



A Decadal Prediction Case Study: Late Twentieth-Century North Atlantic Ocean Heat Content

STEPHEN YEAGER, ALICIA KARSPECK, GOKHAN DANABASOGLU, JOE TRIBBIA, AND HAIYAN TENG

National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 12 October 2011, in final form 25 January 2012)

ABSTRACT

An ensemble of initialized decadal prediction (DP) experiments using the Community Climate System Model, version 4 (CCSM4) shows considerable skill at forecasting changes in North Atlantic upper-ocean heat content and surface temperature up to a decade in advance. Coupled model ensembles were integrated forward from each of 10 different start dates spanning from 1961 to 2006 with ocean and sea ice initial conditions obtained from a forced historical experiment, a Coordinated Ocean-Ice Reference Experiment with Interannual forcing (CORE-IA), which exhibits good correspondence with late twentieth-century ocean observations from the North Atlantic subpolar gyre (SPG) region. North Atlantic heat content anomalies from the DP ensemble correlate highly with those from the CORE-IA simulation after correcting for a drift bias. In particular, the observed large, rapid rise in SPG heat content in the mid-1990s is successfully predicted in the ensemble initialized in January of 1991. A budget of SPG heat content from the CORE-IA experiment sheds light on the origins of the 1990s regime shift, and it demonstrates the extent to which low-frequency changes in ocean heat advection related to the Atlantic meridional overturning circulation dominate temperature tendencies in this region. Similar budgets from the DP ensembles reveal varying degrees of predictive skill in the individual heat budget terms, with large advective heat flux anomalies from the south exhibiting the highest correlation with CORE-IA. The skill of the DP in this region is thus tied to correct initialization of ocean circulation anomalies, while external forcing is found to contribute negligibly (and for incorrect reasons) to predictive skill in this region over this time period.

1. Introduction

Society would benefit tremendously if climate scientists could produce reliable forecasts, years to decades in advance, of changes in regional hurricane activity, rainfall, or the likelihood of extreme events such as severe heat waves. In the relatively new field of decadal climate prediction, work is under way to assess the feasibility of using coupled general circulation models (CGCMs) to generate such forecasts, but significant scientific challenges must be overcome if this effort is to be successful (Meehl et al. 2009). In an effort to spur advances in this novel application of CGCMs, the protocol for the Coupled Model Intercomparison Project phase 5 (CMIP5) of the World Climate Research Programme includes a set of coordinated initialized prediction experiments to be run

by modeling groups around the world. The premise behind the experimental design is that there is potential to improve climate predictions, and in particular regional predictions, on time scales of up to a few decades by reducing the forecast uncertainty associated with intrinsic climate variability (Hawkins and Sutton 2009). Initializing CGCM simulations with conditions that reflect the current observed state of the earth system could enhance predictive skill by synchronizing the slowly evolving internal variations in the model, in particular ocean variations, with those in nature. Initialization could also improve skill simply by reducing model bias in the early years of a prediction experiment (Solomon et al. 2011).

Several noteworthy studies have demonstrated that historical initialization can enhance model skill at predicting observed regional surface temperature variations on decadal time scales above the level that can be achieved using realistic external forcing alone (e.g., Smith et al. 2007; Keenlyside et al. 2008; Pohlmann et al. 2009). In each of these studies, an anomaly initialization technique was used so as to avoid the inevitable drift

Corresponding author address: Stephen Yeager, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000.
E-mail: yeager@ucar.edu

that results when a model is initialized far from its preferred (biased) climatology. Problematic aspects of the anomaly initialization approach have been noted, however, such as the fact that assimilating observed anomalies of ocean temperature and salinity can produce incorrect density anomalies when added to model climatology, due to nonlinearities in the equation of state (Robson 2010). An alternative approach is to initialize the model with full state fields, which are somehow derived from available observations, and then subtract an empirically defined drift to derive a meaningful forecast. This method has been shown to work in the context of seasonal-to-interannual forecasting (Stockdale 1997), and there are suggestions that it could be effective even on decadal time scales (Troccoli and Palmer 2007).

In this paper, we present some promising results obtained using such a full-field initialization approach in a set of CMIP5 experiments run with the Community Climate System Model, version 4 (CCSM4). By choosing this initialization method, we avoid the issues associated with incompatibilities between anomalies and model climatology, as well as those associated with defining a model climatology in the presence of forced climate trends. Moreover, initial conditions for the ocean and sea ice were readily available on the model grid from a historical atmosphere-forced ocean–sea ice simulation run as part of the latest phase of the Coordinated Ocean-ice Reference Experiment (CORE) ocean model development project (Griffies et al. 2009). A complication, however, is that the bias correction issue becomes a critically important part of the analysis. A recent report by the Climate Variability and Predictability Decadal Climate and Prediction Panel (CLIVAR 2011) recommends a technique for bias correction of CMIP5 decadal prediction experiments, which we adopt and describe below in section 3.

An abrupt warming of the subpolar gyre (SPG) region of the North Atlantic Ocean in the 1990s has been identified as a particularly good test for decadal prediction systems because of its large, clear signature in the observed record and that it apparently resulted from years of ocean preconditioning by persistent positive North Atlantic Oscillation (NAO) forcing (see Robson 2010; Robson et al. 2012, and references therein). Our analysis focuses exclusively on this region as a case study of the CCSM4 decadal prediction methodology and results. We use the forced ocean–sea ice simulation (hereafter referred to as the CORE-IA simulation to reflect the fact that CORE interannual state fields were used as forcing) to identify the physical processes that likely explain the late twentieth-century SPG heat content changes seen in observations and compare these with the heat budgets obtained from the prediction simulations

over the same time period. We then identify the tendency terms contributing to predictive skill, as well as those degrading it. These results are contrasted with a similar analysis performed on an ensemble of uninitialized twentieth-century (20C) simulations, which show modest predictive skill in this region, but for incorrect reasons.

The paper is organized as follows. After describing the decadal prediction experimental design and initialization procedure in section 2, we document the method used for drift correction in section 3. Next, we examine the CORE-IA ocean state to gain a physical understanding of the observed upper-ocean changes in the high-latitude North Atlantic in the mid-1990s. In sections 5 and 6, the results of the decadal prediction experiments are assessed relative to the CORE-IA “truth,” with specific attention given to the behavior of the predictions in the 1990s. Section 7 is devoted to a discussion and summary of our findings. Except where otherwise noted, we use a first order autoregressive model to test for statistical significance, and this is described in an appendix.

2. Description of experiments

All simulations use the CCSM4, whose general description is given in Gent et al. (2011). The atmospheric model has a finite-volume dynamical core with a nominal 1° horizontal resolution and 26 vertical levels, and the ocean model has a nominal 1° horizontal resolution with 60 vertical levels (Danabasoglu et al. 2012a). The land and sea ice components share the same horizontal grids as the atmosphere and ocean models, respectively.

The decadal prediction (DP) experiments are a set of initialized, fully coupled integrations that conform to the CMIP5 specifications for “near-term” prediction experiments (Taylor et al. 2012). They are historical in the sense that they include time-dependent external forcings (e.g., greenhouse gas concentrations). They consist of 10-member ensembles for each start date integrated through at least 10 years, following initialization on 1 January of a particular calendar year. While some DP experiments extend beyond 10 years, we limit our analysis here to the first 10 years of the DP runs, for reasons outlined below. There are 10 different start years (1961, 1966, 1971, . . . , 2006), so the present analysis is based on a total of 100 decade-long coupled experiments.

The historical initial conditions for the DP experiments come from a forced ocean–sea ice simulation designed to reproduce the evolution of the ocean and sea ice states from the start of 1948 through the end of 2007. The CORE-IA simulation is obtained by coupling together the CCSM4 ocean and sea ice models and forcing both at the surface with CORE version 2 historical atmospheric datasets. There is no assimilation of

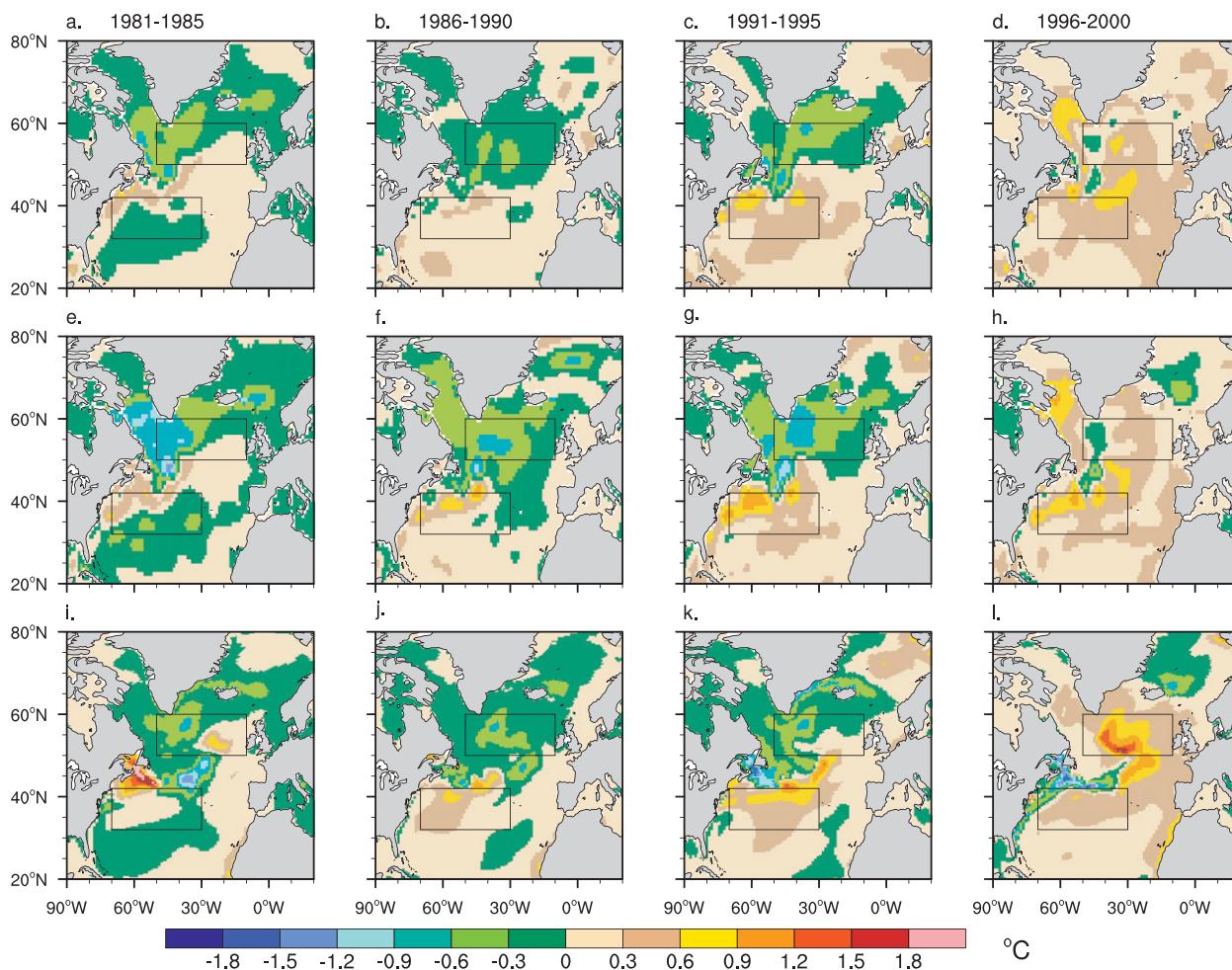


FIG. 1. Pentadal-mean heat content anomalies expressed as the 275-m depth-averaged temperature anomaly relative to 1957–90 climatology from (a)–(d) Ishii and Kimoto (2009), (e)–(h) Levitus et al. (2009), and (i)–(l) CORE-IA. The boxes in each panel demarcate the SPG (50°–10°W, 50°–60°N) and STG (70°–30°W, 32°–42°N) regions.

subsurface observations, and only a weak restoring of model surface salinity to observed climatology is employed. The CORE forcing dataset, which has been adopted by the CLIVAR Working Group on Ocean Model Development for model intercomparisons, imparts realistic surface variability on a range of relevant time scales (Large and Yeager 2004; Large and Yeager 2009), and the resulting simulated upper-ocean variability shows good agreement with a variety of in situ observations (e.g., North Atlantic upper-ocean heat content in Fig. 1). The 240-yr CORE-IA integration is spun up through four consecutive 60-yr cycles of 1948–2007 forcing.

The ocean and sea ice models in DP experiments are initialized with 1 January restart files for a particular year from the last (fourth) cycle of the CORE-IA simulation. No attempt is made to initialize the atmosphere and land models to historical states. Instead, the initial conditions for these component models are taken from corresponding

years of a six-member ensemble of twentieth-century (20C) runs. Specifically, the 10-member DP ensembles are generated by randomly selecting atmosphere and land initial conditions from different 20C runs and/or from different days in the month of January. The reader is referred to Gent et al. (2011) and Meehl et al. (2012) for complete descriptions of the CCSM4 twentieth- and twenty-first-century control simulations and forcing details. We refer to all coupled experiments initialized from CORE-IA, whether of past or future time periods, as decadal prediction experiments.

The DP experiments differ from 20C runs (and their future scenario extensions) in terms of initialization procedure and length of integration but are otherwise subject to the same external forcings of solar irradiance, greenhouse gases, aerosols, and volcanic activity. The forcings used are identical to those used in 20C experiments through 2005, and thereafter, the RCP 4.5 future

emissions scenario is used to force DP experiments. To assess the relative impacts of initialization versus external forcing, we compare the DP runs with a six-member ensemble of 20C experiments generated by branching from a preindustrial (1850 conditions) control run at six different times late in the control, and then integrating with time-varying forcing between 1850 and 2005 (see Gent et al. 2011).

3. Drift correction method

The simulated climate of CCSM4 exhibits systematic differences from both the forced ocean model hindcast and the observed climate. Over multiple decades, coupled predictions that are initialized from the CORE-IA solution tend to drift toward the CCSM4 climatology. For decadal forecasting the climate signal is conflated with the model drift, making it necessary to correct this drifting bias before validating or interpreting the forecasts. We follow the protocol suggested for bias correction of CMIP5 experiments (CLIVAR 2011).

Our approach is to correct the drift relative to the CORE-IA solutions, which we treat as a close approximation to the historic ocean state, because this allows us to correct and analyze the full suite of output fields generated by the CCSM4 ocean model. In particular, this allows us to compute closed ocean heat budgets from the DP simulations. We focus here on the annual average forecasts, but a similar method could be used for monthly data. The underlying assumption in the drift correction methodology is that the bias in the annual average prediction is only a function of forecast lag and does not depend on the model state or the external radiative forcing. The validity of this assumption is certainly questionable and it is the subject of ongoing research in the field, but we adopt it for present purposes. This choice is supported by the fact that, for the field and region of interest here, the drift far exceeds the observed trend (Fig. 4). For a given forecast year τ ($=1, \dots, m$), this allows us to pool the forecast error relative to CORE-IA from all initialization times and compute the model bias as the expected value (or average) of the sample. For a given field, the model drift is thus computed as

$$d_{i\tau} = \langle X_{j\tau} - Y_{j+\tau-1} \rangle_i, \quad (1)$$

where $X_{j\tau}$ is the annual mean from a DP experiment initialized at the beginning of year j , averaged over forecast year τ , and $Y_{j+\tau-1}$ is the annual mean from the CORE-IA simulation at the corresponding time (i.e., calendar year $j + \tau - 1$). The $\langle \dots \rangle_i$ operator is an average over all ensemble members for all initialization

years j , exclusive of year i . Thus, we compute $d_{i\tau}$ separately for each 10-member forecast ensemble so as to avoid overfitting in the drift computation.

The bias-corrected forecast is then obtained by subtracting the estimated drift from the raw forecast:

$$X_{i\tau}^* = X_{i\tau} - d_{i\tau}. \quad (2)$$

The size of the sample available to compute $d_{i\tau}$ is between seven and nine for $\tau \leq 10$ but drops to two for larger forecast lags.¹ This is because 1) we exclude the forecasts to be corrected in the Eq. (1) averaging, 2) only three of the DP ensembles were integrated beyond 10 years, and 3) the CORE-IA only extends through 2007. We therefore limit consideration to forecasts out to $\tau = 10$.

4. The 1990s regime shift in the North Atlantic

The North Atlantic is a region of particular interest for decadal prediction because there is evidence that large-amplitude sea surface temperature (SST) variations there modulate the summer climate in North America and Europe as well as rainfall and hurricane activity in the tropical Atlantic (Sutton and Hodson 2005; Knight et al. 2006; Zhang and Delworth 2006). Furthermore, it has recently been demonstrated that skillful prediction of Atlantic tropical storm activity is dependent upon skillful prediction of SST in the Atlantic subpolar gyre (Smith et al. 2010; Dunstone et al. 2011).

Gridded analyses of North Atlantic hydrography compiled by Ishii and Kimoto (2009) and Levitus et al. (2009) clearly show a large, sudden warming of the high-latitude upper ocean in the mid-1990s (Fig. 1), which has been linked to similarly abrupt shifts in observed North Atlantic surface circulation and sea surface height (Flatau et al. 2003; Hakkinen and Rhines 2004), marine fauna (Hatun et al. 2009), and Greenland glacier melt (Holland et al. 2008). The abrupt warming was preceded by anomalously cool conditions in the subpolar gyre (SPG) region (50° – 10° W, 50° – 60° N; see Fig. 1) and growth of anomalously warm conditions in the subtropical gyre (STG) region (70° – 30° W, 32° – 42° N) between about 1981 and 1995. The surface-forced CORE-IA simulation reproduces the overall pattern and magnitude of observed pentadal-mean 275-m heat content change in the late twentieth century (Figs. 1i–l). Apart from an overly robust warming in the 1996–2000 pentad, the CORE-IA

¹ Note that here the term sample refers to the collection of start dates that can be included for a given τ . Since there are 10 members per start date, multiply by 10 to get the total number of integrations included in the $d_{i\tau}$ computation.

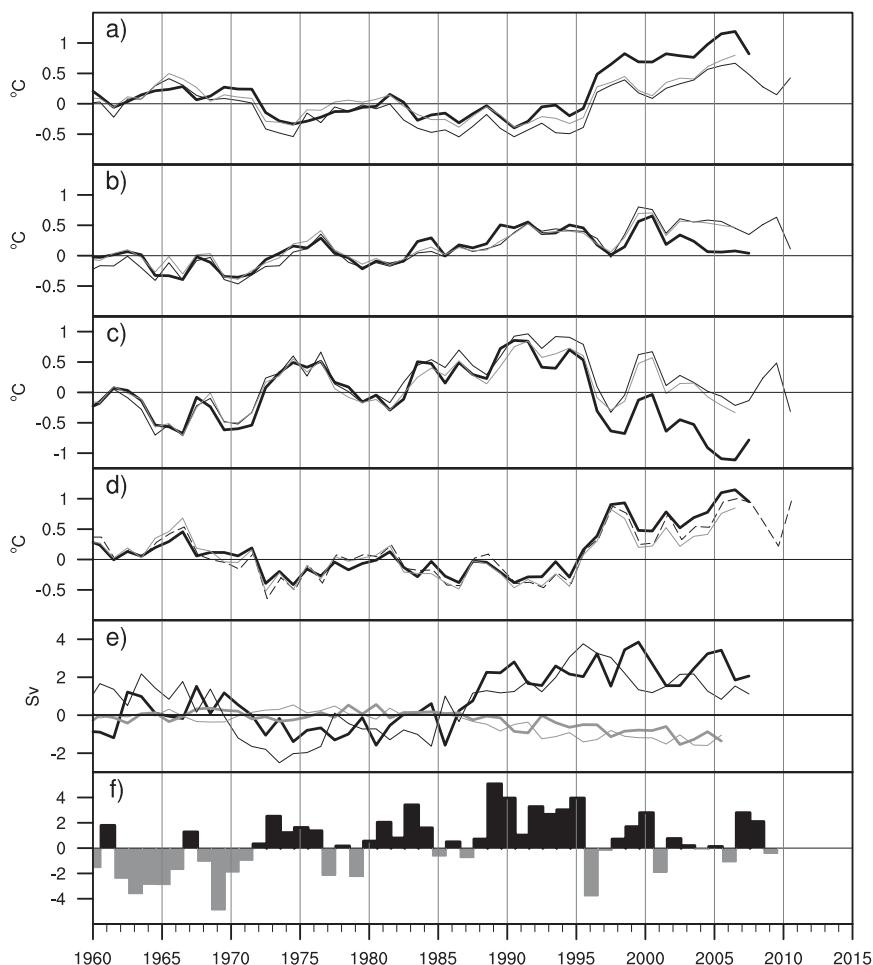


FIG. 2. Annual mean time series of (a) 275-m heat content anomaly in the SPG box (SPG, °C), (b) 275-m heat content anomaly in the STG box (STG, °C), (c) heat content dipole strength (STG – SPG, °C), (d) SST anomaly in SPG box (°C), (e) AMOC strength at 37.5°N and 900-m depth (thick curves) and mean barotropic streamfunction in the SPG box (thin curves) ($\text{Sv} \equiv 10^6 \text{ m}^3 \text{ s}^{-1}$), and (f) observed winter (DJFM) NAO index. In (a)–(d), the thick black curve is CORE-IA and the thin curves are from the following observational datasets: Levitus (thin solid black), Ishii and Kimoto (thin solid gray), and Hurrell (thin dashed black). See text for references. In (e) the black curves are from CORE-IA, the gray curves are from the 20C ensemble, and the barotropic streamfunction anomaly has been multiplied by -1 so that positive values indicate anomalously strong circulation.

discrepancies with observed hydrography are roughly the same magnitude as the uncertainty in the observation-based gridded products.²

The time series of regionally averaged SPG heat content anomaly (Fig. 2a) further demonstrates the abruptness of the high-latitude warming in the mid-1990s following a long cooling trend. The CORE-IA time series matches the observed year-to-year variations in heat content, including the sharp rise in 1996, but shows much more positive

anomalies than either of the observed time series from about 1996 onward. In contrast, the STG region is characterized by a long warming trend between about 1970 and 2000 (Fig. 2b), which mirrors the positive trend in the winter NAO index over the same time period [Fig. 2f; observed winter December–March (DJFM) NAO index provided by the Climate Analysis Section, National Center for Atmospheric Research at Boulder, Colorado, Hurrell (1995)]. The strength of the intergyre heat content gradient, quantified here simply as the difference of the STG and SPG heat content time series, exhibits an even better correspondence with the winter NAO index (Fig. 2c). Positive values of this STG–SPG index indicate

² All heat content anomalies are relative to a 1957–90 climatology that is computed separately for each individual dataset.

that there is an anomalously strong meridional heat content gradient in the North Atlantic with a dipole structure such as can be seen in the 1991–95 panels of Fig. 1. The 1960–2006 correlations of CORE-IA heat content with Ishii and Kimoto (2009) are 0.90, 0.85, and 0.82 in Figs. 2a–c, respectively. The correspondence of CORE-IA with observations in the SPG region is even better for SST, which closely reflects the upper-ocean heat content variations (Fig. 2d). The 1960–2006 correlations of CORE-IA SST with fields from Ishii and Kimoto (2009) and Hurrell et al. (2008) are 0.93 and 0.96, respectively. All of these correlations are significant at the 95% confidence level based on a two-sided Student's *t* test with autocorrelation taken into account.

The dipolar heat content anomaly that preceded the late 1990s warming reached maximum strength in the early 1990s (Fig. 2c) and is generally understood to be the ocean signature of persistent positive NAO (NAO+) forcing in the years leading up to the shift (Marshall et al. 2001; Eden and Willebrand 2001; Lozier et al. 2008; Lohmann et al. 2009b; Lohmann et al. 2009a; Robson et al. 2012). Our CORE-IA results are consistent with many of the ideas advanced in these studies: increasingly strong winter NAO conditions from the early 1970s through 1995 (Fig. 2f) resulted in a decades-long spinup of the Atlantic meridional overturning circulation (AMOC) and subpolar gyre barotropic streamfunction (BSF) strength (Fig. 2e) and anomalously cool conditions at subpolar latitudes initially (Fig. 2a). The increase in northward heat transport associated with a strengthening AMOC contributed to the steady warming of the STG region (Fig. 2b) and, thus, enhancement of the STG–SPG heat content dipole (Fig. 2c). The SPG warming was triggered abruptly in 1996 when winter NAO conditions weakened dramatically, but Lohmann et al. (2009a,b) contend that strong ocean preconditioning (reflected here in the dipole strength index) made significant SPG warming all but inevitable. This is because the enhanced northward heat transport associated with NAO+ at some point overwhelms the surface cooling, leading to a reversal of the high-latitude heat content tendency even if the NAO remains strong and positive. The implication of the Lohmann et al. results is that the large magnitude warming of the high-latitude North Atlantic Ocean in the 1990s was highly predictable, given the state of the Atlantic Ocean in the early 1990s.

A heat budget of the SPG region (to 275-m depth) from the CORE-IA simulation provides insight into the physical mechanisms that account for the rapid late twentieth-century SPG warming. The calculation is similar to that described in detail in Danabasoglu et al. (2012b) for the Labrador Sea region. Imbalances between the net surface heat flux (SFLX), which cools the SPG box, and the net advective and diffusive fluxes

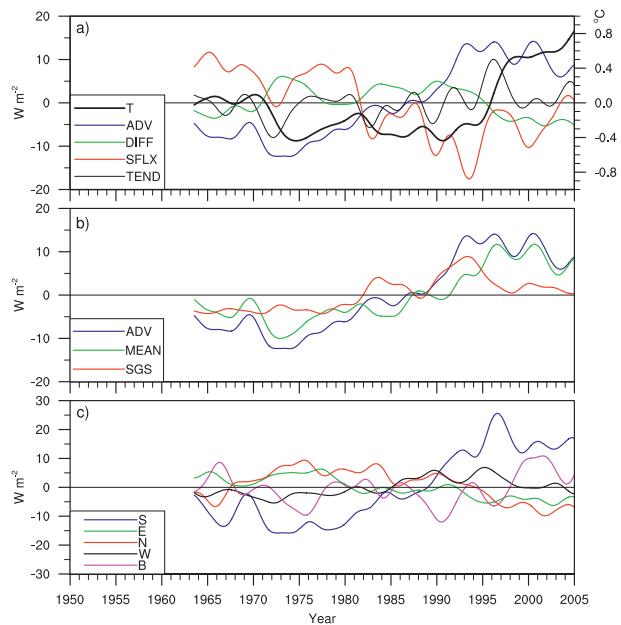


FIG. 3. Time series of 275-m heat budget terms from CORE-IA in the SPG box, with positive values indicating heat gain by the box. (a) The net advective (ADV), diffusive (DIFF), and surface (SFLX) heat fluxes are plotted together with the heat content tendency (TEND) and mean SPG heat content (expressed as the volume-average temperature and referring to scale on right). (b) ADV is replotted together with its decomposition into mean (MEAN) and subgrid-scale (SGS) components. (c) The components of ADV through the south (S), east (E), north (N), and west (W), and bottom (B) faces of the SPG box are plotted. All curves are based on monthly mean data low-pass filtered with a 5-yr cutoff frequency. To get units of W m^{-2} , the terms have been scaled by the surface area of the SPG box and are plotted as anomalies from 1961–2007 climatology.

(ADV and DIFF), which both warm the SPG box, generate nonzero heat content tendency ($\text{TEND} = \text{SFLX} + \text{ADV} + \text{DIFF}$). The 1961–2007 averages are 1, -60 , 47 , and 14 W m^{-2} for TEND, SFLX, ADV, and DIFF, respectively, so there is a slight upward trend in SPG temperature during the 47-yr period being analyzed. The heat budget components in Fig. 3 are plotted as anomalies from long-term mean values computed over this time period.³ The increasingly strong NAO+ forcing of the 1980s and early 1990s (Fig. 2f) gave rise to opposing trends in the SFLX and ADV terms, such that the heat content tendency remained quite low from the mid-1970s up until the mid-1990s. By the early 1990s, very anomalous surface cooling was largely balanced by unusually

³ For the heat budget analysis, the climatological period is chosen to be 1961–2007 because this is the period used for defining d_{it} . With this choice, the climatological budget computed across the DP ensembles matches that computed from CORE-IA.

strong advective and diffusive heating of the SPG box (Fig. 3a). The weak NAO in the winter of 1996 coincided with an extreme, positive heat content tendency in the SPG because the surface cooling was suddenly insufficient to match the strong advective heating that had been established over the preceding decades. Thus, both surface and advective heat fluxes played crucial roles in the SPG warming, but the abruptness and timing of the regime shift is attributable to the sudden change in atmospheric state.

The upward trend in ADV from the mid-1970s to the mid-1990s was primarily due to a pronounced increase in heat flux through the south face of the SPG box (Fig. 3c). A decomposition of CORE-IA mean northward heat transport into gyre and overturning components (see for example Eden and Willebrand 2001; Johns et al. 2011) indicates that, at the southern edge of the box (50°N), the gyre component accounts on average for about 83% of the meridional transport of heat across that latitude (0.55 PW of the the 1961–2007 mean net heat transport of 0.66 PW). Therefore, the trend in heat advection through the south face of the SPG box (Fig. 3c) is a direct reflection of the trend in the strength of the subpolar gyre circulation (Fig. 2e). However, at 45°N the overturning component accounts for about 60%, and at 40°N more than 100%, of roughly the same net meridional heat transport. The boundary between the gyres at latitudes of 40° – 50°N is a transition zone between the overturning-dominated heat transport in the subtropics and the gyre-dominated heat transport at subpolar latitudes, and these clearly covary on long time scales (Fig. 2e). These decadal variations are primarily driven by NAO-related buoyancy forcing (Robson et al. 2012).

While most of the trend in net ADV was due to changes in the mean (MEAN), that is, resolved, circulation (Fig. 3b), subgrid-scale (SGS) eddy heat fluxes also contributed to the nearly steady increase in ADV heating. The periods of intense NAO+ in the early 1980s and 1990s were followed by surges of SGS heat flux into the SPG region while the MEAN heat flux actually dropped. We speculate that the decrease in MEAN advective heating during these periods is associated with anomalously southward Ekman transports due to strong westerly winds and that the rise in the eddy heat transport is explained by parameterizations acting to flatten the steeper isopycnal slopes associated with the enhanced heat content dipole.

While the CORE-IA shows a notable surge in northward heat transport following the NAO+ to NAO– switch, which contributed to the elevated ADV and TEND (see curve S in Fig. 3c), this flux reached a maximum about a year after TEND reached a maximum in early 1996 and it was largely compensated by increased

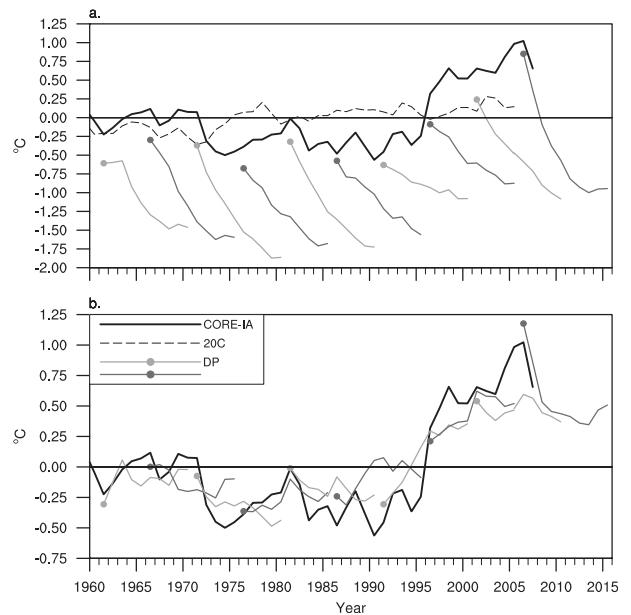


FIG. 4. Annual mean 275-m heat content anomaly ($^{\circ}\text{C}$) in the SPG box for (a) CORE-IA (thick solid black), the 20C 6-member ensemble mean (thin dashed black), and the raw DP 10-member ensemble means (gray curves, alternating shades for clarity) and (b) CORE-IA and the bias-corrected DP ensemble means (note change in scale). Large circles indicate the first-year ($\tau = 1$) average of each DP ensemble. The CORE-IA and 20C anomalies are computed relative to climatologies over the reference periods 1961–2007 and 1961–2005, respectively; DP anomalies are computed relative to the CORE-IA climatology.

heat advection out of the north, east, and bottom faces of the box. Thus, the dramatic rise in SPG temperature in early 1996 in the CORE-IA simulation cannot be attributed primarily to a surge in oceanic heat transport convergence, as Robson et al. (2012) conclude from their budget analysis. Rather, the CORE-IA heat budget indicates a complex evolution of the various forcing terms that resulted in elevated net advective heating of the SPG region throughout the 1990s; the surge in heat content tendency in 1996 is best understood as the loss of balance between strong ADV heating and SFLX cooling when NAO weakened suddenly.

5. Decadal prediction results

The raw time series of ensemble-mean SPG heat content from the prediction experiments are dominated by very large and unrealistic cooling trends in the first decade of integration (Fig. 4a). However, after drift correction, the ensemble-mean heat content time series show a remarkable correspondence with the CORE-IA time series (regarded here as the truth) in the SPG region (Fig. 4b). We focus exclusively hereafter on bias-corrected results. Taken as a whole, the DP runs reproduce the

TABLE 1. First and second pentad correlations between regionally averaged SPG ocean fields from DP and various representations of historical truth (by row, see text for references). Values in parentheses give the 95% confidence level based on a first-order autoregressive parametric bootstrap (see appendix for details).

		$\tau = 1-5$	$\tau = 6-10$
Heat Content	CORE-IA	0.95 (0.78)	0.92 (0.59)
	Levitus et al.	0.74 (0.77)	0.76 (0.60)
	Ishii and Kimoto	0.75 (0.77)	0.76 (0.60)
SST	CORE-IA	0.93 (0.76)	0.94 (0.59)
	Hurrell et al.	0.88 (0.73)	0.89 (0.60)
	Ishii and Kimoto	0.80 (0.73)	0.84 (0.58)

low-frequency evolution from generally cool conditions in the subpolar region between the early 1960s and early 1990s to anomalously warm conditions in the late 1990s and early twenty-first century. While this might be expected just from thermal inertia of the upper ocean after initialization, most ensembles also show skill in predicting the mean heat content trend, especially in the first pentad of simulation. There are, however, notable discrepancies in first pentad trend such as in the 1986- and 2001-initialized DP ensembles (hereafter referred to as 1986DP, 2001DP, etc). A particularly interesting result is obtained from the 1991DP, which predicts a sharp warming around the time of the observed regime shift. Comparison with the uninitialized 20C ensemble appears to suggest that the external forcing contributes at least some skill in capturing the low-frequency trend in this region, but we show below that this conclusion is unwarranted.

To quantify DP skill in the SPG region, we correlate regionally-averaged heat content and SST from the first and second pentads of the prediction simulations (forecast years 1–5 and 6–10, respectively) with the corresponding 5-yr averages from CORE-IA as well as from a collection of observational datasets (Table 1). The correlation with CORE-IA of first pentad heat content forecasts is high ($r = 0.95$) and only slightly lower for second pentad forecasts ($r = 0.92$). The DP skill scores for SST (relative to CORE-IA) are comparably high in both pentads. Note that, because the CORE-IA time series only extends through 2007, the sample size is nine for first pentad correlations and eight for second pentad correlations. To assess the significance of the pentadal correlations, we compare to an alternative forecast method that uses a first-order autoregressive [AR(1)] model trained on one of the 20C simulations (see appendix for details). The correlations between the DP ensembles and CORE-IA of both first and second pentad SPG heat content and SST are significant at well above the 95% confidence level relative to this reasonable null hypothesis.

Relative to the Levitus et al. (2009) and Ishii and Kimoto (2009) observations of heat content in the SPG region, the DP skill is considerably lower than when computed relative to CORE-IA. The correlations of roughly 0.75 are significant at the 95% confidence level only in the second pentad (Table 1). However, for reasons that are not fully understood, correlations with observed SST are higher and more significant than those obtained for heat content. The pentadal correlations with SST from Ishii and Kimoto (2009) are significant in both pentads at ≥ 0.8 , and the skill scores relative to the SST dataset of Hurrell et al. (2008) are only slightly lower than those relative to CORE-IA. [We do not include Levitus et al. (2009) in the SST metrics because this heat content data product did not include an SST field.] While true predictive skill is measured relative to observations, it must be borne in mind that the benchmark datasets included in Table 1 are global products whose construction requires extensive temporal and spatial filling. We presume that the higher SST correlations (both in Table 1 and Fig. 2) are attributable to the fact that remote sensing results in much greater observational data coverage at the ocean surface, so there is considerably less processing required in the generation of gridded, observed SST. As suggested by Robson et al. (2012) in their comparison of in situ data with processed Met Office ocean analyses (Smith and Murphy 2007) in the North Atlantic, gridded observational products may underestimate the magnitude of the mid-1990s heat content change in the SPG as a result of data processing. Thus, the discrepancies between the CORE-IA and the observed heat content time series in Fig. 2a are not wholly indicative of model error, and relatively low correlations between DP heat content and observations in Table 1 should be interpreted with caution. High pentadal correlations with the Hurrell et al. (2008) SST product, which incorporates satellite data after 1981, greatly bolsters our confidence in the DP predictive skill implied by the high CORE-IA correlations. To illuminate the mechanisms behind the apparent skill of our decadal predictions, we now focus solely on DP correlations with the CORE-IA simulation.

Comparing the SPG heat budget terms from the DP ensembles with those from the CORE-IA simulation (Fig. 5) shows that the significant skill in predicting the heat content tendency (TEND) in the first pentad of the forecast ($r = 0.60$) can be attributed primarily to highly correlated advective heat flux anomalies ($r = 0.91$). The 5-yr mean ADV anomalies have the largest range of any of the budget terms over the historical period analyzed, in both CORE-IA and DP, and this large variability can be traced mainly to variations in the MEAN advective heat flux through the south face of the box (Fig. 6). The

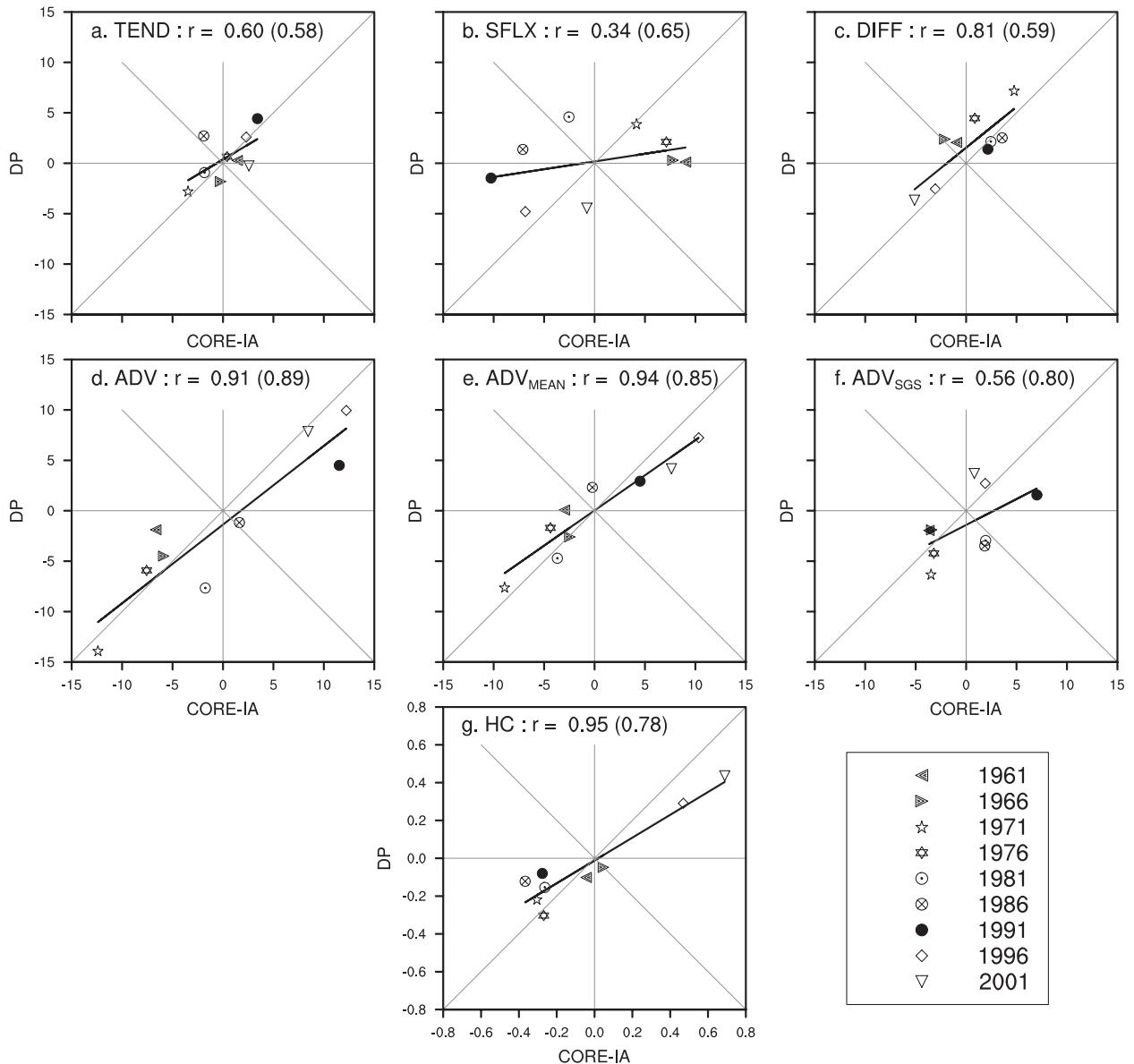


FIG. 5. Comparison of mean SPG heat budget terms from the first pentad ($\tau = 1-5$) of the DP predictions (y axis) with corresponding 5-yr means from the CORE-IA ocean state (x axis). All panels show anomalies relative to the 1961–2007 CORE-IA climatology. The terms plotted are (a) heat content tendency, (b) surface heat flux, (c) diffusive heat flux, (d) advective heat flux, (e) mean component of advective heat flux, (f) subgrid-scale component of advective heat flux. Units are (a)–(f) W m^{-2} and (g) $^{\circ}\text{C}$. The symbols reflect the first year of a given DP ensemble and also the first year of the pentad. Correlation coefficients (r) and regression lines are indicated in each panel. Correlations in parentheses give the 95% confidence level based on the null hypothesis that the predictive skill can be explained by a first-order autoregressive process (see text and appendix for details).

skill in reproducing variations in the net ADV term is enhanced by quite accurate prediction of SGS heat fluxes through the south, north, and especially the bottom face of the SPG box (Fig. 6). In net terms, however, the heat content tendency due to MEAN advection is predicted much better than that due to SGS advection (cf. Figs. 5e,f). The large range of pentadal-mean ADV terms between 1961 and 2005 (Fig. 5d) reflects the low frequency

variations in ocean circulation strength over this time period (Fig. 2e), although Robson et al. (2012) suggest that variations in the temperature field probably also play a role in setting heat transport variations at this latitude.

The diffusive heat flux term shows more predictability than might have been expected in the first pentad of prediction (Fig. 5c). The high correlation of DIFF anomalies ($r = 0.81$) helps augment the TEND

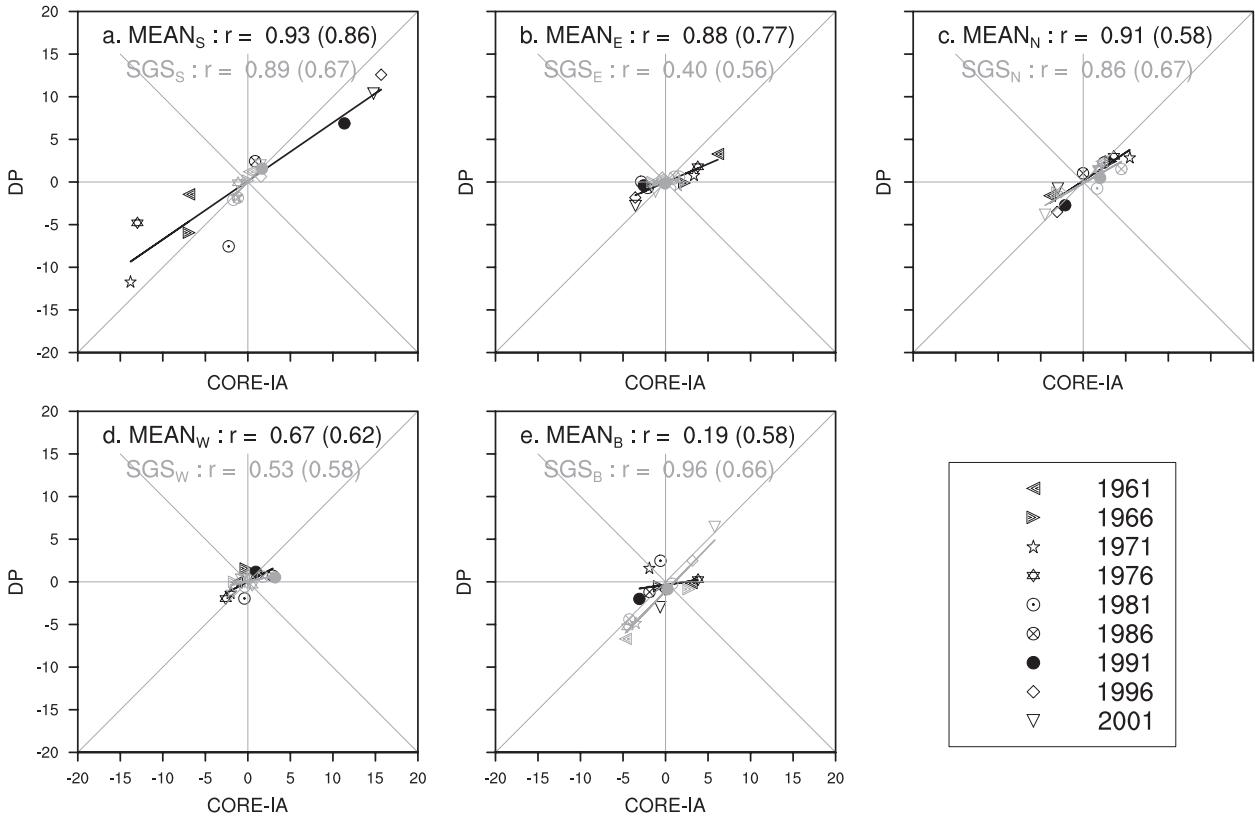


FIG. 6. As in Fig. 5 but showing the various contributions to the advective (ADV) heat budget term averaged over $\tau = 1-5$ for both MEAN and SGS heat flux components through (a)–(e) the south (S), east (E), north (N), west (W), and bottom (B) faces of the SPG box, respectively, are shown.

correlation, but only modestly because these are relatively small budget terms. In CORE-IA, the SFLX terms largely counterbalance the large ADV anomalies but they show much less pentadal variation than ADV in the DP ensembles. As a result, the SFLX term ($r = 0.34$) explains most of the discrepancy in TEND in the first five years. The fact that the range of DP SFLX anomalies in Fig. 5 is about half that of CORE-IA is partly a result of DP ensemble averaging, which tends to reduce the DP range of all heat budget terms and particularly those terms that have a large spread within an ensemble. However, the low SFLX variation across the DP experiments is probably also a reflection of a rather anemic simulation of NAO-related air–sea heat exchange in the CCSM4 model (a point we return to below).

The very high correlation of first pentad heat content anomalies ($r = 0.95$, Fig. 5g) certainly derives in part from the fact that there is a very large range in heat content initial conditions, with most of the DP ensembles initialized prior to the mid-1990s regime shift and two, 1996DP and 2001DP, initialized after. Nevertheless, the forecasts based on our AR(1) null hypothesis indicate that the DP approach yields predictive skill that

is significantly better than a damped persistence model (the 95% confidence correlation is only 0.78). This implies that the mean tendency (Fig. 5a) is not well modeled as a damped trend back toward the climatological mean. This is evident in both Figs. 4 and 5 where examples can be found of first pentad heat content anomalies of both signs that coincide with tendencies driving the anomalies even further away from the long-term mean (e.g., note 1971DP and 1996DP). The high DP skill evident in Fig. 5g is thus to a large extent associated with the skill in getting TEND correct in the first years of the prediction simulations, which we attribute to appropriately initialized ocean dynamics.

The skill in predicting TEND clearly varies with the start date. The negative mean tendency of 1971–75 is skillfully predicted by 1971DP, which exhibits perhaps the best match to CORE-IA of any of the prediction ensembles in terms of first pentad heat budget anomalies (Fig. 5). The 1996DP is also noteworthy for its success in predicting the correct positive tendency between 1996 and 2000, with nearly the correct combination of ADV, SFLX, and DIFF anomalies. The 1991DP predicts nearly the right magnitude positive tendency for 1991–95, but

this is due to nearly compensating errors in SFLX and ADV. The 1981DP predicts a negative tendency in the first pentad that nearly matches CORE-IA but for incorrect reasons: the ADV anomaly is much too negative and the SFLX anomaly is the wrong sign. The worst result is generated by the 1986DP, which predicts an erroneous positive TEND driven by large diffusive fluxes (Figs. 4 and 5). This discrepancy relates to poor skill at predicting the intense NAO+ of the late 1980s (Fig. 2f) as evidenced by the poor SFLX prediction in this ensemble.

Time integration converts the skill in predicting TEND into skill in predicting heat content changes at later times. Thus, the significant correlation of TEND in the first five years of the predictions, mainly attributable to the fact that initialization of anomalous ocean circulation sets the magnitude of MEAN heat advection through the south face, is seen reflected in the high SPG heat content correlation in the second pentad of the DP predictions (Fig. 7g). The significant heat content correlation ($r = 0.92$) is even more distinguishable from damped persistence than in the first pentad and is again seen to derive from the large heat content range. However, this range is now less a reflection of initial conditions than of the successful prediction of large heat content increases between the first and second pentads of 1991DP and 1996DP. These increases are related to the large, positive northward heat flux anomalies in the early years of those ensembles (Fig. 6a). Also, predictions of anomalously negative heat content in the latter halves of 1966DP, 1971DP, 1976DP, and 1981DP (Fig. 7g) are associated with skillful predictions of first pentad TEND in those ensembles (Fig. 5a). Therefore, the DP skill in the second pentad cannot be attributed simply to persistence of initialized heat content anomalies.

The DP ensembles show good skill at predicting SPG heat content at lags 6–10 despite poor skill ($r = -0.44$) at predicting the mean heat content tendency over this latter pentad (Fig. 7a). The breakdown of heat budget terms in the second pentad (Figs. 7 and 8) shows a degradation in predictive skill in almost every component of tendency. The SFLX skill drops to essentially zero (Fig. 7b), and there is only weak lingering skill in ADV (Fig. 7d), which is no longer significant at the 95% level. What skill remains is mainly associated with the DIFF heat flux (Fig. 7c) as well as the MEAN advective heat flux through the south face (Fig. 8a). Interestingly, the correlations of second pentad DIFF (Fig. 7c) and SGS advective flux through the bottom (Fig. 8e) are only slightly lower than the corresponding first pentad correlations. The skill in these terms arises, we think, as a consequence of getting high heat content correlations (i.e., these tendency terms are more a response to SPG heat content changes than a driver of them).

The low skill in predicting TEND in the second pentad suggests that extending these DPs beyond 10 years would probably yield substantially lower heat content correlations with the CORE-IA truth than seen in the first decade of simulation. On the other hand, the analysis implies that improved decadal prediction of SPG heat content, out to 10 years and beyond, might well be possible if skill in forecasting the ADV heating term could be maintained out to greater lags. As discussed above, this term (and in particular the northward component) largely explains the success of the DP experiments in the first and second pentads despite the evidently poor skill in predicting SFLX changes (i.e., poor NAO prediction). The skillful prediction of the oceanic advective heat convergence term (Fig. 5d) derives mostly from good initialization of low-frequency, large-amplitude ocean circulation anomalies (Fig. 2e) and the persistence of those anomalies in the first pentad. The correlations of AMOC strength at 37°N (near the AMOC maximum) between the DP ensembles and CORE-IA are $r = 0.9$ and $r = 0.6$ in the first and second pentad, respectively. These scores are comparable to those obtained for ADV (Figs. 5d and 7d) and ADV from the south (Figs. 6a and 8a). At 45°N, closer to the south face of the SPG box, the AMOC correlations are lower: $r = 0.6$ and $r = 0.0$, respectively. However, the gyre component accounts for most of the northward heat transport at this latitude (section 4), and the pentadal correlations of regionally averaged SPG barotropic streamfunction are $r = 0.96$ and $r = 0.37$, respectively. We speculate that improved prediction of circulation variations into the second pentad would likely translate into skillful TEND predictions in the second pentad; this in turn would allow for skillful forecasts of SPG heat content more than a decade in advance.

6. The 1991-initialized decadal prediction ensemble

The 1991DP ensemble is of particular interest because the observed regime shift occurred around the midpoint of the decade-long prediction, in the winter of 1995/96. The success of this ensemble in reproducing large heat content changes between the first and second pentads contributes significantly to the high correlation of second pentad heat content (Fig. 7g). In both 1991DP and CORE-IA, the 5-yr period from 1991 through 1995 had the highest positive heat content tendency of any of the pentads considered in this analysis (Fig. 5a). The CORE-IA budget shows that the positive tendency was concentrated in late 1995 (Fig. 3a) when the NAO switch occurred (Fig. 2f). In the 1991DP, the first pentad TEND is well-represented, but only because both the ADV and

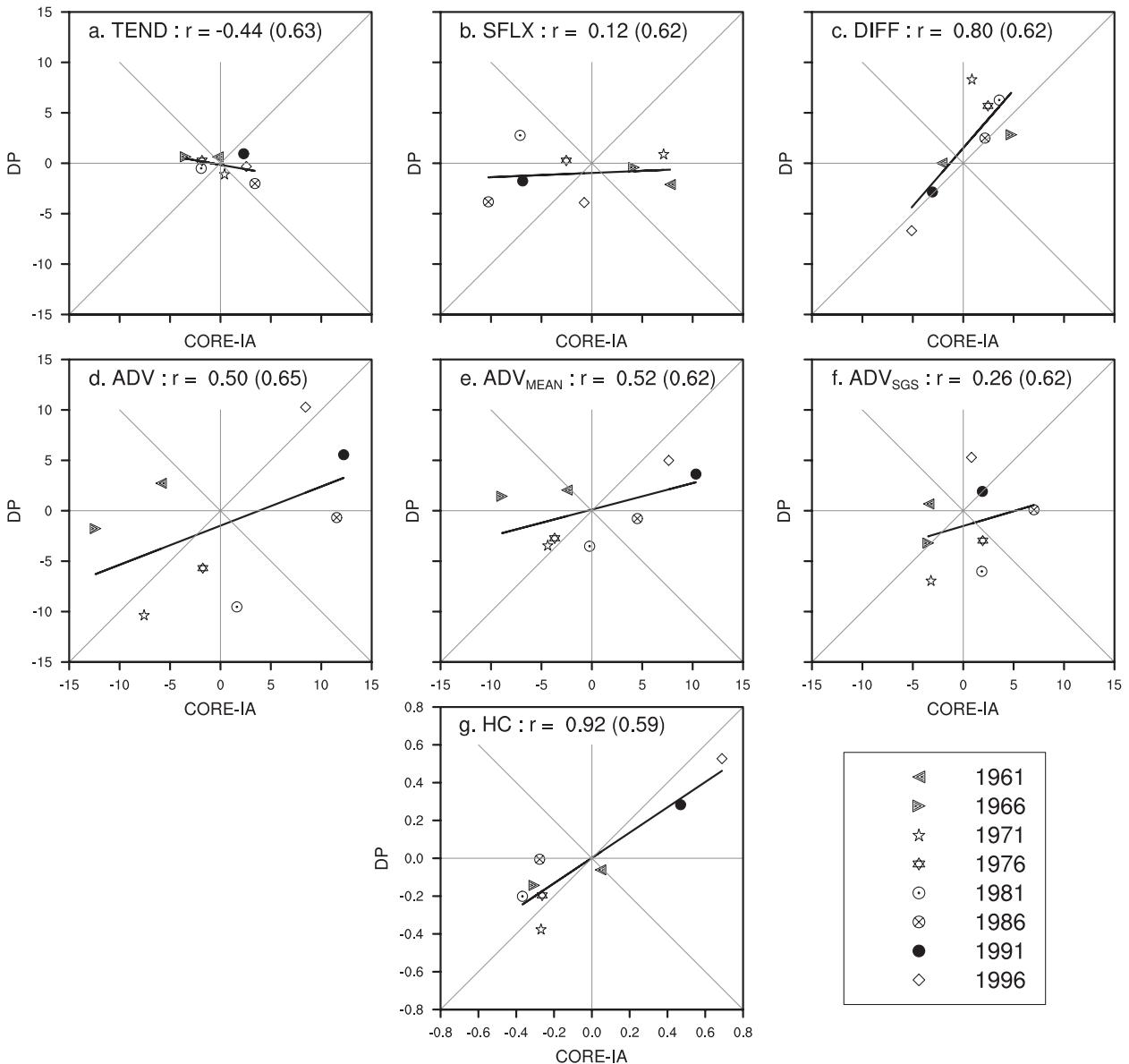


FIG. 7. As in Fig. 5 but for the second pentad ($\tau = 6-10$) of the DP predictions.

SFLX terms are underestimated when averaged across the individual ensemble members (Figs. 5b,d).

As discussed in section 4, the abrupt shift from strong NAO+ conditions to NAO- in late 1995 appears to have set the timing of the SPG regime shift in nature. Thus, we would expect the behavior of the individual members of the 1991DP to differ depending upon their respective NAO predictions. Indeed, ensemble members that simulate strong positive winter NAO conditions in the first pentad of prediction [as measured by the mean January–March (JFM) zonal wind stress anomaly time-averaged over 1991–95 and regionally-averaged over the SPG box region] do best at reproducing

both the magnitude and timing of the heat content shift (Fig. 9a). Ensemble members that simulate only moderate NAO+ or anomalously weak NAO conditions (Figs. 9b,c, respectively) tend to predict a heat content rise earlier than observed, with a lower overall amplitude of heat content increase in the 1990s. Note that regardless of NAO conditions all members predict an increase in SPG heat content. The resulting ensemble-mean heat content rise (Fig. 9b) is thus strongly predetermined by the ocean initial conditions of January 1991. This result lends support to the idea of ocean preconditioning by persistent NAO+ (Lohmann et al. 2009a,b).

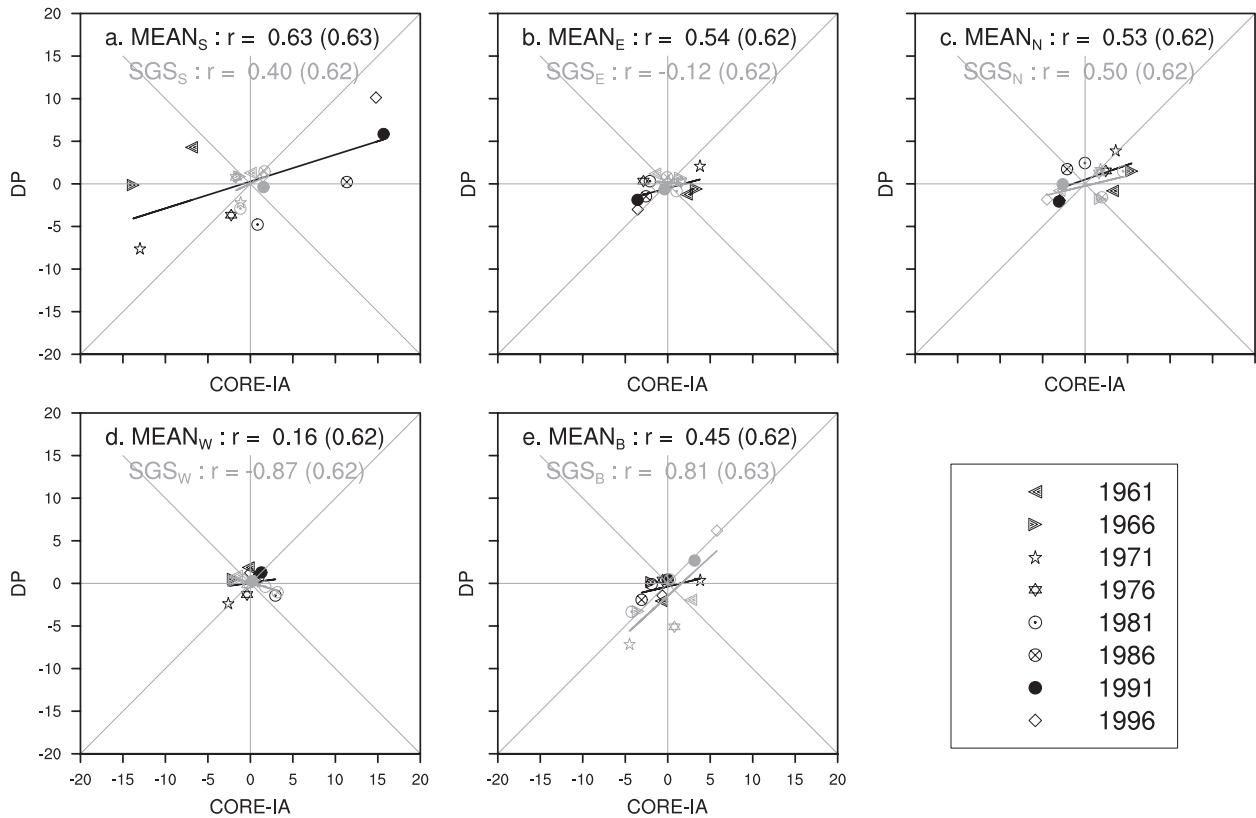


FIG. 8. As in Fig. 6 but for the second pentad ($\tau = 6-10$) of the DP predictions.

7. Discussion and conclusions

The raw predictions of twentieth-century North Atlantic Ocean SPG heat content from full-field-initialized CCSM4 ensembles are dominated by large, systematic drifts toward the model climatology. However, given a sufficient number of ensemble realizations, a mean drift (d_{ir}) can be defined and subtracted from the DP ocean fields to yield good predictions of heat content in the SPG box. Furthermore, the physical mechanisms governing heat content change in this region in the DP experiments are consistent with those diagnosed from a forced ocean–sea ice simulation that exhibits good correspondence with observations.

The CORE-IA analysis suggests that the abrupt 1990s heat content regime shift was the end result of a long period of AMOC increase driven by increasingly intense negative buoyancy forcing in the subpolar region, in general agreement with a similar analysis by Robson et al. (2012). Between about 1970 and 1995, a steady rise in advective heating of the SPG region was largely counterbalanced by the enhanced surface cooling, so that the heat content tendency remained generally low (Fig. 3a), apart from relatively minor negative tendencies in the 1970s and 1980s that were apparently associated with

episodic intensifications of the surface cooling. The large positive heat content tendency of the mid-1990s was associated with an abrupt return to climatological surface cooling over the subpolar gyre, at a time when advective heating of the region (from the south) was anomalously high.

The close correspondence between the SPG heat content evolution from CORE-IA and that predicted in the initialized CMIP5 experiments, which is visually apparent in Fig. 4b and quantified as pentadal-mean correlations in Figs. 5g and 7g, is related to the fact that the DP experiments do a fair job of predicting the correct heat content tendency, at least in the early years of the simulations (Fig. 5a). Although the prediction quality varies with the start date, when taken as a whole, the DP ensembles exhibit enough skill at reproducing large ADV tendencies that heat content is driven toward correct second pentad values (Fig. 7g).

We have identified the MEAN advection of heat through the south face of the SPG box as the predominant budget term that accounts for the success of the DP ensembles in the subpolar region. The large spread in this term across start dates, with negative first pentad anomalies in the early ensembles and positive anomalies in the later ensembles (Fig. 6a), is a reflection of the long upward trend in AMOC

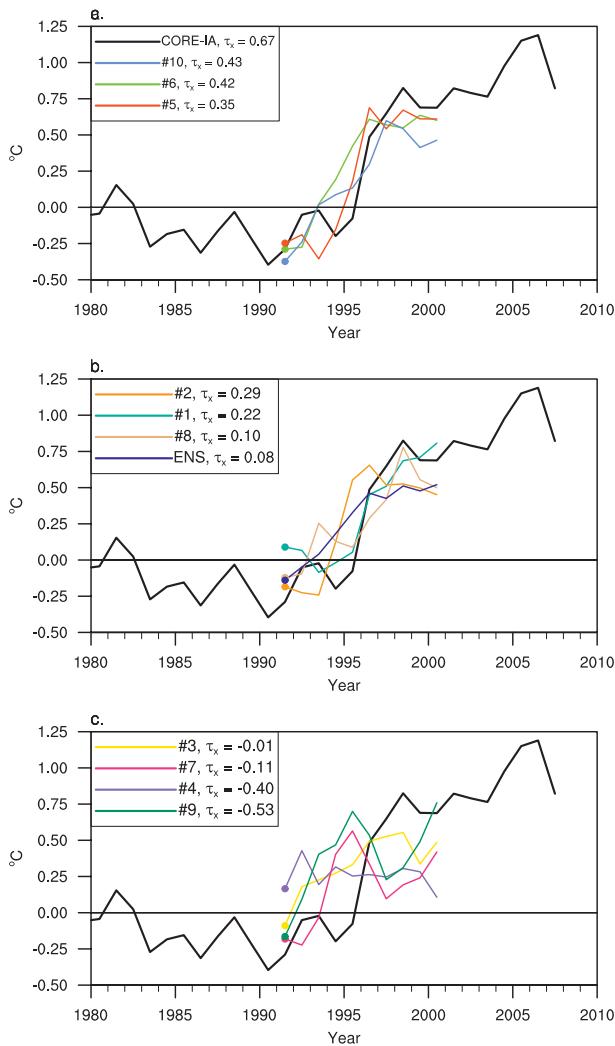


FIG. 9. Annual mean time series of 275-m heat content anomaly in the SPG box from CORE-IA as well as from the ensemble of 1991-initialized DP simulations. The experiments are grouped according to the magnitude of the mean JFM zonal wind stress anomaly (τ_x , N m^{-2}) between 1991 and 1995, regionally averaged over the SPG box and computed relative to the CORE-IA 1961–2007 climatology. The panels are thus meant to correspond to (a) strong, (b) medium, and (c) weak winter NAO conditions in the pentad leading up to the observed regime shift. The ensemble mean is included in (b).

and subpolar gyre strength over the latter part of the twentieth century (Fig. 2e). The fidelity of low frequency variations in North Atlantic circulation strength, which we suspect are largely driven by NAO-related forcing, in the ocean state used for initialization would thus appear to be of critical importance for skillful decadal prediction in the North Atlantic region. In uninitialized 20C runs, both AMOC and the subpolar gyre BSF decrease over the time period considered here (Fig. 2e) because the external forcing adds positive buoyancy to the surface

waters, which damps both deep convection and high-latitude gyre circulation. Thus, the weak but positive correlation of SPG heat content from the 20C ensemble with CORE-IA ($r = 0.31$ and 0.33 in first and second pentads, respectively) seen in Fig. 4a, which might be interpreted as a baseline predictive skill associated with the external forcing, is in fact obtained through incorrect physical mechanisms. In the 20C ensemble, the upward trend in heat content is associated with a positive trend in SFLX (reduced surface cooling) that exceeds the negative trend in ADV heating (associated with the weakening AMOC and SPG gyre circulation). This is the opposite of what we believe actually occurred in nature between about 1970 and 2000 (Fig. 3a).

The 1991DP is a particularly revealing example of how the information embedded in the ocean state can translate into predictive skill on decadal time scales. Despite widely disparate predictions of atmospheric conditions in the 1990s by individual members (Fig. 9), the ensemble mean predicts an unambiguous ocean warming associated with large, preconditioned advective and diffusive heat fluxes (Fig. 5). However, the prognosed atmospheric conditions influence the timing of the warming (Fig. 9), so the lack of skill in predicting NAO conditions is the explanation for why the 1991DP predicts an early, less-abrupt regime shift. We speculate that DP ensembles initialized in any January between about 1989 and 1995 would yield a heat content evolution quite similar to that seen in the 1991DP because the ocean state over these years exhibits the strong heat content dipole structure (Fig. 2c) resulting from NAO preconditioning and because there would be low skill in predicting the observed NAO conditions in subsequent years. This speculation could be tested in the future with a full suite of DP ensembles initialized each successive year. Robson (2010) discusses such behavior in the DePreSys system of the Met Office, but attributes early regime shifts to errors in the assimilated North Atlantic density anomalies in the early 1990s, rather than to poor NAO prediction as we hypothesize here.

We have shown that bias correction methods routinely used in seasonal forecasting can be used in the decadal prediction context to derive skillful forecasts of climatically relevant fields out to at least 10 years. This case study leads us to conclude that decadal prediction with CCSM4 using a full-field initialization approach offers significant promise, although by considering only upper-ocean temperature in the North Atlantic we have admittedly chosen both the field and region expected to offer the greatest decadal-scale predictability in the climate system (Branstator and Teng 2010). The limited focus has allowed us to probe the mechanisms at work and thus to gain a deeper understanding of DP results

and shortcomings. We find that external forcing does not appear to be contributing to skill in the SPG region in a manner consistent with the actual historical mechanisms. The skill derives mainly from accurate simulation of large-scale heat advection, which we think can be attributed primarily to the good representation of ocean circulation anomalies in the initial conditions used. There is generally poor skill at predicting NAO-related surface heat and wind stress fluxes and, in our view, this represents the greatest obstacle to prediction on time scales longer than 10 years in this region. This deficiency may be partly due to the fact that atmosphere and land initial conditions were not carefully chosen for historical accuracy, but there is also a suggestion from this analysis that North Atlantic surface buoyancy forcing is probably too anemic in the CCSM4 model. For example, none of the individual ensemble members of the 1991DP was able to simulate persistent positive NAO between 1991 and 1995 at a level to match that in CORE-IA (Fig. 9). There are known problems with the representation of planetary boundary layers in the CCSM4 model (Bates et al. 2012) that could explain this shortcoming, so ongoing model development efforts offer potential for improved decadal prediction in the years to come.

Acknowledgments. We would first like to acknowledge the dedicated effort of all the scientists and software engineers who contributed to the development of the CCSM4. We also thank three anonymous reviewers for their insightful comments on an earlier version of this paper. This work was supported by the NOAA Climate Program Office under Climate Variability and Predictability Program Grant NA09OAR4310163, by the Office of Science, Biological and Environmental Research, U.S. Department of Energy, Cooperative Agreement No. DE-FC02-97ER62402, and by the National Science Foundation through its sponsorship of the National Center for Atmospheric Research. This research used computing resources provided by the NCAR Computational and Information Systems Laboratory and by the Oak Ridge Leadership Computing Facility, which is supported by the Office of Science of the U.S. Department of Energy under Contract DE-AC05-00OR22725.

APPENDIX

Significance Testing for Correlation Skill

From a statistical perspective, when we ask whether or not a given result is significant, we are inquiring about the likelihood that it would have occurred by chance. Making this assessment requires that we have a probabilistic null

hypothesis against which we are measuring our result. It is of particular importance to choose a null hypothesis that represents a reasonable alternative explanation for the result in question. Equally important is the choice of a test statistic that quantifies our result in a meaningful way and can be computed for the null hypothesis. Throughout this paper, we have chosen to use the Pearson correlation coefficient as the measure of our skill, and hence as our test statistic.

For the purposes of this work, we have chosen a null hypothesis that can be stated as follows: the predictive skill (as measured by our pentadal correlation test statistic) can be explained by a first order autoregressive process initialized along the hindcast trajectory, but without explicit ocean dynamics. We can express this statistical model as

$$z_t = f_t + \alpha z_{t-1} + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2). \quad (\text{A1})$$

Here z_t is the annual average of the variable of interest over a given year t , α is the damping coefficient of the AR(1) model, and f_t is a linear trend. The probabilistic variable ε is normally distributed with zero mean and variance σ^2 . A statistical model of this form is chosen because it takes into account the variance in the time series as well as the autocorrelation that stems both from short-term persistence and from long-term trends in the data. It also implicitly accounts for a baseline level of skill that will stem from initializing in either anomalously high or low states.

We use a single 155-yr time series from one 20C realization to estimate the parameters of this statistical model for each of the SPG heat budget terms outlined in the text. Training the model on 60-yr CORE-IA time series did not give meaningfully different results, so the longer 20C time series was chosen for robustness. The linear trend is determined independently from 1850 to 1960 and from 1961 to 2005 by simple least squares minimization. We fix the pivot point in 1960 a priori because of a change in trend around that time. Once the trend has been removed, we estimate the remaining parameters α and σ^2 . Details of the least squares fitting algorithm for the AR(1) parameters are documented in Schneider and Neumaier (2001).

The strategy for assessing the significance of a correlation skill is to build a distribution of possible correlation scores that could be realized by applying the AR(1) model described above instead of the dynamical prediction with CCSM4. In statistical parlance, this is called a parametric bootstrap. Consistent with the DP forecasting method, the statistical model is initialized at each of the 10 start years along the hindcast trajectory. For each initialization date, the mean of 10 realizations

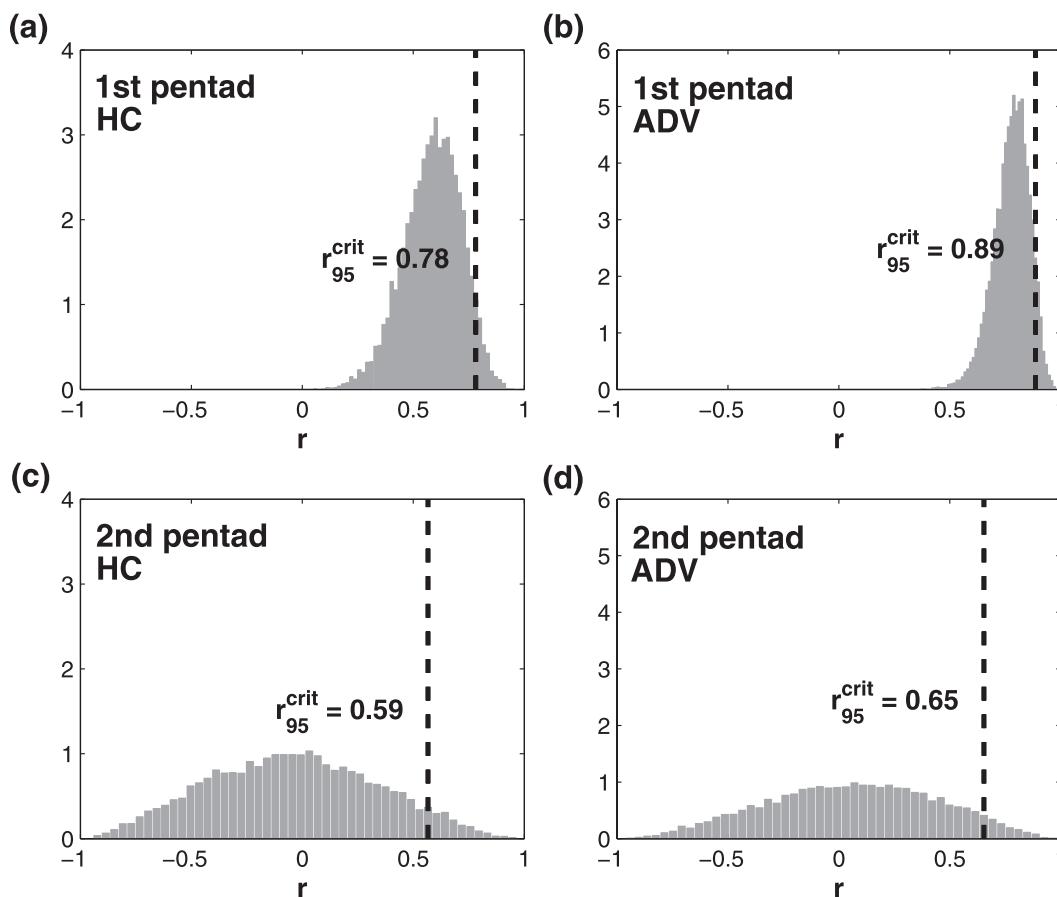


FIG. A1. Bootstrapped probability density functions for the correlations of our autoregressive statistical forecasts with CORE-IA heat content and total ADV tendency for the SPG region: (a),(c) the distributions for the first and second pentads for heat content and (b),(d) distributions for the ADV term. The dashed lines mark the critical values below which 95% of the bootstrapped correlations are found. These distributions suggest that there is a higher baseline level of predictability in the first pentad that stems from anomalous initial conditions.

(ensemble members) of (A1) is computed for the first and second pentad following the initialization. The correlation score for the first pentad is computed from the collection of first pentad forecasts ($n = 9$) and the corresponding CORE-IA pentads. The correlation score for the second pentad is computed similarly except that $n = 8$.

As an illustration, Fig. A1 shows bootstrapped probability density functions (PDFs) for SPG heat content and the total ADV tendency term. Ninety-five percent of the bootstrapped correlations are less than the critical value marked by the dashed line (i.e., it is the 95% confidence level at which the null hypothesis may be rejected). Note that in the example variables we show here, the null hypothesis forecast is different for the first and second pentad. The positive mean in the first pentad reflects the baseline level of predictive skill that can emerge simply from statistical persistence of anomalous initial conditions. Because the parameters of the statistical model (and the hindcast initialization states) are

different for each of the variables considered in this paper, the critical value corresponding to a 95% confidence level is variable specific. These are listed in Table 1 and Figs. 5–8 for each variable referenced in the body of the paper for both first and second pentad predictions.

REFERENCES

- Bates, S. C., B. Fox-Kemper, S. R. Jayne, W. G. Large, S. Stevenson, and S. G. Yeager, 2012: Mean biases, variability, and trends in air–sea fluxes and sea surface temperature in the CCSM4. *J. Climate*, in press.
- Branstator, G., and H. Teng, 2010: Two limits of initial-value decadal predictability in a CGCM. *J. Climate*, **23**, 6292–6311.
- CLIVAR, 2011: Data and bias correction for decadal climate predictions. CLIVAR Publ. Series 150, International CLIVAR Project Office, 4 pp.
- Danabasoglu, G., S. Bates, B. Briegleb, S. Jayne, M. Jochum, W. Large, S. Peacock, and S. Yeager, 2012a: The CCSM4 ocean component. *J. Climate*, **25**, 1361–1389.

- , S. G. Yeager, Y.-O. Kwon, J. Tribbia, A. Phillips, and J. Hurrell, 2012b: Variability of the Atlantic meridional overturning circulation in CCSM4. *J. Climate*, **25**, 2077–2103.
- Dunstone, N. J., D. M. Smith, and R. Eade, 2011: Multi-year predictability of the tropical Atlantic atmosphere driven by the high latitude North Atlantic Ocean. *Geophys. Res. Lett.*, **38**, L14701, doi:10.1029/2011GL047949.
- Eden, C., and J. Willebrand, 2001: Mechanism of interannual to decadal variability of the North Atlantic circulation. *J. Climate*, **14**, 2266–2280.
- Flatau, M. K., L. Talley, and P. P. Niiler, 2003: The North Atlantic oscillation, surface current velocities, and SST changes in the subpolar North Atlantic. *J. Climate*, **16**, 2355–2369.
- Gent, P. R., and Coauthors, 2011: The Community Climate System Model version 4. *J. Climate*, **24**, 4973–4991.
- Griffies, S. M., and Coauthors, 2009: Coordinated ocean-ice reference experiments (COREs). *Ocean Modell.*, **26**, 1–46.
- Hakkinen, S., and P. B. Rhines, 2004: Decline of subpolar North Atlantic circulation during the 1990s. *Science*, **304**, 555–559.
- Hatun, H., and Coauthors, 2009: Large bio-geographical shifts in the northeastern Atlantic Ocean: From the subpolar gyre, via plankton, to blue whiting and pilot whales. *Prog. Oceanogr.*, **80**, 149–162.
- Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1107.
- Holland, D. M., R. H. Thomas, B. de Young, M. H. Ribergaard, and B. Lyberth, 2008: Acceleration of Jakobshavn Isbrae triggered by warm subsurface waters. *Nature*, **1**, 659–664.
- Hurrell, J. W., 1995: Decadal trends in the North Atlantic oscillation: Regional temperatures and precipitation. *Science*, **269**, 676–679.
- , J. J. Hack, D. Shea, J. M. Caron, and J. Rosinski, 2008: A new sea surface temperature and sea ice boundary dataset for the community atmosphere model. *J. Climate*, **21**, 5145–5153.
- Ishii, M., and M. Kimoto, 2009: Reevaluation of historical ocean heat content variations with time-varying XBT and MBT depth bias corrections. *J. Oceanogr.*, **65**, 287–299.
- Johns, W. E., and Coauthors, 2011: Continuous, array-based estimates of the Atlantic Ocean heat transport at 26.5°N. *J. Climate*, **24**, 2429–2449.
- Keenlyside, N., M. Latif, J. Jungclaus, L. Kornbleuh, and E. Roeckner, 2008: Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature*, **453**, 84–88.
- Knight, J. R., C. K. Folland, and A. A. Scaife, 2006: Climate impacts of the Atlantic multidecadal oscillation. *Geophys. Res. Lett.*, **33**, L17706, doi:10.1029/2006GL026242.
- Large, W. G., and S. G. Yeager, 2004: Diurnal to decadal global forcing for ocean and sea ice models: The datasets and climatologies. NCAR Tech. Note NCAR/TN-460+STR, 105 pp.
- , and —, 2009: The global climatology of an interannually-varying air–sea flux data set. *Climate Dyn.*, **33**, 341–364.
- Levitus, S., J. I. Antonov, T. P. Boyer, R. A. Locarnini, H. E. Garcia, and A. V. Mishonov, 2009: Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems. *Geophys. Res. Lett.*, **36**, L07608, doi:10.1029/2008GL037155.
- Lohmann, K., H. Drange, and M. Bentsen, 2009a: A possible mechanism for the strong weakening of the north atlantic subpolar gyre in the mid-1990s. *Geophys. Res. Lett.*, **36**, L15602, doi:10.1029/2009GL039166.
- , —, and —, 2009b: Response of the North Atlantic subpolar gyre to persistent North Atlantic oscillation like forcing. *Climate Dyn.*, **32**, 273–285, doi:10.1007/s00382-008-0467-6.
- Lozier, M. S., S. Leadbetter, R. G. Williams, V. Roussenov, M. S. C. Reed, and N. J. Moore, 2008: The spatial pattern and mechanisms of heat-content change in the North Atlantic. *Science*, **319**, 800–803, doi:10.1126/science.1146436.
- Marshall, J., H. Johnson, and J. Goodman, 2001: A study of the interaction of the North Atlantic Oscillation with ocean circulation. *J. Climate*, **14**, 1399–1421.
- Meehl, G. A., W. M. Washington, J. M. Arblaster, A. Hu, H. Teng, C. Tebaldi, W. G. Strand, and J. B. White III, 2012: Climate system response to external forcings and climate change projections in CCSM4. *J. Climate*, **25**, 3661–3683.
- Meehl, J., and Coauthors, 2009: Decadal prediction: Can it be skillful? *Bull. Amer. Meteor. Soc.*, **90**, 1467–1485.
- Pohlmann, H., J. H. Jungclaus, A. Köhl, D. Stammer, and J. Marotzke, 2009: Initializing decadal climate predictions with the GECCO oceanic synthesis: Effects on the North Atlantic. *J. Climate*, **22**, 3926–3938.
- Robson, J. I., 2010: Understanding the performance of a decadal prediction system. Ph.D. thesis, University of Reading, 233 pp.
- , R. Sutton, K. Lohmann, and D. Smith, 2012: Causes of the rapid warming of the North Atlantic Ocean in the mid-1990s. *J. Climate*, **25**, 4116–4134.
- Schneider, T., and A. Neumaier, 2001: Algorithm 808: Arfit—A Matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Trans. Math. Software*, **27**, 58–65.
- Smith, D. M., and J. M. Murphy, 2007: An objective ocean temperature and salinity analysis using covariances from a global climate model. *J. Geophys. Res.*, **112**, C02022, doi:10.1029/2005JC003172.
- , 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.
- , R. Eade, N. J. Dunstone, D. Fereday, J. M. Murphy, H. Pohlmann, and A. A. Scaife, 2010: Skillful multi-year predictions of Atlantic hurricane frequency. *Nat. Geosci.*, **3**, 846–849, doi:10.1038/NNGEO1004.
- Solomon, A., and Coauthors, 2011: Distinguishing the roles of natural and anthropogenically forced decadal climate variability: Implications for prediction. *Bull. Amer. Meteor. Soc.*, **92**, 141–156.
- Stockdale, T., 1997: Coupled ocean–atmosphere forecasts in the presence of climate drift. *Mon. Wea. Rev.*, **125**, 809–818.
- Sutton, R. T., and D. L. R. Hodson, 2005: Atlantic Ocean forcing of North American and European summer climate. *Science*, **309**, 115–118.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498.
- Troccoli, A., and T. N. Palmer, 2007: Ensemble decadal predictions from analyzed initial conditions. *Philos. Trans. Roy. Soc. London*, **A365**, 2179–2191.
- Zhang, R., and T. Delworth, 2006: Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes. *Geophys. Res. Lett.*, **33**, L17712, doi:10.1029/2006GL026267.