes. *Global Planet. Change*, **150**, 24–45, https://doi.org/10.1016/j.gloplacha.2017.01.006.

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.

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.

Wang, J., B. Yang, F. C. Ljungqvist, J. Luterbacher, T. J. Osborn, K. R. Briffa, and E. Zorita, 2017a: Internal and external forcing of multidecadal Atlantic climate variability over the past 1,200 years. *Nat. Geosci.*, **10**, 512–517, https://doi.org/10.1038/ngeo2962.

Wang, X., J. Li, C. Sun, and T. Liu, 2017b: 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., 1997: Resampling hypothesis tests for autocorrelated fields. *J. Climate*, **10**, 65–82, https://doi.org/10.1175/1520-0442(1997)010<0065:RHTFAF>2.0.CO;2.

Williams, R. G., V. Roussenov, D. Smith, and M. S. Lozier, 2014: Decadal evolution of ocean thermal anomalies in the North Atlantic: The effects of Ekman, overturning, and horizontal transport. *J. Climate*, **27**, 698–719, https://doi.org/10.1175/JCLI-D-12-00234.1.

——, M. S. Lozier, and D. Smith, 2015: Mechanisms of heat content and thermocline change in the subtropical and subpolar North Atlantic. *J. Climate*, **28**, 9803–9815, https://doi.org/10.1175/JCLI-D-15-0097.1.

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

Xu, X., E. P. Chassignet, and F. Wang, 2019: On the variability of the Atlantic meridional overturning circulation transports in coupled CMIP5 simulations. *Climate Dyn.*, **52**, 6511–6531, https://doi.org/10.1007/s00382-018-4529-0.

Yan, X., R. Zhang, and T. R. Knutson, 2017: The role of Atlantic overturning circulation in the recent decline of Atlantic major hurricane frequency. *Nat. Commun.*, **8**, 1695, https://doi.org/10.1038/s41467-017-01377-8.

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., 2015: Mechanisms for low-frequency variability of summer Arctic sea ice extent. *Proc. Natl. Acad. Sci. USA*, **112**, 4570–4575, https://doi.org/10.1073/PNAS.1422296112.

——, 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 T. L. Delworth, 2006: Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes. *Geophys. Res. Lett.*, **33**, L17712, https://doi.org/10.1029/2006GL026267.

——, 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.

——, R. Sutton, G. Danabasoglu, T. L. Delworth, W. M. 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.