A Predictable AMO-Like Pattern in the GFDL Fully Coupled Ensemble Initialization and Decadal Forecasting System

XIAOSONG YANG,*,+ ANTHONY ROSATI,* SHAOQING ZHANG,* THOMAS L. DELWORTH,*
RICH G. GUDGEL,* RONG ZHANG,* GABRIEL VECCHI,* WHIT ANDERSON,* YOU-SOON CHANG,*,+
TIMOTHY DELSOLE,# KEITH DIXON,* RYM MSADEK,*,+ WILLIAM F. STERN,* ANDREW WITTENBERG,*
AND FANRONG ZENG*

* NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey
+ University Corporation for Atmospheric Research, Boulder, Colorado
# George Mason University, Fairfax, Virginia, and Center for Ocean–Land–Atmosphere Studies, Calverton, Maryland

ABSTRACT

The decadal predictability of sea surface temperature (SST) and 2-m air temperature (T2m) in the Geophysical Fluid Dynamics Laboratory (GFDL) decadal hindcasts, which are part of the Fifth Coupled Model Intercomparison Project experiments, has been investigated using an average predictability time (APT) analysis. Comparison of retrospective forecasts initialized using the GFDL Ensemble Coupled Data Assimilation system with uninitialized historical forcing simulations using the same model allows identification of the internal multidecadal pattern (IMP) for SST and T2m. The IMP of SST is characterized by an interhemisphere dipole, with warm anomalies centered in the North Atlantic subpolar gyre region and North Pacific subpolar gyre region, and cold anomalies centered in the Antarctic Circumpolar Current region. The IMP of T2m is characterized by a general bipolar seesaw, with warm anomalies centered in Greenland and cold anomalies centered in Antarctica. The retrospective prediction skill of the initialized system, verified against independent observational datasets, indicates that the IMP of SST may be predictable up to 4 (10) yr lead time at 95% (90%) significance level, and the IMP of T2m may be predictable up to 2 (10) yr at the 95% (90%) significance level. The initialization of multidecadal variations of northward oceanic heat transport in the North Atlantic significantly improves the predictive skill of the IMP. The dominant roles of oceanic internal dynamics in decadal prediction are further elucidated by fixed-forcing experiments in which radiative forcing is returned abruptly to 1961 values. These results point toward the possibility of meaningful decadal climate outlooks using dynamical coupled models if they are appropriately initialized from a sustained climate observing system.

1. Introduction

Multiyear to decadal climate prediction, characterized by combined signals from external radiative forcing changes and internal climate variations, is an initiative of the fifth Coupled Model Intercomparison Project (CMIP5), which will be assessed in the Intergovernmental Panel on Climate Change's Fifth Assessment Report (IPCC AR5) (Meehl et al. 2009; Taylor et al. 2012). Although the capability to provide meaningful decadal climate outlooks using dynamical models has yet to be

firmly established, the pioneering decadal hindcast experiments using initialized coupled models appear promising (Keenlyside et al. 2008; Smith et al. 2007).

A potentially large source of multiyear to decadal predictability may come from fluctuations in the Atlantic meridional overturning circulation (AMOC) (Delworth and Mann 2000; Knight et al. 2005; Zhang and Delworth 2006). The multidecadal variations of the basin-scale North Atlantic sea surface temperature (SST), generally referred to as the Atlantic multidecadal oscillation (AMO) (Enfield et al. 2001), have been hypothesized to be associated with AMOC fluctuations (Knight et al. 2005; Solomon et al. 2011). Zhang (2008) showed that the AMO is closely linked to North Atlantic subsurface temperature in observations and the Geophysical Fluid Dynamics Laboratory Climate Model version 2.1 (GFDL

CM2.1) control simulation with constant external conditions, and it was suggested that the AMO has similar predictability as subsurface temperature and AMOC on the order of 10 years in perfect model predictability experiments using CM2.1 (Msadek et al. 2010). The potential predictability of North Atlantic upper-ocean temperature $O(10\text{ yr})$ was found in the CM2.1 long control integrations using an initial-value predictability metric (Branstator et al. 2012). AMO-like SST patterns were found in the uninitialized CMIP3 simulations for the twentieth-century, twenty-first-century, and pre-industrial eras, while most models tend to produce AMO of shorter time scales and less periodic than suggested by observations (Ting et al. 2011). However, the existence of the AMO-like SST patterns in realistic decadal hindcasts has not been well established, and the retrospective prediction skill of such patterns using the GFDL fully coupled ensemble initialization and decadal forecasting system has not been assessed.

The expected gain from initialized decadal hindcasts is to capture the evolution of slow internal variations (e.g., AMO) of the climate system in addition to the relatively robust response to external forcings, so it is of necessity to detect the existence of those internal variations in the initialized hindcasts. The detection involves separating natural internal variability from anthropogenic forced response (Solomon et al. 2011). Here we take the approach of comparing parallel sets of initialized decadal hindcasts and uninitialized historical forcing simulations made with the same model. If the external forcing is identical, then the difference between the two sets of hindcasts, called internal residuals here, isolates the impact of initialization. We employ a new method, called the average predictability time (APT) optimization (DelSole and Tippett 2009a,b), to identify characteristic patterns of predictable components in the internal residuals of decadal hindcasts. The method successfully identified an internal multidecadal SST pattern from the CMIP3 multimodel unforced simulations (DelSole et al. 2011).

In this study, we apply APT analysis to investigate the decadal predictability of the annual mean SST and 2-m surface temperature in the GFDL IPCC CMIP5 hindcast experiments. Our main goals are to identify the internal multidecadal patterns in the initialized decadal hindcasts and assess the prediction skill of those patterns. Details of the hindcasts, observational datasets, and methods are discussed in section 2. In section 3, the internal multidecadal patterns for SST and T2m are identified by the APT analysis, the retrospective prediction skill is assessed using observations, and roles of ocean internal dynamics in decadal prediction are investigated. Conclusions and discussions are given in section 4.

2. Decadal hindcasts, datasets, and methods

The decadal hindcasts were initialized by the GFDL ensemble coupled data assimilation (ECDA) system. The ECDA employs an ensemble-based filtering algorithm applied to the GFDL fully coupled climate model, CM2.1, which is one of two GFDL CMIP3 models (Delworth et al. 2006). More details of ECDA can be found in Zhang and Rosati (2010) and Zhang et al. (2007), and a comprehensive assessment of oceanic variability from the latest version of the ECDA analyzed from 1960 to 2010 can be found in Chang et al. (2012). This fully coupled model methodology was chosen to produce a balanced state between the atmosphere and ocean. The 10-member ensemble decadal hindcasts were initialized on 1 January every year from 1961 to 2011 and integrated for 10 years with temporally varying anthropogenic and natural forcing, giving a total of 5100 years of hindcasts.

For the historical forcing simulations, the 10 ensemble members using CM2.1 were integrated using temporally varying anthropogenic and natural forcing for the 1861–2020 period (Knutson et al. 2006). Note that the temporally varying anthropogenic and natural forcings between 1961 and 2020 are exactly the same for the historical forcing simulations and decadal hindcasts. To elucidate the role of initialization in predicting the internal multidecadal climate variations, we conducted a set of 10-member ensemble fixed-forcing decadal prediction experiments initialized on 1 January every five years from 1965 to 2010. In the fixed-forcing experiments, the values of the anthropogenic and natural forcings are returned to 1961 conditions, and the initial conditions are exactly the same as in the decadal hindcasts.

The mean forced response was obtained from the ensemble mean of the 10-member historical forcing simulations. The decadal hindcast anomalies for each variable were obtained by subtracting out the lead-time-dependent climatology from hindcasts, which effectively removes the climate drift assuming that the climate drift is systematic with the forecast lead time. Then, the internal residuals were computed by subtracting the mean forced response from hindcast anomalies. We diagnose the predictability in the internal residuals as the predictability gain arising purely from initialization. Mathematically, the internal residual (I) for a hindcast variable X is defined as

$$I_\tau = X_\tau - \overline{X}_\tau - R_\tau, \qquad (1)$$

where R is the mean forced response, and the overbar denotes the climatology of the hindcasts, which is a function of lead time τ.

The observational datasets for evaluating prediction skill include the sea surface temperature from the U.K.
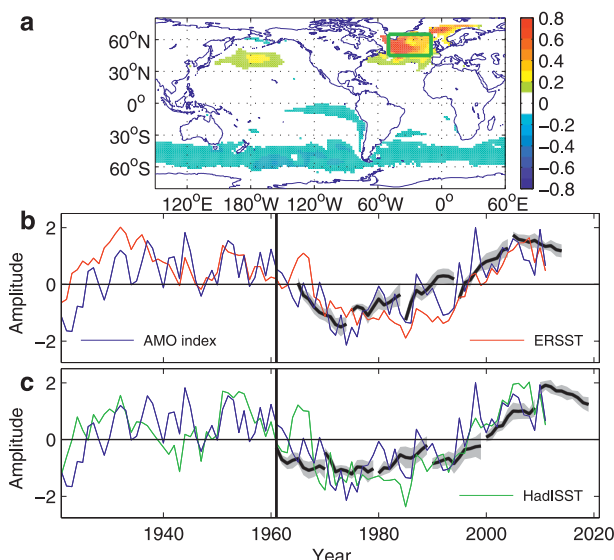
FIG. 1. (a) The spatial structure of the component that maximized the average predictability time of SST in the decadal hindcasts, which is called the internal multidecadal pattern (IMP). (b) The ensemble mean (black solid) and spread (gray shading) time series of the IMP as a function of forecast lead time for the decadal hindcasts initialized on 1 Jan every 10 yr from 1965 to 2005, the time series for projecting the ERSST data onto IMP (red solid), and the normalized AMO index (blue solid) from 1920 to 2010. (c) As in (b) but for hindcasts initialized on 1 Jan 1961 and every 10 yr from 1970 to 2010. The green line denotes the projected time series of HadISST data onto IMP.

Met Office Hadley Centre's global Sea Ice and Sea Surface Temperature dataset, version 1.1 (HadISST 1.1) (available online at http://badc.nerc.ac.uk/data/hadisst/) (Rayner et al. 2003) and the Extended Reconstruction Sea Surface Temperature (ERSST) analysis version 3b (Smith and Reynolds 2004); the 2-m temperature (T2m) from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis (NNR) (Kalnay et al. 1996; Kistler et al. 2001) from 1948 to 2011; and the Twentieth-Century Reanalysis (20CR) from 1900 to 2010 (Compo et al. 2011). Consistent with the hindcasted internal residuals, we obtain observed internal residuals by subtracting out the long-term linear trend covering the whole dataset for each variable. Van Oldenborgh et al. (2012) reported that verification of the decadal climate prediction skill does not depend strongly on the definition of the trend, and we obtain similar results of decadal predictive skill when a quadratic trend is removed from observations, so we choose removing the long-term linear trend from observations in this study.

The observed AMO index used in Figs. 1 and 4 was downloaded from http://www.esrl.noaa.gov/psd/data/timeseries/AMO/. The AMO index is defined as the detrended, area-weighted-average SST over the North Atlantic from 0° to 70°N using the Kaplan SST dataset (Kaplan et al. 1998). The index was annually averaged and normalized to unit variance.

Complete details of the average predictability time (APT) can be found in DelSole and Tippett (2009a,b). Briefly, the method is to optimize APT, which is defined as the integral over lead time of the "signal to total" ratio of a forecast mode:

$$\text{APT} = 2 \int_0^\infty \frac{\sigma_{\text{signal}}^2(\tau)}{\sigma_{\text{total}}^2} \, d\tau, \qquad (2)$$

where $\sigma_{\text{signal}}^2(\tau)$ is the variance of the ensemble mean at fixed lead time τ and σ_{total}^2 is the corresponding total variance of the forecast ensembles. For the ensemble forecasts, the signal and total covariance can be approximated by the corresponding ensemble covariances. Following DelSole and Tippett (2009a), maximizing APT in ensemble forecasts leads to the generalized eigenvalue problem

$$\left(2 \sum_{\tau=1}^\infty \mathbf{\Sigma}_{\text{signal}}(\tau) \right) \mathbf{q} = \lambda \mathbf{\Sigma}_{\text{total}} \mathbf{q}, \qquad (3)$$

where $\mathbf{q}$ is the projection vector, $\mathbf{\Sigma}_{\text{signal}}(\tau)$ is the forecast ensemble mean covariance matrix at the lead time τ for a given variable, and $\mathbf{\Sigma}_{\text{total}}$ is the total ensemble covariance matrix over all lead times. The eigenvectors $\mathbf{q}$ provide the basis for decomposing the multivariate time series into a complete, uncorrelated set of components ordered such that the first maximizes APT, the second maximizes APT subject to being uncorrelated with the first, and so on. The eigenvalues of (3) correspond to the APT values of each component. This decomposition based on APT is analogous to empirical orthogonal function (EOF) analysis except that we decompose predictability instead of decomposing variance. More detailed descriptions of the APT calculations are given in the first section of the appendix.

3. Results

a. The IMP of SST

We first apply APT analysis to the internal residuals of the annual mean SST. As discussed in section 2, using the internal residuals ensures that the obtained predictability is due to internal variability and not to natural and anthropogenic forcing. The component with maximum APT in decadal hindcasts is shown in Fig. 1a. The pattern is predominantly of a general interhemisphere dipole with warm anomalies centered in the North Atlantic subpolar
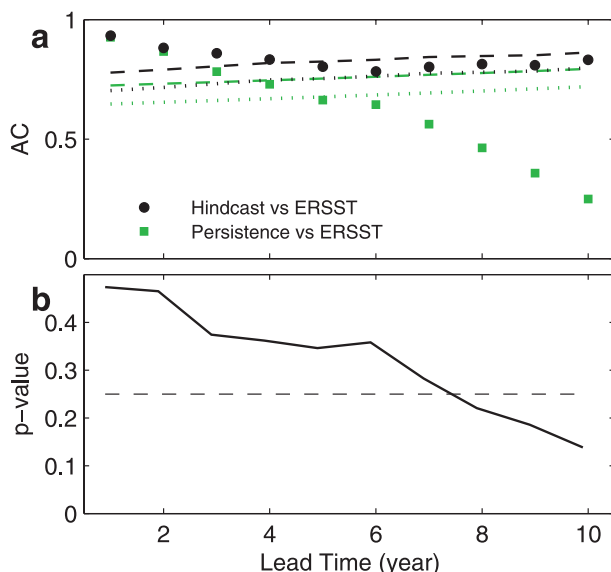
FIG. 2. (a) The anomaly correlation (AC) coefficients between forecasts and observations for the model hindcasts (black dots) and the persistence forecasts (green squares) of the SST IMP verified by the ERSST observations as a function of the forecast lead time. The dashed and dotted lines indicate the critical values at the 95% and 90% significance levels of AC, respectively, for model hindcasts (black) and persistence forecasts (green). (b) The p value from the statistical significance test that the two anomaly correlation coefficients for model hindcasts and persistence forecasts are not different from each other. The dashed line denotes the p value 0.25.

gyre region and North Pacific subpolar gyre region and cold anomalies centered in the Antarctic Circumpolar Current region. Note that the amplitudes of the pattern are considerably larger in the North Atlantic subpolar gyre region (over 0.5°C) than those in the North Pacific subpolar gyre and the Antarctic Circumpolar Current (about 0.1°–0.2°C) regions. The APT for this component is 18.9 yr, and the fraction of explained annual mean variance by this component is about 6.8%. The APT of this component is found to be statistically significant from a white noise process (see the first section of the appendix). Note that APT measures the time scale of predictability in a perfect model scenario, whereas the quantification of the actual predictive skill for this component remains to be verified against observations. Since the component has loadings concentrated in the North Atlantic subpolar gyre (SPG) and has temporal variations closely following the AMO index as shown next, we refer to this component as the AMO-like internal multidecadal pattern (IMP).

The APT value is probably underestimated because we only integrate APT up to 10-yr lead time in (3) due to the 10-yr upper limit of the hindcasts. If we had longer forecasts, we would expect to obtain components with at

least as much APT as found in the 10-yr case. The APT also could depend significantly on model: DelSole et al. (2011) found that the APT estimated from preindustrial control runs of the CMIP3 dataset varied by more than a factor of 4 (see also Branstator et al. 2012). However, the model-to-model predictability variations in decadal hindcasts could be reduced because all model states are initialized from the same climate observing system, although the initialization methods vary among model hindcasts. The sensitivity of APT to different model hindcasts is beyond the scope of this study.

The time series of the component with maximum APT in the decadal hindcasts as a function of initialized years and lead time every five years from 1961 to 2010 are shown by shading in Figs. 1b,c. To assess the forecast skill of the component, we project the detrended ERSST and HadISST data onto the eigenvector **q** with maximum APT from (3) to obtain the observed time series, which are indicated by solid lines in Figs. 1b,c. A striking aspect of two observed time series is the multidecadal oscillations with negative anomalies during 1965–95 and positive anomalies during 1925–60 and 1996–2010. Also, the two observed time series are strongly correlated with the annual mean observed Atlantic multidecadal oscillation index (the correlation coefficients are 0.75 and 0.77 for ERSST and HadISST datasets, respectively). Interestingly, the initialized forecasts of the component closely follow the observed AMO index during the negative (1961–94) and positive (1996–2010) phases, and even the phase shift around 1995.

To facilitate comparison with statistical prediction methods, we chose a simple persistence model, a commonly used benchmark for skill, which assumes the forecast equals the initial condition. The anomaly correlation (AC) coefficients between forecasts and ERSST observation as a function of lead time, shown in Fig. 2a, are statistically significant up to 4-yr lead time at 95% significance level and up to 10 yr at 90% significance level for the hindcasts, while they are statistically significant up to 3-yr lead time at 95% significance level and up to 4 yr at 90% significance level for the persistence forecasts. Verification of predictions against HadISST data yields similar predictive skill (figure not shown). The p value of testing significance difference between two AC coefficients for model hindcasts and persistence forecasts, shown in Fig. 2b, tends to decrease from about 0.45 to 0.15 with forecast lead time. Note that the critical values and p values were computed by adjusting the effective sample size accounting for autocorrelations in the data (see the second section of the appendix). Although 41–50 forecasts were used to compute the skill of forecasts at given lead time, the effective sample size is generally less than 10 owing to the strong autocorrelations

of the forecasts and observations, resulting in relatively large p values (larger than 0.1). Given these relatively large p values, we choose 0.25 as the threshold p value for testing the difference between two correlation coefficients. We conclude that the prediction skill of model hindcasts is statistically better than that of persistence forecasts for 8–10-yr lead times at 75% significance level, and not statistically different for 1–7-yr lead times.

The covarying structure of the AMO-like IMP in the North Atlantic and North Pacific is similar to the IMP identified from the CMIP3 multimodel unforced simulations using the APT analysis (DelSole et al. 2011), which may be explained as the impact of AMO on North Pacific climate variability through atmospheric teleconnections (Zhang and Delworth 2007). It was also indicated in previous studies that the strong signals of SST variability associated with AMO exist in the North Atlantic subpolar gyre in observations as well as model simulations (Folland et al. 1986; Kushnir 1994; Delworth and Mann 2000). Previous model studies indicated that multidecadal Atlantic meridional overturning circulation variations force interhemispheric dipolar SST anomalies (Collins et al. 2006; Latif et al. 2006b), and the out-of-phase relationship of SST variations between the Antarctic Circumpolar Current region and the North Atlantic subpolar gyre region associated with the Atlantic multidecadal oscillation was also revealed in observations (Latif et al. 2006a; Wu et al. 2011). The advance of this study is that the AMO-like IMP is not only successfully identified in decadal hindcasts using APT analysis but this pattern as a whole may be predictable up to 4 yr at 95% significance level and up to 10 yr at 90% significance level in the GFDL fully coupled ensemble initialization and decadal forecasting system.

To further confirm that the predictable signals are not statistical artifacts of the APT technique, we directly analyze predictability of anomalous SST averaged over the North Atlantic subpolar gyre region (marked as a box bounded by 45°–65°N, 50°–10°W in Fig. 1a), where the strong predictable signals were indicated by APT analysis. The time series of anomalous SST in the North Atlantic subpolar gyre region in the decadal hindcasts as a function of initialized years and lead time every five years from 1961 to 2010 are shown by shading in Figs. 3a, b. The initialized forecasts closely follow the observations during the negative (1961–94) and positive (1996–2010) phases, and even the phase shift around 1995. Also, the ensemble mean of uninitialized historical forcing simulations cannot produce multidecadal variations associated with realistic AMO, indicating that the decadal predictability arises from initialization. When forecasts are verified against ERSST, the SST in the North Atlantic subpolar gyre region is predictable up to

4 yr at 95% significance level and up to 7 yr at the 90% significance level in the model hindcasts, while the persistence model has skill only up to 3 yr at 95% significance level and up to 4 yr at 90% significance level (Fig. 3c). Verification of model hindcasts and persistence forecasts against HadISST data yields similar predictive skill (figure not shown). The predictive skill of model hindcasts is statistically better than that of persistence forecasts for 3–10-yr lead times at 75% significance level, and they are not statistically different for 1–2-yr lead times. This local predictability analysis confirms the robustness of the APT diagnosis for identifying the AMO-like IMP.

b. The IMP of T2m

The skill in predicting the IMP of SST on decadal time scales would imply skill in predicting the air surface temperature as a response to SST forcing; thus we apply the same APT analysis as described above to the internal residuals of the annual mean 2-m temperature. The component with maximum APT in decadal hindcasts is shown in Fig. 4a. The pattern is predominantly of a general bipolar seesaw with warm anomalies extending from the North Atlantic subpolar gyre region to Greenland and the Arctic Ocean and cold anomalies centered in Antarctica. The APT value for this component is 17.8 yr, and the fraction of explained annual mean variance by this component is about 4.5%. The APT of this component is found to be statistically significant. We project the detrended NNR and 20CR data onto the leading eigenvector $\mathbf{q}$ for T2m from (3) to obtain the observed time series, shown in Figs. 4b,c. Two observed time series for the T2m pattern show similar multidecadal variations in phase with the AMO as those for the SST IMP, for example, negative anomalies during 1965–95 and positive anomalies during 1925–60 and 1996–2010, suggesting the multidecadal variability of T2m arises from multidecadal SST variations. The anomaly correlation coefficients between forecasts and observations as a function of lead time, shown in Fig. 5, indicate that the T2m pattern as a whole may be predictable up to 2 yr at 95% confidence level and up to 10 yr at 90% significance level verified against the NNR data (except the 6-yr lead time). The p values of testing significance difference between 6-yr lead time AC coefficient and 8–10-yr lead AC coefficients are larger than 0.3, so the nominal increase of AC coefficients after 6-yr lead time could be due to sampling errors. Verification of the same predictions against 20CR data yields similar predictive skill (figure not shown). The predictive skill of model hindcasts is statistically better than that of persistence forecasts for 7–10-yr lead times at 75% significance level, and not statistically different for 1–6-yr lead times (Fig. 5).
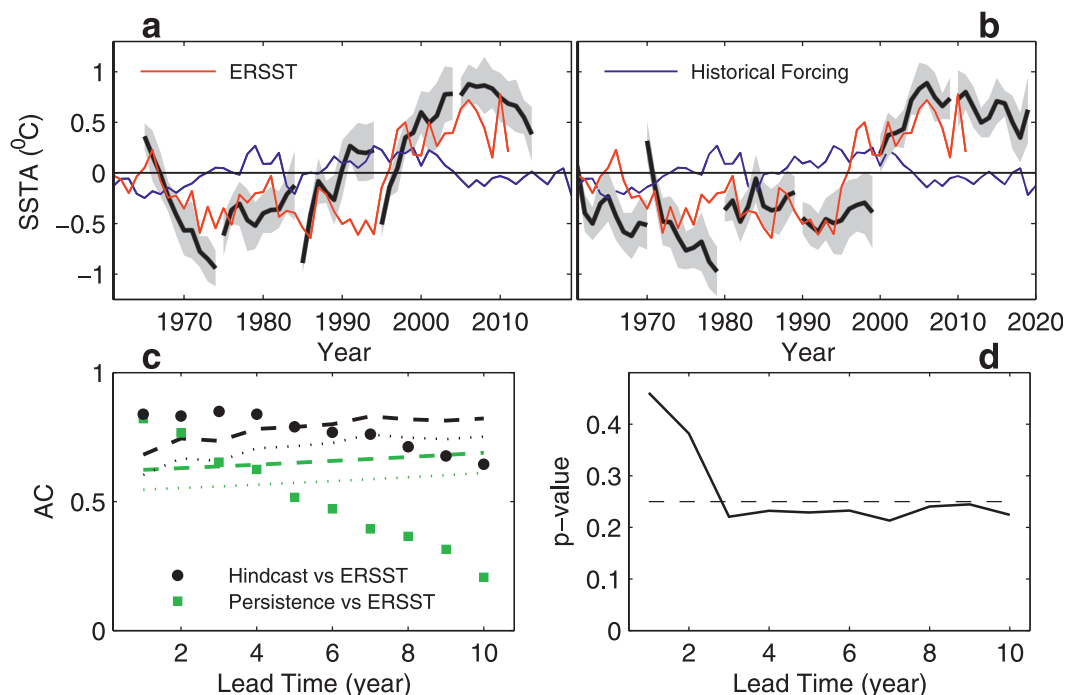
FIG. 3. (a) The ensemble mean (black solid) and spread (gray shading) time series of SST internal residuals in the North Atlantic SPG region (the green boxed region in Fig. 1) as a function of forecast lead time for the decadal hindcasts initialized on 1 Jan every 10 years from 1965 to 2005, the time series for the ERSST (red) data from 1920 to 2010, and the time series of the historical forcing simulations (blue) are plotted for reference. (b) As in (a) but for hindcasts initialized on 1 Jan 1961 and every 10 years from 1970 to 2010. (c) AC coefficients for the model hindcasts (black dots) and the persistence forecasts (green squares) verified by ERSST observations as a function of the forecast lead time. The dashed and dotted lines indicate the critical values at 95% and 90% significance levels of AC, respectively, for model hindcasts (black) and persistence forecasts (green). (d) The *p* value of the statistical test that the two anomaly correlation coefficients for model hindcasts and persistence forecasts are not different from each other. The dashed line denotes the *p* value 0.25.

Although the so-called "bipolar seesaw" pattern has been observed in the Greenland and Antarctica ice core data on millennial time scales (Blunier and Brook 2001; Blunier et al. 1998), Chylek et al. (2010) identified a bipolar seesaw pattern in the twentieth-century Arctic and Antarctic instrumental temperature records on multidecadal time scales, and found that the Arctic (Antarctic) detrended temperatures are highly correlated (anticorrelated) with the AMO index, suggesting the Atlantic Ocean as a possible link between the climate variability of the Arctic and Antarctic regions. Our results not only confirmed the existence of the bipolar seesaw pattern on multidecadal time scales in phase with AMO, but progressed toward predicting the pattern as a whole on multiyear time scales in the GFDL fully coupled ensemble initialization and forecasting system.

c. *Roles of ocean internal dynamics in decadal prediction*

Previous studies with coupled models assuming perfect ocean initial conditions indicate that accurate initialization of the Atlantic MOC may allow Atlantic multidecadal variability to be predicted a decade or more in advance (Collins et al. 2006; Msadek et al. 2010), and the AMO is suggested to be induced by Atlantic MOC variations and associated oceanic heat transport fluctuations (Knight et al. 2005), so we examine the North Atlantic oceanic heat transport in ECDA and decadal hindcasts. The regression coefficients of North Atlantic northward oceanic heat transport between 30° and 70°N onto the time series of normalized North Atlantic subpolar gyre region SST anomalies for ECDA, shown in Fig. 6a, tend to be positive, indicating that ECDA captures the anomalous northward (southward) heat transport into the North Atlantic subpolar gyre associated with the positive (negative) phase of AMO. These results indicate that the heat transport variations in the North Atlantic are in phase with the AMO in both ECDA and hindcasts, which is consistent with that the AMOC variations are in phase with the observed AMO using observational and modeling results (Zhang 2007, 2008).
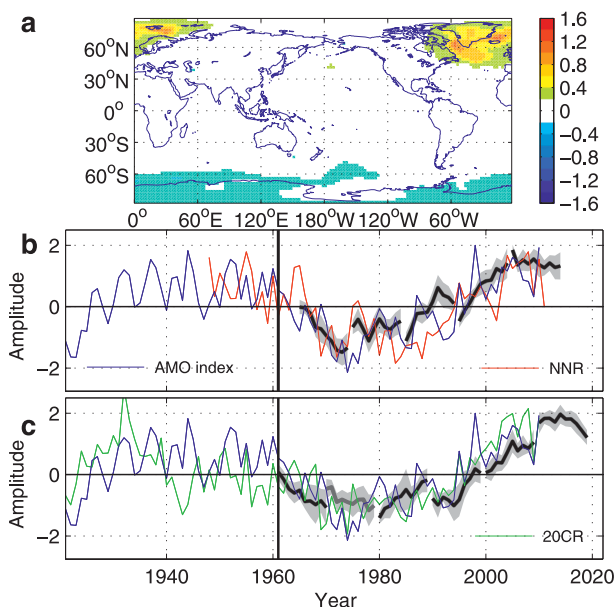
FIG. 4. (a) The spatial structure of the IMP for the 2-m temperature in the decadal hindcasts. (b) The ensemble mean (black solid) and spread (gray shading) time series of IMP as a function of forecast lead time for the decadal hindcasts initialized on 1 Jan every 10 years from 1965 to 2005, the time series for projecting the NNR data onto IMP (red solid), and the normalized AMO index (blue solid) from 1920 to 2010. (c) As in (b) but for hindcasts initialized on 1 Jan 1961 and every 10 years from 1970 to 2010. The green line denotes the projected time series of 20CR data onto IMP.
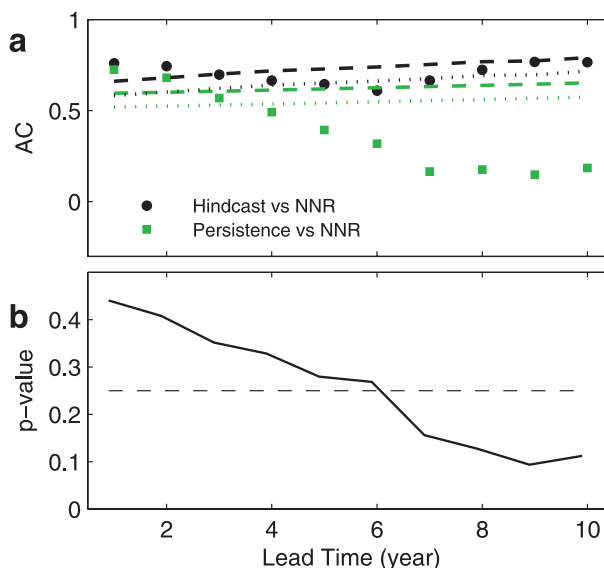
FIG. 5. (a) AC coefficients between forecasts and observations for the model hindcasts (black dots) and the persistence forecasts (green squares) of the T2m IMP verified by the NNR observations as a function of the forecast lead time. The dashed and dotted lines indicate the critical values at the 95% and 90% significance levels of AC, respectively, for model hindcasts (black) and persistence forecasts (green). (b) The *p* value from the statistical significance test that the two anomaly correlation coefficients for model hindcasts and persistence forecasts are not different from each other. The dashed line denotes the *p* value 0.25.

Consequently, the decadal hindcasts initialized by ECDA also show similar anomalous northward heat transport into the North Atlantic subpolar gyre region. The time series of the anomalous heat transports averaged between 35° and 65°N in the decadal hindcasts as a function of initialized years and lead time for 11 hindcast cases from 1961 to 2010, shown in Figs. 6b,c, generally have the same sign as the anomalous heat transports in the ECDA, and show anomalous southward heat transports from the 1960s until the early 1980s and anomalous northward heat transports from the late 1990s until 2010. In contrast, the oceanic heat transport in the uninitialized historical forcing simulations shows a secular weakening trend from 1960 to 2020 without any multidecadal variations. The predictive skill of the heat transport is verified against ECDA since there is no observational long-term data available for oceanic heat transport. The model hindcasts have skill up to 2 yr at 95% confidence level, while the historical forcing simulations show negative correlation with ECDA (Fig. 7). These results suggest that the initialization of multidecadal variations of northward oceanic heat transport in the North Atlantic significantly improves the predictive skill of SST.

The forecast anomalies of the North Atlantic SST and North Atlantic oceanic heat transport averaged between 35° and 65°N for the fixed-forcing decadal prediction experiments and the decadal hindcasts are shown in Fig. 8. The fixed-forcing forecasts of the North Atlantic SST show similar multidecadal variations as the decadal hindcasts as well as the phase transition around 1995 (Fig. 8a), indicating that the internal signals due to initialization dominate over the forced signal on the decadal time scale in the model. Meanwhile, the fixed-forcing forecasts of the North Atlantic SST initialized in 1995 and 2005 show a cooling during about 3–10-yr lead time resulting from the return of radiative forcing to 1961 values. This is similar to the fast cooling of the global mean temperature as a response to an instantaneous return to preindustrial forcing (Held et al. 2010), but the initial decay rate of the North Atlantic SST is much slower than that of the global mean SST (Fig. 8c). Note that the global mean SST in the fixed-forcing experiments shows a similar fast cooling as the global mean temperature in the experiments by returning radiative forcing to preindustrial values (Held et al. 2010). The slower decay of North Atlantic SST than the global mean SST is due to the fact that the northward heat transports in the North Atlantic evolve similarly during
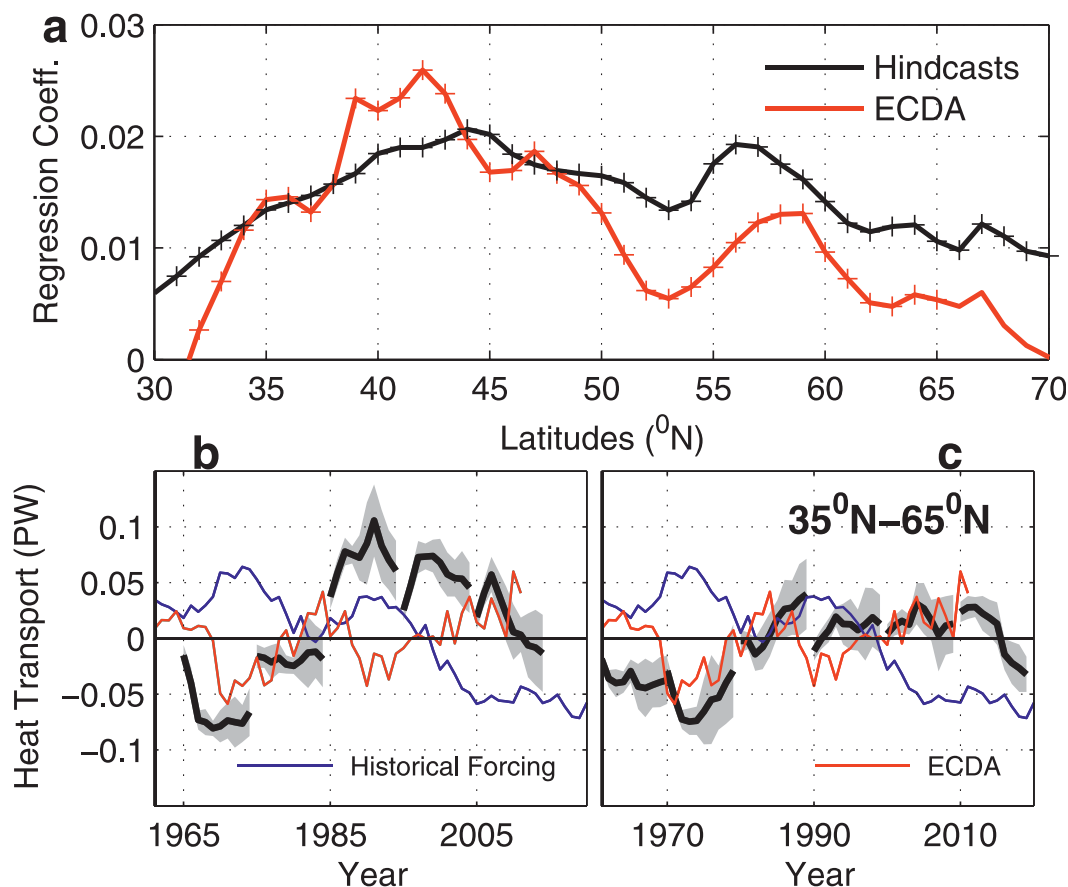
FIG. 6. (a) The regression coefficients of anomalous northward oceanic heat transport (PW) from 30° to 70°N in the North Atlantic onto the normalized time series of the North Atlantic subpolar gyre SST anomalies for decadal hindcasts and ECDA. (b) The ensemble mean (black solid) and spread (gray shading) time series of the anomalous oceanic heat transport averaged over the latitude belt between 35° and 65°N as a function of forecast lead time for the decadal hindcasts initialized on 1 Jan every 10 years from 1965 to 2005, and the time series for the historical forcing simulations (blue) and ECDA (red) are plotted for reference. (c) As in (b) but for hindcasts initialized on 1 Jan 1961 and every 10 years from 1970 to 2010. The plus symbol marker in (a) indicates the regression coefficient is significant at 90% level.

the 10-yr forecasts in the two experiments (Fig. 8b), suggesting that the internal oceanic heat transport compensates the fast cooling of the North Atlantic SST. The fixed-forcing experiments further demonstrate that the ocean internal dynamics play dominant roles on decadal prediction over the external forcings.

The latest forecast shows that the AMO-like SST pattern is weakening in the coming decade 2011–20 but is still in the warm phase (Fig. 1). Consistently, the bipolar seesaw pattern of T2m is predicted to be weakening but still above the climatology in the coming decade (Fig. 4). The dynamical basis for these forecasts is that the North Atlantic northward heat transport is predicted to be weakening (Fig. 6). This predicted weakening of AMO-like patterns in the coming decade using a dynamical model is consistent with the statistically predicted weakening of the AMOC

using subsurface and surface fingerprints (Mahajan et al. 2011).

4. Conclusions and discussion

The decadal predictability of SST and the 2-m air temperature (T2m) in the GFDL decadal hindcasts has been investigated using the average predictability time (APT) analysis. By diagnosing the internal residuals between initialized hindcasts and uninitialized historical forcing simulations using the same model (GFDL CM2.1), internal multidecadal patterns (IMP) for SST and T2m were identified. The IMP of SST is predominantly of an interhemisphere dipole with warm anomalies centered in the North Atlantic subpolar gyre region and North Pacific subpolar gyre region and cold anomalies centered in the Antarctic Circumpolar Current
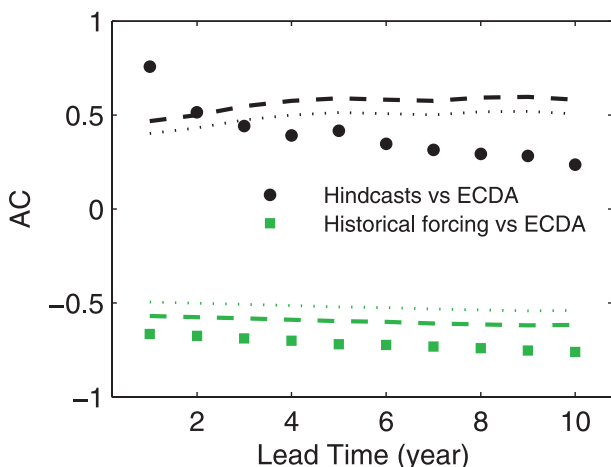
FIG. 7. (a) AC coefficients between forecasts and ECDA for the model hindcasts (black dots) and the historical forcing simulations (green squares) of the anomalous oceanic heat transport averaged over the latitude belt between 35° and 65°N in the North Atlantic as a function of the forecast lead time. The dashed and dotted lines indicate the critical values at the 95% and 90% significance levels of AC, respectively, for model hindcasts (black) and historical forcing simulations (green).
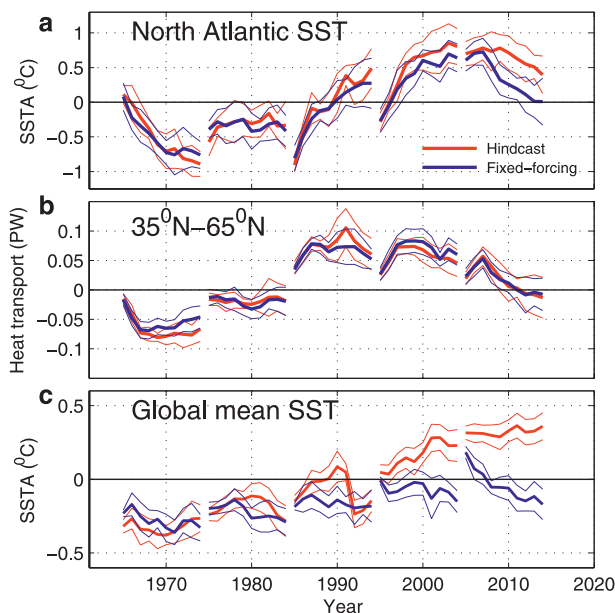

FIG. 8. (a) The ensemble mean (thick) and spread (thin) time series of the anomalous SST in the North Atlantic subpolar gyre region as a function of forecast lead time for the decadal hindcasts (red) and fixed-forcing (blue) experiments initialized every 10 years from 1965 to 2005. (b) As in (a) but for the anomalous oceanic heat transport averaged over the latitude belt between 35° and 65°N in the North Atlantic. (c) As in (a) but for the anomalous global mean SST.

region. The IMP of T2m is predominantly of a general bipolar seesaw with warm anomalies extending from the North Atlantic subpolar gyre region to Greenland and the Arctic Ocean and cold anomalies centered in Antarctica.

The projected time series of IMP onto observations closely follow the multidecadal oscillations, with negative anomalies during 1965–95 and positive anomalies during 1925–60 and 1996–2010, in phase with the observed Atlantic multidecadal oscillation (AMO) index. The retrospective prediction skills of the IMP for SST and T2m were verified against independent observational datasets, revealing that the SST pattern is predictable up to 4-yr lead time at 95% significance level and up to 10 yr at 90% significance level, and the T2m pattern is predictable up to 2-yr lead time at 95% significance level and up to 10 yr at 90% significance level. Further analysis suggests that the initialization of multidecadal variations of northward oceanic heat transport in the North Atlantic significantly improves the predictive skill of the AMO-like IMP. The fixed-forcing decadal prediction experiments, in which radiative forcing is returned abruptly to 1961 values, further elucidate roles of ocean internal dynamics in the decadal prediction.

Beyond the robust long-term anthropogenic signal predicted by uninitialized CMIP3 climate models (Hegerl et al. 2007), the results presented here from the GFDL CMIP5 decadal hindcasts show that the coupled climate model, initialized by advanced data assimilation methods, may be capable of predicting AMO-like internal multidecadal pattern over multiyears to a decade, thus pointing toward the possibility of meaningful decadal climate outlooks using dynamical models if they are appropriately initialized by the climate observing system. However, the predictable signal of SST is primarily located in the North Atlantic subpolar gyre, and not a basin-scale signal observed in the North Atlantic. The predictable signals of T2m are mainly over the two polar regions, but not over other continents indicated by previous studies (Hermanson and Sutton 2010). This discrepancy may be attributable to model biases and lack of deep-ocean observations in the twentieth century. Another possibility is that the continental predictability may be related to some other SST pattern that is less predictable than the AMO-like pattern (Jia and DelSole 2011).

A challenge for decadal prediction is the lack of many decadal-scale events (e.g., AMO transitions) for assessing its reliability. In contrast, there are many observed ENSO cases for assessing seasonal climate prediction. In this study, the hindcast period from 1961 to 2011 covers only one AMO episode, so the reliability of the predictive skill presented here needs to be evaluated by hindcasts covering more AMO episodes. More research is needed
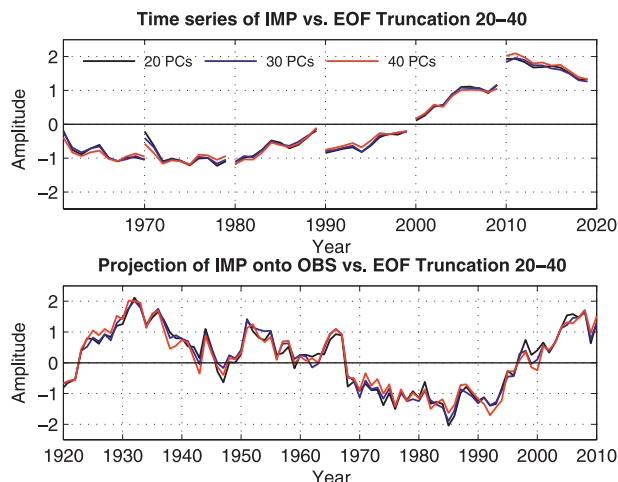
FIG. A1. (a) The ensemble mean time series of the internal multidecadal pattern (IMP) as a function of forecast lead time for the decadal hindcasts initialized on 1 Jan 1961 and every 10 years from 1970 to 2010 using 20, 30, and 40 leading PCs. (b) The time series for projecting the ERSST data onto IMP from 1920 to 2010 using 20, 30, and 40 leading PCs.

on producing a larger set of initial conditions covering more AMO episodes for assessing decadal climate prediction. Nevertheless, the skill of the dynamical model prediction is consistently better than that of the persistence forecasts, especially for forecast lead times longer than 5 years, thus the results presented here are encouraging for CMIP5 decadal prediction initiatives using initialized dynamical models.

APPENDIX

Details of Calculation

a. APT

For solving the APT optimization problem [Eq. (2) in the main paper] in practice, the data are first projected onto the leading principal components (PC) (DelSole et al. 2011). The sensitivity of APT to the truncation of PCs for SST, shown in Fig. A1, indicates that the results for the time series of IMP and projection of IMP onto observations are virtually independent of the number of
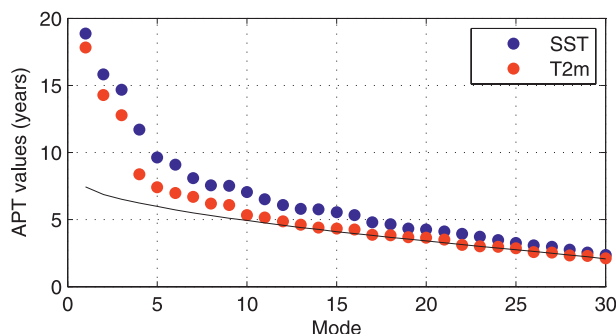

FIG. A2. The average predictability time (APT) using 30 leading PCs for SST and T2m. Solid horizontal line is the 5% significance level of the APT values.

PCs in the range of 20–40 PCs, presumably because the time series of 5100 years are relatively long. We obtained similar results for the APT analysis of T2m (figures not shown). We choose 30 EOFs for displaying results for both SST and T2m in the paper.

Following (DelSole et al. 2011), the statistical significance test of APT was estimated by Monte Carlo methods. The null hypothesis for the test is that the data are drawn from a white noise process when sampled every 2 yr. Accordingly, we generate a 30×2500 data matrix by drawing independent random numbers from a normal distribution with zero mean and unit variance. The time dimension of the data was grouped as a set of 25 10-yr forecasts with 10 ensemble members. Thirty APT values were then determined. This procedure was repeated 1000 times to generate 1000×30 APT values. The upper fifth percentile of the 1000×30 APT values was then determined as the threshold values, plotted in Fig. A2 as the horizontal line. The figure shows that the first 25 (10) components of SST (T2m) have statistically significant APT values. However, only the leading component has multiyear predictive skill verified against observations, so we only focus on the leading component in this study.

b. Statistical tests

Two statistical significance tests were performed in this study. One null hypothesis is that the anomaly correlation coefficient is zero, and critical values at 95% (90%) significance levels are computed for assessing the predictive skill at each lead time. The other null hypothesis is that two anomaly correlation coefficients are not different from each other. We compute the p value of testing the difference of two anomaly correlation coefficients for the decadal hindcasts and the persistence forecasts at each lead time.

Autocorrelation in the data was accounted for by adjusting the effective sample size (N^*) using the

following procedure (Trenberth 1984; Bretherton et al. 1999):

$$N^* = \frac{N}{1 + 2 \sum_{j=1}^{N-1} (1 - j/N) \rho_{xx}(j) \rho_{yy}(j)}, \quad (A1)$$

where N is the number of sample pairs, and $\rho_{xx}(j)$ and $\rho_{yy}(j)$ are the sample autocorrelation of x and y at lag j. The effective N^* was used for computing the critical values of testing the significance difference of the anomaly correlation from zero, and p values of testing the significance difference of two correlation coefficients. Although 41–50 samples were used to compute the skill of forecasts at each lead time, we found that the effective sample size is generally less than 10 owing to the strong autocorrelations of the forecasts and observations.

REFERENCES

Blunier, T., and E. J. Brook, 2001: Timing of millenial-scale climate change in Antarctica and Greenland during the last glacial period. *Science,* **291,** 109–112, doi:10.1126/science.291. 5501.109.

——, and Coauthors, 1998: Asynchrony of Antarctic and Greenland climate change during the last glacial period. *Nature,* **394,** 739–743, doi:10.1038/29447.

Branstator, G., H. Teng, G. A. Meehl, M. Kimoto, J. R. Knight, M. Latif, and A. Rosati, 2012: Systematic estimates of initial-value decadal predictability for six AOGCMs. *J. Climate,* **25,** 1827–1846.

Bretherton, C. S., M. Widmann, V. P. Dymnikov, J. M. Wallace, and I. Bladè, 1999: The effective number of spatial degrees of freedom of a time-varying field. *J. Climate,* **12,** 1990–2009.

Chang, Y.-S., S. Zhang, A. Rosati, T. L. Delworth, and W. F. Stern, 2012: An assessment of oceanic variability for 1960–2010 from the GFDL ensemble coupled data assimilation. *Climate Dyn.,* doi:10.1007/s00382-012-1412-2, in press.

Chylek, P., C. K. Folland, G. Lesins, and M. K. Dubey, 2010: Twentieth century bipolar seesaw of the Arctic and Antarctic surface air temperatures. *Geophys. Res. Lett.,* **37,** L08703, doi:10.1029/2010GL042793.

Collins, M., and Coauthors, 2006: Interannual to decadal climate predictability in the North Atlantic: A multimodel-ensemble study. *J. Climate,* **19,** 1195–1203.

Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis project. *Quart. J. Roy. Meteor. Soc.,* **137,** 1–28.

DelSole, T., and M. K. Tippett, 2009a: Average predictability time. Part I: Theory. *J. Atmos. Sci.,* **66,** 1172–1187.

——, and ——, 2009b: Average predictability time. Part II: Seamless diagnosis of predictability on multiple time scales. *J. Atmos. Sci.,* **66,** 1188–1204.

——, ——, and J. Shukla, 2011: A significant component of unforced multidecadal variability in the recent acceleration of global warming. *J. Climate,* **24,** 909–926.

Delworth, T. L., and M. E. Mann, 2000: Observed and simulated multidecadal variability in the Northern Hemisphere. *Climate Dyn.,* **16,** 661–676.

——, and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate,* **19,** 643–674.

Enfield, D. B., A. M. Mestas-Nuñez, and P. J. Trimble, 2001: The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental U.S. *Geophys. Res. Lett.,* **28,** 2077–2080.

Folland, C. K., T. N. Palmer, and D. E. Parker, 1986: Sahel rainfall and worldwide sea temperatures, 1901–85. *Nature,* **320,** 602–607.

Hegerl, G. C., and Coauthors, 2007: Understanding and attributing climate change. *Climate Change 2007: The Physical Science Basis,* S. Solomon et al., Eds., Cambridge University Press, 663–745.

Held, I. M., M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis, 2010: Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *J. Climate,* **23,** 2418–2417.

Hermanson, L., and R. Sutton, 2010: Case studies in interannual to decadal climate predictability. *Climate Dyn.,* **35,** 1169–1189.

Jia, L., and T. DelSole, 2011: Diagnosis of multiyear predictability on continental scales. *J. Climate,* **24,** 5108–5124.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856–1991. *J. Geophys. Res.,* **103** (C9), 18 567–18 589.

Keenlyside, N. S., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner, 2008: Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature,* **453,** 84–88, doi:10.1038/nature06921.

Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.,* **82,** 247–267.

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

Knutson, T. R., and Coauthors, 2006: Assessment of twentieth-century regional surface temperature trends using the GFDL CM2 coupled models. *J. Climate,* **19,** 1624–1651.

Kushnir, Y., 1994: Interdecadal variations in North Atlantic sea surface temperature and associated atmospheric conditions. *J. Climate,* **7,** 141–157.

Latif, M., C. Böning, J. Willebrand, A. Biastoch, J. Dengg, N. Keenlyside, U. Schweckendiek, and G. Madec, 2006a: Is the thermohaline circulation changing? *J. Climate,* **19,** 4631–4637.

——, M. Collins, H. Pohlmann, and N. Keenlyside, 2006b: A review of predictability studies of Atlantic sector climate on decadal time scales. *J. Climate,* **19,** 5971–5987.

Mahajan, S., R. Zhang, T. L. Delworth, S. Zhang, A. Rosati, and Y.-S. Chang, 2011: Predicting Atlantic meridional overturning circulation (AMOC) variations using subsurface and surface fingerprints. *Deep-Sea Res. II,* **58,** 1895–1903.

Meehl, G. A., and Coauthors, 2009: Decadal prediction: Can it be skillful? *Bull. Amer. Meteor. Soc.,* **90,** 1467–1485.

Msadek, R., K. W. Dixon, T. L. Delworth, and W. Hurlin, 2010: Assessing the predictability of the Atlantic meridional overturning circulation and associated fingerprints. *Geophys. Res. Lett.,* **37,** L19608, doi:10.1029/2010GL044517.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, 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.

Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy, 2007: Improved surface temperature prediction for the coming decade from a global climate model. *Science,* **317,** 796–799, doi:10.1126/science. 1139540.

Smith, T. M., and R. W. Reynolds, 2004: Improved extended reconstruction of SST (1854–1997). *J. Climate,* **17,** 2466–2477.

Solomon, A., and Coauthors, 2011: Distinguishing the roles of natural and anthropogenically forced decadal climate variability. *Bull. Amer. Meteor. Soc.,* **92,** 141–156.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of the CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.,* **93,** 485–498.

Ting, M., Y. Kushnir, R. Seager, and C. Li, 2011: Robust features of Atlantic multi-decadal variability and its climate impacts. *Geophys. Res. Lett.,* **38,** L17705, doi:10.1029/2011GL048712.

Trenberth, K. E., 1984: Some effects of finite sample size and persistence on meteorological statistics. Part I: Autocorrelations. *Mon. Wea. Rev.,* **112,** 2359–2368.

van Oldenborgh, G. J., F. J. Doblas-Reyes, B. Wouters, and W. Hazeleger, 2012: Decadal prediction skill in a multi-model

ensemble. *Climate Dyn.,* **38,** 1263–1280, doi:10.1007/s00382-012-1313-4.

Wu, Z., N. E. Huang, 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.

Zhang, R., 2007: Anticorrelated multidecadal variations between surface and subsurface tropical North Atlantic. *Geophys. Res. Lett.,* **34,** L12713, doi:10.1029/2007GL030225.

——, 2008: Coherent surface-subsurface fingerprint of the Atlantic meridional overturning circulation. *Geophys. Res. Lett.,* **35,** L20705, doi:10.1029/2008GL035463.

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

——, and ——, 2007: Impact of the Atlantic Multidecadal Oscillation on North Pacific climate variability. *Geophys. Res. Lett.,* **34,** L23708, doi:10.1029/2007GL031601.

Zhang, S., and A. Rosati, 2010: An inflated ensemble filter for ocean data assimilation with a biased coupled GCM. *Mon. Wea. Rev.,* **138,** 3905–3931.

——, M. J. Harrison, A. Rosati, and A. T. Wittenberg, 2007: System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Mon. Wea. Rev.,* **135,** 3541–3564.