Estimating Decadal Predictability for the Southern Ocean Using the GFDL CM2.1 Model

Liping Zhang,[a,b] Thomas L. Delworth,[b] Xiaosong Yang,[b,c] Richard G. Gudgel,[b] Liwei Jia,[a,b] Gabriel A. Vecchi,[b] and Fanrong Zeng[b]

[a] *Atmospheric and Oceanic Science, Princeton University, Princeton, New Jersey*
[b] *NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey*
[c] *University Corporation for Atmospheric Research, Boulder, Colorado*

ABSTRACT

This study explores the potential predictability of the Southern Ocean (SO) climate on decadal time scales as represented in the GFDL CM2.1 model using prognostic methods. Perfect model predictability experiments are conducted starting from 10 different initial states, showing potentially predictable variations of Antarctic bottom water (AABW) formation rates on time scales as long as 20 years. The associated Weddell Sea (WS) subsurface temperatures and Antarctic sea ice have potential predictability comparable to that of the AABW cell. The predictability of sea surface temperature (SST) variations over the WS and the SO is somewhat smaller, with predictable scales out to a decade. This reduced predictability is likely associated with stronger damping from air–sea interaction. As a complement to this perfect predictability study, the authors also make hindcasts of SO decadal variability using the GFDL CM2.1 decadal prediction system. Significant predictive skill for SO SST on multiyear time scales is found in the hindcast system. The success of the hindcasts, especially in reproducing observed surface cooling trends, is largely due to initializing the state of the AABW cell. A weak state of the AABW cell leads to cooler surface conditions and more extensive sea ice. Although there are considerable uncertainties regarding the observational data used to initialize the hindcasts, the consistency between the perfect model experiments and the decadal hindcasts at least gives some indication as to where and to what extent skillful decadal SO forecasts might be possible.

1. Introduction

Climate variability can be generated by both internal interactions and external forcing (Latif et al. 2013). The former refers to interactions within or between the individual climate system components that include atmosphere, ocean, land, and sea ice, while the latter points to responses to changes in anthropogenic greenhouse gas/aerosol concentration as well as variations in solar irradiance and volcanic eruptions. The El Niño–Southern Oscillation (ENSO; Philander 1990), the Pacific decadal oscillation (PDO; Mantua et al. 1997; Kwon and Deser 2007; Zhang and Delworth 2015, 2016a), the Atlantic multidecadal oscillation (AMO; Knight et al. 2005; Delworth and Mann 2000), and the Southern Ocean (SO) decadal to centennial variability (Martin et al. 2013; Latif et al. 2013; O'Kane et al. 2013; Le Bars et al. 2016) are typical examples of internal variability.

Internal climate fluctuations can sometimes strongly project onto global or regional climate change, thereby masking the effects of external forcing. These internal variabilities are also quite different in different climate models in terms of time scale and spatial structure, especially over the SO (Monselesan et al. 2015).

Over the most recent two decades, the SO SST has shown a broad cooling (e.g., Purkey and Johnson 2010, 2012; Zhang et al. 2017), along with an expansion of Antarctic total sea ice area (e.g., Comiso and Nishio 2008; Cavalieri and Parkinson 2008) despite highly regional heterogeneity (e.g., Holland and Kwok 2012; Matear et al. 2015). However, climate models forced with historical changes in radiative forcing typically do not reproduce the observed cooling around the Antarctic. Instead, they show a slow but steady SO warming and total Antarctic sea ice area loss (Purich et al. 2016). Potential explanations for the discrepancy between observations and model projections are that the Antarctic ice sheet meltwater is absent in climate models

(Bintanja et al. 2013, 2015) or the natural internal variability plays a large role (Zunz et al. 2013; Polvani and Smith 2013). It is therefore of great importance to understand the detailed dynamics and predictability of internal variability over the SO.

In the instrumental period, the SO SST exhibits a pronounced multidecadal internal variability (Latif et al. 2013; Monselesan et al. 2015), albeit with some uncertainties due to short observation records. The most recent warm phase of this multidecadal variability coincides with a prominent Weddell polynya event (1974–76) that displayed a large area of open water within the ice-covered Weddell Sea and was accompanied with strong deep convection (e.g., Gordon 1978, 1982). This potential linkage between the SO multidecadal variability and Weddell Sea deep convection in observations also appears in fully coupled climate models (e.g., Martin et al. 2013; Galbraith et al. 2011; Zhang and Delworth 2016a,b). Martin et al. (2013) suggested that the SO internal multidecadal variability in the Kiel Climate Model (KCM) is mainly driven by the Weddell Sea deep convection; in addition, they attributed the multidecadal time scale to the slow accumulation of heat advected into the Weddell Sea at middepth by the lower limb of the Atlantic meridional overturning circulation (AMOC). Similar variability, including characteristics and mechanisms, is found in the Geophysical Fluid Dynamics Laboratory Climate model, version 2.1 (GFDL CM2.1).

Considering the unforced or internal predictability of the system, the pronounced low-frequency multidecadal variability over the SO mentioned above corresponds to a large value of potential predictability variance fraction (ppvf; Boer 2004, 2011). However, SO predictability/prediction has received less attention compared to the North Atlantic and North Pacific Oceans, partly due to the lack of observations, especially in the subsurface ocean. Given this reason, we used a diagnostic approach [averaged predictability time (APT); DelSole and Tippett 2009] to identify the most predictable components of decadal SST variations over the SO in a 4000-yr control simulation of the GFDL CM2.1 model. We found that the most predictable component is closely related to the mature phase of a mode of internal variability in the SO that is associated with fluctuations of deep ocean convection, which has a significant predictive skill as long as 20 years. The second most predictable component has a significant predictive skill up to 6 years and is associated with a transition between phases of the dominant pattern of SO internal variability. The current paper is a follow-up study of our previous work, which continues to examine the SO climate predictability but changes from a diagnostic

perspective to a prognostic perspective. In the prognostic approach, the SO predictability is estimated with so-called perfect model ensemble experiments. The fully coupled atmosphere–ocean general circulation model (AOGCM) is initialized by identical oceanic and perturbed atmospheric conditions. We further extend this method to decadal hindcasts/forecasts that are initialized with reanalysis data/observations, although past observations of the SO are sparse. The prediction skill is assessed by analyzing how well the time evolving variables produced by the initialized model match the observations.

2. Model, experiments, and methods

a. "Perfect model" predictability experiments

The model we use is the Geophysical Fluid Dynamics Laboratory Coupled Model version 2.1 (Delworth et al. 2006). The atmospheric component, AM2.1, has a horizontal resolution of $2° \times 2°$ with 24 levels in the vertical. The ocean and ice models, MOM4, have a horizontal resolution of $1°$ in the extratropics, with meridional grid linearly decreasing to $\frac{1}{3}°$ near the equator. The ocean model has 50 levels in the vertical, with 22 evenly spaced levels over the top 220 m. A 2000-yr control simulation is conducted with atmospheric constituents and external forcing held constant at 1860 conditions (this is conducted on a recently installed supercomputer and used in this study; we also have a 4000-yr simulation conducted on an older machine).

From the last quasi-equilibrium 500 years of the control run, 10 different years are randomly selected and used as initial conditions to perform the so-called perfect model ensemble experiments. Each ensemble consists of 10 members with perturbed atmospheric states but with the same oceanic initial conditions. We generate the perturbed atmospheric states by taking the atmosphere from different years in the control run excluding the current year. Note that these oceanic and atmospheric initial conditions are purely from the model. Each experiment integrates forward for 30 years. These ensemble experiments are referred to as predictability experiments. The spread within the ensemble is interpreted as an estimate of predictability.

The normalized ensemble variance (NEV) (Griffies and Bryan 1997) and prognostic potential predictability (PPP) (Pohlmann et al. 2004) are used to quantitatively estimate the potential predictability of climate variable X over the SO. The NEV as a function of the prediction period t is defined as

$$\text{NEV}(t) = \frac{\frac{1}{N(M-1)} \sum_{j=1}^{N} \sum_{i=1}^{M} [X_{ij}(t) - \overline{X}_j(t)]^2}{\sigma^2}, \quad (1)$$

where X_{ij} is the ith member of the jth ensemble, $\overline{X}_j(t)$ is the jth ensemble mean, N is the number of ensemble ($N = 10$), M is the number of members that also include the control run ($M = 11$), and σ^2 is the variance in the control experiment. The PPP then has a form of

$$\text{PPP} = 1 - \text{NEV}. \quad (2)$$

PPP amounts to a value of 1 for a perfect predictability and a value of 0 when the ensemble spread equals the variance from the control integration. The statistical significance of NEV and PPP is estimated by the F test as suggested by von Storch and Zwiers (1999). The damped persistence based on a red noise null hypothesis is also used for comparison. The NEV and PPP of a hypothetical ensemble generated by stochastic process are given by $1 - e^{-2[(1-a)/(1+a)]t}$ and $e^{-2[(1-a)/(1+a)]t}$ (Griffies and Bryan 1997), respectively, where a is the lag-1 autoregressive coefficient of variable X.

In comparison with PPP, we also calculate the diagnostic potential predictability (DPP), which is defined as a fraction of long time scale (or low frequency) variability with respect to the total variability (σ_L^2/σ^2) (Boer 2004); σ_L^2 is the variance of the m-yr mean SST, where m can be selected as any integer number (we select m to be 10, 20, and 30 in the present paper). The high DPP regions identify those areas in which long time scale variability stands out clearly from short time scale variability, and thus variability in these regions may be at least potentially predictable.

b. Decadal prediction experiments

The CM2.1 decadal prediction experiments (e.g., Yang et al. 2013) are initialized by the GFDL ensemble coupled data assimilation (ECDA) system (Zhang et al. 2007). ECDA applies the ensemble adjustment Kalman filter (Anderson 2001) to the fully coupled climate model CM2.1. The atmosphere assimilates the NCEP atmospheric analysis (Kalnay et al. 1996) and the ocean assimilates observations of SST from satellite [optimum interpolation SST (OISST)] and temperature and salinity from the *World Ocean Database 2009* (Boyer et al. 2009), which includes profiles from expendable bathythermograph temperature (XBT) and Argo profiles after year 2000. Ten-member ensemble decadal hindcasts were initialized on 1 January every year from 1961 to 2016 and integrated for 10 years. These experiments include the effects of changing radiative forcing from anthropogenic and natural sources. We also run a 10-member

ensemble of simulations using the same model with the same radiative forcing, but without initialization. The ensemble mean response of these uninitialized historical simulations can be taken as the externally forced response.

The CM2.1 predictions use a full-field assimilation method and have a systematic model drift due to model bias, common to many other models (Smith et al. 2013). We therefore subtract the lead-time-dependent climatology from each variable in the hindcast simulations, which effectively removes the climate drift [see details in Yang et al. (2013)]. Our primary interest is in the prediction skill from internal decadal variations in the SO. To distinguish internal variability from externally forced variations, we subtract the historical ensemble mean forced response from both the hindcasts and the output of the ECDA. In addition to the ECDA reanalysis, we also use several observational datasets for evaluating prediction skill, including the Extended Reconstruction Sea Surface Temperature (ERSST) analysis version 3b (Smith and Reynolds 2004), the UK Met Office Hadley Centre's Sea Ice and Sea Surface Temperature dataset (HadISST; Rayner et al. 2003), the National Centers for Environmental Prediction (NCEP) atmosphere reanalysis (Kalnay et al. 1996), and the Goddard Institute for Space Studies (GISS) surface temperature analysis for the globe (Hansen et al. 2001).

We use the traditional correlation and mean square skill score (MSSS) (Goddard et al. 2013) to measure the prediction skill. The MSSS is defined as

$$\text{MSSS} = 1 - \frac{\sum_n (O_n - H_n)^2}{\sum_n (O_n - \overline{O})^2}, \quad (3)$$

where O_n and H_n are the observation and hindcast predicted values for the nth year, respectively, and $\overline{O}$ is the climatological mean value of observation. The MSSS has a value of one for a perfect prediction and becomes negative when the hindcast error is larger than the error of a prediction using observed climatology.

3. Multidecadal variability over the Southern Ocean

In the GFDL CM2.1 model, the internal multidecadal variability over the SO is dominated by fluctuations in the Weddell Sea (WS) deep convection (Zhang and Delworth 2016b). The deep convection status can be well represented by the strength of the Antarctic Bottom Water (AABW) cell. Figures 1a and 1b show the global meridional overturning circulation (GMOC) streamfunction in depth space and density space, respectively.
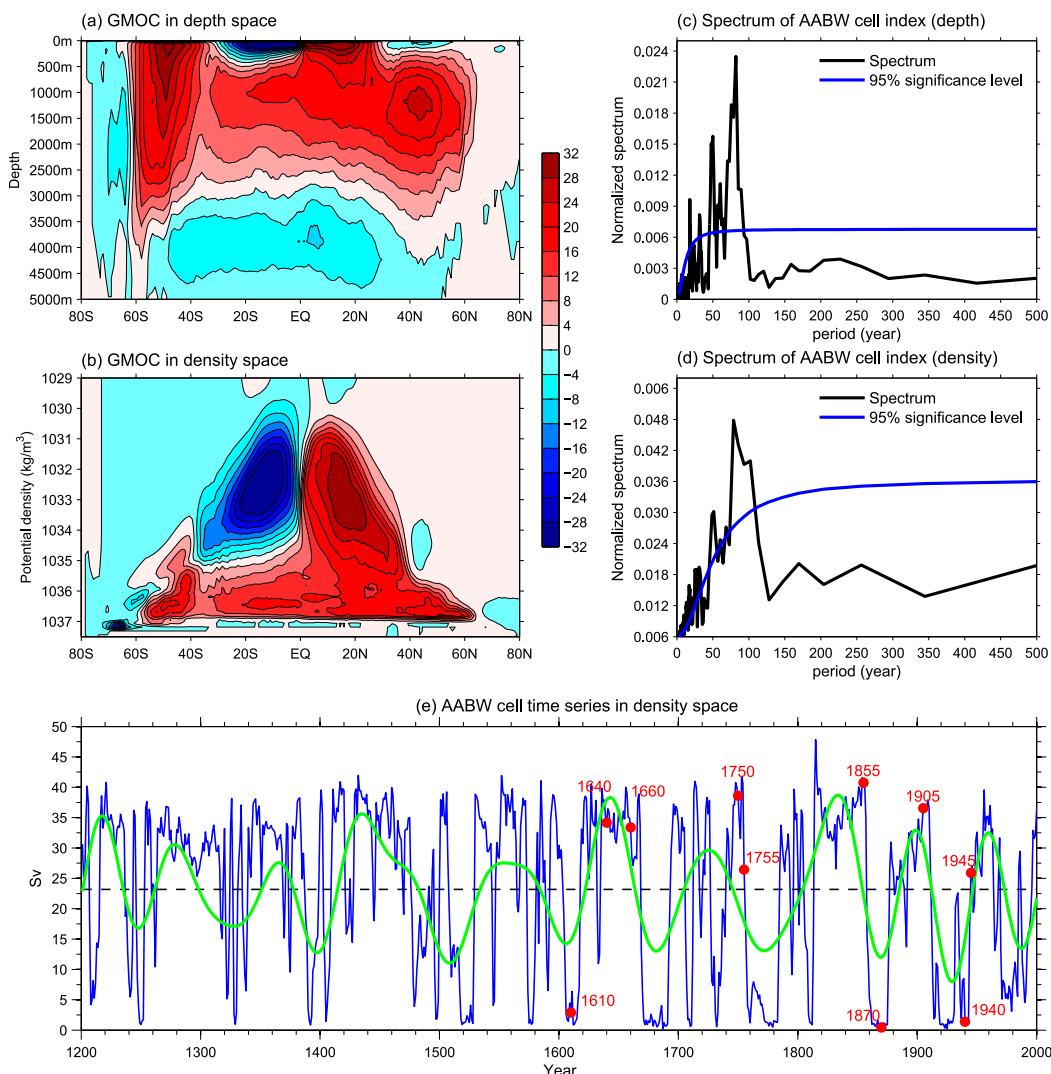
FIG. 1. Long-term mean global meridional overturning circulation (GMOC) streamfunction in (a) depth space and (b) density space. Unit is Sv. Power spectrum of the normalized Antarctic bottom water (AABW) cell index defined as the absolute value of minimum streamfunction across 67°S section in (c) depth and (d) density spaces, respectively. The blue line indicates the 95% significance level based on the red noise null hypothesis. (e) Annual mean time series of the AABW cell index in density space from year 1200 to 2000 (blue line) with starting points of "perfect ensemble" predictability experiments shown as red dots. The green line denotes the 25-yr low-pass filtered AABW cell index, while the dashed black line denotes the 800-yr mean of the AABW cell index. Unit is Sv.

Note that our model calculates the streamfunction in density space online, which considers high-frequency information and does not do any approximation. The negative streamfunction south of 60°S denotes the anti-clockwise AABW cell. Here, the AABW cell strength index is defined as the absolute value of the minimum streamfunction across 67°S section where is the location of maximum long-term mean overturning. The long-term mean AABW cell in depth space is approximately 8 Sv ($1 \, \mathrm{Sv} \equiv 10^6 \, \mathrm{m}^3 \, \mathrm{s}^{-1}$) using this definition, while it is much stronger in density space with a value of approximately 23 Sv that is within the range of observed estimates

($\sim$20 Sv; Lumpkin and Speer 2007). The AABW cell index in the last 500 years of the control integration shows pronounced multidecadal variability in both depth and density spaces, with peak periods around 60–120 years (Figs. 1c,d). These internal low-frequency fluctuations are also seen from the AABW cell time series (Fig. 1e). Although the AABW cell magnitude in density space is more realistic than in depth space, their time series are almost in phase (not shown). The overlapping red points in Fig. 1e denote the starting years of perfect predictability experiments. Some experiments start from a year corresponding to relatively weak AABW cell
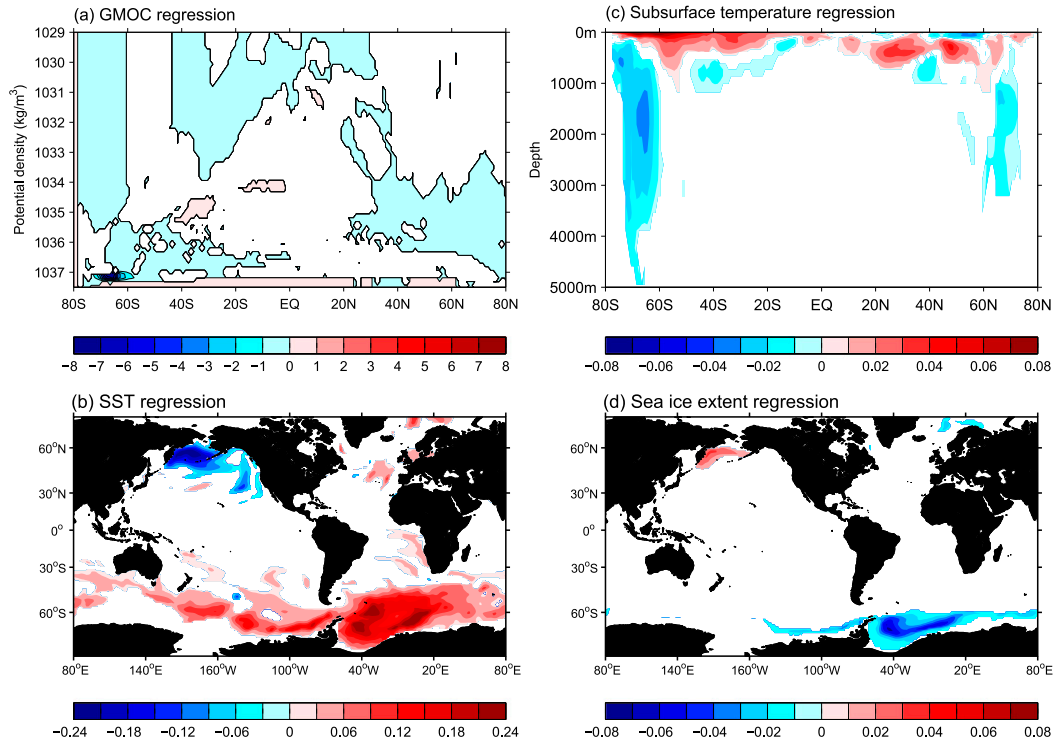
FIG. 2. Climate variability over the Southern Ocean (SO). Regressions of (a) the GMOC (Sv) in density space, (b) SST (°C), (c) zonal mean temperature profile (°C), and (d) sea ice concentration (100%) against the normalized AABW cell index. The shading only shows values significant at 95% significance according to a Student's *t* test.

conditions (years 1610, 1870, and 1940), some from strong conditions (years 1640, 1660, 1750, 1855, and 1905), and some from intermediate conditions (years 1755 and 1945).

Figure 2 shows the impact of multidecadal AABW cell fluctuations on the SST, subsurface temperature, and sea ice. Associated with a stronger-than-normal AABW cell (Fig. 2a), the SST shows a broad warming anomaly over the SO, with maximum value over the WS (Fig. 2b). This is consistent with our physical understanding: the spinup of the AABW cell is associated with increased vertical convection, which brings SO subsurface warm water to the surface and thereby leads to positive (negative) temperature anomalies at the surface (subsurface) (Fig. 2c). The warm surface water reduces the sea ice concentration around the Antarctic (Fig. 2d). The opposite is also established when the AABW cell spins down.

4. Prognostic potential predictability using perfect ensemble approaches

a. Decadal predictability of the AABW cell

We show in Fig. 3 the AABW cell predictability for the 10 sets of the 10-member ensembles. We identify the

experiments by the year at which initial conditions are taken from the Control simulation—the years listed are from the Control simulation, and have no relationship to calendar years. Experiments for years 1610, 1870, and 1940 (Figs. 3a, 3g, and 3i, respectively) start from the mature negative phase of the AABW cell index. Using the long-term mean AABW cell index in control run (yellow line in Fig. 3) as a reference to depict below-, near-, and above-normal AABW cell strength, we can see that the ensemble means in those panels (red line in Fig. 3) tend to follow the AABW cell variations of the respective control run (blue line in Fig. 3) for almost three decades. The predictability starting from year 1870 (Fig. 3g) seems to be the largest among three cases, due to its small ensemble spread. The ensemble spread is relatively small before year 1885 and gradually increases thereafter. In contrast, the ensemble spread does not change very much after the initial 5 years in the 1940 case. In the 1610 case, the ensemble spread increases much faster and earlier than that in the 1870 case. The above diversity suggests that the predictability more or less depends on the AABW cell initial state.

Figures 3b, 3c, 3d, 3f, and 3h display the AABW cell predictability initialized from its stronger-than-normal status. Again, all ensemble means track the initial
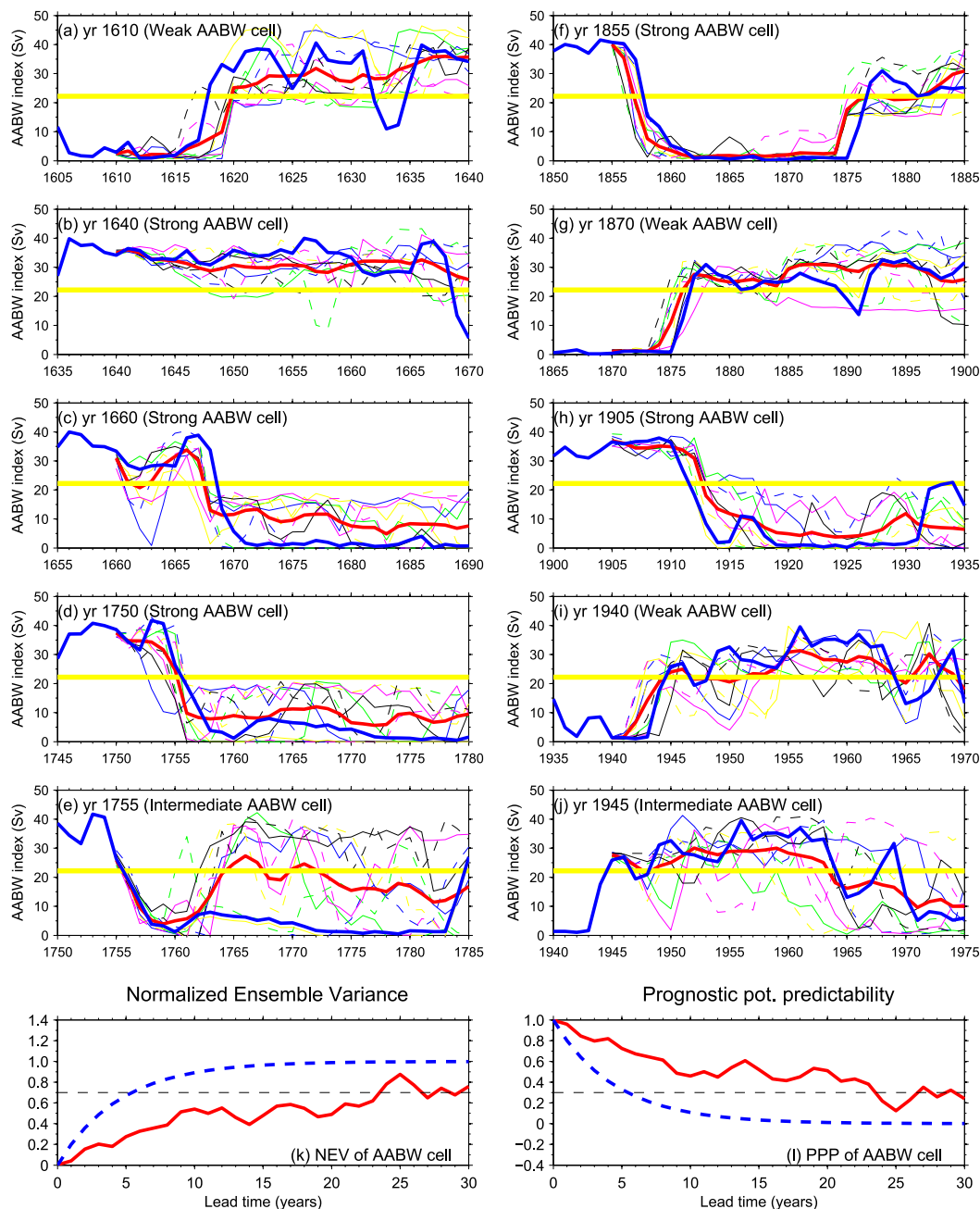
FIG. 3. AABW cell index (density space) trajectories in the control run (bold blue line, starts 5 years from the beginning of the ensemble run) and in the perfect ensemble predictability experiments (10 thin colored lines) initialized in 1 January of year (a) 1610, (b) 1640, (c) 1660, (d) 1750, (e) 1755, (f) 1855, (g) 1870, (h) 1905, (i) 1940, and (j) 1945. The bold yellow line denotes the long-term mean AABW cell index averaged in the last 500 years in the control run. Unit is Sv. (k) Normalized ensemble variance (NEV) and (l) prognostic potential predictability (PPP) averaged over 10 ensemble experiments (bold red line), along with the red noise null hypothesis (dashed blue line). The dashed black horizontal line denotes the 95% significance level.

control run relatively well, although with less variability due to averaging effect. In particular, the 1855 case has the highest predictability, in which the ensemble mean almost overlaps the control run in the first two decades with small ensemble spread. In the 1640 case, the ensemble mean keeps the positive phase of AABW cell for almost the whole prediction period, in agreement with the control run. In the rest of the three experiments, the

perfect ensemble experiments basically capture the AABW cell transition from the positive phase to the negative phase, but with a relatively large ensemble spread compared to the 1855 and 1640 cases.

We finally show in Figs. 3e and 3j the AABW cell predictability starting from intermediate conditions. Case 1755 initializes from an intermediate condition as the AABW cell transits from the positive phase to the negative phase. In the first 5 years, the ensemble mean matches the control run very well with a small ensemble spread. After that, the ensemble mean gradually deviates from the control run and the ensemble spread increases as well. For case 1945, the AABW cell starts from an average condition that coincides with a transition from negative phase to positive phase. Although the 10 ensemble means track the control run fairly well for the entire prediction period (30 yr), the ensemble spread is quite large even at the beginning. These results suggest that forecasts starting from intermediate conditions of natural internal variability of AABW cell seem to be less skillful than those initialized from mature phases. A caveat is that the limited number of experiments may limit the robustness of this demonstration. However, this preliminary claim is consistent with the argument proposed by Zhang et al. (2017, manuscript submitted to *J. Climate*), who used a diagnostic APT method to examine the SO SST predictability.

To get a quantitative estimate of the potential predictability of the AABW cell, the averaged skills (NEV and PPP) from 10 start dates as a function of lead time up to three decades are exhibited in Figs. 3k and 3l. The NEV of the AABW cell index indicates a statistically significant predictability up to 23 years. The ensemble variance increases more slowly than the red noise null hypothesis, suggesting that predictability exceeds that predicted by the damped persistence forecast. Accordingly, the PPP shows a decreasing skill as the lead time increases, with a loss of predictability after about 23 years. The PPP also exhibits better skill than that of damped persistence. The PPP value averaged over the first decade is as high as 0.72, which decreases to 0.5 in the second decade. The AABW cell predictability here is mainly attributed to its oscillatory characteristics (Figs. 1c,d).

b. Decadal predictability of the AABW cell fingerprints

The AABW cell, mainly representing deep ocean signals, is very difficult to observe on long time scales, therefore limiting the ability to validate model predictions. In the CM2.1 model, a stronger-than-normal (weaker-than-normal) AABW cell is accompanied by a broad SST warming (cooling) over the SO, with

maximum anomalies over the WS (Fig. 2b). Thus, SST anomalies averaged over the WS (75°–55°S, 52°W–30°E) and SO (70°–50°S, 0°–360°E), which are relatively easier to monitor, can be used as fingerprints of the AABW cell variations. We describe the SST trajectories in cases 1870, 1905, and 1945 (Figs. 4a–c and 4e–g), as they are typical examples initialized from below-, above-, and near-normal AABW cell strengths. Similar conclusions hold for the other seven ensembles. Figures 4a–c show that the ensemble means of the WS averaged SST track the variations of the control run in all three cases, with variations that mirror those of the AABW cell. This implies that the WS SST is predictable when the AABW cell is predictable. A close inspection finds that the SST trajectories are noisier than those of the AABW cell (cf. Figs. 3 and 4). This is not surprising, since SST is more affected by the high-frequency atmosphere forcing and radiative effects. The corresponding PPP calculated from 10 ensemble sets shows a statistically significant skill on time scales of about 15 years, which is shorter than the AABW cell (Fig. 4d).

Similarly, the SO area averaged SST predicted from the perfect ensemble experiments basically tracks the control run fluctuations (Figs. 4e–g), particularly the ensemble means. The SO SST has weak predictability based on PPP (Fig. 4h), which is also implied from the large ensemble spread. Figure 4h shows that the PPP has a value as high as 0.8 in the first 5 years, but then quickly drops to 0.3 around the 10th year and loses significance thereafter. This suggests that the whole SO averaged SST is less predictable than the WS averaged SST. Note that the WS is the main deep convection region in the CM2.1 model, which acts as a window connecting the deep ocean signal to the surface and thus local SST has longer memory and predictability than other places. This is also why SST change associated with the AABW cell fluctuation has its maximum value over the WS (Fig. 2b).

We further show in Figs. 5a–c the PPP spatial map of SST averaged over the 10 ensemble experiments in the first, second, and third prediction period. Averaged over the first decade, the most predictable regions appear in the North Atlantic (Labrador Sea), the Nordic Seas, and the SO WS (Fig. 5a). These high PPP regions coincide with the model's deep convection sites where the deep ocean long memory can be transmitted to the surface. Over the SO, the PPP value quickly drops beyond the WS. Obviously, the PPP value averaged over the WS exceeds that averaged over the whole SO, in agreement with Fig. 4. The PPP value of the second decade is everywhere less significant than that of the first decade (cf. Figs. 5a and 5b). However, the predictability remains significant in part of the Labrador and Weddell Seas,
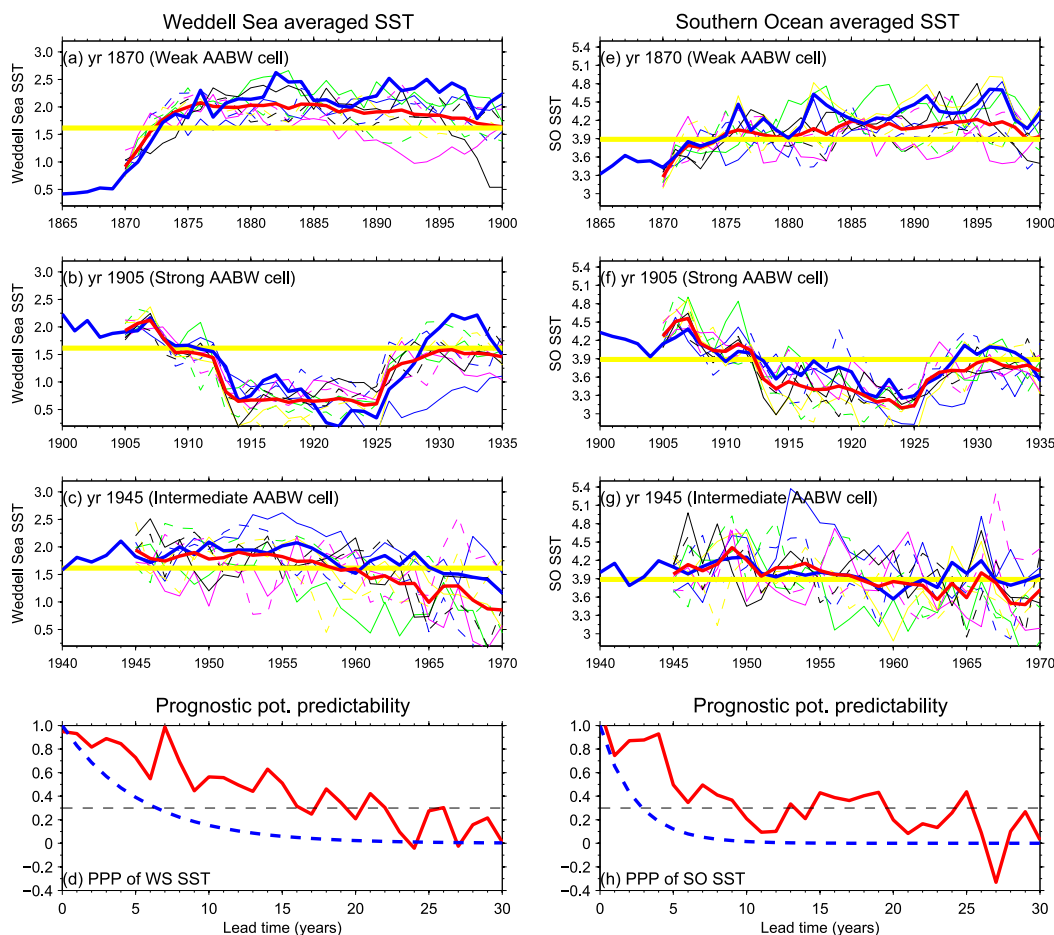
FIG. 4. As in Figs. 3a–j, but for the (a)–(c) Weddell Sea (WS) (75°–55°S, 52°W–30°E) and (e)–(g) SO (50°–70°S, 0°–360°E) averaged SST trajectories initialized in year 1870, 1905, and 1945. (d),(h) As in Fig. 3l, but for the PPP of WS and SO averaged SST anomalies.

indicating multidecadal predictability of SST in these areas. In the third decade, the PPP value further decreases everywhere (Fig. 7c). The North Atlantic SST totally loses its predictability, while the predictability is still significant in a small area within the WS. This suggests that SST has a longer predictability over the WS than that over the North Atlantic. The SST predictability over the SO and North Atlantic in CM2.1 model originates from oscillatory characteristics of the AABW cell and AMOC, respectively. The AABW cell has a longer period than that of AMOC (80 vs 20 yr), and thus the former has a longer predictability than the latter.

We also compare the PPP spatial pattern with the diagnostic potential predictability map of decadal means SST calculated from the control run (cf. Figs. 5a–c and Figs. 5d–f). The DPP pattern in CM2.1 model was already shown in our previous paper, which attempts to quantify the fraction of long-term predictable variance (10-, 20-, and 30-yr averaged SST) with respect to the

total variability. The high predictability regions identified by PPP method coincide to some extent with those of highest DPP scores. These regions mainly include the convection sites: North Atlantic Ocean and SO. There are also some discrepancies in the North Pacific Ocean. The SST predictability over the North Pacific Ocean identified by PPP is much weaker than those over the North Atlantic and SO. The DPP method, however, shows that the North Pacific western subpolar SST predictability is comparable to (much larger than) that over the SO (North Atlantic). The reason for these differences remains to be clarified in future. The similarities between the high predictability region (WS) identified by two approaches (PPP and DPP) (Fig. 5) and the region (WS) significantly sensitive to the low-frequency AABW cell fluctuations (Fig. 2b) further confirm that the SO internal SST predictability arises from the deep convection memory.
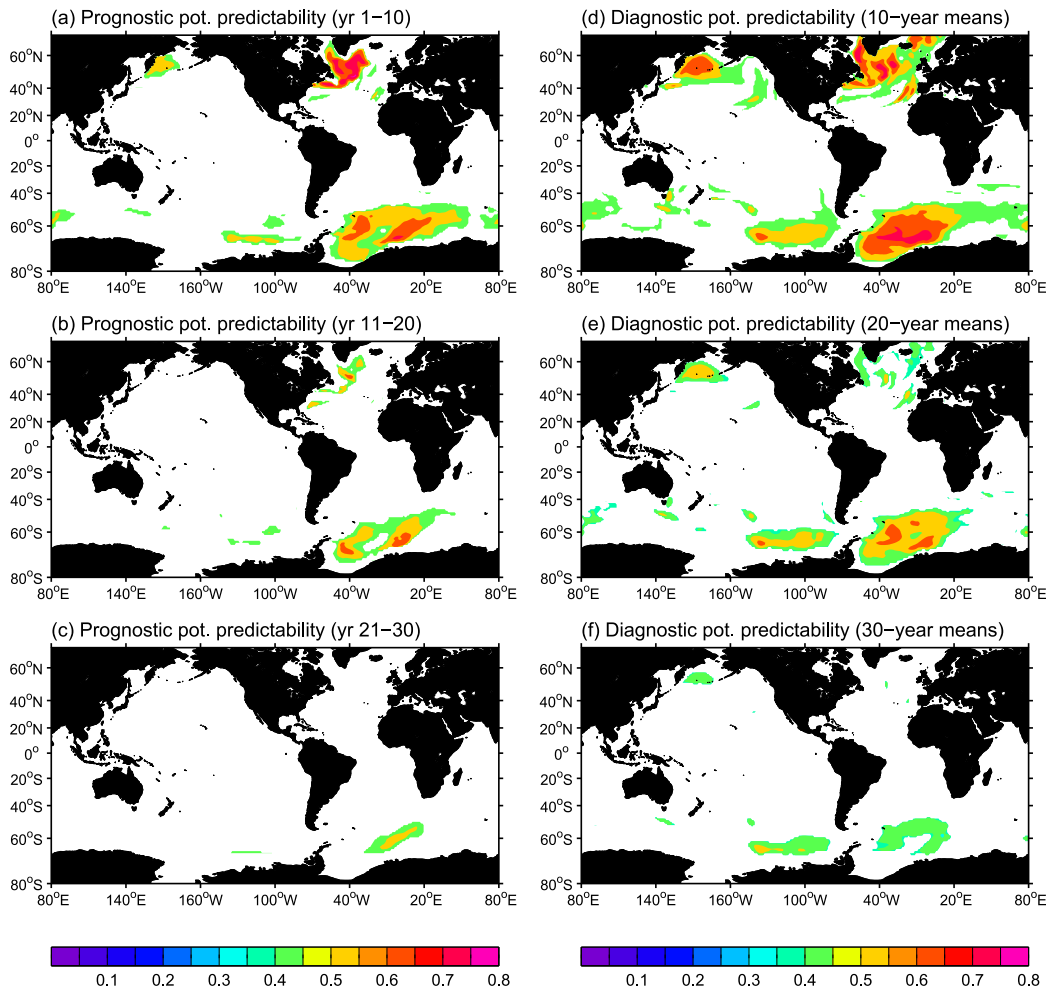
FIG. 5. PPP of SST averaged over the 10 ensemble experiments and the (a) first, (b) second, and (c) third prediction decade. Diagnostic potential predictability (DPP) of SST based on (d) 10-yr, (e) 20-yr and (f) 30-yr means. The shading only shows values significant at 95% significance according to the F test.

The AABW cell fluctuations are also associated with the WS subsurface temperature and Antarctic sea ice changes (Figs. 2c,d). These two fingerprints are expected to have higher skill than the SST, since they are less affected by the surface atmospheric perturbations. Consistent with our physical understanding, the subsurface temperature over the WS and Antarctic sea ice trajectories in three ensembles are out of phase with the SST and AABW cell fluctuations (Figs. 6a–c and Figs. 6e–g). The ensemble means generally follow the variations of the control run. Compared to SST, the trajectories of these two variables are much smoother and the ensemble spreads are much smaller (cf. Figs. 6 and 7). Accordingly, PPP shows high skills for the subsurface temperature and sea ice, comparable to that of the AABW cell (Figs. 6d,h).

Using perfect model predictability experiments, we find that the internal SO climate fluctuations in the CM2.1 model, including SST, ocean subsurface temperature, and

Antarctic sea ice, are predictable up to a decade or more. These decadal predictabilities primarily arise from the oscillatory characteristics of deep convection over the WS. It is important to keep in mind that we have assessed the upper limit of the AABW cell predictability as well as its climate signatures when we apply it to reality, since we assume the coupled model here is perfect and could be perfectly initialized with three-dimensional observational fields. Actually, this is impossible. The CM2.1 model has significant biases over the SO compared to observations (Delworth et al. 2006). Long-term observations are also rare over the SO. In the next section, we attempt to examine the SO prediction skill in a real CM2.1 decadal forecast system that is initialized by ECDA reanalysis. Although the reanalysis data over the SO are not perfect, especially before the satellite era starting in 1979, the following analysis may give some indications as to where and to what extent skillful decadal forecasts might be possible.
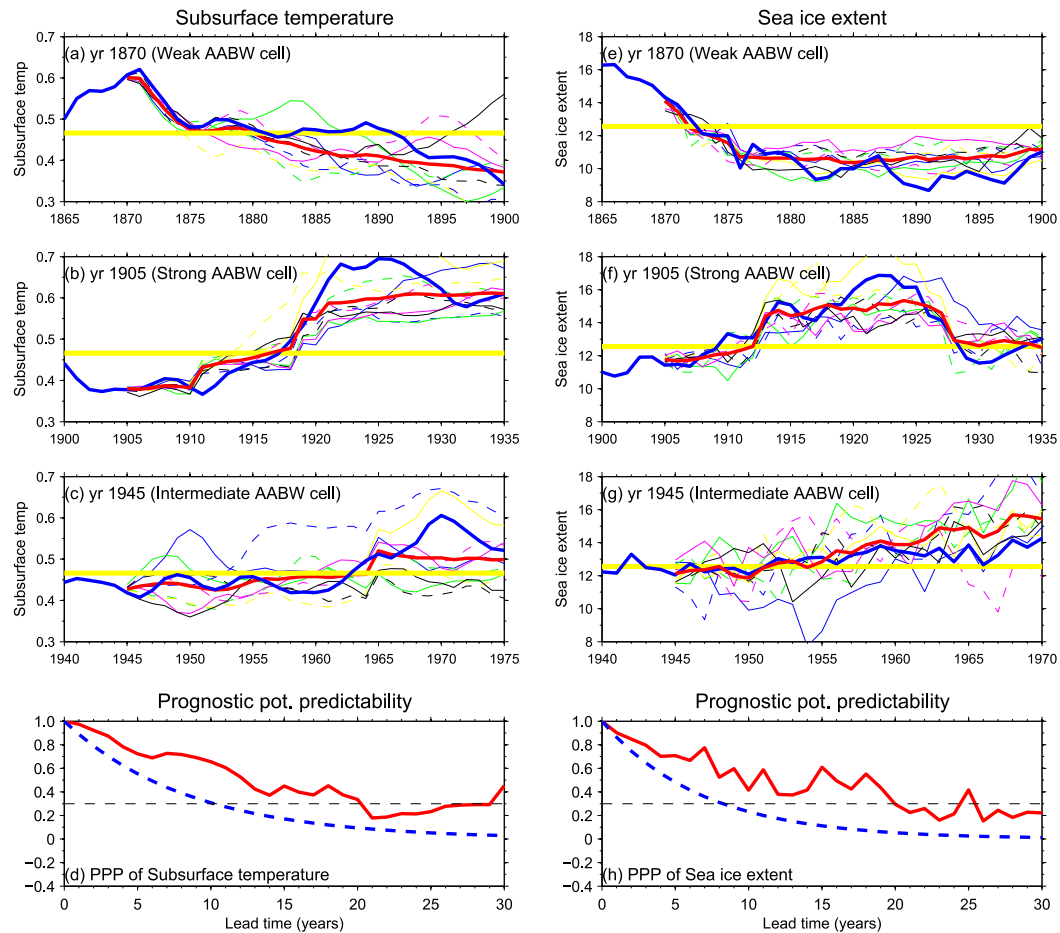
FIG. 6. As in Fig. 4, but for the (a)–(d) WS averaged subsurface temperature (°C) at 850 m and (e)–(h) Antarctic sea ice extent (10^{12} m^2).

5. Prediction skill in the decadal hindcast run

a. Skill assessment of the Southern Ocean SST

With regard to the relative reliability of SST data in reanalysis/observation, we first look at the SST prediction skill in the hindcast run. Figure 7 shows the correlation coefficient of SST anomalies between the CM2.1 hindcasts and ECDA reanalysis at each grid point as a function of lead time from 2 to 10 years over the period 1961–2014 (left panels) and 1980–2014 (right panels). Consistent with previous studies (e.g., Yang et al. 2013), SST anomalies over the North Atlantic subpolar region can be predicted up to a decade in advance, while the North Pacific SST shows lower prediction skill (Figs. 7a–e). Around the SO, the hindcasted SST variations are significantly correlated with the ECDA with several years lead time, with a maximum over the WS along with a secondary maximum north of the Ross Sea (Figs. 7a–e). These significant predictive skills in the SO SST suggest that initialization enables us

to predict the SO SST on decadal time scales. The SO prediction skill is still apparent if we only examine hindcasts from the relatively short satellite era (Figs. 7f–j), albeit with a somewhat lower correlation after 6-yr lead. We also assess the predictive skill using ERSST data and find similar results (not shown).

We show in Fig. 8 the area averaged SST time series for the 1961–2014 period as well as their prediction skill over the WS and SO, respectively. Figures 8a and 8b show that the recent negative SST anomalies over the SO are retrospectively well predicted by the 5-yr mean initialized hindcasts. The correlations over the WS are all above 0.6 at lead 1–10 years and significant at the 90% confidence level (Fig. 8c). Moreover, these hindcast skills are higher than persistence, demonstrating the important role of initialization for these predictions. The SO averaged time series exhibit less prediction skills than that in the WS after 6-yr lead (Fig. 8d), consistent with Fig. 7. The correlation gradually decreases as the lead time increases and the significance is lost after 7-yr lead. We also use the
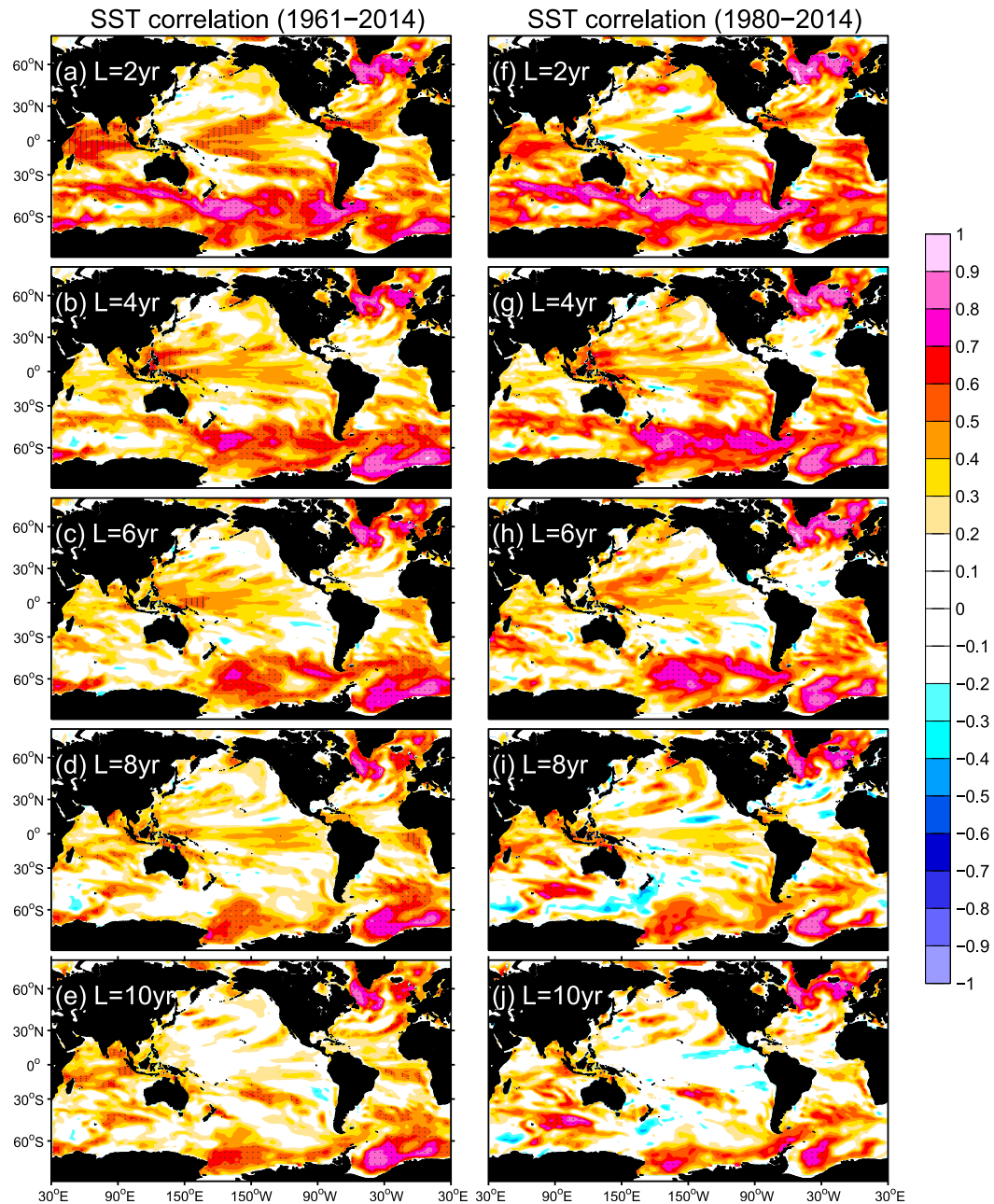
FIG. 7. Skill in hindcasting internal SST variations in terms of anomaly correlation coefficients at each grid point. Correlation between the hindcast data initialized 2–10 yr in advance and ECDA reanalysis from (a)–(e) 1961–2014 and (f)–(j) 1980–2014. The external forcing is removed by subtracting ensemble mean SST of the uninitialized run from the total signal in both hindcasts and ECDA. The black dots overlapped on the shading indicate the correlations are significant at 90% confidence level based on a bootstrap approach using 1000 sets of reordered values.

mean square skill MSSS to measure the prediction skill. The MSSS shows positive values in all lead years for the WS and in the first seven lead years for the SO, indicating an improved prediction skill using the initialized model relative to the uninitialized climatology. Again, the SO SST prediction skill after 1980 shows similar results compared to the whole period discussed here (not shown).

b. Possible mechanisms explaining the successful prediction

The SST prediction skill in the decadal hindcast experiments (Fig. 7) is very similar to the PPP skill estimated from perfect model predictability experiments and the DPP skill diagnosed from the control run (Fig. 5). All
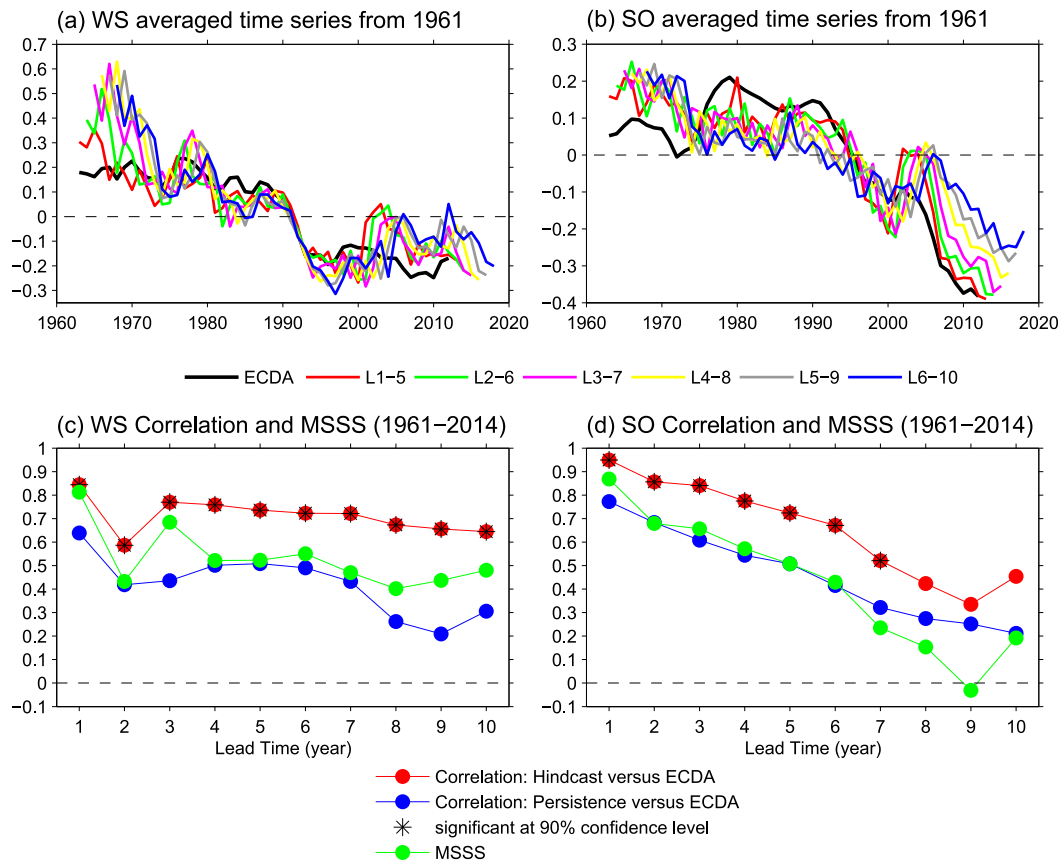
FIG. 8. Time series of 5-yr mean initialized hindcasts (1961–2014) of SST anomalies (°C) averaged over the (a) Weddell Sea (75°–55°S, 52°W–30°E) and (b) Southern Ocean (70°–50°S, 0°–360°E) compared with ECDA (black line). (c),(d) Anomaly correlation and mean squared skill score (MSSS) between the hindcasted SST and ECDA as a function of lead time. The persistence skill is also shown.

three methods show the highest skill over the WS where internal SST variability is very sensitive to deep convection or AABW cell fluctuations. Thus, it is reasonable to speculate that the SO SST skill in the decadal hindcasts may arise from the correct synchronization of the slowly evolving AABW cell internal variability in the model.

For the AABW cell, we focus on the period 1980–2014 when observations are somewhat more reliable than earlier periods. The AABW cell derived from the ECDA shows a decreasing trend in recent decades (Fig. 9a), which corresponds to a weakening of deep convection, thereby isolating the subsurface warm water from the surface, leading to a cooling trend for SST (Figs. 9b,c). Figure 9a shows that the 7-yr mean initialized AABW fluctuations agree well with that in ECDA, with ensemble spread small compared to the signal. This 7-yr hindcast skill is largely due to capturing the long-term decreasing trend, with weaker AABW formation after 1995 than before 1995. The SST hindcast skill over the WS and SO with a 7-yr lead is thus consistent with

predicting AABW cell variations (Figs. 9b,c). Consistent with the SO SST, Antarctic sea ice shows an expansion that can be predicted approximately 7-yr in advance (Fig. 9d). This is again very likely to be attributed to skill in predicting the AABW cell variations. The important role of initializing AABW is further investigated by examining predictions started in 1993 (Fig. 10). The AABW is initialized as anomalously weak in the hindcast runs (Fig. 10a), corresponding to weak surface cooling and shoaled mixed layer depth (Figs. 10c,e). The AABW anomaly persists for several years due to the large inertia of the ocean (Fig. 10b), which is also indicated from mixed layer depth change (Fig. 10f). The system successfully predicts negative (positive) temperature anomalies in the surface (subsurface) in the following years (Fig. 10d).

c. Climate impacts

To explore the predicted climate impacts associated with the SO cooling trend, CM2.1 hindcasts that predict a
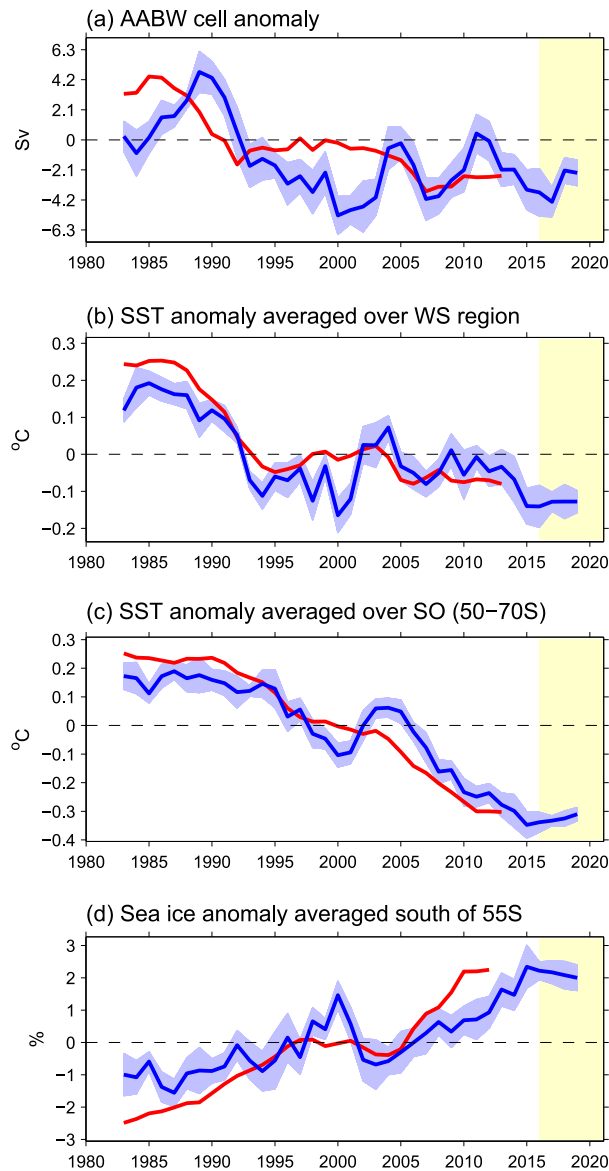
FIG. 9. Hindcast/forecast 7-yr (lead 1–7) means of selected internal time series, compared with ECDA/observation. (a) AABW cell strength (Sv). (b) WS averaged SST index (°C). (c) SO averaged SST index (°C). (d) Antarctic (south of 55°S) sea ice concentration (%). Red lines indicate time series in reanalysis, while blue lines denote the hindcasted time series with ensemble spread (shading) overlapped. Yellow shading begins in 2016, indicating the start of the true forecast period. AABW and SST are from ECDA, sea ice from HadISST, and wind stress from NCEP.

cold SST (1996–2010) are compared to earlier hindcasts (1980–91). We present here the averaged results from forecast years 2–7 to highlight where the impact of initialization remains beyond year 1. As suggested by Robson et al. (2013), comparing differences between hindcasts for the same lead times removes the need to do bias correction. Subtracting the uninitialized historical forecast

from the initialized hindcasts and observations removes an estimate of the externally forced trend and is better able to reveal the impact of initialization. In austral winter [June–August (JJA)], CM2.1 hindcasts predict broad sea ice increases over the SO (Fig. 11a). There is an exception over the northern Weddell Sea where the sea ice response shows a decrease. The sea ice decrease is primarily associated with the warm temperature advection by the anomalous northwesterly wind (see the sea level pressure in Fig. 11a). Cold surface air temperature (SAT) anomalies are predicted over most of the Antarctic continent (Fig. 11c). The warm SAT around 30°W is due to warm advection from the north, and coincides well with the sea ice decrease. The sea ice and SAT predictions are consistent with the negative SST anomalies during the same period in the model (Fig. 9c). Further inspection finds that the predictions of sea ice and SAT are generally in agreement with those observed in the SO (Figs. 11a,c vs Figs. 11b,d), although the magnitudes in the hindcast simulations in many places are relatively small.

6. Discussion and summary

In this study, we use prognostic methods to estimate the decadal predictability of SO deep convection and its associated climate impacts, which complement our previous paper that uses a diagnostic method to examine the SO SST predictability. We conduct perfect model predictability experiments using the GFDL CM2.1 climate model, with 10 ensembles starting from 10 different initial states. Results show that SO deep convection (or AABW cell) fluctuations, associated with internal climate variability, are potential predictable up to 20 years in advance. The associated WS subsurface temperatures and Antarctic sea ice have comparable potential predictability as the AABW cell. These two fingerprints are relatively easy to monitor by direct measurements, thereby providing another approach to assess current and future variability of the SO deep convection. A close relationship between the AABW cell and SST also exists. Because of the close contact with high-frequency atmospheric perturbations, the SST variability over the WS and SO is less predictable, but still predictable for lead times up to a decade. Despite the limited number of perfect model ensembles, the AABW cell and its associated climate impacts seem to have higher predictability when initialized from mature phases of the internal variability as compared to those initialized from intermediate conditions. This result is consistent with the diagnostic analysis in our previous paper but needs to be confirmed in subsequent studies by more experiments specifically designed to investigate the role of initial conditions.
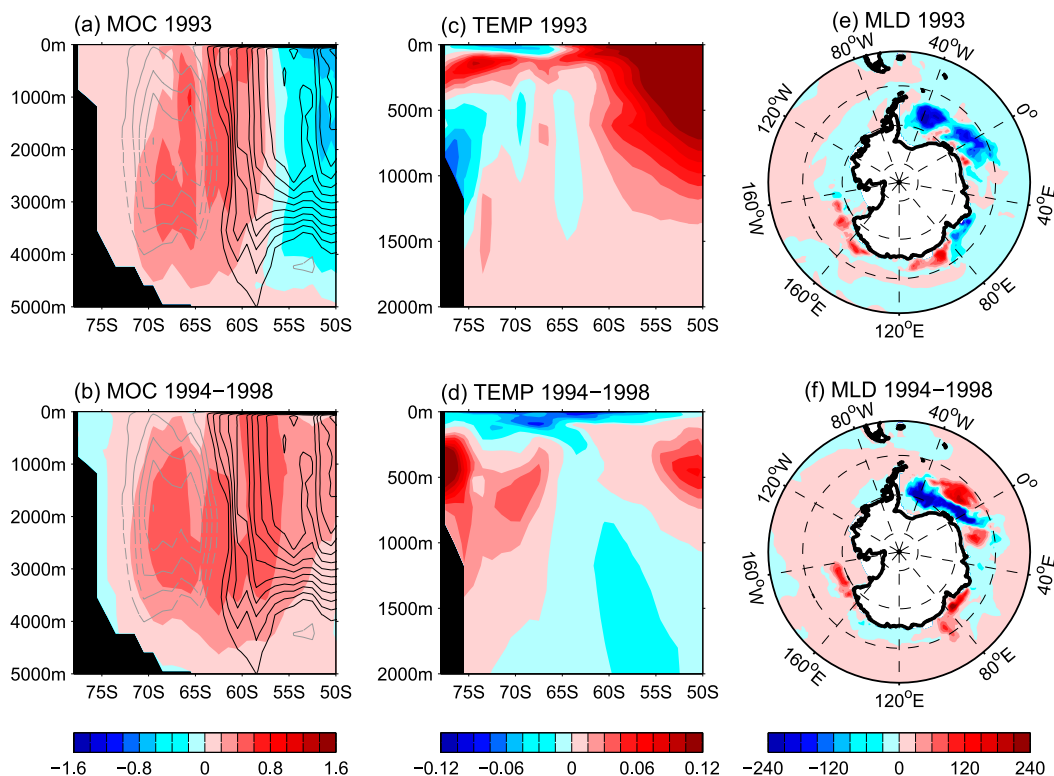
FIG. 10. The (a) first year and (b) 2–6-yr averaged GMOC (Sv) anomalies in GFDL CM2.1 hindcasts initialized in year 1993. The contour imposed on the shading denotes the long-term mean GMOC. The dashed gray (solid black) contour denotes negative (positive) streamfunction (contour interval 2 Sv). (c),(d) As in (a),(b), but for zonal mean temperature (°C). (e),(f) As in (a),(b), but for mixed layer depth (m).

The perfect ensemble experiments provide an upper limit of the SO climate predictability since we assume a perfect model and perfect knowledge of the initial conditions. Indeed, the GFDL CM2.1 model has significant biases over the SO compared to observation. The approximately 80-yr internal cycle of the AABW cell in the CM2.1 model also greatly influences the predictability time scale. Different models could have different time scales of internal variability over the SO, such as the 300-yr time scale variations in the Kiel Climate Model (Martin et al. 2013).

In reality we do not have perfect observations with which to initialize prediction models. Thus, to what extent we can apply our conclusions to the real climate system remain a big challenge. Regarding these issues, we have extended our perfect model prognostic study to real decadal hindcast/forecast experiments that are initialized from observations through an ECDA reanalysis system. This at least can give us some indications as to where and to what extent skillful decadal forecasts might be possible. We found that the internal variability component of SO SST fluctuations can be predicted several years in advance. The success of the hindcast

predictions, especially of the recent SO cooling that appears in both observations and the ECDA analysis, is largely due to initialization of the strength of AABW formation. The GFDL CM2.1 hindcasts also predict significant changes in Antarctic sea ice and SAT associated with the recent SO shifts, including both hemispheric mean and detailed spatial pattern, which generally agree with the observed changes. In contrast to the long-term expansion trend (Fig. 9d), the Antarctic sea ice extent in the last three months shows a record low. This short-term monthly sea ice changes seem to be more related to the anomalous wind than deep convection changes, which could be examined in our seasonal forecast system and will be our future work.

We have also made true predictions for the next decade (2016–25). We find that AABW formation is predicted to be weak for the next few years but has a tendency to transition to a neutral state (Fig. 9a). Thus, the negative SST anomalies and above normal sea ice over the SO may well continue for the next several years, although the amplitude of the anomalies will tend to decrease (Figs. 9c,d). In brief, the conclusions from decadal hindcasts/forecasts are generally consistent with
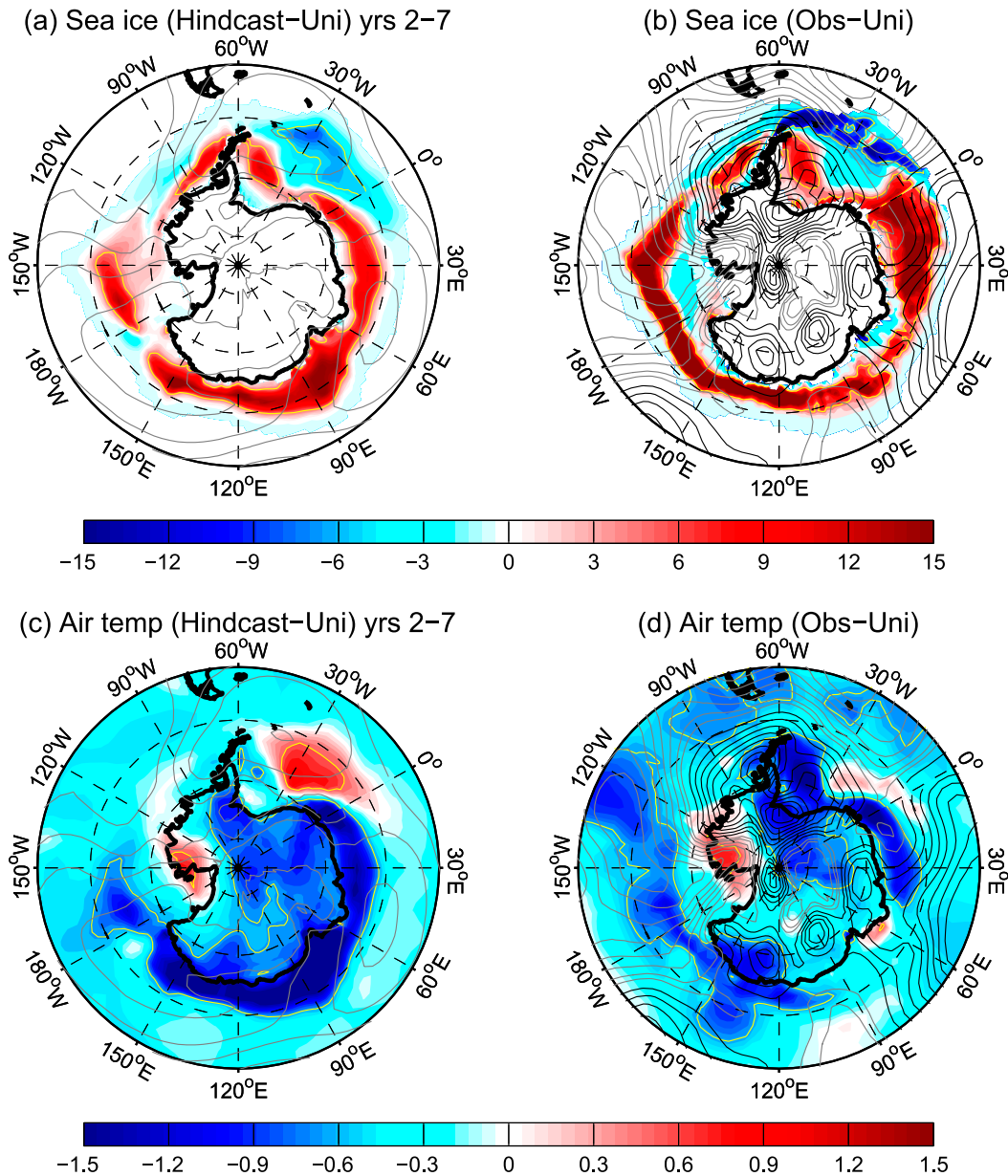
FIG. 11. Comparison of the observed and predicted (years 2–7) (a),(b) sea ice and (c),(d) surface air temperature (SAT) before and after the mid-1990s cooling of the SO, e.g., the difference of year 2–7 mean sea ice (%) and air temperature (°C) in hindcast initialized in 1996–2010 minus year 2–7 from hindcast initialized from 1980–91. The observed difference is taken from the 6-yr running mean in the same period. Difference is then made relative to the same difference in uninitialized run. The contours overlapping on the shading denote the sea level pressure (SLP) difference. The contour interval is 0.3 hPa for the SLP [solid (dashed) line denotes positive (negative) SLP]. The observed sea ice is from HadISST, SAT from GISS surface temperature analysis, and SLP from NCEP. The shading inside yellow contours indicates that the sea ice–SAT difference is significant at a 90% confidence level.

what we obtained from the perfect ensemble predictability experiments. This provides us some level of confidence that some characteristics and skills that appeared in the current paper have some basis in reality.

While past observations of the strength of AABW formation are sparse, our results suggest that substantial decadal predictive skill could be realized with adequate observations of the subsurface SO. The fact that the decadal predictions shown here have meaningful skill suggests that our initialization system may have captured some aspects of the time-varying rate of AABW formation, despite the paucity of observations. Further, our

results suggest that improved and sustained measurements of the subsurface SO could yield substantial benefits in terms of contributing to successful decadal predictions of the SO.

Given the importance of ocean circulation initialization in the hindcast skills, the SO predictability is very likely dependent on the ocean model resolution. Note that the ocean model in CM2.1 is approximately 1°, and it therefore cannot explicitly simulate mesoscale eddies in the SO. In addition, the initial conditions from ECDA we used in the hindcast experiments have considerable uncertainties, especially arising from changing observational networks. Thus, it would be very useful to repeat such hindcast experiments with other models and initialization methods to assess the robustness of the results, especially using models with substantially higher ocean resolution.

REFERENCES

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

Bintanja, R., G. J. van Oldenborgh, S. S. Drijfhout, B. Wouters, and C. A. Katsman, 2013: Important role for ocean warming and increased ice-shelf melt in Antarctic sea-ice expansion. *Nat. Geosci.*, **6**, 376–379, doi:10.1038/ngeo1767.

——, ——, and C. A. Katsman, 2015: The effect of increased fresh water from Antarctic ice shelves on future trends in Antarctic sea ice. *Ann. Glaciol.*, **56**, 120–126, doi:10.3189/2015AoG69A001.

Boer, G. J., 2004: Long time-scale potential predictability in an ensemble of coupled climate models. *Climate Dyn.*, **23**, 29–44, doi:10.1007/s00382-004-0419-8.

——, 2011: Decadal potential predictability of twenty-first century climate. *Climate Dyn.*, **36**, 1119–1133, doi:10.1007/s00382-010-0747-9.

Boyer, T. P., and Coauthors, 2009: *World Ocean Database 2009.* NOAA Atlas NESDIS 66, 217 pp. [Available online at ftp://ftp.nodc.noaa.gov/pub/WOD09/DOC/wod09_intro.pdf.]

Cavalieri, D. J., and C. L. Parkinson, 2008: Antarctic sea ice variability and trends, 1979–2006. *J. Geophys. Res.*, **113**, C07004, doi:10.1029/2007JC004564.

Comiso, J. C., and F. Nishio, 2008: Trends in the sea ice cover using enhanced and compatible AMSR-E, SSM/I, and SMMR data. *J. Geophys. Res.*, **113**, C02S07, doi:10.1029/2007JC004257.

DelSole, T., and M. K. Tippett, 2009: Average predictability time. Part I: Theory. *J. Atmos. Sci.*, **66**, 1172–1187, doi:10.1175/2008JAS2868.1.

Delworth, T. L., and M. E. Mann, 2000: Observed and simulated multidecadal variability in the Northern Hemisphere. *Climate Dyn.*, **16**, 661–676, doi:10.1007/s003820000075.

——, and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate*, **19**, 643–674, doi:10.1175/JCLI3629.1.

Galbraith, E. D., and Coauthors, 2011: Climate variability and radiocarbon in the CM2Mc Earth system model. *J. Climate*, **24**, 4230–4254, doi:10.1175/2011JCLI3919.1.

Goddard, L., and Coauthors, 2013: A verification framework for interannual-to-decadal predictions experiments. *Climate Dyn.*, **40**, 245–272, doi:10.1007/s00382-012-1481-2.

Gordon, A. L., 1978: Deep Antarctic convection west of Maud Rise. *J. Phys. Oceanogr.*, **8**, 600–612, doi:10.1175/1520-0485(1978)008<0600:DACWOM>2.0.CO;2.

——, 1982: Weddell Deep Water variability. *J. Mar. Res.*, **40**, 199–217.

Griffies, S. M., and K. Bryan, 1997: A predictability study of simulated North Atlantic multidecadal variability. *Climate Dyn.*, **13**, 459–487, doi:10.1007/s003820050177.

Hansen, J., R. Ruedy, M. Sato, M. Imhoff, W. Lawrence, D. Easterling, T. Peterson, and T. Karl, 2001: A closer look at United States and global surface temperature change. *J. Geophys. Res.*, **106**, 23 947–23 963, doi:10.1029/2001JD000354.

Holland, P. R., and R. Kwok, 2012: Wind-driven trends in Antarctic sea-ice drift. *Nat. Geosci.*, **5**, 872–875, doi:10.1038/ngeo1627.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

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

Kwon, Y.-O., and C. Deser, 2007: North Pacific decadal variability in the Community Climate System model version 2. *J. Climate*, **20**, 2416–2433, doi:10.1175/JCLI4103.1.

Latif, M., T. Martin, and W. Park, 2013: Southern Ocean sector centennial climate variability and recent decadal trends. *J. Climate*, **26**, 7767–7782, doi:10.1175/JCLI-D-12-00281.1.

Le Bars, D., J. P. Viebahn, and H. A. Dijkstra, 2016: A Southern Ocean mode of multidecadal variability. *Geophys. Res. Lett.*, **43**, 2102–2110, doi:10.1002/2016GL068177.

Lumpkin, R., and K. Speer, 2007: Global ocean meridional overturning. *J. Phys. Oceanogr.*, **37**, 2550–2562, doi:10.1175/JPO3130.1.

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

Martin, T., W. Park, and M. Latif, 2013: Multi-centennial variability controlled by Southern Ocean convection in the Kiel Climate Model. *Climate Dyn.*, **40**, 2005–2022, doi:10.1007/s00382-012-1586-7.

Matear, R. J., T. J. O'Kane, J. S. Risbey, and M. Chamberlain, 2015: Sources of heterogeneous variability and trends in Antarctic sea-ice. *Nat. Commun.*, **6**, 8656, doi:10.1038/ncomms9656.

Monselesan, D. P., T. J. O'Kane, J. S. Risbey, and J. A. Church, 2015: Internal climate memory in observations and models. *Geophys. Res. Lett.*, **42**, 1232–1242, doi:10.1002/2014GL062765.

O'Kane, T. J., R. J. Matear, M. A. Chamberlain, J. S. Risbey, B. M. Sloyan, and I. Horenko, 2013: Decadal variability in an

OGCM Southern Ocean: Intrinsic modes, forced modes and metastable states. *Ocean Modell.*, **69**, 1–21, doi:10.1016/j.ocemod.2013.04.009.

Philander, S. G. H., 1990: *El Niño, La Niña, and the Southern Oscillation*. Academic Press, 289 pp.

Pohlmann, H., M. Botzet, M. Latif, A. Roesch, M. Wild, and P. Tschuck, 2004: Estimating the decadal predictability of a coupled AOGCM. *J. Climate*, **17**, 4463–4472, doi:10.1175/3209.1.

Polvani, L. M., and K. L. Smith, 2013: Can natural variability explain observed Antarctic sea ice trends? New modeling evidence from CMIP5. *Geophys. Res. Lett.*, **40**, 3195–3199, doi:10.1002/grl.50578.

Purich, A., W. Cai, M. H. England, and T. Cowan, 2016: Evidence for link between modelled trends in Antarctic sea ice and underestimated westerly wind changes. *Nat. Commun.*, **7**, 10409, doi:10.1038/ncomms10409.

Purkey, S. G., and G. C. Johnson, 2010: Warming of global abyssal and deep Southern Ocean waters between the 1990s and 2000s: Contributions to global heat and sea level rise budgets. *J. Climate*, **23**, 6336–6351, doi:10.1175/2010JCLI3682.1.

——, and ——, 2012: Global contraction of Antarctic Bottom Water between the 1980s and 2000s. *J. Climate*, **25**, 5830–5844, doi:10.1175/JCLI-D-11-00612.1.

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

Robson, J. I., R. T. Sutton, and D. M. Smith, 2013: Predictable climate impacts of the decadal changes in the ocean in the 1990s. *J. Climate*, **26**, 6329–6339, doi:10.1175/JCLI-D-12-00827.1.

Smith, D. M., R. Eade, and H. Pohlmann, 2013: A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Climate Dyn.*, **41**, 3325–3338, doi:10.1007/s00382-013-1683-2.

Smith, T. M., and R. W. Reynolds, 2004: Improved extended reconstruction of SST (1854–1997). *J. Climate*, **17**, 2466–2477, doi:10.1175/1520-0442(2004)017<2466:IEROS>2.0.CO;2.

von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.

Yang, X., and Coauthors, 2013: A predictable AMO-like pattern in the GFDL fully coupled ensemble initialized and decadal forecasting system. *J. Climate*, **26**, 650–661, doi:10.1175/JCLI-D-12-00231.1.

Zhang, L., and T. L. Delworth, 2015: Analysis of the characteristics and mechanisms of the Pacific decadal oscillation in a suite of coupled models from the Geophysical Fluid Dynamics Laboratory. *J. Climate*, **28**, 7678–7701, doi:10.1175/JCLI-D-14-00647.1.

——, and ——, 2016a: Simulated response of the Pacific decadal oscillation to climate change. *J. Climate*, **29**, 5999–6018, doi:10.1175/JCLI-D-15-0690.1.

——, and ——, 2016b: Impact of the Antarctic bottom water formation on the Weddell Gyre and its northward propagation characteristics in GFDL CM2.1 model. *J. Geophys. Res. Oceans*, **121**, 5825–5846, doi:10.1002/2016JC011790.

——, ——, and F. Zeng, 2017: The impact of multidecadal Atlantic meridional overturning circulation variations on the Southern Ocean. *Climate Dyn.*, **48**, 2065–2085, doi:10.1007/s00382-016-3190-8.

Zhang, S., M. J. Harrison, A. Rosati, and A. T. Wittenberg, 2007: System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Mon. Wea. Rev.*, **135**, 3541–3564, doi:10.1175/MWR3466.1.

Zunz, V., H. Goosse, and F. Massonnet, 2013: How does internal variability influence the ability of CMIP5 models to reproduce the recent trend in Southern Ocean sea ice extent? *Cryosphere*, **7**, 451–468, doi:10.5194/tc-7-451-2013.