

Decadal prediction skill in the GEOS-5 forecast system

Yoo-Geun Ham · Michele M. Rienecker · Max J. Suarez ·
Yury Vikhliav · Bin Zhao · Jelena Marshak ·
Guillaume Vernieres · Siegfried D. Schubert

Received: 9 July 2012 / Accepted: 24 June 2013 / Published online: 14 December 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract A suite of decadal predictions has been conducted with the NASA Global Modeling and Assimilation Office's (GMAO's) GEOS-5 Atmosphere–Ocean general circulation model. The hind casts are initialized every December 1st from 1959 to 2010, following the CMIP5 experimental protocol for decadal predictions. The initial conditions are from a multi-variate ensemble optimal interpolation ocean and sea-ice reanalysis, and from GMAO's atmospheric reanalysis, the modern-era retrospective analysis for research and applications. The mean forecast skill of a three-member-ensemble is compared to that of an experiment without initialization but also forced with observed greenhouse gases. The results show that initialization increases the forecast skill of North Atlantic sea surface temperature compared to the uninitialized runs, with the increase in skill maintained for almost a decade over the subtropical and mid-latitude Atlantic. On the other hand, the initialization reduces the skill in predicting the warming trend over some regions outside the Atlantic. The annual-mean atlantic meridional overturning circulation

index, which is defined here as the maximum of the zonally-integrated overturning stream function at mid-latitude, is predictable up to a 4-year lead time, consistent with the predictable signal in upper ocean heat content over the North Atlantic. While the 6- to 9-year forecast skill measured by mean squared skill score shows 50 % improvement in the upper ocean heat content over the subtropical and mid-latitude Atlantic, prediction skill is relatively low in the subpolar gyre. This low skill is due in part to features in the spatial pattern of the dominant simulated decadal mode in upper ocean heat content over this region that differ from observations. An analysis of the large-scale temperature budget shows that this is the result of a model bias, implying that realistic simulation of the climatological fields is crucial for skillful decadal forecasts.

Keywords Decadal prediction · AMOC · Decadal variability · GEOS-5 AOGCM

Y.-G. Ham (✉) · M. M. Rienecker · M. J. Suarez ·
Y. Vikhliav · B. Zhao · J. Marshak · G. Vernieres ·
S. D. Schubert
NASA Goddard Space Flight Center (GSFC/NASA),
Global Modeling and Assimilation Office, Code 610.1,
Greenbelt, MD, USA
e-mail: ygham@jnu.ac.kr

Y.-G. Ham
Goddard Earth Science Technology & Research,
Morgan State University, Baltimore, MD, USA

Y.-G. Ham
Faculty of Earth Systems and Environmental Sciences,
Chonnam National University, Gwangju, Korea

Y. Vikhliav
Goddard Earth Sciences Technology and Research,
Universities Space Research Association,
Columbia, MD, USA

B. Zhao
Science Applications International Corporation (SAIC),
McLean, VA, USA

G. Vernieres
Science Systems and Applications, Inc.,
Lanham, MD, USA

1 Introduction

The climate research community has increasingly turned its attention to the topic of decadal prediction (Zhang and Delworth 2005; Meehl et al. 2009; van Oldenborgh et al. 2012; Chikamoto et al. 2012). This is in part user driven, due to the potentially large socio-economic impacts of skillful predictions on these time scales. It is also in part driven by the science, since it is a natural extension of the work done by both the seasonal-to interannual (SI) prediction community (extending to longer time scales) and the climate change community (extending to shorter timescales). It is now generally accepted that the decadal problem requires information of both initial and boundary conditions (Smith et al. 2007; Keenlyside et al. 2008). This contrasts with the seasonal-to-interannual prediction problem where the initial condition is the key, and the centennial time-scale problem where changing boundary conditions dominate the projections. In that sense, the decadal prediction problem bridges the gap between the two timescales, with the goal of providing useful information on near-future climate.

However, it has only been in the last several years that a few pioneering efforts have made progress in assessing the importance of initialization for decadal forecasts (e.g., Smith et al. 2007; Keenlyside et al. 2008; Pohlmann et al. 2009), while others highlighted the importance of changing boundary conditions, such as those associated with anthropogenic greenhouse gases (Wetherald et al. 2001; Meehl et al. 2005). While phase 3 of the coupled model intercomparison project (CMIP3) focused mostly on the forced response of the dynamical models, phase 5 (CMIP5) specifically includes an assessment of the ability of dynamical models to simulate and predict decadal variability including the impact of initialization.

The increase in decadal forecast skill from initialization, especially over the North Atlantic and Pacific Oceans, has been attributed to the local decadal variability in those ocean basins (Read and Gould 1992; Mantua et al. 1997). Keenlyside et al. (2008) showed that a simple initialization process using observed sea surface temperature (SST) leads to improvements in forecast skill particularly in the North Atlantic. Similarly, Robson (2010) found that the regional improvements through initialization were found mainly in the North Atlantic Ocean. It is widely believed that the source of decadal predictability over the Atlantic Ocean originates from decadal fluctuations of the Atlantic meridional overturning circulation (AMOC). This is consistent with previous observational (Enfield et al. 2001; Huang et al. 2012) and modeling studies (Dong and Sutton 2005; Zhang et al. 2007; Hawkins and Sutton 2011), indicating that the strength of the AMOC exhibits a strong multi-decadal variability. For the Pacific Ocean, Mochizuki

et al. (2010) demonstrated that initialization of their dynamical model leads to skillful prediction of upper-ocean temperature in the regions typically affected by the Pacific decadal oscillation (PDO, Mantua et al. 1997).

While the above studies show that some progresses have already been made in demonstrating skill in decadal predictions using dynamical models, there are a number of outstanding problems that limit current skill. In particular, the initialization process used in those studies is quite simple compared to that used for seasonal climate and weather forecasts. For example, Keenlyside et al. (2008) use a simple approach that assimilates only sea surface temperature in order to avoid the difficulties associated with the sparse historical subsurface ocean observations. At the National Center for Atmospheric Research (NCAR) and Max-Planck Institute (MPI), an alternative approach has been tested where atmospheric reanalyses are used to force the subsurface ocean temperature and salinity (Meehl et al. 2009; Matei et al. 2012a). While these efforts avoid having to deal with sparse, and in some cases biased, subsurface ocean data (Ishii and Kimoto 2009), it is nevertheless to be expected that the direct use of subsurface ocean observations would improve forecast skill (Meehl et al. 2009; Doblas-Reyes et al. 2011).

The other challenge in decadal prediction skill arises from model deficiencies, which can degrade forecast skill. For example, model biases can lead to erroneous oceanic responses to imposed anthropogenic forcing, which can, in turn, lead to erroneous decadal modes of variability as a result of the different future climate (Meehl et al. 2007). Even initialized models will evolve to the model's own climate during the forecast (e.g., Lee et al. 2010), resulting in a forecast bias that is a function of the forecast lead time. Particularly for the long leads in decadal prediction, the quality of the model's simulated climatology may determine the realism of the decadal variability and the decadal forecast skill. However, this potential connection between the quality of the simulated climatological fields and the quality of decadal predictions is yet to be examined in any detail.

To contribute to the assessment of the current skill of decadal forecasts and examine the key factors affecting such skill, this study analyzes the decadal forecasts from version 5 of the Goddard Earth Observing System coupled atmosphere–ocean general circulation model (GEOS-5 AOGCM) developed at the NASA Global Modeling and Assimilation Office (GMAO). Note that the initialization process employed for GEOS-5 utilizes most of the available oceanic observations in order to consider the observed subsurface variation as well as the surface variations (Doblas-Reyes et al. 2011). The paper is organized as follows. The model description and experimental design for the decadal forecasts are provided in Sect. 2. Section 3

describes the basic results of the decadal predictions, and Sect. 4 examines the possible relationship between the simulated climatology and decadal modes of variability to determine the quality of decadal prediction skill over the Atlantic. The summary and conclusion are presented in Sect. 5.

2 Model description and experimental design

2.1 Model description

The main components of the GEOS-5 AOGCM are the GEOS-5 atmospheric model and the catchment land surface model (Rienecker et al. 2008, Molod et al. 2012), the Modular Ocean Model, version 4 (MOM4, Griffies et al. 2005) and the Los Alamos sea ice model (CICE model) (Hunke and Lipscomb 2008). The configuration of the atmospheric component uses a 2.5° longitude \times 2° latitude grid, with 72 vertical levels extending to 0.01 hPa. The ocean model configuration is 1° longitude and latitude, with a meridional equatorial refinement to $1/3^\circ$, and 50 vertical levels. The vertical grid spacing is a constant 10 m over the top 225 m, and gradually increases up to 360 m in the deep ocean. These two components exchange fluxes of momentum, heat and fresh water every time step through a skin layer interface. The skin layer includes the parameterization of the diurnal cycle in the near-surface ocean, and the thermodynamics of CICE. The atmospheric component includes a river runoff routing scheme and an aerosol model based on the Goddard Chemistry Aerosol Radiation and Transport (GOCART) model using emissions prescribed by the CMIP5 protocol. Only those volcanic aerosols from continually outgassing volcanoes have been included, i.e., the explosive volcanic eruptions are not included. All components are coupled together using the earth system modeling framework (ESMF) interface.

2.2 Experimental design

2.2.1 Initial conditions for the decadal forecasts

The initial conditions for the decadal forecasts/hind casts are obtained from an ocean and sea-ice assimilation performed using the GEOS-5 AOGCM and a multi-variate ensemble optimal interpolation (EnOI) analysis scheme (Vernieres et al. 2012), while the atmosphere is constrained by MERRA (Rienecker et al. 2011) from 1979 to 2005 and a related atmospheric analysis prior to 1979. The atmospheric analysis prior to 1979 is produced with the same observing and assimilation system as MERRA, but the atmospheric model has coarser resolution. The EnOI for the ocean and sea ice assimilation is a sequential ensemble

assimilation method, where the error covariance is estimated from an ensemble of multi-year simulations. The background error covariances are estimated from a static ensemble, and flow dependency of the background error is obtained through localization in density space. Details are provided in Vernieres et al. (2012).

The assimilated ocean profile observations consist of temperature and salinity profiles from expendable bathythermographs (XBTs) and conductivity temperature depth (CTD) sensors extracted from the EN3 data base (Ingleby and Huddleston 2007) with time-varying XBT corrections applied according to Levitus et al. (2009), temperature observations from the tropical moored buoy array (McPhaden et al. 2010), and Argo temperature and salinity profiles from the Argo Global Data Assembly Center (GDAC, see <http://www.usgodae.org/argo/argo.html>). Along-track sea level anomalies from the archiving, validation and interpretation of satellite oceanographic (AVISO) merged product were also assimilated as was gridded sea surface temperature (SST). For the latter, NOAA daily SST of Reynolds et al. (2007) was used from 1982 to present, and HadISST1 (Rayner et al. 2003; Hurrell et al. 2008) prior to 1982. Unfortunately, the difference in climatology for the updated Reynolds and that used for HadISST1 was not recognized until the forecasts were analyzed. Sea-ice concentration from the National Snow and Ice Data Center (NSIDC) from 1979 onwards and the CMIP5 sea ice concentration prior to 1979 were also assimilated. Prior to the Argo period, the model biases were corrected by assimilating 10 % of the temperature and salinity profiles, randomly selected, from the World Ocean Atlas 2009 (WOA09) gridded climatology (Antonov et al. 2010; Locarnini et al. 2010) every 10 days with very small weights. The assimilation product is used to provide the initial conditions for decadal forecasts, as well as to evaluate the results. In the following, we refer to this assimilation product as the GMAO ocean reanalysis (ORA).

The ORA was processed in three separate streams, each initialized from the WOA09 climatology and spun up by assimilating available observations prior to the official starting date of the stream. Stream 1 is for the period January 1, 1959 to December 31, 1980, followed by stream 2 from January 1, 1981 to December 31, 1997, and stream 3 from January 1, 1997 to December 31, 2010. A series of 10-year hind casts were initialized from this ORA and the corresponding atmospheric and land analyses, starting every December 1 from 1959 to 1980 and from 1985 to 2010. Results from forecasts from the initial conditions spanning December 1981 to 1984 are not included in the analysis because it was found that by 1981 the ORA fields from the first stream had degraded due to a loss of water volume and those from the second stream were still

undergoing a spin-up adjustment. Unfortunately, because of a change in the ORA software, the missing years were not able to be re-processed.

Predictions were made with three ensemble members for each start date. One ensemble member started from the assimilated state, the other two started with perturbations to the assimilated state, with perturbations generated using a breeding approach. In this study, the breeding method, which has been shown to be beneficial to ensemble forecasting at various time-scales (Toth and Kalnay 1993, 1997; Yang et al. 2008; Ham et al. 2012a, b), was applied with a 5-year rescaling interval starting from the year 1959. A two-sided breeding method was used so that the rescaling norm was selected as the root mean square (RMS) difference of the monthly-mean heat content averaged from the surface to 500 m over the Atlantic (90°W–20°E, 20–70°N) from the two perturbed integrations. The use of two-sided breeding removes the influence of model drift during the integration. At the end of every breeding cycle, the amplitude was reduced to 10 % of the natural variability over the region used for the norm. The choice of a 5-year rescaling interval was based on the results from Vikhliav et al. (2007) and on sensitivity tests that showed that the growth rate saturated on that time scale. The initial perturbation for the atmosphere in the first breeding cycle was taken from a 6-hour difference on December 1, 1954. There was no initial perturbation in the ocean at the first breeding cycle, i.e., the initial ocean states of the ocean for the bred vectors (BV) and control runs were identical.

2.2.2 Twentieth century simulations

To assess the model's response to increased anthropogenic forcing during the Twentieth century, a three-member ensemble of coupled model simulations was conducted with observed greenhouse gases (GHG, i.e. CO₂, BC, OC, CH₄, Sulfur, NO_x, VOC, CO and NH₃) from 1950 to 2010, as specified by the CMIP5 protocol. Initial conditions for the Twentieth century simulations were taken from a 200-year simulation with perpetual 1950 boundary conditions. Each ensemble member was initialized with same year-date but 1 year apart. In the following, this experiment will be denoted as C20C simulations.

3 Decadal predictability in the GMAO forecast system

Figure 1 shows the time series of annual mean SST averaged over the region 0–360°E, 60°S–60°N from the reanalysis, the ensemble mean C20C simulation, and the ensemble mean decadal predictions. The upper panel shows the raw value, and the lower panel shows the

anomaly, which has been bias-corrected by removing the mean bias away from the reanalysis. For the decadal prediction, the bias-corrected anomaly is calculated as $\hat{Y}_{jt} = Y_{jt} - \sum_{k=1}^N (Y_{kt} - O_{kt})/N - \bar{O}$, where Y_{jt} , and \hat{Y}_{jt} are the raw and bias-corrected anomalous predictions, respectively, at the j th initialized year and forecast lead year t (ICPO 2011; Eq. (1) of Smith et al. 2013). Here, $\bar{O} = \sum_{k=1960}^{2010} O_k/N$ is the climatology from the reanalysis (in this study, GMAO ORA) or from SST observations. The anomaly of the C20C simulation and reanalysis is obtained by subtracting its own climatology from 1960 to 2010 (i.e. $\hat{X}_j = X_j - \bar{X}$, where \hat{X}_j , and X_j are the raw and anomalous value, and $\bar{X} = \sum_{k=1960}^{2010} X_k/N$).

The HadISST1 (Rayner et al. 2003) shows a general global warming signal beginning in the 1980s. This is less so for the GMAO ORA, which shows a cooling during the early 1980s after which the warming is in general agreement with the HadISST1. This apparent cooling signal results from the switch from the HadISST1 to the daily Reynolds et al. (2007) data set in 1982, as described above. However, the variability is quite similar in both, implying a similarity in decadal variability between the two.

The time series of the annual mean (from January to December) from the first year of the forecast (labeled year 1 lead forecast) exhibits a warm bias although the warming trend is similar to that of the ORA. Also, the year-to-year variability is again quite similar to that of the ORA. The time series of the 2-year lead forecasts of annual mean SST shows a stronger warm bias and global warming signal than the 1-year lead forecasts. This bias is not shown in the bias-corrected anomaly. In contrast to the 1-year-lead forecast, a sudden drop of global mean temperature between forecast starting at 1981 to 1984 in the 2-year lead forecasts is clear, due to problems in the initial condition as mentioned earlier. Note that forecasts initialized from 1981 to 1984 have not been used to validate the decadal forecast skill. The tendency for the global warming trend to become stronger as the forecast lead year increases is consistent with the excessive global warming trend in the C20C simulation. In the forecasts after 6-year leads, there is weak year-to-year variability and only the warming trend remains. This indicates that the SST response to the increased greenhouse gas emission is excessive in this configuration of the GEOS-5 AOGCM.

Figure 2 shows the spatial pattern of the annual mean SST bias as a function of forecast lead year using the raw prediction. That is, the bias is defined as $\sum_{k=1}^N (Y_{kt} - O_{kt})/N$. The four-year average (i.e. average for years 2–5 and 6–9) is used for easy comparison with other studies (e.g., Goddard et al. 2013; Kim et al. 2012). The bias has been averaged to a 5° latitude × 5° longitude

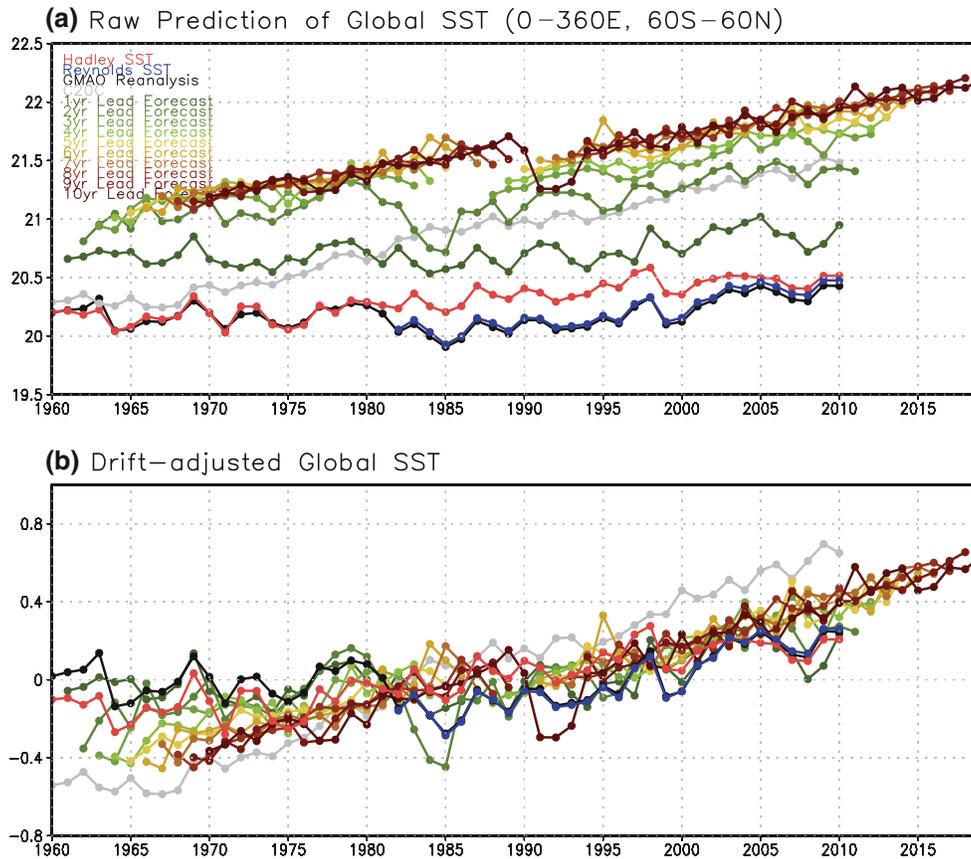


Fig. 1 The annual mean SST (°C) averaged over 0–360°E, 60°S–60°N from the HadISST1 (red), Reynolds SST (blue), the GMAO ocean reanalysis (black), the GEOS C20C simulation (gray) and decadal forecasts as a function of the forecast lead time

grid, and is defined from the time-averaged values from 1960 to 2010. Note that the HadISST1 is used as verifying SST. The SST averaged over forecast leads of 2–5 years shows a warm bias of about 2 °C over the far eastern Pacific and Atlantic. In fact, a warm bias occurs over much of the equatorial and subtropical regions, while there is a cold bias over the high-latitudes poleward of about 50°. The biases increase with the forecast lead, with the magnitude reaching 3 °C at 6- to 9-year leads when the spatial pattern and magnitude of the SST bias is similar to that of C20C simulation. We note that the spatial pattern of the bias in the long-term simulation of the model with fixed (1950) CO₂ concentration is quite similar to that in the 2- to 5-year lead forecast, except for Indian Ocean and mid-latitude eastern Pacific, where the bias is nearly zero and slightly negative, respectively, and the higher latitudes where the bias is more strongly negative.

Two skill measures are used to examine whether the initialization of the decadal forecasts increases the forecast skill compared to the C20C simulation, the anomaly correlation coefficient (ACC) and a mean squared skill score (MSSS; Murphy 1988; Goddard et al. 2013). Note that both are calculated using the bias-corrected anomaly shown in

Fig. 1b (Goddard et al. 2013), and this is the case for all subsequent figures except for the climatological values (i.e., Fig. 11). The MSSS is defined using the mean squared error of the C20C and decadal forecasts as follows:

$$MSSS = \frac{MSE_{C20C} - MSE_{F_{cst}}}{MSE_{C20C}}$$

where MSE_{C20C} and $MSE_{F_{cst}}$ denote the mean-squared-error (MSE) of anomalies in the C20C simulation and decadal forecast, respectively. The upper limit of this value is 1 when the decadal forecast is perfect, and a positive value denotes decadal forecast skill that is better than that of the C20C simulation in terms of MSE. In the following figures, only correlation coefficients that are non-zero with 90 % confidence level are shown, based on a one-tailed *t* test. For this test, the degrees of freedom are defined using the auto-correlation of the ORA. That is, the number of degrees of freedom is defined as the total number of samples (i.e., 51) divided by *n*, where the auto-correlation is above the 99 % confidence level for the (*n* – 1)-year lag, but not for the (*n*)-year lag. For example, where the 4-year-lag auto-correlation in the ORA is above the 99 % confidence level, and 5-year-lag auto-correlation

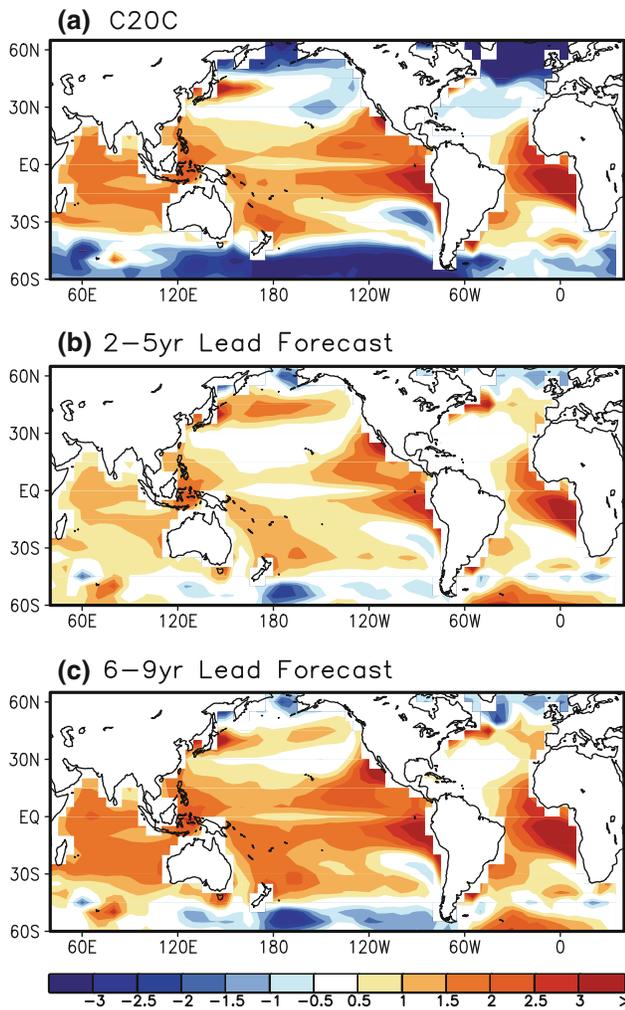


Fig. 2 The SST bias ($^{\circ}\text{C}$) relative to the HadISST1: **a** C20C simulation, **b** 2–5 year lead initialized forecast, **c** 6–9 year lead initialized forecast. The period for the verification is from 1960 to 2010

is below that level, the degree of freedom is defined as 10 (i.e. 51/5).

Figure 3 shows the ACC of the 4-year-averaged SST from the C20C simulations and the difference of the ACC for the initialized decadal forecasts from that for the C20C simulations. Note that the HadISST1 is used for validation since it is a gridded synthesis of available observations and provides the appropriate metric for the C20C simulations. Based on the ACC, there is a predictable signal in the C20C simulation over the Indian Ocean, western and south-central Pacific, and equatorial Atlantic. In addition, the ACC exceeds 0.6 over the mid-latitude western Atlantic. This skill is entirely due to the external forcing because the C20C simulation is not initialized. Therefore, the comparison of the skill in the C20C simulation to that in the decadal forecasts provides some information on added skill or on skill degradation due to initialization.

According to the anomaly correlation measure, the initialized prediction skill of 2- to 5-year lead forecasts over the subtropical, mid- and eastern North Atlantic is greater than that for the C20C simulation. This is significant with 90 % confidence level using the bootstrap approach as in Smith et al. (2010). This implies that the initialization has been successful in enhancing the prediction skill of Atlantic decadal variability. In addition, the prediction skill over the off-equatorial western Pacific is higher by up to 0.2 in the initialized decadal forecasts. In contrast, the skill in the subtropical southeastern Pacific and the extratropics in the southern Pacific and Indian sectors are lower in the initialized forecasts. This reduction in skill might be related to a distortion of the trend signal by the decadal variation in the initialization of forecasts over some areas in the Pacific and Indian Oceans. In addition, biases in the dominant decadal modes would also lead to a reduction in skill. The deficiency at high southern latitudes is related to the changes in the prescribed SST analysis from the HadISST1 to Reynolds for initialization in 1982. With the GMAO ORA used for validation, the correlation skill is systematically higher than that using HadISST1 and the negative difference over the southern hemisphere is not apparent (not shown here).

The MSSS shows more robust improvement with initialization compared with the anomaly correlation metric. The MSSS values are positive over most regions of the globe. However, exceptions are evident in localized areas over the mid-North Pacific and parts of the subpolar North Atlantic.

For the 6- to 9-year lead forecasts, the improvement in ACC has decreased over the subpolar North Atlantic compared with years 2–5, while that over the mid-latitude Atlantic and Pacific is even higher than in the C20C simulations. This indicates that the GEOS-5 AOGCM has some success in predicting the Atlantic climate up to about 10-year lead times except for the subpolar regions (Matei et al. 2012b). The MSSS results are similar for both leads. Thus, by these measures, initialization gives a systematic increase of prediction skill for SST over many areas of the globe, with the notable exception of the subpolar North Atlantic.

Kim et al. (2012) summarized the current prediction skill in CMIP5 decadal prediction systems. All of the prediction results show that the skill over the Pacific is lower than that over the Indian Ocean and Atlantic. This is consistent with other results showing that the prediction skill for the SST over the North Pacific is lower than that over the North Atlantic (e.g., van Oldenborgh et al. 2012). The spatial pattern of ACC over the Atlantic in the GEOS-5 system is similar to that in several models presented in Kim et al. (2012): three of the seven models shown in that study have relatively lower prediction skill over the subpolar (or

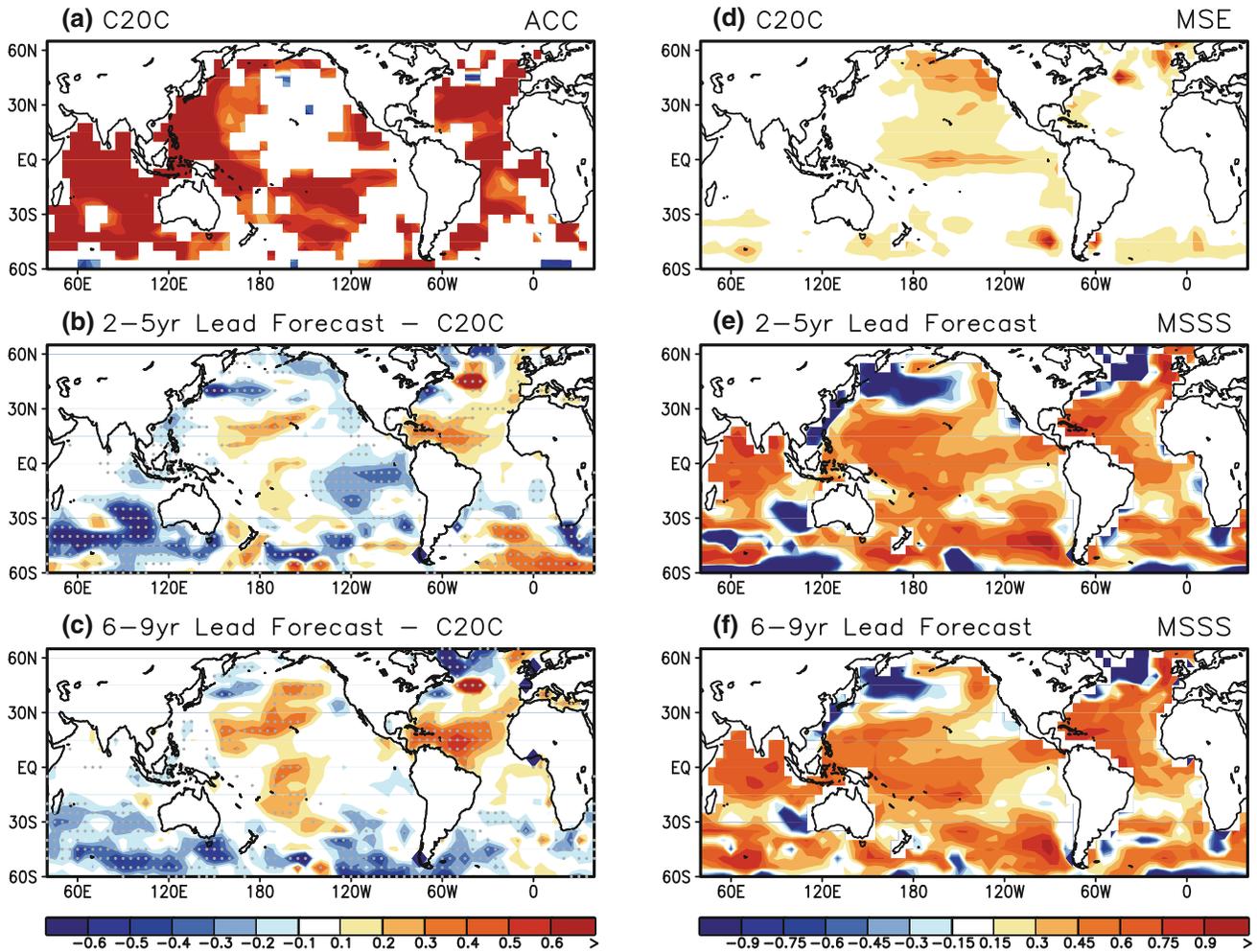


Fig. 3 **a** The anomaly correlation of 4-year moving averaged SST from the HadISST1 with the C20C simulation. The difference of anomaly correlation in the **b** 2–5-year lead forecast and **c** 6–9-year lead forecast from that in C20C. The right panel shows the **d** Mean square error (MSE) of the C20C simulation, Mean squared skill score

(MSSS) for the **e** 2–5-year lead forecast and **f** 6–9-year lead forecast. The *grey dots* **b**, and **c** denote that the difference in correlation is significant with 90 % confidence level. The procedure for the significant test is the same as in Smith et al. (2010)

western subpolar) Atlantic compared with other areas of the Atlantic, while the other models show similar skills over all regions of the Atlantic (Yeager et al. 2012; Robson et al. 2012a). This means that several decadal prediction systems have common issues, although the causes of such deficiencies are likely to be model-dependent. One possible reason for this problem in the GEOS-5 system will be discussed in the next section.

Figure 4 shows the skill measures for the 4-year-averaged heat content in the upper 500 m (defined as the average temperature from the surface to 500 m, hereafter HC500). The GMAO ORA is used as a validating product. Based on the anomaly correlation, there is significant skill in the C20C simulation in regions of the Atlantic and along the northern and eastern rim of the Pacific. The skill elsewhere is not significant and is even negative in a few places such as the high southern latitudes. The

improvement in ACC in the initialized decadal prediction of HC500 at 2- to 5-year leads is apparent in the central basins and high latitudes as well as the western Indian Ocean. On the other hand, the ACC over the eastern Indian Ocean and the eastern equatorial Pacific as well as over the western and subpolar North Atlantic has degraded in the initialized forecasts. It is not surprising that the areas of significant skill for HC500 are larger than for SST because the subsurface ocean varies more slowly than the ocean surface (Griffies and Bryan 1997). Degradations in skill by the initialization might occur in predicting the warming trend over some regions, presumably from internal adjustments after the initialization process. It is also possible that the decadal variability simulated in the model is systematically incorrect over some areas. The MSSS values, including over the subpolar North Atlantic, show patterns that are consistent with those from the ACC. The

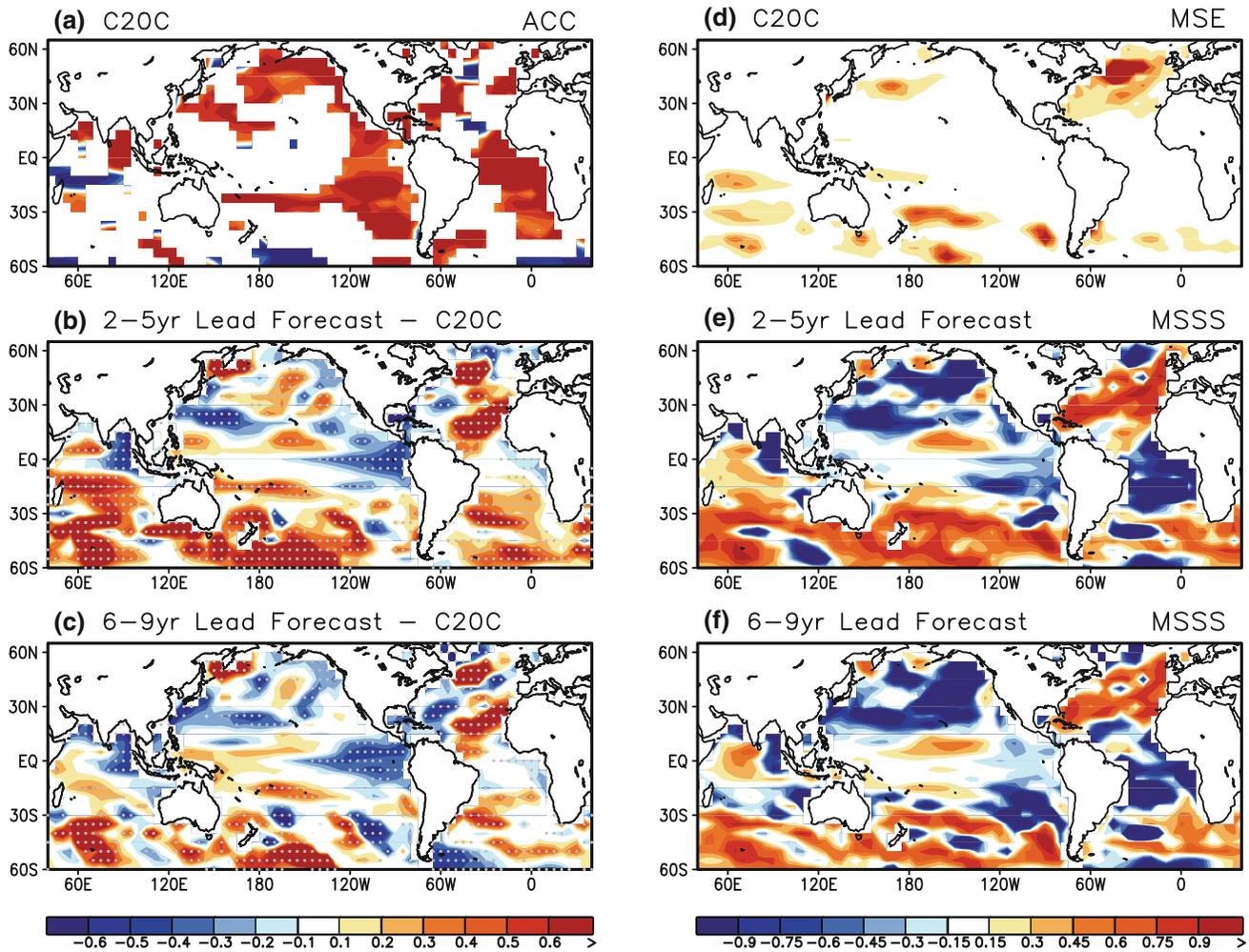


Fig. 4 As for Fig. 3, but for HC500

subpolar North Atlantic will be discussed in more detail below.

The patterns at 6- to 9-year leads are not very different from those at 2- to 5-year leads. There is still some predictable signal over the subtropical and mid-latitude Atlantic, suggesting that the decadal variability has some predictability up to 10-year leads over a limited area of the Atlantic.

One may ask why the MSSS score for SST is higher than that for HC500, since the subsurface is generally thought to have higher predictability (Griffies and Bryan 1997). In fact, the ACC skill of HC500 is generally higher than that of SST especially during early forecast years, consistent with previous studies. The higher MSSS score for SST compared to HC500 is due to the large MSE of SST in the C20C simulation, not due to the small MSE of SST in the decadal forecasts. The large MSE of SST in the C20C simulation is due to the excessive trend in the simulation, which is mainly in the surface layer. Therefore, the MSE is higher for SST than for HC500 (Figs. 3d, 4d).

Since the MSSS score denotes the improvement relative to the C20C simulation, it can be higher for SST even though the MSE of SST for the decadal forecasts is smaller than that of the HC500.

Up to now, we have compared the skill of the initialized decadal forecasts to that of the C20C simulations in order to separate the predictable signal due to the decadal variability from that due to the external forcing. An alternative approach is to calculate the prediction skill after removing the global warming trend. This is done here by removing the spatial pattern regressed onto observed GHG emissions (the prescribed forcing for both the decadal forecasts and the C20C simulations), following the approach of van Oldenborgh et al. (2012). Note that the GHG emissions used were those recommended for the CMIP5 experiments (<http://cmip-pcmdi.llnl.gov/cmip5/forcing.html#concentrations>). Due to the different magnitude of the trend, the de-trending is applied separately for each forecast lead and reanalysis output.

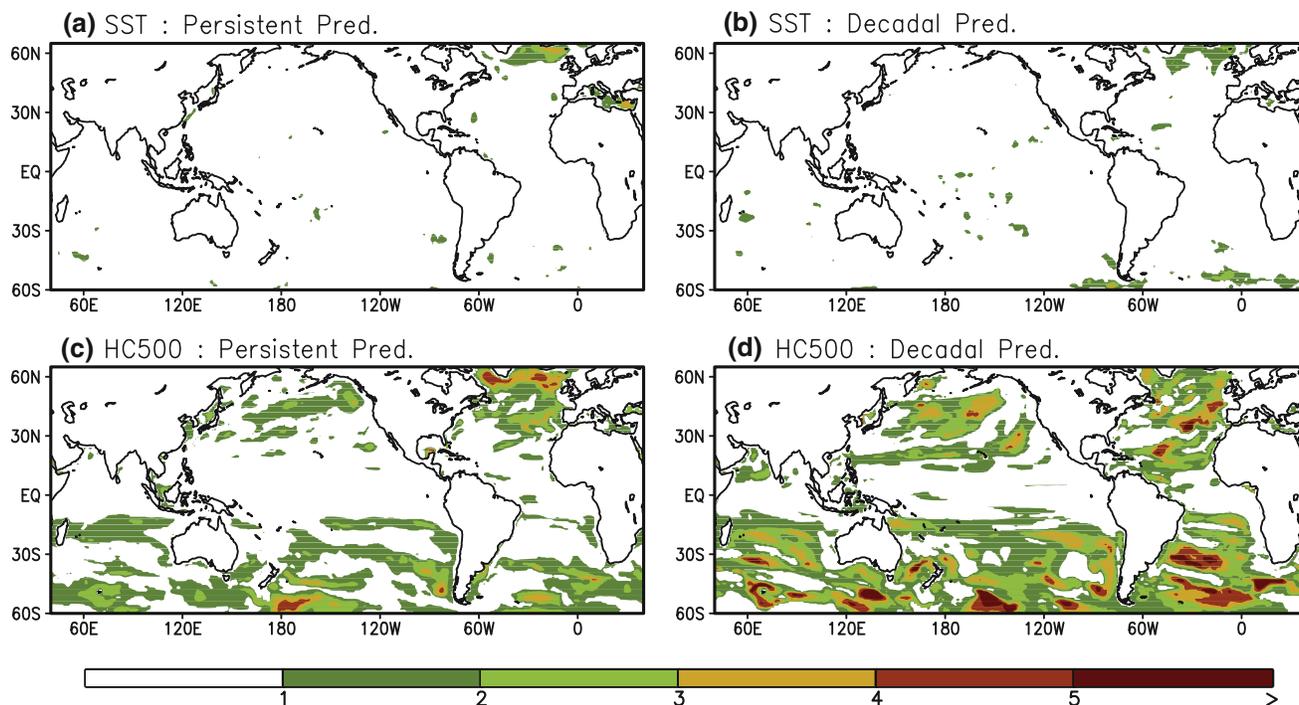


Fig. 5 The predictable year of de-trended annual mean SST in **a** persistence forecasts, and **b** decadal forecasts. The predictable year of annual mean HC500 in persistence forecasts and decadal forecasts

is shown in **c** and **d**, respectively. Note the predictable year is defined as the number of years for which the correlation skill is higher than 0.35

Figure 5 shows the number of years in which the ACC of the de-trended annual mean values is higher than 0.35 for the decadal hindcasts to measure the e-folding time scale (i.e. $1/e$ of the initial correlation value). This is compared with the ACC of persistence forecasts using the de-trended HadSST1 and GMAO ORA output for SST and HC500, respectively. The persistence forecasts show skill for SST for up to 1 or 2 years, primarily over the subpolar Atlantic region. In the case of HC500, the predictable signal extends from subpolar regions to the mid-latitude Atlantic. There is also some skill out to 3-year leads over the North Pacific. The skill from persistence apparently reflects the large decadal variability over the North Pacific and Atlantic Oceans. In addition, there is some predictability up to 2 or 3 years over some areas of the southern hemisphere between $30\text{--}60^\circ\text{S}$, possibly due to the low-frequency variation of the subtropical gyre over the southern hemisphere (Roemmich et al. 2007; Gille 2008).

The annual mean SST is predictable in the decadal forecasts only for about 1-year lead times over small areas of the subpolar Atlantic. The decadal forecasts of HC500 show more widespread areas where the ACC exceeds 0.35 for 1 or more years longer than in the persistence forecasts. For example, the areas of skill for HC500 extend over the lower-latitude Atlantic and over the Southern Hemisphere in the decadal forecasts. Two areas that stand out are the North Pacific, and mid-latitude eastern Atlantic where the

skill extends out to 5 years. This is in contrast with the subpolar Atlantic where the skill extends out to only about 2 years lead time both for SST and HC500, which is systematically shorter than for the persistence forecasts. This suggests that model deficiencies are generating errors in the prediction of decadal variability over the subpolar Atlantic.

Figure 6 shows the time series of HC500 over the North Pacific area ($170\text{--}150^\circ\text{W}$, $35\text{--}50^\circ\text{N}$), and the subpolar Atlantic ($40\text{--}10^\circ\text{W}$, $55\text{--}65^\circ\text{N}$). Over the North Pacific, it is likely that there is a dominant mode of decadal variability, which is consistent with previous studies investigating the PDO (Mantua et al. 1997; Mochizuki et al. 2010). Mochizuki et al. (2010), for example, showed that the PDO signal is predictable almost 6 years in advance at a 70 % confidence level in a prediction system using model for interdisciplinary research on climate (MIROC). Consistent with that study, the HC500 over the North Pacific is predictable to some extent in the GEOS-5 decadal forecasts. For example, the increase in HC500 from 1960 to 1975 is predicted out to about a 5-year lead time. Similarly, the evolution from 1990 to 1995, and 2005 to 2010 in the GMAO ORA is also reproduced well in the model hindcasts. However, during the 2000s the predictions are quite poor beyond a 4-year lead time, especially for the HC500 predictions during 2003 and 2004. This is potentially because the variability captured with Argo data in 2003 and 2004 was not represented in the analysis of earlier years

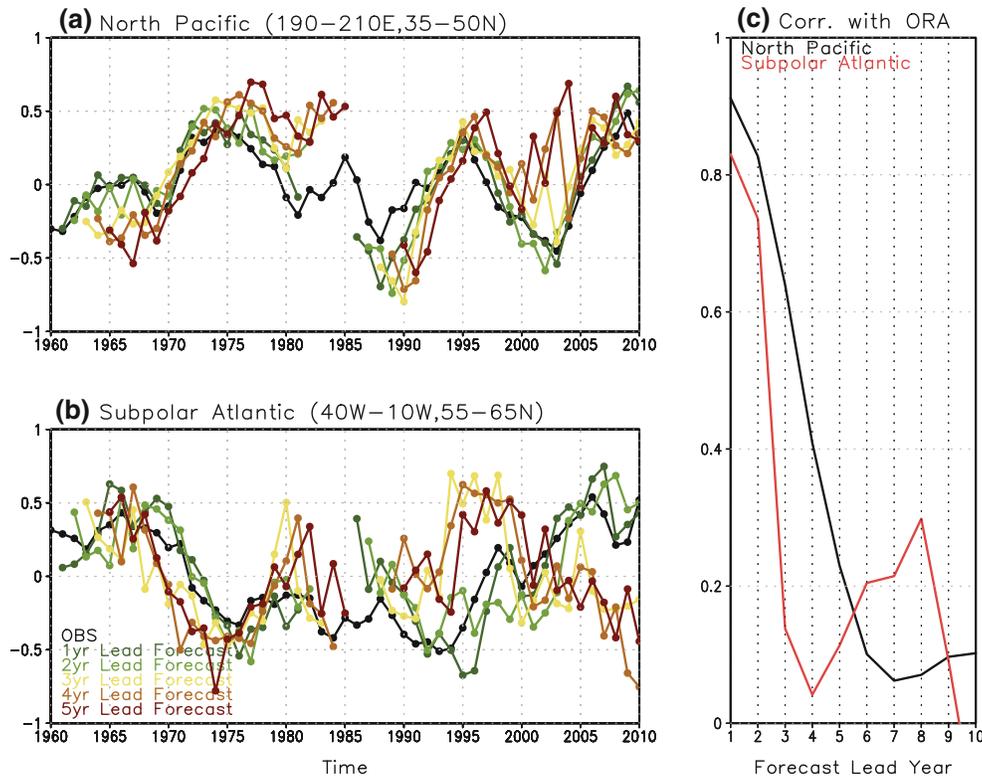


Fig. 6 The time series of de-trended anomalous annual mean HC500 ($^{\circ}\text{C}$) in the GMAO ORA (*black*) and decadal forecasts averaged over the **a** North Pacific ($170\text{--}150^{\circ}\text{W}$, $35\text{--}50^{\circ}\text{N}$), and **b** subpolar Atlantic

($40\text{--}10^{\circ}\text{W}$, $55\text{--}65^{\circ}\text{N}$). **c** The correlation between the GMAO ORA and the decadal forecasts over North Pacific (*black*) and subpolar Atlantic regions (*red*) as a function of the forecast lead year

(used for initialization) when Argo data were not available. The overall ACC of HC500 predictions is significant at the 90 % confidence level, with values remaining above 0.6 for up to 3-year lead times.

Over the subpolar Atlantic, the ORA shows a dominant decadal mode of variability (Smith et al. 2007; Keenlyside et al. 2008). The anomalous HC500 is positive during the 1960s and 2000s, while it is negative from the 1970s to the 1990s. The transition from a positive to a negative phase between 1960 and 1975 is relatively well simulated in the decadal hindcasts, while the warming trend in the mid-1990s and transition to the positive phase during the 2000s is successfully simulated only up to 2-year leads. The ACC between the reanalysis and the predictions is above 0.5 for lead times out to 2 years.

The discussion above was concerned with the overall quality of decadal predictions. In the next section we examine several of those results in more detail. In particular we examine why (1) the skill of the HC500 hindcasts is systematically higher over the mid- and north- Atlantic compared with the C20C simulation, and (2) the predictable HC500 signal extends to 5 years over the North Pacific and Atlantic, while over the subpolar Atlantic it is less skillful than a persistence forecast.

4 Simulation of Atlantic decadal variability

4.1 The Atlantic meridional circulation

Many studies have pointed out the role of the AMOC as a source of decadal predictability in the Atlantic (e.g., Delworth and Mann 2000; Knight et al. 2005; Dijkstra et al. 2006; van Oldenborgh et al. 2012). Therefore, prior to examining regional details in the prediction skill, we examine the predictability of the AMOC strength. Figure 7a shows the time series of the anomalous AMOC index (defined here as the maximum of the zonally-integrated annual mean overturning streamfunction averaged over $41\text{--}43^{\circ}\text{N}$) using the ORA and forecast output. Note that the trend has not been removed. The time series is a 3-year moving average to remove high-frequency variations. The definition used here is similar to that in Huang et al. (2012), although they selected 44°N for the latitude of their AMOC index since that was where the 1st empirical orthogonal function (EOF) pattern of their reanalysis stream function (derived from the Global Ocean Data Assimilation System from NOAA's National Centers for Environmental Prediction) reached its maximum value. Similarly, the latitude for the AMOC index used here was

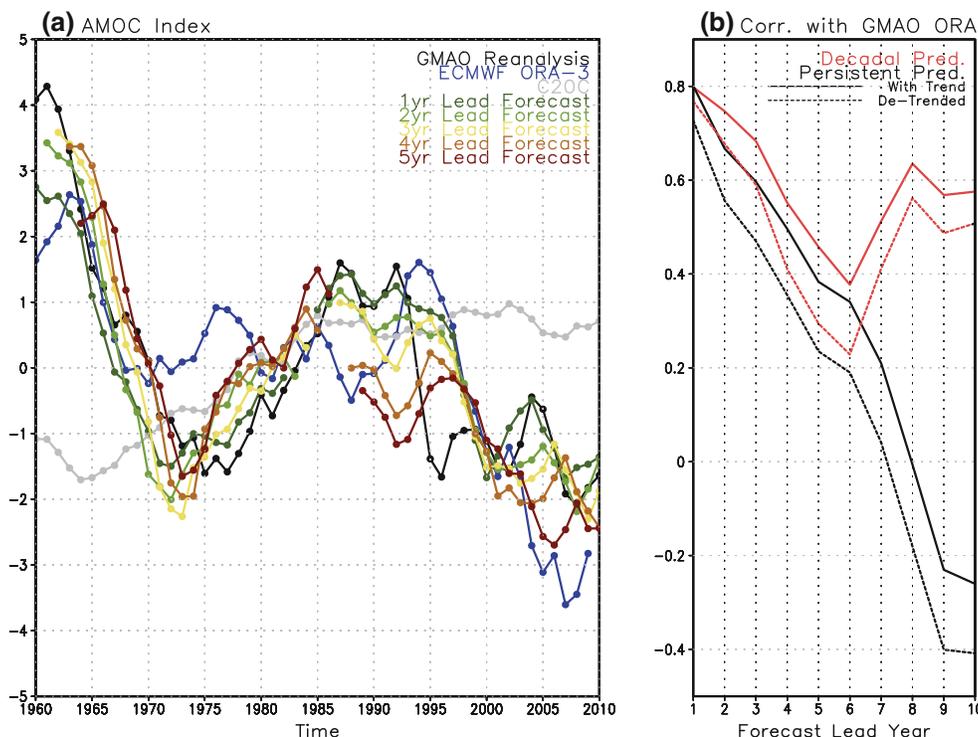


Fig. 7 **a** The time series of anomalous AMOC index (SV) in the GMAO ORA (black), ECMWF ORA-S3 (blue), the GEOS C20C simulation (gray), and decadal forecasts. Note that the AMOC index is defined as the maximum of the zonally-integrated annual mean overturning stream function averaged over 41–43°N. The time-series

shown is a 3-year moving average and is not de-trended. **b** The correlation of annual-mean AMOC index from the GMAO ORA and decadal forecasts (red) and persistence forecast (i.e. auto-correlation; black). The solid line denotes the correlation with trend, and dashed line denotes the correlation after removing the trend

selected as the location where the 1st EOF of stream function from the GMAO ORA reached its maximum value (not shown). However, note that the AMOC index is not very sensitive to the slight change in latitude choice between 40–45°N. Bingham et al. (2007) shows that AMOC variations farther south are dominated by interannual variations rather than decadal variations as in the subpolar branch.

In the GMAO ORA, there is a weak decreasing trend in the AMOC at this latitude. This decreasing trend is also seen in the System 3 ORA (ORA-S3) from the European Centre for Medium-Range Weather Forecasts (ECMWF) (blue line, Balmaseda et al. 2008), which exhibits very low transport in 2007. The decadal variations are generally consistent with those from 1985 to 2003 in Bingham et al. (2007) using an ocean circulation and climate advanced modeling project model (OCCAM) simulation. The OCCAM simulation shows AMOC variations north of 40°N that increase in strength from the late 1980s until the mid-1990s and then afterwards gradually decrease, continuing that trend for the rest of the simulation. Reichler et al. (2012) also shows that in a multi-reanalysis from 1979 to 2010 AMOC variations at 45°N peaked in the mid-1990s, then gradually decreased. Relative to the mean anomaly from the multi-reanalysis in the 1990s, the

GMAO ORA peaks too early and decays too early. However, Fig. 1b in Reichler et al. (2012) and Fig. 1 in Pohlmann et al. (2013) shows that there is a lot of spread in the ocean reanalyses (since the analyses are not always well constrained by observations, especially during periods of sparse observations) so the differences between ORAs seen in Fig. 7 are not uncommon.

In the forecasts, the overall variation, with a minimum in the mid-1970s and a decrease from the mid-1990s, is captured at all lead times up to 5 years; however, the details of shorter timescale variations change with lead time. Interestingly, the forecasts display a decline in transport in the late 1990's with a timing more consistent with ORA-S3 than with the GMAO ORA. The correlation of the forecast time series with the ORA transport is over 0.5, which is significant at about the 90 % confidence level, according to the one-tail *t* test as described earlier, for leads up to 4 years. The correlation skill of a persistence forecast shows slightly lower predictability. Similar results are found if the ECMWF ORA-S3 is used for validation. Interestingly, the correlation skill in the initialized decadal forecasts rebounds after 6 years. This implies that the predictability of the AMOC is systematically higher in the dynamical forecast system, and the AMOC variability is likely to be predictable up to almost a decade. Note that the

C20C simulation shows slight increases in the AMOC index with time and cannot simulate the decadal variations. The correlation skill of the initialized decadal predictions calculated after removing the trend is systematically lower than that with the trend, however, the behavior is basically the same.

4.2 The upper oceanic heat content (HC500)

The successful prediction of the AMOC transport does not guarantee a skillful forecast of the patterns of temperature change in the Atlantic. Figure 8 shows the correlation skill of the de-trended annual mean HC500 as a function of the forecast lead year. The warming trend has been removed in the same way as for Figs. 5 and 6 to focus on the low-frequency variability. Only correlations significant at the

90 % level are shown. Note that the degrees of freedom at each grid point are calculated using auto-correlation as done in previous figures. In the C20C simulation, the correlation skill is not significant anywhere over the Atlantic except for the Labrador Sea and southwest of the Iberian Peninsula. This correlation might be due to a nonlinear response to increasing GHG or merely to chance because of the small number of ensemble members.

In the 1-year lead initialized forecasts, the correlation skill is above 0.6 over most of the Atlantic, systematically higher than that in the C20C simulation. During the second year, the correlation skill drops over the Gulf Stream and in parts of the subpolar gyre. During the third year, the correlation skill remains above 0.6 over the eastern Atlantic, and the rim of the subpolar gyre. In addition, the correlation skill over the western Atlantic around 30°N is almost

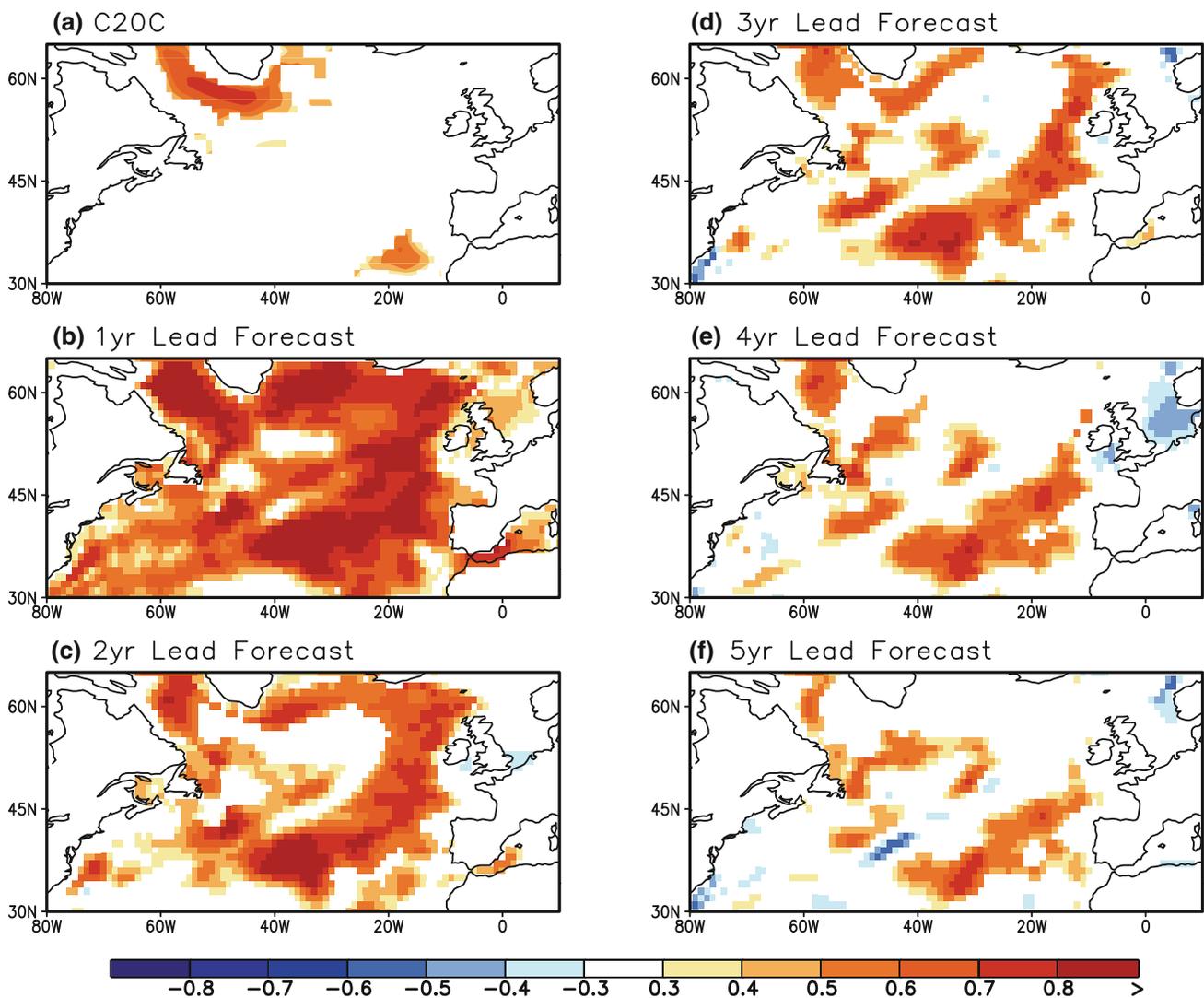


Fig. 8 The correlation skill of the de-trended annual mean HC500 anomaly in **a** the C20C simulation, **b** 1-year lead forecasts, **c** 2-year lead forecasts, **d** 3-year lead forecasts, **e** 4-year lead forecasts, and **f** 5-year lead forecasts

zero. During the fourth and fifth years, the correlations decay further although the areas of significant correlation remain unchanged except for the region southeast of Greenland, in the Irminger Current, where the correlation has dropped below the 90 % significance level.

Why does the correlation skill start to drop relatively faster over some areas of the Atlantic? Here we investigate how this decay is related to the realism of the simulated decadal modes over the Atlantic and show that the forecast skill degrades where the spatial pattern of the dominant decadal mode in the predictions departs from that in the ORA.

Figure 9 shows the first EOF and principal component (PC) of the de-trended HC500 from the ORA. Note that the EOF analysis performed is based on the Atlantic region (80°W–20°E, 20–65°N). The first EOF is comprised of a dipole pattern over the North Atlantic, with correlated variability of opposite signs in the subpolar gyre and the northern recirculation region of the Gulf Stream. It explains almost 28 % of the total de-trended variability. This dipole pattern is consistent with the dominant decadal mode in several observational and modeling studies (Hakkinen and Rhines 2004; Dong and Sutton 2005; Hatun et al. 2005; Zhang 2008), and the EOF analysis using ECMWF ORA-3, EN3 (Ingleby and Huddleston 2007, <http://www.metoffice.gov.uk/hadobs/en3/>), or grid point data from NOAA's National Oceanographic Data Center (NODC) (http://www.nodc.noaa.gov/OC5/3M_HEAT_CONTENT/) (plots not shown). From an analysis of HadCM3 (the climate prediction model of the UK Met Office) output, Dong and Sutton (2005) argue that temperature and salinity advection is a crucial factor for the maintenance and transition of this decadal mode. Zhang (2008) finds that shifts in the location of the Gulf Stream are tied to this decadal dipole mode. Consistent with their studies, the positive values in the first EOF of HC500 in Fig. 9 are located along the Gulf Stream between 30°N and 45°N, implying that oceanic advection by the Gulf Stream and North Atlantic Current play an important role in producing and maintaining this decadal mode in the GEOS-5 AOGCM.

The PC of the first EOF shows multi-decadal time-scales, with negative anomalies from 1960 to 1970, and in the 2000s, and positive anomalies between 1970 and 2000. This variation is also consistent with the results from ECMWF ORA-S3, EN3, and NODC output. (The correlation between any of these time series is at least 0.89.). The variations are closely aligned to those in the subpolar branch of the AMOC (grey line in Fig. 9b), though with less prominent peaks. The lagged correlation between the two time series is 0.55 at 8 years with HC500 leading. As found in the modeling study by Zhang (2008), the HC500 anomalies lead the AMOC anomalies and there is an out-of-phase relationship between the AMOC and the subpolar

gyre. Thus, strengthening of the AMOC is associated with the strengthening of the northern recirculation gyre (lower HC500 near the Gulf Stream path) and the weakening of the subpolar gyre (higher HC500 in the subpolar gyre). Consistent with the lag between HC500 and AMOC anomalies noted above, the patterns from the regression of the AMOC time series onto the time series HC500 and the corresponding time series for salinity and density 8 years earlier show that salinity variations are well correlated with density variations in the Labrador Sea and northern recirculation gyre while the temperature variations are well correlated in the subpolar gyre (not shown). Negative HC500 anomalies over the subpolar area along with positive salinity anomalies, both of which are positive density anomalies, are related to a stronger AMOC through enhanced deep water formation.

The ACC of the forecast PC is calculated with respect to the GMAO ORA in Table 1. The forecast PC is calculated by pattern regression of the ORA HC500 EOF onto the predicted HC500 anomaly. The ACC of the PC as a function of forecast lead time is above 0.6 for up to 6-year lead times, indicating that the dominant mode is quite predictable because of its decadal time-scale. However, as mentioned earlier, it does not guarantee a skillful forecast over the entire Atlantic, because the spatial pattern of the dominant simulated mode is different from the EOF obtained from the GMAO ORA.

Figure 10 shows the first EOF of HC500 in the ORA and the regression of its PC onto current vectors averaged from the surface to 500 m, from the ORA as well as from the forecast fields at leads of 1–5 years. In the ORA, there is an anticyclonic flow around the positive HC500, i.e., the negative density anomaly related to the positive HC500 anomaly produces an anticyclonic geostrophic current (Dong and Sutton, 2005). This anticyclonic flow produces an eastward current over the subpolar Atlantic, and southwestward current over the mid-eastern Atlantic. Note that the negative HC500 anomaly on the northern limb of the subpolar gyre is coincident with an acceleration of the subpolar gyre (Zhang 2008; Lohmann et al. 2009; Hakkinen et al. 2011; Robson et al. 2012b).

During the first year of the forecast, the overall pattern is quite similar to that in the GMAO ORA. However, the positive HC500 anomaly is generally thicker than the ORA, and is not as strong in the western Atlantic as in ORA. The negative HC500 anomaly is slightly weaker than that in the ORA. During the second year of the forecast, the positive HC500 anomaly extends farther to the north above 50°N (i.e. the red box), and the negative HC500 anomaly in the subpolar gyre is split into two separate regions. The magnitude of the anomalous current is weaker than that in the ORA. During the third year of the forecast, the positive HC500 anomaly extends far north along the North Atlantic

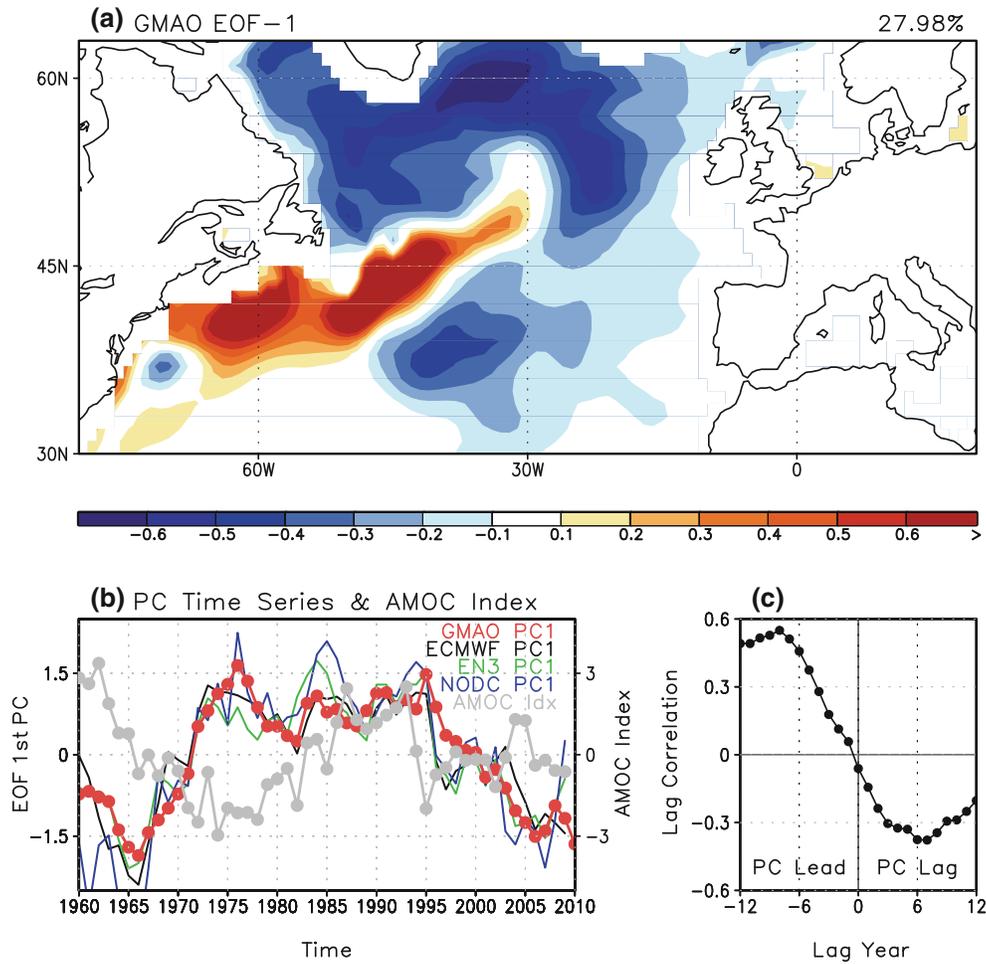


Fig. 9 **a** The first EOF of the GMAO ORA de-trended annual mean HC500 anomaly (°C). **b** The PC of the first EOF of HC500 in the GMAO ORA (red), ECMWF ORA-S3 (black), EN3 (green), and NODC output (blue), and the de-trended AMOC index (gray). **c** The lag-correlation between the PC and AMOC index (unitless quantity)

Table 1 The anomaly correlation between the PC of the 1st EOF from the GMAO ORA and the time series from the regression of the ORA EOF on the forecast, shown as a function of the forecast lead year

Forecast lead time	Decadal forecast
1 year	0.93
2 years	0.87
3 years	0.78
4 years	0.73
5 years	0.68
6 years	0.61
7 years	0.54
8 years	0.51
9 years	0.26
10 years	0.10

The bold values indicate that the correlation coefficient is over 90 % confidence level

and Irminger Currents, above 50°N. The anomalous geostrophic currents related to this positive HC500 anomaly are primarily meridional. Correspondingly, the negative HC500 anomaly in the Greenland Sea is quite weak, and the negative HC500 anomaly on the southern branch of the subpolar gyre is advected farther eastward. These anomalies gradually decay in later years.

These differences that emerge in the dominant decadal mode as the forecast evolves are consistent with the spatial patterns of correlation skill as a function of forecast lead time. For example, the forecast skill is relatively low over the subpolar central Atlantic in the 2-year-lead forecasts because the positive HC500 related to the dominant decadal mode is extended too far north. The ACC over the subpolar Atlantic drops quickly because the negative HC500 anomaly over the subpolar region becomes weak. Similarly, the confinement of the higher correlations to the rim of the basin in the forecasts at 5-year lead times is

and Irminger Currents, above 50°N. The anomalous geostrophic currents related to this positive HC500 anomaly are primarily meridional. Correspondingly, the negative HC500 anomaly in the Greenland Sea is quite weak, and the negative HC500 anomaly on the southern branch of the subpolar gyre is advected farther eastward. These anomalies gradually decay in later years.

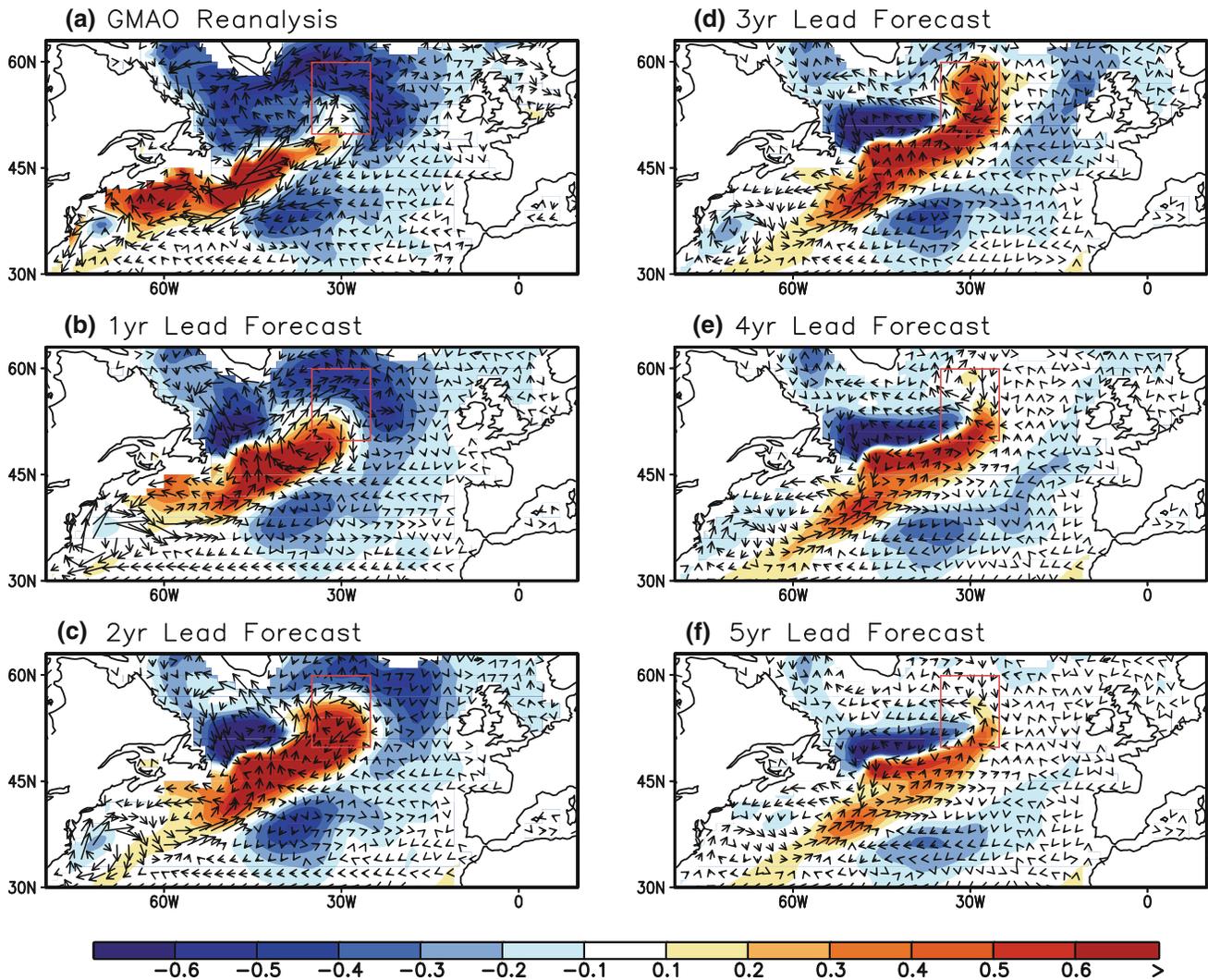


Fig. 10 The pattern of de-trended HC500 (°C) and currents (m/s) averaged from the surface to 500 m regressed onto the 1st PC from the GMAO ORA in **a** the GMAO ORA, **b** 1-year lead forecasts,

c 2-year lead forecasts, **d** 3-year lead forecasts, **e** 4-year lead forecasts, and **f** 5-year lead forecasts

consistent with the representation of dominant decadal mode at that lead.

Why is the spatial pattern of the simulated decadal mode different from that in reanalysis output? Motivated by previous studies that emphasized the role of the strength of mean current on the dominant decadal mode (e.g., Zhang 2008), we compare the simulated climatology of HC500 and currents averaged between the surface and 500 m to the ORA (Fig. 11). Although, the overall patterns of the analyzed climatological HC500 and currents are well simulated by the decadal forecasts, the systematic biases increase with forecast lead. One of biggest differences is the strength of North Atlantic Current (NAC), continuing the Gulf Stream northeast above 45°N, which is gradually stronger at longer forecast leads. After 3 years, the strength of the NAC in the forecasts is almost twice that in the ORA

and the zonal temperature gradient over the subpolar region near 30°W is stronger.

This is important because it can lead to stronger advection of positive HC500 anomalies along the NAC, and ultimately cause the excessive northward extension of positive anomalies in HC500 shown in Fig. 10c, d. The anomalous meridional current would not in turn generate HC500 anomalies since the current anomalies tend to be along the climatological temperature gradient in the later forecasts.

To summarize the results of Figs. 10 and 11, we analyze the advection terms in the HC500 budget over 35–25°W, 50–60°N (i.e., the red box in Figs. 10, 11), where the simulated decadal mode has an unrealistic northward extension of positive anomalies in HC500. Figure 12 shows the HC500 advection analysis related to the

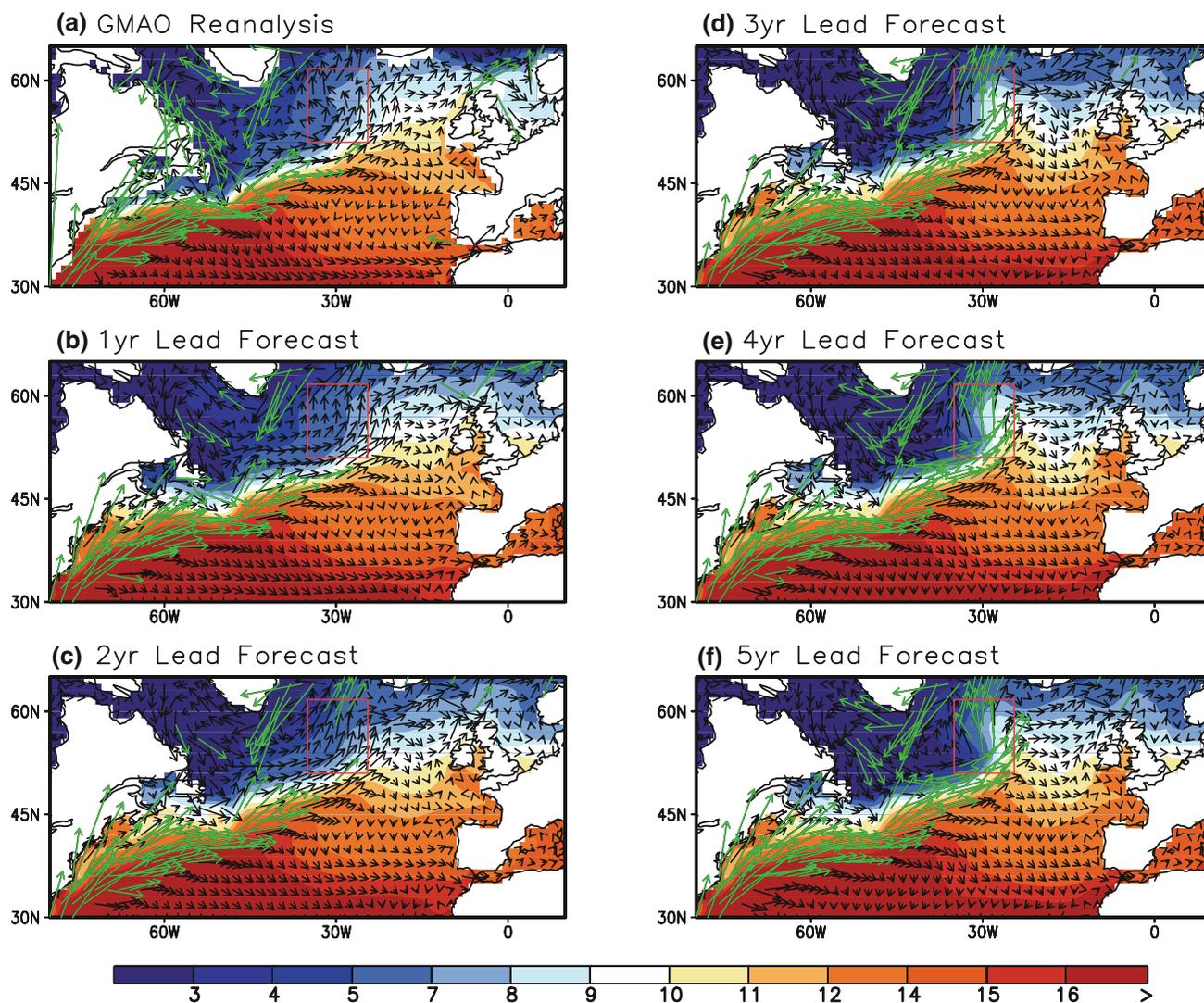


Fig. 11 As for Fig. 10, but climatological HC500 ($^{\circ}\text{C}$) and currents (m/s) averaged from the surface to 500 m. The *green* vectors denote currents with a speed above 0.04 m/s

dominant EOF in the GMAO ORA (denoted as 0 on the x-axis), and from the decadal forecasts as a function of forecast lead year (denoted as 1–5 on the x-axis). For this calculation, we regressed current anomalies onto the PC of the first EOF of HC500, then calculated the HC500 advection terms: $-u' \frac{\partial \bar{T}}{\partial x}$, $-v' \frac{\partial \bar{T}}{\partial y}$, $-\bar{u} \frac{\partial T'}{\partial x}$, $-\bar{v} \frac{\partial T'}{\partial y}$, where u' , v' , and T' are the zonal current, meridional current, and temperature anomalies related to the first EOF, and \bar{u} , \bar{v} , and \bar{T} denote the climatological values. The analysis is conducted as a function of forecast lead year. All fields were averaged from the surface to 500 m prior to regression. Then, the area-averaged value over $30^{\circ}\text{--}20^{\circ}\text{W}$, $50^{\circ}\text{--}60^{\circ}\text{N}$ was calculated. Note that the term $-v' \frac{\partial \bar{T}}{\partial y}$ is not shown because the regression was not significant at the 95 % confidence level.

In the ORA, the total HC500 advection (black) is negative, consistent with Fig. 10a. However, in first year of the

forecasts, the total HC500 advection is positive due to stronger advection by the climatological currents. The difference becomes robust during the second year of the forecasts, consistent with the increasing tendency of positive HC500 anomaly up to the third year of the forecasts (i.e. difference Fig. 10c from b, or d from c). In later forecast years, the total advection terms are negative, again consistent with the decreasing positive HC500 after the 3-year lead forecast in Fig. 10c, d, e.

In the ORA, the negative HC500 anomaly is mainly produced by the anomalous advection of mean HC500 (blue bar), i.e., the anomalous eastward current that advects cold water from the Labrador Sea. The contribution of this advection term is overwhelmed by the other terms in forecasts after the first year although the magnitude remains comparable to that in the ORA. In the first year of the forecast, the mean zonal advection of anomalous

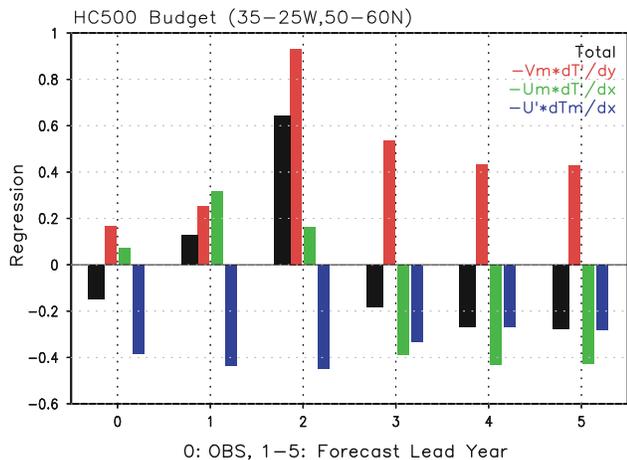


Fig. 12 The magnitude of HC500 advection terms ($^{\circ}\text{C}/\text{year}$) related to the first EOF averaged over $35\text{--}25^{\circ}\text{W}$, $50\text{--}60^{\circ}\text{N}$ (i.e. the red box in Figs 10, 11) from the ORA (0 in x -axis), and each forecast lead years (1–5 in x -axis) using the regression analysis. The total advection term (black bar), advection of anomalous HC500 due to mean meridional current ($-\bar{v} \frac{\partial T'}{\partial y}$, red bar), that due to mean zonal current ($-\bar{u} \frac{\partial T'}{\partial x}$, green bar), the advection of mean HC500 due to anomalous zonal current ($-u' \frac{\partial T}{\partial x}$, blue bar) is denoted. Note that the term $-v' \frac{\partial T}{\partial y}$ is not shown because it is not over the 95 % confidence level for the reanalysis and any of forecasts

HC500 (green bar), related to the eastward extension of the Gulf Stream, becomes significant. In the second year of the forecast, the contribution from the mean meridional current (red bar) peaks, caused by the strong NAC over the mid Atlantic above 45°N . Thus, the excessive northeast extension of positive HC500 anomalies related to the dominant EOF is due to the stronger climatological meridional currents in the decadal forecasts. The advection due to the mean current becomes negative from year 3 because the anomalous temperature gradient changes its sign from negative to positive as the center of positive HC500 shifts slightly to the east. Because the simulated NAC increases in strength during the first few years of the forecast, the advection of positive HC500 anomalies by the mean meridional current remains strong in later years even though the magnitude of the HC500 anomaly becomes slightly weaker later in the forecast due to the negative contribution from the other advection terms.

In summary, the dominant decadal mode simulated in the decadal forecasts is different from that in the ORA especially along the NAC, and this leads to a relatively rapid drop in prediction skill over this region. While the dominant EOF in the ORA exhibits a negative HC500 anomaly over the subpolar regions related to anomalous eastward advection of HC500, in the second year of the forecast there is a positive HC500 anomaly in the same region that is tied to a too strong climatological NAC. Thus, the mechanism that leads to negative HC500

anomalies in the GMAO ORA is not operating in the decadal forecasts. This implies that the realistic simulation of the climatological circulation is crucial to skillful decadal forecasts.

The impact of the stronger NAC on other EOFs in the decadal forecasts is examined in Fig. 13, following the same procedure as used for the first EOF. Clearly the same degradation in the patterns is seen as a function of forecast lead. This implies that climatological biases in this region, particularly in the North Atlantic Current, is the one of main reasons for the loss of prediction skill over the North Atlantic.

The dominant decadal mode is strongly coupled with the North Atlantic Oscillation (NAO)-like atmospheric variations in observations (Lohmann et al. 2009). Unfortunately, the NAO is not successfully simulated in the decadal forecast (not shown). Since the NAO is important for the subpolar North Atlantic ocean (Hakkinen et al. 2011; Yeager and Danabasoglu 2012), the weak NAO-like representation in the decadal forecast system might be another reason of the abrupt degradation of forecast skill over the subpolar regions.

5 Summary and conclusion

In this study, the forecast skill of decadal hindcasts from 1960 to 2010 using the GMAO’s GEOS-5 AOGCM has been evaluated. The mean of decadal forecasts using three ensemble members shows a warm bias between $50^{\circ}\text{S}\text{--}50^{\circ}\text{N}$ and a cold bias poleward of those latitudes. The warm bias increases with longer forecast lead times, consistent with the fact that the global warming trend in the AOGCM is stronger than observed. The forecast skill of 4-year-averaged SST shows that there are improvements in MSSS up to 10-year lead forecasts over the Atlantic, Indian Ocean, and the tropical western Pacific, indicating the benefits of initializing the decadal forecasts. The improvement in the initialized decadal forecast is also due to the reduction in the global warming signal in the initialized forecast relative to the C20C run. In particular, the forecast skill of 4-year-averaged SST is systematically higher over the North Atlantic with the aid of initialization. The increase of skill, measured by both anomaly correlation and MSSS, in the decadal forecast is maintained for almost a decade over the subtropical and mid-latitude Atlantic.

Even though the skill in upper ocean heat content is not as widespread as for SST, the MSSS shows there is about 50 % improvement over the subtropics and mid-latitude Atlantic up to lead times of 6–9 years with initialization by successful simulation of Atlantic decadal variability. On the other hand, the MSSS value is negative over the North Pacific and southern Atlantic, implying that the

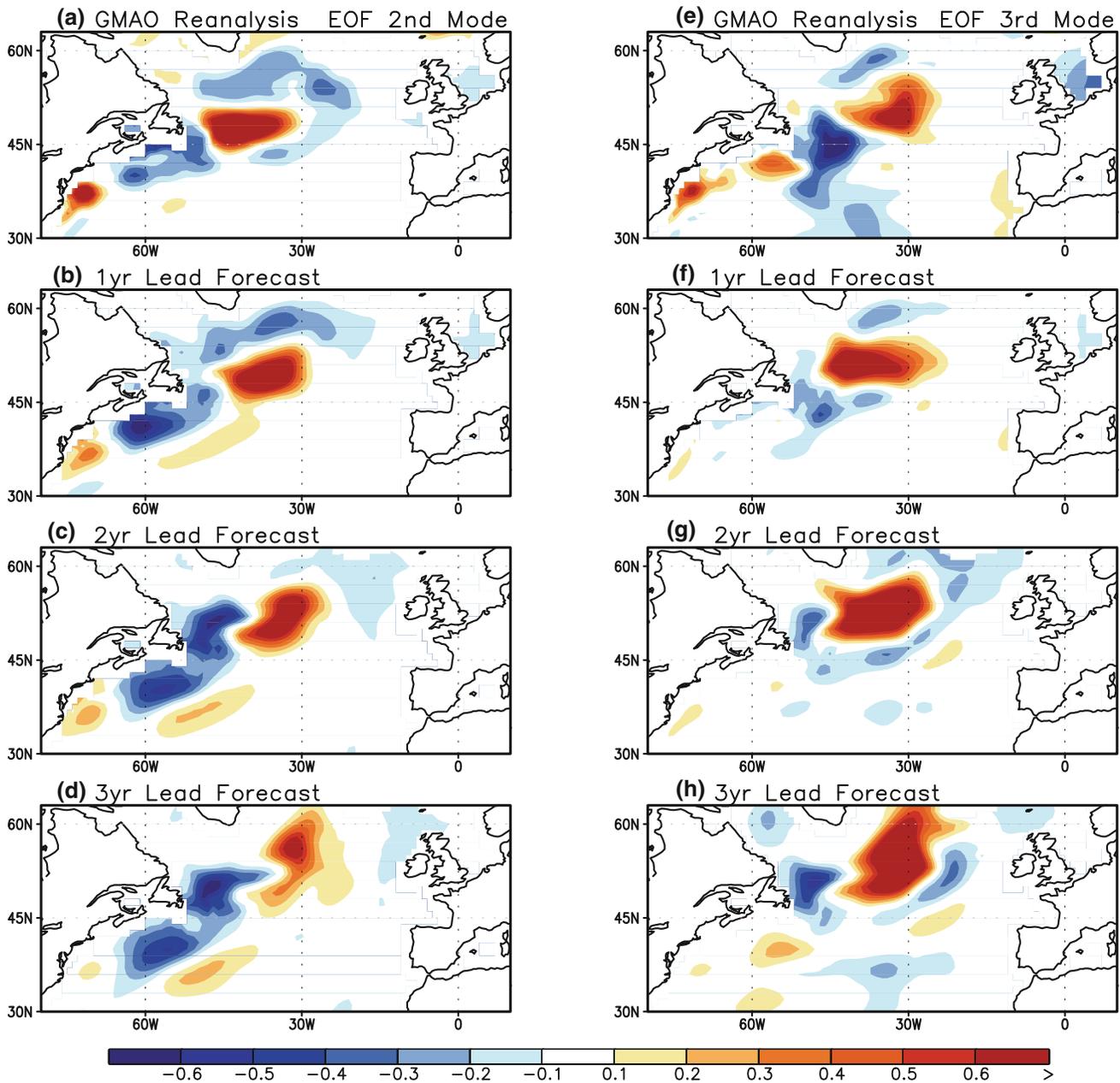


Fig. 13 The de-trended annual mean HC500 anomaly regressed onto the second (*left panel*), and third (*right panel*) EOF in the ORA and the forecasts at 1- to 3-year lead times

initialization reduces the prediction skill over some regions outside the Atlantic. The annual-mean AMOC index for the subpolar branch of the AMOC is predictable for up to 4-year leads, consistent with the predictable signal in subsurface temperature over the mid-latitude and North Atlantic. However, the predictable signal in the HC500 over the Atlantic is relatively lower in the subpolar gyre.

The predictability of the decadal signal was investigated by detrending the decadal forecasts prior to calculating the decadal anomalies. The system has lower prediction skill for detrended decadal heat content anomalies in the

subpolar North Atlantic than for other ocean regions and the predictability time scale (or e-folding time scale) is lower in this region than for persistence. This low skill is related to the fact that the spatial pattern of the dominant decadal mode in HC500 over this region in the ORA is quite different from that in the forecasts beyond the first year. An analysis of the advection terms as a function of forecast lead shows that biases in the climatological NAC, which is too strong in the decadal forecasts, are responsible for the excessive northward extension of mid-latitude anomalies to the subpolar Atlantic. This implies that the

realistic simulation of the climatological circulation is crucial for skillful decadal forecasts.

This study reveals the relationship between the simulated climatology and decadal modes, and shows how this relationship directly impacts the decadal prediction skill of upper ocean heat content in the North Atlantic in the GEOS-5 AOGCM. This study gives some potential insights into the performance of anomaly initialization (Schneider et al. 1999; Smith et al. 2007; Keenlyside et al. 2008; Pohlmann et al. 2009; Smith et al. 2013). Anomaly initialization has the advantage of reducing the initialization shock, however, it has drawbacks by using the climatology in the model right from the outset. Hence, depending on the model biases, it might be expected that anomaly initialization might lead to lower skill at short forecast leads than full initialization. However, Smith et al. (2013) find that the differences in skill between full-field and anomaly initialization are generally not significant. The best method for initializing decadal forecasts is still an active area of research as is the identification of predictable decadal signals.

Acknowledgments This study was supported by NASA's Modeling, Analysis and Prediction program. Computer time was provided by the NASA Center for Climate Simulation at NASA Goddard Space Flight Center (GSFC). Support from Tony Rosati and colleagues from NOAA's Geophysical Fluid Dynamics Laboratory in the configuration of MOM4 are gratefully acknowledged, as is that from Elizabeth Hunke at Los Alamos National Laboratory in the use of CICE. Arlindo da Silva and Peter Colarco at GSFC configured the aerosol component of the AOGCM. We are grateful for the comments from two anonymous reviewers that helped improve the manuscript.

References

- Antonov JI et al (2010) World Ocean Atlas 2009 vol. 2: Salinity. In: S. Levitus (ed) NOAA Atlas NESDIS 69 US. Government Printing Office, Washington, D.C. 184 pp
- Balmaseda MA, Vidard A, Anderson DLT (2008) The ECMWF ocean analysis system: ORA-S3. *Mon Weather Rev* 136:3018–3034. doi:10.1175/2008MWR2433.1
- Bingham RJ, Hughes CW, Roussenov V, Williams RG (2007) Meridional coherence of the North Atlantic meridional overturning circulation. *Geophys Res Lett* 34:L23606. doi:10.1029/2007GL031731
- Chikamoto Y, Kimoto M, Ishii M, Mochizuki T, Sakamoto TT, Tatebe H, Komuro Y, Watanabe M, Nozawa T, Shiogama H, Mori M, Yasunaka S, Imada Y (2012) An overview of decadal climate predictability in a multi-model ensemble by climate model MIROC. *Clim Dyn*. doi:10.1007/s00382-012-1351-y
- Delworth TL, Mann ME (2000) Observed and simulated multi decadal variability in the Northern Hemisphere. *Clim Dyn* 16:661–676
- Dijkstra HA, te Raa L, Schmeits M, Gerrits J (2006) On the physics of the Atlantic multi decadal oscillation. *Ocean Dyn* 56:36–50
- Doblas-Reyes FJ, Balmaseda MA, Weisheimer A, Palmer TN (2011) Decadal climate prediction with the European centre for medium-range weather forecasts coupled forecast system: impact of ocean observations. *J Geophys Res* 116:D19111. doi:10.1029/2010JD015394
- Dong B, Sutton RT (2005) Mechanism of interdecadal thermohaline circulation variability in a coupled ocean–atmosphere GCM. *J Clim* 18:1117–1135. doi:10.1175/JCLI3328.1
- Enfield DB, Mestas-Núñez AM, Trimble PJ (2001) The Atlantic multi decadal oscillation and its relation to rainfall and river flows in the continental US. *Geophys Res Lett* 28:2077–2080
- Gille ST (2008) Decadal-scale temperature trends in the Southern hemisphere ocean. *J Clim* 21:4749–4765. doi:10.1175/2008JCLI2131.1
- Goddard L et al (2013) A verification framework for interannual-to-decadal predictions experiments. *Clim Dyn* 40:245–272. doi:10.1007/s00382-012-1481-2
- Griffies SM, Bryan K (1997) Predictability of North Atlantic multidecadal climate variability. *Science* 275:181–184
- Griffies SM et al (2005) Formulation of an ocean model for global climate simulations. *Ocean Science* 1:45–79
- Hakkinen S, Rhines PB (2004) Decline of subpolar North Atlantic circulation during the 1990s. *Science* 304:555–559
- Hakkinen S, Rhines PB, Worthen DL (2011) Warm and saline events embedded in the meridional circulation of the northern North Atlantic. *J Geophys Res* 116:C03006. doi:10.1029/2010JC006275
- Ham YG, Kug JS, Kang IS (2012a) Coupled bred vectors in the tropical Pacific and their application to ENSO prediction. *Prog Oceanogr*. doi:10.1016/j.pocan.2012.04.005
- Ham YG, Rienecker MM, Schubert S, Marshak J, Yeh SW, Yang SC (2012b) The decadal modulation of coupled bred vectors. *Geophys Res Lett* 39:L20712. doi:10.1029/2012GL053719
- Hatun H, Sandø AB, Drange H, Hansen B, Valdimarsson H (2005) Influence of the Atlantic subpolar gyre on the thermohaline circulation. *Science* 309:1841–1844
- Hawkins ED, Sutton R (2011) The potential to narrow uncertainty in projections of regional precipitation change. *Clim Dyn* 37:407–418
- Huang B, Xue Y, Kumar A, Behringer DW (2012) AMOC variations in 1979–2008 simulated by NCEP operational ocean data assimilation system. *Clim Dyn* 38:513–525. doi:10.1007/s00382-011-1035-z
- Hunke EC, Lipscomb WH (2008) CICE: the Los Alamos Sea ice model, documentation and software manual, version 4.0. Technical report, Los Alamos National Laboratory
- Hurrell JW, Hack JJ, Shea D, Caron JM, Rosinski J (2008) A new sea surface temperature and sea ice boundary dataset for the community atmosphere model. *J Clim* 21:5145–5153. doi:10.1175/2008JCLI2292.1
- ICPO (International CLIVAR Project Office) (2011) Data and bias correction for decadal climate predictions. International CLIVAR Project Office, CLIVAR publication series no. 150, p 6
- Ingleby B, Huddleston M (2007) Quality control of ocean temperature and salinity profiles—historical and real-time data. *J Mar Syst* 65:158–175. doi:10.1016/j.jmarsys.2005.11.019
- Ishii M, Kimoto M (2009) Reevaluation of historical ocean heat content variations with an XBT depth bias correction. *J Oceanogr* 65:287–299
- Keenlyside NS, Latif M, Jungclaus J, Kornbluh L, Roeckner E (2008) Advancing decadal-scale climate prediction in the north Atlantic sector. *Nature* 453:84–88. doi:10.1038/nature06921
- Kim HM, Webster PJ, Curry JA (2012) Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts. *Geophys Res Lett* 39:L10701. doi:1029/2012GL051644
- Knight JT, Allan RJ, Folland CK, Vellinga M, Mann ME (2005) A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophys Res Lett* 32:L20708. doi:1029/2005GL024233
- Lee J-Y et al (2010) How are seasonal prediction skills related to models' performance on mean state and annual cycle? *Clim Dyn* 35:267–283. doi:10.1007/s00382-010-0857-4

- Levitus S, Antonov JJ, Boyer TP, Locarnini RA, Garcia HE, Mishonov AV (2009) Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems. *Geophys Res Lett* 36:L07608. doi:[10.1029/2008GL037155](https://doi.org/10.1029/2008GL037155)
- Locarnini RA, Mishonov AV, Antonov JJ, Boyer TP, Garcia HE, Baranova OK, Zweng MM, Johnson DR (2010) *World Ocean Atlas 2009, Volume 1: Temperature*. S. Levitus, Ed. NOAA Atlas NESDIS 68, U.S. Government Printing Office, Washington, D.C., p 184
- Lohmann K, Drange H, Bentsen M (2009) Response of the North Atlantic subpolar gyre to persistent North Atlantic Oscillation like forcing. *Clim Dyn* 32:273–285
- Mantua NJ, Hare SR, Zhang Y, Wallace JM, Francis RC (1997) A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull Am Meteorol Soc* 78:1069–1079
- Matei D, Baehr J, Jungclaus J, Haak H, Muller WA, Marotzke J (2012a) Multiyear prediction of monthly mean atlantic meridional overturning circulation at 26.5°N. *Science* 335:76–79
- Matei D, Pohlmann H, Jungclaus J, Wolfgang M, Haak H, Marotzke J (2012b) Two tales of initializing decadal climate prediction experiments with the ECHAM5/MPI-OM model. *J Clim*. doi:[10.1175/JCLI-D-11-00633.1](https://doi.org/10.1175/JCLI-D-11-00633.1)
- McPhaden M et al. (2010) The global tropical moored buoy array. In: Hall J, Harrison DE, Stammer D (eds) *Proceedings of OceanObs'09: sustained ocean observations and information for society* (vol. 2), Venice, Italy, 21–25 September 2009, ESA Publication WPP-306. doi:[10.5270/OceanObs09.cwp.61](https://doi.org/10.5270/OceanObs09.cwp.61)
- Molod A et al (2012) The GEOS-5 atmospheric general circulation model: mean climate and development from MERRA to Fortuna., NASA technical report series on global modeling and data assimilation, V28
- Meehl GA, Washington WM, Collins WD, Arblaster JM, Hu AX, Buja LE, Strand WG, Teng HY (2005) How much more global warming and sea level rise? *Science* 307:1769–1772
- Meehl GA et al (2007) Global climate projections. *Climate change 2007*. In: Solomon S et al (eds) *The physical science basis*. Cambridge University Press, Cambridge, pp 747–845
- Meehl GA et al (2009) Decadal prediction: can it be skillful? *Bull Am Meteorol Soc* 90:1467–1485. doi:[10.1175/2009BAMS2607.1](https://doi.org/10.1175/2009BAMS2607.1)
- Mochizuki T et al (2010) Pacific decadal oscillation hind casts relevant to near-term climate prediction. *Proc Natl Acad Sci USA* 107:1833–1837
- Murphy JM (1988) The impact of ensemble forecasts on predictability. *Q J R Meteorol Soc* 114:463–493
- Pohlmann H, Jungclaus JH, Kohl A, Stammer D, Marotzke J (2009) Initializing decadal climate predictions with the GECCO oceanic synthesis: effects on the North Atlantic. *J Clim* 22(14): 3926–3938. doi:[10.1175/2009JCLI2535.1](https://doi.org/10.1175/2009JCLI2535.1)
- Pohlmann H, Smith DM, Balaseda MA, Kennlyside NS, Masina S, Matei D, Muller WA, Rogel P (2013) Predictability of the mid-latitude Atlantic meridional overturning circulation in a multi-model system. *Clim Dyn*. doi:[10.1007/s00382-013-1663-6](https://doi.org/10.1007/s00382-013-1663-6)
- Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J Geophys Res* 108:4407. doi:[10.1029/2002JD002670](https://doi.org/10.1029/2002JD002670)
- Read JF, Gould WJ (1992) Cooling and freshening of the subpolar North Atlantic ocean since the 1960s. *Nature* 360:55–57
- Reichler T, Kim J, Manzini E, Kroger J (2012) A stratospheric connection to Atlantic climate variability. *Nat Geosci*. doi:[10.1038/ngeo1586](https://doi.org/10.1038/ngeo1586)
- Reynolds RW, Smith TM, Liu C, Chelton DB, Casey KS, Schlax MG (2007) Daily high-resolution blended analyses for sea surface temperature. *J Clim* 20:5473–5496
- Rienecker MM et al. (2008) The GEOS-5 Data assimilation system—documentation of versions 5.0.1, 5.1.0, and 5.2.0, NASA Technical report series on global modeling and data assimilation, V27
- Rienecker MM et al (2011) MERRA—NASA’s modern-era retrospective analysis for research and applications. *J Clim* 24:3624–3648. doi:[10.1175/JCLI-D-11-00015.1](https://doi.org/10.1175/JCLI-D-11-00015.1)
- Robson JJ (2010) Understanding the performance of a decadal prediction system. Ph.D. thesis, University of Reading, UK
- Robson J, Sutton R, Smith D (2012a) Initialized decadal predictions of the rapid warming of the North Atlantic Ocean in the mid 1990s, 39: L19713. doi:[10.1029/2012GL053370](https://doi.org/10.1029/2012GL053370)
- Robson J, Sutton R, Lohmann K, Smith D (2012b) Causes of the rapid warming of the North Atlantic Ocean in the mid-1990s. *J Clim* 25:4116–4134
- Roemmich D, Gilson J, Davis R, Sutton P, Wijffels S, Piser S (2007) Decadal spinup of the South Pacific subtropical gyre. *J Clim* 20:162–173
- Schneider EK, Huang B, Zhu Z, Dewitt DG, Kinter JL, Kirtman BP, Shukla J (1999) Ocean data assimilation, initialization, and predictions of ENSO with a coupled GCM. *Mon Weather Rev* 127:1187–1207
- Smith DM, Cusack S, Colman AW, Folland CK, Harris GR, Murphy JM (2007) Improved surface temperature prediction for the coming decade from a global climate model. *Science* 317:796–799. doi:[10.1126/science.1139540](https://doi.org/10.1126/science.1139540)
- Smith DM, Eade R, Pohlmann H (2013) A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Clim Dyn*. doi:[10.1007/s00382-013-1683-2](https://doi.org/10.1007/s00382-013-1683-2)
- Smith DM, Eade R, Dunstone NJ, Fereday D, Murphy JM, Pohlmann H, Scaife AA (2010) Skilful multi-year predictions of Atlantic hurricane frequency. *Nat Geosci* 3(12):846–849
- Toth Z, Kalnay E (1993) Ensemble forecasting and NMC: the generation of perturbations. *Bull Am Meteorol Soc* 74:2317–2330
- Toth Z, Kalnay E (1997) Ensemble forecasting at NCEP and the breeding method. *Mon Weather Rev* 125:3297–3319
- van Oldenborgh GJ, Doblas-Reyes FJ, Wouters B, Hazeleger W (2012) Decadal prediction skill in a multi-model ensemble. *Clim Dyn* 38:1263–1280
- Vernieres G, Keppenne C, Rienecker MM, Jacob J, Kovach R (2012) The GEOS-ODAS, description and evaluation. NASA technical report series on global modeling and data assimilation, NASA/TM-2012-104606, Vol. 30
- Vikhliav Y, Kirtman B, Schopf P (2007) Decadal North Pacific bred vectors in a coupled GCM. *J Clim* 20:5744–5764
- Wetherald RT, Stouffer RJ, Dixon KW (2001) Committed warming and its implications for climate change. *Geophys Res Lett* 28:1535–1538
- Yang S-C, Kalnay E, Cai M, Rienecker MM (2008) Bred vectors and tropical pacific forecast errors in the NASA coupled general circulation model. *Mon Weather Rev* 136:1305–1326
- Yeager S, Danabasoglu G (2012) Sensitivity of Atlantic meridional overturning circulation variability to parameterized Nordic Sea overflows in CCSM4. *J Clim* 25:2077–2103. doi:[10.1175/JCLI-D-11-00149.1](https://doi.org/10.1175/JCLI-D-11-00149.1)
- Yeager S, Karspeck A, Danabasoglu G, Tribbia J, Teng H (2012) A decadal prediction case study: late 20th century North Atlantic ocean heat content. *J Clim* 25:5173–5189
- Zhang R (2008) Coherent surface-subsurface fingerprint of the Atlantic meridional overturning circulation. *Geophys Res Lett* 35:L20705. doi:[10.1029/2008GL035463](https://doi.org/10.1029/2008GL035463)
- Zhang R, Delworth TL (2005) Simulated tropical response to a substantial weakening of the Atlantic thermohaline circulation. *J Clim* 18:1853–1860. doi:[10.1175/JCLI3460.1](https://doi.org/10.1175/JCLI3460.1)
- Zhang R, Delworth TL, Held IM (2007) Can the Atlantic Ocean drive the observed multidecadal variability in Northern Hemisphere mean temperature? *Geophys Res Lett* 34:L02709. doi:[10.1029/2006GL028683](https://doi.org/10.1029/2006GL028683)