

1 **A predictable AMO-like pattern in GFDL's fully-coupled**
2 **ensemble initialization and decadal forecasting system**

3 XIAOSONG YANG^{*1,2}, ANTHONY ROSATI¹, SHAOQING ZHANG¹,
4 THOMAS L. DELWORTH¹, RICH G. GUDGEL¹, RONG ZHANG¹,
5 GABRIEL VECCHI¹, WHIT ANDERSON¹, YOU-SOON CHANG^{1,2},
6 TIMOTHY DELSOLE³, KEITH DIXON¹, RYM MSADEK^{1,2},
7 WILLIAM F. STERN¹, ANDREW WITTENBERG¹, FANRONG ZENG¹

¹*NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey*

²*University Corporation for Atmospheric Research, Boulder, Colorado*

³*George Mason University, Fairfax, Virginia and Center for Ocean-Land-Atmosphere Studies, Calverton, Maryland*

Revised for *Journal of Climate*

July 16, 2012

* *Corresponding author address:* Xiaosong Yang, NOAA/Geophysical Fluid Dynamics Laboratory, 201 Forrester Rd., Princeton, NJ, 08540
E-mail: Xiaosong.Yang@noaa.gov

ABSTRACT

9 The decadal predictability of sea surface temperature (SST) and 2m air temperature (T2m)
10 in Geophysical Fluid Dynamics Laboratory (GFDL)'s decadal hindcasts, which are part of
11 the Fifth Coupled Model Intercomparison Project experiments, has been investigated us-
12 ing an average predictability time (APT) analysis. Comparison of retrospective forecasts
13 *initialized* using the GFDL's Ensemble Coupled Data Assimilation system with *uninitial-*
14 *ized* historical forcing simulations using the same model, allows identification of internal
15 multidecadal pattern (IMP) for SST and T2m. The IMP of SST is characterized by an
16 inter-hemisphere dipole, with warm anomalies centered in the North Atlantic subpolar gyre
17 region and North Pacific subpolar gyre region, and cold anomalies centered in the Antarctic
18 Circumpolar Current region. The IMP of T2m is characterized by a general bi-polar seesaw,
19 with warm anomalies centered in Greenland, and cold anomalies centered in Antarctica.
20 The retrospective prediction skill of the initialized system, verified against independent ob-
21 servational datasets, indicates that the IMP of SST may be predictable up to 4 (10) year
22 lead time at 95% (90%) significance level, and the IMP of T2m may be predictable up to 2
23 (10) years at 95% (90%) significance level. The initialization of multidecadal variations of
24 northward oceanic heat transport in the North Atlantic significantly improves the predic-
25 tive skill of the IMP. The dominant roles of oceanic internal dynamics in decadal prediction
26 are further elucidated by fixed-forcing experiments, in which radiative forcing is returned
27 abruptly to 1961 values. These results point towards the possibility of meaningful decadal
28 climate outlooks using dynamical coupled models, if they are appropriately initialized from
29 a sustained climate observing system.

1. Introduction

Multi-year 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 multi-year 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 multi-decadal 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 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 on the order of 10 years 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 20th-century, 21st-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 GFDL’s fully-coupled ensemble initialization and decadal forecasting

57 system has not been assessed.

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

71 In this study, we apply APT analysis to investigate the decadal predictability of the an-
72 nual mean SST and 2-m surface temperature in GFDL’s IPCC CMIP5 hindcast experiments.
73 Our main goals are to identify the internal multidecadal patterns in the initialized decadal
74 hindcasts and assess the prediction skill of those patterns. Details of the hindcasts, observa-
75 tional datasets and methods are discussed in section 2. In section 3, the internal multidecadal
76 patterns for SST and T2m are identified by APT analysis, the retrospective prediction skill
77 is assessed using observations, and roles of ocean internal dynamics on decadal prediction
78 are investigated. Conclusions and discussions are given in section 4.

79 **2. Decadal Hindcasts, Datasets and Methods**

80 The decadal hindcasts were initialized by the GFDL’s ensemble coupled data assimilation
81 (ECDA) system. The ECDA employs an ensemble-based filtering algorithm applied to the

82 GFDL’s fully coupled climate model, CM2.1, which is one of two GFDL CMIP3 models
 83 (Delworth et al. 2006). More details of ECDA can be found in Zhang and Rosati (2010)
 84 and Zhang et al. (2007), and a comprehensive assessment of oceanic variability from the
 85 latest version of the ECDA analyzed from 1960 to 2010 can be found in Chang et al. (2012).
 86 This fully coupled model methodology was chosen to produce a balanced state between the
 87 atmosphere and ocean. The 10-member ensemble decadal hindcasts were initialized on 1
 88 January every year from 1961 to 2011 and integrated for 10 years with temporally varying
 89 anthropogenic and natural forcing, giving a total of 5100 yr of hindcast.

90 For the historical forcing simulations, the 10 ensemble members using CM2.1 were inte-
 91 grated using temporally varying anthropogenic and natural forcing for the 1861-2020 period
 92 (Knutson et al. 2006). Note that the temporally varying anthropogenic and natural forcings
 93 between 1961 to 2020 are exactly the same for the historical forcing simulations and decadal
 94 hindcasts. To elucidate the role of initialization in predicting the internal multidecadal cli-
 95 mate variations, we conducted a set of 10-member ensemble *fixed-forcing* decadal prediction
 96 experiments initialized on 1 January every 5 years from 1965 to 2010. In the fixed-forcing
 97 experiments, the values of the anthropogenic and natural forcings are returned to 1961 con-
 98 ditions, and the initial conditions are exactly the same as in the decadal hindcasts.

99 The mean forced response was obtained from the ensemble mean of 10-member histor-
 100 ical forcing simulations. The decadal hindcast anomalies for each variable were obtained
 101 by subtracting out the lead-time dependent climatology from hindcasts, which effectively
 102 removes the climate drift assuming that the climate drift is systematic with the forecast lead
 103 time. Then, the internal residuals were computed by subtracting the mean forced response
 104 from hindcast anomalies. We diagnose the predictability in the *internal residuals* as the
 105 predictability gain arising purely from initialization. Mathematically, the internal residual
 106 (IR) for a hindcast variable \mathbf{X} is defined as

$$107 \quad \mathbf{IR}_\tau = \mathbf{X}_\tau - \bar{\mathbf{X}}_\tau - \mathbf{R}_\tau, \quad (1)$$

108 where \mathbf{R} is the mean forced response, the overbar denotes the climatology of hindcasts, which

109 is a function of lead time τ .

110 The observational datasets for evaluating prediction skill include the sea surface tem-
111 perature (SST) from the United Kingdom Meteorological Office Hadley Centre’s Global
112 sea-ice coverage and SST analysis (HadISST 1.1; available online at [http://badc.nerc.
113 ac.uk/data/hadisst/](http://badc.nerc.ac.uk/data/hadisst/)) (Rayner et al. 2003) and the Extended Reconstruction Sea Surface
114 Temperature (ERSST) analysis version 3b (Smith and Reynolds 2004), the 2-m temperature
115 (T2m) from the NCEP-NCAR Reanalysis (NNR) (Kalnay and Coauthors 1996; Kistler et al.
116 2001) from 1948 to 2011, and the 20-century reanalysis (20CR) from 1900 to 2010 (Compo
117 et al. 2011). Consistent with the hindcasted internal residuals, we obtain observed internal
118 residuals by subtracting out the long-term linear trend covering the whole dataset for each
119 variable. van Oldenborgh et al. (2012) reported that verification of the decadal climate pre-
120 diction skill does not depend strongly on the definition of the trend, and we obtain similar
121 results of decadal predictive skill when a quadratic trend is removed from observations, so
122 we choose removing the long-term linear trend from observations in this study.

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

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

$$130 \quad APT = 2 \int_0^{\infty} \frac{\sigma_{signal}^2(\tau)}{\sigma_{total}^2} d\tau, \quad (2)$$

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

135 generalized eigenvalue problem

$$136 \quad \left(2 \sum_{\tau=1}^{\infty} \Sigma_{signal}(\tau) \right) \mathbf{q} = \lambda \Sigma_{total} \mathbf{q}, \quad (3)$$

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

146 **3. Results**

147 *a. The IMP of SST*

148 We first apply APT analysis to the *internal* residuals of the annual mean SST. As dis-
149 cussed in section 2, using the internal residuals ensures that the obtained predictability is
150 due to internal variability and not to natural and anthropogenic forcing. The component
151 with maximum APT in decadal hindcasts is shown in Fig 1a. The pattern is predominantly
152 of a general inter-hemisphere dipole with warm anomalies centered in the North Atlantic
153 subpolar gyre region and North Pacific subpolar gyre region, and cold anomalies centered
154 in the Antarctic Circumpolar Current region. Note that the amplitudes of the pattern are
155 considerably larger in the North Atlantic subpolar gyre region (over $0.5^{\circ}C$) than those in
156 the North Pacific subpolar gyre and the Antarctic Circumpolar Current (about $0.1^{\circ} - 0.2^{\circ}C$)
157 regions. The APT for this component is 18.9 yr, and the fraction of explained annual-mean
158 variance by this component is about 6.8%. The APT of this component is found to be sta-

159 tistically significant from a white noise process (see the first section of the appendix). Note
160 that APT measures the time scale of predictability in a perfect model scenario, whereas the
161 quantification of the actual predictive skill for this component remains to be verified against
162 observations. Since the component has loadings concentrated in the North Atlantic SPG
163 and has temporal variations closely following the AMO index as shown next, we refer to this
164 component as the AMO-like internal multidecadal pattern (IMP).

165 The APT value is probably underestimated, because we only integrate APT up to 10
166 year lead time in (3) due to the 10-year upper limit of hindcasts. If we had longer forecasts,
167 we would expect to obtain components with at least as much APT as found in the 10-year
168 case. The APT also could depend significantly on model: DelSole et al. (2011) found that the
169 APT estimated from pre-industrial control runs of the CMIP3 data set varied by more than a
170 factor of four (see also Branstator et al. (2012)). However, the model-to-model predictability
171 variations in decadal hindcasts could be reduced because all model states are initialized from
172 the same climate observing system, although the initialization methods vary among model
173 hindcasts. The sensitivity of APT to different model hindcasts is beyond the scope of this
174 study.

175 The time series of the component with maximum APT in the decadal hindcasts as a func-
176 tion of initialized years and lead time every five years from 1961 to 2010 are shown by shading
177 in Fig. 1b and c. To assess the forecast skill of the component, we project the detrended
178 ERSST and HadISST data onto the eigenvector \mathbf{q} with maximum APT from (3) to obtain
179 the observed time series, which are indicated by solid lines in Fig. 1b and c. A striking aspect
180 of two observed time series is the multidecadal oscillations with negative anomalies during
181 1965-95 and positive anomalies during 1925-60 and 1996-2010. Also, the two observed time
182 series are strongly correlated with the annual mean observed Atlantic multidecadal oscilla-
183 tion (AMO) index (the correlation coefficients are 0.75 and 0.77 for ERSST and HadISST
184 datasets respectively). Interestingly, the initialized forecasts of the component closely follow
185 the observed AMO index during the negative (1961-1994) and positive (1996-2010) phases,

186 and even the phase shift around 1995.

187 To facilitate comparison with statistical prediction methods, we chose a simple persis-
188 tence model, a commonly used benchmark for skill, which assumes the forecast equals the
189 initial condition. The anomaly correlation (AC) coefficients between forecasts and ERSST
190 observation as a function of lead time, shown in Fig. 2a, are statistically significant up to 4
191 year lead time at 95% significance level and up to 10 years at 90% significance level for the
192 hindcasts, while they are statistically significant up to 3 year lead time at 95% significance
193 level and up to 4 years at 90% significance level for the persistence forecasts. Verification
194 of predictions against HadISST data yields similar predictive skill (figure not shown). The
195 p-value of testing significance difference between two AC coefficients for model hindcasts
196 and persistence forecasts, shown in Fig. 2b, tends to decrease from about 0.45 to 0.15 with
197 forecast lead time. Note that the critical values and p-values were computed by adjusting
198 the effective sample size accounting for autocorrelations in the data (see the second section
199 of the appendix). Although 41-50 forecasts were used to compute the skill of forecasts at
200 given lead time, the effective sample size is generally less than 10 due to the strong auto-
201 correlations of the forecasts and observations, resulting in relatively large p-values (larger
202 than 0.1). Given these relatively large p-values, we choose 0.25 as the threshold p-value for
203 testing the difference between two correlation coefficients. We conclude that the prediction
204 skill of model hindcasts is statistically better than that of persistence forecasts for 8-10 years
205 lead times at 75% significance level, and not statistically different for 1-7 years lead times.

206 The covarying structure of the AMO-like IMP in the North Atlantic and North Pacific
207 is similar to the IMP identified from the CMIP3 multimodel unforced simulations using the
208 APT analysis (DelSole et al. 2011), which may be explained as the impact of AMO on North
209 Pacific climate variability through atmospheric teleconnections (Zhang and Delworth 2007).
210 It was also indicated in previous studies that the strong signals of SST variability associ-
211 ated with AMO exist in the North Atlantic subpolar gyre in observations as well as model
212 simulations (Folland et al. 1986; Kushnir 1994; Delworth and Mann 2000). Previous model

213 studies indicated that multidecadal Atlantic MOC variations force inter-hemispheric dipolar
214 SST anomalies (Collins et al. 2006; Latif et al. 2006b), and the out-of-phase relationship of
215 SST variations between the Antarctic Circumpolar Current region and the North Atlantic
216 subpolar gyre region associated with AMO was also revealed in observations (Latif et al.
217 2006a; Wu et al. 2011). The advance of this study is that the AMO-like IMP is not only suc-
218 cessfully identified in decadal hindcasts using APT analysis, but this pattern as a whole may
219 be *predictable* up to 4 years at 95% significance level and up to 10 years at 90% significance
220 level in GFDL’s fully-coupled ensemble initialization and decadal forecasting system.

221 To further confirm that the predictable signals are not statistical artifacts of the APT
222 technique, we directly analyze predictability of anomalous SST averaged over the North
223 Atlantic subpolar gyre region (marked as a box bounded by 50°W-10°W and 45°N-65°N
224 in Fig. 1a), where the strong predictable signals were indicated by APT analysis. The
225 time series of anomalous SST in the North Atlantic subpolar gyre region in the decadal
226 hindcasts as a function of initialized years and lead time every five years from 1961 to 2010
227 are shown by shading in Fig. 3a and b. The initialized forecasts closely follow the observations
228 during the negative (1961-1994) and positive (1996-2010) phases, and even the phase shift
229 around 1995. Also, the ensemble mean of uninitialized historical forcing simulations cannot
230 produce multidecadal variations associated with realistic AMO, indicating that the decadal
231 predictability arises from initialization. When forecasts are verified against ERSST, the SST
232 in the North Atlantic subpolar gyre region is predictable up to 4 years at 95% significance
233 level and up to 7 years at 90% significance level in the model hindcasts, while the persistence
234 model has skill only up to 3 years at 95% significance level and up to 4 years at 90%
235 significance level (Fig. 3c). Verification of model hindcasts and persistence forecasts against
236 HadISST data yields similar predictive skill (figure not shown). The predictive skill of model
237 hindcasts is statistically better than that of persistence forecasts for 3-10 years lead times at
238 75% significance level, and they are not statistically different for 1-2 years lead times. This
239 local predictability analysis confirms the robustness of the APT diagnosis for identifying the

240 AMO-like IMP.

241 *b. The IMP of T2m*

242 The skill in predicting the IMP of SST on decadal timescales would imply skill in predict-
243 ing the air surface temperature as a response to SST forcing, thus we apply the same APT
244 analysis as described above to the *internal* residuals of the annual mean 2-m temperature.
245 The component with maximum APT in decadal hindcasts is shown in Fig. 4a. The pat-
246 tern is predominantly of a general bi-polar seesaw with warm anomalies extending from the
247 North Atlantic subpolar gyre region to Greenland and the Arctic Ocean, and cold anomalies
248 centered in Antarctica. The APT value for this component is 17.8 yr, and the fraction of
249 explained annual-mean variance by this component is about 4.5%. The APT of this compo-
250 nent is found to be statistically significant. We project the detrended NNR and 20CR data
251 onto the leading eigenvector \mathbf{q} for T2m from (3) to obtain the observed time series, shown
252 in Fig. 4b and c. Two observed time series for the T2m pattern show similar multidecadal
253 variations in phase with AMO as those for the SST IMP, e.g., negative anomalies during
254 1965-95 and positive anomalies during 1925-60 and 1996-2010, suggesting the multidecadal
255 variability of T2m arises from multidecadal SST variations. The anomaly correlation coeffi-
256 cients between forecasts and observations as a function of lead time, shown in Fig. 5, indicate
257 that the T2m pattern as a whole may be predictable up to 2 years at 95% confidence level
258 and up to 10 years at 90% significance level verified against the NNR data (except the 6
259 year lead time). The p-values of testing significance difference between 6 year lead time AC
260 coefficient and 8-10 year lead AC coefficients are larger than 0.3, so the nominal increase of
261 AC coefficients after 6 year lead time could be due to sampling errors. Verification of the
262 same predictions against 20CR data yields similar predictive skill (figure not shown). The
263 predictive skill of model hindcasts is statistically better than that of persistence forecasts for
264 7-10 years lead times at 75% significance level, and not statistically different for 1-6 years
265 lead times (Fig. 5).

266 Although the so-called “bi-polar seesaw” pattern has been observed in the Greenland and
267 Antarctica ice core data on millennial time scales (Blunier and Brook 2001; Blunier et al.
268 1998), Chylek et al. (2010) identified a bi-polar seesaw pattern in the 20th century Arctic and
269 Antarctic instrumental temperature records on multidecadal time scales, and found that the
270 Arctic (Antarctic) de-trended temperatures are highly correlated (anti-correlated) with the
271 Atlantic Multidecadal Oscillation (AMO) index suggesting the Atlantic Ocean as a possible
272 link between the climate variability of the Arctic and Antarctic regions. Our results not
273 only confirmed the existence of the bi-polar seesaw pattern on multidecadal time scales in
274 phase with AMO, but progressed towards predicting the pattern as a whole on multiyear
275 time scales in GFDL’s fully-coupled ensemble initialization and forecasting system.

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

277 Previous studies with coupled models assuming perfect ocean initial conditions indicate
278 that accurate initialization of Atlantic MOC may allow Atlantic multidecadal variability to
279 be predicted a decade or more in advance (Collins et al. 2006; Msadek et al. 2010), and the
280 AMO is suggested to be induced by Atlantic MOC variations and associated oceanic heat
281 transport fluctuations (Knight et al. 2005), so we examine the North Atlantic oceanic heat
282 transport in ECDA and decadal hindcasts. The regression coefficients of North Atlantic
283 northward oceanic heat transport between 30°N and 70°N onto the time series of normalized
284 North Atlantic subpolar gyre region SST anomalies for ECDA, shown in Fig. 6a, tend to
285 be positive, indicating that ECDA captures the anomalous northward (southward) heat
286 transport into the North Atlantic subpolar gyre associated with the positive (negative) phase
287 of AMO. These results indicate that the heat transport variations in the North Atlantic are
288 in phase with AMO in both ECDA and hindcasts, which is consistent with that AMOC
289 variations are in phase with the observed AMO using observational and modeling results
290 (Zhang 2007, 2008).

291 Consequently, the decadal hindcasts initialized by ECDA also show similar anomalous

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

305 The forecast anomalies of the North Atlantic SST and North Atlantic oceanic heat trans-
306 port averaged between 35°N and 65°N for the fixed-forcing decadal prediction experiments
307 and the decadal hindcasts are shown in Fig. 8. The fixed-forcing forecasts of the North
308 Atlantic SST show similar multidecadal variations as the decadal hindcasts as well as the
309 phase transition around 1995 (Fig. 8a), indicating that the internal signals due to initializa-
310 tion dominate over the forced signal on the decadal time scale in the model. Meanwhile, the
311 fixed-forcing forecasts of the North Atlantic SST initialized in 1995 and 2005 show a cooling
312 during about 3 to 10 year lead time resulting from the return of radiative forcing to 1961
313 values. This is similar to the fast cooling of global mean temperature as a response to an
314 instantaneous return to preindustrial forcing (Held et al. 2010), but the initial decay rate of
315 the North Atlantic SST is much slower than that of the global mean SST (Fig. 8c). Note
316 that the global mean SST in the fixed-forcing experiments shows a similar fast cooling as the
317 global mean temperature in the experiments by returning radiative forcing to preindustrial
318 values (Held et al. 2010). The slower decay of North Atlantic SST than the global mean SST

319 is due to the fact that the northward heat transports in the North Atlantic evolve similarly
320 during the 10 yr forecasts in the two experiments (Fig. 8b), suggesting that the internal
321 oceanic heat transport compensates the fast cooling of the North Atlantic SST. The fixed-
322 forcing experiments further demonstrate that the ocean internal dynamics play dominant
323 roles in decadal prediction over the external forcings.

324 The latest forecast shows the AMO-like SST pattern is weakening in the coming decade
325 2011-2020 but still in the warm phase (Fig. 1). Consistently, the bi-polar seesaw pattern
326 of T2m is predicted to be weakening but still above the climatology in the coming decade
327 (Fig. 4). The dynamical basis for these forecasts is that the North Atlantic northward heat
328 transport is predicted to be weakening (Fig. 6). The predicted weakening of AMO-like
329 patterns in the coming decade by using dynamical models is consistent with the statistically
330 predicted weakening of AMOC using subsurface and surface fingerprints (Mahajan et al.
331 2011).

332 4. Conclusions and discussions

333 The decadal predictability of SST and T2m in GFDL's decadal hindcasts has been in-
334 vestigated using the APT analysis. By diagnosing the internal residuals between *initialized*
335 hindcasts and *uninitialized* historical forcing simulations using the the same model (GFDL
336 CM2.1), internal multidecadal patterns (IMP) for SST and T2m were identified. The IMP
337 of SST is predominantly of an inter-hemisphere dipole with warm anomalies centered in
338 the North Atlantic subpolar gyre region and North Pacific subpolar gyre region, and cold
339 anomalies centered in the Antarctic Circumpolar Current region. The IMP of T2m is pre-
340 dominantly of a general bi-polar seesaw with warm anomalies extending from the North
341 Atlantic subpolar gyre region to Greenland and the Arctic Ocean, and cold anomalies cen-
342 tered in Antarctica.

343 The projected time series of IMP onto observations closely follow the multidecadal oscil-

344 lations, with negative anomalies during 1965-95 and positive anomalies during 1925-60 and
345 1996-2010, in phase with the observed AMO index. The retrospective prediction skills of the
346 IMP for SST and T2m were verified against independent observational datasets, revealing
347 that the SST pattern is *predictable* up to 4 year lead time at 95% significance level and up to
348 10 years at 90% significance level, and the T2m pattern is predictable up to 2 year lead time
349 at 95% significance level and up to 10 years at 90% significance level. Further analysis sug-
350 gests that the initialization of multidecadal variations of northward oceanic heat transport
351 in the North Atlantic significantly improves the predictive skill of the AMO-like IMP. The
352 fixed-forcing decadal prediction experiments, in which radiative forcing is returned abruptly
353 to 1961 values, further elucidate roles of ocean internal dynamics in the decadal prediction.

354 Beyond the robust long-term anthropogenic signal predicted by uninitialized CMIP3 cli-
355 mate models (Hegerl et al. 2007), the results presented here from GFDL’s CMIP5 decadal
356 hindcasts show that the coupled climate model initialized by advanced data assimilation
357 methods, may be capable of predicting AMO-like internal multidecadal pattern over mul-
358 tiyears to a decade, thus pointing towards the possibility of meaningful decadal climate
359 outlooks using dynamical models if they are appropriately initialized by the climate observ-
360 ing system. However, the predictable signal of SST is primarily located in the North Atlantic
361 subpolar gyre, and not a basin-scale signal observed in the North Atlantic. The predictable
362 signals of T2m mainly are in two polar regions, but not over other continents indicated by
363 previous studies (Hermanson and Sutton 2010). This discrepancy may be attributable to
364 model biases and lack of deep-ocean observations in the 20th century. Another possibility
365 is that the continental predictability may be related to some other SST pattern that is less
366 predictable than the AMO-like pattern (Jia and DelSole 2011).

367 A challenge for decadal prediction is the lack of many decadal-scale events (e.g., AMO
368 transitions) for assessing its reliability. In contrast, there are many realistic ENSO cases
369 for assessing seasonal climate prediction. In this study, the hindcast period from 1961 to
370 2011 covers only one AMO episode, so the reliability of the predictive skill presented here

371 needs to be evaluated by hindcasts covering more AMO episodes. More research is needed
372 on producing a larger set of initial conditions covering more AMO episodes for assessing
373 decadal climate prediction. Nevertheless, the skill of the dynamical model prediction is
374 consistently better than that of the persistence forecasts, especially for forecast lead times
375 longer than 5 years, thus the results presented here are encouraging for CMIP5 decadal
376 prediction initiatives using initialized dynamical models.

377 *Acknowledgments.*

378 We thank Isaac Held, Takeshi Doi for helpful reviews of an earlier draft, and three
379 anonymous reviewers for constructive comments that helped to improve the manuscript.
380 This research was supported by the Visiting Scientist Program at the National Oceanic and
381 Atmospheric Administration's Geophysical Fluid Dynamics Laboratory, administered by the
382 University Corporation for Atmospheric Research.

383

384

385

Details of calculation

386 *a. APT*

387 For solving the APT optimization problem (Equ. (2) in the main paper) in practice, the
388 data are first projected onto the leading principal components (PC) (DelSole et al. 2011).
389 The sensitivity of APT to the truncation of PCs for SST, shown in Fig. 9, indicates that
390 the results for the time series of IMP and projection of IMP onto observations are virtually
391 independent of the number of PCs in the range of 20-40 PCs, presumably because the time
392 series of 5100 years are relatively long. We obtained similar results for the APT analysis of
393 T2m (figures not shown). We choose 30 EOFs for displaying results for both SST and T2m
394 in the paper.

395 Following (DelSole et al. 2011), the statistical significance test of APT was estimated
396 by Monte Carlo methods. The null hypothesis for the test is that the data are drawn from
397 a white noise process when sampled every 2 years. Accordingly, we generate a 30×2500
398 data matrix by drawing independent random numbers from a normal distribution with zero
399 mean and unit variance. The time dimension of the data was grouped as a set of 25 10-yr
400 forecasts with 10 ensemble members. 30 APT values were then determined. This procedure
401 was repeated 1000 times to generate 1000×30 APT values. The upper fifth percentile of
402 the 1000×30 APT values was then determined as the threshold values, plotted in Fig. 10 as
403 the horizontal line. The figure shows that the first 25 (10) components of SST (t2m) have
404 statistically significant APT values. However, only the leading component has multi-year
405 predictive skill verified against observations, so we only focus on the leading component in
406 this study.

407 *b. Statistical tests*

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

414 Autocorrelation in the data was accounted for by adjusting the effective sample size (N^*)
 415 using the following procedure (Trenberth 1984; Bretherton et al. 1999):

$$416 \quad N^* = \frac{N}{1 + 2 \sum_{j=1}^{N-1} (1 - \frac{j}{N}) \rho_{xx}(j) \rho_{yy}(j)} \quad (A1)$$

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

REFERENCES

- 425 Blunier, T. and E. Brook, 2001: Timing of millennial-scale climate change in Antarctica and
426 Greenland during the last glacial period. *Science*, **291**, 109–112, doi:10.1126/science.291.
427 5501.109.
- 428 Blunier, T., et al., 1998: Asynchrony of Antarctic and Greenland climate change during the
429 last glacial period. *Nature*, **394**, 739–743, doi:10.1038/29447.
- 430 Branstator, G., H. Teng, G. A. Meehl, M. Kimoto, J. R. Knight, M. Latif, and A. Rosati,
431 2012: Systematic estimates of initial-value decadal predictability for six AOGCMs. *J.*
432 *Clim.*, **25**, 1827–1846.
- 433 Bretherton, C. S., M. Widmann, V. P. Dymnikov, J. M. Wallace, and I. Bladè, 1999: The
434 effective number of spatial degrees of freedom of a time-varying field. *J. Clim.*, **12**, 1990–
435 2009.
- 436 Chang, Y.-S., S. Zhang, A. Rosati, T. L. Delworth, and W. F. Stern, 2012: An assessment
437 of oceanic variability for 1960–2010 from the GFDL ensemble coupled data assimilation.
438 *Clim. Dyn.*, **In press**, doi:10.1007/s00382-012-1412-2.
- 439 Chylek, P., C. K. Folland, G. Lesins, and M. K. Dubey, 2010: Twentieth century bipolar
440 seesaw of the Arctic and Antarctic surface air temperatures. *Geophys. Res. Lett.*, **37**,
441 L08703, doi:10.1029/2010GL042793.
- 442 Collins, M., et al., 2006: Interannual to decadal climate predictability in the North Atlantic:
443 A multimodel-ensemble study. *J. Clim.*, **19**, 1195–1203.
- 444 Compo, G., et al., 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Me-*
445 *teor. Soc.*, **137**, 1–28.

- 446 DelSole, T. and M. K. Tippett, 2009a: Average predictability time. part I: Theory. *J. Atmos.*
447 *Sci.*, **66**, 1172–1187, doi:10.1175/2008JAS2868.1.
- 448 DelSole, T. and M. K. Tippett, 2009b: Average predictability time. part II: seamless di-
449 agnosis of predictability on multiple time scales. *J. Atmos. Sci.*, **66**, 1188–1204, doi:
450 10.1175/2008JAS2869.1.
- 451 DelSole, T., M. K. Tippett, and J. Shukla, 2011: A significant component of unforced
452 multidecadal variability in the recent acceleration of global warming. *J. Clim.*, **24**, 909–
453 926.
- 454 Delworth, T. L. and M. E. Mann, 2000: Observed and simulated multidecadal variability in
455 the Northern Hemisphere. *Clim. Dyn.*, **16**, 661–676.
- 456 Delworth, T. L., et al., 2006: GFDL’s CM2 global coupled climate models. part I: Formula-
457 tion and simulation characteristics. *J. Clim.*, **19**, 643–674.
- 458 Enfield, D. B., A. M. Mestas-Nuñez, and P. J. Trimble, 2001: The Atlantic Multidecadal
459 Oscillation and its relation to rainfall and river flows in the continental U.S. *Geophys. Res.*
460 *Lett.*, **28**, 2077–2080, doi:10.1029/2000GL012745.
- 461 Folland, C. K., T. N. Palmer, and D. E. Parker, 1986: Sahel rainfall and worldwide sea
462 temperatures, 1901–85. *Nature*, **320**, 602–607.
- 463 Hegerl, G., et al., 2007: Understanding and attributing climate change. *Climate Change*
464 *2007: The Physical Science Basis*, S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis,
465 K. Averyt, M. Tignor, and H. Miller, Eds., Cambridge University Press, 663–745.
- 466 Held, I. M., M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis, 2010: Probing
467 the fast and slow components of global warming by returning abruptly to preindustrial
468 forcing. *J. Climate*, **23**, 2418–2417.

- 469 Hermanson, L. and R. Sutton, 2010: Case studies in interannual to decadal climate pre-
470 dictability. *Clim. Dyn.*, **35**, 1169–1189.
- 471 Jia, L. and T. DelSole, 2011: Diagnosis of multiyear predictability on continental scales. *J.*
472 *Climate*, **24**, 5108–5124.
- 473 Kalnay, E. and Coauthors, 1996: The NCEP/NCAR 40-Year Re-analysis Project.
474 *Bull. Am. Meteor. Soc.*, **77**, 437–471.
- 475 Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan,
476 1998: Analyses of global sea surface temperature 1856-1991. *J Geophys. Res.-Oceans*, **103**
477 **(C9)**, 18 567–18 589.
- 478 Keenlyside, N. S., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner, 2008: Advancing
479 decadal-scale climate prediction in the North Atlantic sector. *Nature*, **453 (7191)**, 84–88,
480 doi:10.1038/nature06921.
- 481 Kistler, R., et al., 2001: The NCEP/NCAR 50-Year Reanalysis: Monthly Means CD-ROM
482 and Documentation. *Bulletin of the American Meteorological Society*, **82 (2)**, 247–267.
- 483 Knight, J. R., R. J. Allan, C. K. Folland, M. Vellinga, and M. E. Mann, 2005: A signature
484 of persistent natural thermohaline circulation cycles in observed climate. *Geophys. Res.*
485 *Lett.*, **32**, L20708, doi:10.1029/2005GL024233.
- 486 Knutson, T. R., et al., 2006: Assessment of twentieth-century regional surface temperature
487 trends using the GFDL CM2 coupled models. *J. Clim.*, **19**, 1624–1651.
- 488 Kushnir, Y., 1994: Interdecadal variations in North Atlantic sea surface temperature and
489 associated atmospheric conditions. *J. Clim.*, **7**, 141–157.
- 490 Latif, M., C. Böning, J. Willebrand, A. Biastoch, J. Dengg, N. Keenlyside, U. Schweckendiek,
491 and G. Madec, 2006a: Is the thermohaline circulation changing? *J. Clim.*, **19**, 4631–4637.

- 492 Latif, M., M. Collins, H. Pohlmann, and N. Keenlyside, 2006b: A review of predictability
493 studies of Atlantic sector climate on decadal time scales. *J. Clim.*, **19**, 5971–5987.
- 494 Mahajan, S., R. Zhang, T. L. Delworth, S. Zhang, A. Rosati, and Y.-S. Chang, 2011: Pre-
495 dicting Atlantic meridional overturning circulation (AMOC) variations using subsurface
496 and surface fingerprints. *Deep-Sea Research, Part II*, **58**, 1895 – 1903.
- 497 Meehl, G. A., et al., 2009: Decadal prediction: Can it be skillful? *Bull. Amer. Meteor. Soc.*,
498 **90**, 1467–1485, doi:10.1175/2009bams2778.1.
- 499 Msadek, R., K. W. Dixon, T. L. Delworth, and W. Hurlin, 2010: Assessing the predictability
500 of the Atlantic meridional overturning circulation and associated fingerprints. *Geophys.*
501 *Res. Lett.*, **37**, L19608, doi:10.1029/2010GL044517.
- 502 Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, and
503 A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine
504 air temperature since the late nineteenth century. *J. Geophys. Res.*, **108** (D14), 4407,
505 29PP., doi:10.1029/2002JD002670.
- 506 Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy,
507 2007: Improved surface temperature prediction for the coming decade from a global climate
508 model. *Science*, **317** (5839), 796–799, doi:10.1126/science.1139540.
- 509 Smith, T. M. and R. W. Reynolds, 2004: Improved extended reconstruction of SST (1854-
510 1997). *J. Clim.*, **17**, 2466–2477.
- 511 Solomon, A., et al., 2011: Distinguishing the roles of natural and anthropogenically forced
512 decadal climate variability. *Bull. Amer. Meteor. Soc.*, **92**, 141–156.
- 513 Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of the CMIP5 and the
514 experiment design. *Bull. Amer. Meteor. Soc.*, doi:10.1175/BAMS-D-11-00094.1.

515 Ting, M., Y. Kushnir, R. Seager, and C. Li, 2011: Robust features of Atlantic multi-
516 decadal variability and its climate impacts. *Geophys. Res. Lett.*, **38**, L17705, doi:
517 10.1029/2011GL048712.

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

520 van Oldenborgh, G. J., F. J. Doblas-Reyes, B. Wouters, and W. Hazeleger, 2012: Decadal
521 prediction skill in a multi-model ensemble. *Clim. Dyn.*, **38**, 1263–1280, doi:10.1007/
522 s00382-012-1313-4.

523 Wu, Z., N. E. Huang, J. M. Wallace, B. V. Smoliak, and X. Chen, 2011: On the time-varying
524 trend in global-mean surface temperature. *Clim. Dyn.*, **37**, 759–773.

525 Zhang, R., 2007: Anticorrelated multidecadal variations between surface and subsurface
526 tropical North Atlantic. *Geophys. Res. Lett.*, **34**, L12 713, doi:10.1029/2007GL030225.

527 Zhang, R., 2008: Coherent surface-subsurface fingerprint of the Atlantic meridional over-
528 turning circulation. *Geophys. Res. Lett.*, **35**, L20 705, doi:10.1029/2008GL035463.

529 Zhang, R. and T. L. Delworth, 2006: Impact of atlantic multidecadal oscillations on
530 India/Sahel rainfall and Atlantic hurricanes. *Geophys. Res. Lett.*, **33**, L17712, doi:
531 10.1029/2006GL026267.

532 Zhang, R. and T. L. Delworth, 2007: Impact of the Atlantic Multidecadal Oscillation
533 on North Pacific climate variability. *Geophys. Res. Lett.*, **34**, L23708, doi:10.1029/2007/
534 GL031601.

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

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

540 List of Figures

- 541 1 **a**, The spatial structure of the component that maximized the average pre-
542 dictability time of SST in the decadal hindcasts, which is called IMP. **b**, The
543 ensemble mean (black solid) and spread (gray shading) time series of IMP
544 as a function of forecast lead time for the decadal hindcasts initialized on
545 1 January every 10 years from 1965 to 2005, the time series for projecting
546 the ERSST data onto IMP (red solid) and the normalized AMO index (blue
547 solid) from 1920 to 2010. **c**, Same as **b** but for hindcasts initialized on 1
548 January 1961 and every 10 years from 1970 to 2010. The green line denotes
549 the projected time series of HadISST data onto IMP. 28
- 550 2 **a**, The anomaly correlation (AC) coefficients between forecasts and observa-
551 tions for the model hindcasts (black dots) and the persistence forecasts (green
552 squares) of the SST IMP verified by the ERSST observations as a function
553 of the forecast lead time. The dashed and dotted lines indicate the critical
554 values at the 95% and 90% significance levels of AC respectively for model
555 hindcasts (black) and persistence forecasts (green). **b**, the p-value from the
556 statistical significance test that the two anomaly correlation coefficients for
557 model hindcasts and persistence forecasts are not different from each other.
558 The dashed line denotes the p-value of 0.25. 29

559 3 **a**, The ensemble mean (black solid) and spread (gray shading) time series of
560 SST internal residuals in the North Atlantic SPG region (the green boxed
561 region in Fig. 1) as a function of forecast lead time for the decadal hindcasts
562 initialized on 1 January every 10 years from 1965 to 2005, the time series for
563 the ERSST (red) data from 1920 to 2010, and the time series of the historical
564 forcing simulations (blue) are plotted for reference. **b**, The same as **a** but
565 for hindcasts initialized on 1 January 1961 and every 10 years from 1970 to
566 2010. **c**, The anomaly correlation (AC) coefficients for the model hindcasts
567 (black dots) and the persistence forecasts (green squares) verified by ERSST
568 observations as a function of the forecast lead time. The dashed and dotted
569 lines indicate the critical values at 95% and 90% significance levels of AC
570 respectively for model hindcasts (black) and persistence forecasts (green). **d**,
571 The p-value of the statistical test that the two anomaly correlation coefficients
572 for model hindcasts and persistence forecasts are not different from each other.
573 The dashed line denotes the p-value of 0.25. 30

574 4 **a**, The spatial structure of IMP for the 2-m temperature in the decadal hind-
575 casts. **b**, The ensemble mean (black solid) and spread (gray shading) time
576 series of IMP as a function of forecast lead time for the decadal hindcasts
577 initialized on 1 January every 10 years from 1965 to 2005, the time series for
578 projecting the NNR data onto IMP (red solid) and the normalized AMO index
579 (blue solid) from 1920 to 2010. **c**, Same as **b** but for hindcasts initialized on
580 1 January 1961 and every 10 years from 1970 to 2010. The green line denotes
581 the projected time series of 20CR data onto IMP. 31

- 582 5 **a**, The anomaly correlation (AC) coefficients between forecasts and observa-
583 tions for the model hindcasts (black dots) and the persistence forecasts (green
584 squares) of the T2m IMP verified by the NNR observations as a function of
585 the forecast lead time. The dashed and dotted lines indicate the critical values
586 at the 95% and 90% significance levels of AC respectively for model hindcasts
587 (black) and persistence forecasts (green). **b**, the p-value from the statistical
588 significance test that the two anomaly correlation coefficients for model hind-
589 casts and persistence forecasts are not different from each other. The dashed
590 line denotes the p-value of 0.25. 32
- 591 6 **a**, The regression coefficients of anomalous northward oceanic heat transport
592 (PW) from 30⁰N to 70⁰N in the North Atlantic onto the normalized time series
593 of the North Atlantic subpolar gyre SST anomalies for decadal hindcasts and
594 ECDA. **b**, The ensemble mean (black solid) and spread (gray shading) time
595 series of the anomalous oceanic heat transport averaged over the latitude belt
596 between 35⁰N and 65⁰N as a function of forecast lead time for the decadal
597 hindcasts initialized on 1 January every 10 years from 1965 to 2005, and the
598 time series for the historical forcing simulations (blue) and ECDA (red) are
599 plotted for reference. **c**, The same as **b** but for hindcasts initialized on 1
600 January 1961 and every 10 years from 1970 to 2010. The + marker in **a**
601 indicates the regression coefficient is significant at 90% level. 33
- 602 7 **a**, The anomaly correlation (AC) coefficients between forecasts and ECDA for
603 the model hindcasts (black dots) and the historical forcing simulations (green
604 squares) of the anomalous oceanic heat transport averaged over the latitude
605 belt between 35⁰N and 65⁰N in the North Atlantic as a function of the forecast
606 lead time. The dashed and dotted lines indicate the critical values at the 95%
607 and 90% significance levels of AC respectively for model hindcasts (black) and
608 historical forcing simulations (green). 34

609	8	<p>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, Same as a but for the anomalous oceanic heat transport averaged over the latitude belt between 35⁰N and 65⁰N in the North Atlantic. c, Same as a but for the anomalous global mean SST.</p>	35
610			
611			
612			
613			
614			
615	9	<p>a, The ensemble mean time series of IMP as a function of forecast lead time for the decadal hindcasts initialized on 1 January 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.</p>	36
616			
617			
618			
619			
620	10	<p>The APT values using 30 leading PCs for SST and T2m. Solid horizontal line is the 5% significance level of the APT values.</p>	37
621			

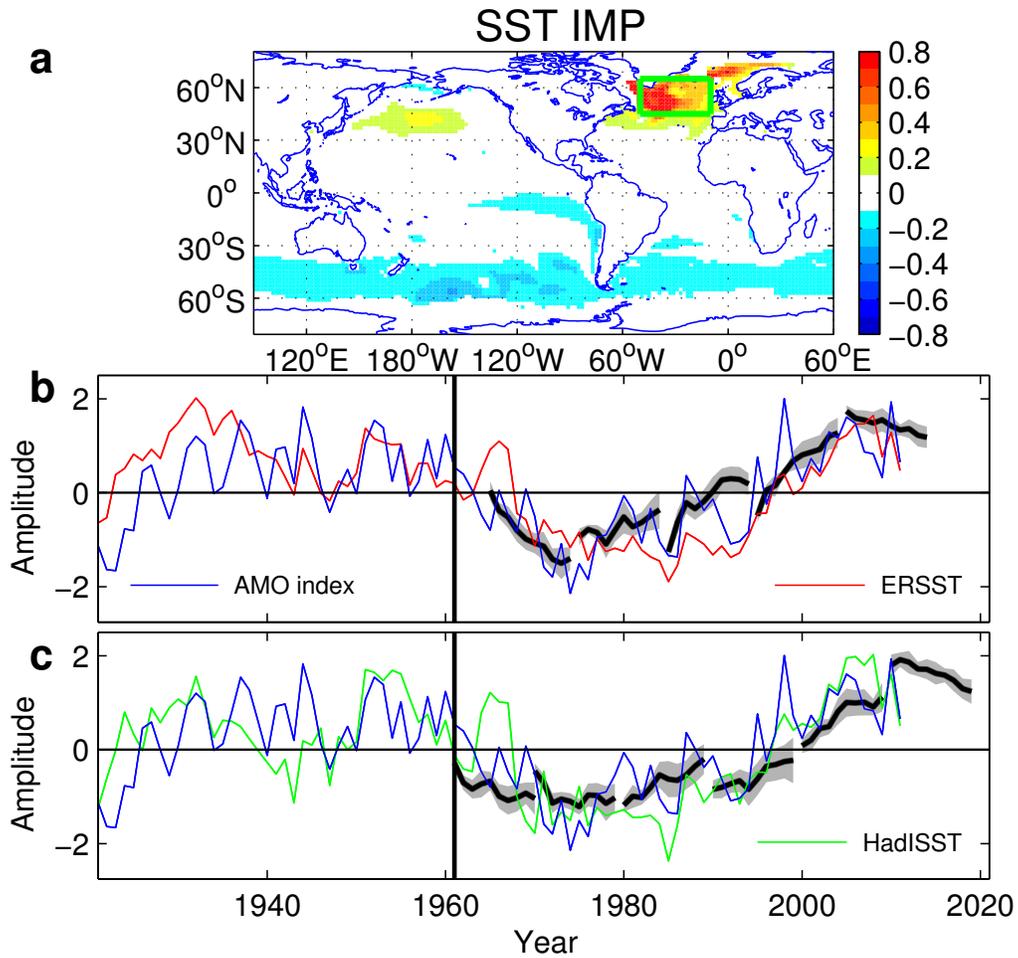


FIG. 1. **a**, The spatial structure of the component that maximized the average predictability time of SST in the decadal hindcasts, which is called IMP. **b**, The ensemble mean (black solid) and spread (gray shading) time series of IMP as a function of forecast lead time for the decadal hindcasts initialized on 1 January every 10 years 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**, Same as **b** but for hindcasts initialized on 1 January 1961 and every 10 years from 1970 to 2010. The green line denotes the projected time series of HadISST data onto IMP.

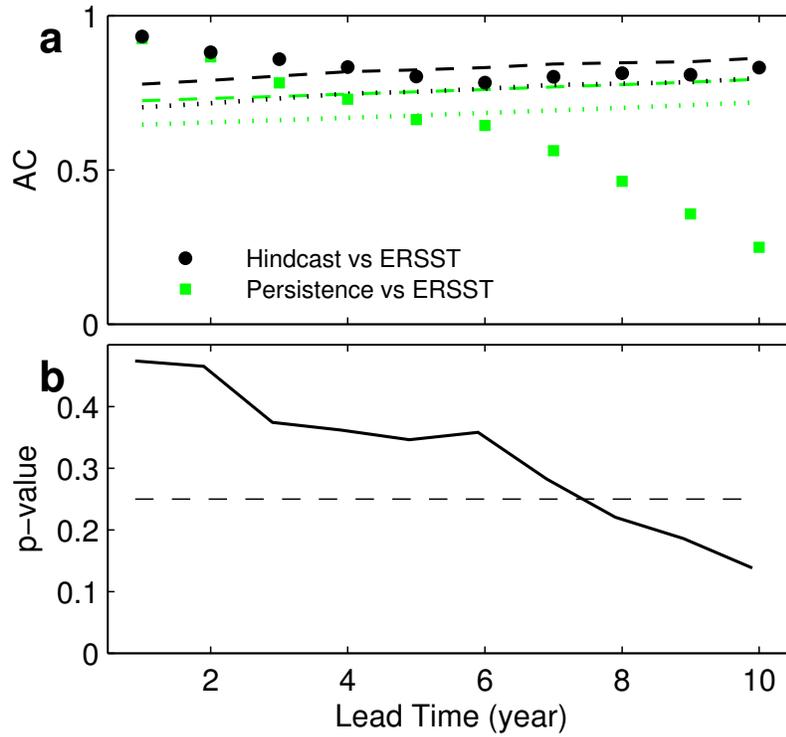


FIG. 2. **a**, The anomaly correlation (AC) coefficients between forecasts and observations for the model hindcasts (black dots) and the persistence forecasts (green squares) of the SST IMP verified by the ERSST observations as a function of the forecast lead time. The dashed and dotted lines indicate 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 of 0.25.

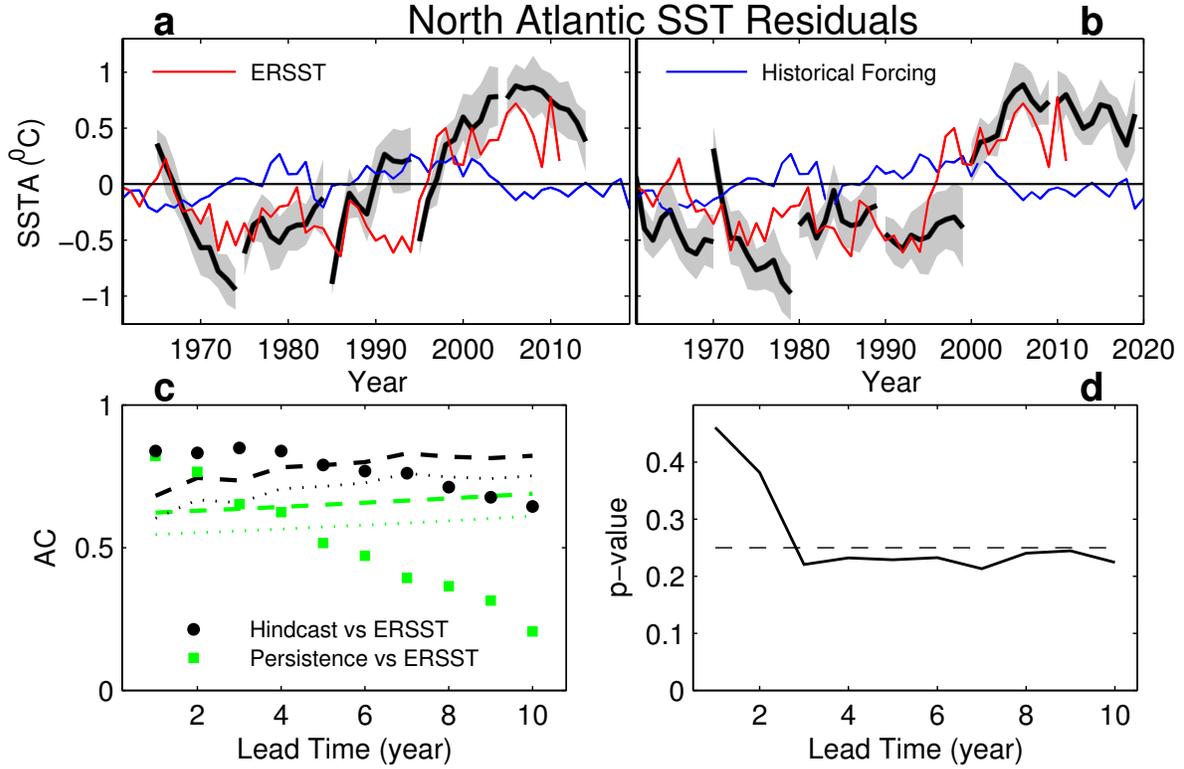


FIG. 3. **a**, The ensemble mean (black solid) and spread (gray shading) time series of SST internal residuals in the North Atlantic SPG region (the green boxed region in Fig. 1) as a function of forecast lead time for the decadal hindcasts initialized on 1 January every 10 years from 1965 to 2005, the 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**, The same as **a** but for hindcasts initialized on 1 January 1961 and every 10 years from 1970 to 2010. **c**, The anomaly correlation (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 of 0.25.

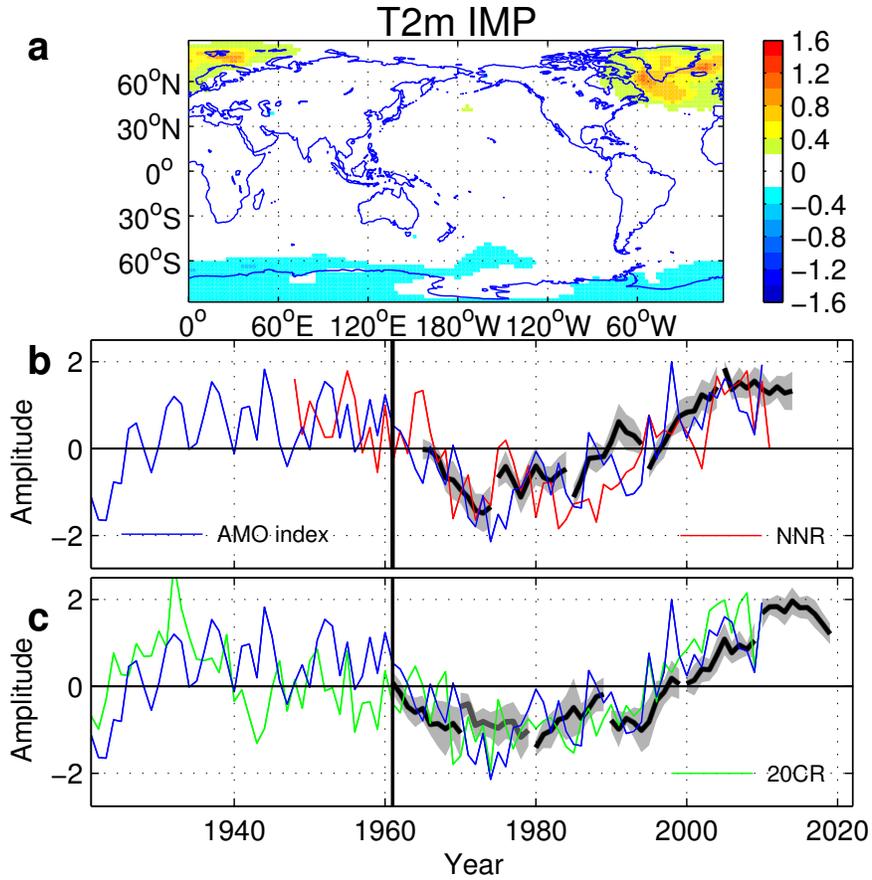


FIG. 4. **a**, The spatial structure of IMP for the 2-m temperature in the decadal hindcasts. **b**, The ensemble mean (black solid) and spread (gray shading) time series of IMP as a function of forecast lead time for the decadal hindcasts initialized on 1 January every 10 years from 1965 to 2005, the time series for projecting the NNR data onto IMP (red solid) and the normalized AMO index (blue solid) from 1920 to 2010. **c**, Same as **b** but for hindcasts initialized on 1 January 1961 and every 10 years from 1970 to 2010. The green line denotes the projected time series of 20CR data onto IMP.

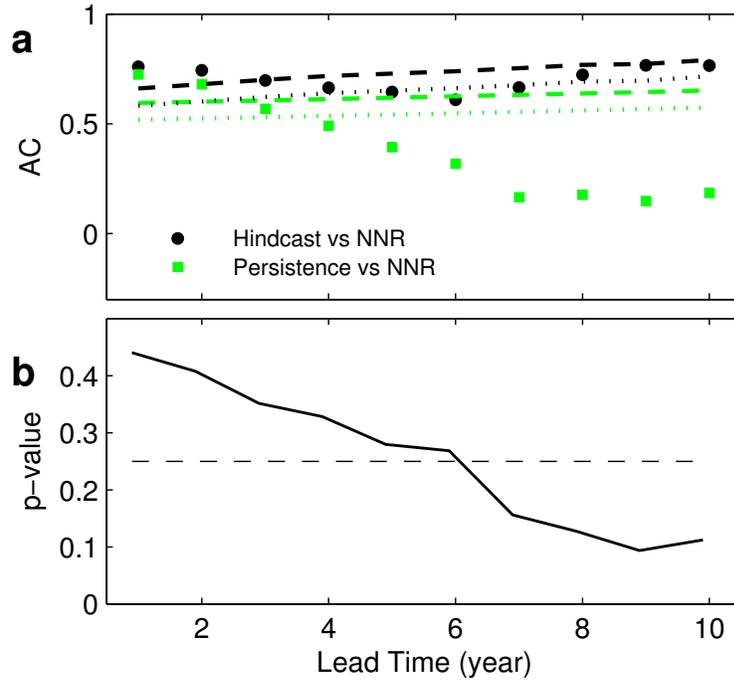


FIG. 5. **a**, The anomaly correlation (AC) coefficients between forecasts and observations for the model hindcasts (black dots) and the persistence forecasts (green squares) of the T2m IMP verified by the NNR observations as a function of the forecast lead time. The dashed and dotted lines indicate 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 of 0.25.

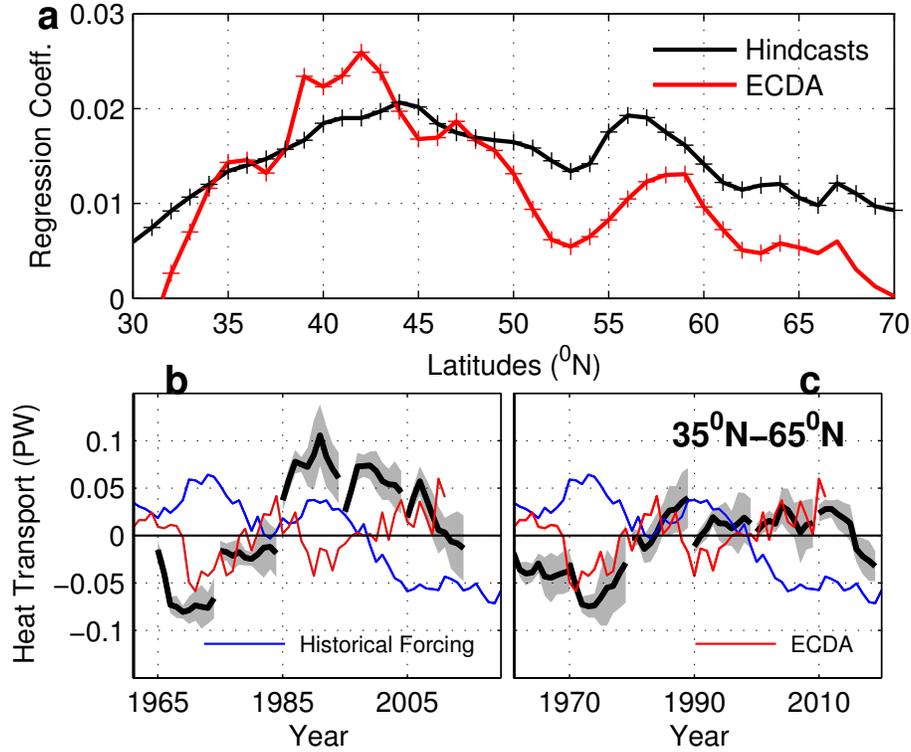


FIG. 6. **a**, The regression coefficients of anomalous northward oceanic heat transport (PW) from 30° N to 70° N in the North Atlantic onto the normalized time series of the North Atlantic subpolar gyre SST anomalies for decadal hindcasts and ECDA. **b**, The ensemble mean (black solid) and spread (gray shading) time series of the anomalous oceanic heat transport averaged over the latitude belt between 35° N and 65° N as a function of forecast lead time for the decadal hindcasts initialized on 1 January 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**, The same as **b** but for hindcasts initialized on 1 January 1961 and every 10 years from 1970 to 2010. The + marker in **a** indicates the regression coefficient is significant at 90% level.

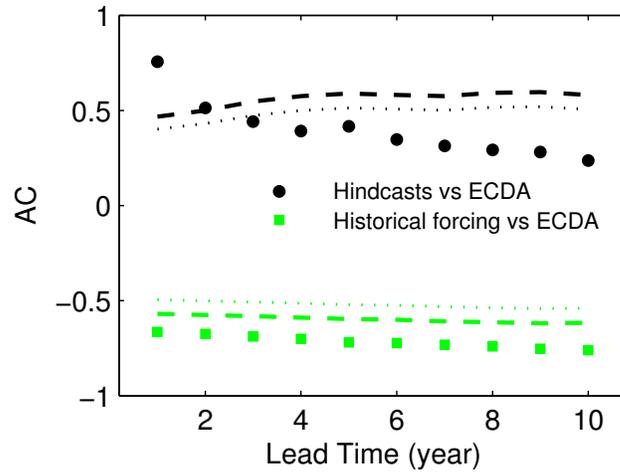


FIG. 7. **a**, The anomaly correlation (AC) coefficients between forecasts and ECDA for the model hindcasts (black dots) and the historical forcing simulations (green squares) of the anomalous oceanic heat transport averaged over the latitude belt between 35°N and 65°N in the North Atlantic as a function 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).

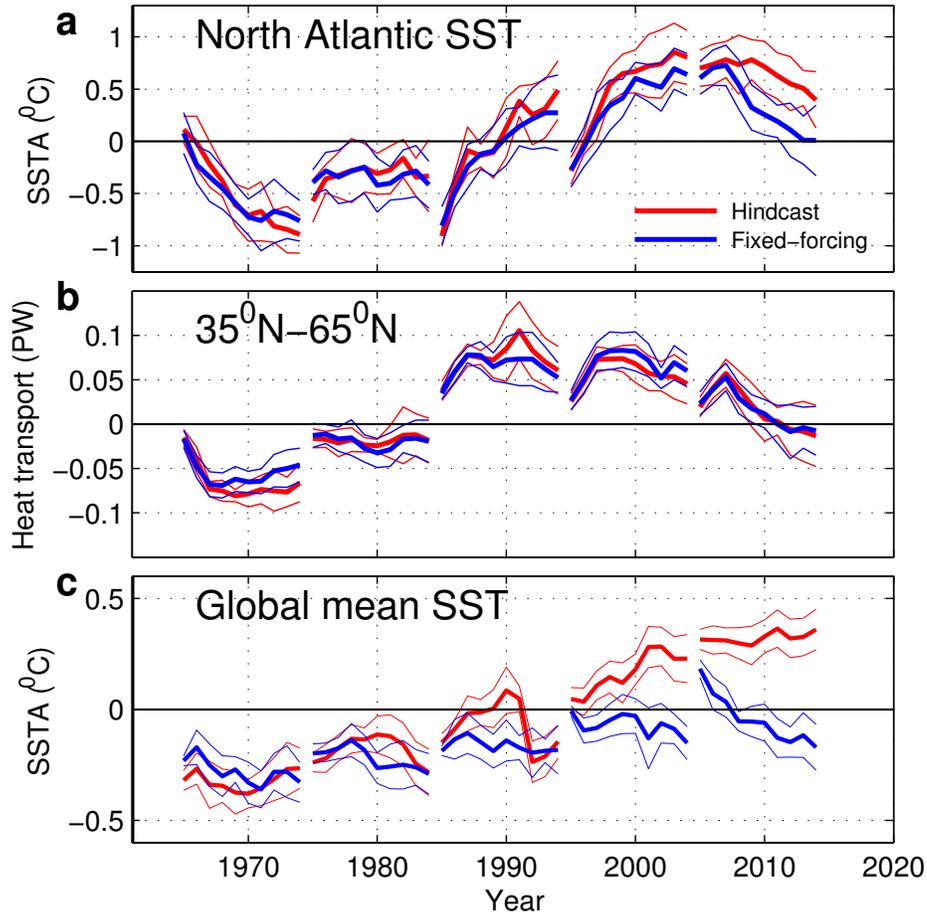


FIG. 8. **a**, The ensemble mean (thick) and spread (thin) time series of the anomalous SST in the North Atlantic subpolar gyre region as a function of forecast lead time for the decadal hindcasts (red) and fixed-forcing (blue) experiments initialized every 10 years from 1965 to 2005. **b**, Same as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

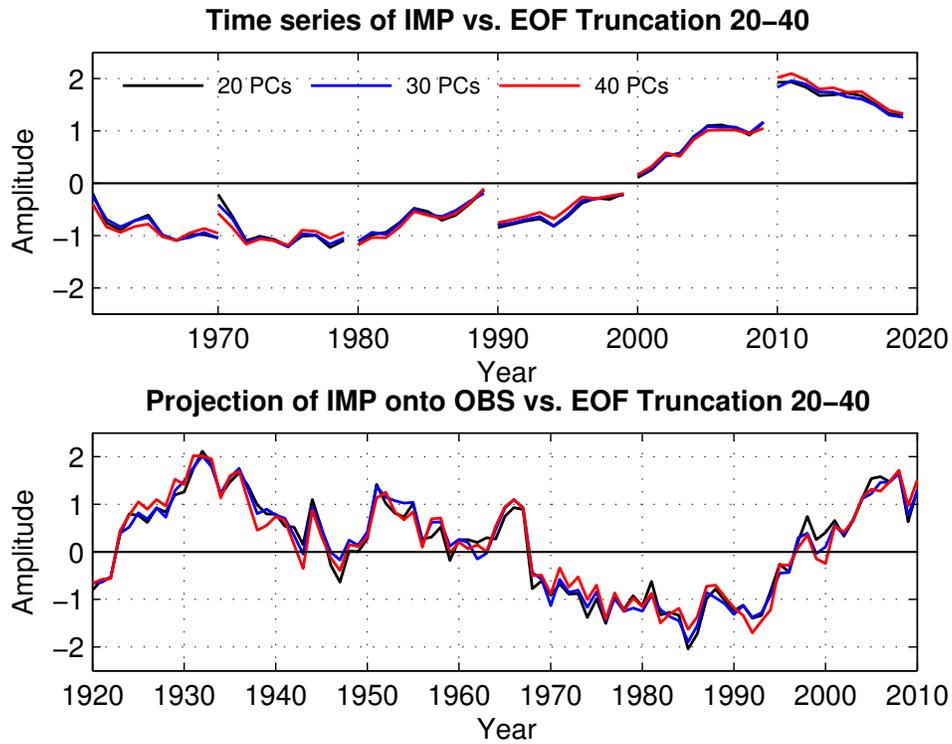


FIG. 9. **a**, The ensemble mean time series of IMP as a function of forecast lead time for the decadal hindcasts initialized on 1 January 1961 and every 10 years from 1970 to 2010 using 20, 30 and 40 leading PCs. **b**, the time series for projecting the ERSST data onto IMP from 1920 to 2010 using 20, 30 and 40 leading PCs.

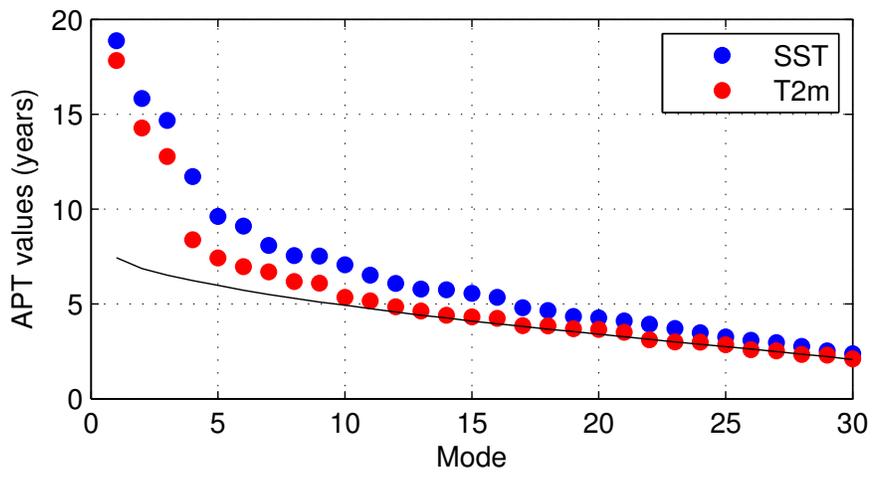


FIG. 10. The APT values using 30 leading PCs for SST and T2m. Solid horizontal line is the 5% significance level of the APT values.