Low-Pass Filtering, Heat Flux, and Atlantic Multidecadal Variability

MARK A. CANE

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

AMY C. CLEMENT AND LISA N. MURPHY

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

KATINKA BELLOMO

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

(Manuscript received 12 November 2016, in final form 8 June 2017)

ABSTRACT

In this model study the authors explore the possibility that the internal component of the Atlantic multidecadal oscillation (AMO) sea surface temperature (SST) signal is indistinguishable from the response to white noise forcing from the atmosphere and ocean. Here, complex models are compared without externally varying forcing with a one-dimensional noise-driven model for SST. General analytic expressions are obtained for both unfiltered and low-pass filtered lead–lag correlations. It is shown that this simple model reproduces many of the simulated lead–lag relationships among temperature, rate of change of temperature, and surface heat flux. It is concluded that the finding that at low frequencies the ocean loses heat to the atmosphere when the temperature is warm, which has been interpreted as showing that the ocean circulation drives the AMO, is a necessary consequence of the fact that at long periods the net heat flux (ocean plus atmosphere) is zero to a good approximation. It does not distinguish between the atmosphere and ocean as the source of the AMO and is consistent with the hypothesis that the AMO is driven by white noise heat fluxes. It is shown that some results in the literature are artifacts of low-pass filtering, which creates spurious low-frequency signals when the underlying data are white or red noise. It is concluded that in the absence of external forcing the AMO in most GCMs is consistent with being driven by white noise, primarily from the atmosphere.

1. Introduction

Our working hypothesis in this paper is that in CMIP preindustrial (PI) simulations and, more generally, in the absence of variable external forcing, the Atlantic multidecadal oscillation (AMO) sea surface temperature (SST) is a response to white noise forcing reddened by the heat capacity of the ocean mixed layer. The noise source may be the atmosphere or the ocean. The North Atlantic Oscillation (NAO) is a well-known source of atmospheric noise, but it is not the only atmospheric noise the ocean feels. Ocean noise includes the variations in the mixed layer, which are largely induced from wind and buoyancy forcing from above, along with high-frequency fluctuations in heat transport convergences into the mixed layer that may be forced by the wind stress as well as by internal ocean fluctuations, such as eddies. It will turn out that most of the white noise forcing appears to come from the atmosphere. The important finding is that the forcing may all be noise. No systematic long-period influence from the ocean circulation is needed.

Our hypothesis is motivated by findings in Clement et al. (2015, hereafter C15). C15 showed that CMIP3 atmospheric GCMs coupled to slab oceans produce the same spatial patterns of the AMO SST as fully coupled models and that both do a reasonable job of simulating the observed pattern. The slab ocean configuration does not have an active ocean circulation; the only "ocean circulation" is the unvarying climatological q flux that maintains the SST climatology. Beyond that, the slab models have a specified mixed layer depth and do not include the physics to vary the mixed layer depth or allow mixing from below.

The temporal behavior of the AMO in models also motivates our noise-driven hypothesis. The spectra in

Corresponding author: Mark A. Cane, mcane@ldeo.columbia.edu

DOI: 10.1175/JCLI-D-16-0810.1

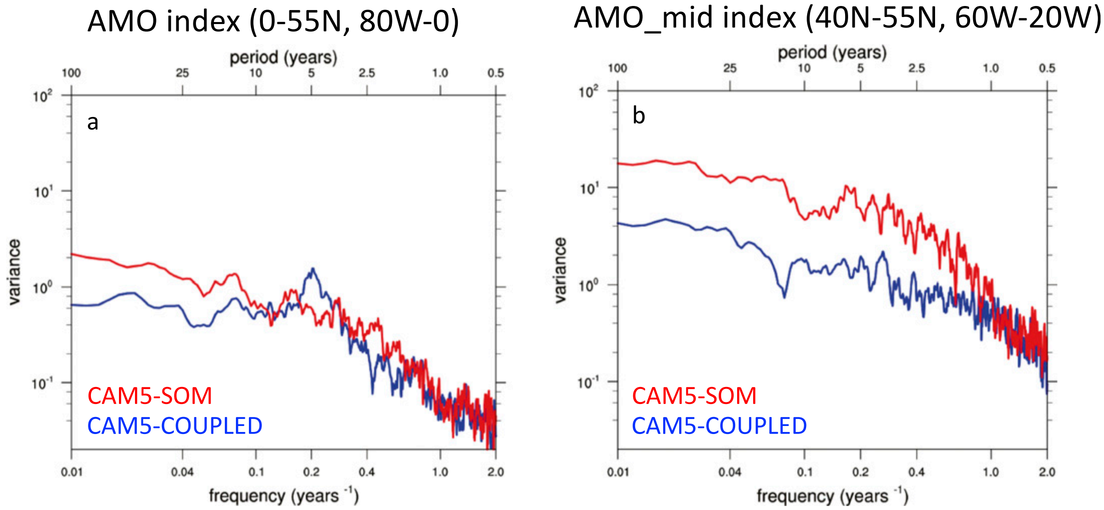

FIG. 1. Spectra for 500 yr from the fully coupled PI simulation of CESM1(CAM5) (blue) and 500 yr from CAM5-SOM (red) of (a) the AMO temperature index (all of the North Atlantic from 0° to 55°N) and (b) AMO_{mid}, the average SST over the region 40°–55°N, 20°–60°W.

Fig. 1 look like red noise and do not show any pronounced multidecadal peaks. The spectra in C15's Figs. 2 and 3 include the multimodel mean (MMM) and individual CMIP3 models, showing that the featureless spectra in our Fig. 1 are not an idiosyncrasy of the one model shown here, CESM1(CAM5). Additional examples of red noise drawn from CMIP5 PI runs appear in Fig. 2 of Ba et al. (2014) and Fig. 2 of Peings et al. (2016).[1] In addition, the spectra in C15 and in Fig. 1 show no structural difference between the AMO in the coupled models and that in the slab ocean model. Since the atmospheric forcing in these PI runs is more or less white in the atmosphere, and since adding an active ocean makes no essential difference, it suggests that whatever forcing comes from the ocean is also more or less white noise. On the basis of these results, C15 concluded that the ocean circulation is not an essential driver of the AMO.

The experimental protocol that leads to this conclusion is in the mold of the classic model-based attribution study: in one set of numerical experiments we include factor X, and in the other set we exclude it. Features that are the same in the two sets are not attributable to X. This sort of counterfactual test of causality goes back at least to David Hume in the 18th century: Y is caused by X if and only if X was not to occur, then Y would not occur (Hannart et al. 2016). In the present case Y = ''AMO-like variability'' and X = ''ocean circulation together with most of mixed layer physics.'' In the slab ocean model experiments, X does not occur, but Y does occur, so we conclude that ocean circulation (X) is not the cause of AMO-like variability (Y). Perhaps the real world is different, but this is what the models tell us in the absence of time varying external forcing.

Nonetheless, in reality and in the fully coupled models there surely is an ocean circulation and mixed layer physics that influence the SSTs in the North Atlantic. Thus, it is a given that the heat budgets determining SST cannot be exactly the same in the slab models as they are in the coupled models, but this need not mean that the AMO mechanism differs in an essential way from the suggestion in C15 that the AMO is driven by atmospheric white noise. In their model studies of decadal variability in the Atlantic, Fan and Schneider (2012) and Schneider and Fan (2012) conclude that the primary forcing is atmospheric noise, though they leave open a possible role for the ocean circulation. Others go further in their advocacy for the idea that the low-frequency ocean circulation, most prominently the Atlantic meridional overturning circulation (AMOC), is an essential player in multidecadal Atlantic SST variability, an idea with august beginnings in Bjerknes (1964) and Kushnir (1994), continuing to Zhang et al. (2016) and O'Reilly et al. (2016).[2] This does not go so far as to eliminate a role for the atmosphere, but it does anoint low-frequency heat convergence organized by the ocean

[1] In Peings et al. (2016), 2 of the 23 PI CMIP5 models do have significant multidecadal peaks. These models may have important ocean circulation features, though in the context of examining so many models and so many spectral bands without an a priori hypothesis, the significance of these peaks may be questioned.

[2] For a fuller set of references, see the review by Buckley and Marshall (2016).

circulation as the key driver of the AMO. Advocates often point to a negative relationship between surface heat flux and SST at long periods. They interpret this as showing the atmosphere responding to SST changes that must be driven by ocean heat flux convergence that is low frequency, which implies that this convergence is organized by low-frequency variations in ocean circulation. Recent examples include Gulev et al. (2013), Brown et al. (2016), Zhang et al. (2016), O'Reilly et al. (2016), and Drews and Greatbatch (2016). All these analyses are based on low-pass filtered data. We will show below that these low-passed results admit other interpretations, ones that do not require any nonrandom or dominant influence from the ocean circulation.

We are less certain that the observed AMO is red noise. There may be a real multidecadal peak, as was argued by Schlesinger and Ramankutty (1994), for example, but the record is too short to be certain. Even if real, a peak need not be a sign of internal variability since the instrumental period of observations has been marked by external forcing due to volcanic eruptions, solar variability, anthropogenic aerosol, and greenhouse gases. Models run with historical forcing have more power at low frequencies than preindustrial runs. This may be seen in C15 (cf. their Fig. 2c with their Figs. 2a,b) for multimodel means and Fig. 2 of Peings et al. 2016 for individual models. Our hypothesis, pursued further in Murphy et al. (2017) and Bellomo et al. (2017), is that the observations are best explained as largely a response to external forcing from aerosol and greenhouse gases. We will return to this issue in the discussion section, but throughout the rest of the paper we will concern ourselves with variability generated within the models' climate system in PI and slab ocean model (SOM) runs.

We begin in section 2 with the simple point model for sea surface temperature T that contains only a damping term and white noise forcing from both atmosphere and ocean. Such a model and its application to SST have a more or less unbroken 40-yr lineage going back to Hasselmann (1976) and Frankignoul and Hasselmann (1977). Among the long line of papers that follow, we single out Frankignoul et al. (1998) for its derivation of the simple model in Eq. (1) from a complete heat equation and for its derivation of some of the covariances in Eqs. (A2) and (A3). In this simple model, the damping term arises from the tendency of surface fluxes, especially the turbulent latent and sensible heat fluxes, to adjust to remove departures from the equilibrium value of T at zero total heat flux. It is generally accepted that the atmospheric forcing is white in time or nearly so (Wunsch1999; Stephenson et al. 2000; and see Fig. 2). The oceanic heat flux also appears to be nearly

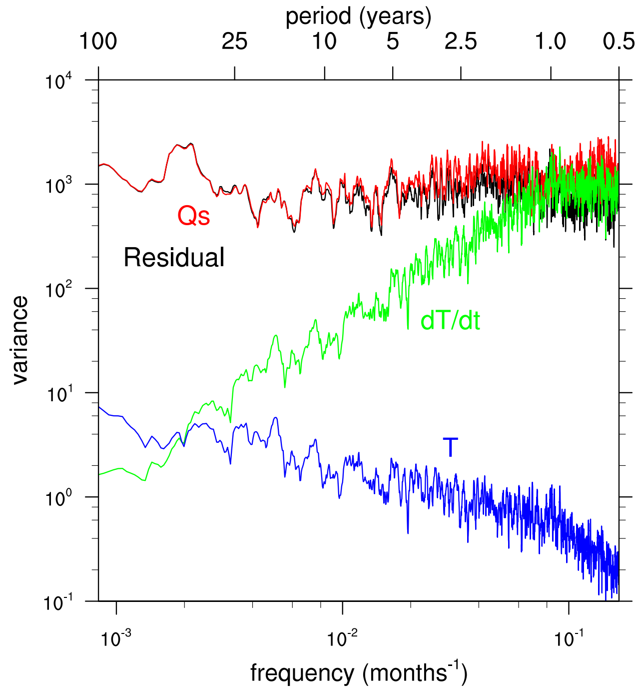

FIG. 2. (a) Spectra from a PI run of the CESM1(CAM5) coupled mode for quantities averaged over the $\mathrm{AMO_{mid}}$ region. The $\mathrm{AMO_{mid}}$ index is the blue curve. Also shown are temperature tendency dT/dt (green), surface heat flux Q_s (red), and ocean heat flux (black). As in many other presentations of ocean heat flux, it is actually the residual $\rho C_p h dT/dt - Q_s$, the difference between the rate of change of SST and the surface heat exchange with the atmosphere. In addition to heat convergence associated with the ocean circulation, it includes the heating due to mixed layer processes and computational error.

white in time, though this is less widely appreciated. Figure 2 shows the spectrum of oceanic heat flux derived from a PI run of the fully coupled model CESM1(CAM5) along with spectra of the surface flux Q_s, the surface temperature T, and the temperature tendency dT/dt, all for the $\mathrm{AMO_{mid}}$ region (20°–60°W, 40°–55°N).[3] As in many other calculations of oceanic heat flux (e.g., Zhang et al. 2016), it is actually the residual $\rho C_p h dT/dt - Q_s$, the difference between the rate of change of SST and the surface heat exchange with the atmosphere (h is mixed layer depth taken as 50 m). Figure 2 shows it to be approximately white. (The figure also shows that at low frequencies the ocean and atmospheric heat fluxes have the same power. At low frequencies, the temperature tendency is very small. Consequently, these fluxes must sum to near zero and so

[3] O'Reilly et al. (2016) used a region with the same limits, except that theirs goes to 60°N; we stop at 55°N to reduce the confounding influence of sea ice in winter. Gulev et al. (2013) used the average over the region east–west across the Atlantic from 35° to 50°N.

are equal and opposite. The implications of this will be considered below.). Recent observational results show that the ocean has power at all observed frequencies (Lozier 2012), though these series are not long enough to say whether or not the spectrum is truly white out to multidecadal periods.

We put the simple model forward as a way of understanding the complex model simulations. In section 2, after defining the model, we spend some time on the properties of low-pass filters, especially when applied to white or red noise. Analytic results for the unfiltered and low-pass (LP) filtered correlations that appear in the figures are derived in the appendix. These results are summarized and discussed at the end of section 2. In section 3, we first see if this simple model can account for properties of the coupled and slab models, particularly correlations involving SST and heat fluxes. We then consider results in the literature that are held up as evidence that the ocean driving is important for the AMO in order to see if our analysis of the simple red noise T model can account for them. We close this section with a consideration of low-frequency forcing that is distinct from white noise. A discussion section that includes some consideration of the twentieth-century observational record follows in section 4.

We use monthly CMIP5 data (found at http://cmip-pcmdi.llnl.gov/cmip5/data_portal.html). The CESM1(CAM5) monthly averages are available online (http://www.cesm.ucar.edu/projects/community-projects/LENS/data-sets.html). The models used and the length of each simulation are reported in Table 1. The surface flux Q_s is smoothed with a 1–2–1 filter and derivatives are approximated by centered second-order differences. For most purposes, we focus on the CMIP5 MMM. We single out the CESM1(CAM5) coupled model because very long simulations are available and, most relevant here, it is the only CMIP5 model for which we have access to a long SOM simulation. The CAM5-SOM simulation takes the mixed layer depths to be the mean seasonally varying depths in the corresponding coupled model, CESM1(CAM5). CESM1(CAM5) is generally similar to the CMIP5 MMM, but it does have a lower variance than some of the other models (Table 1). An account of this model's performance in the North Atlantic is given by Danabasoglu et al. (2012).

2. Noise-forced model

a. The model

We will work with the simple red noise model:

$$\frac{dT}{dt} = -\alpha T + q_T, \qquad (1)$$

where T is temperature, α^{-1} is the damping time, and $q_T = q_a + q_o$ is the total noise forcing, with q_a being the atmospheric and q_o the oceanic contribution. Here and throughout, we adopt the convention that heat fluxes into the ocean mixed layer (downward in the atmosphere) are positive. The net surface heat flux is

$$Q_s = -\alpha T + q_a. \qquad (2)$$

Note that in such a simple model the only distinction between q_a and q_o is that the former is included in Q_s and the latter is not. Usually we mean Q_s to be the total surface heat exchange, but in some literature Q_s is just the turbulent (latent and sensible) components, excluding the radiative contributions. In this interpretation, $-\alpha T$ is the feedback from changes in surface temperature and q_a accounts for all other influences on turbulent fluxes, such as changing wind speed or relative humidity or cloud cover. In our version, q_a also includes radiative changes due to varying cloud cover, aerosol influences, and solar variability. We assume that q_a and q_o are uncorrelated white noise of amplitudes a and b. Denoting expected value of the covariance of x, y by $\mathbb{E}\{x, y\}$,

$$\mathbb{E}\{q_a(t_1)q_a(t_2)\} = a^2\delta(t_1 - t_2), \mathbb{E}\{q_o(t_1)q_o(t_2)\} = b^2\delta(t_1 - t_2), \mathbb{E}\{q_a(t_1)q_o(t_2)\} = 0, \qquad (3)$$

where δ is the Dirac delta function and t_1, t_2 are arbitrary times. Consequently

$$\mathbb{E}\{q_T(t_1)q_T(t_2)\} = (a^2 + b^2)\delta(t_1 - t_2).$$

The assumption in Eq. (3) that q_a and q_o are uncorrelated makes the model as simple as possible and yet distinguishes atmosphere from ocean. Since the atmosphere term q_a is composed of radiative effects and influences on turbulent fluxes, such as changing wind speed or relative humidity or cloud cover, it is hard to see a direct connection with ocean circulation that is not mediated by surface temperature. In this model, anything that works through the surface temperature is accounted for by the temperature feedback term. There are also wind effects that influence both heat exchange at the surface and the ocean circulation, but the connection involves time lags (e.g., Czaja and Marshall 2000), and we leave this for future work. For the present, we are trying to see how much of the unforced model behavior can be captured by the simple model [Eqs. (1)–(3)], equations that deliberately do not include any time delayed mechanism that might be due to ocean circulation.[4]

[4] In the appendix, we briefly explore the case where $\mathbb{E}\{q_a(t_1)q_o(t_2)\} \neq 0$ if $t_1 = t_2$.

TABLE 1. Coupled models included here in the CMIP5 MMM. The second column gives the length of the available PI simulation for each model. The next column is the number of (nonindependent) 140-yr-long samples used in correlation statistics here and in Fig. 8. (For MIROC5, we delete the first 85 yr when the model is still spinning up.) All statistics are for the $\mathrm{AMO_{mid}}$ index, the average SST over the region 20°–60°W, 40°–55°N. Var T' is the variance of $\mathrm{AMO_{mid}}$. Correlations ρ are the low-pass correlations between temperature and the surface heat flux $\rho_{\mathrm{LP}}(T, Q_s)$. The "all" correlation is over the length of the simulation with the first and last 10 yr omitted to preserve the order of the filter. Mean, median, max, and min are taken over the set of 140-yr-long samples for each model.

Model name	Length (yr)	No. of samples	Var T'	ρ all	ρ mean	ρ median	ρ max	ρ min	ρ^2 all	ρ^2 mean	ρ^2 median
ACCESS1.0	500	340	0.29	−0.35	−0.38	−0.35	−0.21	−0.59	0.13	0.14	0.12
ACCESS1.3	500	340	0.23	−0.01	0.02	0.01	0.59	−0.66	0.00	0.00	0.00
CanESM2	996	836	0.23	−0.23	−0.23	−0.26	0.63	−0.72	0.05	0.05	0.07
CCSM4	501	341	0.17	−0.16	−0.22	−0.15	0.45	−0.64	0.03	0.05	0.02
CESM1(CAM5)	1801	1641	0.16	−0.28	−0.28	−0.31	0.31	−0.78	0.08	0.08	0.10
CMCC-CM	330	170	0.29	−0.59	−0.64	−0.64	−0.57	−0.74	0.35	0.41	0.41
CMCC-CMS	500	340	0.29	−0.70	−0.64	−0.66	−0.50	−0.80	0.49	0.41	0.44
CMCC-CESM	277	117	0.38	−0.67	−0.67	−0.67	−0.56	−0.84	0.45	0.45	0.45
CNRM-CM5–2	359	199	0.49	−0.29	−0.25	−0.24	0.35	−0.59	0.09	0.06	0.06
CNRM-CM5	850	690	0.32	−0.41	−0.28	−0.35	0.45	−0.77	0.17	0.08	0.12
CSIRO Mk3.6.0	500	340	0.24	−0.28	−0.28	−0.25	0.23	−0.70	0.08	0.08	0.06
FGOALS-s2	500	340	0.32	−0.70	−0.73	−0.73	−0.34	−0.89	0.50	0.53	0.53
GFDL CM3	500	340	0.24	−0.36	−0.24	−0.34	0.26	−0.54	0.13	0.06	0.12
GFDL-ESM2G	500	340	0.31	−0.31	−0.41	−0.44	−0.16	−0.62	0.10	0.17	0.20
GFDL-ESM2M	500	340	0.14	−0.20	−0.11	−0.13	0.33	−0.62	0.04	0.01	0.02
GISS-E2-H	780	620	0.12	0.00	−0.10	−0.05	0.37	−0.53	0.00	0.01	0.00
GISS-E2-H-CC	251	91	0.17	−0.71	−0.68	−0.69	−0.50	−0.75	0.50	0.46	0.48
HadGEM2-CC	190	30	0.54	0.11	−0.09	−0.11	0.05	−0.14	−0.01	0.01	0.01
INM-CM4.0	500	340	0.18	−0.66	−0.52	−0.59	0.14	−0.79	0.44	0.27	0.35
IPSL-CM5A-LR	430	270	0.61	−0.76	−0.72	−0.71	−0.62	−0.85	0.58	0.52	0.51
IPSL-CM5A-MR	300	140	0.34	−0.71	−0.67	−0.70	−0.43	−0.77	0.51	0.45	0.49
IPSL-CM5B-LR	300	140	0.62	−0.27	−0.21	−0.20	0.00	−0.40	0.07	0.04	0.04
MIROC5	670	425	0.23	−0.68	−0.64	−0.66	−0.45	−0.76	0.46	0.41	0.43
MIROC-ESM	630	470	0.20	−0.53	−0.47	−0.52	0.03	−0.82	0.28	0.22	0.27
MPI-ESM-LR	1000	840	0.28	−0.67	−0.60	−0.63	−0.20	−0.81	0.44	0.36	0.39
MPI-ESM-MR	1000	840	0.27	−0.69	−0.66	−0.68	−0.38	−0.84	0.47	0.43	0.46
MPI-ESM-P	1156	996	0.27	−0.60	−0.58	−0.65	−0.07	−0.84	0.35	0.34	0.42
MRI-CGCM3	500	340	0.64	−0.19	−0.11	−0.12	0.17	−0.37	0.04	0.01	0.01
NorESM1-M	501	341	0.11	−0.14	−0.10	−0.10	0.18	−0.36	0.02	0.01	0.01
NorESM1-ME	252	92	0.13	−0.16	−0.29	−0.30	−0.07	−0.45	0.03	0.08	0.09
CMIP5 MMM											
Mean	586	423	0.29	−0.41	−0.39	−0.41	−0.02	−0.67	0.23	0.21	0.22
Median	500	340	0.27	−0.36	−0.33	−0.35	0.01	−0.73	0.13	0.11	0.12
Std dev	337	337	0.15	0.26	0.23	0.24	0.38	0.18	0.21	0.19	0.20
Min	190	30	0.11	−0.76	−0.73	−0.73	−0.62	−0.89	−0.01	0.00	0.00
Max	1801	1641	0.64	0.11	0.02	0.01	0.63	−0.14	0.58	0.53	0.53

For the present, our goal is to make a very simple model and note where it fails convincingly enough to demand the addition of other processes. That it lacks the verisimilitude one hopes for in a GCM is obvious. The question is whether such simple and transparent physics is adequate to account for much of the behavior of complex models.

Returning to Eq. (1), it is convenient to scale time by α^{-1} (the only time scale in this model) to obtain

$$\frac{dT}{dt} = -T + q_T = -T + q_a + q_o; \tag{4}$$

$$T(t) = T(t_0)e^{-(t-t_0)} + \int_{t_0}^{t} e^{-(t-t')} q_T(t')\,dt'. \tag{5}$$

We also scale the heat fluxes so that $a^2 + b^2 = 1$, which means that the division between ocean and atmosphere is determined by a single parameter (a or b), and, regardless of the division, the expected value of the total noise forcing is fixed: $\mathbb{E}\{q_T^2\} = 1$. Thus, $a^2(b^2)$ is the percentage of the total noise variance attributable to the atmosphere (ocean).

We will want to compare various correlations involving temperature and heat fluxes in the noise forced model [Eq. (1)] with results from GCMs. Rather than just find these by numerical simulations of the model equation, we derive analytic results for expected values for correlations of unfiltered variables in section a of the appendix. In keeping with the fact that the forcing introduces no special time scales, all of these correlations are substantially different from zero only for leads and lags within a few e-folding time scales of $1/\alpha$ (i.e., less than a year). Some are delta functions at zero lag but smoothed over the minimal time resolution of the data. We will see in the figures to follow that the unfiltered correlations from the CMIP5 MMM and from CESM1(CAM5) also have appreciable correlations only at small lags and that some of the correlations resemble delta functions in being very localized at zero lag.

b. Low-pass linear filtering

Literature on the AMO almost always involves some LP linear filter on the data used to calculate correlations and examine the lead–lag relationships among multidecadal variables, such as temperature and heat fluxes. Hence, we will examine the impact of LP filtering on the conclusions drawn in this literature. We begin by calculating the correlations that result from low-pass filtering the variables in the red noise model [Eq. (1)]. A linear filter L applied to a time series $f(t)$ may be written

$$L[f(t)] \equiv \int_{-\infty}^{+\infty} L(\tau)f(t-\tau)\,d\tau, \qquad (6)$$

where, with a slight abuse of notation, we use L to denote both the operator $L[f]$ on the time series and the function $L(\tau)$ that specifies the filter weights. The general idea of a low-pass filter is to divide the signal spectrally into a low-frequency passband and a higher-frequency stopband. ''Low'' and ''high'' frequencies are separated by a filter parameter $\tilde{\omega}_c$ with dimensions of frequency. For a Butterworth filter, $\tilde{\omega}_c$ is taken as the half power point, while for an n-yr running mean $\tilde{\omega}_c$ is $(n\text{ years})^{-1}$. In our context, ''low pass'' means that the cutoff period is long compared to the damping time. We formalize this condition as $\tilde{\omega}_c^2 \ll \alpha^2$; nondimensionally $\omega_c^2 \ll 1$, where $\omega_c \equiv \tilde{\omega}_c/\alpha$.

Write $\hat{L}$ for the Fourier transform of the low-pass filter L. Then $\hat{R} = \hat{L}\hat{L}^* = |\hat{L}|^2$ is the power spectrum of the transfer function of $L(t)$. Using the notation

$$\mathbb{E}\{f(t+t')g(t')\} \equiv \int_{-\infty}^{+\infty} f(t+t')g(t')\,dt' \qquad (7)$$

for the expected value of the lagged covariance of f and g, we define notation for the low-pass covariance by

$$\mathbb{E}_{\mathrm{LP}}\{f,g\} \equiv \mathbb{E}\{L[f(t+\tau)]L[g(t)]\} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{L}[\hat{f}(\omega)]\{\hat{L}[\hat{g}(\omega)]\}^* e^{-i\omega\tau}\,d\omega = \frac{1}{2\pi}\int_{-\infty}^{+\infty} [\hat{L}(\omega)\hat{L}(\omega)^*][\hat{f}(\omega)\hat{g}(\omega)^*]e^{-i\omega\tau}\,d\omega;$$

$$\mathbb{E}_{\mathrm{LP}}\{f,g\} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{R}(\omega)[\hat{f}(\omega)\hat{g}(\omega)^*]e^{-i\omega\tau}\,d\omega = \int_{-\infty}^{+\infty} R(t-\tau)\mathbb{E}\{f(t+t')g(t')\}\,dt. \qquad (8)$$

We made use of the fact that $\hat{R}$ is real so $\hat{R} = \hat{R}^*$ to obtain the convolution in the last integral, which says that the low-pass covariance of functions f, g is the convolution of R with their unfiltered covariance [Eq. (7)]. The inverse transform of $\hat{R}(\omega)$, $R(\tau)$, is defined by

$$R(\tau) \equiv \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{R}(\omega)e^{-i\omega\tau}\,d\omega = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{L}(\omega)\hat{L}(\omega)^* e^{-i\omega\tau}\,d\omega = \int_{-\infty}^{+\infty} L(t+\tau)L(t)\,dt, \qquad (9)$$

which shows that R is the autocovariance function for the filter $L(t)$. It follows that R is real and an even function $[R(\tau) = R(-\tau)]$, and therefore so is $\hat{R}(\omega)$.

To calculate the low-pass covariances, we will make use of the fact that in LP filters $L(\tau)$ the time argument τ appears only in the combination $s = \omega_c\tau$. For example, for an n-month running mean filter $\omega_c = 1/(n\text{ months})$ and

$$L = \omega_c H(1 - 2|\omega_c\tau|) = \omega_c H(1 - 2|s|), \qquad (10)$$

where H is the Heaviside function [$H(x) = 1$ if $x > 0$, $H(x) = 0$ if $x < 0$]. In common with other filters, one changes the time of the cutoff period by changing the parameter ω_c without changing the functional form. Therefore, we may write $R = R(s)$ with $s = \omega_c\tau$. Hence $dR(s)/d\tau = \omega_c R_s(s)$. One advantage to using low-pass time s as the argument instead of τ is that R_s and R_{ss} are $O(R)$ while R_τ and $R_{\tau\tau}$ are $O(\omega_c R)$ and

$O(\omega_c^2 R)$, respectively, so using s makes the size of terms more transparent. The transform function corresponding to $R(s)$ is $\hat{R}(\omega/\omega_c)$. The combinations $\omega_c\tau$ and ω/ω_c are both dimensionless and have the same values whether the variables are taken as dimensional or dimensionless.

The low-pass correlations we need follow from the low-pass covariances derived in section b of the appendix. To the leading two orders in ω_c they are

$$\rho_{\mathrm{LP}}(T,T) = R(s)/R(0), \tag{11a}$$

$$\rho_{\mathrm{LP}}(T_t,T) = R_s(s)/[-R_{ss}(0)R(0)]^{1/2}, \tag{11b}$$

$$\begin{aligned}\rho_{\mathrm{LP}}(Q_s,T) &= -[b^2R(s) - \omega_c a^2 R_s(s)]/[R(0)Q_0]^{1/2}\\ &\approx -bR(s)/R(0),\end{aligned} \tag{11c}$$

$$\begin{aligned}\rho_{\mathrm{LP}}(T_t,Q_s) &= -[b^2R_s(s) + \omega_c a^2 R_{ss}(s)]/[-R_{ss}(0)Q_0]^{1/2}\\ &\approx -bR_s(s)/[-R_{ss}(0)R(0)]^{1/2}, \quad \text{and}\end{aligned} \tag{11d}$$

$$\rho_{\mathrm{LP}}(q_o,T) = b[R(s) + \omega_c R_s(s)]/R(0) \approx bR(s)/R(0), \tag{11e}$$

where $Q_0 = \sigma^2(Q_s) = b^2R(0) - \omega_c^2(a^2 - b^2)R_{ss}(0)$. The approximations shown in Eqs. (11c) and (11d) hold if $b^2 \gg \omega_c a^2$. For $b = 0$, the leading-order terms in Eqs. (11c) and (11d) and Q_0 drop out, and

$$\begin{aligned}\rho_{\mathrm{LP}}(Q_s,T) &= R_s(s)/[-R(0)R_{ss}(0)]^{1/2},\\ \rho_{\mathrm{LP}}(T_t,Q_s) &= -R_{ss}(s)/|R_{ss}(0)|,\end{aligned} \tag{12}$$

so that the structure of the correlations involving Q_s is drastically different with no ocean forcing, the noise-forced model (NFM) analog of a slab ocean model. Thus there is a striking difference in these correlations between the case with some ocean forcing [that is, with b merely small (i.e. $b^2 \ll a^2$ but $b^2 > \omega_c a^2$), and almost no ocean forcing [b very close to zero ($b^2 \ll \omega_c a^2$)]. A notable feature is that the zero lag correlation between temperature and the total atmospheric forcing (i.e., the surface heat flux Q_s) is zero if and only if there is no forcing from the ocean. Beyond that, Eq. (11c) shows that the strength of this correlation between SST and the atmospheric heat flux measures the amplitude $b = |q_o|$ of the ocean forcing. At the same time, this correlation tells us nothing about the impact of the ocean on SST; Eqs. (11a) and (11b) show that the presence or absence of ocean forcing makes no difference to the autocorrelation structure of the SST itself. The figures below will illustrate these general points. Section c of the appendix extends these results with a consideration of how the correlations change if the atmospheric heat flux q_a covaries with the ocean heat flux q_o.

c. Summary of analytic results for the NFM

The model presented in this section is quite simple: a one-dimensional temperature [Eq. (1)] forced by white noise from the atmosphere q_a and ocean q_o and damped by a linear feedback with time scale α^{-1}. Consequently, the temperature T has a red spectrum $\sim(\alpha^2 + \omega^2)^{-1}$. The only distinguished time scale in such a model is the damping time α^{-1}. Observational estimates for α^{-1} are less than one year, so it is well separated from multidecadal time scales. We find that unfiltered lag covariances and correlations die off exponentially with decay time α^{-1}. There are no significant relationships at long time scales, though such relationships can be created by low-pass filtering. [The appendix of Trenary and DelSole (2016) has a nice example of how LP filtering can create what is not in the data; in their example, it falsely exaggerates predictability times. Foukal and Lozier (2016) show how LP filtering falsely creates a propagating anomaly in SST observations.]

A low-pass filter divides the signal spectrally into a low-frequency passband and a higher-frequency stopband. ‘‘Low’’ and ‘‘high’’ frequencies are separated by a filter parameter $\tilde{\omega}_c$ with dimensions of frequency. The nondimensional cutoff is $\omega_c = \tilde{\omega}_c/\alpha$ and $\tau = \alpha t$ is nondimensional time. An important variable in our analysis is low-pass time $s = \tilde{\omega}_c t = \omega_c\tau$, which is the same in terms of dimensional or nondimensional variables. The key equation of our low-pass analysis is Eq. (8), which says that the low-pass covariance of any two variables u, v is the convolution of the autocovariance of the filter with the unfiltered covariance of u, v. Using this equation we are able to derive generic formulas for low-pass covariances [Eq. (A11)] and correlations [Eq. (11)] that are independent of the specific type of filter. The most striking thing about these low-pass formulas is their strong dependence on the filter autocovariance $R(s) = R(\omega_c\tau)$ and its derivatives. The structure of the unfiltered solution is almost entirely lost. This should be expected. The filtering is being applied to either white noise (as in the forcing) or red noise (as for T). The former just returns the filter autocovariance since the Fourier transform of white noise = 1 [viz., Eq. (8)]. If red noise is involved, the fact that the filter is low pass means that $\alpha^2 \gg \tilde{\omega}_c^2$ $(1 \gg \omega_c)$ so that, for frequencies $\tilde{\omega}$ below the cutoff $\tilde{\omega}_c$, we have, for a red noise spectrum f with amplitude A,

$$|\hat{f}|^2 = \frac{|A|^2}{\alpha^2 + \tilde{\omega}^2} \approx \frac{|A|^2}{\alpha^2}, \tag{13}$$

which means that the part of the spectrum that is passed by the low-pass filter is nearly white. Hence, in all cases the covariance that appears in Eq. (8) convolved with R

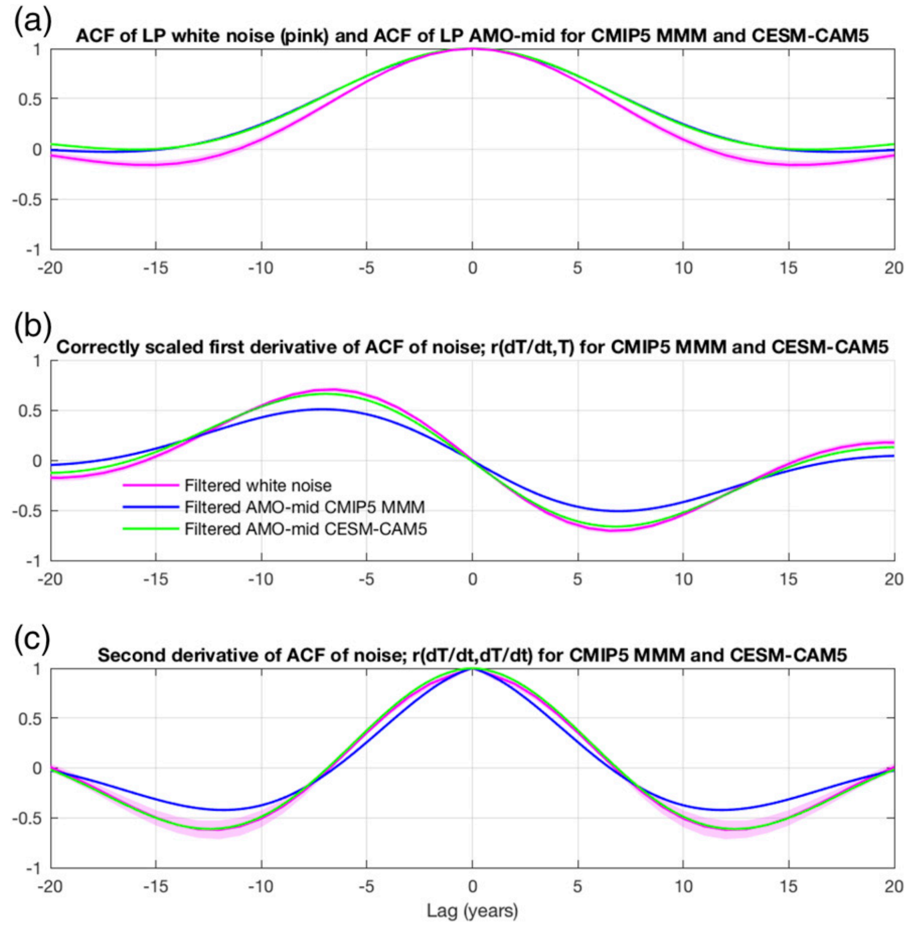

FIG. 3. (a) The autocorrelation function $R(t)/R(0)$ (magenta) for the fourth-order 20-yr Butterworth filter. Also shown is the autocorrelation of temperature filtered by the same Butterworth filter in the AMO_{mid} region in the coupled model CESM1(CAM5) (green) and in the CMIP5 MMM (blue). (b) The $dR(\tau)/d\tau/[-d^2R(0)/d\tau^2R(0)]^{1/2}$ [see Eq.(11b)] and the lead–lag correlations $\rho(dT/dt,\ T)$ in the two coupled models. (c) The $d^2R(\tau)/d\tau^2/d^2R(0)/d\tau^2$ and the lead–lag correlations $\rho(dT/dt,\ (dT/dt)$ in the two coupled models. The autocorrelation function R is calculated by applying the same filter software used throughout to a white noise sequence. The agreement among the curves is predicted by Eqs. (11a), (11b), and (A11f) if the temperature in the coupled models is red noise.

is essentially white noise, and the filtering returns the autocovariance of the filter or its derivatives. In particular, the lagged correlation of T with itself $\sim R(s)$, and that of dT/dt and $T \sim R_s(s)$, illustrate that the LP filter has created long lead correlations not found in the unfiltered model data. The more complicated correlations involving heat fluxes will be discussed in context in the next section.

Figure 3 displays the autocovariance $R(s = \omega_c\tau)$ and its first two derivatives for a fourth-order 20-yr Butterworth filter. These are calculated by filtering a monthly white noise series, then finding the lagged covariance and taking its derivatives numerically in the usual way. We have scaled R, R_s, R_{ss} as in Eqs. (11a), (11b), and (A11f) to make the three panels comparable to low-pass lead–lag correlations $(T,\ T)$, $(dT/dt,\ T)$, and $(dT/dt,\ dT/dt)$. The following features are evident in Fig. 3. Since R is an even function, $R(0)$ is an extremum and $R_s(0) = 0$. Since we expect a maximum correlation at $t = 0$, the extremum is a maximum so that $R_{ss}(0) < 0$.[5] Since we are dealing with low-pass filters, we expect that $R(s)$ will be fairly smooth in a sizable neighborhood around $\tau = 0$. In more detail, since $R_s(s = 0) = 0$, we expect $R(s)$ to vary slowly for $s \lesssim 1$; in terms of dimensional variables, this neighborhood is $|\tau| \lesssim \tilde{\omega}_c^{-1}$. Hence, in this neighborhood we expect R, which is an even function, to be positive and relatively large, and R_s to be positive for $\tau < 0$ and negative for $\tau > 0$. These are

[5] We do not digress to discuss what happens if R or R_s is discontinuous.

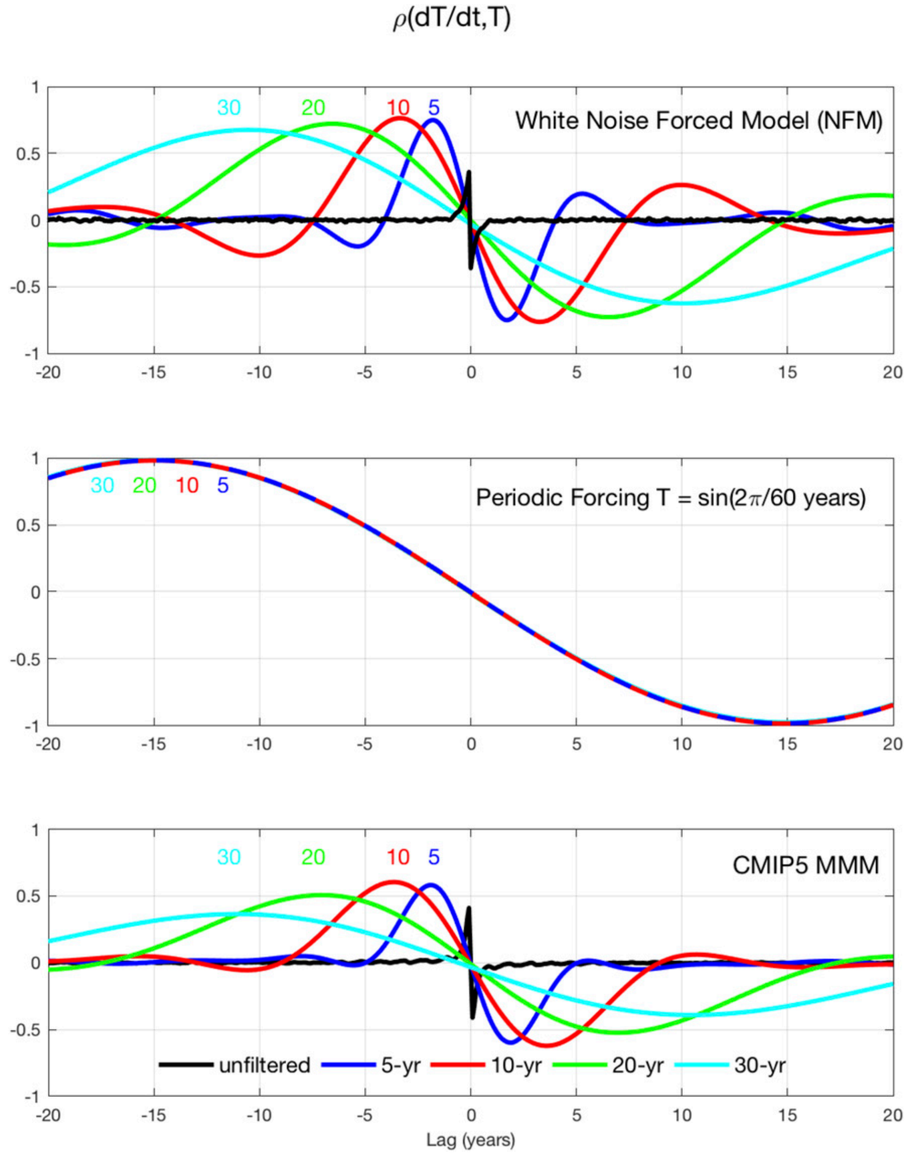

FIG. 4. (top) The correlation $\rho(dT/dt,\ T)$ for $T(t)$ a low-pass filtered time series of monthly white noise. The different curves are the fourth-order LP Butterworth filter with half-power periods of 5, 10, 20, and 30 yr. The movement of the peak is linear with the period, as expected for white noise. (middle) The same correlation filtered in the same ways but applied to a time series $T = \sin(2\pi t/60\,\mathrm{yr})$. In this case, the curves for different cutoff periods lie on top of one another, and, as expected, for a true low-frequency sinusoidal signal the peak holds fast at 15 yr, a quarter cycle. (bottom) Correlation of dT/dt and T for the $\mathrm{AMO_{mid}}$ index in the CMIP5 MMM.

generic expectations for all low-pass filters and suggest that these low-pass correlations will not be very sensitive to the form of the filter (Hamming, Lanczos, etc.). They will, however, be quite sensitive to the cutoff $\tilde{\omega}_c$ since the lag τ where maxima, minima, and zeros occur depends on $s = \tilde{\omega}_c\tau$ and hence inversely on the cutoff frequency $\tilde{\omega}_c$ or linearly on the cutoff period.

The effect of changing the filter cutoff is illustrated in Fig. 4. The top panel shows the correlation of dT/dt and T when $T(t)$ is red noise [e.g., if T is generated from Eq. (1)]. The different curves are for filter cutoff periods of 5, 10, 20, and 30 yr. The lead or lag where the extrema occur move linearly with the cutoff period, as expected for red or white noise. The middle panel of Fig. 4 shows the correlation for a low-frequency signal $T = \sin(2\pi t/60\,\mathrm{yr})$ with the same set of 5-, 10-, 20-, and 30-yr filters. With a true low-frequency signal—any low-frequency signal, not just the simple sinusoid—the different filters leave the signal unchanged so the different curves all lie on top of one another. In particular, the well-known quadrature between dT/dt and T, which gives a peak when the tendency leads by a quarter

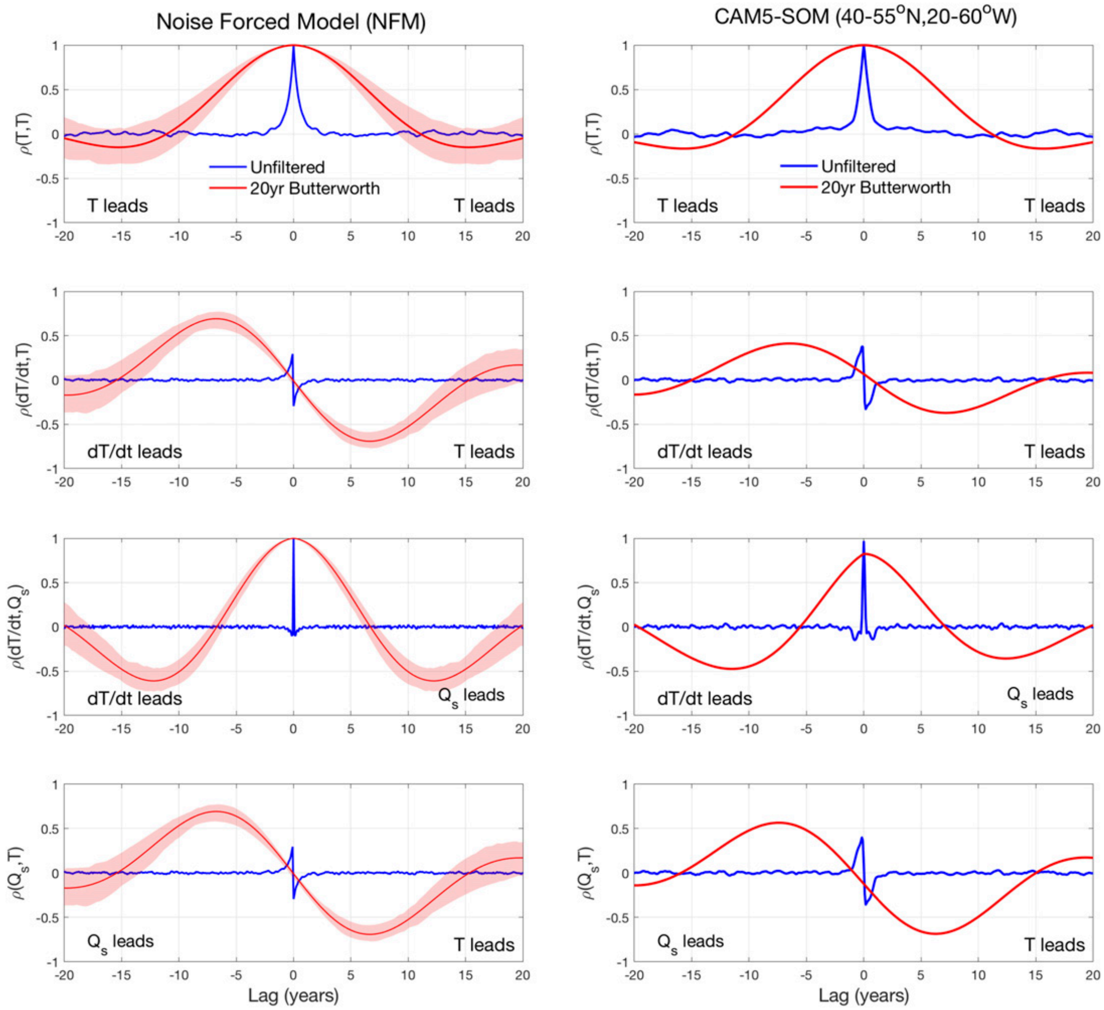

FIG. 5. Lead–lag correlations for (top)–(bottom) $\rho(T, T)$, $\rho(T_t, T)$, $\rho(T_t, Q_s)$, and $\rho(Q_s, T)$. (left) The simple NFM with parameters $a = 1$, $b = 0$, $\alpha = 13\,\mathrm{W\,m^{-2}\,K^{-1}} = (0.5\,\mathrm{yr})^{-1}$ for a 50-m mixed layer. (right) The CAM5-SOM for the $\mathrm{AMO_{mid}}$ region. In both models, all the forcing comes from the atmosphere. Correlations for unfiltered (blue) and low-pass filtered (20-yr fourth-order Butterworth; red) data are shown. The shading indicates the ±5% confidence limits based on 1000 simple model runs of 140 yr each.

period (15 yr in this example) is a robust feature and does not change when the filter length changes. In contrast, if the peak does move when the filter length changes, as in the top panel of Fig. 4, then this is indicative of white or red noise, and the peak is an artifact of the filter. Such movement may be used as a diagnostic to distinguish noise from a true low-frequency signal.

3. Comparison of the white noise–forced model with GCM and observational results

a. *Is the white noise–forced model relevant for GCMs?*

Our first task is to establish that the white noise–forced model (NFM hereafter) presented in the previous section is relevant for interpreting the GCM results. We focus on lead–lag correlations involving temperature and heat fluxes, which appear often in the literature. The right side of Fig. 5 shows the CAM5-SOM lead–lag correlations for the unfiltered results in blue and the low-pass results in red. For comparison, the left side of Fig. 5 shows the correlations from the NFM when $a = 1$ so that all the forcing is from heat exchanges with the atmosphere, as for SOM. The SOM figures are based on an index $\mathrm{AMO_{mid}}$, the average SST over the region 20°–60°W, 40°–55°N. The NFM figures are computed from data taken from a long run of the simple model [Eq. (1)]. All the filtered results were obtained using a fourth-order 20-yr low-pass Butterworth filter. Other choices for an LP filter would make no important qualitative changes, though, as noted above, changing the frequency cutoff does create significant differences. We estimated a value of the damping time α^{-1} from the

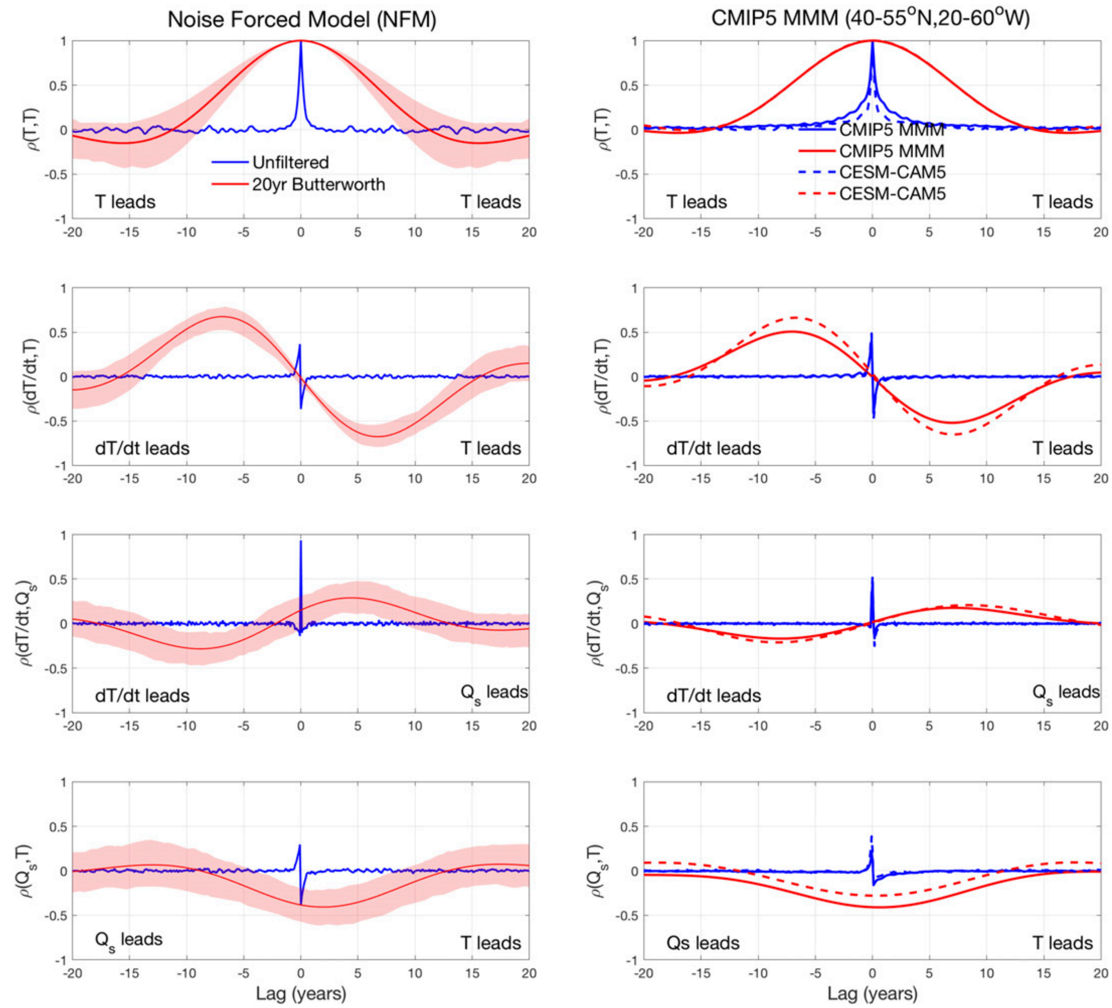

FIG. 6. As in Fig. 5, but for (left) the simple model with parameters $a^2 = 0.85$, $b^2 = 0.15$, $\alpha = 20\ \mathrm{W\,m^{-2}\,K^{-1}} = (4\ \mathrm{months})^{-1}$ for a 50-m mixed layer; and (right) the CESM1(CAM5) fully coupled model and the CMIP5 MMM.

unfiltered autocorrelation of temperature in SOM (Fig. 5, top). The value obtained is 0.5 yr, corresponding to a heat flux per degree of $13\ \mathrm{W\,m^{-2}\,K^{-1}}$ for a 50-m deep mixed layer, which is close to but smaller than the values estimated by Park et al. (2005) and Frankignoul and Kestenare (2002) from observed data using the methods devised by Frankignoul et al. (1998).

The similarities between the corresponding curves in the SOM and the NFM with $a = 1$ are compelling, as also found by O'Reilly et al. (2016). Starting at the top, the unfiltered (blue) curves for $\rho(T,\ T)$ are identical as far as the eye can tell. Both have the form predicted by Eq. (A7a). The implication is that, for SOM as for NFM, the temperature behaves like damped white noise: a red spectrum $\sim(\alpha^2 + \omega^2)^{-1}$. The LP filtered correlation (red curve), which is the same for SOM and NFM, is predicted by Eq. (11a) to be the autocorrelation function of the filter. We verify this in the top panel of Fig. 3, which shows the autocorrelation for the same Butterworth filter used for Fig. 5. The correlations $\rho(T_t,\ T)$ in the second panels from the top are again quite similar. The SOM $\rho(T_t,\ T)$ is otherwise indistinguishable from the NFM. As expected from Eq. (11b), the LP filtered correlation (red curve) is just the scaled first derivative R_s of the filter autocovariance (viz., Fig. 3).

The right side of Fig. 6 is as the right side of Fig. 5, but for the coupled model CESM1(CAM5) and the CMIP5 MMM for the models in Table 1. We estimate α to be $20\ \mathrm{W\,m^{-2}\,K^{-1}}$, which is comparable to observed values (Frankignoul et al. 1998; Park et al. 2005) and slightly higher than for the slab ocean. For a 50-m mixed layer this corresponds to a damping time of 0.325 years or slightly less than 4 months. The NFM case in Fig. 6 we show as comparable has $a^2 = 85\%$ of the forcing variance coming from the atmosphere and only 15% from the ocean. (Taking $a^2 = 95\%$ yields qualitatively similar results.) Note that Eqs. (A7a) and (11a) show that, for the NFM, the temperature correlations $\rho(T,\ T)$ and

$\rho(T_t, T)$ are independent of a. As with the SOM results in the top two panels of Fig. 5 and the NFM results in Fig. 5, we find that the fully coupled model results for temperature agree quite well with the white noise–forced model. Thus, as with the SOM, the temporal structure of temperature T in CESM and in the MMM are consistent with a damped response to white noise forcing. Both the unfiltered and LP filtered T (top panel of Fig. 6) have the same structure as the comparable SOM (Fig. 5) or the NFM (in either Fig. 5 or Fig. 6). As suggested by the analysis in section 2, given that the unfiltered T from the coupled model looks like red noise with a damping time short compared to the multidecadal time scales of interest for the AMO, the LP filtered correlation $\rho(T, T)$ (red, top panel of Fig. 6) is just the filter autocovariance R, while the LP $\rho(T_t, T)$ is the first derivative of R. The MMM looks to be the same. The red noise structure of the unfiltered data has been bleached by the filtering.

Zhang et al. (2016) interpret the LP lagged correlations between T and T_t as meaningful, but in view of the dependence of this on the filter [Fig. 3; Eq. (11b)], this is questionable. The bottom panel of Fig. 4 clearly shows that in the CMIP5 MMM the midlatitude temperature $\mathrm{AMO_{mid}}$ behaves like the white noise in the top panel. (This is also true of the CESM; not shown.) The same conclusion applies to the ensemble mean of 10 coupled models studied by Zhang et al. (2016). They state that the maximum of $\rho_{\mathrm{LP}}(d\mathrm{AMO}/dt, \mathrm{AMO})$ is at 4 yr for a 10-yr low-pass filter (LF),[6] while "with 20- or 30-year LF, the multi-model mean correlation between $d\mathrm{AMO}/dt$ and the AMO peaks at longer lead times (8 or 10 to 11 years, respectively) due to the broad AMO spectra in CGCM" (Zhang et al. 2016, p. 1527-a). Our theory predicts a linear increase with the cutoff period—the peak at 4, 8, and 12 yr for cutoffs of 10, 20, and 30 yr—if the AMO is approximately white in the passband.

With the correlation $\rho(T_t, Q_s)$ shown in the middle panels of Figs. 5 and 6, we come to a quantity that does depend on the percentage of the total forcing attributed to the atmosphere a^2. In both of the NFM cases and the coupled and SOM GCM runs, the unfiltered results show the delta function behavior at zero lag expected for the NFM from Eq. (A7d). The SOM and NFM with $a = 1$ (Fig. 5) show the symmetry about $\tau = 0$ for both the unfiltered and filtered curves expected from Eqs. (A7d) and (12) when there is no forcing from the ocean ($b = 0$). The unfiltered results for CESM and the CMIP5 MMM (Fig. 6) have the positive spike at $\tau = 0$ but differ from the $b = 0$ cases of Fig. 5 in having a negative dip for small positive τ. The NFM $a^2 = 0.85$ case (Fig. 6, left) has similar behavior, and, for both the coupled models and the NFM, the filtered correlation is somewhat asymmetric. This requires that there be some ocean forcing ($b \neq 0$) but also a high value of a so the symmetric term R_{ss} in Eq. (11d) is evident despite being at a lower order in ω_c than the antisymmetric term R_s. The NFM case with 85% of the forcing coming from the atmosphere and 15% from the ocean resembles the MMM and CESM LP correlations more than cases with stronger ocean forcing (not shown).

Thus, there is nothing so far to rule out the hypothesis that the coupled model SST response is primarily a response to white noise forcing from the atmosphere with additional random forcing from the ocean. The ocean need not do anything systematic and need not be the dominant forcing. However, there is a considerable literature offering arguments that the AMO is forced by ocean circulation. We next review some of that literature with the aid of the simple noise-forced model.

b. SST changes and surface heat flux

Zhang et al. (2016) studied the relation between T_t and Q_s in a different way by regressing both LP filtered terms on T, the AMO index. They then looked at these two regressed variables at a 4-yr lead, picked because it is where $\rho(T_t, T)$ is a maximum for the 10-yr LP filter they use. (As shown in Fig. 4, the peak time changes with the filter cutoff; it is an artifact of the LP filter and is not robust or physical.) At a 4-yr lead over T, they find that in the multimodel mean of a set of CMIP3 coupled models the LP tendency T_t and the LP surface heat flux Q_s have opposite sign. They interpret this to mean that the (negative) surface flux is not the cause of the (positive) temperature change. They contrast this with the SOM case, where T_t and Q_s at a 4-yr lead are both positive. One can see the same thing in Figs. 5 and 6 here. When dT/dt leads T by 4 yr in the SOM case and the coupled case (Figs. 5 and 6, second panel from the top; lag $= -4$ yr) it is positive, for both the GCM and the NFM. The bottom panels show that at the same lead Q_s is positive in the SOM case of Fig. 5 but negative in the coupled case of Fig. 6. The NFM is forced by white noise, and there is no long-period ocean circulation to account for this behavior. None is needed.

As noted by Clement et al. (2016), it is evident from Fig. 1 of Zhang et al. (2016) that in both the SOM and coupled simulations the temperature tendency T_t is small relative to the coupled model surface heat flux Q_s or ocean heat flux q_o (which is actually computed as the residual $\rho C_p h T_t - Q_s$ and so in addition to convergence of ocean heat transport comprises ocean mixed layer

[6] The type of filter used is not stated, but the analysis in section 2c is robust to filter type as long as it is low pass.

processes and computational error). Figure 1 in Zhang et al. (2016) shows that the atmospheric forcing and the ocean forcing are approximately in balance in the coupled case, while, in the SOM case, the only forcing, the one from the atmosphere, is approximately zero.

This must be so. Consider the dimensional Eq. (1), which is generic enough to apply to the GCMs as well as the simple model if we do not restrict the atmospheric forcing q_a and ocean forcing q_o to be white noise. It is still valid to write part of the surface flux forcing as a temperature-dependent negative feedback term $-\alpha T$ since this is true of the turbulent transfer terms (the latent and sensible heat fluxes). Since all variables are low-pass filtered, T_t is solely low frequency relative to the damping time; that is, $T_t \ll \alpha T$, so the lhs of Eq. (1) may be neglected: $0 \approx Q_s + q_o = (-\alpha T + q_a) + q_o$. Therefore,

$$Q_s \approx -q_o, \tag{14a}$$

$$T \approx (q_a + q_o)/\alpha, \tag{14b}$$

and it follows that

$$\mathbb{E}_{\mathrm{LP}}\{Q_s, T\} \approx \mathbb{E}_{\mathrm{LP}}\{-q_o, T\} = -\mathbb{E}_{\mathrm{LP}}\{q_o, T\} \tag{15a}$$

$$\approx -\mathbb{E}_{\mathrm{LP}}\{q_o, (q_a + q_o)/\alpha\} = -\alpha^{-1}\mathbb{E}_{\mathrm{LP}}\{q_o, q_o\} \le 0. \tag{15b}$$

In the special case that the ocean forcing q_o is white noise, then according to Eqs. (15b) and (A11d), $\mathbb{E}_{\mathrm{LP}}\{Q_s, T\} = -b^2 R(s)$, consistent with the leading-order term of Eq. (A11g). However, the relations in Eqs. (14) and (15) do not require white noise forcing. The derivation of Eq. (14) assumes only that variables are low frequency, which is guaranteed—imposed, in fact—by the low-pass filtering. Equation (15a) then follows from Eq. (14), though the final step in Eq. (15b) also assumes that q_a and q_o are uncorrelated. If they are correlated, Eq. (15a) still holds, so $\mathbb{E}_{\mathrm{LP}}\{Q_s, T\}$ will again be negative unless the temperature and ocean forcing are negatively correlated; that is, unless the temperature moves systematically opposite to the ocean heat flux so that, for example, when the ocean provides more heat the surface temperature goes down. This is not anyone's idea of what is meant by the ocean driving the temperature. Moreover, what evidence there is (viz., Brown et al. 2016; Bellomo et al. 2016) suggests that a convergence of ocean heat results in a positive radiative feedback (i.e., that q_a and q_o covary positively). The appendix (section c) contains a brief account of how our results are altered when the two heat fluxes are white noise but covary.

Equation (15) means that, as long as there is *any* ocean forcing ($q_o \neq 0$) to leading order, the LP relation between T and Q_s will be negative for a range of lags and leads around zero. It also must be that $\mathbb{E}_{\mathrm{LP}}\{T_t, T\} > 0$ for T_t leading T. There is nothing physical about this: as discussed in section 2c, it is a mathematical consequence of the fact that the autocorrelation of T—or *any* variable—is an even function of lag with a maximum at zero lag. So T_t and Q_s necessarily have opposite signs for T_t leading (e.g., by 4 years). If instead one had taken T_t lagging by 4 yr, one could be tempted to conclude that, since the surface heat flux has the same sign as the temperature tendency, it must be the atmosphere that is the driver for the surface temperature, not the ocean. The same holds for the multimodel mean presented in Zhang et al. (2016) if one looks at the regressions with T_t lagging T. This was shown in Fig. 1 of Clement et al. (2016) and is evident in Fig. 6 here, for both the coupled model and the NFM. The arbitrary choice of lead or lag dictates the interpretation.

Neither interpretation is justified. The low-frequency relations say there is an approximate balance between heating from the atmosphere and heating from the ocean, a near equilibrium. One cannot infer causality from the sign of the correlations; they tell what the balance is at equilibrium but not what was responsible for creating the equilibrium state. Causality is no more revealed by this balance than geostrophic balance tells whether the state was achieved by the pressure adjusting to the winds or the winds adjusting to the pressure. The NFM case in Fig. 6 is an illustration. The atmospheric forcing is 6 times the ocean forcing, and yet the LP correlation between temperature and surface heating is negative at and near zero lag. The correlation remains negative even when the forcing is overwhelmingly from the surface (e.g., 95% atmosphere and only 5% ocean).

The atmosphere-only (SOM) case is singular in that the surface heat flux Q_s by itself must be approximately zero. Figure 1d of Zhang et al. (2016) shows that this holds for the CMIP3-SOM multimodel mean. Equation (12) shows that, if there is no ocean flux ($b = 0$), then the correlations depend on lower-order terms. More directly, if there is no contribution from the ocean, then $T_t = Q_s$, so $\rho(Q_s, T) = \rho(T_t, T)$ [cf. Eqs. (11b) and (12)]. In particular, the correlations are zero at zero lag, negative when T leads, and positive when T lags, unlike the situation when the ocean flux is nonzero. Without question, the oceans do something in both the real coupled system and the coupled models. The issue is whether or not that something is essential for the AMO. A diagnosis of the quasi-equilibrium state does not speak to this issue.

In an important and influential paper, Gulev et al. (2013) examined the relation between SST and surface heat flux in observational data. Their intent was to examine the Bjerknes (1964) hypothesis that at short time

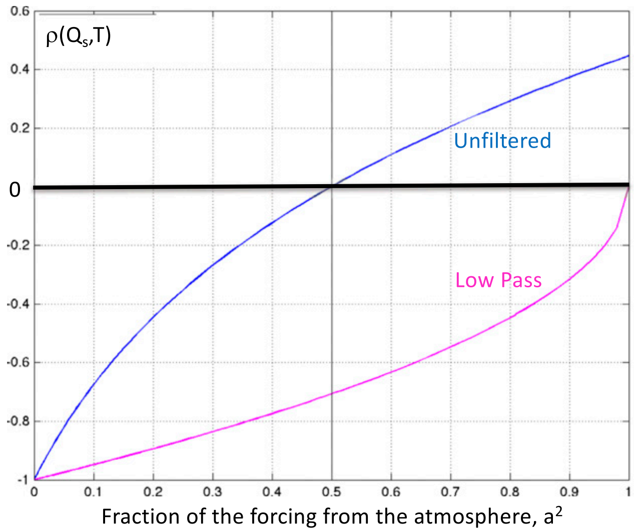

FIG. 7. The unfiltered [blue curve; Eq. (A7c)] and low-pass filtered [magenta curve; Eq.(11c)] correlation $\rho(Q_s, T)$ at zero lag for the simple noise-forced model as a function of a^2, the fraction of the forcing variance coming from the atmosphere. Note that the two curves have opposite sign for $a^2 > 0.5$ or, equivalently, $a^2 > b^2$ (i.e., when the atmospheric forcing is greater than the ocean forcing).

scales the atmosphere drives the ocean, whereas at multidecadal time scales the ocean is the driver. The ''surface heat flux'' data they use are an estimate of the turbulent fluxes only; they omit the radiative components of atmospheric fluxes, such as variations in cloud cover (Bellomo et al. 2016). After the data were low passed by applying an 11-yr running mean, they found that the correlation at zero lag between T and Q_s is such that warmer temperatures mean the ocean loses heat (a negative correlation using our convention that heat into the ocean is positive).[7] They interpret this to mean that the ocean is forcing the atmosphere at the low frequencies passed by the filter. In contrast, when the data are high-pass filtered, they find that the correlation is positive, which is interpreted to mean that the atmosphere is forcing the ocean. They conclude that this is evidence in support of the Bjerknes hypothesis.

We have already seen that the sign of the correlation of T and Q_s at low frequencies cannot tell us whether the forcing resides in the ocean or the atmosphere. In Fig. 7, we plot the low-pass correlation of T and Q_s in the white noise–forced model as a function of the atmospheric forcing variance a^2. The LP correlation is given by Eq. (11c); at zero lag, $\rho_{\mathrm{LP}}(T, Q_s) = -b$ independent of the form of the filter. Thus, $\rho_{\mathrm{LP}}(T, Q_s) < 0$ if the ocean forcing is nonzero. We conclude that the low-frequency result in Gulev et al. (2013) cannot be taken as evidence for the Bjerknes hypothesis. Nor is it evidence against it. As above, it is uninformative about the nature of the forcing.

We may make further use of the low-pass relation [Eq. (11c)] shown in Fig. 7 to estimate the relative magnitudes of the ocean and atmospheric forcing. O'Reilly et al. (2016) found that the low-pass correlation $\rho_{\mathrm{LP}}(T, Q_s)$ is $\approx$0.4 for the multimodel mean of CMIP5 models. From Eq. (11c), this means $b = \rho_{\mathrm{LP}}(T, Q_s) \approx 0.4$, so b^2, which is the fraction of the forcing variance from the ocean, $\approx$16%. Such a low value for the ocean forcing is consistent with the Clement et al. (2015) finding for the CMIP3 MMM that removing the ocean circulation (as in the slab model) had little effect on the solution for SST.

Figure 8a shows the range of values of the correlation $\rho_{\mathrm{LP}}(T, Q_s)$ for the 30 different CMIP5 coupled models in Table 1. For most (16 of 30) $b^2 = \rho_{\mathrm{LP}}^2 < 0.15$ [i.e., according to Eq. (11c), the ocean forcing is less than 15% of the total]. The CMIP5 MMM value is 16% (as in O'Reilly et al. 2016), and the median value is 13%. In only 2 of 30 models does this measure suggest that the ocean forcing is greater than half the total. In summary, the MMM and most of the individual models appear to be largely forced by the atmosphere, but there are some exceptions. It is intriguing that seemingly similar models (e.g., GISS-E2-H and GISS-E2-H-CC; IPSL-CM5A-LR and IPSL-CM5B-LR) fall at opposite ends of the distribution of correlation coefficients, but exploring this further is beyond the scope of the present study.[8]

Figure 8a shows distributions of values in box-and-whisker format. The distributions were created from all subsamples 140 years long, the approximate length of the observational record and one allowing at least a few tens of samples for all models. The spreads are quite large, but even accounting for sampling issues the correlations are almost always negative, consistent with Eq. (15). Again, while this further demonstrates consistency with the white noise forcing hypothesis, it does not rule out other possible explanations. Moreover, even if the NFM assumptions (including q_a and q_o being uncorrelated) are completely satisfied, this sampling variation creates considerable uncertainty in estimating the ocean forcing b^2 from a sample correlation. Figure 8b is in the same format as Fig. 8a, but for 500-yr runs of the NFM. The different box and whiskers are for different

[7] Gulev et al's (2013) sign convention is the opposite of ours in that a positive heat flux is out of the ocean. In what follows, we describe their results using our sign convention: heat flux is positive into the ocean.

[8] In other words, a look at the documentation for these differences did not bring enlightenment to us.

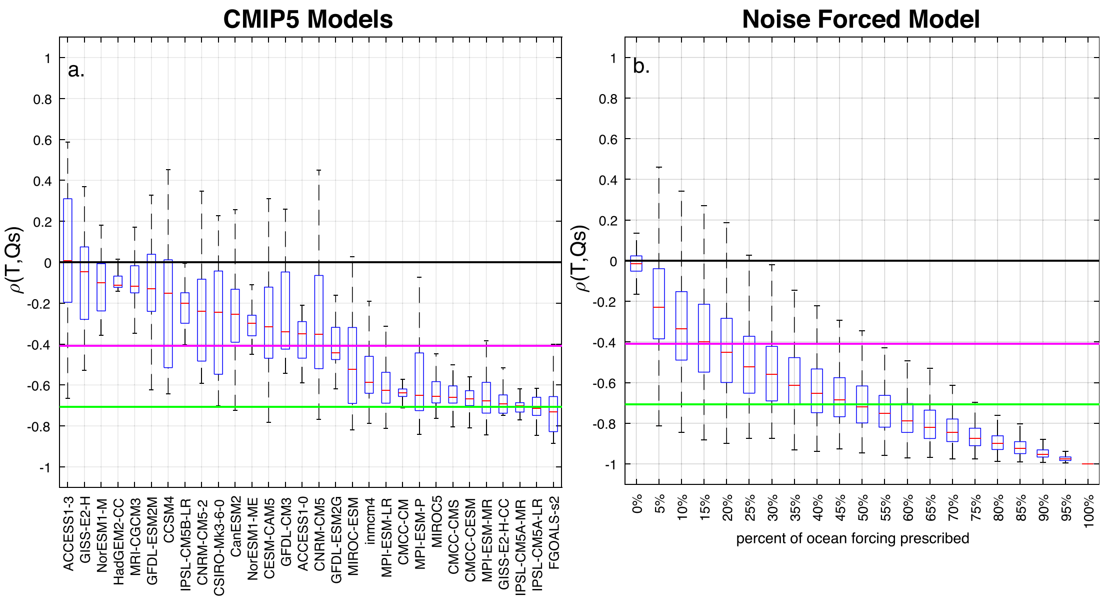

FIG. 8. (a) Low-pass correlation of surface heat flux Q_s and temperature for CMIP5 model PI simulations. The box and whiskers give the median (magenta lines in the boxes), upper and lower quartiles (boxes), and range for samples of 140 yr. (See Table 1 for the number of samples; the first and last 10 yr of each simulation are left off to accommodate the LP filter.) (b) Similar plots, but for the simple NFM. Samples are drawn from 500-yr simulations (the median length of the CMIP PI runs). Each box-and-whiskers plot is for a different value of the percent of forcing from the ocean. The horizontal red lines are at the CMIP MMM; the horizontal green lines mark the values where the correlation indicates that the ocean is one-half of the total forcing ($\rho_{\mathrm{LP}} = 1/\sqrt{2}$).

specified values of the ocean forcing b^2. The spreads are large, and an "observed" sample value of $\rho_{\mathrm{LP}} = 1/\sqrt{2}$—the green line—has a chance of being found when the ocean forcing is anywhere from 5% to 75%, though if we insist the chance be as high as 1 in 4, then the range narrows to 35%–60%, which is still broad. The NFM plot suggests that the small correlation coefficients above the magenta line (the CMIP5 MMM) found in the majority of CMIP5 coupled models are not likely unless the ocean forcing is less than 20%. This of course does not rule out the possibility that one of the models in the more strongly ocean-forced minority is closer to the truth.

The high-pass data are computed as the difference of the unfiltered and the low-pass data. High-pass covariances are approximately equal to the unfiltered covariance for the NFM since the LP covariance is typically lower order $O(\omega_c)$. The high-pass correlation is complicated by the fact that the unfiltered covariance [Eq. (A3c)] and the correlation $\rho(Q_s, T)$ [Eq. (A7c)] are discontinuous at zero lag, with different values at lags of 0_-, 0, and 0_+. In Fig. 7, we plot the unfiltered correlation for the NFM at 0_-, that is, with Q_s infinitesimally leading T. This choice is consistent with the understanding that at short time scales the heating is driving the temperature and also with the explicit time differencing used in models. It is the choice that agrees with Gulev et al. (2013) in finding a positive correlation, but here the correlation is positive only if $a^2 > 1/2$. Only if the greater share of the forcing comes from the atmosphere does the NFM meet the criteria that were interpreted as support for the Bjerknes hypothesis. Note that in the NFM there is no change in the relative forcing of the atmosphere and ocean as the frequency changes. The difference between high- and low-pass behavior is created by the filter.

Our need to choose one of the several disparate values of the unfiltered correlation at and near zero lag is a sign that the high-pass correlation is problematic. In a valuable study, O' Reilly et al. (2016) extended Gulev et al. (2013) by calculating the same quantities in CMIP models. They found the same LP correlation between T and turbulent surface fluxes in observationally based estimates and in models but remark on the inability of the multimodel mean to reproduce the positive relation that they and Gulev et al. (2013) find in the fluxes estimated from observations. Cayan (1992) studied the relation between turbulent fluxes and temperature tendency and found that, for unfiltered monthly data in midlatitudes, the relation typically shows the atmosphere to be heating the ocean. Perhaps this is closer to what Bjerknes (1964) had in mind than the relation of fluxes to temperature. Figure 6 shows that we find the same positive relation between dT/dt and Q_s as Cayan (1992) in the coupled model and in the NFM, and Fig. 5

shows that this is also true of the SOM. Yu et al. (2011) considered the covariability of T and surface heat fluxes in reanalyses and CMIP3 models. (Their Q_s includes radiative as well as turbulent fluxes.) For the North Atlantic (30°–60°W, 30°–50°N) at zero lag, they find a positive relation in the ERA-40 reanalysis, but a zero covariance for the NCEP-2 reanalysis (see their Fig. 6). The multimodel mean has negative covariance at zero lag, with individual models being both positive and negative. They do find a positive relation for all models and both reanalyses when the surface flux leads temperature by one month. O'Reilly et al. (2016) have the same result for Q_s leading by a month. We conclude that, while the zero lead relation found by Gulev et al. (2013) is not robust across other observational products or models, the idea of the atmospheric flux forcing the ocean at short time scales is confirmed by other diagnostics.

Gulev et al. (2013) identify a spectral peak in the observations for the 128 years 1880–2007 at a 50–70-yr period. They find the peak to be significant at the 95% level, and, while other spectral methods do not yield the same significance (e.g., viz., Fig. 2 in Ba et al. 2014 or Fig. 2 of Peings et al. 2016), we believe the peak is more likely than not to be real (for reasons discussed below). A low-frequency peak raises the question of whether our NFM is as relevant to the observations as it appears to be for the models. To address this concern, we add a low-frequency periodic forcing $F = c\sin\phi t$ with $\phi < \omega_c \ll 1$ to Eq. (4), the nondimensional version of Eq. (1). O'Reilly et al. (2016) studied a similar model. We do not assert that this is the right form of forcing; we are just exploring the effects of low-frequency (e.g., multidecadal) forcing, and, as so often done, we use a single frequency as an example. Equation (4) is now

$$\frac{dT}{dt} = -T + q_a + q_o + c\sin(\phi t). \tag{16}$$

Since the noise variance $\mathbb{E}\{(q_a + q_o)^2\} = 1$ and the variance of $\sin(\phi t) = 1/2$, we interpret $c^2/2$ as the signal-to-noise (S/N) ratio. The appendix gives the solution and mathematical expressions for relevant covariances and correlations. Because LP filtering leaves the low-frequency signal almost unchanged from the unfiltered version, the high-pass signal, which is the difference of the two, is not affected by the inclusion of a low-frequency forcing.[9] Equation (15) tells us immediately that adding F as ocean forcing strengthens the conclusion that at zero lag $\mathbb{E}_{\rm LP}\{Q_s, T\} < 0$ because it increases $\mathbb{E}_{\rm LP}\{q_o, q_o\}$. In short, it does not matter whether the ocean forcing is noise or signal: as long as it does something—anything—the LP filtered surface flux will be out of the ocean when temperature is high. Further, if the surface flux Q_s includes only turbulent fluxes as in Gulev et at (2013), then leaving the atmospheric radiative fluxes out of Q_s ensures that $\mathbb{E}_{\rm LP}\{Q_s, T\} < 0$ even if the ocean does nothing. The case of low-frequency forcing is illustrated in Fig. 9. For this example, we took the ocean to contribute only 15% to the noise forcing ($b^2 = 0.15$) and the S/N ratio, $c^2/2 = 0.05$. The top panel shows that the temperature has the signal imposed by the periodic forcing plus noise and that the LP filtered temperature is almost all periodic signal. In the next panel, the correlation of T_t and T is a maximum at a quarter-wavelength lead (15 yr). In contrast to the noise-only case (dotted line), the locations of the extrema do not change if the cutoff of the LP filter changes. The third panel shows that the correlation of Q_s and T at zero lag is negative, as expected from Eq. (15) and the appendix.

The more interesting case where the periodic forcing is added to the atmosphere—to the surface heat flux—is also illustrated in Fig. 9. The top two panels are the same as for the ocean forcing; the temperature does not care where the heat comes from and carries no telltale saying "atmosphere" or "ocean." In this simple model, the difference between the two cases lies only in our bookkeeping, in whether the periodic forcing is counted as part of the surface heat flux or the ocean heat flux. The bottom panel of Fig. 9 shows that the LP correlation between T and Q_s is again negative at zero lag, as it must be according to Eq. (15). Clearly, this relationship is not proof that the ocean is driving the temperature since it is obvious from the time series of $T(t)$ (top panel of Fig. 9) that the periodic atmospheric forcing is the driver.

However, the correlation is more negative when the periodic signal is in the ocean rather than the atmosphere. The context here is a very long time series generated by a model with a simple structure. In reality and in complex models, the low-frequency signal (if any) is not simply periodic, the other forcings are not as simple as those in our NFM, and the time series are short. There is only a distant prospect of using this difference in correlation strength to detect the source of the signal in a more complex setting with many processes and a time series short compared to the signal periods.

O'Reilly et al. (2016) introduced a simple model much like ours, but the only ocean forcing they consider is a periodic low-frequency signal; that is, their model is our

[9] This is not quite true for a running mean filter, which alters the signal even at frequencies deep into the passband.

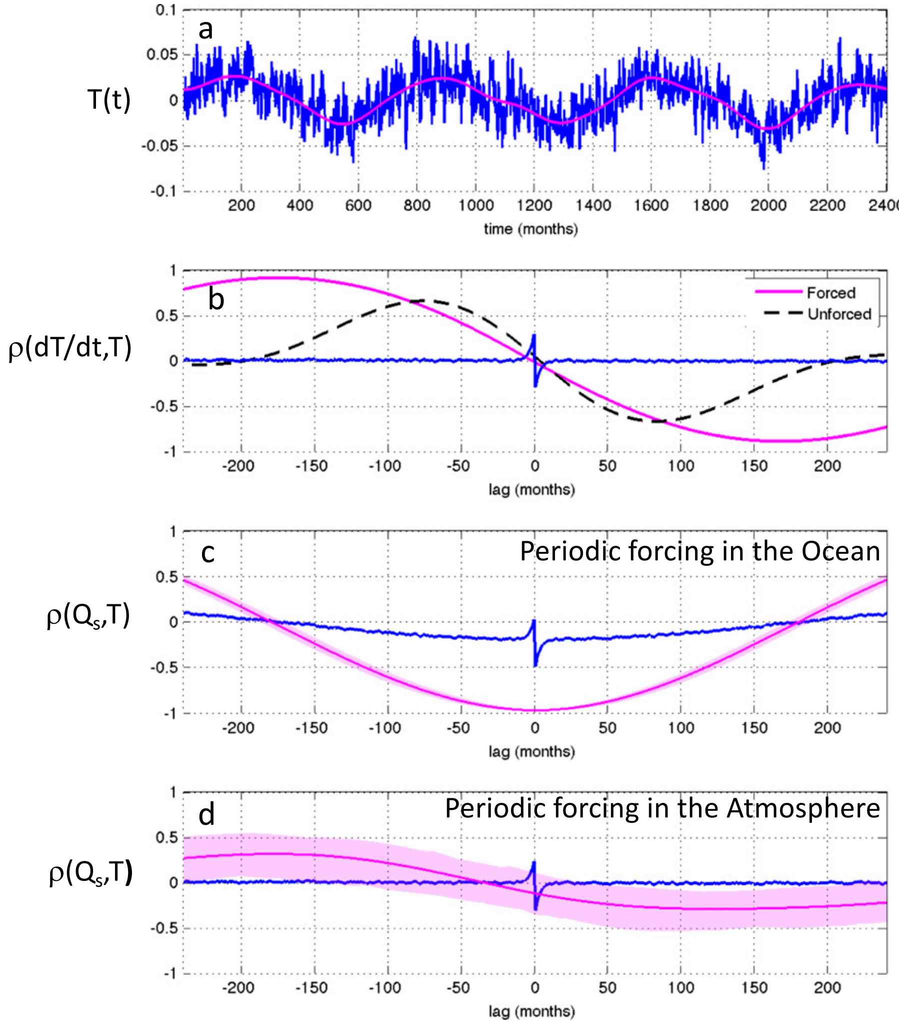

FIG. 9. Correlations in the simple NFM with a periodic forcing added. Equation (16) with $a^2 = 0.85, b^2 = 0.15,$ and $c^2 = 0.10$, so the S/N ratio is 0.05. Compare to Figs. 5 and 6. (a) The $\rho(T,\ T)$; (b) $\rho(T_t,\ T)$. The dashed curve is the correlation without the periodic forcing. (c) The $\rho(Q_s,\ T)$ with the periodic forcing in the ocean; (d) the $\rho(Q_s,\ T)$ with the periodic forcing in the atmosphere. Note that (a) and (b) are the same regardless of where the periodic forcing is. The temperature does not care where the heat comes from.

Eq. (16) with $q_o = 0$. They use an S/N ratio of 0.08, a period of 60 yr, and a damping time of $\alpha^{-1} = 4$ yr, justifying such a large value by appealing to the reemergence mechanism for ocean mixed layer anomalies. We do not agree that reemergence is properly viewed as a damping, but the long damping time is an incidental issue here. Because the only ocean forcing they consider is low frequency, they overlook the possibility that *any* ocean forcing, including white noise, is sufficient to make the LP correlation of surface flux and temperature negative when the flux leads. The conclusion that "these relationships . . . rely crucially on low-frequency forcing of SST" is too specific (O'Reilly et al. 2016, p. 2810). All that is needed from the ocean is a modicum of white noise.

4. Summary and discussion

We have used the simple damped noise-forced model (NFM) of Eq. (1) to help us interpret some results found in the literature that are often cited as evidence that the ocean is driving the AMO. In addition to a damping term that is linearly proportional to temperature, this model's temperature tendency is driven by white noise forcings in the atmosphere and ocean that are independent of one another. The analytic solutions we derived for both unfiltered and low-pass filtered correlations (e.g., of temperature and temperature tendency and of heat fluxes and temperature) provide some generic results that hold for all low-pass filters. We compare the correlations in this model to those for preindustrial (i.e., constant external forcing) runs of a coupled GCM [CESM1(CAM5)], those for the CMIP5 MMM, and for an atmospheric model (CAM5) coupled to a slab ocean model (SOM).

We find that the unfiltered correlations fall off rapidly since the *e*-folding time is just the damping time, which is less than a year. This is physical, but after applying an LP filter, the structure depends on the filter, not the physics. The low-pass correlations for this model depend on the autocorrelation function of the LP filter, and their lead–lag structures depend on the cutoff period used in the filter. For example, if the peak value of the LP correlation between dT/dt and T is found at 4 yr for a 10-yr filter cutoff, then it will be at 8 yr for a 20-yr filter and 12 yr for a 30-yr filter (see Fig. 4). While the location of the peak does vary somewhat with the type of filter (e.g., Butterworth, Hamming, and running mean), the linear variation of the peak and other properties with the period of the filter is a general result for all low-pass filters. (Judging by the many papers that do not specify what type of filter is used, our literature seems to have taken to heart the notion that the type of filter does not matter.)

We find that the autocorrelation in the coupled and SOM models of temperature for the $\mathrm{AMO_{mid}}$ index (the average temperature over the region 20°–60°W, 40°–55°N) is exceedingly similar to that of the noise-forced model (viz., Figs. 5, 6). This holds for both the unfiltered and the LP correlations, suggesting that in these complex models the temperature is also determined by white noise forcing. This behavior is a robust property of red noise, and Figs. 2 and 3 of C15, Fig. 2 in Ba et al. (2014), and Fig. 2 of Peings et al. (2016) show that many models have this red spectrum. Moreover, the correlation of dT/dt and T in the CESM1 and the SOM are also indistinguishable from the NFM (Figs. 5, 6). We also note that Zhang et al. (2016) report that, for this correlation in the multimodel mean of 10 CMIP3 models, the peak correlation moves as in the NFM, a behavior that distinguishes the response to white noise forcing from the response to a true low-frequency forcing. A caveat is that not all CMIP models have unfiltered correlations in such good agreement with the NFM. Some do not exhibit the simple exponential decay in the autocorrelation of $T(t)$ that we see here (Fig. 6) for CCSM and the CMIP5 multimodel mean as well as for the NFM. These differences will be the subject of future investigations.

We have seen that, when the underlying record is white or red noise, the LP filter will create long-period lead–lag structures constructed from the autocovariance of the filter and its derivatives. It would be good practice to check and see if there is a distinct low-frequency signal before applying a low-pass filter. The obvious check is to see if the spectrum is statistically indistinguishable from red or white noise. Another is to see if there is the tendency for the peak lag correlation (e.g., between dT/dt and T) or the lag at which a correlation crosses zero to move when the filter cutoff period is changed. It would be generally useful to try different filter time scales to see which features are robust. Another caution is that low-pass filters greatly reduce the effective sample size, so the resulting correlation estimates may be quite uncertain, as shown in Fig. 8.

A negative LP correlation between sea surface temperature T and surface heat flux into the ocean Q_s at or near zero lag has been interpreted as meaning that since this is the surface heating responding to T, it must be the ocean that is driving the AMO variations. This interpretation has been applied to observations (Gulev et al. 2013; O'Reilly et al. 2016) and models (O'Reilly et al. 2016; Zhang et al. 2016; Drews and Greatbatch 2016). Here we show that this negative correlation is a necessary consequence of the surface heat balance at long time scales being in near equilibrium: that is, a state where the temperature tendency is small compared to the components of the heat fluxes that individually would induce large temperature changes. It is dictated by the negative temperature feedback associated with turbulent surface fluxes acting rapidly to balance the other heat fluxes. It tells us only that the net ocean plus atmosphere heat flux is near zero, but it is uninformative as to how the temperature got to be what it is. We show examples in Figs. 6 and 9 where the temperature is strongly driven by the atmosphere and yet the correlation between T and Q_s is negative at and near-zero lag. All that is needed is some forcing from the ocean; anything will do. In our model, the ocean forcing is white noise and provides no systematic signal. That is all it takes to make $\rho_{\mathrm{LP}}(T, Q_s) < 0$. It does not take much ocean forcing either: we find that white noise forcing that is 85% atmosphere and only 15% ocean is a good fit to the CMIP5 MMM and to most individual CMIP5 models. The negative LP correlation of T and Q_s is consistent with the forcing being all white noise, coming mostly from the atmosphere. The negative correlation does not demonstrate that the AMO must be driven by the ocean, as is often claimed.

It does imply that the ocean does something to SST. If the ocean did absolutely nothing then $\rho_{\mathrm{LP}}(T, Q_s) \approx 0$ because the surface flux Q_s, being the only flux, must be approximately zero since the temperature tendency is small. As shown by Eq. (12), the case of no ocean forcing whatsoever is singular in that it eliminates leading-order terms, changing the structure of the lead–lag correlations between Q_s and T or dT/dt. The ocean forcing has to be very small indeed for this to happen. For example, with a 10-yr running mean filter and a feedback of $20\,\mathrm{W\,m^{-2}\,K^{-1}}$ it follows from Eq. (12) that the ocean heating variance would have to be $\lesssim 2\%$ of that from the atmosphere.

Of course, there is no doubt that there is an ocean in both reality and coupled models. Here, we are saying that 1) the sign of this correlation cannot tell us whether the ocean or the atmosphere is driving the AMO; and 2) the lead–lag correlations of (T, T), $(dT/dt, T)$, $(Q_s, dT/dt)$, (Q_s, T) in the coupled model are consistent with the idea that the driving is primarily atmospheric white noise with a small addition of white noise from the ocean. The ocean contribution is not needed to explain the temperature structure in the models, but it is needed to explain the $\rho_{\mathrm{LP}}(T, Q_s)$ diagnostic that has incorrectly been taken as indicating that the ocean drives the AMO.

Our thinking is along the lines of the paradigm in Barsugli and Battisti (1998): the AMO pattern is a response to the atmospheric forcing that least damps that forcing. The surface heat exchange between ocean and atmosphere adjusts to what the ocean imposes in order to create or maintain that pattern. In some regions, the ocean heat convergence is helpful, in others harmful, but the surface heat exchange adjusts to do what is needed. We do not prove this hypothesis in this paper but mention it here as a suggestion of how the atmospheric driving of the AMO can carry the day in the context of an active ocean circulation.

While there is a traditional idea that the ocean is most active at long periods, recent observations show clearly that the ocean has power at all frequencies. Figure 2 shows that the ocean forcing in the preindustrial CGCM is indistinguishable from white noise. We do not have a long enough observational dataset to know if this is true of the real ocean, but recent observations from the Atlantic, such as those from the Rapid Climate Change programme (RAPID) support the idea that the ocean circulation has considerable power at all observed frequencies (e.g., Lozier 2012; McCarthy et al. 2015; Zhao and Johns 2014). At present, we have no direct observational evidence to say whether or not the heat flux convergence in the North Atlantic Ocean has more power at multidecadal time scales than expected from white noise. As noted above, however, the instrumental record of SST does exhibit greater power in the unfiltered AMO (or $\mathrm{AMO_{mid}}$) index at periods of about 70 yr, at least in some analyses (Gulev et al. 2013; Zhang and Wang 2013). In Murphy et al. (2017) and Bellomo et al. (2017), we add model-based evidence to a variety of other studies (Mann and Emanuel 2006; Otterå et al. 2010; Booth et al. 2012), suggesting that this long-period variability seen in the ~140 yr of instrumental data is externally forced by volcanic and anthropogenic aerosols, greenhouse gases, and variations in solar radiance, none of which are present in the PI models analyzed in this paper. Perhaps ocean dynamics influences the SST response to external forcing, but this has not yet been demonstrated, and it may be that ocean dynamics is not required for a plausible explanation of twentieth-century variability in North Atlantic average SST.

Understanding the AMO will be greatly helped by analysis of mechanisms in observations and coupled models, as recommended by Zhang et al. (2016). One of the few examples in the literature is Fig. 7 of Buckley and Marshall (2016), a heat budget analysis based on an analysis of the ECCO v4 ocean state estimate that is limited by being only a few decades long. It shows that over most of the North Atlantic the variance of ocean heat convergences is small compared to the local atmospheric heating, but the opposite is true along the western boundary, where the Gulf Stream turns offshore, and in parts of the subpolar gyre around Greenland and Iceland in particular. Overall, this is consistent with the idea that the SST is largely driven by the atmosphere, but it challenges us to accommodate the regions where ocean heat convergence is dominant.

The unanswered question revised by our work here is "just what is the role of the ocean in the AMO?" Unquestionably, the ocean is active. It would help to clarify what is meant by the claim that "the ocean drives the AMO." Does it mean that ocean physics amplifies atmospheric forcing, as suggested, for example, by Delworth et al. (2016)? Would a time varying atmospheric forcing produce a stronger AMO than the same run with a slab ocean? Clement et al. (2016) indicate otherwise. Does it mean that the ocean does not amplify atmospheric noise such as the NAO but does amplify low-frequency external forcings, for example, from aerosols and greenhouse gases? Does "the ocean drives the AMO" mean that variability entirely internal to the ocean does it? At the extreme, does an ocean model generate an AMO if coupled to a climatological atmosphere, a setup that would exclude atmospheric variability? To our knowledge, no such experiment has been done with the recent generations of models. If so, we would have to conclude that the ocean alone is sufficient for the AMO, just as we concluded in C15 that the atmosphere alone is sufficient. There may be more than one way to generate the AMO. We are unaware of any recent advocacy for the proposal that the ocean alone is sufficient, though much of the thinking about Atlantic variability stems from the seminal box model papers of Stommel (1961) and Rooth (1982) and the early coupled models in which regular multidecadal oscillations were attributed to the ocean circulation (e.g., Delworth et al. 1993).

It is well to remember that the models are all imperfect, particularly in the North Atlantic, where the mean state in models has a cold bias (Wang et al. 2014). In a higher-resolution coupled model, Siqueria and Kirtman (2016) showed that interactions with the mean state produced decadal time scale variability in the North Atlantic that is absent in a version of that model with a lower resolution, one comparable to the CMIP models presented in all prior work cited here. Recently, Drews and Greatbatch (2016) showed that these surface flux diagnostics were different in a model with a corrected mean state, suggesting that improvements to the model may change the influence of the ocean. However, in that study the only evidence for the change in behavior was in the correlation of surface fluxes with temperature. As we have shown here, the correct interpretation of that diagnostic is that the ocean is doing something, but not that it is important to the simulation of the climate variability as represented by the surface temperature. A simpler and more straightforward test that an ocean circulation is essential for the model's surface temperature response is to obtain a different SST when the active ocean is removed and replaced by a slab ocean.

Acknowledgments. We are grateful for fruitful conversations with Yochanan Kushnir, Mingfang Ting, Claude Frankignoul, Rong Zhang, Carl Wunsch, and Michael Tippett. We thank Laure Zanna and three anonymous reviewers for helpful reviews. MAC is supported by the Office of Naval Research under the Research Grant MURI (N00014-12-1-0911) and by the National Science Foundation Grant OCE-1657209, ACC and LNM are supported by National Science Foundation Grant NSF-1304540, and KB is supported by a cooperative agreement between NASA and Columbia University (NNX15AJ05A).

APPENDIX

Mathematical Derivations

a. Unfiltered covariances and correlations

We will need to calculate various covariances (or correlations) to compare with the results from the coupled and slab models. One could compute covariances from the solution [Eq. (5)] for $T(t)$, but it is easier to use the convolution theorem to compute expected values:

$$\begin{aligned}\mathbb{E}\{f(t+\tau)g(t)\} &= \int_{-\infty}^{+\infty} f(t+\tau)g(t)\,dt\\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{f}(\omega)\hat{g}^*(\omega)e^{-i\omega\tau}\,d\omega,\end{aligned}$$

where a hat denotes the Fourier transform and an asterisk the complex conjugate. We first take the Fourier transform of Eq. (4),

$$\begin{aligned}-i\omega\hat{T} &= -\hat{T}+\hat{q}_T = -\hat{T}+\hat{q}_a+\hat{q}_o;\\ \hat{T} &= \frac{\hat{q}_T}{1-i\omega}.\end{aligned}\tag{A1}$$

We can now calculate the various lagged covariances we will need.[10] In the formulas below sgn, H, and δ are the standard sign, Heaviside, and impulse (Dirac delta) functions, respectively. Only a few require the inversion of the Fourier integrals:

$$\begin{aligned}\mathbb{E}\{T(t+\tau)T(t)\} &= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{\hat{q}_T\hat{q}_T^*}{1+\omega^2}e^{-i\omega\tau}\,d\omega\\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{1}{1+\omega^2}e^{-i\omega\tau}\,d\omega = \frac{1}{2}e^{-|\tau|},\end{aligned}\tag{A2a}$$

$$\begin{aligned}\mathbb{E}\{T(t+\tau)q_a(t)\} &= a^2\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{1}{1-i\omega}e^{-i\omega\tau}\,d\omega\\ &= a^2e^{-|\tau|}H(\tau),\quad\text{and}\end{aligned}\tag{A2b}$$

$$\begin{aligned}\mathbb{E}\{T(t+\tau)q_o(t)\} &= b^2\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{1}{1-i\omega}e^{-i\omega\tau}\,d\omega\\ &= b^2e^{-|\tau|}H(\tau).\end{aligned}\tag{A2c}$$

There is ambiguity in the definition of $H(\tau)$ at $\tau=0$. Evaluation of the integrals in Eqs. (A2b) and (A2c) at $\tau=0$ shows that the appropriate interpretation here is $H(0)=1/2$. Using Eq. (A2), the other relationships we need may be derived with a bit of algebra and the substitution of $-T+q_T$ for T_t [viz. Eq. (4)]. We adopt the shorthand notation $\mathbb{E}\{f, g\}$ for $\mathbb{E}\{f(t+\tau)g(t)\}$:

$$\begin{aligned}\mathbb{E}\{T_t(t+\tau)T(t)\} &= \mathbb{E}\{-T+q_T, T\}\\ &= -\frac{1}{2}e^{-|\tau|}+e^{-|\tau|}H(\tau) = -\frac{1}{2}\operatorname{sgn}(\tau)e^{-|\tau|},\end{aligned}\tag{A3a}$$

[10] Frankignoul et al. (1998) give a different derivation of some of these results.

$$\begin{aligned}\mathbb{E}\{T_t(t+\tau)T_t(t)\} &= \mathbb{E}\{-T+q_T, -T+q_T\}\\ &= \frac{1}{2}e^{-|\tau|} - [e^{-|\tau|}H(\tau) + e^{-|\tau|}H(-\tau)] + \delta(\tau)\\ &= \delta(\tau) - \frac{1}{2}e^{-|\tau|},\end{aligned} \tag{A3b}$$

$$\begin{aligned}\mathbb{E}\{T(t+\tau)Q_s(t)\} &= -\mathbb{E}\{T(t+\tau)T(t)\} + \mathbb{E}\{T(t+\tau)q_a(t)\}\\ &= -\frac{1}{2}e^{-|\tau|} + a^2e^{-|\tau|}H(\tau),\end{aligned} \tag{A3c}$$

$$\begin{aligned}\mathbb{E}\{T_t(t+\tau)Q_s(t)\} &= \mathbb{E}\{-T+q_T, -T+q_a\}\\ &= \left[\frac{1}{2}\operatorname{sgn}(\tau) - a^2H(\tau)\right]e^{-|\tau|} + a^2\delta(\tau),\end{aligned} \tag{A3d}$$

and

$$\begin{aligned}\mathbb{E}\{Q_s(t+\tau)Q_s(t)\} &= -\mathbb{E}\{T, Q_s\} + \mathbb{E}\{q_a, -T+q_a\}\\ &= \left[\frac{1}{2} - a^2H(\tau)\right]e^{-|\tau|} - a^2H(-\tau)e^{-|\tau|}\\ &\quad + a^2\delta(\tau) = \left(\frac{1}{2} - a^2\right)e^{-|\tau|} + a^2\delta(\tau).\end{aligned} \tag{A3e}$$

To compute correlations involving $f(t)$, we will need $\sigma^2(f) = \mathbb{E}\{f(t)^2\}$:

$$\begin{aligned}\sigma^2(T) &= \frac{1}{2};\ \sigma^2(T_t) = \delta(\tau) - \frac{1}{2};\ \sigma^2(Q_s)\\ &= \left(\frac{1}{2} - a^2\right) + a^2\delta(\tau);\ \sigma^2(q_o) = b^2\delta(\tau).\end{aligned} \tag{A4}$$

The presence of delta functions in these formulas is indicative of a mismatch between the continuous time formulation used here and the finite time steps necessary for computer models like GCMs or even for the numerical integration of the simple red noise model [Eq. (1)]. In the finite time difference models, the noise term must be constant for at least the duration of a time step, in contrast to the idealized noise that is uncorrelated from instant to instant [viz. Eq.(3)]. The continuous results above have to be adjusted before they can be compared to the finite difference models. One approach would be to redo the above with noise of finite duration: for example, having the noise stay constant over each time interval of length ε and solving Eq. (5) accordingly (as in Miller and Cane 1989). However, this greatly complicates the calculation, and we are addicted to the efficacy of the differential calculus. Instead, it is sufficient to average the solutions over a (small) time interval $(-\varepsilon,\ +\varepsilon)$:

$$\bar{f}(\tau) = \frac{1}{2\varepsilon}\int_{-\varepsilon}^{+\varepsilon} f(\tau + t')\,dt'. \tag{A5}$$

For small ε and absent a δ function this averaging makes little difference: $\bar{f} \approx f$. If there is a δ function (which is nonzero only at $\tau = 0$), then the term with the δ function dominates, and, with $f(\tau) = \delta(\tau)$,

$$\bar{f}(\tau) = \frac{1}{2\varepsilon}\int_{-\varepsilon}^{+\varepsilon} \delta(\tau + t')\,dt' = (2\varepsilon)^{-1}\delta_\varepsilon(\tau), \tag{A6}$$

so the function $\delta_\varepsilon(\tau) = 1$ if $|\tau| < \varepsilon$ and $\delta_\varepsilon(\tau) = 0$ otherwise. The averaging means we replace $\delta(\tau)$ with $(2\varepsilon)^{-1}\delta_\varepsilon(\tau)$ in all formulas. This averaging procedure introduces a new parameter, the averaging interval ε, but it is effectively chosen by the model we are trying to emulate. If we have model output for each time step Δt, then we could take $\varepsilon = \Delta t$, but if all that is available is monthly data, then the averaging interval might as well be one month.

We can now compute the correlations corresponding to what is shown in the figures. We will write all variables without the bar, but in fact all should be understood as the averages defined in Eqs. (A5) and (A6). Writing $\rho(f, g)$ for the correlation between $f(t+\tau)$ and $g(t)$,

$$\rho(T, T) = e^{-|\tau|}, \tag{A7a}$$

$$\rho(T_t, T) = \frac{-\frac{1}{2}\operatorname{sgn}(\tau)e^{-|\tau|}}{\sqrt{\frac{1}{2}\left[(2\varepsilon)^{-1} - \frac{1}{2}\right]}} = -\sqrt{\frac{\varepsilon}{1-\varepsilon}}\operatorname{sgn}(\tau)e^{-|\tau|}, \tag{A7b}$$

$$\begin{aligned}\rho(Q_s, T) &= \frac{\sqrt{\varepsilon}[2a^2H(-\tau) - 1]}{[a^2 + \varepsilon(b^2 - a^2)]^{1/2}}e^{-|\tau|}\\ &\approx \frac{\sqrt{\varepsilon}[2a^2H(-\tau) - 1]e^{-|\tau|}}{a},\end{aligned} \tag{A7c}$$

$$\rho(T_t, Q_s) = \frac{a^2\delta_\varepsilon(\tau) + \varepsilon[\operatorname{sgn}(\tau) - 2a^2H(\tau)]e^{-|\tau|}}{\{(1-\varepsilon)[a^2 + \varepsilon(b^2 - a^2)]\}^{1/2}} \approx a\delta_\varepsilon(\tau), \tag{A7d}$$

and

$$\rho(q_o, T) = 2\sqrt{\varepsilon}be^{-|\tau|}H(-\tau). \tag{A7e}$$

The two approximate formulas require $a \neq 0$; if $a = 0$, then $b = 1$ and

$$\rho(Q_s, T) = -e^{-|\tau|};\quad \rho(T_t, Q_s) = \sqrt{\frac{\varepsilon}{1-\varepsilon}}\operatorname{sgn}(\tau)e^{-|\tau|}. \tag{A8}$$

b. *Covariances after low-pass linear filtering*

We begin by rewriting Eq. (8) for the low-pass covariance $\mathbb{E}_{\rm LP}\{f, g\}$ in terms of the low-pass time $s = \omega_c t$:

$$\mathbb{E}_{\rm LP}\{f,g\} = \omega_c^{-1}\int_{-\infty}^{+\infty} R(s'-s)\mathbb{E}\{f(\omega_c^{-1}s'+t')g(t')\}\,ds'. \tag{A9}$$

The unfiltered covariance $\mathbb{E}$ in Eq. (A9) is a function of $\tau = \omega_c^{-1}s$. Inspection of Eqs. (A3) and (A2) shows that all the covariances listed can be constructed as linear combinations of just three functions: $\delta(\tau)$, $H(+\tau)e^{-|\tau|}$, and $H(-\tau)e^{-|\tau|}$. In view of Eq. (A9), we can build all LP covariances from the three integrals

$$P^0 = \omega_c^{-1}\int_{-\infty}^{+\infty} R(s'-s)\delta(s'/\omega_c)\,ds' = R(-s) = R(s), \tag{A10a}$$

$$\begin{aligned} P^+ &= \omega_c^{-1}\int_0^{\infty} R(s'-s)e^{-s'/\omega_c}\,ds' \sim \sum_{n=0}^{\infty}\omega_c^n R^{(n)}(-s) \\ &= \sum_{n=0}^{\infty}(-1)^n\omega_c^n R^{(n)}(s), \quad \text{and} \end{aligned} \tag{A10b}$$

$$\begin{aligned} P^- &= \omega_c^{-1}\int_{-\infty}^{0} R(s'-s)e^{-|s'/\omega_c|}\,ds' \\ &= \omega_c^{-1}\int_0^{\infty} R(-s'-s)e^{-|s'/\omega_c|}\,ds' \\ &= \omega_c^{-1}\int_0^{\infty} R(s'+s)e^{-s'/\omega_c}\,ds' \sim \sum_{n=0}^{\infty}\omega_c^n R^{(n)}(s), \end{aligned} \tag{A10c}$$

where $R^{(n)} \equiv d^nR/ds^n$. We used Watson's Lemma to obtain the sums in the last two equations, exploiting the fact that $1/\omega_c \gg 1$. We also made liberal use of the fact that R is an even function so $R(-s) = R(s)$ and $R^{(n)}(-s) = (-1)^n R^{(n)}(s)$.

We use the integrals [Eq. (A10)] and the unfiltered covariances [Eqs. (A3) and (A2)] in Eq. (A9) to obtain the LP covariances:

$$\mathbb{E}_{\rm LP}\{T,T\} = \frac{1}{2}(P^+ + P^-) = R(s) + O(\omega_c^2), \tag{A11a}$$

$$\mathbb{E}_{\rm LP}\{T,q_a\} = a^2P^+ = a^2R(s) - a^2\omega_c R_s(s) + O(\omega_c^2), \tag{A11b}$$

$$\mathbb{E}_{\rm LP}\{T,q_o\} = b^2P^+ = b^2R(s) - b^2\omega_c R_s(s) + O(\omega_c^2), \tag{A11c}$$

$$\mathbb{E}_{\rm LP}\{q_a,q_a\} = a^2P^0 = a^2R(s); \quad \mathbb{E}_{\rm LP}\{q_o,q_o\} = b^2R(s), \tag{A11d}$$

$$\mathbb{E}_{\rm LP}\{T_t,T\} = -\frac{1}{2}(P^+ - P^-) = \omega_c R_s(s) + O(\omega_c^3), \tag{A11e}$$

$$\mathbb{E}_{\rm LP}\{T_t,T_t\} = P^0 - \frac{1}{2}(P^+ + P^-) = -\omega_c^2 R_{ss}(s) + O(\omega_c^4), \tag{A11f}$$

$$\begin{aligned} \mathbb{E}_{\rm LP}\{T,Q_s\} &= -\frac{1}{2}(P^+ + P^-) + a^2P^+ = -b^2R(s) \\ &\quad - a^2\omega_c R_s(s) + O(\omega_c^2), \end{aligned} \tag{A11g}$$

$$\begin{aligned} \mathbb{E}_{\rm LP}\{T_t,Q_s\} &= \frac{1}{2}(P^+ - P^-) - a^2P^+ + a^2P^0 \\ &= -b^2\omega_c R_s(s) - a^2\omega_c^2 R_{ss}(s) + O(\omega_c^3), \quad \text{and} \end{aligned} \tag{A11h}$$

$$\begin{aligned} \mathbb{E}_{\rm LP}\{Q_s,Q_s\} &= \left(\frac{1}{2} - a^2\right)(P^+ + P^-) + a^2P^0 \\ &= b^2R(s) - (a^2 - b^2)\omega_c^2 R_{ss}(s) + O(\omega_c^4). \end{aligned} \tag{A11i}$$

In all cases, we include only the two highest-order terms in ω_c, but the expression for $\mathbb{E}_{\rm LP}\{Q_s, Q_s\}$ is taken to a higher order to avoid a degeneracy when $b = 0$.

c. *If q_a and q_o covary*

We assumed in Eq. (3) that q_a and q_o are uncorrelated. Now suppose that they covary:

$$\mathbb{E}\{q_a,q_o\} = \mu\tilde{g}^2, \tag{A12}$$

where $\mu = \pm 1$ to allow covariances to be positive or negative. We set

$$q_a = \tilde{a}\zeta_1 + \tilde{g}\zeta_2; \quad q_o = \tilde{b}\zeta_3 + \mu\tilde{g}\zeta_2, \tag{A13}$$

where the ζ's are white in time-independent random numbers with zero mean and unit variance: $\mathbb{E}\{\zeta_i\} = 0; \mathbb{E}\{\zeta_i, \zeta_j\} = \delta_{i,j}$. Hence, with $q_T \equiv q_a + q_o = \tilde{a}\zeta_1 + (1+\mu)\tilde{g}\zeta_2 + \tilde{b}\zeta_3$,

$$|q_a|^2 = \tilde{a}^2 + \tilde{g}^2, \tag{A14a}$$

$$|q_o|^2 = \tilde{b}^2 + \tilde{g}^2, \quad \text{and} \tag{A14b}$$

$$|q_T|^2 = \tilde{a}^2 + \tilde{b}^2 + (1+\mu)^2\tilde{g}^2 = \tilde{a}^2 + \tilde{b}^2 + 2(1+\mu)\tilde{g}^2. \tag{A14c}$$

As before, we will normalize so that $q_T^2 = 1$. Note that now $|q_T|^2 \neq |q_a|^2 + |q_o|^2$. Define

$$a^2 = \tilde{a}^2 + (1+\mu)\tilde{g}^2, \quad b^2 = \tilde{b}^2 + (1+\mu)\tilde{g}^2. \tag{A15}$$

Then $a^2 + b^2 = q_T^2 = 1$. In the main text where q_a and q_o are uncorrelated, we had $|q_a|^2 = a^2$ and $|q_o|^2 = b^2$. Comparison of Eqs. (A14a) and (A14b) with Eq. (A15) shows this is not true here (for $\tilde{g} \neq 0$).

Working through the algebra shows that, with the different definitions of Eq. (A15) used here for a and b, the unfiltered covariances look exactly the same as Eqs. (A2) and

(A3), but in Eq. (A3e) the last term is now $|q_a|^2\delta(\tau)$ instead of $a^2\delta(\tau)$. Similarly, the low-pass filtered covariances have the same form as in Eq. (A11), but for Eq. (A11i) for $\mathbb{E}_{\rm LP}\{Q_s, Q_s\}$, in which $b^2R(s)$ is replaced by $|q_o|^2R(s)$. Consequently, Q_0, which appears in the denominator of the correlations in Eqs. (11c) and (11d), changes to

$$Q_0 = \sigma^2(Q_s) = |q_o|^2R(0) - \omega_c^2(a^2 - b^2), \quad \text{(A16)}$$

and in the approximate forms of Eqs. (11c) and (11d), the correlations involving Q_s, b is replaced by $b^2/|q_o| = |q_o| + \mu\tilde{g}^2/|q_o|$. In the main text, we used the correlation $\rho_{\rm LP}(Q_s, T)$ to estimate $b = |q_o|$. Here we see that this LP correlation overestimates $|q_o|$ if q_a and q_o covary positively ($\mu = +1$) and underestimates it if they covary negatively ($\mu = -1$). However, interpretation of $|q_o|^2$ as the fractional contribution of the ocean heat flux is no longer clear cut. If q_a and q_o covary negatively, then Eq. (A14c) shows that the common variance $\tilde{g}^2$ cancels out of the total variance q_T even as Eq. (A14b) (correctly) counts it as part of the ocean heat flux variance. Thus, $|q_a|^2 + |q_o|^2 > |q_T|^2$. If they covary positively, then Eq. (A14) shows that $|q_a|^2 + |q_o|^2 < |q_T|^2$, so this time the separate heat fluxes fall shy of the total. So perhaps a and b are better measures of the atmosphere and ocean contributions. As with $|q_o|$, $\rho_{\rm LP}(Q_s, T)$ overestimates b if $\mu = +1$ and underestimates it if $\mu = -1$.

We have limited discussion to the case where q_a and q_o have a simultaneous correlation rather than one with a time delay as, for example, in Czaja and Marshall (2000). Time delays allow a rich class of models that need careful constraint from observations and are beyond the scope of the present paper.

d. *Low-frequency periodic forcing*

Here we develop solutions for Eq. (16), the noise model with an added low period forcing. Since the equation is linear, we can add solutions for the different forcing terms. After the initial conditions are forgotten, the solution to Eq. (16) with $q_a = q_o = 0$ is

$$T(t) = c\sin(\phi t + \hat{\phi}); \quad \hat{\phi} = \arctan(\phi) \approx \phi. \quad \text{(A17)}$$

Hence,

$$\mathbb{E}_{\rm LP}\{T, T\} = \mathbb{E}\{T(t), T(t-\tau)\} = c^2\mathbb{E}\{\sin\phi(t-1), \sin\phi(t-\tau-1)\} = \frac{c^2}{2}\cos\phi\tau, \quad \text{(A18a)}$$

$$\mathbb{E}_{\rm LP}\{c\sin\phi t, T\} = \mathbb{E}\{c\sin\phi t, T(t-\tau)\} = c^2\mathbb{E}\{\sin\phi t, \sin\phi(t-\tau-1)\} = \frac{c^2}{2}\cos\phi(\tau+1), \quad \text{and} \quad \text{(A18b)}$$

$$\mathbb{E}_{\rm LP}\{T_t, T\} = \frac{d}{d\tau}\mathbb{E}\{T(t+\tau), T(t)\} = -\frac{c^2}{2}\phi\sin\phi\tau. \quad \text{(A18c)}$$

We may also add the covariance of the periodic solution [Eq. (A18)] to that for the noise-driven solution [Eq. (A11g)]. If the periodic forcing is in the ocean,

$$\mathbb{E}_{\rm LP}\{Q_s, T\} = -b^2R(s) + a^2\omega_cR_s(s) - \frac{c^2}{2}\cos\phi\tau; \quad \text{(A19)}$$

if it is in the atmosphere,

$$\begin{aligned}\mathbb{E}_{\rm LP}\{Q_s, T\} &= -b^2R(s) + a^2\omega_cR_s(s) + \frac{c^2}{2}[-\cos\phi\tau + \cos\phi(\tau+1)]\\ &= -b^2R(s) + a^2\omega_cR_s(s) + \frac{c^2}{2}(-\cos\phi\tau + \cos\phi\tau\cos\phi - \sin\phi\tau\sin\phi)\\ &= -b^2R(s) + a^2\omega_cR_s(s) - \frac{c^2}{2}\left[\phi\sin\phi\tau + \frac{\phi^2}{2}\cos\phi\tau + O(\phi^3)\right]. \end{aligned} \quad \text{(A20)}$$

Correlations also involve

$$\sigma^2(T) = R(0) + \frac{c^2}{2} = \omega_c + \frac{c^2}{2}; \quad \sigma^2(T_t) \approx -\omega_c^2R_{ss}(0) + \frac{c^2}{2}\phi^4 = \omega_c^3r_2 + \frac{c^2}{2}\phi^4;$$

$$\sigma^2(Q_s) = b^2R(0) - (2a^2-1)\omega_c^2R_{ss}(0) + \frac{c^2}{2} = b^2\omega_c + (2a^2-1)\omega_c^3r_2 + \frac{c^2}{2} \quad \text{(ocean)};$$

$$\sigma^2(Q_s) = b^2R(0) - (2a^2-1)\omega_c^2R_{ss}(0) + \frac{c^2\phi^2}{4} = b^2\omega_c + (2a^2-1)\omega_c^3r_2 + \frac{c^2\phi^2}{4} \quad \text{(atmosphere)},$$

where $r_2 = -R_{ss}(0)/\omega_c$ is $\lesssim O(1)$ and we have taken $R(0) = \omega_c$ [which is true for the running mean; for the general LP filter, $R(0) = O(\omega_c)$].

When the periodic forcing is in the ocean, then at $\tau = 0$, $\mathbb{E}_{\mathrm{LP}}\{Q_s, T\} = -b^2\omega_c - c^2/2$, which is negative unless there is no forcing from the ocean whatsoever. If it is in the atmosphere, then at $\tau = 0$, $\mathbb{E}_{\mathrm{LP}}\{Q_s, T\} = -b^2\omega_c - c^2\phi^2/4$. This too is negative.

REFERENCES

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

Barsugli, J. J., and D. S. Battisti, 1998: The basic effects of atmosphere–ocean thermal coupling on midlatitude variability. *J. Atmos. Sci.*, **55**, 477–493, doi:10.1175/1520-0469(1998)055<0477:TBEOAO>2.0.CO;2.

Bellomo, K., A. C. Clement, L. N. Murphy, L. M. Polvani, and M. A. Cane, 2016: New observational evidence for a positive cloud feedback that amplifies the Atlantic multidecadal oscillation. *Geophys. Res. Lett.*, **43**, 9852–9859, doi:10.1002/2016GL069961.

——, L. N. Murphy, M. A. Cane, A. C. Clement, and L. M. Polvani, 2017: Historical forcings as main drivers of the Atlantic multidecadal oscillation in the CESM large ensemble. *Climate Dyn.*, doi:10.1007/s00382-017-3834-3, in press.

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

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, doi:10.1038/nature10946.

Brown, P. T., M. S. Lozier, R. Zhang, and W. Li, 2016: The necessity of cloud feedback for a basin-scale Atlantic multidecadal oscillation. *Geophys. Res. Lett.*, **43**, 3955–3963, doi:10.1002/2016GL068303.

Buckley, M. W., and J. Marshall, 2016: Observations, inferences, and mechanisms of the Atlantic meridional overturning circulation: A review. *Rev. Geophys.*, **54**, 5–63, doi:10.1002/2015RG000493.

Cayan, D. R., 1992: Latent and sensible heat flux anomalies over the northern oceans: The connection to monthly atmospheric circulation. *J. Climate*, **5**, 354–369, doi:10.1175/1520-0442(1992)005<0354:LASHFA>2.0.CO;2.

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

——, M. A. Cane, L. N. Murphy, K. Bellomo, T. Mauritsen, and B. Stevens, 2016: Response to comment on "The Atlantic multidecadal oscillation without a role for ocean circulation." *Science*, **352**, 1527, doi:10.1126/science.aaf2575.

Czaja, A., and J. Marshall, 2000: On the interpretation of AGCMS response to prescribed time-varying SST anomalies. *Geophys. Res. Lett.*, **27**, 1927–1930, doi:10.1029/1999GL011322.

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

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

——, F. Zeng, G. A. Vecchi, X. Yang, L. Zhang, and R. Zhang, 2016: The North Atlantic Oscillation as a driver of rapid climate change in the Northern Hemisphere. *Nat. Geosci.*, **9**, 509–512, doi:10.1038/ngeo2738.

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, doi:10.1002/2016GL069815.

Fan, M., and E. K. Schneider, 2012: Observed decadal North Atlantic tripole SST variability. Part I: Weather noise forcing and coupled response. *J. Atmos. Sci.*, **69**, 35–50, doi:10.1175/JAS-D-11-018.1.

Foukal, N. P., and M. S. Lozier, 2016: No inter-gyre pathway for sea-surface temperature anomalies in the North Atlantic. *Nat. Commun.*, **7**, 11333, doi:10.1038/ncomms11333.

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

——, and E. Kestenare, 2002: The surface heat flux feedback. Part I: Estimates from observations in the Atlantic and the North Pacific. *Climate Dyn.*, **19**, 633–647, doi:10.1007/s00382-002-0252-x.

——, A. Czaja, and B. L'Heveder, 1998: Air–sea feedback in the North Atlantic and surface boundary conditions for ocean models. *J. Climate*, **11**, 2310–2324, doi:10.1175/1520-0442(1998)011<2310:ASFITN>2.0.CO;2.

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, doi:10.1038/nature12268.

Hannart, A., J. Pearl, F. E. L. Otto, P. Naveau, and M. Ghil, 2016: Causal counterfactual theory for the attribution of weather and climate-related events. *Bull. Amer. Meteor. Soc.*, **97**, 99–110, doi:10.1175/BAMS-D-14-00034.1.

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

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

Lozier, M. S., 2012: Overturning in the North Atlantic. *Annu. Rev. Mar. Sci.*, **4**, 291–315, doi:10.1146/annurev-marine-120710-100740.

Mann, M. E., and K. A. Emanuel, 2006: Atlantic hurricane trends linked to climate change. *Eos, Trans. Amer. Geophys. Union*, **87**, 233–241, doi:10.1029/2006EO240001.

McCarthy, G., and Coauthors, 2015: Measuring the Atlantic meridional overturning circulation at 26°N. *Prog. Oceanogr.*, **130**, 91–111, doi:10.1016/j.pocean.2014.10.006.

Miller, R. N., and M. A. Cane, 1989: A Kalman filter analysis of sea level height in the tropical Pacific. *J. Phys. Oceanogr.*, **19**, 773–790, doi:10.1175/1520-0485(1989)019<0773:AKFAOS>2.0.CO;2.

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

O'Reilly, C. H., 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, doi:10.1002/2016GL067925.

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

Park, S., C. Deser, and M. A. Alexander, 2005: Estimation of the surface heat flux response to sea surface temperature anomalies over the global oceans. *J. Climate*, **18**, 4582–4599, doi:10.1175/JCLI3521.1.

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

Rooth, C., 1982: Hydrology and ocean circulation. *Prog. Oceanogr.*, **11**, 131–149, doi:10.1016/0079-6611(82)90006-4.

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

Schneider, E. K., and M. Fan, 2012: Observed decadal North Atlantic tripole SST variability. Part II: Diagnosis of mechanisms. *J. Atmos. Sci.*, **69**, 51–64, doi:10.1175/JAS-D-11-019.1.

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

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

Stommel, H. M., 1961: Thermohaline convection with two stable regimes of flow. *Tellus*, **13**, 224–230, doi:10.3402/tellusa.v13i2.9491.

Trenary, L., and T. DelSole, 2016: Does the Atlantic multidecadal oscillation get its predictability from the Atlantic meridional overturning circulation? *J. Climate*, **29**, 5267–5280, doi:10.1175/JCLI-D-16-0030.1.

Wang, C., L. Zhang, S.-K. Lee, L. Wu, and C. R. Mechoso, 2014: A global perspective on CMIP5 climate model biases. *Nat. Climate Change*, **4**, 201–205, doi:10.1038/nclimate2118.

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

Yu, B., G. Boer, F. Zwiers, and W. Merryfield, 2011: Covariability of SST and surface heat fluxes in reanalyses and CMIP3 climate models. *Climate Dyn.*, **36**, 589–605, doi:10.1007/s00382-009-0669-6.

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, doi:10.1002/jgrc.20390.

Zhang, R., 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, doi:10.1126/science.aaf1660.

Zhao, J., and W. Johns, 2014: Wind-forced interannual variability of the Atlantic meridional overturning circulation at 26.5°N. *J. Geophys. Res. Oceans*, **119**, 2403–2419, doi:10.1002/2013JC009407.