Geophysical Research Letters

RESEARCH LETTER
10.1029/2022GL097957

No Internal Connections Detected Between Low Frequency Climate Modes in North Atlantic and North Pacific Basins

T. Fenske[1] **and A. Clement**[1]

[1]Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL, USA

Abstract Previous studies documented possible connections between low frequency climate modes in the Northern Hemisphere ocean basins. We use observed sea surface temperatures and 270 large ensemble climate model simulations, which allows for improved methods of separating external and internal variability, such as removing the ensemble mean from each simulation. Detrending methods for observations have also improved since some of these previous studies were conducted. We also devise a modified statistical test using bootstrapping that is tuned specifically to this analysis. With these tools, we reexamine relationships among these modes. While previous studies have argued for the existence of an inter-basin link, our results suggest that any internal connections between these modes are indistinguishable from random noise. Further, we show that external forcing affects each region in similar ways. This suggests that anthropogenic warming can cause an indirect link between the two basins, confounding the interpretation of a potential relationship.

Plain Language Summary We reexamine possible connections between climate patterns in the North Atlantic and North Pacific oceans. Improved climate model simulations and appropriate statistical methods allow us to build on previous research of these linkages. In contrast to previous studies, no natural connections are detected. However, global warming is shown to affect each region in similar ways, suggesting that climate change could cause an indirect link between the two basins.

Key Points:

- An analysis of relationships between leading Northern Hemisphere climate modes does not show any internal connections
- External forcing such as global warming is shown to be a possible confounding factor in climate relationships

Supporting Information:

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

Correspondence to:

T. Fenske,
tyler.fenske@rsmas.miami.edu

Citation:

Fenske, T., & Clement, A. (2022). No internal connections detected between low frequency climate modes in North Atlantic and North Pacific basins. *Geophysical Research Letters*, *49*, e2022GL097957. https://doi.org/10.1029/2022GL097957

Received 18 JAN 2022
Accepted 17 FEB 2022

1. Introduction

Low frequency internal climate modes in the Northern Hemisphere oceans operate on similar timescales to various external forcing sources. This can make it difficult to determine their causes: are these modes completely internal, or are they also driven by external forcing? In either case, can a deeper understanding of the sources offer any predictability?

To help address these questions, we examine the time history of the dominant modes of variability in the North Atlantic and North Pacific to test whether some of the variability can be explained by interactions between the different ocean basins. These long-distance effects are referred to as teleconnections and are driven by atmospheric Rossby wave propagation, also known as atmospheric bridges (Alexander et al., 2002; Liu & Alexander, 2007). Sea-surface temperature (SST) variability associated with a particular climate mode is coupled to the atmosphere, allowing the ocean to change the overlying atmospheric circulation. This signal is transported through the atmosphere to other regions, where it can influence SSTs in a distant location, potentially imprinting on or exciting a different climate mode (e.g., Dommenget & Latif, 2008; Liu & Alexander, 2007). Teleconnections can also exhibit variability, such as changing magnitude over time (Raible et al., 2014). The possibility of climate mode interactions must be considered to fully understand the sources of climate variability. We focus our interest on a particular potential teleconnection that may link the North Atlantic (NA) and North Pacific (NP) ocean basins. Do the climate modes in these basins interact with each other? And if so, does this mean that an inter-basin connection exists?

In the NA, low-frequency variability is captured via the Atlantic Multidecadal Oscillation/Variability (AMO or AMV; Enfield et al., 2001). In the NP, two modes are used to capture the low-frequency variability. The Pacific Decadal Oscillation (Mantua et al., 1997) and the Victoria Mode (VM; Bond et al., 2003) or North Pacific Gyre Oscillation (NPGO; Di Lorenzo et al., 2008) are the leading modes of decadal and multidecadal variability respectively. Other methods of capturing variability in these basins exist (e.g., Årthun et al., 2021; Curry & McCartney, 2001; Eden & Jung, 2001; Gastineau & Frankignoul, 2015; Martin et al., 2019; Menary et al., 2015; Nigam et al., 2011; Nigam et al., 2020; Taylor & Stephens, 1998; Salinger et al., 2001; Sun et al., 2015), although convention maintains usage of the AMV and PDO.

Several previous studies have argued that there is a significant relationship between the NA and NP modes. This question was initially examined by d'Orgeville and Peltier (2007) and R. Zhang and Delworth (2007). Both studies relate the first two Empirical Orthogonal Functions (EOFs) of the Hadley Center's observed NP SSTs (Rayner et al., 2003) to a metric for the AMV (first EOF of NA SSTs, NA area mean SSTs respectively), although they utilize those EOFs in different ways. d'Orgeville and Peltier combine them, then isolate the 20-year (decadal) and 60-year (multidecadal) period wavelets, cross correlate the time series from each basin, and conclude that the variability in both regions is connected. Zhang and Delworth also use a 10-member ensemble of one model (GFDL CM2.1, Delworth et al., 2006) where the NA is constrained to match observations to try and reveal teleconnections from the NA to other regions. They conclude that the Atlantic Meridional Overturning Current (AMOC) drives the AMV, which in turn drives the PDO/VM through atmospheric teleconnections. These studies are limited to a relatively short observational period, which makes accurate analysis of low frequency modes difficult. The role of low frequency external forcing may not be adequately removed, particularly in the NA (L. N. Murphy et al., 2017). Ten ensemble members is also near the minimum number needed to accurately estimate the forced response over the ocean (Milinski et al., 2020).

A number of subsequent studies supported the findings of the 2007 papers. Wu et al. (2011) come to a similar conclusion as the prior studies, using a modest but statistically significant cross correlation to argue that the AMV can be used to predict the PDO on annual timescales. Their detrending method is a third order polynomial, which at most grid points is nearly equivalent to quadratic detrending. Marini and Frankignoul (2014) use several methods to deconstruct the AMV origins, such as dynamical filtering and removing different trends. Their analysis includes a cross correlation of the AMV and PDO time series, again showing similar results to previous studies. Nigam et al. (2020) use the first 11 global rotated EOFs to represent major global modes. Their AMV and PDO cross correlation also matches previous results. An et al. (2021) use ensemble pacemaker experiments to suggest that multidecadal Pacific variability is generated by AMV forcing and local air-sea interactions. Most previous studies on the AMV-PDO link agree that a modest but significant correlation can be detected, with the AMV leading the PDO by 12–14 years. All these studies use the student's *t*-test (or a similar test) for significance. Wu et al. (2011) also use customized bootstrap methods.

Calculating significance thresholds for these relatively short timeseries can be challenging. Additionally, since these previous studies were conducted, improved detrending methods have been developed that more accurately separate the forced signal from the internal variability. Frankignoul et al. (2017) used two large ensembles to find more accurate detrending methods in the NA and NP basins, focusing on the PDO and AMV time series. They briefly mention that their AMV-PDO cross correlations show no statistical significance. To expand on this work, we reexamine potential NA-NP connections with observations and six large ensembles of climate model simulations that feature improved methods for separating external and internal variability, and a modified, more appropriate statistical test to reexamine relationships among these modes.

We begin by assessing the statistical tools necessary for our study, leading us to use non-parametric statistics and to modify an existing test so that it is more appropriately optimized for this analysis. Next, we focus on observations, and using our statistical tools we suggest little to no statistically significant relationship can be detected from 121 years of data. We also argue that external forcing may confound the relationship if not accurately removed. Finally, we support this argument by repeating our observational analysis on a set of large ensembles, where we can accurately and completely remove the externally forced response. These results further suggest that no inter-basin connections can be detected on decadal to multidecadal timescales and that external forcing can cause a spurious significant correlation to arise.

2. Methods

For observations, we use the UK Met Office's Hadley Centre Sea Ice and SST data set (HADISST; Rayner et al., 2003). 121 years of annual data from 1900 to 2020 are included. Other SST datasets provide no qualitative differences.

We use six ensembles in the Multi-Model Large Ensemble Archive (MMLEA; CANESM2, CESM, CSIRO-MK36, GFDL-CM3, GFDL-ESM2M, and MPI; C. Deser et al., 2020). The MMLEA contains a total of 270 members (50, 40, 30, 20, 30, and 100 members respectively). All ensembles are from the CMIP5 era and use historical and RCP8.5 forcing. The time period for each model is matched to the observed time period. We use each model's

pre-industrial control run (PI), separate from the MMLEA. More metadata on the MMLEA are provided in Table S1 in Supporting Information S1 (C. Deser et al., 2020).

Our mode definitions generally follow those of the AMV, PDO, and VM (Enfield et al., 2001, Mantua et al., 1997, and Bond et al., 2003 respectively). However, instead of applying different methods (EOF for NP, area mean for NA) to each basin, we apply EOF analysis to each basin to capture individual modes more consistently. We compute the EOFs over the NP (20°N-65°N, 120°E−100°W) and the NA (0°N-65°N, 120°W-0°), and for clarity and brevity, we only show the first two EOFs from each basin in our observational analysis. Including further EOFs does not affect our conclusions (Figures 5–8 in Supporting Information S1).

All observed SST fields are detrended by regressing the global mean SST onto each grid point (hereafter GMR detrending), following Ting et al. (2009), Frankignoul et al. (2017, the REGR method), and C. lara. Deser and Phillips (2021). MMLEA data is detrended for each of the six models by removing the ensemble mean from every member. The advent of large ensembles has been a primary driver in improving detrending methods due to their ability to accurately estimate a particular model's true forced response by averaging out internal variability (C. Deser et al., 2020; Kay et al., 2015; Lehner et al., 2020). Frankignoul et al. (2017), using two large ensembles, provide a thorough analysis of four relatively simple detrending methods (three more complex detrending methods are also covered, however these are beyond the scope of our study). They show that quadratic detrending and a regression of the global mean temperature (C. Deser and Phillips, 2021; Ting et al., 2009) generally perform better than other simple methods, although their performance does vary spatially. They also show that linear detrending, the simplest and most common method, is largely ineffective at removing the externally forced signal.

We define “ensemble mean” as averaging across MMLEA members, then computing the EOF analysis. We define “composite mean” as EOF analysis performed on each ensemble member once the ensemble mean has been removed (i.e., each member has been detrended) and followed by averaging across MMLEA members. In this study, we use the ensemble mean EOF to represent the forced signal in each model. We use the composite mean EOF exclusively on detrended MMLEA data to represent the average internal variability in each model. Averaging across all ensemble members may mutually cancel out polarized clusters of members, so caution must be used when computing ensemble or composite means (Bellucci et al., 2017).

All SST data are normalized and smoothed at each grid point with a 10-year low-pass Lanczos filter to focus on decadal to multi-decadal variability. We compute spatial pattern correlations to match EOFs for each MMLEA member to their corresponding observed EOF. We then reorder and flip the sign of MMLEA EOFs where necessary to make qualitative assessments clearer and easier.

We use a non-parametric bootstrapping method for significance testing. Figure 1 includes normality tests for our observed time series which show that they are not consistent with a normal distribution, requiring non-parametric statistics. This is further discussed in Results. To summarize, statistical tests used in previous studies rely on assumptions such as consistence with normality and estimated degrees of freedom. Our primary statistical tool is cross-correlation, so we build our method to evaluate the significance of a “real” cross-correlation with as few assumptions as possible. We create sets of phase randomized (Ebisuzaki, 1997) time series by shuffling each unfiltered time series 1000 times. Each shuffled time series is filtered, and each possible pair of modes is cross-correlated for the entire set. These 1000 cross-correlations are used to compute 0.1 significance levels for the “real” cross-correlation.

We also use sets of white noise, AR1 red noise (Katz, 1982), and quantile-mapped sets (Maraun, 2013). While these methods provide no qualitative differences, we do note that the bootstrapped-white-noise method produces the lowest thresholds of the four methods. Any model that is significant with phase randomization, AR1 or quantile-mapped statistics would also be significant with white noise testing. We elect to use the phase randomized method here as it provides the most physically realistic random time series, although our conclusions would not change if we used any other methods.

We note that the phase randomized significance levels are dependent on the frequency spectra and length of the time series. The former dependency can be seen clearly in Figure 2, where the significance level for the upper left panel is higher than the rest. This is because the non-detrended NA1 and NP1 time series are dominated by a trend, so the frequency spectra is that of a time series with a clear, strong trend. This means that all phase randomized time series will exhibit a similar trend. This common trend drives the correlation higher, resulting in a

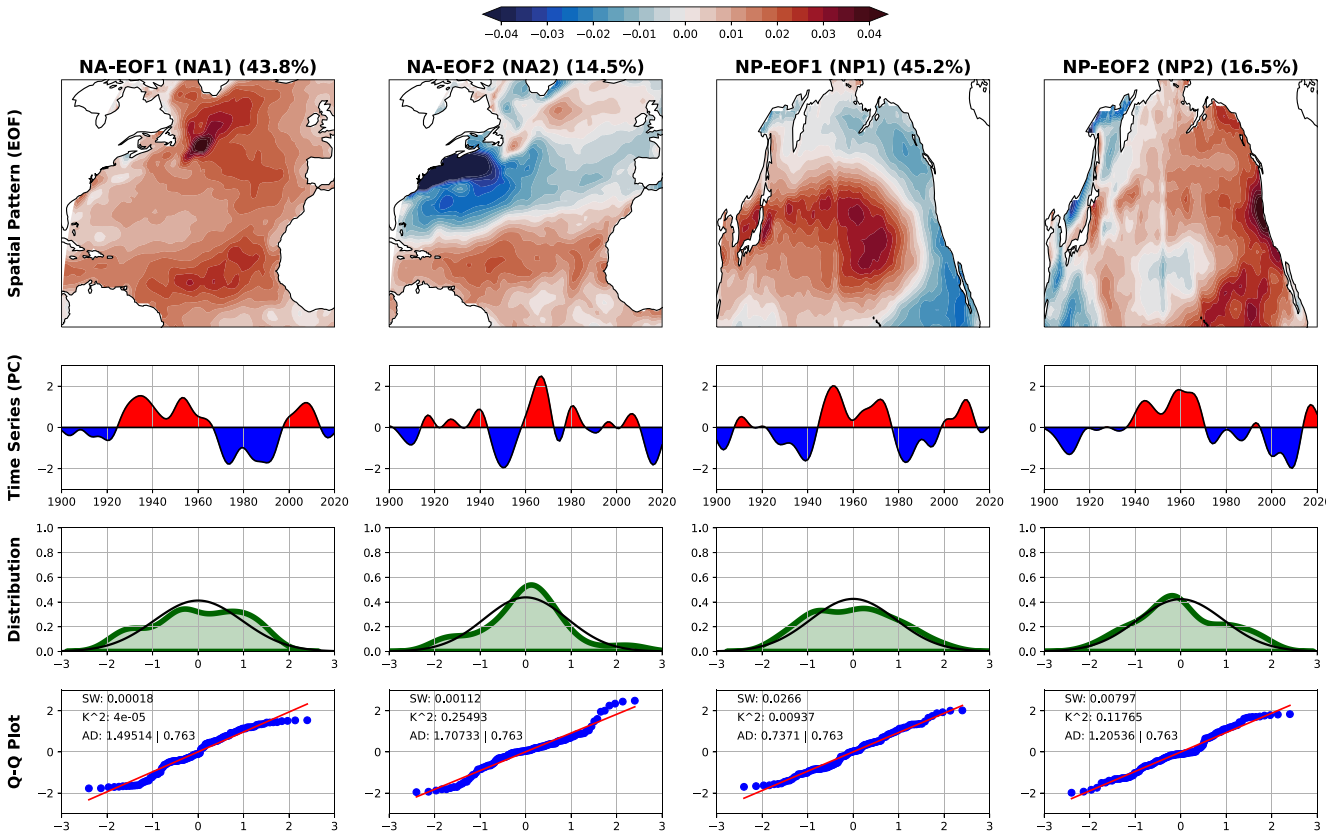

Figure 1. Empirical Orthogonal Functions (EOF) modes and normality analysis of observed detrended (via global mean regression) HadISST North Atlantic and North Pacific SSTs. From left to right: NA-EOF1 (NA1), NA-EOF2 (NA2), NP-EOF1 (NP1), and NP-EOF2 (NP2). The percentage represents the variance explained by each EOF. Rows from top to bottom: 1) Spatial patterns of each EOF. Red (blue) corresponds to warming (cooling) when time series is positive (negative). 2) Filtered normalized time series of each EOF. 3) Probability distribution of time series are shown in shaded green; corresponding standard normal distributions are shown in black. 4) Quantile-quantile plots, data shown in blue. Standard normal distribution is shown in red. Three quantitative assessments of normality are applied: 1) The Shapiro-Wilk Test (SW). The p-value is listed, with p-values >0.05 implying consistence with normality. 2) D'Agostino's K-Squared Test (K^2). The p-value is listed, with p-values > 0.05 implying consistence with normality. 3) The Anderson-Darling Test (AD). First value is measure of consistence with normality; second value is 95% critical value for the AD test. If the first value < critical value, consistence with normality is implied.

higher significance level. The dependency on time series length would suggest that the significance level should be lower near 0 lag (where both time series are 121 years in length) and higher near the maximum lags (where shifted time series only overlap for 81 years). Testing reveals that this dependency is relatively small (not shown) and does not qualitatively affect our results.

This method is similar to the bootstrap used by Wu et al. (2011), although our method differs slightly. Wu et al. (2011) calculate the 0.05 significance level at each specific lag in their cross correlation (hereafter the "point test"). However, their interest is not on a particular lag, but on the peaks of the cross-correlation that are significant. The specific lag these peaks occur at is unimportant – whether it is at 0- or 30-year or any other lag, their conclusions would remain the same. Their statistical test is computed at each individual lag, so that the significance level at each lag is independent from the rest. Their statistics can be interpreted as meaning that the lag at which the peaks occur *is* important, contrary to their conclusions. This is an a priori test with an a posteriori conclusion, suggesting their significance thresholds may not be appropriate.

Alternatively, a "peak test" can be used. Instead of calculating the significance thresholds at each lag, we choose the maximum value of each random cross-correlation to compute the thresholds from. In essence, the minimum and maximum values from each of the 1000 cross correlations can be used to compute significance thresholds that test the observed cross correlation peaks appropriately. The effect of this difference is seen at the 0.05 significance level, where 5% of random cross-correlations have any points that are significant when using the peak test, while our testing shows that ∼50% have significant points when using the point test (examples shown

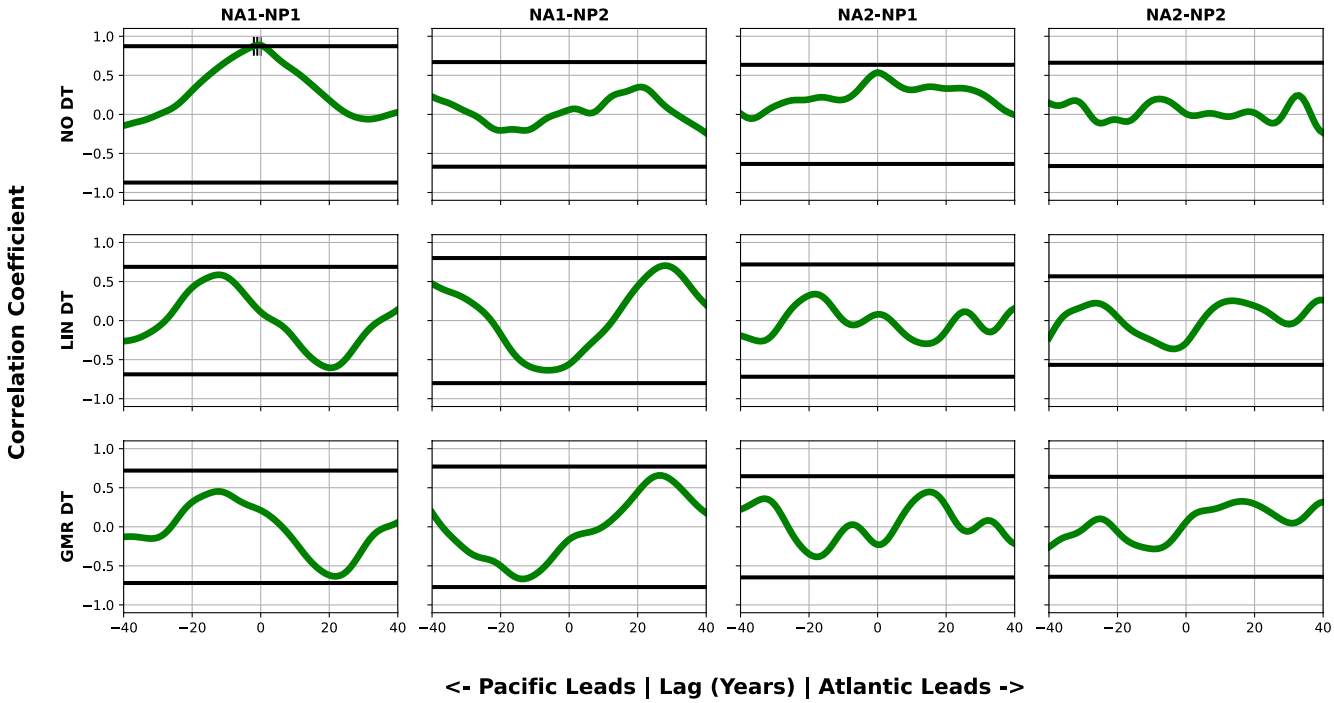

Figure 2. Cross-correlation matrix of HADISST observed time series with various detrending methods. All possible cross-correlations between each of the two Empirical Orthogonal Functions modes in each basin are shown. From left to right: NA1-NP1, NA1-NP2, NA2-NP1, and NA2-NP2. Detrending methods are shown on the *y*-axis, from top to bottom: no detrending, linear detrending, and global mean regression detrending. Black horizontal lines are 0.1 statistical significance levels as calculated by the "peak" test. Vertical ticks show where cross-correlations are significant. For positive (negative) lags, the NA (NP) mode leads.

in Supplemental Figure 1). Additionally, because the point test is computed at each lag, a number of points may be spuriously significant depending on the significance level (e.g., with 81 lags and 0.05 significance, 4 points can be significant without any meaning). The effect of filtering must also be considered (Cane et al., 2017). This is typically considered by estimating the degrees of freedom, although here we can simply filter the randomized time series the same way and automatically account for filtering.

In essence, the statistical tests from previous studies (the point test) are not incorrect, they simply require careful consideration of their context. The peak test accounts for most of this context automatically, allowing for a more objective conclusion. This shows that using an improper significance test can result in many spurious significant points on any given cross-correlation. All significance levels shown here are computed using the peak test. We note that we use the 0.1 significance level since we are trying to show that our cross correlations are **not** significant. A lower significance level therefore strengthens our conclusions.

3. Results

Figure 1 shows the spatial patterns (first row) and time series (second row) of these EOF modes. NA-EOF1 (NA1, NA2 for NA-EOF2) represents the AMV, characterized by a tripole pattern and multi-decadal time series. NP-EOF1 (NP1) represents the PDO, characterized by a dipole pattern with warming (cooling) in the central NP and cooling (warming) in the east, with a decadal time series. NP-EOF2 (NP2) represents the VM, characterized by a dipole pattern offset to the west from the PDO's, and a decadal time series. Supplemental Figure 2 shows the observed spatial patterns and time series for the non-detrended, linear detrended, and GMR detrended observations. Supplemental Figure 3 shows the composite mean spatial patterns of each mode compared to observations. The similarities suggest that the models are simulating realistic modes.

To determine if standard parametric statistics can be used, the normality of our time series is assessed. While non-Gaussian parametric statistics exist, Gaussian assumptions are common with geophysical time series analysis and are used by the previous studies on the NA-NP inter-basin relationships (d'Orgeville & Peltier, 2007, R. Zhang and Delworth, 2007, Wu et al., 2011, Marini & Frankignoul, 2014, and Nigam et al., 2020). We test the

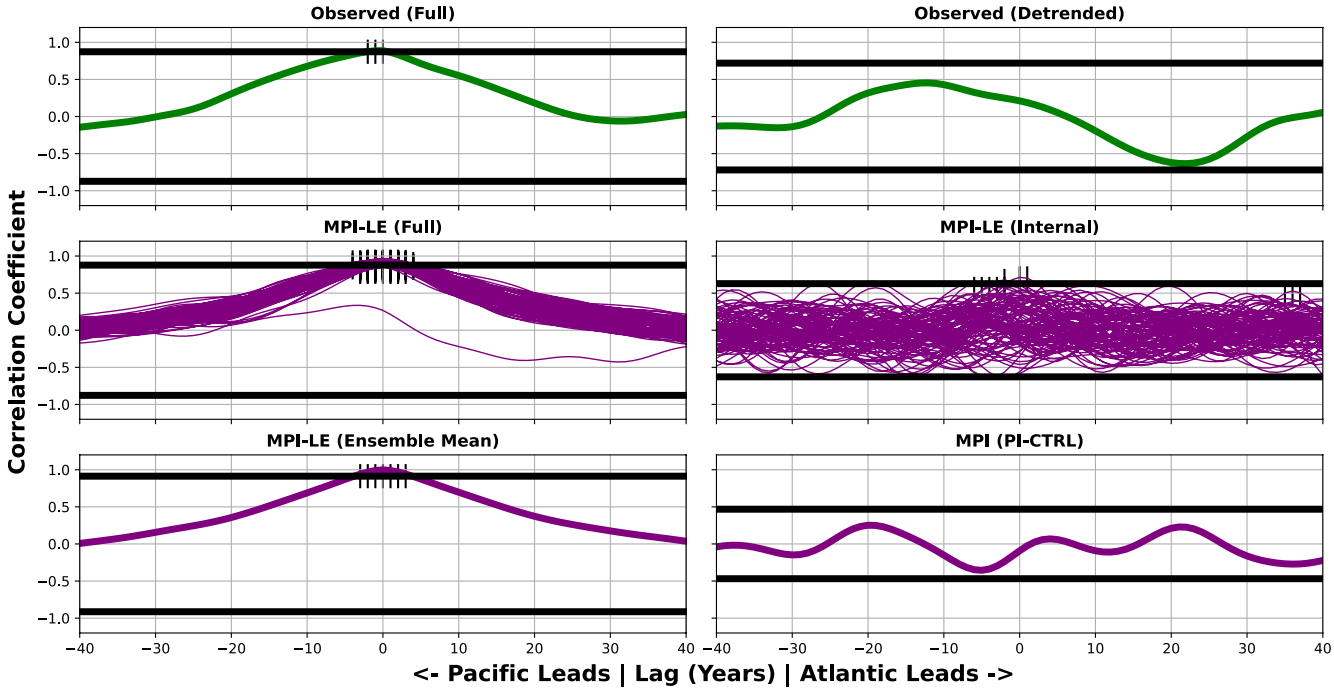

Figure 3. Cross-correlations of NA1-NP1 mode relationships for the MPI-LE. Left panels show relationships that contain full (non-detrended) forced variability. Right panels show only internal (ensemble mean detrended) variability. Top row: observations. Middle row: MPI-LE from the MMLEA. Bottom row: MPI-LE ensemble mean (left) and MPI pre-industrial control run (right). Black horizontal lines are,1 statistical significance levels as calculated by the "peak" test. Vertical ticks show where cross-correlations are significant. For positive (negative) lags, the NA (NP) mode leads.

four modes' filtered time series with five normality tests as recommended by Yap and Sim (2011) and Ghasemi and Zahediasl (2012). Two are qualitative assessments: a histogram with a standard normal curve fitted to the data (Figure 1, third row), and a quantile-quantile plot (Figure 1, fourth row). Three are quantitative: the Shapiro-Wilk test (SW), the D'Agostino skewness test (K^2), and the Anderson-Darling test (AD; shown as text in Figure 1, fourth row). Only NA2 passes more than a single quantitative test. The combination of tests suggests that only NA2 of the filtered time series can be described as consistent with normality, so Gaussian assumptions cannot be made. Non-parametric statistics are required for significance testing.

Figure 2 shows observed cross-correlations for the four possible relationships: NA1-NP1, NA1-NP2, NA2-NP1, and NA2-NP2. The top row shows the full, or non-detrended, observed data, the middle row shows the linearly detrended data, and the bottom row shows the GMR detrended data. Horizontal black lines represent 0.1 significance levels. NA1-NP1 with no detrending shows the only cross-correlation with any statistically significant points, with a nearly perfect correlation very near 0 lag. NA2-NP1 shows a roughly similar cross-correlation, but with a smaller maximum correlation that is not significant. NA1-NP2 and NA2-NP2 show no significant points. Our cross correlations match the results from previous studies (e.g., d'Orgeville & Peltier, 2007; R. Zhang and Delworth, 2007), although our results show no statistical significance because of our modified significance testing.

We also tested quadratic and global mean subtraction detrending methods and find that the observed results are not sensitive to the detrending method. Therefore, we show only the linear and GMR methods in Figure 1. Given that only the full NA1-NP1 relationship shows any significance, we choose to show only the NA1-NP1 relationship for our MMLEA analysis. All other relationships in the MMLEA support similar conclusions as the observations (Figures 5–8 in Supporting Information S1) - that no statistically significant connections are present, either in observations or models.

It is difficult to make any robust conclusions on low-frequency modes using only the relatively short observational period. The true forced signal must be estimated as well, leaving room for errors to arise from imperfect detrending. MMLEA data helps to address these challenges by offering 270 simulations and each model's true forced response. Figure 3 focuses on the NA1-NP1 relationships for the MPI-LE, with full variability (non-detrended)

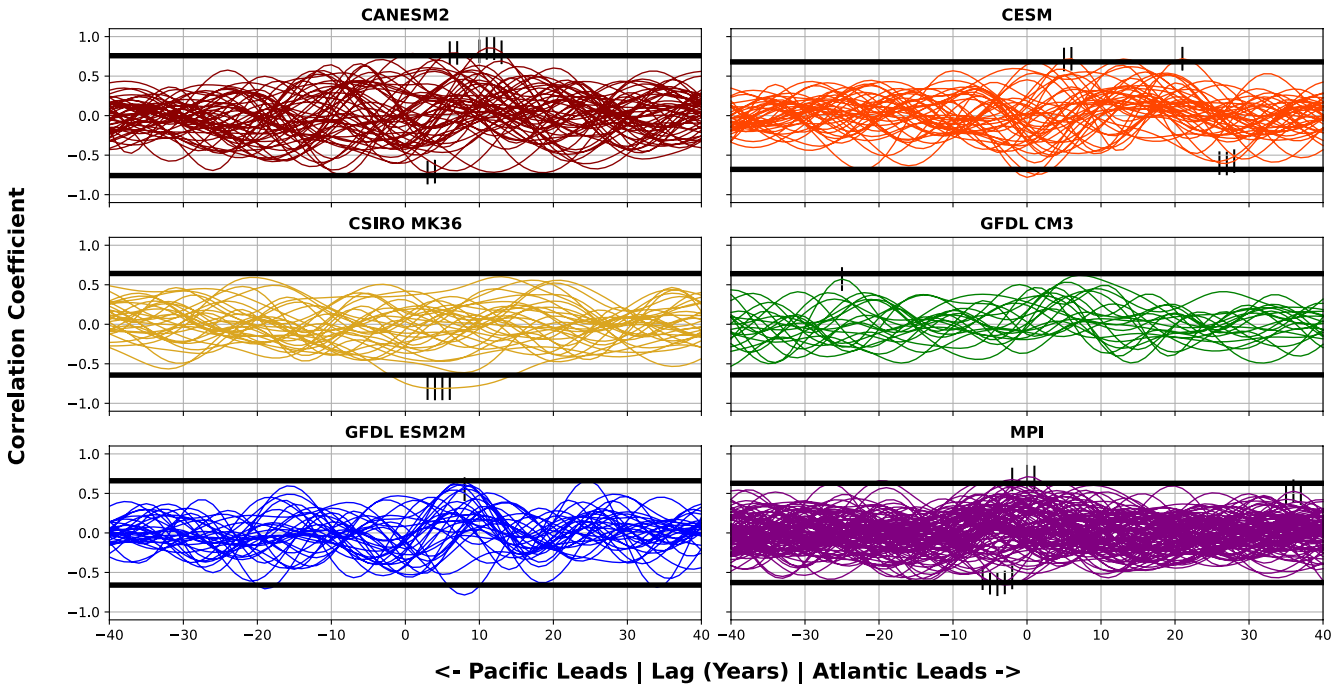

Figure 4. Cross-correlations of MMLEA internal only (ensemble mean detrended) NA1-NP1 mode relationships for all LEs. Black horizontal lines are 0.1 significance levels as calculated by the "peak" test. Vertical ticks show where cross-correlations are significant. For positive (negative) lags, the NA (NP) mode leads.

on the left and internal variability (detrended) on the right. We choose the MPI-LE as it contains 100 members, over one third of the MMLEA's total members, therefore providing the largest sample size for a single LE. In the first row are observations, the same as the top left and bottom left panels from Figure 2. The second row shows the MPI-LE cross-correlations. The full, non-detrended cross-correlations show agreement with the non-detrended observations, with a significant peak near 0 lag. Almost all members match each other and 93 have significant peaks, showing a clear pattern. In contrast, the internal only variability has only four statistically significant peaks. Clearly, in the absence of external forcing, no significant connection between the two basins can be detected in the MPI-LE.

The third row shows NA1-NP1 of the ensemble mean for MPI-LE, which is the model's true forced signal. In every model, the first EOFs (NA1 and NP1) of the ensemble mean contains >95% of the variance explained, meaning that it is represents virtually all of the forced signal. Other methods (e.g., area mean, linear regressions) of representing the forced signal produce similar spatial patterns and time series. Again, we see a clear resemblance to the other full cross-correlations. This suggests that those relationships are dominated by external forcing. In the same row is also the MPI pre-industrial control run, which, despite a much lower significance level due to its much longer period, has no significant points. We also compute the internal composite mean (not shown), or the average internal variability, in the MPI-LE, and again find no significance.

Finally, Figure 4 shows the internal only (ensemble mean detrended) NA1-NP1 relationships for all members of the MMLEA, organized by LE. Note that the middle right panel in Figure 3 and the bottom right panel in Figure 4 are the same. Supplemental Figure 4 shows the same, but for full (non-detrended) variability to show the effect of including forcing. We do note that the full relationships show model dependence, with the CSIRO-LE and (to a lesser extent) the CESM-LE appearing less cohesive and noisier. This is due to the NP1 time series, as these two models show more NP1 variance from member to member than the other four models. The cause of this increased variance may be attributed to differences in model physics and forcing parameters, but this is beyond our scope. Across all 270 members, only 13 (4.8%) show an internal connection, while 189 (70%) members show a significant relationship in the non-detrended data. Thus, the only relationships that can be considered significant are those that contain external forcing. With more appropriate statistical analysis and separation of forced and internal variability (and even imperfect separation, in the case of observations), there are no significant internal connections between the NA and NP detected.

4. Discussion

Our results suggest that, for the NA and NP basins, an internal connection between the two cannot be detected currently. These findings may also have broader implications regarding the roles of external forcing and internal variability as drivers of climate modes. Present theories on climate mode drivers focus on varying roles for the ocean, atmosphere, internal variability, and external forcing (e.g., Clement et al., 2015; Klavans et al., 2022; L. N. Murphy et al., 2021; Newman et al., 2016; O'Reilly et al., 2019; Wills et al., 2018; Y. Zhang et al., 2018; Zhang et al., 2019). External forcing is particularly challenging, due to the direct/linear and indirect/non-linear effects on climate variability (Frankignoul et al., 2017; Li et al., 2020). Additionally, there is debate about the role and importance of these internal modes in an increasingly forced global climate (Ting et al., 2009; DelSole et al., 2011; Haustein et al., 2019; Mann et al., 2020, etc.).

Our results suggest that traditional observed AMV and PDO definitions contain at least some external forcing. Care must be taken when using these mode definitions to properly remove the forced signal to isolate the internal variability. However, this remains a challenging task outside of the realm of large ensembles. Tools such as the MMLEA will be vital in making progress toward isolating the forced response in observations, as they can possibly average out various model differences and provide a closer analog to the observed forced signal. These modeling tools are especially useful due to the relatively short observed period, which may not be sufficient to adequately observe variability on multi-decadal timescales, such as those studied here.

Finally, our findings do not rule out other regions or modes driving variability in the NA and NP. Other mode relationships have been shown to exist, such as how ENSO helps drive the PDO (Newman et al., 2003). Further work can include a complex matrix of potential relationships between global modes as in Shin et al. (2010) or using global EOFs such as Nigam et al. (2020). Methods presented here can assist in a thorough decomposition of sources of variability in a particular region. This can lead to better understanding of variability drivers, which can ultimately result in improved climate models and more accurate climate forecasts.

Data Availability Statement

Data Access: MMLEA data were downloaded from https://www.cesm.ucar.edu/projects/community-projects/MMLEA/(C. Deser et al., 2020). Hadley Center Observed SST data (HADISST) were downloaded from https://www.metoffice.gov.uk/hadobs/hadisst/ (Rayner et al., 2003).

Acknowledgments
The authors gratefully acknowledge funding from the NSF Climate and Large Scale Dynamics program and NOAA Climate Program Office that supported this work. We also acknowledge all modelling groups and the US CLIVAR Working Group on Large Ensembles for making their data available in the Multi-Model Large Ensemble data repository. We thank Brian Mapes and his Applied Data Analysis class for inspiration and discussions on proper statistical analysis. We would also like to gratefully acknowledge the anonymous reviewers, whose comments have significantly improved the manuscript.

References

Alexander, M. A., Bladé, I., Newman, M., Lanzante, J. R., Lau, N.-C., & Scott, J. D.. (2002). The atmospheric bridge: The influence of ENSO teleconnections on air–sea interaction over the global oceans. *Journal of Climate 15*(16). 2205–2231. < https://doi.org/10.1175/1520-0442(2002)015<2205:TABTIO>2.0.CO;2>

An, X., Wu, B., Zhou, T., & Liu, B. (2021). Atlantic multidecadal oscillation drives interdecadal Pacific variability via tropical atmospheric bridge. *Journal of Climate*, *34*(13), 5543–5553. https://doi.org/10.1175/jcli-d-20-0983.1

Årthun, M., Wills, R. C. J., Johnson, H. L., Chafik, L., & Langehaug, H. R. (2021). Mechanisms of decadal North Atlantic climate variability and implications for the recent cold anomaly. *Journal of Climate*, *34*(9), 3421–3439.

Bellucci, A., Mariotti, A., & Gualdi, S. (2017). The role of forcings in the twentieth-century North Atlantic multidecadal variability: The 1940–75 North Atlantic cooling case study. *Journal of Climate*, *30*(18), 7317–7337. https://doi.org/10.1175/jcli-d-16-0301.1

Bond, N. A., Overland, J. E., Spillane, M., & Stabeno, P. (2003). Recent shifts in the state of the North Pacific. *Geophysical Research Letters*, *30*, 2183. https://doi.org/10.1029/2003gl018597

Cane, M. A., Clement, A. C., Murphy, L. N., & Bellomo, K. (2017). Low-pass filtering, heat flux, and Atlantic multidecadal variability. *Journal of Climate*, *30*(18), 7529–7553. https://doi.org/10.1175/jcli-d-16-0810.1

Clement, A., Bellomo, K., Murphy, L. N., Cane, M. A., Mauritsen, T., Radel, G., & Stevens, B. (2015). The Atlantic Multidecadal Oscillation without a role for ocean circulation. *Science*, *350*(6258), 320–324. https://doi.org/10.1126/science.aab3980

Curry, R. G., & McCartney, M. S., (2001). Ocean gyre circulation changes associated with the North Atlantic Oscillation. *Journal of Physical Oceanography*, *31*, 3374–3400. https://doi.org/10.1175/1520-0485(2001)031<3374:ogccaw>2.0.co;2

DelSole, T., Tippett, M. K., & Shukla, J. (2011). A significant component of unforced multidecadal variability in the recent acceleration of global warming. *Journal of Climate*, *24*(3), 909–926. https://doi.org/10.1175/2010jcli3659.1

Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A., et al. (2006). GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *Journal of Climate*, *19*(5), 643–674. https://doi.org/10.1175/jcli3629.1

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., et al. (2020). Insights from Earth system model initial-condition large ensembles and future prospects. *Nature climate change*, *10*, 277–286. https://doi.org/10.1038/s41558-020-0731-2

Deser, C., & Phillips, A. S. (2021). Defining the internal component of Atlantic multidecadal variability in a changing climate. *Geophysical Research Letters*, *48*(22), e2021GL095023. https://doi.org/10.1029/2021gl095023

Dommenget, D., & Latif, M. (2008). Generation of hyper climate modes. *Geophysical Research Letters*, *35*(2), L02706. https://doi.org/10.1029/2007GL031087

d'Orgeville, M., & Peltier, W. R. (2007). On the Pacific decadal oscillation and the Atlantic multidecadal oscillation: Might they be related?: Pdo and AMO related? *Geophysical Research Letters*, *34*, L23705. https://doi.org/10.1029/2007GL031584

Ebisuzaki, W. (1997). A method to estimate the statistical significance of a correlation when the data are serially correlated. *Journal of Climate*, *10*(9), 2147–2153. https://doi.org/10.1175/1520-0442(1997)010<2147:amtets>2.0.co;2

Eden, C., & Jung, T. (2001). North Atlantic interdecadal variability: Oceanic response to the North Atlantic oscillation (1865–1997). *Journal of Climate*, *14*, 16. https://doi.org/10.1175/1520-0442(2001)014<0676:naivor>2.0.co;2

Enfield, D. B., Mestas-Nuñez, A. M., & Trimble, P. J. (2001). The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental U.S. *Geophysical Research Letters*, *28*(10), 2077–2080. https://doi.org/10.1029/2000gl012745

Frankignoul, C., Gastineau, G., & Kwon, Y.-O. (2017). Estimation of the SST response to anthropogenic and external forcing and its impact on the Atlantic multidecadal oscillation and the Pacific decadal oscillation. *Journal of Climate*, *30*(24), 9871–9895. https://doi.org/10.1175/jcli-d-17-0009.1

Gastineau, G., & Frankignoul, C. (2015). Influence of the North Atlantic SST variability on the atmospheric circulation during the twentieth century. *Journal of Climate*, *28*, 1396–1416, https://doi.org/10.1175/JCLI-D-14-00424.1

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, *10*(2), 486–489. https://doi.org/10.5812/ijem.3505

Haustein, K., Otto, F. E. L., Venema, V., Jacobs, P., Cowtan, K., Hausfather, Z., et al. (2019). A limited role for unforced internal variability in twentieth-century warming. *Journal of Climate*, *32*(16), 4893–4917. https://doi.org/10.1175/jcli-d-18-0555.1

Katz, R. W. (1982). Statistical evaluation of climate experiments with general circulation models: A parametric time series modeling approach. *Journal of the Atmospheric Sciences*, *39*(7), 1446–1455. https://doi.org/10.1175/1520-0469(1982)039<1446:seocew>2.0.co;2

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The community earth system model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349. https://doi.org/10.1175/bams-d-13-00255.1

Klavans, J. M., Clement, A. C., Cane, M. A., & Murphy, L. N. (2022). The evolving role of external forcing in North Atlantic SST variability over the last millennium. *Journal of Climate*, 1, 44. (published online ahead of print). https://doi.org/10.1175/JCLI-D-21-0338.1

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., et al. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth System Dynamics*, *11*(2), 491–508. https://doi.org/10.5194/esd-11-491-2020

Li, S., Wu, L., Yang, Y., Geng, T., Cai, W., Gan, B., et al. (2020). The Pacific Decadal Oscillation less predictable under greenhouse warming. *Nature Climate Change*, *10*(1), 30–34. https://doi.org/10.1038/s41558-019-0663-x

Liu, Z., & Alexander, M. (2007). Atmospheric bridge, oceanic tunnel, and global climatic teleconnections. *Reviews of Geophysics*, *45*(2), RG2005. https://doi.org/10.1029/2005RG000172

Lorenzo, E. D., Schneider, N., Cobb, K. M., Franks, P. J. S., Chhak, K., Miller, A. J., et al. (2008). North Pacific Gyre Oscillation links ocean climate and ecosystem change. *Geophysical Research Letters*, *35*, L08607. https://doi.org/10.1029/2007gl032838

Mann, M. E., Steinman, B. A., & Miller, S. K. (2020). Absence of internal multidecadal and interdecadal oscillations in climate model simulations. *Nature Communications*, *11*(1), 49. https://doi.org/10.1038/s41467-019-13823-w

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997). A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society*, *78*(6), 1069–1079. https://doi.org/10.1175/1520-0477(1997)078<1069:apicow>2.0.co;2

Maraun, D. (2013). Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *Journal of Climate*, *26*(6), 2137–2143. https://doi.org/10.1175/jcli-d-12-00821.1

Marini, C., & Frankignoul, C. (2014). An attempt to deconstruct the Atlantic multidecadal oscillation. *Climate Dynamics*, *43*(3), 607–625. https://doi.org/10.1007/s00382-013-1852-3

Martin, T., Reintges, A., & Latif, M. (2019). Coupled North Atlantic subdecadal variability in CMIP5 models. *Journal of Geophysical Research: Oceans*, *124*(4), 2404–2417. https://doi.org/10.1029/2018jc014539

Menary, M. B., Hodson, D. L., Robson, J. I., Sutton, R. T., & Wood, R. A. (2015). A mechanism of internal decadal variability in a high-resolution coupled climate model. *Journal of Climate*, *28*, 7764–7785. https://doi.org/10.1175/JCLI-D-15-0106.1

Milinski, S., Maher, N., & Olonscheck, D. (2020). How large does a large ensemble need to be? *Earth System Dynamics*, *11*(4), 885–901. https://doi.org/10.5194/esd-11-885-2020

Murphy, L. N., Bellomo, K., Cane, M., & Clement, A. (2017). The role of historical forcings in simulating the observed Atlantic multidecadal oscillation. *Geophysical Research Letters*, *44*(5), 2472–2480. https://doi.org/10.1002/2016gl071337

Murphy, L. N., Klavans, J. M., Clement, A. C., & Cane, M. A.. (2021). Investigating the roles of external forcing and ocean circulation on the Atlantic multidecadal SST variability in a large ensemble climate model hierarchy. *Journal of Climate*, *34*(12). 4835–4849. https://doi.org/10.1175/JCLI-D-20-0167.1

Newman, M., Alexander, M. A., Ault, T. R., Cobb, K. M., Deser, C., Di Lorenzo, E., et al. (2016). The Pacific decadal oscillation, revisited. *Journal of Climate*, *29*(12), 4399–4427. https://doi.org/10.1175/jcli-d-15-0508.1

Newman, M., Compo, G. P., & Alexander, M. A. (2003). ENSO-forced variability of the Pacific decadal oscillation. *Journal of Climate*, *16*(23), 3853–3857. https://doi.org/10.1175/1520-0442(2003)016<3853:evotpd>2.0.co;2

Nigam, S., Guan, B., & Ruiz-Barradas, A. (2011). Key role of the Atlantic multidecadal oscillation in 20th century drought and wet periods over the Great Plains. *Geophysical Research Letters*, *38*, L16713. https://doi.org/10.1029/2011GL048650

Nigam, S., Sengupta, A., & Ruiz-Barradas, A. (2020). Atlantic–pacific links in observed multidecadal SST variability: Is the Atlantic multidecadal oscillation's phase reversal orchestrated by the Pacific decadal oscillation? *Journal of Climate*, *33*(13), 5479–5505. https://doi.org/10.1175/jcli-d-19-0880.1

O'Reilly, C. H., Zanna, L., & Woollings, T. (2019). Assessing external and internal sources of Atlantic multidecadal variability using models, proxy data, and early instrumental indices. *Journal of Climate*, *32*(22), 7727–7745.

Raible, C. C., Lehner, F., González-Rouco, J. F., & Fernández-Donado, L. (2014). Changing correlation structures of the Northern Hemisphere atmospheric circulation from 1000 to 2100 AD. *Climate of the Past*, *10*(2), 537–550. https://doi.org/10.5194/cp-10-537-2014

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*, 4407. https://doi.org/10.1029/2002jd002670

Salinger, M. J., Renwick, J. A., & Mullan, A. B. (2001). Interdecadal Pacific oscillation and south Pacific climate. *International Journal of Climatology*, *21*(14), 1705–1721. https://doi.org/10.1002/joc.691

Shin, S.-I., Sardeshmukh, P. D., & Pegion, K. (2010). Realism of local and remote feedbacks on tropical sea surface temperatures in climate models. *Journal of Geophysical Research*, *115*, D21110. https://doi.org/10.1029/2010jd013927

Sun, C., Li, J., & Jin, F.-F. (2015). A delayed oscillator model for the quasi-periodic multidecadal variability of the NAO. *Climate Dynamics*, *45*(7–8), 2083–2099. https://doi.org/10.1007/s00382-014-2459-z

Taylor, A. H., & Stephens, J. A. (1998). The North Atlantic oscillation and the latitude of the gulf stream. *Tellus*, *50*, 134–142. https://doi.org/10.1034/j.1600-0870.1998.00010.x

Ting, M., Kushnir, Y., Seager, R., & Li, C. (2009). Forced and internal twentieth-century SST trends in the North Atlantic. *Journal of Climate*, *22*(6), 1469–1481. https://doi.org/10.1175/2008jcli2561.1

Wills, R. C., Schneider, T., Wallace, J. M., Battisti, D. S., & Hartmann, D. L. (2018). Disentangling global warming, multidecadal variability, and el niño in Pacific temperatures. *Geophysical Research Letters*, *45*(5), 2487–2496. https://doi.org/10.1002/2017gl076327

Wu, S., Liu, Z., Zhang, R., & Delworth, T. L. (2011). On the observed relationship between the Pacific decadal oscillation and the Atlantic multi-decadal oscillation. *Journal of Oceanography*, *67*(1), 27–35. https://doi.org/10.1007/s10872-011-0003-x

Yap, B. W., & Sim, C. H. (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, *81*(12), 2141–2155. https://doi.org/10.1080/00949655.2010.520163

Zhang, R., & Delworth, T. L. (2007). Impact of the Atlantic multidecadal oscillation on North Pacific climate variability: Impact ON North Pacific variability. *Geophysical Research Letters*, *34*. L23708. https://doi.org/10.1029/2007gl031601

Zhang, Y., Xie, S.-P., Kosaka, Y., & Yang, J.-C. (2018). Pacific decadal oscillation: Tropical Pacific forcing versus internal variability. *Journal of Climate*, *31*(20), 8265–8279. https://doi.org/10.1175/jcli-d-18-0164.1

Zhang, R., Sutton, R., Danabasoglu, G., Kwon, Y., Marsh, R., Yeager, S. G., et al. (2019). A Review of the Role of the Atlantic Meridional Overturning Circulation in Atlantic Multidecadal Variability and Associated Climate Impacts. *Reviews of Geophysics*, *57*(2). 316–375. https://doi.org/10.1029/2019rg000644