Estimation of the SST Response to Anthropogenic and External Forcing and Its Impact on the Atlantic Multidecadal Oscillation and the Pacific Decadal Oscillation

Claude Frankignoul, Guillaume Gastineau, Young-Oh Kwon

Estimation of the SST Response to Anthropogenic and External Forcing and Its Impact on the Atlantic Multidecadal Oscillation and the Pacific Decadal Oscillation

CLAUDE FRANKIGNOUL AND GUILLAUME GASTINEAU

Sorbonne Universités, UPMC/CNRS/IRD/MNHN, LOCEAN/IPSL, Paris, France

YOUNG-OH KWON

Woods Hole Oceanographic Institution, Woods Hole, Massachusetts

(Manuscript received 6 January 2017, in final form 30 August 2017)

ABSTRACT

Two large ensembles (LEs) of historical climate simulations are used to compare how various statistical methods estimate the sea surface temperature (SST) changes due to anthropogenic and other external forcing, and how their removal affects the internally generated Atlantic multidecadal oscillation (AMO), Pacific decadal oscillation (PDO), and the SST footprint of the Atlantic meridional overturning circulation (AMOC). Removing the forced SST signal by subtracting the global mean SST (GM) or a linear regression on it (REGR) leads to large errors in the Pacific. Multidimensional ensemble empirical mode decomposition (MEEMD) and quadratic detrending only efficiently remove the forced SST signal in one LE, and cannot separate the short-term response to volcanic eruptions from natural SST variations. Removing a linear trend works poorly. Two methods based on linear inverse modeling (LIM), one where the leading LIM mode represents the forced signal and another using an optimal perturbation filter (LIMopt), perform consistently well. However, the first two LIM modes are sometimes needed to represent the forced signal, so the more robust LIMopt is recommended. In both LEs, the natural AMO variability seems largely driven by the AMOC in the subpolar North Atlantic, but not in the subtropics and tropics, and the scatter in the AMOC–AMO correlation is large between individual ensemble members. In three observational SST reconstructions for 1900–2015, linear and quadratic detrending, MEEMD, and GM yield somewhat different AMO behavior, and REGR yields smaller PDO amplitudes. Based on LIMopt, only about 30% of the AMO variability is internally generated, as opposed to more than 90% for the PDO. The natural SST variability contribution to global warming hiatus is discussed.

1. Introduction

The global mean surface temperature has risen in the last 100 years, but it has not done so uniformly in time, reflecting changes in the emission of greenhouse gases and aerosols, fluctuations in solar activity, and also internally generated variability driven by the interactions among the atmosphere, ocean, cryosphere, and land. The variability driven by Earth interactions, which would occur in the absence of external forcing, will be called natural climate variability. Separating the forced variability from natural variability is needed for detecting and attributing climate changes, estimating the climate response to external forcing, and identifying the unforced climate variability. This is of much interest since the natural climate variability substantially contributes to regional and perhaps global temperature changes on interannual–multidecadal time scales, as during the so-called warming hiatus or pause in surface warming seen in globally averaged surface temperature in the first decade of the twenty-first century. The pause in surface warming has been largely attributed to natural variability in the tropical Pacific Ocean (e.g., Meehl et al. 2013; Kosaka and Xie 2013; Trenberth et al. 2014; England et al. 2014), perhaps in part driven by Atlantic changes (McGregor et al. 2014; Chen and Tung 2014; Li et al. 2015; Ruprich-Robert et al. 2017). However, a recent SST bias correction suggests that there was no global warming slowdown (Karl et al. 2015). The

Corresponding author: Claude Frankignoul, cf@locean-ipsl.upmc.fr

DOI: 10.1175/JCLI-D-17-0009.1

substantial role of natural variability is also illustrated by the large spread in large ensembles of scenario simulations (Deser et al. 2014; Wettstein and Deser 2014).

Because of the short duration of the observational record, a careful removal of the forced signal is key to better understanding of the dynamical and thermodynamical processes that drive the main modes of low-frequency sea surface temperature (SST) variability, such as the Atlantic multidecadal oscillation (AMO) and the Pacific decadal oscillation (PDO). For instance, it is generally considered that the Atlantic meridional overturning circulation (AMOC) in part drives the AMO (e.g., Delworth et al. 1993; Knight et al. 2005), and most climate models indeed show that the AMOC leads the AMO, albeit by a model-dependent time lag (Medhaug and Furevik 2011; Gastineau and Frankignoul 2012). The integral response of the upper ocean to stochastic atmospheric forcing also contributes to the AMO variability, as recently emphasized by Clement et al. (2015). Hence, it is of interest to compare the AMO and the AMOC SST fingerprint and investigate their relationship. Marini and Frankignoul (2014) showed that removing the forced signal in historical simulations with three climate models lead to a lag AMOC–AMO correlation that was more similar to that found in control simulations, and the agreement depended on how the forced SST signal was estimated. Tandon and Kushner (2015) also showed that external forcing interferes with the AMOC–AMO relationship and that removing forced variations leads to closer agreement between models or between ensemble members in a large ensemble. The link between the AMO and the PDO is also of interest. D'Orgeville and Peltier (2007) found observational evidence that the AMO leads the PDO by 13 yr and the PDO leads the (negative) AMO by 17 yr, arguing that they are signature of the same oscillation cycle. However, by removing the global warming signal with linear inverse modeling, Marini and Frankignoul (2014) suggested that the AMO plays the leading role in the interbasin connection since the correlation remains strong when the AMO leads but decreases when the PDO leads. The relation between low-frequency modes is thus sensitive to the method used for separating forced and natural climate variability.

Two different approaches have been used to estimate the space–time pattern of the forced climate response, one that solely relies on observations and one that uses the response patterns of historical or scenario runs with state-of-the art climate models. The latter approach is based on multimodel simulations, sometimes with multiple ensemble members, and it involves statistical methods to better identify the forced component, such as scaling factors (Franckombe et al. 2015), signal-to-noise-maximizing empirical orthogonal functions (EOFs) (Ting et al. 2009), or discriminant analysis and maximization of the average predictability time (Delsole et al. 2011). Although it has been argued that a multimodel approach should strongly reduce the impact of individual model biases, there is no guarantee that it will yield unbiased results, since most climate models share similar components and parameterizations (Knutti et al. 2013) and might have been tuned to reproduce the observed twentieth-century warming (Hourdin et al. 2017). In addition, there remain substantial uncertainties in the anthropogenic and external forcing prescribed in historical runs, and current climate models generally do not capture regional trend patterns well (Shin and Sardeshmukh 2011). A variant to the multimodel approach is to use a large ensemble of a single climate model (e.g., Branstator and Selten 2009), which allows a less ambiguous separation between the externally forced climate change and natural climate variability in the model. Such large ensemble can also be used to estimate the thermodynamically induced trend representative of the forced components (Deser et al. 2016).

Here, we solely focus on methods that only rely on observations. They are not affected by model biases but are affected by the substantial observational uncertainties, in particular before the second half of the twentieth century, and by the method of SST or air temperature reconstructions. A common practice in the observation-based approach is to use the linear trend to characterize secular changes, especially in short records (e.g., Swart et al. 2015; Vincent et al. 2015), and to remove it when analyzing the main modes of natural climate variability or their impact (e.g., Knight et al. 2005; Sutton and Hodson 2005; Alexander et al. 2014; Delworth et al. 2017; Lyu et al. 2017; Osborne et al. 2017). This simply assumes that the forced climate evolution is primarily linear in time, even though observations and climate model simulations show that it is not really appropriate. Trenberth and Shea (2006) stressed that the global SST warming has intensified in recent decades, so that the AMO index (the low-pass-filtered SST averaged over the North Atlantic) is better defined by first subtracting the global mean SST, which turned out to substantially reduce the AMO amplitude. A similar strategy had been used by Zhang et al. (1997) to study ENSO-like interdecadal variability, and by Mantua et al. (1997) to define the PDO. Enfield and Cid-Serrano (2010) argued that removing the globally averaged SST or the regression on it also removes part of the natural climate variability; hence, they subtracted a quadratic trend for investigating the natural climate

variability. Several other methods have been proposed to separate external forcing from internal variability. Parker et al. (2007) used the first EOF of low-pass-filtered SST and air temperature to represent the forced SST signal, while Guan and Nigam (2009) used rotated extended EOFs. Ting et al. (2009) used a regression on the global average of the SST or the surface temperature to represent the forced signal. More sophisticated methods have also been used. Compo and Sardeshmukh (2010), Newman (2013), and Marini and Frankignoul (2014) used a dynamical filter based on linear inverse modeling (LIM; e.g., Penland and Matrosova 1994, 2006), which decomposes the SST into nonorthogonal normal modes and can single out the SST signature of ENSO and the global secular trend pattern. Hannachi (2007) introduced trend EOFs obtained using correlations between time positions of the sorted data to capture the different trend patterns of a field, which was generalized by Li et al. (2011) to extract coupled trends between SST and latent heat flux using singular value decomposition (SVD). The empirical mode decomposition (EMD; Huang et al. 1998) and its extensions, the ensemble empirical mode decomposition (EEMD; Wu and Huang 2009), and the multidimensional EEMD (MEEMD; Wu et al. 2009; Ji et al. 2014) have been used to decompose a time series into amplitude–frequency-modulated components of increasing time scales, and applied to temperature datasets.

Here we compare the ability of several of the observation-based methods at estimating and removing the forced climate response, using two large ensembles (LEs) of climate simulations subject to identical natural and anthropogenic forcing, but slightly different initial conditions, as a substitute for the real climate system. The first ensemble was conducted with the IPSL-CM5A-LR, and it contains $n = 30$ members. The second ensemble was conducted with the Community Earth System Model, version 1 (CESM1), and it contains $n = 40$ members (Kay et al. 2015). These LE simulations provide a unique opportunity to systematically assess the efficiency of each statistical method at extracting the forced signal in a single realization (an analogy to the observations). Indeed, the forced climate change signal can be estimated by the ensemble mean, as the noise due to the natural climate variability is reduced by $\sqrt{n}$, so that it is 5.5 or 6.3 times smaller. Furthermore, by comparing LEs using two distinct models, the robustness of the findings may be better judged.

Seven methods, namely based on linear and quadratic trend, global mean SST, regression on the global mean SST, LIM, optimal perturbation filter, and MEEMD are applied to each simulation to extract the forced signal. The estimated forced signal is then compared to the ensemble mean, which is our target truth for the respective LE. We compare the ability of these methods at representing both the long-term climate changes and the short-term response to large volcanic eruptions. As additional criteria, we compare the natural variability of the AMO and the PDO, and the AMOC SST footprint, after removing the estimated forced signal. Finally, we apply the methods to three observation-based SST reconstructions during 1900–2015.

2. Estimation of the forced climate signal

a. Linear trend

In this simplest case, the forced climate signal is represented by a linear trend (hereafter D1), which is estimated by least squares fit. Here we apply it to the yearly mean SST anomaly at each grid point after slight smoothing by a binomial (¼–½–¼) filter. The spatial pattern of the forced signal evolves with time, as the trend differs between grid points.

b. Quadratic trend

Quadratic detrending (hereafter D2) consists of fitting instead a second-order polynomial. The spatial pattern of the estimated forced signal also evolves with time.

c. Global mean SST

Following Trenberth and Shea (2006) but for a slight smoothing, the global forced signal is approximated at each grid point by the yearly averaged global mean SST anomaly smoothed by a binomial (¼–½–¼) filter (hereafter GM). The estimated forced signal is identical at all grid points.

d. Regression on the global mean SST

The nonuniformity of the global warming signal is accounted for by using a regression of the SST anomaly at each grid point on the yearly global mean SST anomaly time series (i.e., the GM time series) after smoothing by the binomial filter (hereafter REGR method). The estimated forced signal has spatially nonuniform amplitude.

e. Linear inverse model

The LIM assumes that the SST anomalies $\mathbf{x}$ are well approximated by a multivariate linear Markov process

$$\frac{d\mathbf{x}}{dt} = \mathbf{B}\mathbf{x} + \mathbf{F}, \tag{1}$$

where $\mathbf{B}$ is a linear operator, and $\mathbf{F}$ is white noise (e.g., Penland and Sardeshmukh 1995). The most probable forward solution of (1) at time $t + \tau$ is $\mathbf{x}(t + \tau) = \mathbf{G}(\tau)\mathbf{x}(t)$,

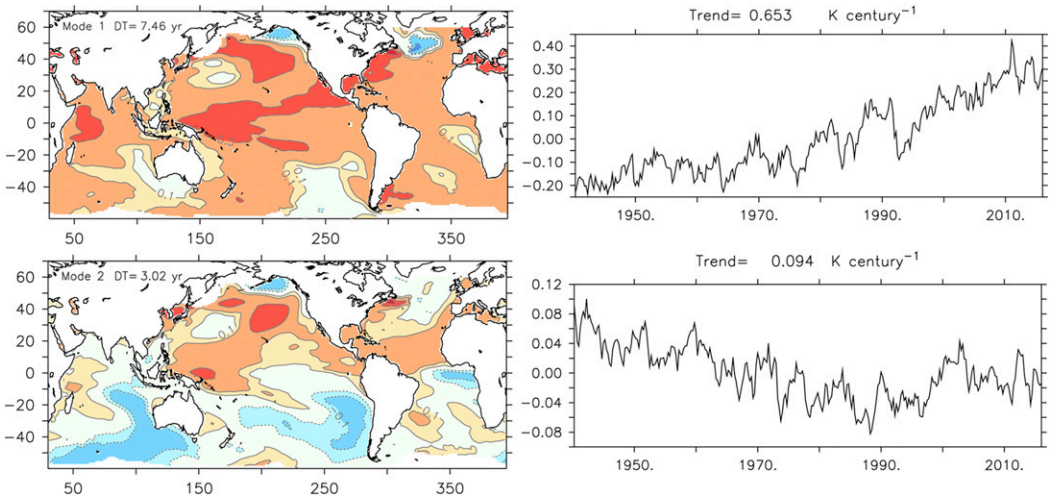

FIG. 1. (left) Spatial pattern of the first two LIM eigenmodes (K) with their damping time (DT; yr), and (right) associated time series for CESM-LE ensemble member 21. The linear trend is given; higher eigenmodes have no secular trend.

where $\mathbf{G}(\tau) = \exp(\mathbf{B}\tau)$. Three-month averages of SST anomalies in January–March (JFM), April–June (AMJ), July–September (JAS), and October–December (OND) are used for $\mathbf{x}$, with $\tau = 3$ months. The dimensionality is reduced by working in a truncated rotated-EOF space, retaining the first 15 rotated near-global (60°S–60°N) EOFs,[1] which account for about 75% of the SST variance. Increasing the number of EOFs leads to similar results, but the LIM modes become less stable among ensemble members. The SST field is then decomposed into a sum of nonnormal eigenmodes (e.g., Penland and Matrosova 1994, 2006):

$$\mathbf{x}(t) = \sum_i \mathbf{u}_i \alpha_i(t), \tag{2}$$

where $\mathbf{u}_i$ is the eigenvector of $\mathbf{B}$ estimated by $\mathbf{B} = \tau^{-1} \ln[\mathbf{C}(\tau)\mathbf{C}(0)^{-1}]$ and associated with the eigenvalue β_i, and $\alpha_i(t)$ is the time series obtained by projecting $\mathbf{x}$ onto the corresponding adjoint eigenvector. Here $\mathbf{C}(\tau)$ is the lag covariance matrix of $\mathbf{x}$ at lag τ. Because $\mathbf{B}$ is a real, negative, but not symmetric matrix, the eigenmodes may be nonorthogonal and all the eigenvalues have negative real parts. There is thus no orthogonality assumption, unlike in EOF analysis. The ith eigenmode of $\mathbf{B}$ is either a stationary damped mode characterized by one pattern and a decay time or an oscillatory mode with two patterns, a period $2\pi/\mathrm{Im}(\beta_i)$, and a decay time $-1/\mathrm{Re}\,(\beta_i)$. In all cases, the first eigenmode (that with least damping) is a weakly damped mode whose time evolution has a large trend, as in Fig. 1; it is taken as the forced climate signal, as in Compo and Sardeshmukh (2010), Newman (2013), and Marini and Frankignoul (2014). In several ensemble members, the second damped mode seems to also contribute to the secular trend but, although we tried several selection criteria, taking the second LIM mode into account degraded the overall results. Hence, the forced signal has the fixed spatial pattern of the first mode.

f. *Linear inverse model with optimal perturbation filter*

Solomon and Newman (2012) showed that an optimal perturbation filter based on LIM was more efficient at removing ENSO than a set of LIM eigenmodes. Here we adapt their method to define a space and time varying estimate of the forced SST signal (hereafter LIMopt). The forced signal $\mathbf{F}(t)$ is represented by the variability from the data that evolves into the maximum possible anomaly after a time τ_e, using $\mathbf{\Phi}_1(\tau_e)$, the optimal initial structure, which is the normalized right singular vector of $\mathbf{G}(\tau_e) = \exp(\mathbf{B}\tau_e)$. Here we use $\tau_e = 2.5$ yr, much longer than in Solomon and Newman (2012), who used $\tau_e = 6$–9 months to estimate ENSO signal. As described in Solomon and Newman (2012), the filter iteratively considers the evolution over a longer period τ_1, taken here to be $\tau_1 = 20$ yr. The forced signal is calculated from the initial condition $t = 0$ by $\alpha(0) = \mathbf{\Phi}_1(\tau_e)\mathbf{x}(0)$:

$$\mathbf{F}(0) = \alpha(0)\mathbf{G}(0)\mathbf{\Phi}_1(\tau_e), \tag{3}$$

for $t = 1$ and $\alpha(1) = \mathbf{\Phi}_1(\tau_e)[\mathbf{x}(1) - \alpha(0)\mathbf{G}(1)\mathbf{\Phi}_1(\tau_e)]$:

[1] Similar results are obtained by calculating the rotated EOFs separately in the Indo-Pacific and the Atlantic, but more EOFs are required to represent a comparable amount of variance.

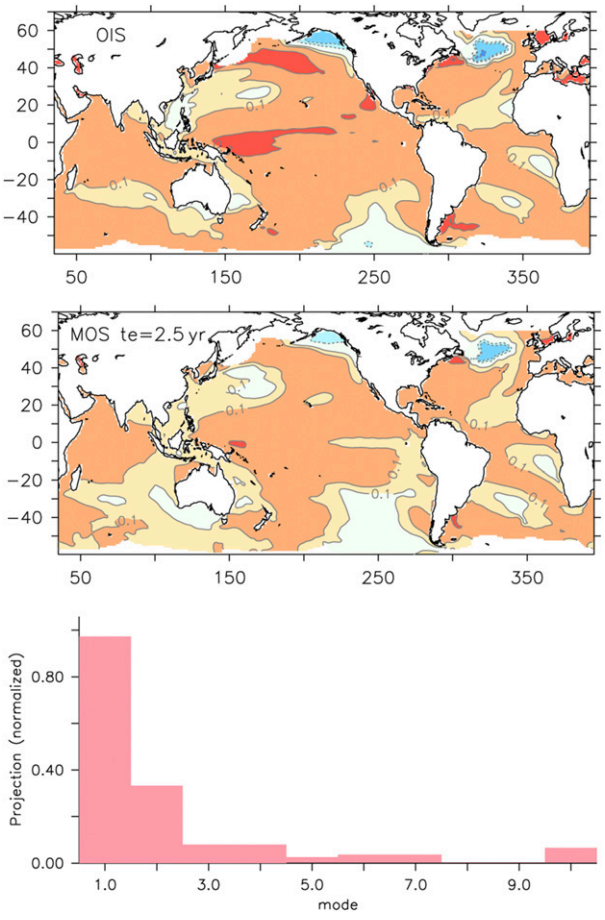

FIG. 2. (top) Optimal initial structure, (middle) its growth after 2.5 years, and (bottom) projection of the first LIM modes onto the optimal initial structure for CESM-LE ensemble member 21.

$$\mathbf{F}(1) = \alpha(0)\mathbf{G}(3)\mathbf{\Phi}_1(\tau_e) - \alpha(3)\mathbf{G}(0)\mathbf{\Phi}_1(\tau_e), \quad (4)$$

and for each $t \geq \tau_1$ and $\alpha(t) = \mathbf{\Phi}_1(\tau_e)\left[\boldsymbol{x}(t) - \sum_{\tau=1}^{\tau_1}\alpha(t-\tau)\mathbf{G}(\tau)\mathbf{\Phi}_1(\tau_e)\right]$:

$$\mathbf{F}(t) = \sum_{\tau=0}^{\tau_1}\alpha(t-\tau)\mathbf{G}(\tau)\mathbf{\Phi}_1(\tau_e), \quad (5)$$

where $\alpha(t-\tau)$ designates the projections on $\mathbf{\Phi}_1(\tau_e)$ at that point in the iteration. For $t < \tau_1$, we use $\tau_1 = t$ in (4). The spinup of the iteration is fast, and we only remove the results from the first year to avoid initial instabilities.

The results are not sensitive to the precise value of τ_e (as long as it is longer than the ENSO time scale) and τ_1. An example of the optimal initial condition and resulting global mean SST after 2.5 years is given in Fig. 2. Note that although the first LIM mode strongly contributes to the optimal initial structure, the method is more general than LIM as it makes no assumption as to which modes represent the forced signal and could include a significant contribution of the second LIM mode, as illustrated by the projection of each LIM mode on the optimal initial structure (Fig. 2, bottom).

g. *Multidimensional ensemble empirical mode decomposition*

EMD and EEMD (e.g., Wu and Huang 2009; Z. Wu et al. 2011) use amplitude–frequency-modulated oscillatory components of increasing time scale obtained adaptively from an anomaly time series $x(t)$, which is decomposed into a finite and often small number of components $c_j(t)$, such that

$$x(t) = \sum_{j=1}^{n} c_j(t) + R_n(t), \quad (6)$$

where $R_n(t)$ is the residual after n intrinsic modes have been extracted. In EMD, each oscillatory component is obtained through a sifting process: the local maxima and minima of $x(t)$ are first found, and the upper envelope $e_u(t)$ and the lower envelope $e_l(t)$ obtained by connecting them via a cubic spline. The local mean $m(t)$ of the two envelopes for the riding wave is then calculated. If it is close enough to zero (the upper and lower envelopes are symmetric with respect to the zero line), the sifting is terminated. Otherwise, $m(t)$ is subtracted from $x(t)$ and the sifting process is repeated until $m(t)$ is close enough to zero. This yields the highest-frequency oscillatory component $c_1(t)$ [the first riding wave component or intrinsic mode function (IMF), which has the shortest time scale]. Subtracting it from $x(t)$ leads to a residual $R_1(t)$, which is treated similarly, yielding a second IMF $c_2(t)$ and a residual $R_2(t)$. The decomposition process stops when the remaining time series $R_n(t)$ is a monotonic function or a function that has at most one internal extremum. As the IMFs are sensitive to noise, more robust estimates are obtained (EEMD method) by adding multiple white-noise realizations to $x(t)$, decomposing the data in each case, and using the ensemble means of the respective IMFs as the final result. The multidimensional generalization of EEMD (i.e., MEEMD) simply pieces together the EEMD components of similar time scale. This allows for both static and spatially propagating signals of naturally determined time scale, at least if there is a clear correspondence between the IMFs at the different grid points. Here we work with annual mean SST anomalies at each grid point, without binomial smoothing. The noise added to each input time series has a standard deviation 5 times smaller than the original time series, and the ensemble number is 400, as in Ji et al. (2014). It is not a priori obvious which residual best represents the forced climate signal. The last

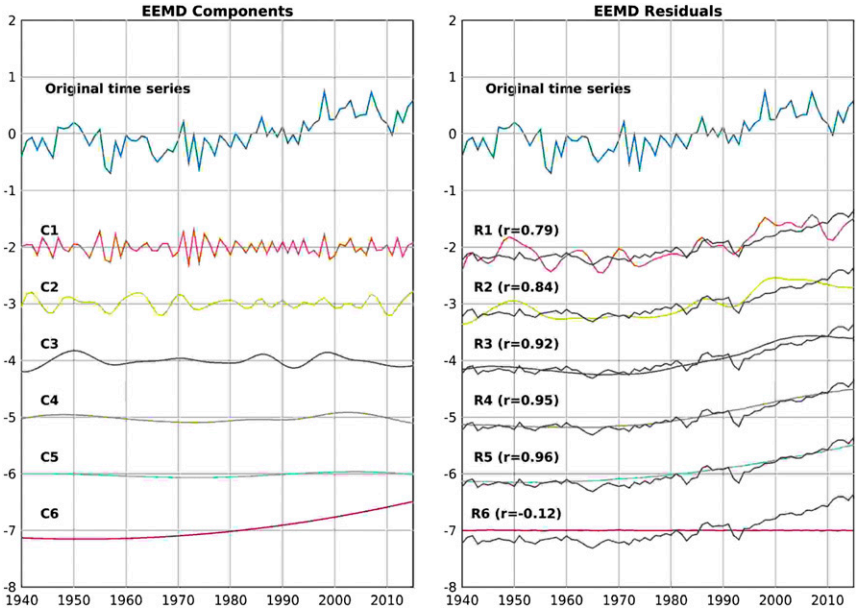

FIG. 3. EEMD decomposition of the annual SST anomaly at 35°N, 60°W for CESM-LE member 10. Each EEMD component $C_n(t)$ and residual $R_n(t)$ are shifted in the y axis for visibility. Units of the y axes are in degrees Celsius. The gray curve is the ensemble-mean SST; its correlation with each residual is indicated.

residual $R_n(t)$ was chosen in Z. Wu et al. (2011) and Ji et al. (2014). This seems to hold at some locations, as in Fig. 3, where $R_5(t)$, practically a quadratic trend, is highly correlated with the ensemble-mean SST, but not at others. It was found that in the two LEs, MEEMD fairs poorly for $R_1(t)$ and $R_2(t)$, better for $R_3(t)$, and best for $R_4(t)$, although $R_5(t)$ works nearly as well. Hence, in the following we use $R_4(t)$. However, using $R_4(t)$ in the SST reconstructions leads to very different AMO evolutions than the other methods, but not $R_5(t)$, which is then preferred. Because the EEMD is applied to each grid point separately, the spatial coherence in MEEMD is not always very good.

3. Datasets

Two LEs are considered. The first ensemble was conducted at Laboratoire d'Océanographie et du Climat: Expérimentation et Approches Numériques (LOCEAN) with the IPSL-CM5A-LR (IPSL-LE) at about 2° ocean and atmosphere resolution (Dufresne et al. 2013), and it contains 30 members for the period 1940–2020. The members start from random initial oceanic and atmospheric conditions selected between 1920 and 1960 in six CMIP5 historical runs. The second ensemble was conducted at the National Center for Atmospheric Research (NCAR) with the CESM1 (CESM-LE) at about 1° ocean and atmosphere resolution (Kay et al. 2015), and it contains 40 members for the period 1920–2100 starting from different atmospheric initial conditions but the same oceanic ones. All IPSL-LE and CESM-LE ensemble members have the same specified external forcing, following the CMIP5 design protocol: historical forcing until 2005, then representative concentration pathway 8.5 (RCP8.5). Details are given in Kay et al. (2015), who showed that the observed global warming anomaly was within the spread predicted by the CESM-LE. Because the oceanic initial conditions were identical in CESM-LE, we do not consider the period prior to 1940 to better distinguish forced and natural variability. Indeed, the AMOC, which largely contributes to the oceanic meridional heat transport, was only uncorrelated between members after the initial 20 years.

Our analysis focuses on monthly SST anomalies during 1940–2015 between 60°S and 60°N extracted from the surface temperature where the land fraction is smaller than 0.3, the SST larger than −2°C, and, in CESM-LE, the monthly sea ice concentration always smaller than 5%. In IPSL-LE, as there is too much sea ice during the cold season (Dufresne et al. 2013), we consider grid points where the sea ice climatology does not exceed 10%. As illustrated by the time evolution of the ensemble mean of the monthly, area-weighted, quasi-globally averaged SST in Fig. 4, the IPSL model shows a much larger sensitivity to the external forcing

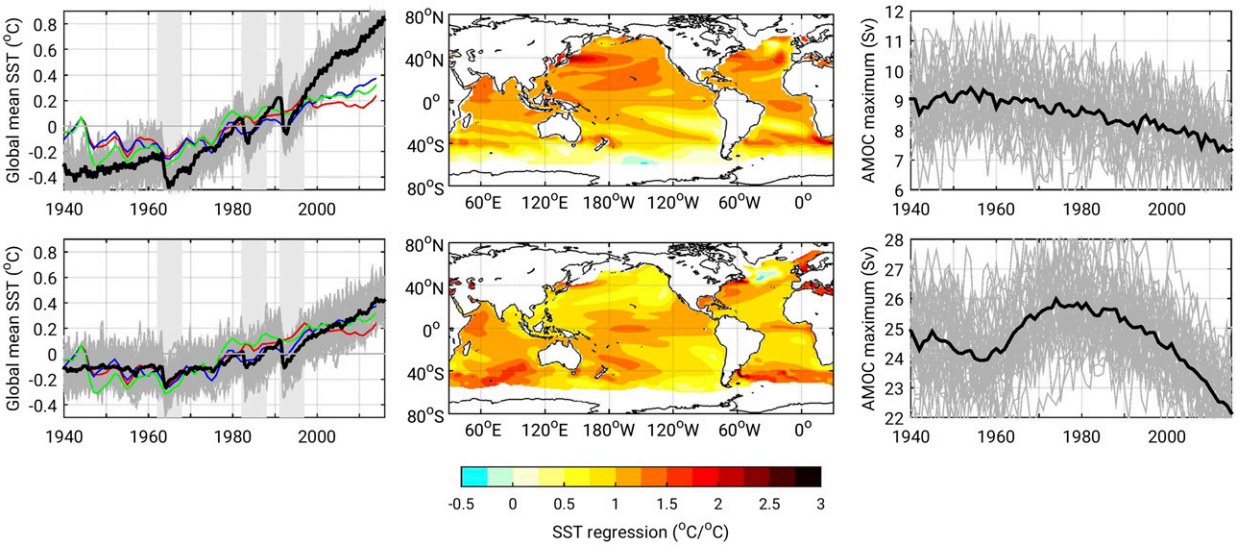

FIG. 4. (left) Area-weighted global mean SST anomaly (°C) for (top) IPSL-LE and (bottom) CESM-LE. The black curve is for the ensemble mean and the gray curves are for each ensemble member. Shading indicates the periods selected for investigating the response to volcanic eruptions. The red, blue, and green curves are the LIMopt estimates for the HadISST, COBE-SST, and ERSST observational reconstructions, respectively. (center) Regression of the ensemble-mean SST on the ensemble global mean SST (°C °C^{-1}). (right) Mean maximum AMOC at 40°N (Sv; 1 Sv $\equiv 10^6$ m^3 s^{-1}).

than CESM1, but the sensitivity of CESM1 is in better agreement with our best estimate of the forced SST signal in the HadISST (red), COBE-SST (blue), and ERSST (green) observational SST reconstructions (taken from Fig. 15). This is consistent with the CMIP5 model study of Andrews et al. (2012), where IPSL-CM5A-LR shows a rather large climate sensitivity, and CESM1 a rather low one. Nonetheless, both models represent the acceleration of the quasi-global mean SST increase, except for a sudden cooling after the three large volcanic eruptions, namely Mount Agung in 1963, El Chichón in 1982, and Mount Pinatubo in 1991. In the two LEs, there is no clear indication of the global warming slowdown from the late 1990s to 2012 that is seen in the HadISST reconstruction, but not in COBE-SST and ERSST (see section 6). There is much scatter between ensemble members, in particular in CESM-LE (gray lines in Fig. 4). The spatial pattern of the global mean SST changes, as obtained by regression, also varies, with a larger equatorial signal in CESM-LE, but larger mid-latitude signals in IPSL-LE (Fig. 4).

For the observations, we consider the 1900–2015 period in three recent SST reconstructions, namely HadISST, version 1.1 (HadISST1.1), on a 1° × 1° grid (Rayner et al. 2003), ERSST, version 4 (ERSST.v4), on a 2° × 2° grid (Huang et al. 2015), and COBE-SST2 on a 1° × 1° grid (Hirahara et al. 2014). The mask using the same criteria as in CESM-LE is used to select open ocean grid points, using the sea ice dataset associated to each SST reconstruction, except for ERSST.v4 where we use the HadISST sea ice mask.

4. Externally forced signal

A synthetic view of method efficiency is provided by comparing the evolution of the area-weighted global mean forced SST during 1940–2015 in each ensemble member against the target truth, i.e., the global mean SST of the ensemble average given in Fig. 4. Because the global mean yearly SST time series show a near monotonous increase over time, except after volcanic eruptions, the correlation r with the ensemble average is very high and basically undistinguishable between methods, except for D1. Indeed, the averaged correlation for the other methods ranges between 0.98 and 0.99 in IPSL-LE, and between 0.94 and 0.97 in CESM-LE, while D1 is clearly inferior ($r = 0.93$ in IPSL-LE and 0.86 in CESM-LE). A more selective comparison is provided by the differences between the estimated yearly forced global mean SST time series in each ensemble member and the global mean SST of the ensemble average, which eliminates the trend that dominates the correlation. To compare the methods, we average over ensemble members (30 or 40) the root-mean-square (rms) difference for the 76 years. By the central limit theorem, this statistic has to a good approximation a Gaussian distribution. As shown in Table 1, LIM and then LIMopt yield the smallest errors, except in CESM-LE where

TABLE 1. The rms of the global mean SST error (10^{-2} K) and 95% confidence interval during 1940–2015.

	D1	D2	GM	REGR	LIM	LIMopt	MEEMD
IPSL-LE	13.27 ± 0.08	6.90 ± 0.41	6.07 ± 0.34	6.08 ± 0.34	5.54 ± 0.41	5.63 ± 0.42	7.12 ± 0.23
CESM-LE	8.57 ± 0.04	4.91 ± 0.32	6.05 ± 0.23	6.04 ± 0.23	4.44 ± 0.25	5.16 ± 0.34	4.26 ± 0.17

MEEMD and D2 fare slightly better or comparably, presumably because the global mean SST approximately increases quadratically with time in this model (Fig. 4). GM and REGR lead to larger errors, albeit only significantly so in CESM-LE, while D1 is clearly inadequate.

The geographical distribution of the averaged (over time and ensemble members) rms differences suggests that regions with a large natural variability, such as near the equator and at high latitude, dominate the global errors for all methods, as illustrated for IPSL-LE (Fig. 5). These regions also dominate their area-weighted spatial average in each ensemble member, whose histogram is given in Fig. 6, although similar results are found with normalized SST data. In IPSL-LE (Figs. 6a–g), the averaged rms error is smallest for LIM and LIMopt, then REGR, and largest for GM and D1. Based on the two-sample Student's *t* test, the errors are smallest in LIM and LIMopt at the 1% significance level, but they cannot be differentiated. In CESM-LE (Figs. 6h–n), the estimation errors and the differences among methods are smaller, although LIM-opt, D2, MEEMD, and LIM are best (and not significantly different), and REGR and D1 yield larger errors. Hence, LIM and LIMopt are the only methods that lead to small errors in both LEs.

To investigate how each method represents short-term forced climate fluctuations, we separately consider the 7-yr periods that follow the three largest volcanic eruptions occurring in the 1940–2015 period, namely Mount Agung (1963), El Chichòn (1982), and Mount Pinatubo (1991). A global measure of method efficiency could be provided as above. However, the rms errors would have two contributions, one corresponding to the initial year of the volcanic eruption, and the other to the SST decrease in the following six years. Hence, they

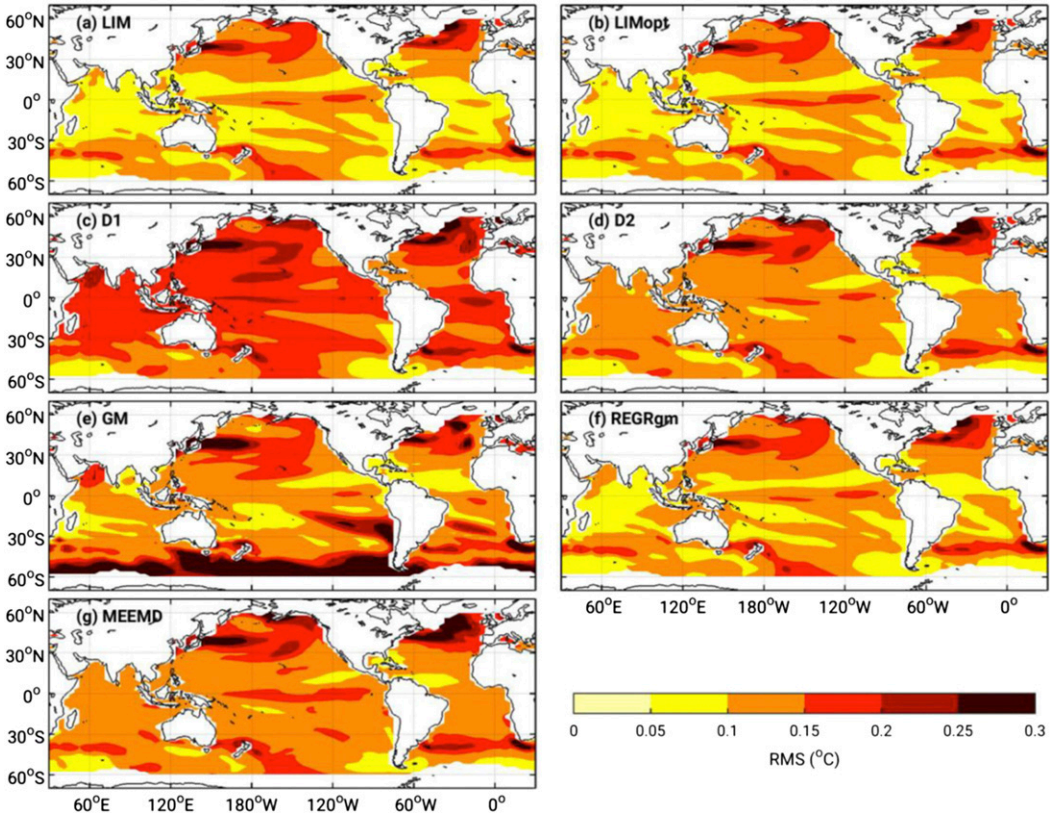

FIG. 5. Ensemble-mean rms difference (°C) between the estimated forced SST signal in each member of IPSL-LE and the ensemble-mean SST in 1940–2015 for the different methods.

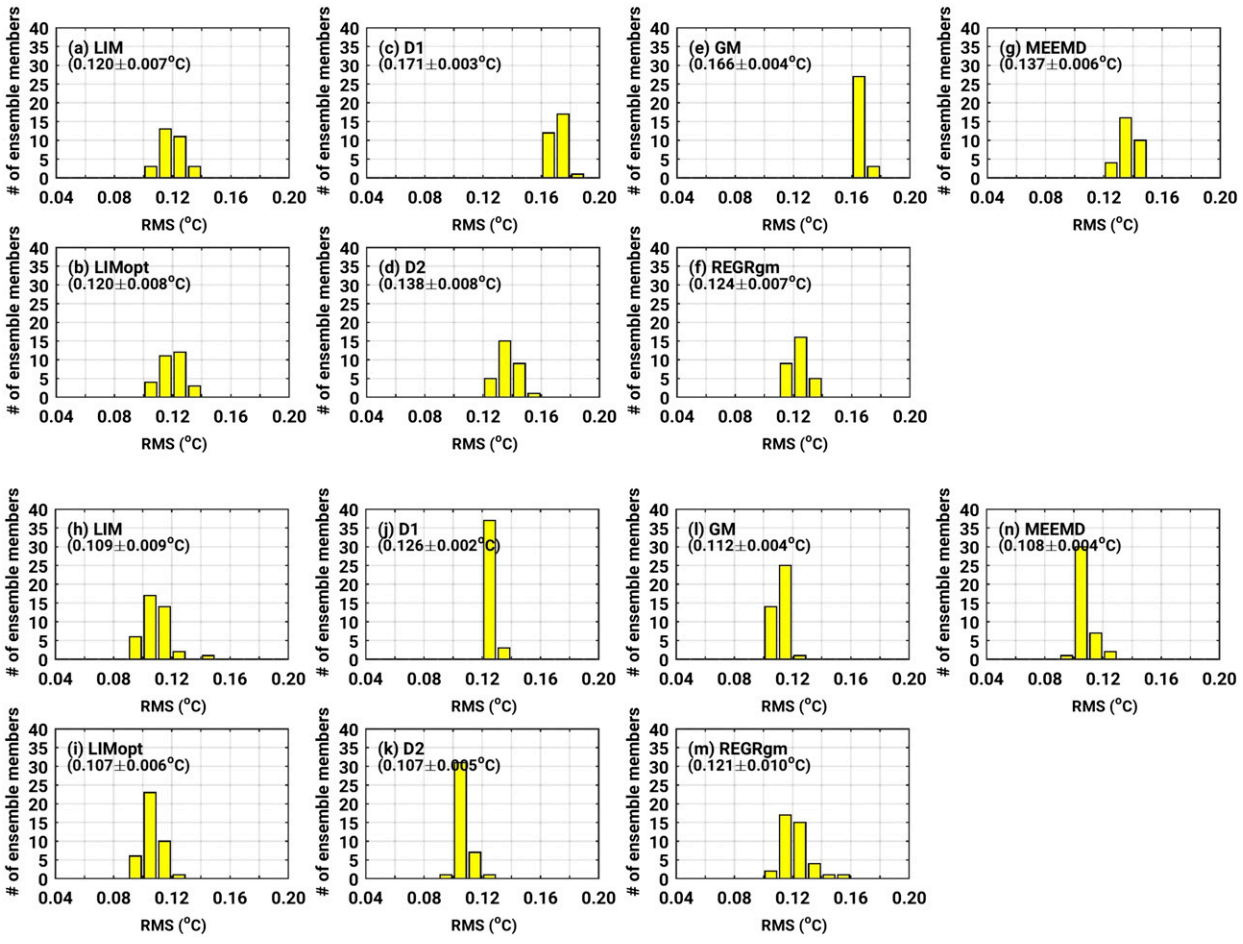

FIG. 6. Histogram of the area-averaged rms difference (°C) between the forced SST in each ensemble member and the ensemble-mean SST in 1940–2015, as given by various methods in (a)–(g) IPSL-LE and (h)–(n) CESM-LE. The mean rms and its standard deviation are indicated.

would not provide an optimal measure of method ability at representing fast changes. Indeed, if a method, say, underestimates the global mean SST in the initial year, it may yield a small error even if the actual volcanic cooling is underestimated. Therefore, we compare instead the SST evolution following the year of each volcanic event in each ensemble member to that of the ensemble mean (Table 2). This is more independent from the statistics given in Table 1, which already include the volcanic periods. In IPSL-LE, the area-weighted mean rms error during the three volcanic periods is smallest in GM, LIMopt, REGR, and LIM, while it is much larger in methods that emphasize the most slowly varying components of the SST evolution (i.e., MEEMD, D1, and D2). In CESM-LE, LIMopt and LIM yield much smaller errors than MEEMD, GM, and REGR, while D1 and D2 again lead to large errors.

5. Natural variability after removing externally forced signals

a. *Atlantic multidecadal oscillation*

The AMO time series in each ensemble member is defined by the low-pass-filtered, area-weighted SST

TABLE 2. Mean rms error (10^{-2} K) and 95% confidence interval for the global mean SST decrease in the six years following the three major volcanic events in IPSL-LE and CESM-LE.

	D1	D2	GM	REGR	LIM	LIMopt	MEEMD
IPSL-LE	16.9 ± 0.11	17.4 ± 0.12	7.14 ± 0.69	7.19 ± 0.70	7.24 ± 0.70	7.20 ± 0.65	13.86 ± 0.11
CESM-LE	9.07 ± 0.06	9.55 ± 0.06	7.78 ± 0.50	7.87 ± 0.51	5.03 ± 0.43	4.88 ± 0.39	7.62 ± 0.05

anomaly average within 0°–60°N, 0°–80°W after subtraction of the forced SST signal, using yearly SST and a 10-yr cutoff (Butterworth filter), with the land and sea ice mask described above. The AMO pattern in each ensemble member is obtained by regressing the yearly mean SST anomalies onto the AMO time series. We assume that the true unforced AMO pattern is given to a good approximation by the average of the patterns obtained in each ensemble member after subtraction of the ensemble-mean SST. Note that there is some scatter in the AMO pattern between ensemble members, reflecting that the record length (76 yr) is too short to well resolve the long AMO time scale. The same strategy is used for each method, except that it is the estimated forced signal that is subtracted in each ensemble member before estimating the AMO.

Method efficiency at representing the natural variability of the AMO is evaluated in two different ways. First, we consider the ensemble average of the AMO time series, which should by construction be, within error bars, equal to zero throughout 1940–2015. The comparison is shown in Figs. 7a,b, where the 95% confidence interval for the ''true'' AMO, as estimated by removing the ensemble mean in each ensemble member, is indicated. For clarity, no confidence interval is given for each method, as it is likely to be similar to, but not independent of, that of the true mean. The poor performance of D1 is striking in both ensembles. In IPSL-LE and, to a lesser extent CESM-LE, MEEMD and D2 are unable to separate the global cooling that follows the two main volcanic eruptions (in the 1960s and the 1990s) from the natural variability of the AMO, yet they lead to the best averaged correlation with the true AMO in CESM-LE (Figs. 7b,d,f,h), as other methods overestimate the AMO in the first decade. In both LEs, LIM and LIMopt lead to small errors. Although GM fares well in the volcanic periods, it poorly separates the forced SST changes from the AMO during the last decade. Indeed, as climate models (and observations) show a minimum warming in the subpolar North Atlantic (e.g., Drijfhout et al. 2012; see also Fig. 4), the removal of a spatially uniform pattern leads to erroneous cold AMO anomalies after 2000, when the external forcing from greenhouse gases has increased. Nonetheless, GM works well in IPSL-LE, and REGR fares nicely in both LEs.

Another comparison is given in Figs. 8 and 9, where the mean (over all ensemble members) AMO pattern estimated by each method is compared to that derived by subtracting the ensemble mean from each ensemble member. The true AMO patterns are fairly realistic (cf. with observational AMO patterns in Figs. 16a,b), although the tropical part of the AMO is shifted northward in IPSL-LE (Gastineau et al. 2013), and in both models there is a weak ENSO-like warming in the Pacific and, in particular in CESM-LE, a PDO-like signal in the North Pacific. This presumably occurs because ENSO teleconnections contribute to driving both the AMO and the positive phase of the PDO in the two models. While the AMO was also shown to drive a negative PDO in sensitivity experiments with CESM1 (Ruprich-Robert et al. 2017), it would lead to a PDO of the opposite sign if it were the dominant interaction. These AMO patterns are almost identical to those derived from 1500-yr control simulations (not shown). The true AMO pattern is generally well represented by each method in the North Atlantic, except D1, but there are larger differences in the Pacific, as GM and REGR do not reproduce the Pacific extensions. To compare the methods more quantitatively, Table 3 gives the mean of the rms differences between the AMO pattern in each member/method and the corresponding true AMO pattern, with the pattern correlation, which takes into account the scatter between ensemble members. In IPSL-LE, LIMopt provides the best results, but they are not significantly different from LIM, D2, and MEEMD, and only significantly better than REGR, GM, and D1. In CESM-LE, D2 and MEEMD are both significantly better than LIM and LIMopt, while GM and REGR fare poorly for the global pattern and D1 provide large errors both in the North Atlantic and globally.

b. *AMOC SST signature*

To determine the SST fingerprint of the AMOC, we define the AMOC by the maximum of the yearly AMOC streamfunction between 500 and 2500 m averaged between 35° and 45°N (hereafter AMOC40), and we remove its ensemble mean, since in both models the AMOC has substantially decreased by 2015 (Fig. 4). In IPSL-LE, the mean AMOC decreases almost linearly after the mid-1950s. In CESM-LE, the mean AMOC, which is much stronger, slightly increases in the 1960s and only decreases after the mid-1980s, reflecting a possible influence of anthropogenic aerosols (Tandon and Kushner 2015). There is much scatter between ensemble members, however, reflecting the large natural variability of the AMOC.

Although the SST fingerprint of the natural variability of the AMOC evolves with time lag, as shown by Gastineau et al. (2013) for IPSL-CM5 and Frankignoul et al. (2015) for an earlier version of CESM1 (CCSM4), a typical low-frequency footprint can be obtained by regressing the low-pass-filtered SST onto the low-pass-filtered AMOC40, after removing the global forced signal, at the lag of maximum AMOC40–AMO correlation. This was done by concatenating all ensemble

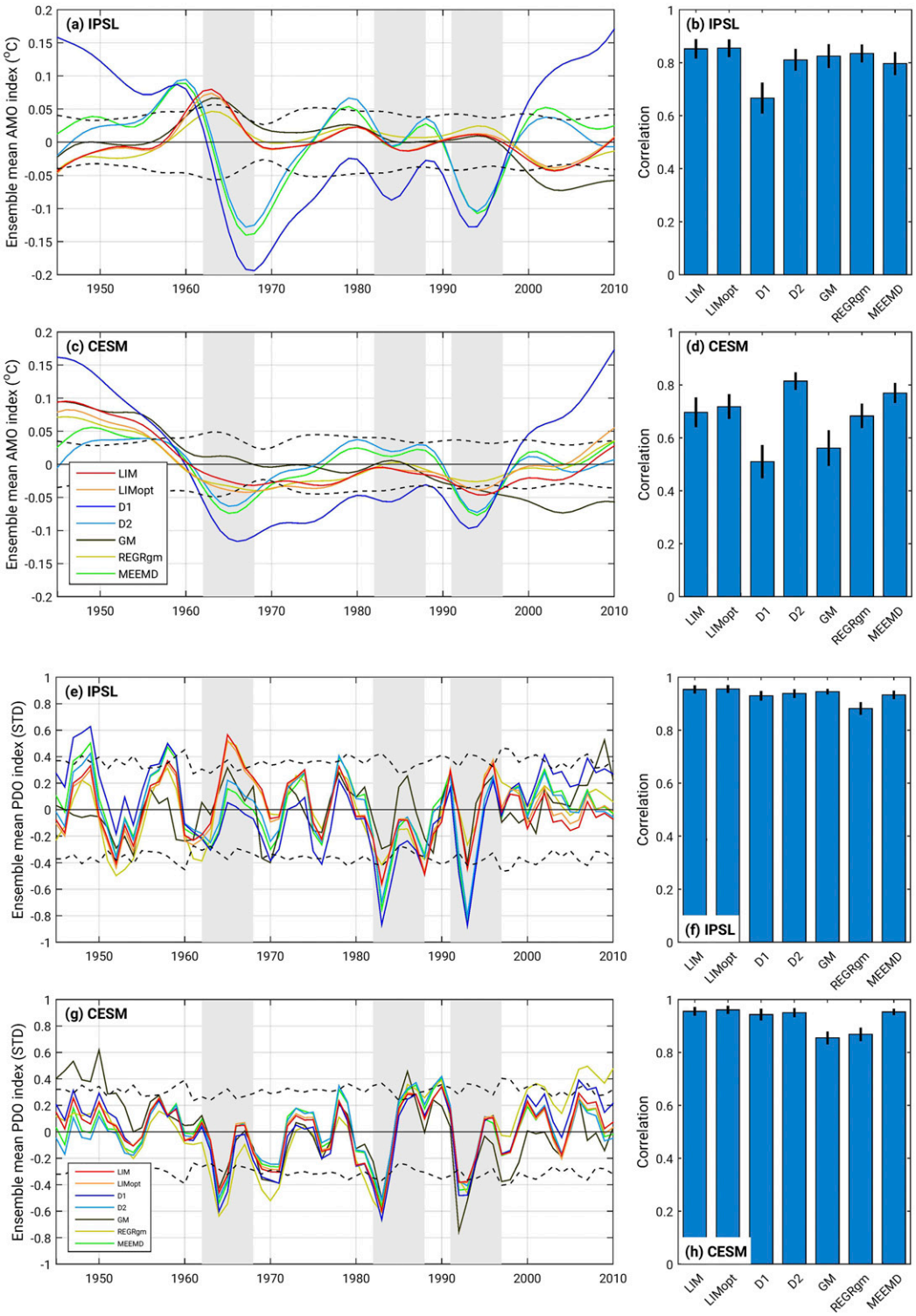

FIG. 7. Ensemble mean of the (a),(c) AMO and (e),(g) PDO natural variability for IPSL-LE in (a),(e) and for CESM-LE in (c),(g), as determined after removing the forced signal by each method. The 95% confidence interval estimated by removing the ensemble mean in each ensemble member is indicated by the dashed lines. The gray shading indicates the seven years following major eruptions. Mean correlation between the (b),(d) AMO and (f),(h) PDO time series estimated by each method and that obtained by removing the ensemble mean in each ensemble member, with 95% confidence interval.

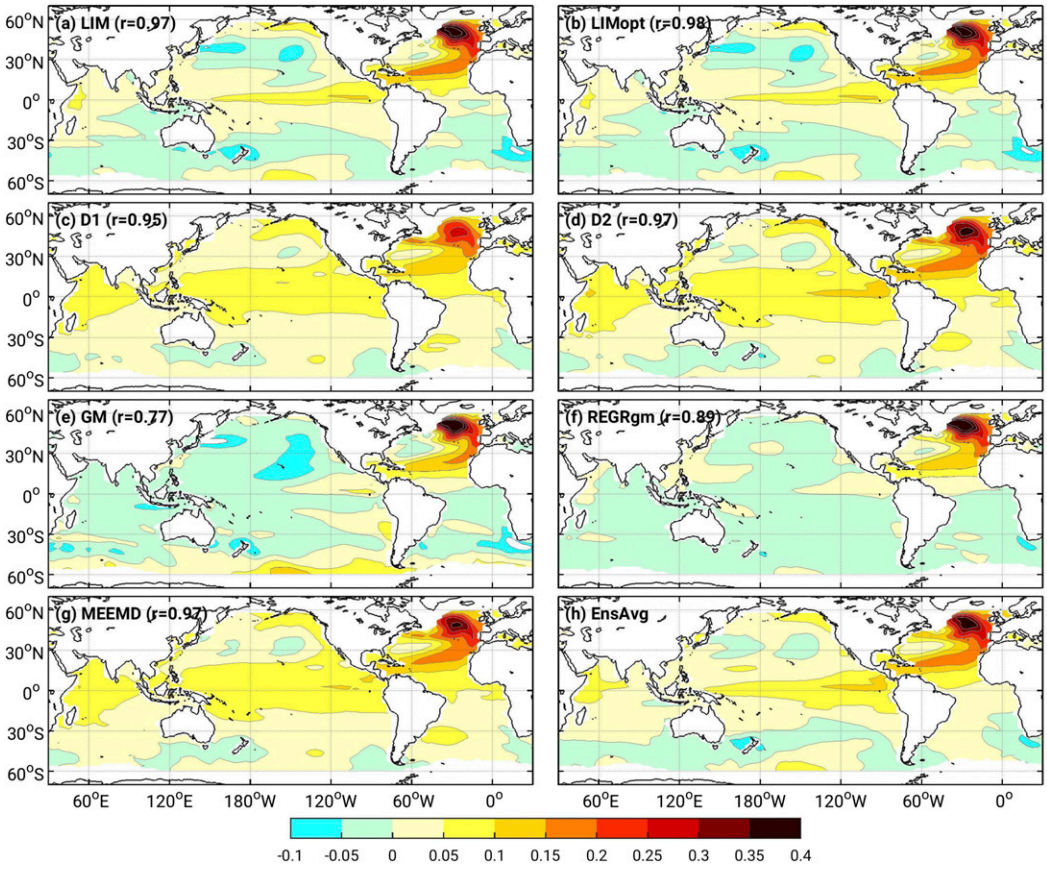

FIG. 8. Mean AMO pattern (K) in IPSL-LE after the externally forced signal estimated by the different methods has been removed from each ensemble member. The regression is calculated on the standardized AMO time series, so the patterns give typical amplitudes. The spatial correlation with the global ensemble average (EnsAvg) AMO pattern (the true pattern) is given.

members into a single realization, with the SST lagging the AMOC by 9 yr for IPSL-LE and 1 yr for CESM-LE (see Fig. 12). The best estimate of the AMOC SST fingerprint, obtained by removing the ensemble-mean SST signal (the true fingerprint), is represented in Figs. 10h and 11h. In both models, there is a strong warming in the subpolar gyre and a cooling along the Gulf Stream, which likely reflects the southward shift of the Gulf Stream due to bottom torque at the crossover of the deep western branch of the AMOC with the Gulf Stream, as in CCSM4 (Frankignoul et al. 2015). There is also a weak cooling in the South Atlantic.

The comparison with the true AMO pattern in Figs. 8 and 9 shows that the AMOC SST fingerprint broadly resembles the AMO north of about 40°N in both models, but differs elsewhere. In both models, the AMOC SST fingerprint lacks the subtropical/tropical North Atlantic warming and the equatorial Pacific warming associated with the AMO, while there is no South Atlantic cooling in the AMO. The AMO also lacks the Gulf Stream cooling seen in the AMOC fingerprint. This suggests that in these two models the AMO, south of about 40°N, is primarily driven by local atmospheric variability and ENSO teleconnections, or by oceanic variability unrelated to the AMOC. Note that the equatorial Pacific warming associated with the AMO is not seen in the observations (see Figs. 16a,b), which suggests that the tropical Pacific influence is perhaps too strong in these models. The AMOC SST fingerprint is generally well reproduced when the forced SST signal is estimated in each ensemble member, although MEEMD fares less well in CESM-LE (Figs. 10 and 11).

The best (true) estimate of the lag correlation between the (low-pass filtered) AMO and AMOC40 is obtained by subtracting the ensemble-mean SST before computing the AMO. Figure 12 shows that the maximum mean correlation occurs when the AMOC leads the AMO by about 9 yr in IPSL-LE, consistent with Gastineau et al. (2013) and Marini and Frankignoul (2014), and by 1 or 2 yr in CESM-LE, as in CCSM4

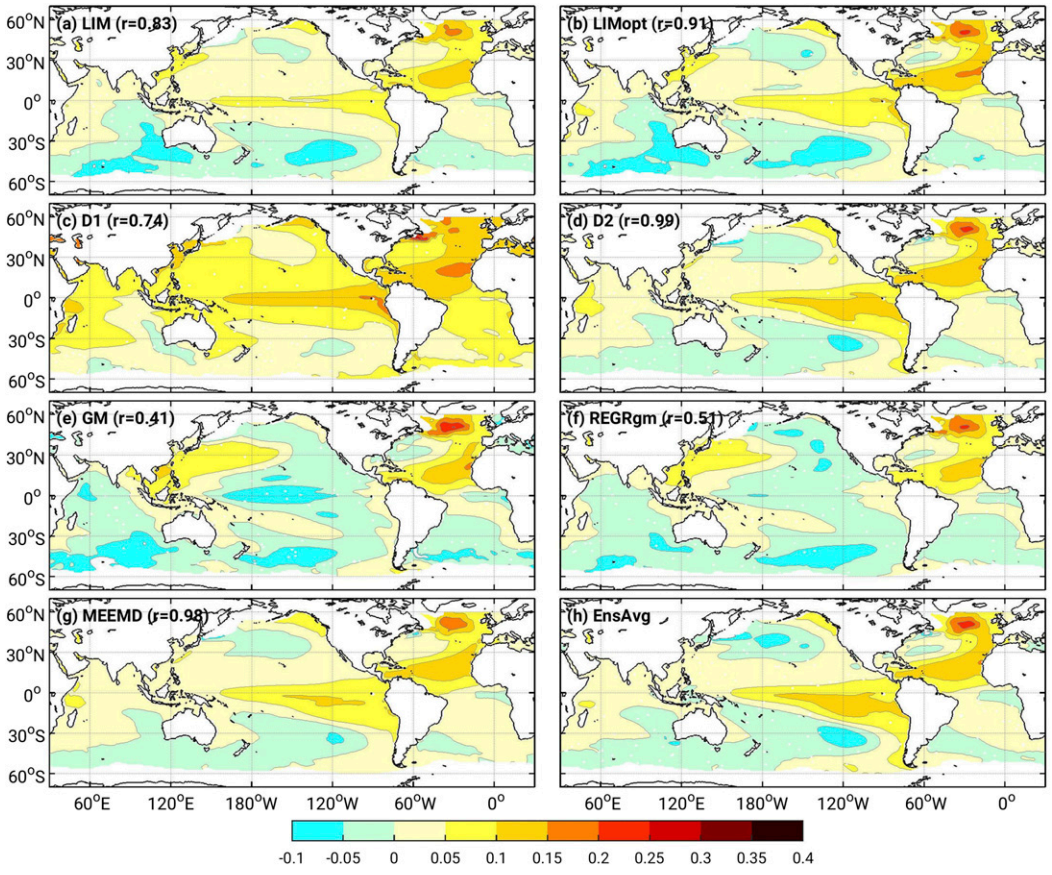

FIG. 9. As in Fig. 8, but for CESM-LE.

(Danabasoglu et al. 2012). The maximum correlation in IPSL-LE ($r = 0.26$) is much smaller than in CESM-LE ($r = 0.40$). There is also a marginally significant hint that a negative AMO precedes an intensification of the AMOC by 10–16 yr in CESM-LE. Yet, there is much scatter among ensembles members, as illustrated by the dispersion of the lag and the value of the largest correlation (blue dots in Fig. 12). A 76-yr record is thus too short to determine with some confidence the relation between the AMOC and the AMO, and the AMOC–AMO correlation is not a powerful criterion to distinguish between methods of forced signal removal. In IPSL-LE, the mean (over ensemble members) AMOC–AMO correlation depends little on the way the forced signal is removed, although D1, D2, and MEEMD yield a smaller maximum; in CESM-LE, the mean AMOC–AMO correlation is generally well reproduced, although the maximum correlation is systematically too

TABLE 3. Mean and its standard error of the rms differences between the AMO pattern from each member/method and that from the corresponding member of EnsAvg (10^{-2} K) in the North Atlantic and the global domain for (left) IPSL-LE and (right) CESM-LE. The mean pattern correlation and its standard error are in parentheses.

	IPSL-LE		CESM-LE	
Method	North Atlantic	Global	North Atlantic	Global
D1	7.86 ± 0.37 (0.74 ± 0.04)	5.61 ± 0.19 (0.75 ± 0.03)	7.58 ± 0.21 (0.44 ± 0.05)	7.70 ± 0.23 (0.55 ± 0. 03)
D2	6.07 ± 0.37 (0.85 ± 0.04)	5.05 ± 0.23 (0.83 ± 0.03)	3.40 ± 0.17 (0.86 ± 0.02)	3.63 ± 0.14 (0.84 ± 0.02)
GM	7.30 ± 0.40 (0.86 ± 0.02)	7.03 ± 0.35 (0.73 ± 0.04)	4.96 ± 0.22 (0.74 ± 0.03)	7.53 ± 0.29 (0.47 ± 0.04)
REGR	6.47 ± 0.40 (0.88 ± 0.02)	5.73 ± 0.22 (0.83 ± 0.02)	4.61 ± 0.20 (0.75 ± 0.03)	6.46 ± 0.29 (0.58 ± 0.04)
LIM	6.15 ± 0.43 (0.87 ± 0.02)	4.83 ± 0.20 (0.87 ± 0.01)	4.99 ± 0.25 (0.68 ± 0.04)	5.46 ± 0.25 (0.66 ± 0.04)
LIMopt	5.90 ± 0.41 (0.88 ± 0.02)	4.75 ± 0.19 (0.87 ± 0.01)	4.39 ± 0.22 (0.78 ± 0.03)	4.94 ± 0.21 (0.74 ± 0.03)
MEEMD	6.26 ± 0.37 (0.82 ± 0.04)	5.26 ± 0.22 (0.80 ± 0.03)	3.74 ± 0.19 (0.83 ± 0.03)	4.05 ± 0.15 (0.80 ± 0.02)

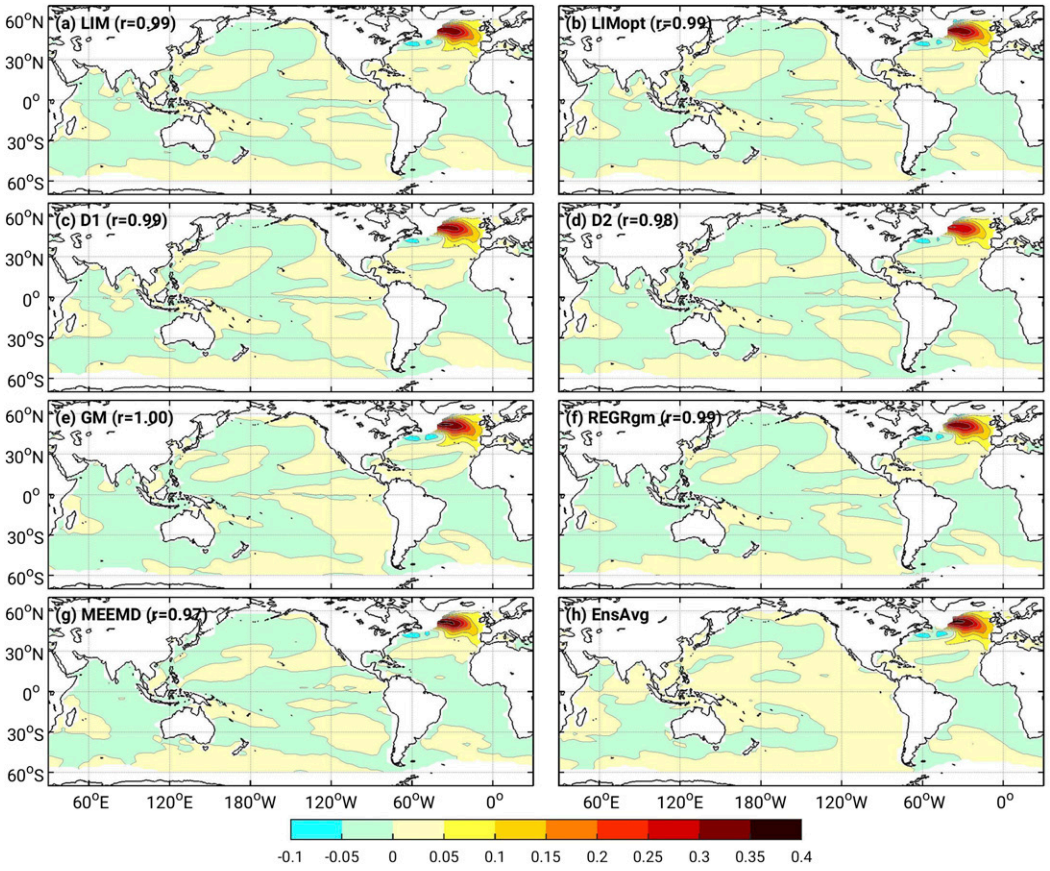

FIG. 10. AMOC fingerprint (K) in IPSL-LE as determined by each method after removal of the estimated forced SST signal, and so-called true fingerprint obtained by using the ensemble average SST as forced signal (EnsAvg). The spatial correlation with the global EnsAvg AMO pattern is given.

small (not shown). Similar results are obtained if AMOC40 is replaced by the averaged meridional heat transport between 35° and 45°N, or if the AMOC variability is represented by its first principal component (not shown).

c. *The Pacific decadal oscillation*

The PDO is defined as the first EOF of yearly averaged North Pacific (20°–60°N) SST variability computed in each ensemble member after removing the forced SST signal. The best (true) estimate of the PDO pattern (extended globally by regression) is obtained by removing the ensemble average SST from each ensemble member (Figs. 13h and 14h). As the mean PDO patterns estimated by each method look pretty similar, we show for comparison their difference with the true PDO (note the change of scale). In both models, the best estimates of the PDO pattern are obtained with LIMopt and LIM (Table 4). MEEMD works well, but D1 leads to a weak cold bias and D2 to much variability among ensemble members in IPSL-LE, although the mean pattern is well estimated. More importantly, REGR substantially underestimates the PDO amplitude, perhaps because PDO and global mean SST are related (Meehl et al. 2013), and GM yields large errors. The comparison between the rms errors in Tables 3 and 4 shows that the global PDO pattern tends to be better estimated than the AMO pattern when the best methods are selected, while the rms errors are otherwise comparable. In fact, the largest rms errors are found for the PDO when using GM or REGR.

Method efficiency at representing the natural variability of the PDO is also evaluated by considering the ensemble-average of the PDO time series, which should by construction be, within error bars, equal to zero throughout 1940–2015. The comparison is shown in Fig. 7 (bottom two panels), where the 95% confidence interval for the ''true'' PDO, as estimated by removing the ensemble mean in each ensemble member, is indicated. The PDO time behavior in each ensemble member is best represented by LIM and LIMopt, although differences with D1, D2, and MEEMD are not

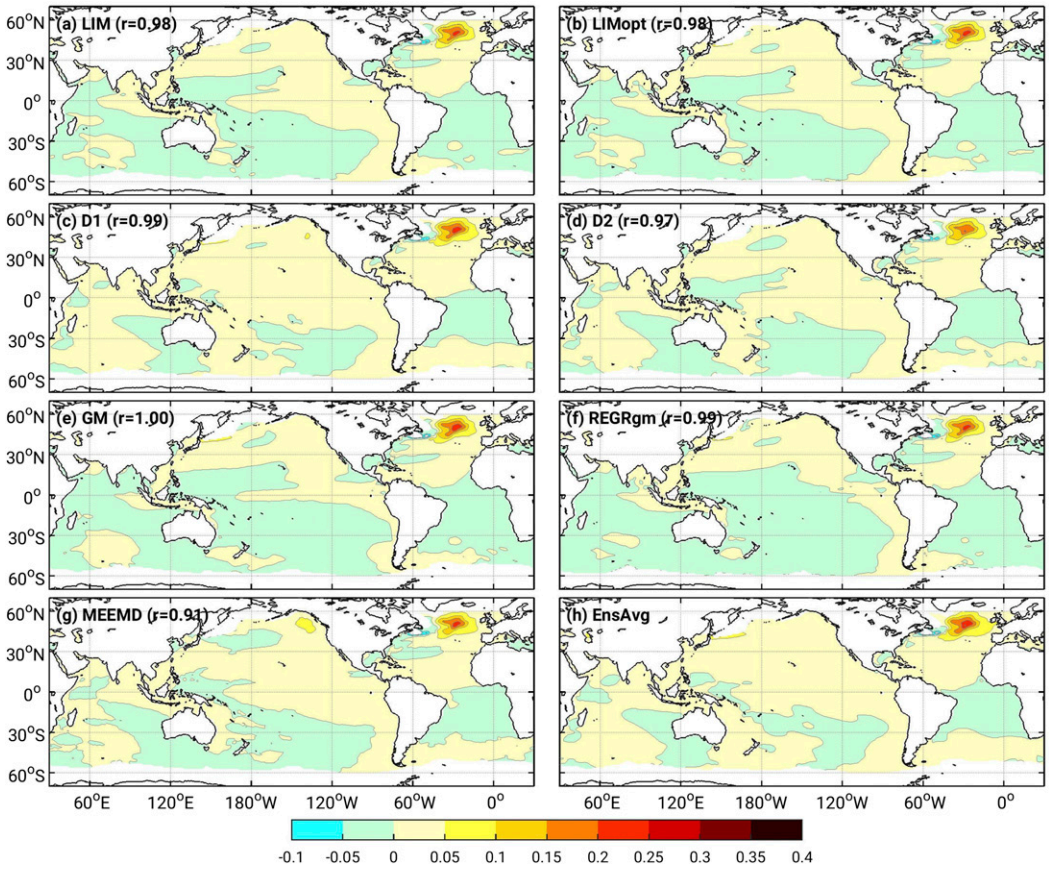

FIG. 11. As Fig. 10, but for in CESM-LE.

5% significant. On the other hand, REGR fares poorly in both LEs, and GM even worse in CESM-LE.

The lag relation between the AMO and the PDO is also considered, using 10-yr low-pass-filtered time series. Even in the true case, the mean lag correlation is not significantly different from zero in both models, and there is much scatter between ensemble members (not shown). Hence, it cannot be used for method validation.

d. *Method summary*

Although LIM and LIMopt do not always provide the smallest errors for the natural climate variability, they lead to consistently good estimates, while GM and

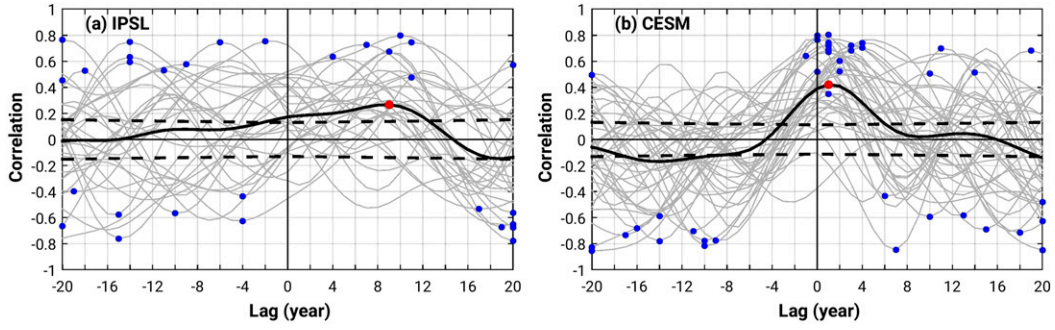

FIG. 12. Lag correlation between the AMO and AMOC40 in (a) IPSL-LE and (b) CESM-LE, after removal of the ensemble-mean signal and low-pass filtering. Gray curves are for each member and the black curves are their average. The dashed lines indicate the 95% confidence interval. The AMOC leads at positive lag (yr). The blue dots indicate the maximum absolute value of each gray curve and the red dot that of the ensemble mean.

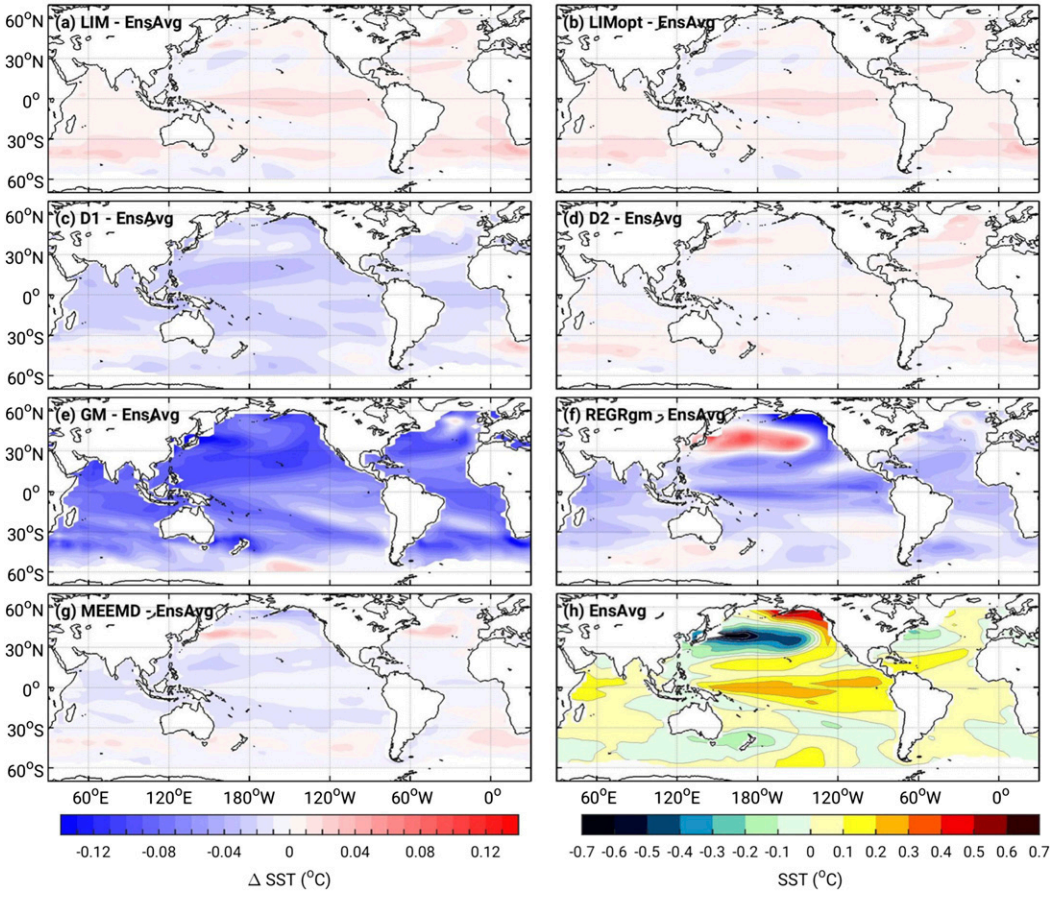

FIG. 13. (a)–(g) Difference between the mean PDO pattern estimated by each method and the true one (left color bar), and (h) best estimate of the true mean PDO pattern in IPSL-LE (°C; right color bar).

REGR may introduce substantial errors in the PDO and the AMO, and D1 often leads to poor estimates.

6. Application to the observations

Each method has been used to estimate and remove the forced signal in the 1900–2015 SST reconstructions. Each reconstruction yields a slightly different quasi-global (between 60°N and 60°S) mean SST evolution (given by GM) that reflects differences in bias correction, interpolation methods, and smoothing (e.g., Hirahara et al. 2014); the differences are at least comparable to those between estimation methods (Fig. 15). Note that the cooling following volcanic eruptions is smaller—or less visible—than in the LEs, although the cooling following the 1963 Mount Agung eruption is enhanced in the LIM and LIMopt estimates. The lack of warming trend between the mid-1940s and the mid-1970s, also seen in land surface air temperature (Hartmann et al. 2013) and sometimes referred to as big hiatus, likely reflects a compensation between the warming from increasing greenhouse gases and the cooling from anthropogenic sulfate aerosols (e.g., Fyfe et al. 2016), followed by the Mount Agung cooling. Consistent with the CMIP5 simulations and the CESM-LE, the natural variability (the difference between GM and the other methods) seems to play little role in the big hiatus, except that it ends up to a decade earlier when the natural variability is removed (i.e., in LIM and LIMopt). This is consistent with Kosaka and Xie (2016), who showed that the prolongation of the big hiatus was due to the change of phase of the interdecadal Pacific oscillation, which is well correlated with the PDO. The global warming slowdown found in global mean surface temperature from the late 1990s to 2012 is seen in the quasi-global mean SST from HadISST, but not COBE-SST or ERSST values (see also Karl et al. 2015). Our analysis suggests that the SST warming slowdown, if any, cannot be directly attributed to the natural variability of the PDO or the AMO, as there are only small differences between GM and LIM or LIMopt. This does not imply that the PDO did not contribute to the slowdown

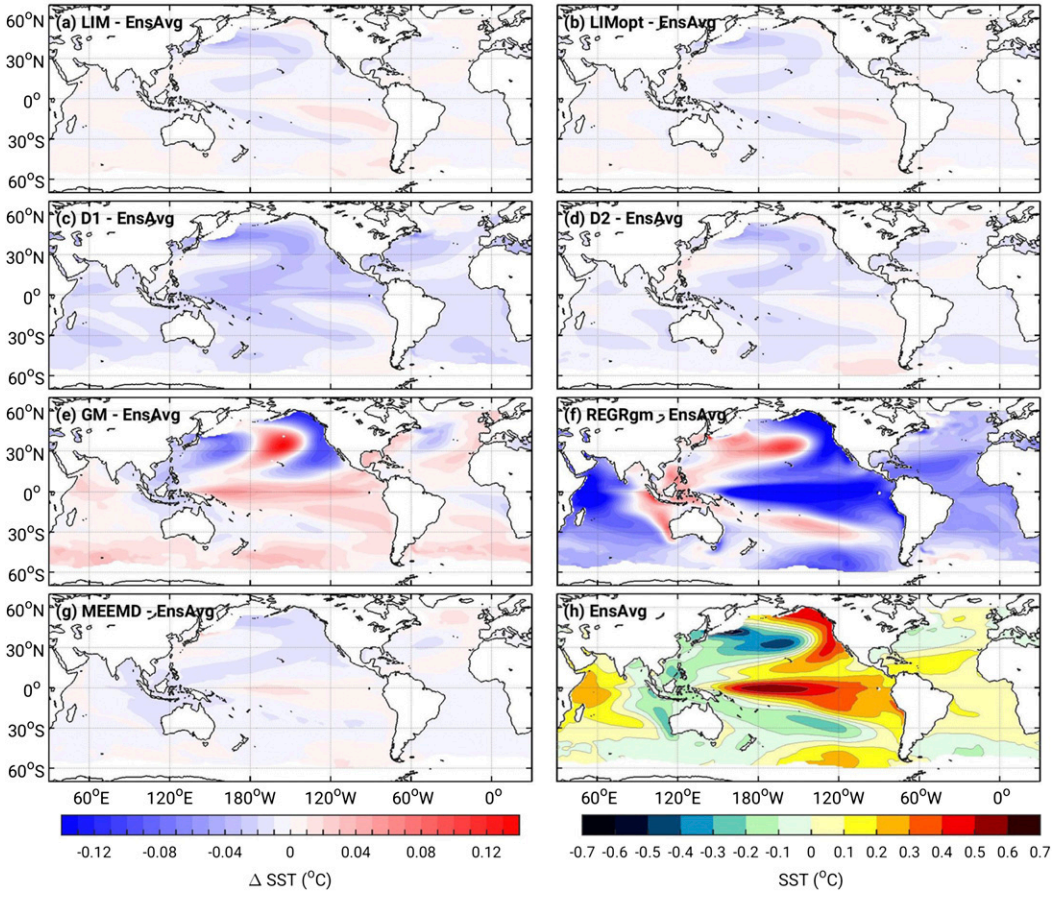

FIG. 14. As in Fig. 13, but for CESM-LE.

of the surface temperature warming, as suggested by Kosaka and Xie (2013, 2016) and England et al. (2014), since the interdecadal Pacific oscillation may have affected surface temperature at higher latitudes and over the continents.

However, the first LIM mode may not always provide a consistent estimate of the forced signal when it is applied to short time periods. This is illustrated for the 1940–2015 period considered in the LEs (dashed line in Fig. 15), as it takes the first two LIM modes to represent the forced signal and provide a similar behavior to LIMopt, in particular in HadISST where the two modes have a similar pattern but nearly opposite time variations (not shown). This reflects that over such short periods the time scale of the forced signal becomes comparable to that of the AMO; also, the forced signal

TABLE 4. Mean and its standard error of the rms differences between the PDO pattern from each member/method and that from the corresponding member of EnsAvg (10^{-2} K) in the North Pacific and the global domain for (left) IPSL-LE and (right) CESM-LE. The mean pattern correlation and its standard error (in parentheses) are shown.

	IPSL-LE		CESM-LE	
Method	North Pacific	Global	North Pacific	Global
D1	6.17 ± 0.57 (0.99 ± 0.00)	5.21 ± 0.42 (0.97 ± 0.00)	4.95 ± 0.49 (0.99 ± 0.01)	3.50 ± 0.23 (0.99 ± 0.00)
D2	5.65 ± 0.47 (0.99 ± 0.00)	4.61 ± 0.36 (0.98 ± 0.00)	4.08 ± 0.42 (0.99 ± 0.00)	2.92 ± 0.18 (0.99 ± 0.00)
GM	10.94 ± 1.19 (0.99 ± 0.00)	8.35 ± 0.92 (0.96 ± 0.01)	8.24 ± 0.46 (0.96 ± 0.01)	5.14 ± 0.26 (0.97 ± 0.01)
REGR	7.32 ± 0.32 (0.99 ± 0.00)	4.68 ± 0.15 (0.96 ± 0.00)	8.77 ± 0.45 (0.97 ± 0.01)	7.59 ± 0.26 (0.96 ± 0.00)
LIM	3.31 ± 0.24 (1.00 ± 0.00)	2.58 ± 0.11 (0.99 ± 0.00)	3.86 ± 0.37 (0.99 ± 0.00)	2.54 ± 0.17 (0.99 ± 0.00)
LIMopt	3.28 ± 0.22 (1.00 ± 0.00)	2.50 ± 0.09 (0.99 ± 0.00)	3.65 ± 0.35 (0.99 ± 0.00)	2.41 ± 0.17 (0.99 ± 0.00)
MEEMD	4.27 ± 0.21 (0.99 ± 0.00)	3.20 ± 0.13 (0.98 ± 0.00)	4.05 ± 0.26 (0.99 ± 0.00)	2.73 ± 0.13 (0.99 ± 0.00)

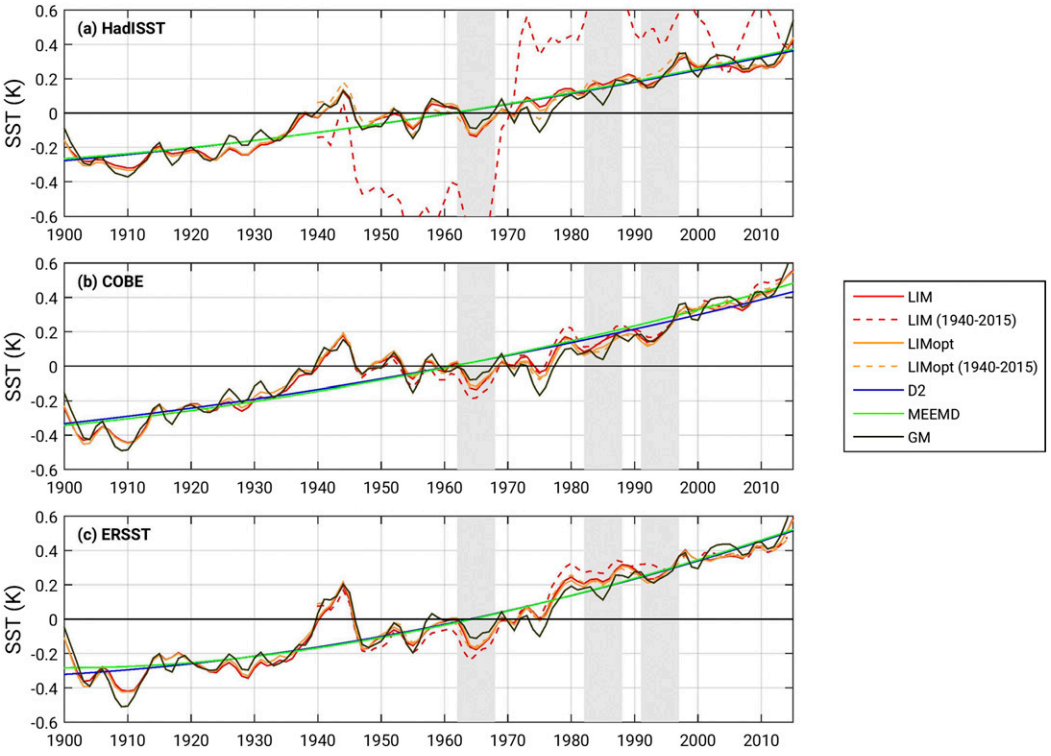

FIG. 15. Area-weighted global mean SST anomaly (K) in (a) HadISST, (b) COBE-SST, and (c) ERSST as estimated by the methods indicated in the right legend.

represents less variance than the main ENSO mode. Both factors make it more difficult for LIM to distinguish between modes of variability. Hence, a careful examination of the main LIM modes is required when representing the forced signal in short records. We recommend using LIMopt instead, as it does not seem affected by the record length.

There are differences in the AMO and PDO patterns between the three SST reconstructions (not shown) that largely reflect their degree of spatial smoothing, namely limited smoothing in HadISST, more substantial smoothing in COBE-SST, and larger smoothing at low frequency in ERSST. Nonetheless, the AMO and the PDO evolve similarly in the three reconstructions. The estimated AMO patterns are illustrated for COBE-SST in Figs. 16a,b, except for D2, which is nearly undistinguishable from D1. They are broadly similar in the North Atlantic, although in COBE-SST (and to a lesser extent in HadISST) the tropical signal is stronger for D1, D2, and MEEMD. These methods also show the largest Pacific anomalies, and some Southern Hemisphere warming instead of a weak cooling. Note that all AMO estimates lack the equatorial Pacific warming seen in the two LEs. Correspondingly, the AMO evolution in each SST reconstruction is very similar for D1 (or D2) and MEEMD, but substantially different for LIM, LIMopt, and REGR, which behave similarly (bottom panels); GM leads to smaller AMO amplitude during the last few decades, in particular in ERSST.

As illustrated for HadISST but found in each SST reconstruction, the PDO patterns are similar among methods (Figs. 17a,b). However, REGR yields smaller amplitude, consistent with the LE results, while MEEMD yields the largest warming off the west coast of North America. The central North Pacific cooling also varies somewhat between methods and datasets. The PDO time evolution at low frequency is similar for each method, so there are fewer differences than for the AMO. LIMopt and REGR lead to very similar behaviors in each reconstruction, despite differences in SST pattern. Why the PDO is more coherent among methods than the AMO may be in part due to the longer AMO time scale, which makes it more difficult to separate it from the slow variations of the forced signal. In addition, as in the LEs, the AMO pattern projects well on the forced SST signal in the Atlantic, whereas the PDO pattern strongly differs from the forced SST signal in the Pacific. Hence, the PDO is more easily separable from the forced signal than the AMO.

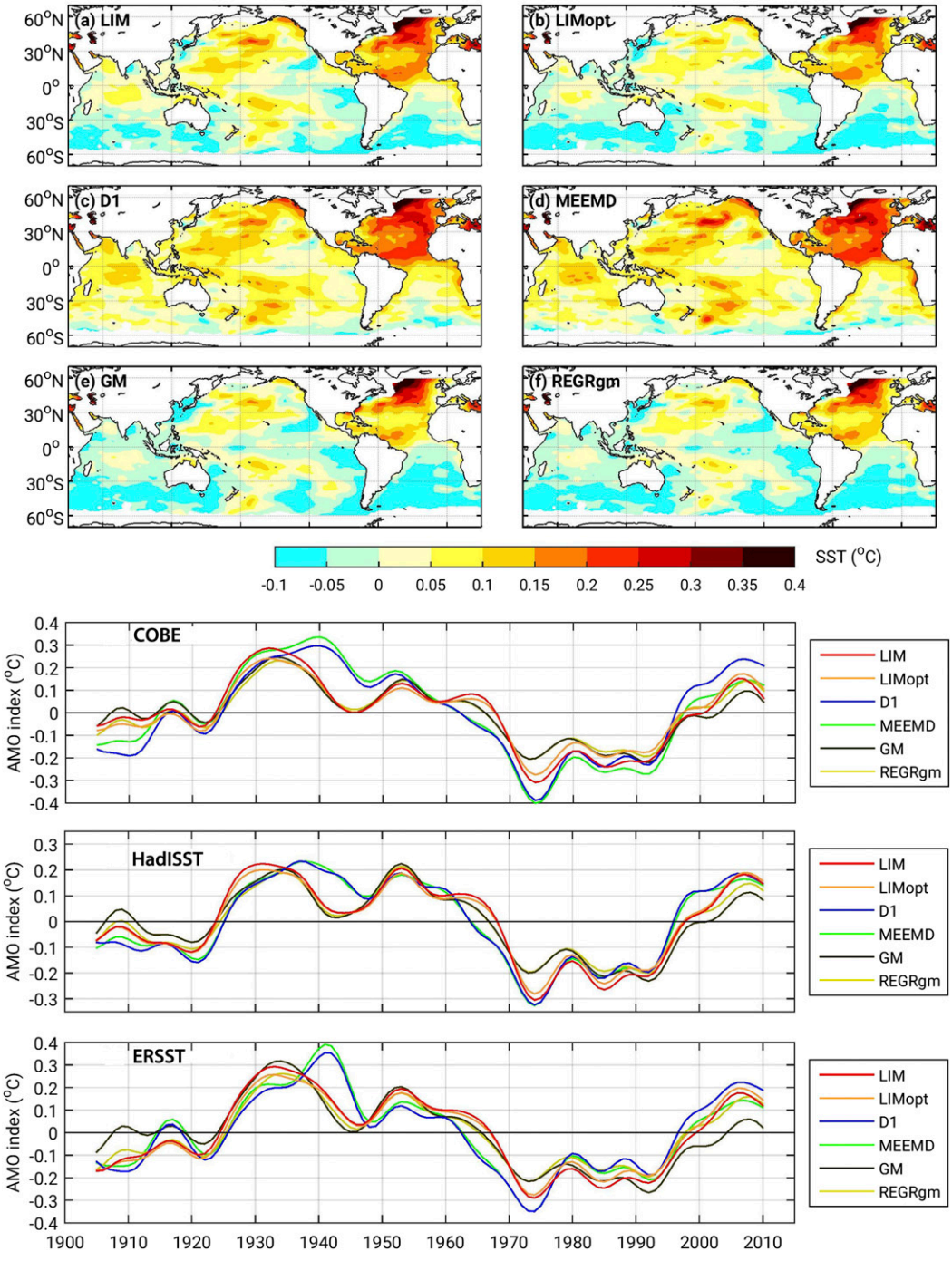

FIG. 16. (a)–(f) AMO pattern (°C) as estimated by each method (except D2, nearly indistinguishable from D1) in COBE-SST. (bottom) Estimated AMO time series in COBE-SST, HadISST, and ERSST.

The lag correlation between the AMO and the low-pass-filtered PDO was also calculated. All methods give rather similar results that are broadly consistent with earlier studies, albeit with large differences between reconstructions, but they are not 10% significant.

7. Summary and conclusions

Two large ensembles of historical simulations with the IPSL-CM5A-LR and the CESM1 climate models are used to assess how well various statistical methods estimate the SST changes due to external and anthropogenic

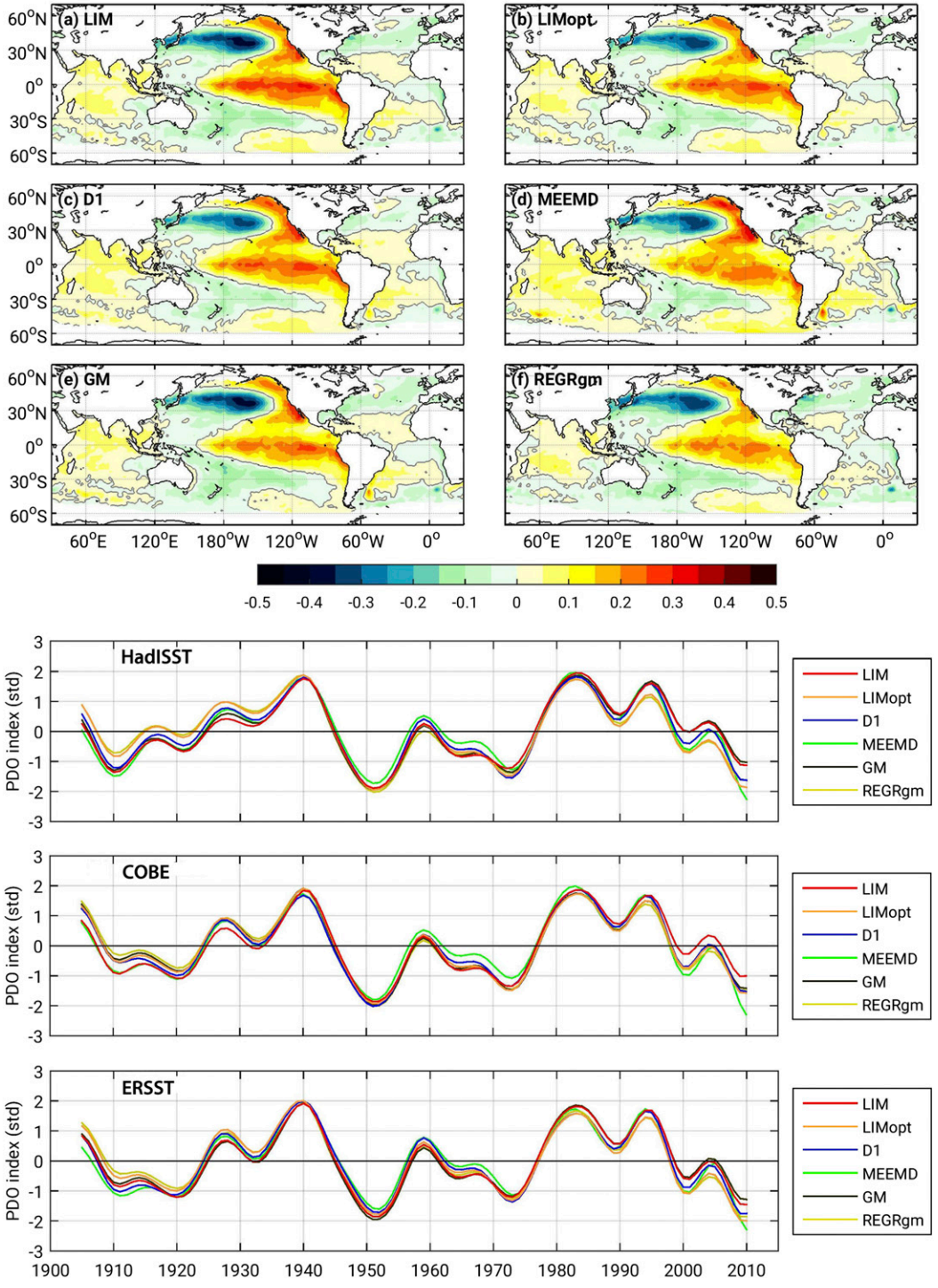

FIG. 17. (a)–(f) PDO pattern (K) as estimated by each method by regression on the standardized annual PDO time series (except D2, nearly indistinguishable from D1) in HadISST and (bottom) estimated PDO time series at low frequency in HadISST, COBE-SST, and ERSST.

forcing during 1940–2015, as well as the main patterns of natural decadal SST variability. The focus was on the AMO, PDO, and SST footprint of the AMOC, which were derived after removing the estimated forced SST signal. Because the amplitude of the natural climate variability is strongly reduced, the ensemble average SST was used as an approximation to the true forced SST signal, thus allowing us to compare the efficiency of each method in estimating the forced SST signal and the natural variability in individual ensemble members.

We have compared seven statistical methods that solely rely on observations to estimate the forced SST signal, thus excluding methods based on multimodel estimates. Two methods based on linear inverse modeling, a standard version (LIM) where the forced signal is taken as the first LIM mode and an optimal perturbation filter (LIMopt), perform consistently well and are often best, leading to small errors in estimating both the forced and the natural SST variability. Removing the forced signal by subtracting the global mean SST (GM) or a linear regression on the global mean SST (REGR) only works well in some cases and leads to large errors in the Pacific and the PDO, perhaps because of the link between global mean SST and PDO (Meehl et al. 2013). Quadratic detrending (D2) efficiently removes the forced SST signal in only one LE, and is unable to represent the SST cooling that follows volcanic eruptions. This is also the case for multidimensional ensemble empirical mode decomposition (MEEMD) because it decomposes signals into components of increasing time scale at each grid point, and hence cannot separate forced SST variations from natural ones with similar time scales but different spatial structures. Removing a linear trend (D1) fares poorly and leads to overestimated AMO fluctuations, so that approach should not be used, even in rather short records.

The two best methods, LIM and LIMopt, provide similar results that are often statistically undistinguishable. LIM is easier to program, but LIMopt is more general as it makes no assumption as to which eigenmode represents the forced SST signal. In fact, when LIM was applied to the ensemble-mean SST, the second LIM mode also contributed to the forced signal, particularly in CESM-LE. This occurred because the natural variability was so strongly reduced that the different SST response patterns expected from different kinds of anthropogenic forcing (Wang et al. 2016) could project on different eigenmodes. The natural variability in individual ensemble members is too strong to allow for such pattern separation, but the first two LIM modes were also needed in HadISST and, to a lesser extent, ERSST when the short period of 1940–2015 was considered. LIMopt is thus more robust since it automatically selected the modes that represent the forced signal.

The LIM methods generally provide a better skill at extracting the forced SST signal probably because they represent the natural climate variability well. LIM decomposes the SST fields into nonorthogonal modes, hence separating the fields into subsets that are not statistically independent. Perhaps more importantly, their success may also arise from the ability of LIM at representing the natural SST variability (the residuals from the global forced signal represented the first LIM mode) by the higher LIM modes. Indeed, midlatitude SST anomalies primarily reflect the response of the upper ocean to stochastic atmospheric forcing and are thus well represented by a first-order Markov process (Frankignoul and Hasselmann 1977; Frankignoul 1985), and tropical SST variability is also well described by a stable linear dynamical system driven by spatially coherent white noise (e.g., Penland and Sardeshmukh 1995; Penland and Matrosova 2006), even on decadal time scales (Newman 2007). Nonetheless, no filter is perfect and the higher LIM modes may retain some forced components, which would explain why LIMopt is more robust than LIM as it makes no a priori assumption as to which modes represent the forced signal. On the other hand, none of the other methods takes advantage of the statistical properties of the natural climate variability.

The comparison between the AMO pattern and the SST fingerprint of the AMOC suggests that in both LEs the AMO is largely driven by the AMOC in the subpolar gyre and the Gulf Stream region, but not in the tropical North Atlantic, which seems influenced by local atmospheric forcing, ENSO teleconnections, or ocean circulation variability unrelated to the AMOC. The maximum mean correlation between the AMOC (or the meridional heat transport) and the AMO occurs when the AMO lags by about 9 yr in IPSL-LE and 1 or 2 yr in CESM-LE, consistent with previous studies. However, even in the ''true'' case, there is much scatter among ensembles members, indicating that a 76-yr record is much too short to determine with some confidence the relation between the AMOC and the AMO. Finally, there is no significant correlation between the AMO and the PDO in either LE.

The methods were applied to the HadISST, COBE-SST, and ERSST SST reconstructions during 1900–2015. The differences between reconstructions are often comparable to or, in the case of the widely reported global warming slowdown from the late 1990s to 2012 that is only seen for the quasi-global mean SST in HadISST, exceed the differences between methods.

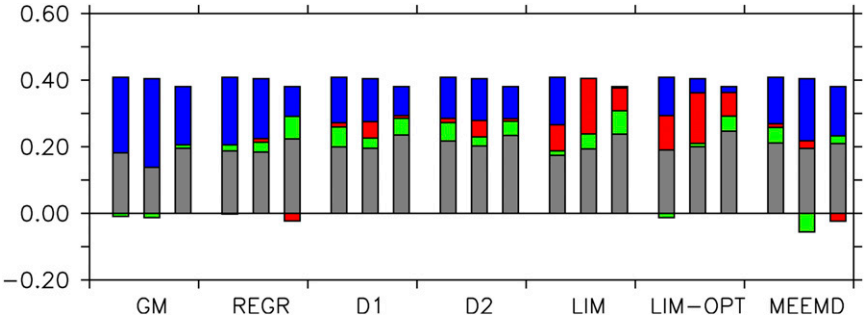

FIG. 18. Tropical transbasin trend (K decade^{-1}) during 1991–2010 in, from left to right, COBE-SST, HadISST, and ERSST, as given by each method. The blue bar indicates the total trend, the red bar is the component due to the forced signal, the green bar is that due to the AMO, and the gray bar is the PDO component. A binomial filter was applied before calculation.

Nonetheless, our estimates of the forced SST signal suggest that the natural SST variability slightly masked the quasi-global cooling following the Mount Agung eruption and only contributed to the big hiatus in surface warming from the mid-1940s to the mid-1970s by prolonging it by up to a decade, consistent with Kosaka and Xie (2016). In addition, natural SST variability plays no direct role in the recent warming slowdown from the late 1990s to 2012.

The estimated spatial patterns of natural variability largely reflect the degree of smoothing in the SST reconstructions. Nonetheless, they share common features in the reconstructions. The AMO time series differ somewhat when estimated by D1, D2, MEEMD, and, to a lesser extent, GM from their estimation by LIM, LIMopt, or REGR. Correspondingly, the AMO patterns are slightly different, in particular for D1, D2, and MEEMD, although they all show a similar warming in the subpolar gyre and along the North Atlantic Current. Interestingly, the amount of AMO variance that is attributed to natural variability and the correlation between forced and natural signals varies between methods and reconstructions. The largest percentage of variance attributed to natural variability is for D1 and D2 (52%, 42%, and 48% for COBE-SST, ERSST, and HadISST, respectively), while providing as expected nearly independent estimates of the free and forced signals (correlation $|r| \leq 0.08$). MEEMD leads to a rather similar ratio, but highly anticorrelated free and forced signals in COBE-SST ($r = -0.29$) and ERSST ($r = -0.2$). The smallest percentage of natural variability is found for REGR (21%, 26%, and 29% for COBE-SST, ERSST, and HadISST, respectively), with nearly independent free and forced AMO time series. The ratios are somewhat larger in GM, but the separation between free and forced signal is poor (r ranging between -0.21 and -0.43), which leads to an overestimation of the forced signal. In LIM, 33%–45% of the AMO variance is attributed to the natural variability, but the separation between free and forced signal is rather poor ($r \approx -0.2$). On the other hand, LIMopt separates well the free and forced signals ($|r| \leq 0.09$) and only attributes a moderate fraction of the AMO variance to natural variability (25%, 34%, and 38% for COBE-SST, ERSST, and HadISST, respectively). This suggests that a substantial fraction of the signals and climatic impacts attributed to the AMO when it is estimated from linearly detrended data (e.g., Knight et al. 2005; Sutton and Hodson 2005; Alexander et al. 2014; Delworth et al. 2017, and many others) results from more global forcing and should be interpreted with care.

The PDO patterns are more comparable among methods, except for REGR, which provides smaller PDO amplitude, and GM, which yields large errors. However, the PDO time series are largely similar. Most of the low-frequency PDO variance is associated with the natural variability, but free and forced signals can be highly anticorrelated; only LIMopt and, to a lesser extent, REGR (besides D1 and D2) provide nearly independent estimates of free and forced low-frequency PDO variability.

These differences between methods are consistent with those found in the LEs, suggesting that the most reliable estimates of the observed natural decadal variability are found by using LIMopt. That the PDO evolution is more coherent among methods than the AMO may be in part due to the longer AMO time scale, which is more comparable to the slow variations of the forced signal. In addition, the AMO pattern projects well on the forced SST signal, while the PDO pattern strongly differs. Hence, the PDO is more easily separable from the forced signal than the AMO.

As the number of LIM modes needed to represent the forced SST signal may vary and LIMopt better separates

free and forced variability, we recommend using LIMopt, which was consistently robust. Solomon and Newman (2012) also noted that the optimal perturbation filter led to more robust representation of ENSO than LIM, and more consistent estimates the ENSO-residual trend in different datasets. On the other hand, using D1, D2, MEEMD, GM, and REGR may lead to biased estimates. Although these differences may not seem striking, they could lead to different interpretations. Consider the tropical transbasin variability, defined by the standardized difference between the yearly mean SST anomaly in the tropical Atlantic–Indian Ocean (15°S–15°N, 40°W–60°E) and the tropical central Pacific (15°S–15°N, 180°–150°W). It approximately increases linearly during 2001–12, but the contributions of the forced signal (Fig. 18, red bar), the AMO (green), and the PDO (gray) differ between estimation methods and, to a lesser extent, datasets. Indeed, the forced signal does not contribute at all to the linear trend in GM and very little in REGR, D1, D2, and MEEMD, but it contributes substantially in LIM and LIMopt; the contribution of the PDO is smallest in GM, and the AMO contribution is small. This could lead to different assessment of forced versus natural climate impacts.

Acknowledgments. Support from the NOAA Climate Program Office Climate Variability and Predictability program (NA13OAR4310139), NSF EaSM2 (OCE-84298900), the European Community Horizon 2020 Framework under Grant Agreement 727852 (Blue-Action), and the ANR MORDICUS project (ANR-13-SENV-0002-02) is gratefully acknowledged. This work was granted access to the HPC resources of TGCC under allocation 2015-017403 made by GENCI. We thank the reviewers for insightful and constructive comments.

REFERENCES

Alexander, M. A., K. H. Kilbourne, and J. A. Nye, 2014: Climate variability during warm and cold phases of the Atlantic multidecadal oscillation (AMO) 1871–2008. *J. Mar. Syst.*, **133**, 14–26, doi:10.1016/j.jmarsys.2013.07.017.

Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor, 2012: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere–ocean climate models. *Geophys. Res. Lett.*, **39**, L09712, doi:10.1029/2012GL051607.

Branstator, G., and F. Selten, 2009: "Modes of variability" and climate change. *J. Climate*, **22**, 2639–2658, doi:10.1175/2008JCLI2517.1.

Chen, X., and K. K. Tung, 2014: Varying planetary heat sink led to global-warming slowdown and acceleration. *Science*, **345**, 897–903, doi:10.1126/science.1254937.

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

Compo, G., and P. Sardeshmukh, 2010: Removing ENSO-related variations from the climate record. *J. Climate*, **23**, 1957–1978, doi:10.1175/2009JCLI2735.1.

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.

DelSole, T., M. K. Tippett, and J. Shukla, 2011: A significant component of unforced variability in the recent acceleration of global warming. *J. Climate*, **24**, 909–926, doi:10.1175/2010JCLI3659.1.

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

Deser, C., A. S. Phillips, M. A. Alexander, and B. V. Smoliak, 2014: Projecting North American climate over the next 50 years: Uncertainty due to internal variability. *J. Climate*, **27**, 2271–2295, doi:10.1175/JCLI-D-13-00451.1.

——, L. Terray, and A. S. Phillips, 2016: Forced and internal components of winter air temperature trends over North America during the past 50 years: Mechanisms and implications. *J. Climate*, **29**, 2237–2258, doi:10.1175/JCLI-D-15-0304.1.

d'Orgeville, M., and W. Peltier, 2007: On the Pacific decadal oscillation and the Atlantic multidecadal oscillation: Might they be related? *Geophys. Res. Lett.*, **34**, L23705, doi:10.1029/2007GL031584.

Drijfhout, S., G. van Oldenborgh, and A. Cimatoribus, 2012: Is a decline of AMOC causing the warming hole above the North Atlantic in observed and modeled warming patterns? *J. Climate*, **25**, 8373–8379, doi:10.1175/JCLI-D-12-00490.1.

Dufresne, J.-L., and Coauthors, 2013: Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5. *Climate Dyn.*, **40**, 2123–2165, doi:10.1007/s00382-012-1636-1.

Enfield, D., and L. Cid-Serrano, 2010: Secular and multidecadal warmings in the North Atlantic and their relationships with major hurricane activity. *Int. J. Climatol.*, **30**, 174–184, doi:10.1002/joc.1881.

England, M., and Coauthors, 2014: Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nat. Climate Change*, **4**, 222–227, doi:10.1038/nclimate2106.

Frankcombe, L. M., M. H. England, M. E. Mann, and B. A. Steinman, 2015: Separating internal variability from the externally forced climate response. *J. Climate*, **28**, 8184–8202, doi:10.1175/JCLI-D-15-0069.1.

Frankignoul, C., 1985: Sea surface temperature anomalies, planetary waves, and air–sea feedback in the middle latitudes. *Rev. Geophys.*, **23**, 357–390, doi:10.1029/RG023i004p00357.

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

——, G. Gastineau, and Y.-O. Kwon, 2015: Wintertime atmospheric response to North Atlantic Ocean circulation variability in a climate model. *J. Climate*, **28**, 7659–7677, doi:10.1175/JCLI-D-15-0007.1.

Fyfe, J. C., and Coauthors, 2016: Making sense of the early-2000s warming slowdown. *Nat. Climate Change*, **6**, 224–228, doi:10.1038/nclimate2938.

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

——, F. D'Andrea, and C. Frankignoul, 2013: Atmospheric response to the North Atlantic Ocean variability on seasonal to decadal time scales. *Climate Dyn.*, **40**, 2311–2330, doi:10.1007/s00382-012-1333-0.

Guan, B., and S. Nigam, 2009: Analysis of Atlantic SST variability factoring interbasin links and the secular trend: Clarified structure of the Atlantic multidecadal oscillation. *J. Climate*, **22**, 4228–4240, doi:10.1175/2009JCLI2921.1.

Hannachi, A., 2007: Pattern hunting in climate: A new method for finding trends in gridded climate data. *Int. J. Climatol.*, **27**, 1–15, doi:10.1002/joc.1375.

Hartmann, D. L., and Coauthors, 2013: Observations: Atmosphere and surface. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 159–254.

Hirahara, S., M. Ishii, and Y. Fukuda, 2014: Centennial-scale sea surface temperature analysis and its uncertainty. *J. Climate*, **27**, 57–75, doi:10.1175/JCLI-D-12-00837.1.

Hourdin, F., and Coauthors, 2017: The art and science of climate model tuning. *Bull. Amer. Meteor. Soc.*, **98**, 589–602, doi:10.1175/BAMS-D-15-00135.1.

Huang, B., and Coauthors, 2015: Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4). Part I: Upgrades and intercomparisons. *J. Climate*, **28**, 911–930, doi:10.1175/JCLI-D-14-00006.1.

Huang, N. E., and Coauthors, 1998: The empirical mode decomposition method and the Hilbert spectrum for non-stationary time series analysis. *Proc. Roy. Soc. London*, **454A**, 903–995, doi:10.1098/rspa.1998.0193.

Ji, F., Z. Wu, J. Huang, and E. P. Chassignet, 2014: Evolution of land surface air temperature trend. *Nat. Climate Change*, **4**, 462–466, doi:10.1038/nclimate2223.

Karl, T. R., and Coauthors, 2015: Possible artifacts of data biases in the recent global surface warming hiatus. *Science*, **348**, 1469–1472, doi:10.1126/science.aaa5632.

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

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

Knutti, R., D. Masson, and A. Gettelman, 2013: Climate model genealogy: Generation CMIP5 and how we got there. *Geophys. Res. Lett.*, **40**, 1194–1199, doi:10.1002/grl.50256.

Kosaka, Y., and S.-P. Xie, 2013: Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature*, **501**, 403–407, doi:10.1038/nature12534.

——, and ——, 2016: The tropical Pacific as a key pacemaker of the variable rates of global warming. *Nat. Geosci.*, **9**, 669–673, doi:10.1038/ngeo2770.

Li, G., B. Ren, J. Zheng, and C. Yang, 2011: Trend singular value decomposition analysis and its application to the global ocean surface latent heat flux and SST anomalies. *J. Climate*, **24**, 2931–2948, doi:10.1175/2010JCLI3743.1.

Li, X., S.-P. Xie, S. T. Gille, and C. Yoo, 2015: Atlantic-induced pan-tropical climate change over the past three decades. *Nat. Climate Change*, **6**, 275–279, doi:10.1038/nclimate2840.

Lyu, K., J.-J. Yu, and H. Paek, 2017: The influences of the Atlantic multidecadal oscillation on the mean strength of the North Pacific subtropical high during boreal winter. *J. Climate*, **30**, 411–426, doi:10.1175/JCLI-D-16-0525.1.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Amer. Meteor. Soc.*, **78**, 1069–1079, doi:10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2.

Marini, C., and C. Frankignoul, 2014: An attempt to deconstruct the Atlantic multidecadal oscillation. *Climate Dyn.*, **43**, 607–625, doi:10.1007/s00382-013-1852-3.

McGregor, S., A. Timmermann, M. F. Stuecker, M. H. England, M. Merrifield, F.-F. Jin, and Y. Chikamoto, 2014: Recent Walker circulation strengthening and Pacific cooling amplified by Atlantic warming. *Nat. Climate Change*, **4**, 888–892, doi:10.1038/nclimate2330.

Medhaug, I., and T. Furevik, 2011: North Atlantic 20th century multidecadal variability in coupled climate models: Sea surface temperature and ocean overturning circulation. *Climate Dyn.*, **7**, 389–404, doi:10.5194/os-7-389-2011.

Meehl, G. A., A. Hu, J. M. Arblaster, J. Fasullo, and K. E. Trenberth, 2013: Externally forced and internally generated decadal climate variability associated with the interdecadal Pacific oscillation. *J. Climate*, **26**, 7298–7310, doi:10.1175/JCLI-D-12-00548.1.

Newman, M., 2007: Interannual to decadal predictability of tropical and North Pacific sea surface temperature anomalies. *J. Climate*, **20**, 2333–2356, doi:10.1175/JCLI4165.1.

——, 2013: An empirical benchmark for decadal forecasts of global surface temperature anomalies. *J. Climate*, **26**, 5260–5269, doi:10.1175/JCLI-D-12-00590.1.

Osborne, J. M., J. A. Screen, and M. Collins, 2017: Ocean–atmosphere state dependence of the atmospheric response to Arctic sea ice loss. *J. Climate*, **30**, 1537–1552, doi:10.1175/JCLI-D-16-0531.1.

Parker, D., C. Folland, A. Scaife, J. Knight, A. Colman, P. Baines, and B. Dong, 2007: Decadal to multidecadal variability and the climate change background. *J. Geophys. Res.*, **112**, D18115, doi:10.1029/2007JD008411.

Penland, C., and L. Matrosova, 1994: A balance condition for stochastic numerical models with application to the ENSO. *J. Climate*, **7**, 1352–1372, doi:10.1175/1520-0442(1994)007<1352:ABCFSN>2.0.CO;2.

——, and P. D. Sardeshmukh, 1995: The optimal growth of tropical sea surface temperature anomalies. *J. Climate*, **8**, 1999–2024, doi:10.1175/1520-0442(1995)008<1999:TOGOTS>2.0.CO;2.

——, and L. Matrosova, 2006: Studies of El Niño and interdecadal variability in tropical sea surface temperatures using a nonnormal filter. *J. Climate*, **19**, 5796–5815, doi:10.1175/JCLI3951.1.

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

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

Shin, S.-I., and P. D. Sardeshmukh, 2011: Critical influence of the pattern of tropical ocean warming on remote climate trends. *Climate Dyn.*, **36**, 1577–1591, doi:10.1007/s00382-009-0732-3.

Solomon, A., and M. Newman, 2012: Reconciling disparate twentieth-century Indo-Pacific ocean temperature trends in the instrumental record. *Nat. Climate Change*, **2**, 691–699, doi:10.1038/nclimate1591.

Sutton, R., and D. Hodson, 2005: Atlantic Ocean forcing of American and European summer climate. *Science*, **309**, 115–118, doi:10.1126/science.1109496.

Swart, N. C., J. C. Fyfe, N. Gillett, and G. J. Marshall, 2015: Comparing trends in the southern annular mode and surface westerly jet. *J. Climate*, **28**, 8840–8859, doi:10.1175/JCLI-D-15-0334.1.

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

Ting, M., Y. Kushnir, R. Seager, and C. Li, 2009: Forced and internal twentieth-century SST trends in the North Atlantic. *J. Climate*, **22**, 1469–1481, doi:10.1175/2008JCLI2561.1.

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

——, J. T. Fasullo, G. Branstator, and A. S. Phillips, 2014: Seasonal aspects of the recent pause in surface warming. *Nat. Climate Change*, **4**, 911–915, doi:10.1038/nclimate2341.

Vincent, L. A., X. Zhang, R. D. Brown, Y. Feng, E. Mekis, E. J. Milewska, H. Wan, and X. L. Wang, 2015: Observed trends in Canada's climate and influence of low-frequency modes. *J. Climate*, **28**, 4545–4560, doi:10.1175/JCLI-D-14-00697.1.

Wang, H., S.-P. Xie, and Q. Liu, 2016: Comparison of climate response to anthropogenic aerosol versus greenhouse gas forcing: Distinct patterns. *J. Climate*, **29**, 5175–5188, doi:10.1175/JCLI-D-16-0106.1.

Wettstein, J. J., and C. Deser, 2014: Internal variability in projections of twenty-first- century Arctic sea ice loss: Role of the large-scale atmospheric circulation. *J. Climate*, **27**, 527–550, doi:10.1175/JCLI-D-12-00839.1.

Wu, Z., and N. E. Huang, 2009: Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.*, **1**, 1–41, doi:10.1142/S1793536909000047.

——, ——, and X. Chen, 2009: The multi-dimensional ensemble empirical mode decomposition method. *Adv. Adapt. Data Anal.*, **1**, 339–372, doi:10.1142/S1793536909000187.

——, ——, J. M. Wallace, B. V. Smoliak, and X. Chen, 2011: On the time-varying trend in global-mean surface temperature. *Climate Dyn.*, **37**, 759–773, doi:10.1007/s00382-011-1128-8.

Zhang, Y., J. M. Wallace, and D. S. Battisti, 1997: ENSO-like interdecadal variability: 1900–93. *J. Climate*, **10**, 1004–1020, doi:10.1175/1520-0442(1997)010<1004:ELIV>2.0.CO;2.