JAMES | Journal of Advances in Modeling Earth Systems®

RESEARCH ARTICLE
10.1029/2021MS002529

Key Points:

- Development of GFDL's new decadal coupled reanalysis/initialization system is documented
- The system captures the multi-decadal variations of Atlantic Meridional Overturning Circulation associated with North Atlantic Oscillation
- The new system significantly outperforms its predecessor in predicting multi-decadal SST trends in the Southern Ocean

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

X. Yang,
Xiaosong.Yang@noaa.gov

Citation:

Yang, X., Delworth, T. L., Zeng, F., Zhang, L., Cooke, W. F., Harrison, M. J., et al. (2021). On the development of GFDL's decadal prediction system: Initialization approaches and retrospective forecast assessment. *Journal of Advances in Modeling Earth Systems*, *13*, e2021MS002529. https://doi.org/10.1029/2021MS002529

Received 3 MAR 2021
Accepted 9 OCT 2021

On the Development of GFDL's Decadal Prediction System: Initialization Approaches and Retrospective Forecast Assessment

Xiaosong Yang[1], Thomas L. Delworth[1], Fanrong Zeng[1], Liping Zhang[1,2], William F. Cooke[1], Matthew J. Harrison[1], Anthony Rosati[1,2], Seth Underwood[1], Gilbert P. Compo[3,4], and Chesley McColl[3,4]

[1]NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA, [2]University Corporation for Atmospheric Research, Boulder, CO, USA, [3]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA, [4]NOAA/Physical Sciences Laboratory, Boulder, CO, USA

Abstract Using GFDL's new coupled model SPEAR, we have developed a decadal coupled reanalysis/initialization system (DCIS) that does not use subsurface ocean observations. In DCIS, the winds and temperature in the atmosphere, along with sea surface temperature (SST), are restored to observations. Under this approach the ocean component of the coupled model experiences a sequence of surface heat and momentum fluxes that are similar to observations. DCIS offers two initialization approaches, called A1 and A2, which differ only in the atmospheric forcing from observations. In A1, the atmospheric winds/temperature are restored toward the JRA reanalysis; in A2, surface pressure observations are assimilated in the model. Two sets of coupled reanalyses have been completed during 1961–2019 using A1 and A2, and they show very similar multi-decadal variations of the Atlantic Meridional Overturning Circulation (AMOC). Two sets of retrospective decadal forecasts were then conducted using initial conditions from the A1 and A2 reanalyses. In comparison with previous prediction system CM2.1, SPEAR-A1/A2 shows comparable skill of predicting the North Atlantic subpolar gyre SST, which is highly correlated with initial values of AMOC at all lead years. SPEAR-A1 significantly outperforms CM2.1 in predicting multi-decadal SST trends in the Southern Ocean (SO). Both A1 and A2 have skillful prediction of Sahel precipitation and the associated ITCZ shift. The prediction skill of SST is generally lower in A2 than A1 especially over SO presumably due to the sparse surface pressure observations.

Plain Language Summary In this paper we document the development and skill assessment of a new coupled decadal prediction system. We have developed a decadal coupled reanalysis/initialization system that does not use subsurface ocean observations. Under this approach the ocean component of the coupled model experiences a sequence of surface heat and momentum fluxes that are similar to observations. The new system offers two initialization approaches which differ only in the atmospheric forcing from observations. One approach utilizes the three-dimensional reanalysis data, while the another only uses the surface pressure observations. We document the process-based diagnosis of the improved representation of multi-decadal variations of the Atlantic Meridional Overturning Circulation associated with the observed North Atlantic Oscillation forcing. The skill assessment of the new decadal retrospective forecasts versus observations supports that the proposed approach is feasible for the decadal climate prediction, increasing diversity of modeling tools for the decadal climate initialization and predictability research. The experimental decadal prediction using the new system will be used for the real-time decadal climate outlooks.

1. Introduction

Multiyear to decadal climate prediction, characterized by combined signals from external radiative forcing changes and internal climate variations, is an important topic in climate research (Smith et al., 2007). A coordinated decadal prediction experiment was part of the fifth Coupled Model Intercomparison Project (CMIP5) (Meehl et al., 2009; Taylor et al., 2012) and assessed in the Intergovernmental Panel on Climate Change's Fifth Assessment Report (IPCC AR5) (Kirtman et al., 2013). With an increasing need for long-lead climate outlooks, the first experimental climate prediction of the coming decade made with multiple

Table 1
Survey of the Subsurface Ocean Data Used in Current Decadal Prediction Systems

Decadal forecast center	Model	Subsurface ocean data	Reference
Barcelona Supercomputing Center, Spain	EC-Earth3	Yes	Bilbao et al. (2021)
Bjerknes Center for Climate Research, Norway	NorCPM1	Yes	Wang et al. (2017)
Canadian Center for Climate Modeling and Analysis, Environment and Climate Change Canada	CanCM5	Yes	Sospedra-Alfonso and Boer (2020)
CSIRO, Australia	CAFE60v1	Yes	O'Kane, Sandery, Kitsios, Sakov, Chamberlain, Collier, et al. (2021) O'Kane, Sandery, Kitsios, Sakov, Chamberlain, Squire, et al. (2021)
Geophysical Fluid Dynamics Laboratory, USA	CM2.1	Yes	Delworth et al. (2006) Chang et al. (2013)
Geophysical Fluid Dynamics Laboratory, USA	SPEAR	No	Delworth et al. (2020) Yang et al. (2021, this paper)
LASG/Institute of Atmospheric Physics, China	FGOALS-s2	Yes	Wu et al. (2015)
Met Office Hadley Center, UK	HadGEM3	Yes	Williams et al. (2017)
Max Planck Institute for Meteorology, Germany	MPI-ESMl.2- HR	Yes	Pohlmann et al. (2019)
National Center for Atmospheric Research, USA	CESM1.1	No	S. G. Yeager et al. (2018)
University of Tokyo, National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, Japan	MIROC6	Yes	Kataoka et al. (2020)

models, initialized with prior observations, was launched in 2013 (Smith, Scaife, et al., 2013). The multi-model decadal prediction with a large ensemble size exhibits robust skill for precipitation over land and atmospheric circulation, in addition to surface temperature (Smith et al., 2019). These real-time experimental multi-model decadal predictions have become operational and are currently being made available through the WMO Lead Center for Annual-to-Decadal Climate Prediction (LC-ADCP) at the U.K. Met Office (www.wmolc-adcp.org). Decadal climate prediction project is also an important component (Boer et al., 2016) of CMIP6 (Eyring et al., 2016).

Although skillful decadal predictions have been achieved by multiple CMIP5 models, the optimal strategy for model initialization, including the data assimilation method and the input data selection from the time-varying climate observational network, is still an active research problem and a major challenge (Karspeck et al., 2017; Meehl et al., 2014). Pohlmann et al. (2013) reported that large disagreement exists among ten oceanic syntheses on reconstructing the historical AMOC variability, although the AMOC signal at 45°N averaged over the ten syntheses shows an increase from 1960s to the mid-1990s and a decrease thereafter. Karspeck et al. (2017) showed that the multiple state-of-the-art ocean reanalysis products, which assimilate the similar historical subsurface ocean observations, do not agree on the year-to-year variations of AMOC during 1960–2014, especially on the decadal to multidecadal time scales, suggesting that the historical reconstruction of AMOC is very sensitive to the assimilation procedure in the ocean reanalysis (Balmaseda et al., 2013), and the AMOC in those reanalyses is not balanced with the corresponding ocean heat content. In contrast, the large and diverse set of ocean-ice models in the Coordinated Ocean-ice Reference Experiments phase II (CORE-II), wherein a common interannually varying atmospheric forcing data set was used to force ocean-ice models, agree very well on the multidecadal AMOC variations (Danabasoglu et al., 2016; Karspeck et al., 2017). Beyond the simulation for understanding climate variability in the ocean-sea ice components in CORE, the forced ocean-sea ice (FOSI) simulation approach has been used for initializing the ocean and ice components in the Community Earth System Model (CESM) for producing a suite of large-ensemble decadal predictions, called CESM-DPLE (Danabasoglu et al., 2016; Yeager & Danabasoglu, 2014; S. G. Yeager et al., 2015, 2018). Among recent decadal prediction systems, mainly from WMO LC-ADCP and CMIP6, CESM-DPLE is the only one which does not use subsurface ocean observations (Table 1). Despite the fact that there is no direct assimilation of either ocean or sea ice observations in this decadal initialization of CESM, the CESM-DPLE exhibits very similar levels of skill for predicting the

annual mean sea surface temperature (SST) over the global ocean as those obtained by CMIP5 decadal prediction models, in which the subsurface ocean observations were directly assimilated (Meehl et al., 2014; S. G. Yeager et al., 2018). Thus, the FOSI approach with using estimates of observed surface fluxes provides an alternative strategy for initializing the coupled model for the decadal climate prediction.

GFDL's CMIP5 decadal prediction system has shown skillful predictions of the North Atlantic SST (Yang et al., 2013), the decadal shift in North Atlantic (Msadek et al., 2014) and the Southern Ocean SST (L. Zhang et al., 2017). It has provided routine decadal forecast to the real-time experimental multi-model decadal predictions since 2011 (Smith, Scaife, et al., 2013). This prediction system was based on GFDL's CM2.1 model (Delworth et al., 2006), and GFDL's ensemble coupled data assimilation (ECDA) with direct assimilation of subsurface ocean observations and atmospheric reanalysis data (Chang et al., 2013; S. Zhang et al., 2007). Recently, a new seasonal-to-centennial modeling and prediction system, called SPEAR (Seamless System for Prediction and EArth System Research), has been developed at GFDL for real-time experimental seasonal-to-decadal climate prediction and future climate projection, with various atmospheric resolutions and the ability to run large ensembles of simulations with available computational resources (Delworth et al., 2020). To leverage the SPEAR model for the seasonal to decadal prediction, the development of the new initialization system for the seasonal climate prediction using SPEAR was documented in Lu et al. (2020). In this manuscript, we document the development of the new coupled initialization system for the decadal climate prediction using SPEAR.

One aspect of the new decadal initialization development is on the historical reconstruction of the multidecadal AMOC variability associated with the observed low-frequency North Atlantic Oscillation (NAO) forcing in the coupled reanalysis, since such multidecadal AMOC variability is not well reconstructed in the GFDL's previous generation decadal prediction system (Karspeck et al., 2017). There exists strong observational and modeling evidence that multidecadal AMOC variability is a crucial driver of the observed Atlantic multidecadal variability (AMV) and associated climate impacts and an important potential source of enhanced decadal predictability and prediction skill (Yeager & Robson, 2017). Since the FOSI approach reconstructed multidecadal variations of AMOC and ocean heat content in the North Atlantic associated with the observed NAO forcing (Danabasoglu et al., 2016), we choose to develop the decadal initialization without direct assimilation of ocean subsurface observations. Specifically, we constrain the atmospheric component by either restoring to reanalysis data or through assimilation of surface pressure observations and constrain the ocean by restoring to observed SST. Physically, by constraining the air-sea interface of the coupled model using observed SST, air temperature and winds, its ocean component of the coupled model is forced by a very similar sequence of surface heat fluxes and wind stresses as observations, thus hopefully reproducing oceanic decadal variations similar to those observed. This approach in the coupled framework was also motivated by the NAO-forced coupled experiments using GFDL's models, in which the surface heat flux anomalies associated with observed NAO index can induce multidecadal variations in AMOC and poleward ocean heat transport in the Atlantic (Delworth et al., 2016).

In this context, the success of twentieth century reanalysis (20CR) project suggests a potential opportunity for a new decadal initialization methodology for coupled model prediction. 20CR fields reconstructing historical weather to climate variations in the troposphere agree very highly with full-input reanalysis data (Compo et al., 2011; Slivinski et al., 2019). 20CR fields of latent heat fluxes and wind speeds over the ocean also agree well with reanalysis and satellite-based estimates (Robertson et al., 2020). We therefore hypothesize that assimilating surface pressure observations in a coupled model will provide adequate information for reconstructing the observed air-sea heat and momentum fluxes, and when combined with SST restoring, may generate suitable initial conditions for coupled decadal predictions. Here we adapt the 20CR surface pressure data assimilation strategy to the coupled model initialization and test this hypothesis.

In Section 2 we describe the SPEAR model, the initialization approaches and the observational datasets. The coupled reanalyses, retrospective forecasts, uninitialized simulations and verification methods are described in Section 3. We present in Section 4 a suite of diagnostics to illustrate the multidecadal variability of AMOC and the subsurface ocean heat content in the coupled reanalysis. In Section 5, the skill assessment of the retrospective forecasts is presented. Conclusions and discussions are presented in Section 6.

2. Prediction Model, Initialization Approaches and Observational Datasets

2.1. SPEAR Model

SPEAR is GFDL's newly developed coupled modeling system for seasonal to multidecadal predictions and projections (Delworth et al., 2020). The SPEAR models incorporated many newly developed components used in the GFDL CM4 model (Held et al., 2019), e.g., the atmosphere and land components from AM4-LM4 (Zhao et al., 2018a, 2018b); and ocean and ice components from the ocean component of CM4 (OM4) (Adcroft et al., 2019). The details of the model configuration, physical parameterization choices and simulation characteristics can be found in Delworth et al. (2020).

SPEAR offers various atmospheric resolutions for users to optimize the model configuration, based on the major research interest and/or prediction focus with the given computational resources (Delworth et al., 2020). For the specific application in the decadal climate prediction, we use the low resolution of SPEAR model (SPEAR_LO) considering the computational efficiency to facilitate large ensembles needed for the ensemble data assimilation and the retrospective forecasts covering about 60 years (1961–2019).

In SPEAR_LO, the horizontal resolution in the ocean and sea ice components is about 1° with meridional refinements to 1/3° in the Tropics. The hybrid vertical coordinate in the ocean model has 75 layers with layer thickness as fine as 2 m near the surface, including 30 layers in the top 100 m. Overall, the physics used in the ocean and sea ice components is similar to OM4 (Adcroft et al., 2019). Since SPEAR uses a 1° horizontal grid spacing, versus the 0.25° grid spacing used in OM4, several related physical schemes including horizontal viscosity, meso-scale eddy and sub-meso-scale eddy parameterizations, are modified for reflecting the horizontal resolution difference (Delworth et al., 2020).

SPEAR_LO uses an atmosphere/land resolution of ∼100 km, and it has 33 vertical levels with model top at 1 hPa. The physical configurations in the atmospheric and land components are identical to those of recently developed GFDL AM4-LM4 model (Zhao et al., 2018a, 2018b), with one exception of using the retuned near-infrared isotropic reflectance parameter for cold snow over glacial surfaces (Delworth et al., 2020).

2.2. Initialization Approaches

We developed two approaches for initializing GFDL's SPEAR decadal prediction system. In the first initialization approach (called A1), the 6-hourly atmospheric winds and temperature were restored toward atmospheric data from the 55-year Japanese Reanalysis (JRA-55, Kobayashi et al., 2015). In addition, sea surface temperature (SST) was restored toward the time-varying observations from NOAA Extended Reconstructed Sea Surface Temperature version 5 (ERSSTv5) data (Huang et al., 2017). The atmospheric restoring is performed for the three-dimensional temperature and winds on a 6-hourly time scale. The SST restoring is confined within 60°S to 60°N. The strength of the restoring is tapered linearly from 1.0 at 55°N(S) to 0.0 at 60°N(S) and is equivalent to a 10-day restoring time scale for a 50-m mixed layer depth, corresponding to an effective restoring rate of about 240 W/(m^2 K). The A1 technique utilizes atmospheric temperature/winds and SST in the restoring process, and can be classified as a so-called "full-field initialization" method (Meehl et al., 2014).

In the second approach (called A2), observations of surface pressure are assimilated in the atmospheric component of the coupled SPEAR model using an ensemble data assimilation (DA) algorithm. The SST restoring in A2 is the same as that in A1, except that the target SST values are the ERSST anomalies plus the SPEAR model climatology. The different choice of SST climatology used in A1 and A2 will be discussed below. In contrast to the 3-dimensional atmospheric observations used in A1, A2 only uses the atmosphere surface pressure observations. The primary goal of A2 here is for the decadal predictability study, and specifically for testing the hypothesis that the surface pressure data assimilation along with SST restoring would provide adequate information for reconstructing the observed air-sea heat and momentum fluxes in the coupled model.

In A2, the ensemble DA solver used for the surface pressure assimilation is the same two-step Ensemble Adjustment Kalman Filter (EAKF) as used in GFDL's CM2.1 ECDA (Anderson, 2003; S. Zhang et al., 2007) and SPEAR-ODA (Lu et al., 2020). The model surface pressure, 3-dimensional winds, temperature and

moisture increments are solved via the ensemble covariance with surface pressure observation and added to the model tendency at each assimilation step. Following the surface pressure assimilation prototype in 20CRv2c and 20CRv3, we apply the covariance inflation and localization in the EAKF. The covariance inflation is the relaxation-to-prior-spread method (Whitaker & Hamill, 2012), which adaptively considers the observation network density in the inflation. We used the same value of covariance inflation parameter as in 20CRv3 (Slivinski et al., 2019). For covariance localization, we apply the same algorithm in both horizontal and vertical directions using the weighting function Equation 4.10 in Gaspari and Cohn (1999).

A paramount step when assimilating surface pressure observations over land is to make appropriate adjustments to these observations to account for the difference between the station elevation and the model orography at the observation location (Compo et al., 2011; Ingleby, 1995). Following the same procedure as in 20CRv2, the adjustment converting the observed values from the station elevation to the model orography was done using the pressure reduction algorithm of Benjamin and Miller (1990). After the surface pressure adjustment step, we also apply similar quality control steps as in 20CRv3 considering the plausible range of observations and the distance between first guess and observations. We use the latest version of the surface pressure observational data set, the International Surface Pressure Databank (ISPD) version 4.7 (Compo et al., 2019; Cram et al., 2015). The surface pressure data from MERRA2 (Gelaro et al., 2017) is used during 2017–2019, since the ISPD version 4.7 data is available up to 2016.

Considering the development feasibility and efficiency with finite computational resource, we use 24 ensemble members for the surface pressure data assimilation in SPEAR, instead of the very large ensemble size used in 20CR, e.g., 56 members in 20CRv2c (Compo et al., 2011) and 80 members in 20CRv3 (Slivinski et al., 2019). To compensate for the medium ensemble size used here, we use a relatively short horizonal localization length, specifically 1,500 km for surface pressure, winds and temperature, and 1,000 km for moisture. As a reference, the horizontal localization length varies with every observation and has 4,000 km maximum in 20CRv3 with 80 members (Slivinski et al., 2019). In the vertical, localization is set to 3.5 scale heights (units of $-\log\left(\frac{p}{ps}\right)$).

The atmospheric temperature increments in the surface pressure assimilation step are produced by the model, so the atmospheric temperature field is radiatively balanced with the model ocean state. Therefore, to maintain the radiative balance between atmosphere and ocean, we use in A2 ERSSTv5 anomalies plus the model climatology as the SST field for restoring. Note that the full-field surface pressure is assimilated in A2, thus A2 can be classified as a mixed full-anomaly field initialization.

2.3. Observational Datasets

For the atmospheric restoring in A1, the 6-hourly 3-dimensional instantaneous temperature and horizontal winds from the JRA-55 reanalysis covering 1958 to 2019 were retrieved from the Research Data Archive at the National Center for Atmospheric Research (Japanese Meteorological Agency, 2013; Kobayashi et al., 2015). For assessment of coupled reanalyses and decadal predictions, monthly NOAA ERSSTv5 SST data (Huang et al., 2017), monthly surface winds, 2-meter air temperature (T2m) and sea level pressure (SLP) from the NCEP/NCAR Reanalysis 1 (NNR) data (Kalnay et al., 1996), and the monthly precipitation data from the Global Precipitation Climatology Project (GPCP) Monthly Analysis Version 2.3 (Adler et al., 2018), were provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA (from their Web site at https://psl.noaa.gov/). Monthly land surface air temperature and precipitation from CRU TSv4 (current version 4.03, Harris et al., 2020) were provided via the Center for Environmental Data Analysis (CEDA), and also at the Climatic Research Unit (CRU) website.

Monthly subsurface ocean temperature and salinity data from the EN4 objective analysis (Good et al., 2013) were downloaded from the Met Office EN4 Web site (https://www.metoffice.gov.uk/hadobs/en4/). The version v2018.2 of RAPID AMOC data (Moat et al., 2020) from 2004 to 2018 was downloaded from http://www.rapid.ac.uk/rapidmoc/.

Table 2
Description of the Experiments of the Uninitialized Large Ensemble Forced Simulations, Decadal Coupled Reanalyses and Initialized Decadal Retrospective Forecasts

Name	Method	Ensembles	Simulation length
Uninitialized large ensemble forced simulation			
SPEAR-LES	SPEAR coupled simulations with CMIP6 historical (1851–2014) and future projected radiative forcing SSP5-8.5 (2015–2100) (Delworth et al., 2020)	30 members starting from 1 January 1851 of the 1850 control simulations	1851–2100
Decadal coupled initialization/reanalysis system (DCIS)			
DCIS-A1	3-d atmospheric restoring toward JRA reanalysis data; SST restoring toward ERSSTv4 data; the same radiative forcing as SPEAR-LES	10 members starting from 1 January 1958 of SPEAR-LES	1958–2019
DCIS-A2	Surface pressure data assimilation; SST restoring toward ESSRTv4 anomalies plus SPEAR model climatology; the same radiative forcing as SPEAR-LES	24 members starting from 1 January 1940 of SPEAR-LES	1940–2019
Decadal retrospective forecasts (DRF)			
DRF-A1	Coupled integration with the same radiative forcing as SPEAR-LES	10 members from every 1 January 1960-2019 of DCIS-A1	10 years for each forecast
DRF-A2	Coupled integration with the same radiative forcing as SPEAR-LES	10 members from every 1 January 1960–2019 of DCIS-A2	10 years for each forecast

3. Coupled Initialization, Retrospective Forecasts and Verification Methods

3.1. Coupled Decadal Initialization System

With the approach A1, we conducted a 10-member coupled reanalysis from 1958 to 2019 (note that JRA-55 data started from 1958). The decadal coupled initialization system using A1 is called DCIS-A1 hereafter. The 10-member initial conditions of DCIS-A1 on 1 January 1958 were taken from the historical large-ensemble simulations using SPEAR (Delworth et al., 2020, called SPEAR-LES hereafter). Beyond the restoring constraint in the atmospheric component and the SST in the ocean component, the ensemble coupled reanalysis includes the time-varying natural and anthropogenic radiative forcings developed in support of CMIP6 (Eyring et al., 2016). Historical forcing is applied for the time period 1958 to 2014, while projection forcing with the Shared Socioeconomic Pathway 5–8.5 (SSP5-8.5) scenario (Kriegler et al., 2017; Riahi et al., 2017) is applied after 2014. Note that the radiative forcing in the coupled initialization is identical to those in the SPEAR large ensemble. Volcanic aerosol forcing is included in both the reanalysis and the predictions. This volcanic forcing would not have been known at the start of each reforecast, so including future volcanic forcing likely overestimates the prediction skill as shown by Timmreck et al. (2016). However, since the volcanic forcing is also included in the historical simulations without initialization, we can still estimate the impact of initialization by comparing the initialized and uninitialized simulations, both of which contain the same volcanic forcing.

For DCIS-A2 we have the flexibility of choosing the initialization period given the very long observational record of measuring the surface pressure since the nineteenth century. The main focus of this study is to produce the initial conditions spanning 1960-present for the decadal prediction, so we generated the DCIS-A2 from 1940. The 24-member initial conditions of DCIS-A2 on 1 January 1940 were taken from the SPEAR-LES. The same historical and projection forcing is applied to DCIS-A2 as those in DCIS-A1. A summary of the two decadal coupled reanalyses is provided in Table 2.

3.2. Retrospective Forecasts

We conducted two sets of decadal retrospective forecasts using the SPEAR_LO model initialized from DCIS-A1 and DCIS-A2, called DRF-A1 and DRF-A2 respectively. For each set of DRF-A1 and DRF-A2, a 10-member ensemble of retrospective forecasts was initialized on 1 January every year from 1961 to 2019 and integrated for 10 years with the temporally varying anthropogenic and natural forcing as in the SPEAR-LES. Note that the 10-member ensemble initial conditions in DRF-A2 are a subset of the 24 ensemble members in the DCIS-A2 ensemble data assimilation. The optimal choice of the ensemble size for the prediction

would be the full ensemble used in the ensemble data assimilation, so this subset ensemble size of 10 is the compromise between the computational cost and development efficiency. A summary of the two sets of decadal retrospective forecasts is provided in Table 2.

A widely studied set of 10-member decadal retrospective forecasts using GFDL CM2.1 (DRF-CM2.1) initialized from GFDL's ECDA (Chang et al., 2013; S. Zhang et al., 2007) were used for comparison with the new decadal prediction system.

3.3. Verification Methods

One important procedure in the forecast skill assessment is a forecast bias adjustment obtained by converting the raw DRF fields into anomaly fields relative to the forecast climatology at each lead time. Specifically, the forecast anomalies for each variable were obtained by subtracting out the lead-time-dependent climatology from forecasts, which effectively removes the climate drift assuming that the climate drift is systematic as a function of forecast lead time. This is a widely used approach in the decadal skill verification (e.g., Boer et al., 2016; Kim et al., 2012; Meehl et al., 2014; Smith, Eade, & Pohlmann, 2013; S. G. Yeager et al., 2018). The same procedure is applied to the observations that exactly match forecasts at each lead time, and the resulting anomalies are compared for verification. Due to the availability of observations up to 2019, only the 50 DRFs initialized during 1961–2010 are used for the skill verification.

The retrospective forecast skill verification in this study follows the metrics outlined in Goddard et al. (2013). The skill metrics include the anomaly correlation coefficient (ACC) and the mean square error (MSE). For a given ensemble mean forecast variable f, the ACC is defined as:

$$\mathrm{ACC}_{\tau} = \frac{\sqrt{\sum_{i=1}^{N}\left(f_{i,\tau}-\overline{f}_{\tau}\right)\left(o_i-\overline{o}\right)}}{\sqrt{\sum_{i=1}^{N}\left(f_{i,\tau}-\overline{f}_{\tau}\right)^2}\sqrt{\sum_{i=1}^{N}\left(o_i-\overline{o}\right)^2}} \tag{1}$$

where τ is theforecast lead time, N is the total number of executed DRFs, $f_{i,\tau}$ represents the ensemble-mean forecast from start year i at lead time τ, and $\overline{f}_{\tau}$ represents the average over N cases at lead time τ. Our main focus here is on predicting the decadal-time-scale low-frequency variability, thus we consider the 5-year or 8-year averaged predictions for isolating the decadal climate signal following S. G. Yeager et al. (2018) and Smith et al. (2019).

The nonparametric block bootstrap approach (Goddard et al., 2013; S. G. Yeager et al., 2018) was used for the significance testing of both ACC and the difference of ACC between DRF and uninitialized SPEAR-LES. For the bootstrap, an additional 2000 resampled hindcasts (or pairs of hindcasts for assessing skill difference) were created with replacement over the time and ensemble dimensions in blocks of five consecutive years for taking autocorrelation into account. The probability distribution from the 2000 times of resampling was used to provide the significance level of the hypothesis that skill (or skill difference) is zero. The block bootstrap is applied at each grid point for the statistical significance test of a given gridded variable.

For a large-scale spatially averaged climate index (e.g., the North Atlantic Subpolar SST and the Southern Ocean SST) with autocorrelation time scales longer than 5 years, we use a t-test with an adjusted degree of freedom for computing the significance testing of ACC. This adjusted degree of freedom for taking autocorrelation into account was widely used for the decadal skill assessment of large-scale climate indices (Chang & Wang, 2020; Msadek et al., 2014; Yang et al., 2013). The effective degree of freedom N* is estimated following Bretherton et al. (1999):

$$N^{*} = \frac{N}{1+2\sum_{j=1}^{N-1}\left(1-\frac{j}{N}\right)\rho_{xx}(j)\rho_{yy}(j)} \tag{2}$$

where N is the number of sample pairs, and $\rho_{xx}(j)$ and $\rho_{yy}(j)$ are the sample autocorrelation of x and y at lag j.

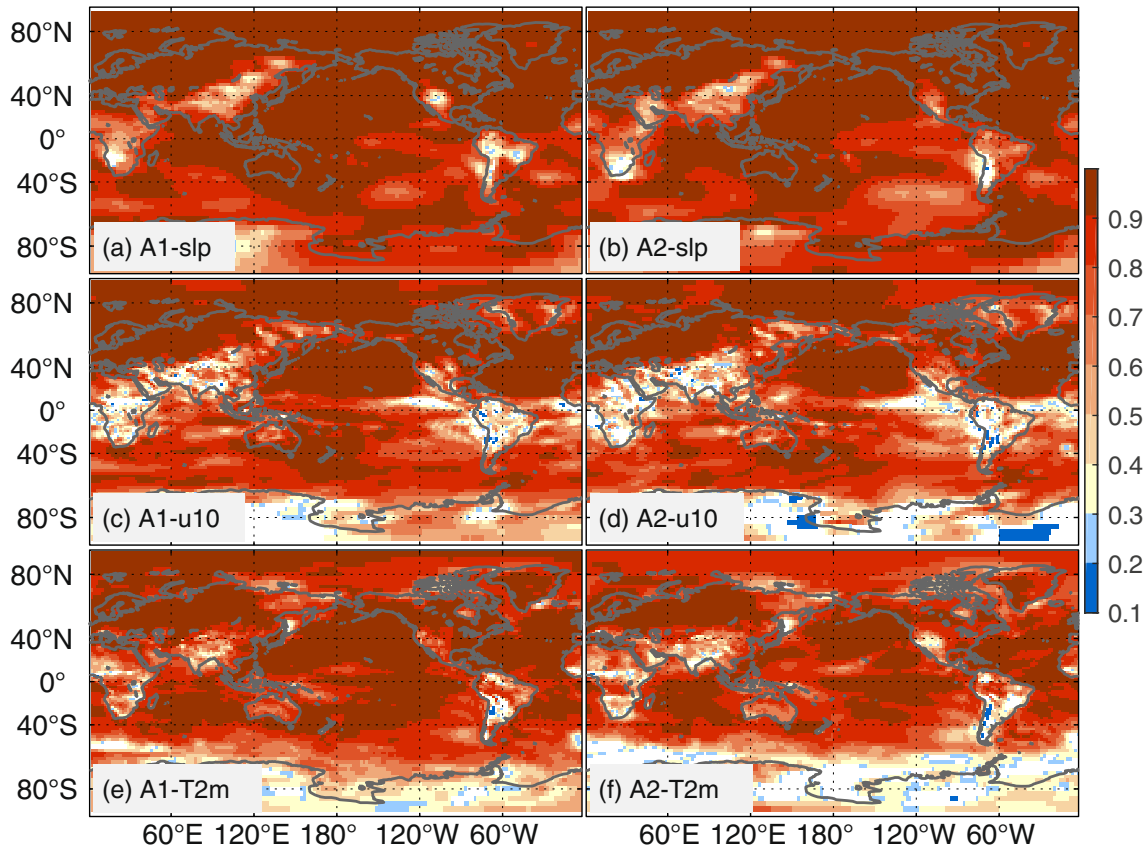

Figure 1. The correlation coefficients at each grid point between the coupled reanalysis and NNR in the December–March season during 1960–2019 for the sea level pressure from (a) DCIS-A1 and (b) DCIS-A2; for the 10-m zonal wind from (c) DCIS-A1 and (d) DCIS-A2; for the 2-m air temperature from (e) DCIS-A1 and (f) DCIS-A2. The correlation coefficients with the statistical p-value larger than 0.05 are not shown.

4. Decadal Variability in the Coupled Reanalysis

4.1. Air-Sea Interface Variability Constrained by Observations

First, we examine major climate indices representing dominant atmospheric surface circulation variability in the coupled initialization in the North Atlantic, North Pacific and Southern Ocean (SO) respectively since there exists substantial decadal to multidecadal climate variability and predictability in the three ocean basins in GFDL's models (Delworth et al., 2020; Yang et al., 2013; L. Zhang et al., 2017; Zhang, Delworth, et al., 2019) and in observations (Newman, 2013). The three climate indices are the NAO (Hurrell & Deser, 2009), the North Pacific Index (NPI) (Trenberth & Hurrell, 1994), and the Southern Annular Mode (SAM) (Marshall, 2003). Both DCIS-A1 and DCIS-A2 can accurately reproduce the observed NAO and NPI in the NNR data with correlation coefficients larger than 0.99 (Figure S1 in Supporting Information S1). The SAM indices estimated from the DCIS-A1 and DCIS-A2 agree very well with that from the NNR observational data with correlation coefficient 0.94 and 0.97 respectively. The ensemble spreads of SAM from DCIS-A2 are relatively large during 1960–1975, indicating that the available surface pressure data are sparse in the Southern Hemisphere during that period (Compo et al., 2011). The very high agreement ($r > 0.95$) of estimating the major climate indices in the three ocean basins among the DCIS-A1, DCIS-A2 and NNR suggests that the surface-pressure-based assimilation is accurate enough for representing the observed dominant surface climate variability in the coupled model.

To further illustrate the climate variability at the air-sea interface when SPEAR is constrained by observations, we show in Figure 1 the correlation coefficients at each grid point between the coupled reanalyses and NNR during 1960–2019 for the December-March SLP, 10-m zonal wind (U10) and 2-m air temperature (T2m) respectively. The interannual variability of SLP estimated from DCIS-A1 and DCIS-A2 agrees very

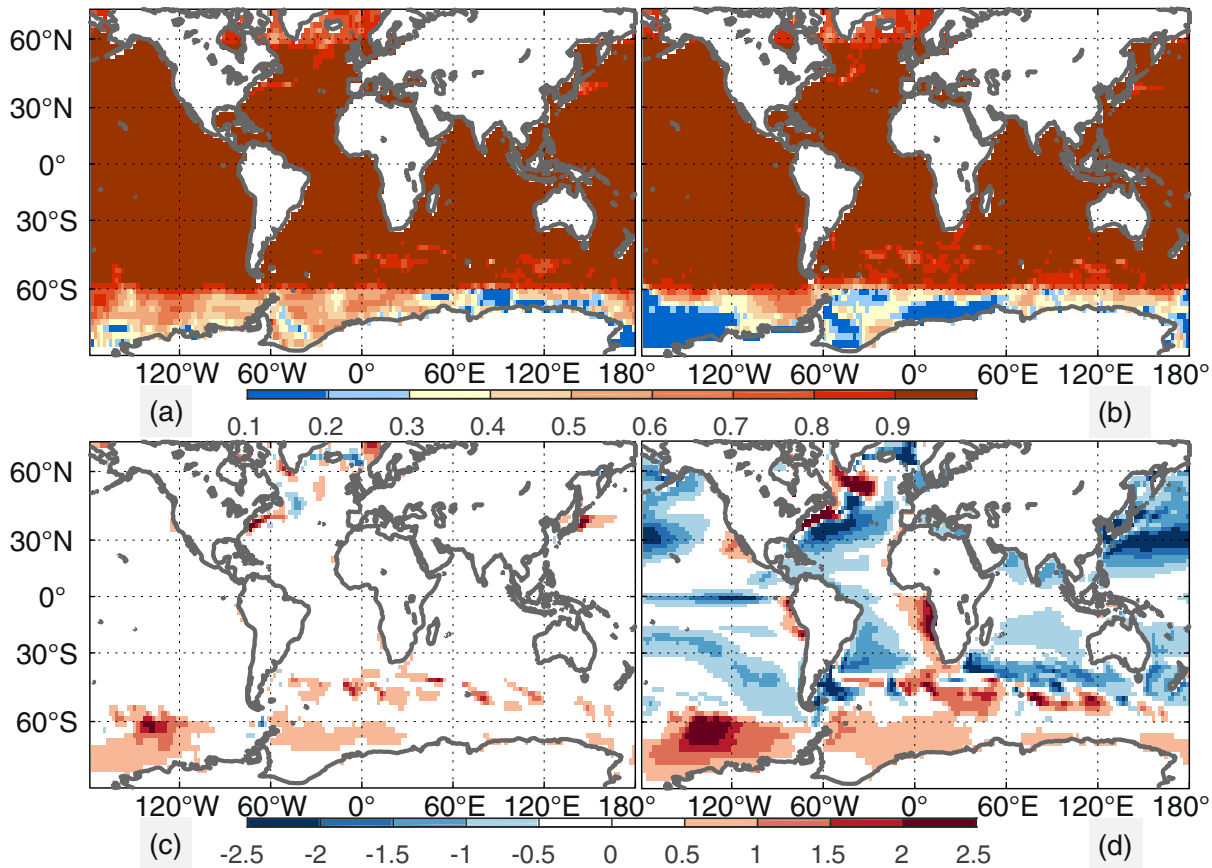

Figure 2. The correlation coefficients of the annual mean sea surface temperature (SST) at each grid point between the analysis and ERSSTv5 data for (a) DCIS-A1 and (b) DCIS-A2 during 1960–2019. The mean bias of SST between the model and ERSSTv5 during 1960–2019 for (c) DCIS-A1 and (d) DCIS-A2. Units are in °C for (c–d).

well with that from NNR over the majority of global land and ocean with correlation coefficients exceeding 0.9, except in the regions with high terrain and part of the SO. Consistent with the relatively lower agreement of the SAM than the NAO and NPI between DCIS-A1/A2 and NNR, the correlation coefficients of SLP in the SO is about 0.1 lower than that in the North Pacific and North Atlantic. Consequently, the surface zonal winds in DCIS-A1/A2 are in very high agreement with those in NNR with correlation coefficients exceeding 0.9 in the North Pacific and North Atlantic, and in high agreement with correlation coefficients exceeding 0.8 over most of the SO. For the surface air temperature, DCIS-A1 is generally in higher agreement with NNR than DCIS-A2, presumably because the 3-dimensional air temperature data from JRA-55 were directly nudged in DCIS-A1. In summary, this global point-to-point correlation analysis provides additional evidence that the surface-pressure-based assimilation in DCIS-A2 is accurate enough for representing the observed climate variability at the air-sea interface in the North Pacific and North Atlantic, and with a comparable performance level as in DCIS-A1, which incorporates the 3-dimensional JRA-55 fields. However, DCIS-A2 has low agreement with NNR in representing the observed T2m variability in the SO. Here we focus on the December-March season, which is the period of maximum deep-water formation in the North Atlantic. Note that the same comparisons have been made for other seasons, and they show similar global patterns of agreement as the boreal winter season (Figures not shown).

The turbulent air-sea heat flux is determined by the surface winds and the temperature difference between SST and surface air temperature through a bulk parameterization in the coupled model, thus we also applied the SST restoring in both DCIS-A1 and DCIS-A2 for representing the SST variability. The correlation coefficients of the annual mean SST between DCIS-A1/A2 and ERSSTv5 data are shown in Figure 2. The SST variability in both DCIS-A1 and DCIS-A2 agrees very well with that in ERSSTv5 data within the region 60°S–60°N, where the SST restoring was applied. Interestingly, both DCIS-A1 and DCIS-A2 show

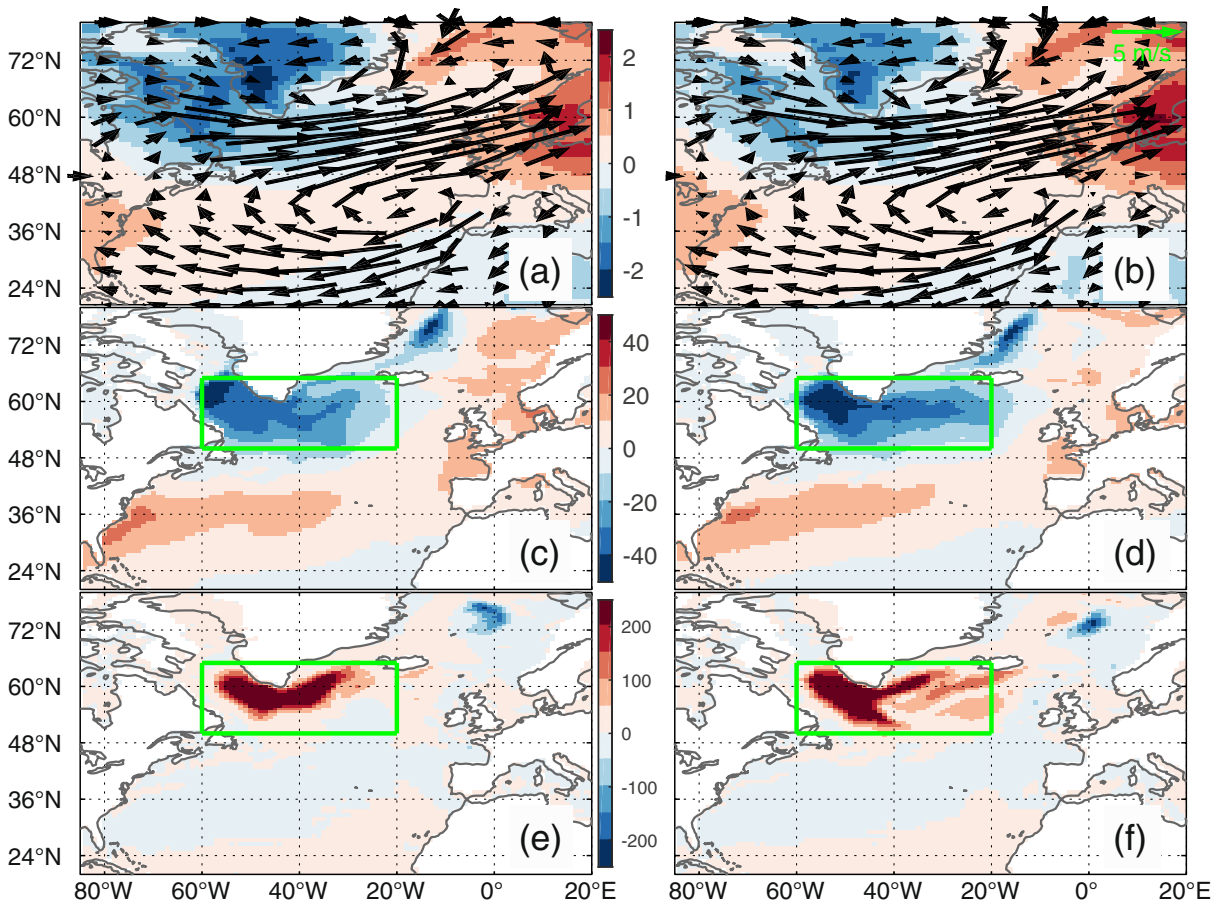

Figure 3. The regression pattern onto the normalized NAO index for the surface winds (vectors) and 2-m air temperature (shading, units in °C) during 1960–2019 from (a) DCIS-A1 and (b) DCIS-A2, for the surface heat flux (shading, units in W/m^2) from (c) DCIS-A1 and (d) DCIS-A2, for the mixed layer depth (shading, units in m) from (e) DCIS-A1 and (f) DCIS-A2.

significant correlation with ERSSTv5 in the polar flank of 60°N solely due to the atmospheric forcing. However, in the polar flank of 60°S, the performance of reproducing observed SST variability in DCIS-A2 is noticeably worse than that in DCIS-A1. This degradation in the far polar flank of the SO in DCIS-A2 is mainly due to the lack of surface pressure observations in the open ocean. This low correlation of SST with ERSSTv5 in DCIS-A2 is also consistent with the low correlation of T2m with NNR (Figure 1f). In contrast, the observed air temperature data used in DCIS-A1 can directly drive the observed SST variability through the air-sea interaction process in the SO (Figures 1e and 2a). Since DCIS-A2 uses anomaly initialization, the global pattern and amplitudes of SST bias in DCIS-A2 are very similar to that in SPEAR's control simulation (Figure 2, Delworth et al., 2020). In contrast, the SST bias is almost absent in DCIS-A1 due to the full initialization (Figures 2c and 2d).

4.2. AMOC Variability

The observed multi-decadal NAO variability can induce the multi-decadal AMOC variations through the NAO-related surface fluxes in the coupled model (Delworth & Zeng, 2016; Delworth et al., 2016, 2017). Here we examine the physical processes connecting NAO forcing and AMOC in the two reanalyses. Figure 3 shows the regression patterns onto the normalized NAO index for the surface winds, T2m, the surface heat flux and the mixed layer depth (MLD) during the winter (December–March) season (the period of maximum deep-water formation in the North Atlantic). The DCIS-A2 with surface pressure assimilation can reproduce the very similar NAO-related surface wind and air temperature patterns as DCIS-A1 (Figures 3a and 3b). Consequently, the NAO-related surface heat flux patterns estimated from DCIS-A1 and DCIS-A2

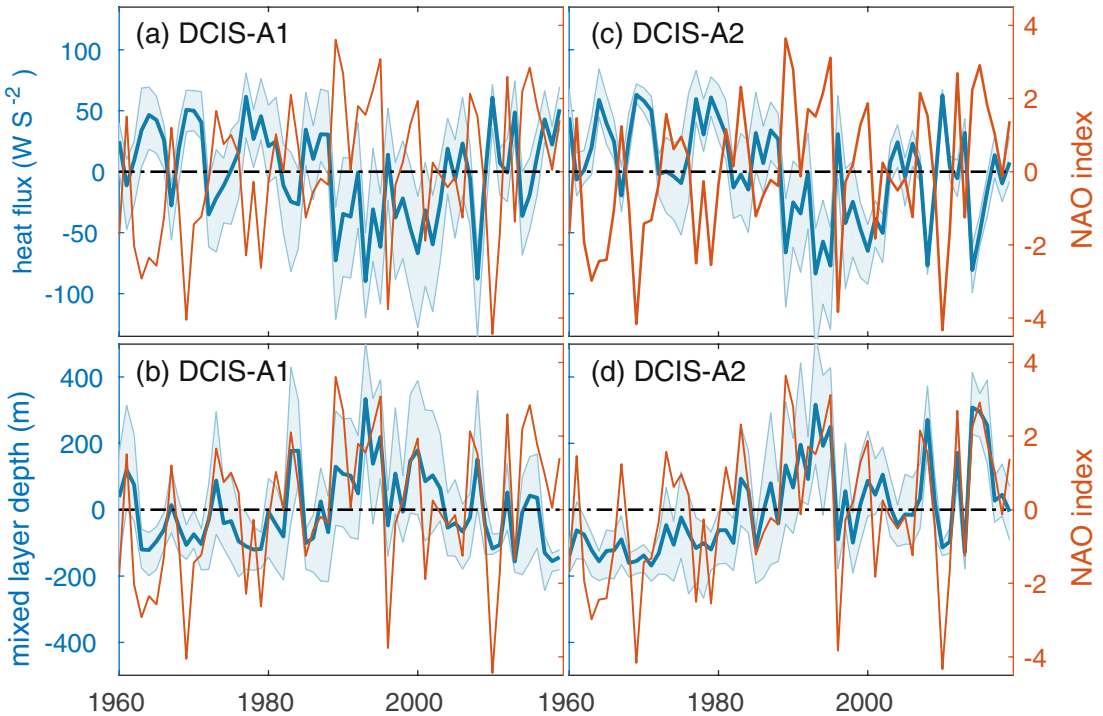

Figure 4. Time series of the ensemble estimates of the surface heat flux and mixed layer depth averaged over the North Atlantic Subpolar Gyre convection region (within 50°N–65°N/60°W–20°W marked with a green box in Figure 3) during the December–March season. The ensemble mean (blue line) and range (blue shading) of surface heat flux calculated from (a) 10 members in DCIS-A1, and (b) 24 members in DCIS-A2. The ensemble mean (blue line) and range (blue shading) of mixed layer depth calculated from (c) 10 members in DCIS-A1, and (d) 24 members in DCIS-A2. The North Atlantic Oscillation index from DCIS-A1 is plotted in (a and b), and from DCIS-A2 is plotted in (c and d).

agree remarkably well with the negative anomalies extending from the Labrador Sea through the central Subpolar Gyre to the Greenland Sea and the positive anomalies extending from the subtropical North Atlantic to the Norwegian Sea (Figures 3c and 3d). Note that the estimated NAO-related surface heat flux patterns are very similar to those from the ECMWF Interim Reanalysis data in terms of the amplitudes and spatial structure (e.g., Figure 1 in Delworth & Zeng, 2016), indicating that the observed dominant variability of the surface heat fluxes is very well reproduced in the two coupled initializations. As a result of the NAO-related flux forcing, the near surface cooling extracting heat from the ocean leads to increased MLDs over the Labrador Sea and the central Subpolar Gyre in both DCIS-A1 and DCIS-A2 (Figures 3e and 3f).

To further illustrate the relationship between the surface heat fluxes and the deep ocean convection over the North Atlantic Subpolar Gyre (NASPG), the time series of the anomalous surface heat flux and MLD estimated from DCIS-A1 and DCIS-A2 are shown in Figure 4. With apparent interannual variations, DCIS-A1 and DCIS-A2 agree well on the estimated surface heat flux with the multi-decadal variations (Figures 4a and 4c). As response to the surface heat flux forcing, the similar multi-decadal variations of the MLD are well reproduced in DCIS-A1 and DCIS-A2 (Figures 4b and 4d). Consistent with the NAO-related regression patterns shown in Figure 3, the NAO index is highly anti-correlated with the surface heat flux over the NASPG with correlation coefficient −0.64 (−0.71) from DCIS-A1 (DCIS-A2), and highly correlated with the NASPG MLD with correlation coefficient 0.60 (0.69) from DCIS-A1 (DCIS-A2). The diagnosed physical processes connecting the surface heat fluxes to the deep ocean convection associated with the NAO are consistent with previous studies (Danabasoglu et al., 2016; Delworth & Zeng, 2016; Delworth et al., 2017), and these processes are the key drivers of the decadal variability of AMOC from the surface flux forcing (Delworth et al., 2016). Note that the surface heat flux and MLD from DCIS-A1 show different behavior from DCIS-A2 during 2015–2019. In particular, during the 2015–2016 winters with strong NAO, the NAO-induced heat flux and MLD anomalies from DCIS-A1 are substantially smaller than those strong NAO winters during late 1980s and early 1990s. The weaker response to the NAO forcing during 2015–16 is related to the stronger mean ocean stratification relative to the earlier period in DCIS-A1 (Figure not shown). In contrast,

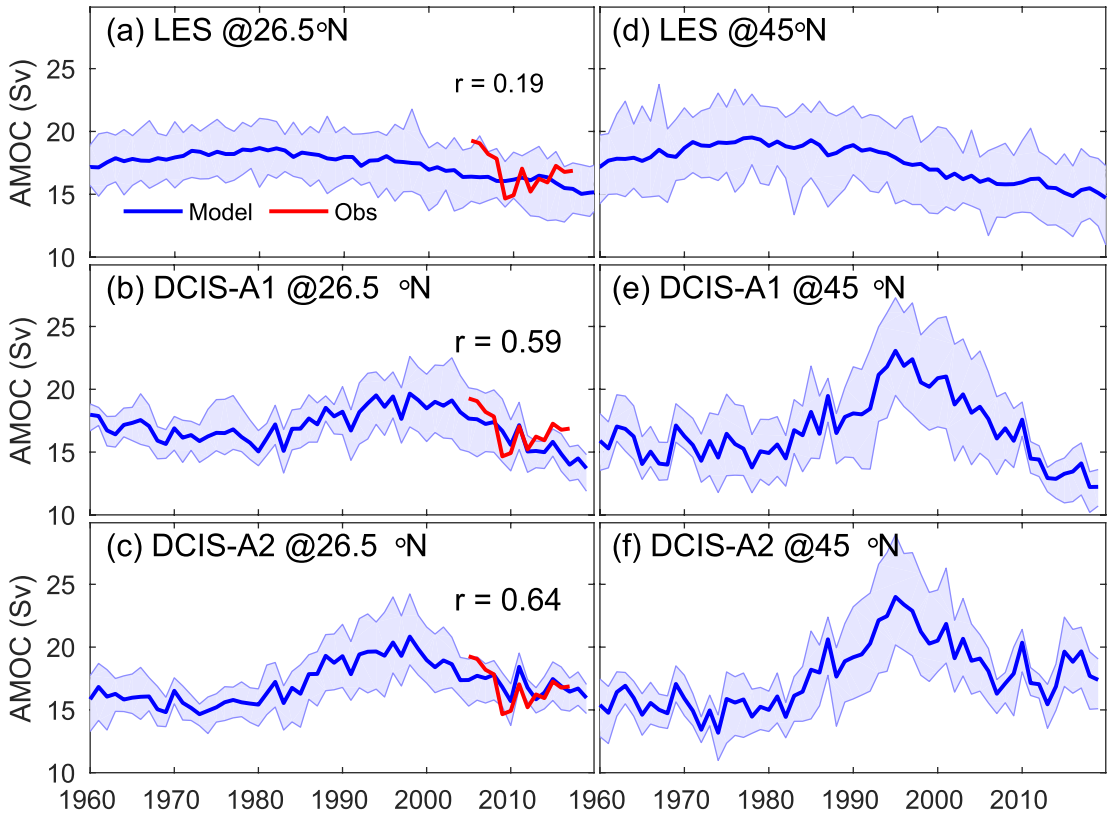

Figure 5. Time series of the ensemble estimates of annual mean Atlantic Meridional Overturning Circulation (AMOC) index during 1960–2019. The index is computed each year as the maximum value of the overturning stream function at a given latitude in the North Atlantic. The AMOC index at 26.5°N estimated from (a) Seamless System for Prediction and EArth System Research (SPEAR) large ensemble simulations, (b) DCIS-A1 and (c) DCIS-A2. The AMOC index at 45°N estimated from (d) SPEAR large ensemble simulations, (e) DCIS-A1 and (f) DCIS-A2. The thick blue lines are the ensemble mean values, and the light blue shading shows the range of values each year across the ensemble members. The red line in (a–c) shows the annual mean values from observations using the RAPID array at 26.5°N in the North Atlantic for 2005–2017. Units are in Sverdrups (Sv), where 1 Sv = 10^6 m^3 s^{-1}.

DCIS-A2 shows similar response of surface flux and MLD to the NAO forcing during the 2015–2016 winters as other historical strong NAO winters.

We show in Figure 5 the time series of the ensemble estimates of annual mean AMOC index at 26.5°N and 45°N during 1960–2019 from the SPEAR-LES, DCIS-A1 and DCIS-A2 respectively. The AMOC index is computed each year as the maximum value of the overturning stream function at a given latitude in the North Atlantic. In contrast to the general weak decline trend after about 1980 in SPEAR-LES, both DCIS-A1 and DCIS-A2 show a remarkable multi-decadal variation of AMOC at both 26.5°N and 45°N, e.g., a weakened state during the 1960s and 1970s, a strengthened state in the late 1980s and 1990s, and a weakened state after 2004. These multi-decadal variations are highly consistent with the multi-decadal variations of the surface heat flux and MLD, indicating that the physical balance between surface forcing and AMOC on decadal time scales is well maintained in the coupled reanalysis for decadal prediction. Both DCIS-A1 and DCIS-A1 show the spatially coherent basin-scale structure of the AMOC difference between the weakened and strengthened states, with the strong signal over the North Atlantic SPG (Figure S2 in Supporting Information S1). This multi-decadal AMOC variability estimated from DCIS-A1 and DCIS-A2 is consistent with that from CORE-II experiments (Danabasoglu et al., 2016) and the coupled model simulation with the imposed observed NAO-forcing (Delworth et al., 2016). Similar AMOC variations are also simulated in a different coupled model in which the 3-dimensional atmosphere and SST are constrained similarly to DCIS-A1 (see Figure 27 in Merryfield et al., 2013)

The RAPID array at 26.5°N since 2004 (Moat et al., 2020) provides in situ observations to assess the realism of the simulated AMOC. As shown in Delworth et al. (2020), the observed AMOC values are within the range of the simulated values from SPEAR-LES. However, the observed inter-annual variability during

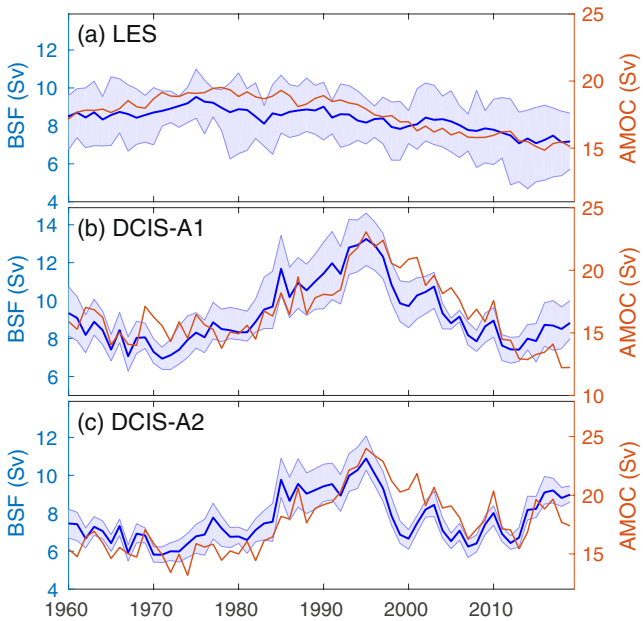

Figure 6. Time series of the ensemble estimates of annual mean barotropic stream function (BSF) index over the North Atlantic Subpolar Gyre (NASPG) during 1960–2019. The index is computed each year as the absolute value of the area mean BSF for the NASPG region defined by 15°–60°W and 48°–60°N. The BSF index estimated from (a) SPEAR LES, (b) DCIS-A1 and (c) DCIS-A2. The thick blue lines are the ensemble mean values, and the light blue shading shows the range of values each year across the ensemble members. The red lines are the ensemble mean values of Atlantic Meridional Overturning Circulation at 45°N.

2005–2017 was not well captured with a correlation coefficient of 0.19 by the SPEAR-LES (Figure 5a). In contrast, the DCIS-A1 and DCIS-A2 show significant correlation with the observations with correlation coefficients of 0.59 and 0.64 respectively. Note that the AMOC from DCIS-A1 shows different behavior from DCIS-A2 during 2015–2019 due to the weaker amplitudes of anomalous surface heat flux and MLD during the 2015/2016 NAO winters from DCIS-A1 than DCIS-A2 (Figure 4). Consequently, the AMOC indices during these two years are weaker from DCIS-A1 than those from DCIS-A2 (Figure 5). The strong deep convection over the NASPG during 2015–2016 from DCIS-A2 is supported by the observational analysis (Yashayaev & Loder, 2017).

Simulated AMOC in GFDL's CM2.1-ECDA (Figure S3 in Supporting Information S1), similar to other ocean reanalysis products with direct subsurface ocean data assimilation, does not show distinct multi-decadal variations during 1960–2014 (Karspeck et al., 2017), and the mean AMOC amplitudes from ECDA were substantially reduced relative to those from the historical CM2.1 simulation (Chang et al., 2013). In this new system, SPEAR-DCIS shows a significant improvement in simulating AMOC climatology and its multidecadal variability as compared to CM2.1-ECDA.

4.3. Circulation Variability in the North Atlantic Subpolar Gyre

The SPG circulation is another key aspect of the decadal variability in the North Atlantic (Danabasoglu et al., 2016; Eden & Willebrand, 2001), and covaries with AMOC on decadal time scales (Yeager, 2015), thus it is crucial to assess the variability of the gyre circulation in the decadal initialization system. Here we use the barotropic stream function (BSF) as a metric of representing the gyre circulation. The BSF is defined as the depth-integrated volume transport. We performed the regression analysis of the wind stress/wind curl and BSF with the normalized NAO index from DCIS-A1 and DCIS-A2 respectively (Figure S4 in Supporting Information S1). The simultaneous and delayed response of circulation to the NAO wind forcing from DCIS-A1 and A2 is consistent with previous study based on the objective data analysis (Eden & Willebrand, 2001).

Figure 6 shows the time series of the ensemble estimates of annual mean BSF index over the NASPG during 1960–2019 for the SPEAR-LES, DCIS-A1 and DCIS-A2 respectively. Note that the BSF index is defined as the absolute values of BSF. In contrast to a relatively flat response curve to the radiative forcing in SPEAR-LES (Figure 6a), both DCIS-A1 and DCIS-A2 have very similar distinct multi-decadal variations (Figures 6b and 6c). The high correlation coefficients between the BSF index and AMOC with $r = 0.77$ (0.72) from DCIS-A1 (DCIS-A2) indicate the largely in-phase strengthening and weakening of AMOC and SPG circulation. We note that the BSF index leads the AMOC index by 2–3 years with a maximum correlation coefficient of about 0.80 and 0.73 from DCIS-A1 and DCIS-A2 respectively. The multidecadal covariations of AMOC and BSF anomalies from DCIS-A1 and A2 are consistent with those from CORE FOSI simulations (Danabasoglu et al., 2016; Yeager, 2015).

4.4. Subsurface Ocean Heat Content Variability

The accurate initialization of the upper ocean heat content (UOHC) is crucial for the decadal climate prediction, so we examine the decadal variability of the UOHC in this subsection. The top 300-m depth-averaged ocean temperature (T300) is used to represent the UOHC variability. We show in Figure 7 the correlation coefficients of T300 at each grid point between DCIS and EN4 data during 1960–2019. Despite the fact there is no data assimilation of the subsurface ocean observations, the UOTC in the DCIS-A1 and DCIS-A2 shows similar variability as in the EN4 data over the equatorial tropical Pacific, the North Pacific and the North Atlantic with correlation coefficients largely exceeding 0.6. In the southern hemisphere, the spatially

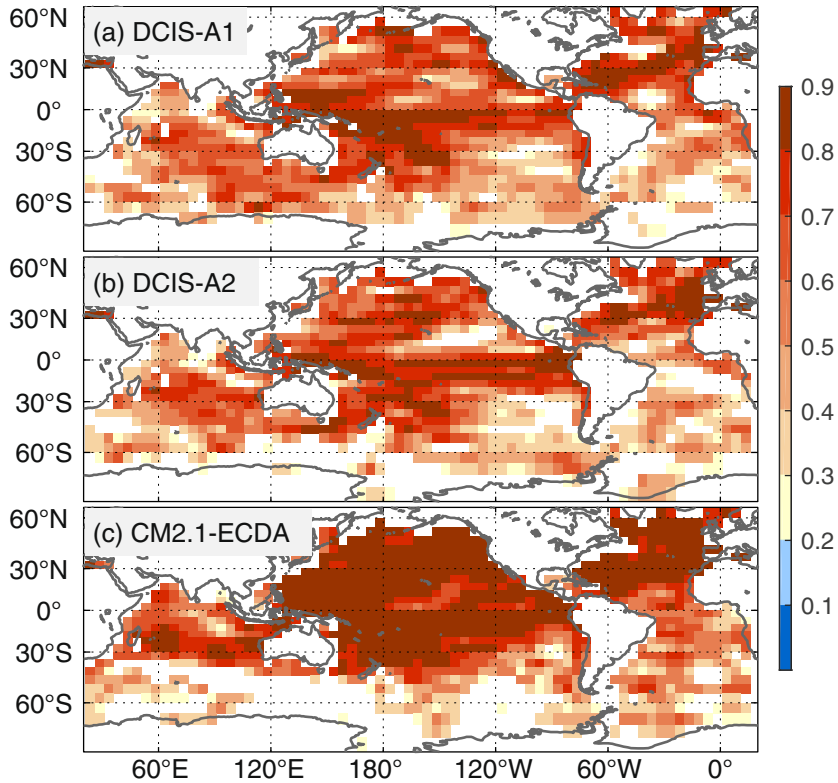

Figure 7. The correlation coefficients of the top 300m depth-averaged ocean temperature at each grid point between the coupled initialization system and EN4 data during 1960–2019 for (a) DCIS-A1, (b) DCIS-A2 and (c) CM2.1-ECDA. The correlation coefficients with the statistical p-value larger than 0.05 are not shown.

coherent region with high correlation with EN4 exists in the southern Indian Ocean and the western South Pacific. Note that CM2.1-ECDA shows consistently higher agreement with EN4 than DCIS-A1/A2 except in the SO due to its direct subsurface ocean data assimilation (Figure 7c). The relatively lower agreement between DCIS and EN4 in the SO in comparison to other ocean basins is mainly due to the sparse observations in both atmospheric and oceanic observation network (Compo et al., 2011; Newman et al., 2019). Due to the observational network changes, the agreement between DCIS and EN4 shows substantial multidecadal variations with a general high agreement during 2000–2019 for which atmospheric and oceanic observations are relatively rich and a general low agreement during 1960–1979 for which observations are relatively sparse (Figure S5 in Supporting Information S1).

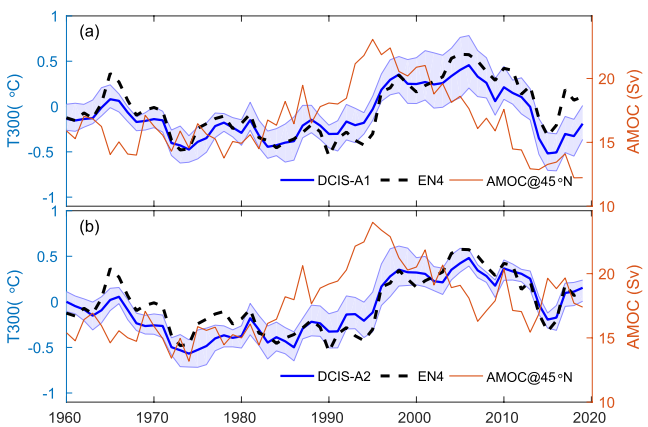

Figure 8. Time series of the EN4 data (black dashed) and ensemble estimates from (a) DCIS-A1 and (b) DCIS-A2 for the top 300 m depth-averaged ocean temperature (T300) anomalies and the Atlantic Meridional Overturning Circulation (AMOC) index at 45°N during 1960–2019. The thick blue lines are the ensemble mean values of T300, and the light blue shading shows the range of values each year across the ensemble members. The red lines are the ensemble mean values of AMOC at 45°N.

The estimated T300 anomalies over the NASPG from DCIS-A1 and A2 agree remarkably well in reproducing the observed multidecadal variations during 1960–2019 (Figure 8). The correlation coefficient of the NASPG T300 anomalies between EN4 and model is 0.87 (0.89) for DCIS-A1 (DCIS-A2). There exists a robust lead-lag relationship between AMOC and the NASPG UOHC on the multidecadal time scale in both DCIS-A1 and DCIS-A2. We also conducted a lag correlation analysis of the time series of AMOC and the NASPG UOHC (figures not shown), with AMOC leading the NASPG UOHC by about 5–10 years. Note that the NASPG T300 from DCIS-A1 is relatively colder compared to DCIS-A2 since 2010, consistent with the relatively weakened AMOC from DCIS-A1 in comparison with DCIS-A2. This physical lead-lag relationship between

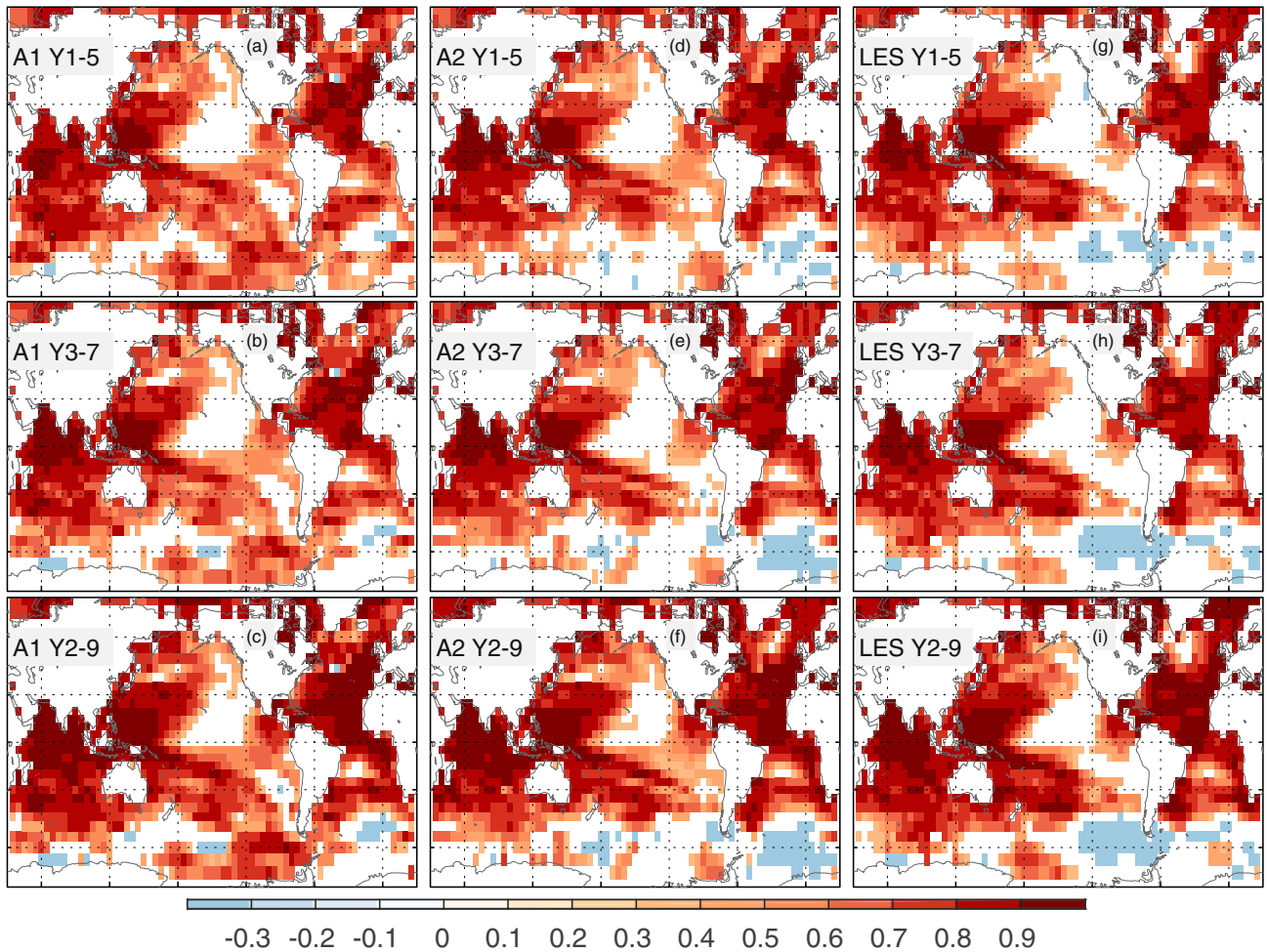

Figure 9. The anomaly correlation coefficient (ACC) of sea surface temperature (a–c) between DRF-A1 and ERSSTv5 observations, (d–f) between DRF-A2 and ERSSTv5, (g–i) between SPEAR-LES and ERSSTv5 for lead times of 1–5, 3–7 and 2–9 years, respectively. All fields were mapped onto a 5° × 5° grid prior to ACC calculation. ACC values with the statistical p-value larger than 0.05 are not shown.

AMOC and UOHC in the NASPG is consistent with the previous studies (Danabasoglu et al., 2016; Delworth et al., 2016, 2017; Robson et al., 2012; S. Yeager et al., 2012; Zhang & Zhang, 2015; Zhang, Sutton, et al., 2019), indicating that both DCIS versions represent the interrelated dynamics of the AMOC and NASPG which are important for decadal prediction. These results also indicate that the surface pressure data assimilation is sufficient for providing the air-sea fluxes driving the multidecadal AMOC and NASPG UOTC variations with the physical lead-lag relationship. Note that the lead-lag relationship between AMOC and OHC was not well represented in GFDL's previous decadal initialization system CM2.1-ECDA, although the NASPG UOHC in CM2.1-ECDA shows higher agreement with EN4 ($r = 0.94$) than that in DCIS-A1/A2 (Figure S3 in Supporting Information S1).

5. Assessment of Decadal Retrospective Forecast

5.1. Decadal Prediction Skill Assessment of SST

We first assess the skill of predicting the SST using the anomaly correlation coefficient (ACC) as shown in Figure 9. In general, both DRF-A1 and A2 show high predictive skill (ACC > 0.6) of SST over the global ocean except the north-east Pacific and parts of the SO (Figures 9a–9f). The uninitialized LES exhibits a very similar global distribution of ACC with notable degrading of skill in the north-east Pacific, the western part of the NASPG and the Drake Passage region in the SO (Figures 9g–9i). Note that the ACC of SST from

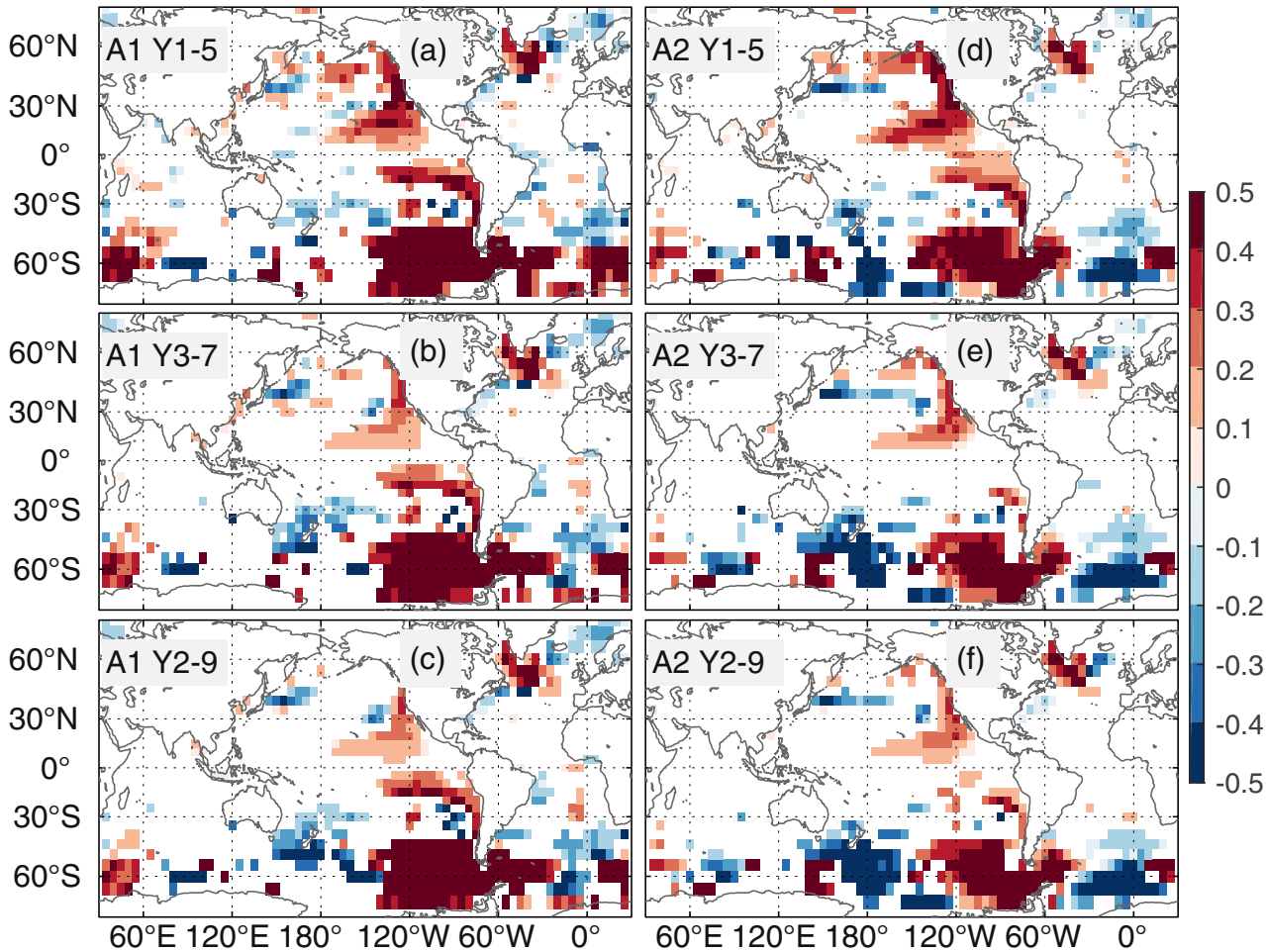

Figure 10. The anomaly correlation coefficient (ACC) skill score difference (a–c) between DRF-A1 and SPEAR-LES and (d–f) between DFR-A2 and SPEAR-LES for lead times of 1–5, 3–7, 2–9 years, respectively. Values of ACC difference with the statistical p-value larger than 0.05 are not shown.

DRF-A2 is coherently lower across different lead years over the Bellinghausen Sea, the Drake Passage region and the Weddell Sea in the SO than that from DRF-A1. This relatively lower skill in the SO from DRF-A2 is largely due to the relatively lower accuracy of the regional initial conditions in the DRF-A2 (Figure 2). Overall, the global pattern and skill level of ACC for SST from DRF-A1 and A2 are very similar to those estimated from the large-ensemble multi-model decadal predictions (Smith et al., 2019).

To further assess the impact of initialization on the prediction skill, we show in Figure 10 the differences of ACC (ΔACC) between the DRF and uninitialized LES. In the North Atlantic, both DRF-A1 and A2 show a substantial increase of skill with ΔACC over 0.4 in the western part of NASPG. The skill increase is coherent with the lead time and the location of the skill increase is coincident with the deep convection center associated with NAO (Figure 3), suggesting that the initialization of the multi-decadal variability of deep-water formation in the NASPG and AMOC plays an important role in this significant decadal skill improvement. In the North Pacific, DRF-A2 shows a significant increase of skill in a coastal zone extending from the Gulf of Alaska along the west coast of North America for all three lead times (Figures 10d–10f), with skill improvement in DRF-A1 that is similar although less evident in the Gulf of Alaska across lead times (Figures 10a–10c). In the SO, DRF-A1 shows a remarkable increase of skill with ΔACC > 0.4 in a band extending westward and across the Drake Passage, the Weddell Sea and Lazarev Sea, while DRF-A2 only shows significant skill increase around the Drake Passage region. Overall, the significant skill improvement due to initialization in the NASPG and SO in our new system is consistent with previous decadal prediction studies, including predictions initialized with subsurface ocean data assimilation (Smith et al., 2019; S. G.

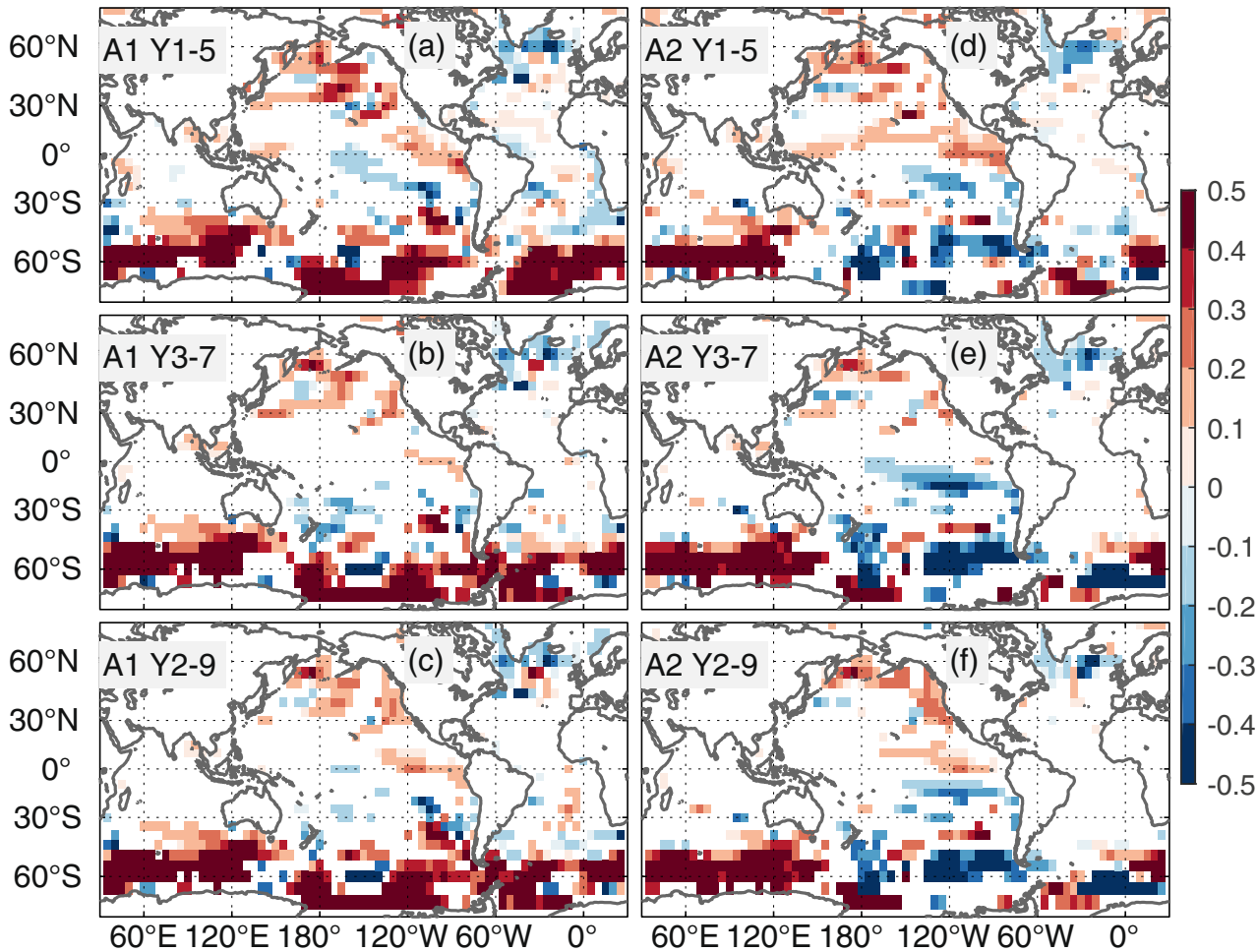

Figure 11. The anomaly correlation coefficient (ACC) skill score difference (a–c) between DRF-A1 and DRF-CM2.1 and (d–f) between DRF-A2 and DRF-CM2.1 for lead times of 1–5, 3–7, 2–9 years, respectively. Values of ACC difference with the statistical *p*-value larger than 0.05 are not shown.

Yeager et al., 2018; L. Zhang et al., 2017), indicating that this new system captures the essential elements of decadal initialization even without the direct subsurface ocean data assimilation.

To compare the prediction skill with GFDL's old DRF-CM2.1 that has directly assimilated the subsurface oceanic observations, we show in Figure 11 ΔACC between the DRF-A1/A2 and DRF-CM2.1. Particularly striking is the substantial increase of skill (ΔACC > 0.4) from DRF-A1 relative to DRF-CM2.1 over almost the entire SO except in the Australian sector of SO (Figures 11a–11c). DRF-A2 shows a significant increase of skill relative to DRF-CM2.1 in a band across the Weddell Gyre to the Indian ocean sector of the SO, but without skill improvement over the Ross Sea Gyre and Amundsen-Bellingshausen Seas as shown from DRF-A1. In the North Atlantic, DRF-A1 shows a skill degrading (ΔACC about −0.1) only in a narrow latitudinal band just at 60°N, while DRF-A2 displays a significant skill degrading only confined in the Iceland basin. There also exist sporadic regions with significant skill improvement in the North Pacific subpolar gyre, eastern North Pacific and the eastern tropical Pacific from DRF-A1 and A2.

To highlight the outperformance of DRF-A1 over DRF-CM2.1 in predicting the SO SST anomalies, we show in Figure 12 the time series of the SO (50°S–70°S/0°–360°) SST anomalies from ERSST, SPEAR-LES, DRF-A1/A2 and DRF-CM2.1. The initialized forecasts from DRF-A1 closely follow the multi-decadal variations of the SO SST anomalies (e.g., a warming trend during 1960–1980 and a cooling trend during 1980–2012) for the lead times of 1–5 years (Figure 12a), although there exists a time delay of predicting the observed SST trend turning point at around 1980 for lead times of 3–7 and 5–9 years. DRF-CM2.1 largely predicts the observed cooling trend during 1980–2012 for lead times of 1–5 and 3–7 years as reported in L.

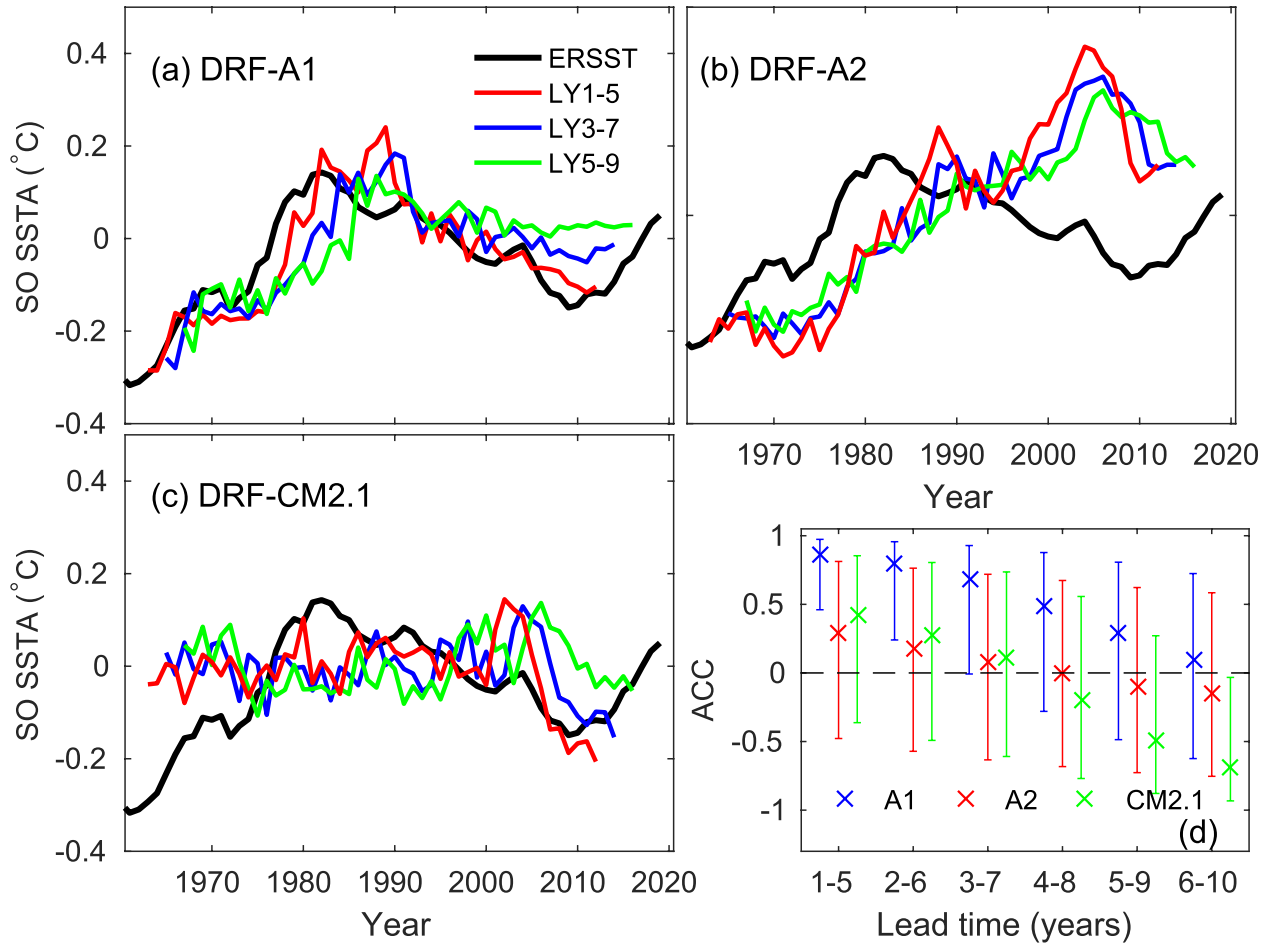

Figure 12. Time series of the sea surface temperature (SST) anomalies averaged over the Southern Ocean (50°S–70°S, 0°–360°) from Extended Reconstructed Sea Surface Temperature (ERSST) (black) and decadal retrospective forecasts. (a) the 5-year-mean of the SST anomalies from DRF-A1 for the lead times of 1–5 (red), 3–7 (blue) and 5–9 (green) years. (b) the same as (a) but from DRF-A2. (c) the same as (a) but from DRF-CM2.1. Anomalies in (a–c) are defined as departures from their time means over 1971–2010 for observations and forecasts respectively. (d) anomaly correlation coefficient (ACC) of the SST anomalies as a function of lead times from DRF-A1 (blue marker), DRF-A2 (red marker) and DRF-CM2.1 (green marker) verified versus ERSST; the error bar in (d) represents the 95% confidence interval of ACC with the adjusted degree of freedom using Equation 2. The time of the 5-year-mean forecasts is aligned to the center of the 5-year window.

Zhang et al. (2017) but fails in predicting the observed warming trend during 1960–1980. DRF-CM2.1 also overpredicts the observations during 2000–2005 for all lead times. DRF-A2 largely predicts the warming trend during 1960–1980 but fails in predicting the observed cooling trend during 1980–2012. Consequently, DRF-A1 shows a high predictive skill of predicting the SO SST anomalies with significant ACC up to lead times of 3–7 years, whereas both DRF-A2 and DRF-CM2.1 do not indicate statistically significant skill for any lead times (Figure 12d).

Southern Ocean convection acts as a driver of the observed recent multi-decadal climate trends (Zhang, Delworth, et al., 2019), so the phase of the SO convection in the initial condition is critical for the accuracy of the SO decadal climate prediction. The MLD as an indicator of SO convection in DCIS-A1 has a peak around 1980 (Figure S6a in Supporting Information S1), corresponding to that the SO SST peak around 1980 (i.e., the SST trend turning point) in DCIS-A1 matches well with observations (Figure S6b in Supporting Information S1 and Figure 12). In contrast, the MLD in DCIS-A2 peaks around 2000 (Figure S6 in Supporting Information S1), resulting in a delayed peak of SO SST at around 2000 in DCIS-A2 relative to the observed one. Therefore, the delayed SO SST peak associated with the delayed SO convection in DCIS-A2 leads to the failure of predicting the observed cooling trend during 1980–2012 in DRF-A2. The contrast between DRF-A1 and A2 in predicting the observed recent climate trends indicates that the surface pressure data

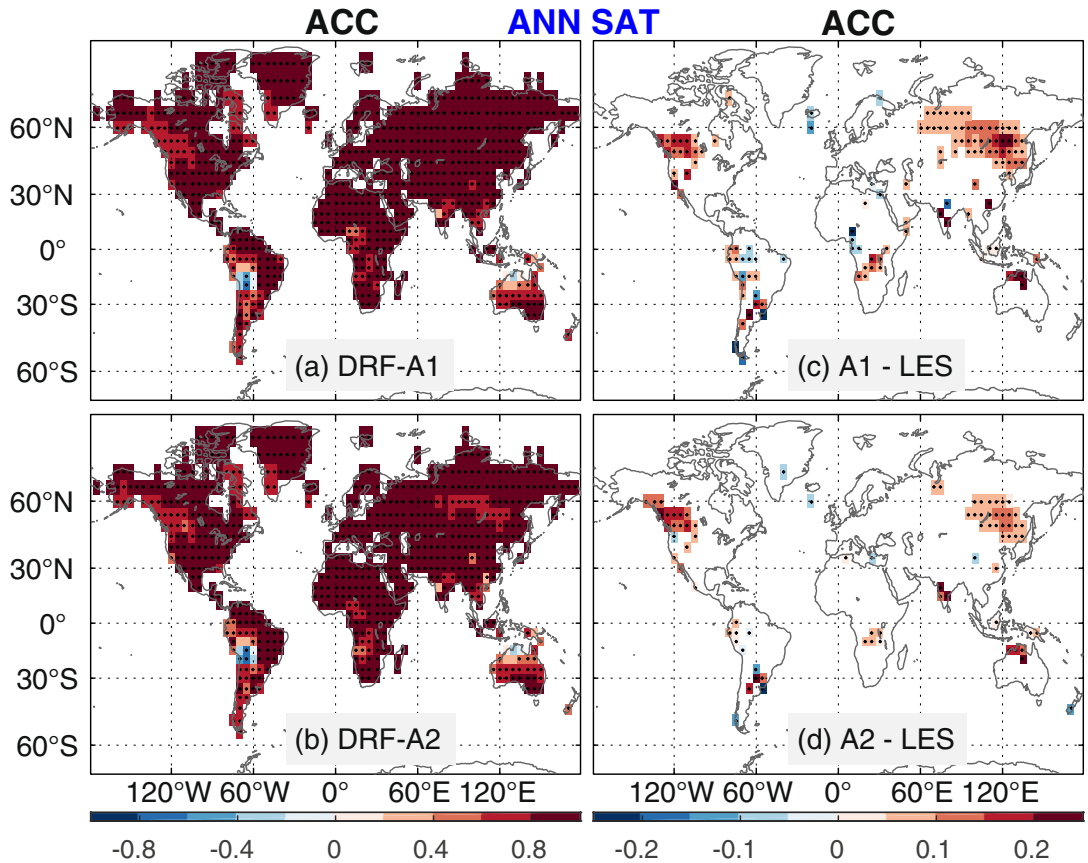

Figure 13. Anomaly correlation coefficient (ACC) of annual land surface air temperature from (a) DRF-A1 and (b) DRF-A2 relative to CRU-TSv4.03 observations for lead times of 2–9 years. ACC skill score differences between (c) DRF-A1 and SPEAR-LES, (d) between DRF-A2 and SPEAR-LES, respectively. All fields were mapped onto a 5° × 5° grid prior to ACC calculation. ACC values and their differences with the statistical p-value larger than 0.05 estimated from the bootstrap method are not shown.

assimilation is not sufficient for the SO decadal initialization. The delayed SO convection peak in DCIS-A2 relative to DCIS-A1 may be linked with the insufficient surface pressure observations before 1980. In addition, the dependence of the SO convection period on the mean ocean stratification (L. Zhang et al., 2021) might influence the phase of SO convection in DCIS-A1 and A2, since DCIS-A1 and A2 use full and anomaly initialization respectively. Further research is needed for unraveling the underlying mechanism on determining the SO convection phase in DCIS-A1 and A2.

5.2. The Decadal Prediction Skill of Hydroclimate Over Land

The gain of skill in predicting SST on decadal time scales due to initialization would imply skill improvement in predicting hydroclimate over land related to the improved SST prediction, thus we assess the decadal prediction skill of hydroclimate over land and the impact of initialization on the hydroclimate prediction. Following Smith et al. (2019), we focus on lead times of 2–9 years for the hydroclimate skill assessment. Beyond the annual mean hydroclimate skill as shown in Smith et al. (2019), we also examine the seasonal dependence of hydroclimate skill for the boreal summer (the mean of June, July, August and September (JJAS)) and boreal winter (the mean of December, January, February and March (DJFM)) respectively. The two initialized hindcasts agree with high skill in predicting annual SAT over land almost everywhere with ACC values exceeding 0.8 (Figures 13a and 13b), with main exceptions of the tropical Andes, the western Canada and the northern Australia. However, the region with coherently significant difference between initialized and uninitialized hindcasts locates only in Western Canada and Siberia (Figures 13c and 13d), indicating that the radiative forcing is the dominant driver for the decadal prediction skill of SAT over the global land. Interestingly, the significant improvement of the annual SAT skill over Western Canada and

Siberia due to initialization mainly arises from the boreal winter season (Figures 13c and 13d and Figures S7c and S7d in Supporting Information S1), but not from the boreal summer season (Figures S8c and S8d in Supporting Information S1).

The average predictability time (APT) analysis (DelSole & Tippett, 2009a, 2009b) has been used for distinguishing the interannual-scale predictable components from the externally forced trend-like response of land SAT in GFDL's seasonal hindcasts (Jia et al., 2015), so we apply APT to the boreal winter SAT over the eastern North Pacific and North America sector for understanding the source of the boreal winter SAT skill improvement over Western Canada. The details of the APT calculation are provided within the supplement (Text S1 in Supporting Information S1). Following Yang et al. (2013), the decadal retrospective forecasts of the boreal winter SAT for all lead times are used to solve the APT, and the APT analysis are conducted separately in DRF-A1 and DRF-A2. The leading predictable component (PrC) of the winter SAT shows a general warming trend pattern with considerably larger amplitudes in the Arctic and North America than North Pacific from both DRF-A1 and A2 (Figure S9 in Supporting Information S1), and it is highly predictable with ACC larger than 0.85 and 0.95 respectively from DRF-A1 and A2 for all lead times (Figure S10 in Supporting Information S1), suggesting the leading PrC is likely due to changes in external radiative forcing.

The spatial patterns of the second PrC from DRF-A1 and DRF-A2 largely agree on warm anomalies in northwestern North America, but the cold anomalous center of the second PrC pattern from DRF-A1 locates in the central North Pacific (Figure 14a), while the corresponding cold anomalous center from DRF-A2 tends to shift southwards and eastwards in the eastern North Pacific (Figure 14b). To explore the possible linkage between the second PrC and the observed North Pacific decadal variability, we compare the leading two principal components (PC) and corresponding EOFs of North Pacific DJFM SST anomalies (Text S2 and Figure S11 in Supporting Information S1) with the second PrC. PC1 is essentially the Pacific Decadal Oscillation (PDO) index as defined by Mantua et al. (1997), and PC2 is another important decadal variability in North Pacific beyond PDO (Bond et al., 2003). The spatial pattern of the second PrC from DRF-A2 is similar to the observed regression pattern of the winter SAT (Figure S12b in Supporting Information S1) onto the observed PDO index, and the pattern correlation coefficient (PCC) between the two patterns (Figure 14b and Figure S12b in Supporting Information S1) is 0.81, and the observed time series of the second PrC by projection onto the JRA55 observation is highly correlated with the observed PDO index ($r = 0.68$), suggesting that this component is likely PDO-related. Note that the observed regression pattern of the winter SAT onto the PDO index is consistent with Figure 3 in Mantua and Hare (2002). The spatial pattern of the second PrC from DRF-A1 resembles the observed regression pattern of the winter SAT onto the PC2 of the North Pacific winter SST anomalies (Figure S12a in Supporting Information S1), and their respective PCC (Figure 14a and Figure S12a in Supporting Information S1) is 0.87. The observed time series of the second PrC from DRF-A1 is highly correlated with the observed PC2 of the North Pacific winter SST ($r = 0.81$). Thus, the second PrC from DRF-A1 is likely associated with second EOF mode of the North Pacific winter SST. Although the second PrC patterns from DRF-A1 and A2 differ in the eastern North Pacific, they largely agree on the strong signals in northwestern North America, corresponding to the improvement of boreal winter SAT skill in Western Canada (Figure 13 and Figure S7 in Supporting Information S1). Unlike the high skill of the leading PrC for all lead times, the skill of the second PrC decreases with lead times and is only significant for lead times of 2–6 years with ACC of 0.51 and 0.50 from DRF-A1 and A2 respectively (Figure S10 in Supporting Information S1), suggesting that the second PrC for lead times of 2–6 years has significant contribution to the SAT skill improvement for 2–9 years. The different temporal and spatial variations of the second PrC in DRF-A1 and A2 are consistent with the winter SST skill difference in the North Pacific (Figure S13 in Supporting Information S1), i e., DRF-A1 has better skill over the central North Pacific and subtropical western North Pacific (strong loadings of the second EOF), while DRF-A2 has better skill of winter SST over the west coast of North America (the coastal frank of PDO) than DRF-A1 especially during 2–6 lead years. The APT analysis suggests that there might exist some connection between the improvement of SAT skill in Western Canada and the SST skill improvement in the Gulf of Alaska and the Pacific west coastal ocean (Figure 10). Note that the above analysis only shows coincidence between SAT predictability in Western Canada in DRF-A1 and A2 and observed decadal variability in the North Pacific, and causal explorations of this connection are left for future research.

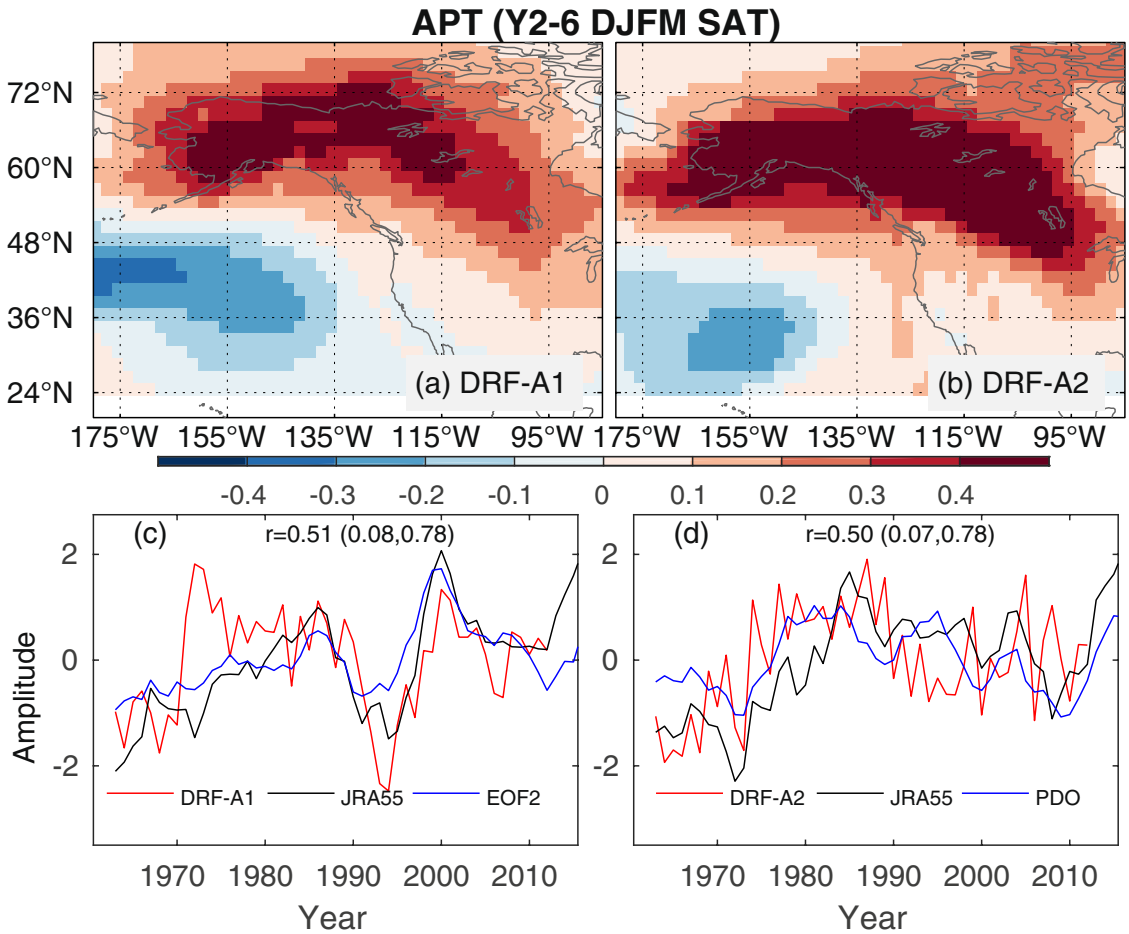

Figure 14. The average predictability time analysis of winter SAT over the North Pacific and North America sector. The spatial pattern of the second predictable component (PrC) for the winter surface air temperature from (a) DRF-A1 and (b) DRF-A2. The ensemble-mean time series (red) of the second PrC from (c) DRF-A1 for lead times of 2–6 years and the projected time series (black) onto the JRA55 observations. The ensemble-mean time series (red) of the second PrC from (d) DRF-A2 for lead times of 2–6 years and the projected time series (black) onto the JRA55 observations. The blue line in (c) shows the time series of the second EOF index of the North Pacific winter sea surface temperature from Extended Reconstructed Sea Surface Temperature; the blue line in (d) shows the time series of the first EOF index. Correlation coefficients with 5–95% confidence intervals are indicated in brackets in (c and d).

The precipitation from the two DRFs shows reasonable skill (ACC > 0.6) in the Sahel and in a broad band across northern Europe and Eurasia and significant skill in southwestern China and parts of North America (Figure S14 in Supporting Information S1). It hardly attributes the high skill in the Sahel, northern Europe and Eurasia to the initialization due to the sporadic distributions with significant difference. To further investigate the seasonal dependence of the precipitation prediction skill, we repeat the same skill assessment as the annual precipitation for the boreal summer and winter precipitation. The significant skill of annual precipitation in northern Europe and western Eurasia is mainly from the boreal winter while the skill over northeastern Eurasia is mainly from the boreal summer season (Figures S15 and S16 in Supporting Information S1). The significant skill of annual precipitation in the Sahel region mainly comes from the boreal summer season (Figures S14–S16 in Supporting Information S1), as the bulk of the annual rainfall falls between June and September (Biasutti, 2019).

To highlight the regional skill in the Sahel region, we show in Figure 15 the ACC of the JJAS land surface precipitation from retrospective forecasts relative to CRU-TSv4.03 observations for lead times of 2–9 years. In comparison with the significant skill over the entire Sahel region from DRF-A1, DRF-A2 and LES, DRF-CM2.1 shows significant skill only in the west Sahel region. For the linearly detrended mean summer precipitation averaged over lead times of 2–9 years, correlations with observations are 0.78, 0.76, 0.69, and 0.30 for DRF-A1, DRF-A2, LES and DRF-CM2.1 respectively, and all except DRF-CM2.1 are significant at the

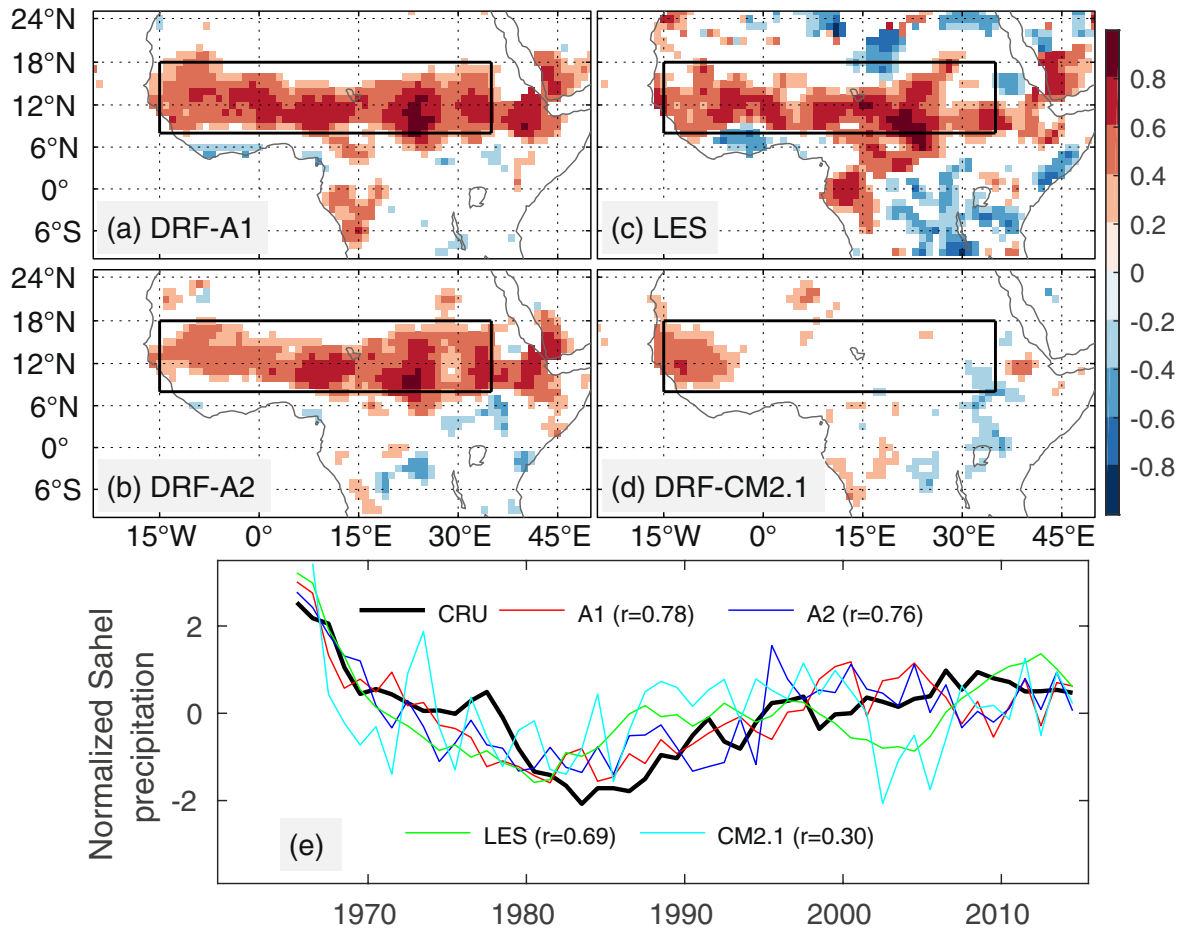

Figure 15. Anomaly correlation coefficient (ACC) of JJAS land surface precipitation from (a) DRF-A1, (b) DRF-A2, (c) LES and (d) DRF-CM2.1 relative to CRU-TSv4.03 observations for lead times of 2–9 years. Boxes in (a–d) outline the Sahel region. (e) The normalized and linearly detrended CRU-TSv4.03 observed (black) and modeled ensemble mean timeseries of summer Sahel (8°N–18°N, 15°W–35°E) rainfall from DRF-A1 (red), DRF-A2 (blue), LES (green) and DRF-CM2.1 (cyan) for lead times of 2–9 years. All fields were mapped onto a 1° × 1° grid prior to ACC calculation. ACC values with the statistical *p*-value larger than 0.05 are not shown.

95% significance level. Note that the correlation skill in both DRF-A1 and A2 is significantly higher than that in DRF-CM2.1 at the 95% significance level. The drought development during 1960s–1980s and the subsequent recovery are well captured by the SPEAR models, while they are not as well represented by DRF-CM2.1. To illustrate the physical process of the Intertropical Convergence Zone (ITCZ) shift associated with the Sahel precipitation on the multi-decadal scale (Sheen et al., 2017), we show in Figure 16 the difference of the composite summer precipitation over the tropical Atlantic and north Africa between the wet period (1995–2010) and the drought period (1975–1990) from observations and retrospective forecasts for lead times of 2–9 years. The three SPEAR retrospective forecasts are able to predict the observed northward migration of the ITCZ during the wet period with a north-south dipole pattern of precipitation anomalies (Figures 16a–16c and 16e–16f), and the spatial pattern correlation coefficient between model and GPCP is 0.62, 0.61, and 0.58 for A1, A2 and LES respectively. In contrast, DRF-CM2.1 is able to predict the observed ITCZ migration in the tropical Atlantic and the west Sahel but fails in predicting the observed ITCZ migration in the east Sahel region (Figures 16d and 16e), resulting in a lack of skill in the east Sahel region from DRF-CM2.1 (Figure 15d). Both initialized (DRF-A1 and A2) and uninitialized (LES) retrospective forecasts produce significantly skillful prediction of the Sahel precipitation, consistent with previous studies that both radiative forcing changes and internal AMV could lead to the Sahel precipitation changes on the multi-decadal scale (Biasutti, 2019; Held et al., 2005; Zhang & Delworth, 2006). Therefore, the skillful prediction of the Sahel summer rainfall in SPEAR decadal prediction system arises both from the radiative forcing and the initialized internal AMV.

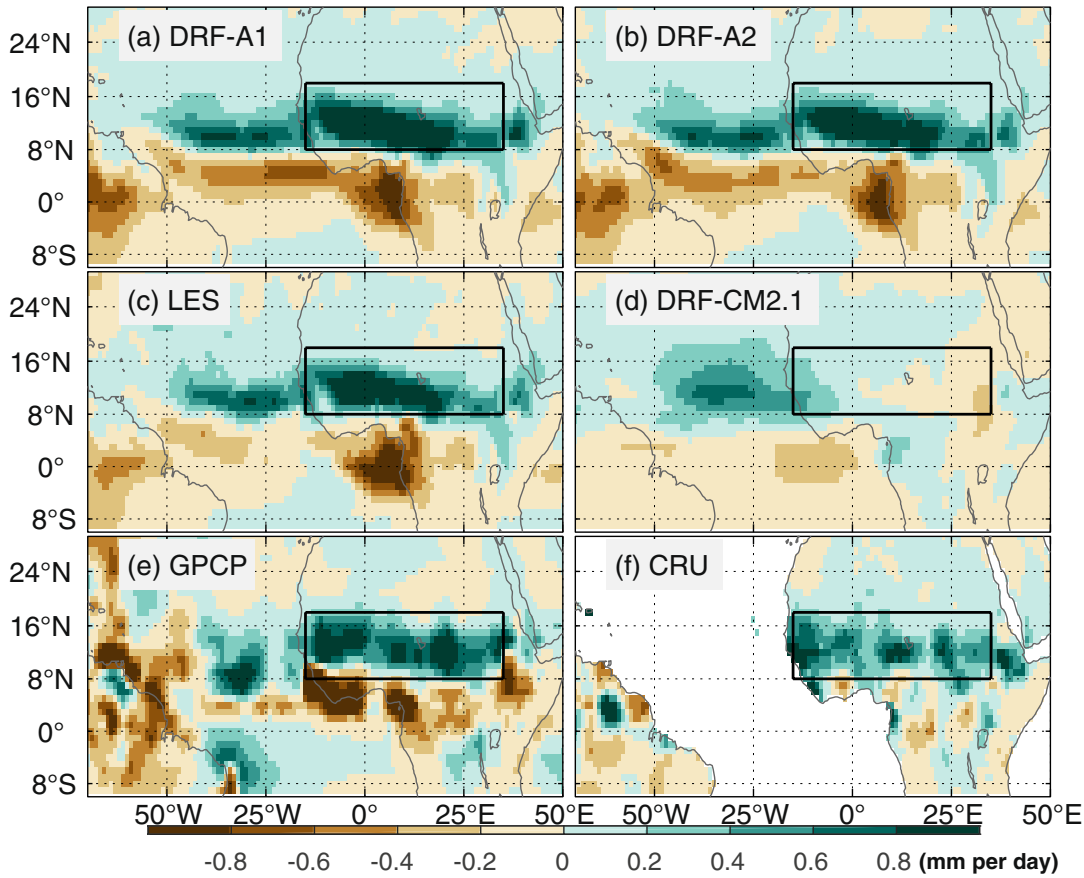

Figure 16. Difference of the composite summer precipitation over the tropical Atlantic and north Africa between the wet period (1995–2010) and the drought period (1975–1990) from (a) DRF-A1, (b) DRF-A2, (c) LES and (d) DRF-CM2.1 for lead times of 2–9 years. The corresponding observed difference from (e) GPCP data and (f) CRU data. The spatial pattern anomaly correlation coefficient between the forecasts (a–d) and GPCP (e) over the figure domain is 0.62, 0.61, 0.58, and 0.17 for DRF-A1, DRF-A2, LES and DRF-CM2.1 respectively.

5.3. Roles of AMOC in Decadal Prediction

We have shown in Section 4 that DCIS-A1 and A2 have very similar multi-decadal variations of AMOC driven by the observation-constrained air-sea fluxes during 1960–2019 and maintain a robust lead-lag relationship between AMOC and UOHC in the NASPG on multi-decadal time scales. Thus, we examine the evolution of AMOC and its physical role in the multi-decadal phase transition in the initialized retrospective forecasts.

Time series of the AMOC index at 45°N from DCIS-A1/A2, uninitialized SPEAR-LES and initialized DRF-A1/A2 during 1960–2019 are shown in Figure 17. The initialized forecasts from DRF-A1 and A2 closely follow the multi-decadal variations of AMOC in the corresponding DCIS for the lead times of 1–5 and 3–7 years (Figures 17a and 17b). For the longer lead time of years 5–9, the multi-decadal variations in both initialized forecasts tend to be weakened toward the forced response in SPEAR-LES. The climatological values of the AMOC index show almost unvarying changes with the forecast lead times, indicating that the physical balance of each model component is well maintained in the initial states of DCIS (Figure 17c). The high prediction skill (ACC > 0.8) verified versus DCIS is up to lead 4–8 years in the two forecasts (Figure 17d). For the longer lead times of 5–9 and 6–10 years, the skill from DRF-A2 tends to drop faster than that from DRF-A1.

To show the NASPG SST anomalies for all 50 retrospective forecasts, we show in Figure 18 the time series of the SST anomalies from ERSST observations, DRF-CM2.1 and DRF-A1/A2 respectively. The initialized forecasts from DRF-A1 and CM2.1 closely follow the multi-decadal variations of the observed NASPG SST anomalies. However, DRF-A2 overpredicted the observed cold anomalies during 1960s and 1970s.

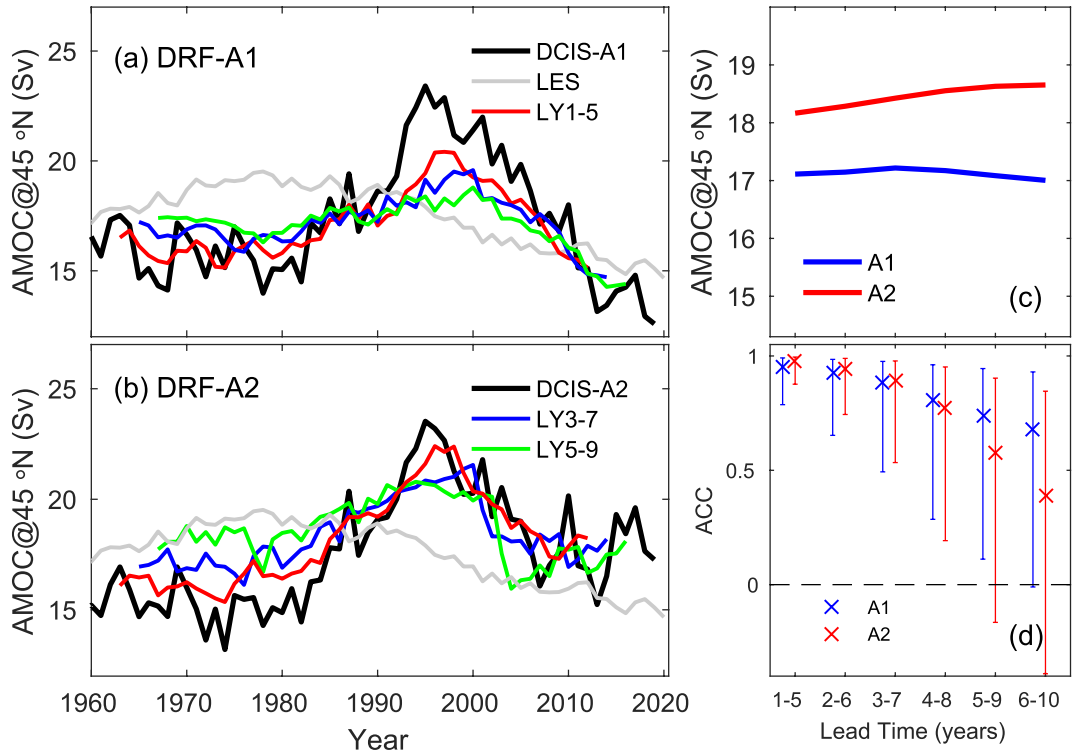

Figure 17. Time series of the Atlantic Meridional Overturning Circulation (AMOC) index at 45°N from the coupled initialization system (black), uninitialized SPEAR-LES (gray) and initialized decadal retrospective forecasts. (a) the 5-year-mean of the AMOC index from DRF-A1 for the lead times of 1–5 (red), 3–7 (blue) and 5–9 (green) years. (b) the same as (a) but from DRF-A2. (c) The climatological values of AMOC index as a function of lead times from DRF-A1 (blue) and A2 (red). (d) Anomaly correlation coefficient (ACC) of the AMOC index as a function of lead times from DRF-A1 (DRF-A2) verified versus DCIS-A1 (DCIS-A2); the error bar in (d) represents the 95% confidence interval of ACC with the adjusted degree of freedom using Equation 2. The time of the 5-year-mean forecasts is aligned to the center of the 5-year window.

Consequently, the skill of predicting the pentadal NASPG SST from DRF-A1 is very high (ACC > 0.9) for all the lead times, while the skill from DRF-A2 is consistently lower than that from DRF-A1. In contrast, the uninitialized LES predicts a linear warming trend. The ACC skill of predicting NASPG SST from the three models is comparable for all lead times (Figure 18d), although DRF-CM2.1 shows much larger bias of NASPG SST than DRF-A1/A2 (Figure S17 in Supporting Information S1).

The high skill of AMOC up to 4–8 years implies that the skillful prediction of NASPG SST in even longer lead times could arise from the lead-lag relationship between AMOC and UOHC in the NASPG (Figure 8). For the distinct phase transition of NASPG SST around mid-1990s, the observed abrupt warming event starting in 1993 was largely predicted from both DRF-A1 and A2 (Figure S17 in Supporting Information S1). This abrupt SST warming event is also associated with the rapid change of UOHC (Figure 8), so the success of predicting this event can be attributed to the skillful prediction of the strong AMOC during the event period through the lead-lag relationship between AMOC and the UOHC over the NASPG. To further highlight the dynamical role of AMOC in the new decadal prediction system, we perform correlation analysis between the initial values of AMOC index at 45°N for each forecast and the corresponding predicted NASPG SST (Figure 19). In DRF-A1 and A2, the predicted values of NASPG SST from all 50 forecasts at three different lead times are highly correlated with the corresponding initial values of AMOC index with significant correlation coefficients larger than 0.75 (Figures 19a and 19b), but there is no statistically significant relationship between the predicted NASPG SST and corresponding initial values of AMOC index from DRF-CM2.1 (Figure 19c). Note that this high correlation between AMOC and predicted NASPG SST in DRF-A1/A2 arises from the lead-lag relationship between AMOC and NASPG UOHC that is manifested in DCIS-A1/A2. Given that the initial condition of UOHC in DRF-A1/A2 is less accurate than that in DRF-CM2.1 (Figure 7), the

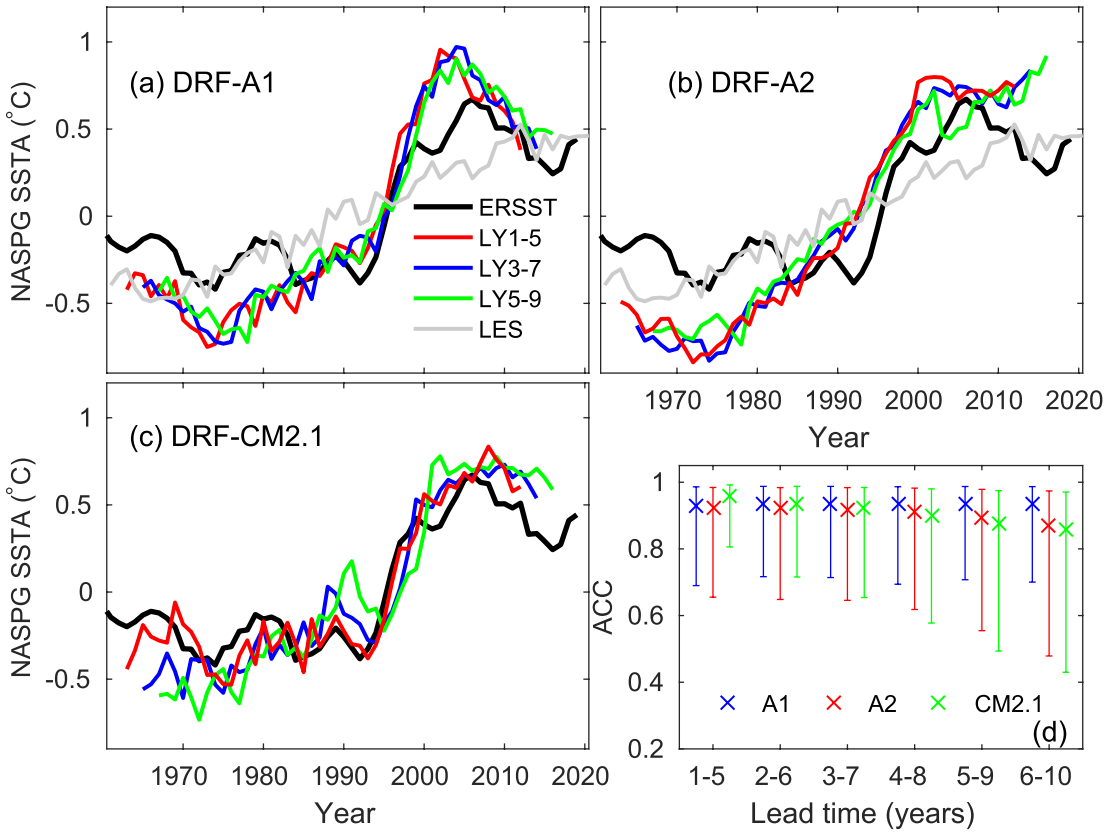

Figure 18. Time series of the sea surface temperature (SST) anomalies averaged over North Atlantic Subpolar Gyre (45°N–60°N, 60°W–0°) from Extended Reconstructed Sea Surface Temperature (ERSST) (black) and decadal retrospective forecasts. (a) the 5-year-mean of the SST anomalies from DRF-A1 for the lead times of 1–5 (red), 3–7 (blue) and 5–9 (green) years, and from uninitialized SPEAR-LES (gray). (b) the same as (a) but from DRF-A2. (c) the same as (a) but from DRF-CM2.1. Anomalies in (a–c) are defined as departures from their time means over 1971–2010 for observations and forecasts respectively. (d) Anomaly correlation coefficient (ACC) of the SST anomalies as a function of lead times from DRF-A1 (blue marker), DRF-A2 (red marker) and DRF-CM2.1 (green marker) verified versus ERSST; the error bar in (d) represents the 95% confidence interval of ACC with the adjusted degree of freedom using Equation 2. The time of the 5-year-mean forecasts is aligned to the center of the 5-year window.

comparable skill between two different initialization systems implies that the skill of predicting NASPG SST in DRF-A1/A2 might arise from both persistence of UOHC and dynamical role of AMOC. Therefore, the dynamical role of AMOC plays important role in the prediction skill from DRF-A1/A2, whereas there is no contribution of AMOC to the prediction skill of NASPG SST in DRF-CM2.1.

6. Conclusions and Discussions

We have documented the development and skill assessment of GFDL's new coupled decadal prediction system using SPEAR model. We have developed a decadal coupled initialization system (DCIS) that does not use subsurface ocean observations. In DCIS, the atmospheric state and SST were constrained by the time-varying observations, and thus the ocean component of the coupled model is forced by a very similar sequence of surface heat fluxes and wind stresses as observations. With the similar physical mechanism of producing the observed low-frequency climate variability as the forced ocean-sea ice (FOSI) approach used in CESM-DP (S. G. Yeager et al., 2018), DCIS allows interactive coupling in the initialization process, thus maintaining physical balance among different components in the coupled model. DCIS offers two initialization approaches, which differ only in the atmosphere state constrains by observations. DCIS-A1 utilized the atmospheric winds and temperature restoring toward the JRA reanalysis, while DCIS-A2 assimilated the surface pressure observations using the ensemble data assimilation technique.

The interannual to multidecadal variability of the essential variables (e.g., the surface winds, surface air temperature and SST) governing the heat and momentum fluxes at the air-sea interface from DCIS-A1 and

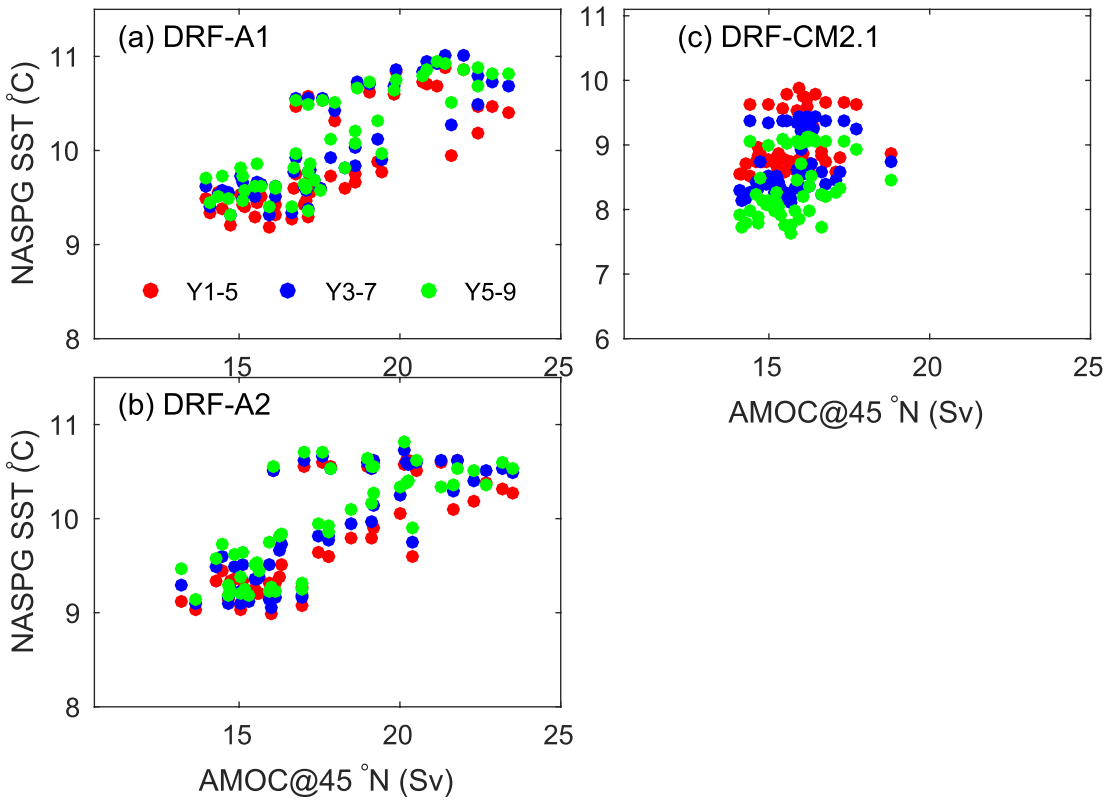

Figure 19. Scatter plot of initial values of Atlantic Meridional Overturning Circulation index at 45°N for each forecast and corresponding predicted sea surface temperature (SST) averaged over North Atlantic Subpolar Gyre (45°N–60°N, 0°–60°E) from decadal retrospective forecasts. (a) the 5-year-mean of the SST anomalies from DRF-A1 for the lead times of 1–5 (red), 3–7 (blue) and 5–9 (green) years. (b) the same as (a) but from DRF-A2. (c) the same as (a) but from DRF-CM2.1.

A2 is in high agreement with the NNR data in the Northern Hemisphere. The two initializations can reproduce the similar physical processes connecting the surface heat fluxes to the deep ocean convection over the North Atlantic SPG associated with the NAO forcing, which are the key drivers of the decadal variability of AMOC from the surface flux forcing (Danabasoglu et al., 2016; Delworth et al., 2016). Consequently, the two initializations show very similar multi-decadal variations of AMOC. In addition, the robust physical lead-lag relationship between AMOC and UOHC in the NASPG was well captured by DCIS-A1 and A2, indicating that the new decadal initialization has largely integrated the fundamental physics of enhancing AMOC's role in the decadal prediction (Delworth et al., 2016, 2017; Robson et al., 2012; Yeager & Robson, 2017; S. Yeager et al., 2012; Zhang & Zhang, 2015).

The decadal prediction skill of SST was assessed versus observations for the two sets of initialized retrospective forecasts and compared with the uninitialized SPEAR-LES. Overall, the significant skill improvement due to initialization in the North Atlantic SPG and SO in our new system is consistent with many previous decadal prediction studies (Smith et al., 2019; S. G. Yeager et al., 2018; L. Zhang et al., 2017), indicating that this new system captures the essential elements of decadal initialization even without the direct subsurface ocean data assimilation. In comparison to GFDL's CM21-based system, the SPEAR decadal prediction system (the A1 approach) shows overall substantial bias reduction, significant skill improvement in the Southern Ocean and comparable or better skill in the other ocean basins. Specifically, the SPEAR DRF-A1 significantly outperforms CM2.1 in better predicting the multi-decadal SST trend before and after 1980 in the Southern Ocean.

The global patterns with the significant decadal skill of SAT and precipitation over land from DRF-A1 and A2 resemble those estimated from the state-of-the-art multi-model decadal predictions (Smith et al., 2019), demonstrating the robustness of the decadal hydroclimate prediction skill from our new system. We found that the significant annual SAT skill improvement due to initialization in western Canada and Siberia

mainly comes from the boreal winter but not from the boreal summer. The APT analysis suggested that SAT skill improvement in western Canada is related to the observed North Pacific decadal variability. In terms of decadal prediction system development advance, the new SPEAR decadal prediction system outperforms the previous CM2.1-based system in predicting the multidecadal variations of Sahel precipitation and the corresponding ITCZ shift. The benefit of initialization might be underestimated using the current method for comparing skill (Smith et al., 2019; Sospedra-Alfonso & Boer, 2020), and we will explore other more powerful methods to assess impact of initialization in the future study.

The multidecadal AMOC is highly predictable when verified versus the DCIS in SPEAR's decadal prediction system, and AMOC's dynamical role in predicting the SST over the North Atlantic SPG can be established through the lead-lag relationship between AMOC and UOHC. The abrupt warming event starting 1993 and subsequent phase transition of the SPG SST were largely predicted by the new system. Even without assimilating the subsurface ocean observations, SPEAR DRF-A1/A2 shows comparable SST predictive skill at all lead years as DRF-CM2.1. The high correlation between the AMOC initial values and forecasted NASPG SST in DRF-A1/A2 indicating the dynamical role of AMOC in enhancing the decadal prediction skill. The results here support that the proposed approach is feasible for the decadal climate prediction, increasing diversity of modeling tools for the decadal climate initialization and predictability research (Table 1).

Since the lead-lag connection between AMOC and UOTC on multidecadal decadal time scales were not established in CM2.1-ECDA, the prediction skill mainly arises from the persistence of the observed ocean heat content from the direct subsurface ocean data assimilation. Therefore, CM2.1-ECDA and SPEAR-DCIS provide two different mechanisms for achieving comparable skill of the decadal prediction in the North Atlantic. In future, we will explore feasible methods for imposing the direct subsurface ocean data assimilation (Lu et al., 2020) in the decadal initialization while maintaining the physical balance of AMOC and ocean heat content as shown in this study. The further increase of decadal predictive skill might be expected if the subsurface ocean initialization is appropriately included in this approach.

The high similarity between the two approaches of reproducing the physical processes governing the multidecadal variations of AMOC and yielding skillful decadal prediction in the North Atlantic, which indicates that the surface pressure data assimilation may provide sufficient atmospheric observational information for the decadal initialization research in the North Atlantic. With the availability of the centennial long record of surface pressure observations, DCIS-A2 provides potential of extending the coupled decadal initialization over the entire twentieth century for the predictability studies, thus allowing more multidecadal episodes for testing the reliability of North Atlantic decadal climate prediction. However, DRF-A2 failed in predicting the observed recent climate trends in SO due to its delayed phase of convection in DCIS-A2, indicates that the surface pressure data assimilation is not sufficient for the SO decadal initialization. Given the fact that the predictive skill of SST is generally higher in DRF-A1 than DRF-A2 especially over Southern Ocean, DRF-A1 will be the approach for SPEAR's real-time decadal prediction.

Data Availability Statement

Model output relevant to this paper can found online (ftp://data1.gfdl.noaa.gov/users/Xiaosong.Yang/SPEAR_DECP_paper/SPEAR).

Acknowledgments
The authors would like to thank Drs. Feiyu Lu and Mitchell Bushuk for their helpful comments on an earlier draft, and William Merryfield and an anonymous reviewer for very constructive comments that helped to improve the manuscript. The authors would also like to thank Dr. Zhi Liang for providing code optimization in the surface pressure data assimilation. Financial support for this work was provided through base funding from the National Oceanic and Atmospheric Administration to the Geophysical Fluid Dynamics Laboratory. GPC and CKM were supported in part by the NOAA Cooperative Agreement with CIRES, NA17OAR4320101.

References

Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C. O., et al. (2019). The GFDL global ocean and sea ice model OM4.0: Model description and simulation features. *Journal of Advances in Modeling Earth Systems*, *11*, 3167–3211. https://doi.org/10.1029/2019MS001726

Adler, R. F., Sapiano, M. R. P., Huffman, G. J., Wang, J.-J., Gu, G., Bolvin, D., et al. (2018). The Global Precipitation Climatology Project (GPCP) Monthly Analysis (New Version 2.3) and a Review of 2017 Global Precipitation. *Atmosphere*, *9*, 138. https://doi.org/10.3390/atmos9040138

Anderson, J. L. (2003). A local least squares framework for ensemble filtering. *Monthly Weather Review*, *131*(4), 634–642. https://doi.org/10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2

Balmaseda, M. A., Mogensen, K., & Weaver, A. T. (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, *139*, 1132–1161. https://doi.org/10.1002/qj.2063

Benjamin, S. G., & Miller, P. A. (1990). An alternative sea level pressure reduction and a statistical comparison of geostrophic wind estimates with observed surface winds. *Monthly Weather Review*, *118*, 2099–2116. https://doi.org/10.1175/1520-0493(1990)118<2099:aaslpr>2.0.co;2

Biasutti, M. (2019). Rainfall trends in the African Sahel: Characteristics, processes, and causes. *WIREs Climate Change*, *10*(e591). https://doi.org/10.1002/wcc.591

Bilbao, R., Wild, S., Ortega, P., Acosta-Navarro, J., Arsouze, T., Bretonnière, P.-A., et al. (2021). Assessment of a full-field initialized decadal climate prediction system with the CMIP6 version of EC-Earth. *Earth System Dynamics*, *12*, 173–196. https://doi.org/10.5194/esd-12-173-2021

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., et al. (2016). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, *9*, 3751–3777. https://doi.org/10.5194/gmd-9-3751-2016

Bond, N. A., Overland, J. E., Spillane, M., & Stabeno, P. (2003). Recent shifts in the state of the North Pacific. *Geophysical Research Letters*, *30*, 218323. https://doi.org/10.1029/2003GL018597

Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M., & Bladé, I. (1999). The effective number of spatial degrees of freedom of a time-varying field. *Journal of Climate*, *12*, 1990–2009. https://doi.org/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2

Chang, C., & Wang, Z. (2020). Multiyear hybrid prediction of Atlantic tropical cyclone activity and the predictability sources. *Journal of Climate*, *33*, 2263–2279. https://doi.org/10.1175/JCLI-D-19-0475.1

Chang, Y.-S., Zhang, S., Rosati, A., Delworth, T. L., & Stern, W. F. (2013). An assessment of oceanic variability for 1960–2010 from the GFDL ensemble coupled data assimilation. *Climate Dynamics*, *40*, 775–803. https://doi.org/10.1007/s00382-012-1412-2

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Allan, R. J., McColl, C., Yin, X., et al. (2019). *The International Surface Pressure Databank Version 3*. NCAR/UCAR Research Data Archive. https://doi.org/10.5065/D6D50K29

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., et al. (2011). The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society*, *137*, 1–28. https://doi.org/10.1002/qj.776

Cram, T. A., Compo, G. P., Yin, X., Allan, R. J., McColl, C., Vose, R. S., et al. (2015). The international surface pressure databank version 2. *Geoscience Data Journal*, *2*, 31–46. https://doi.org/10.1002/gdj3.25

Danabasoglu, G., Yeager, S. G., Kim, W. M., Behrens, E., Bentsen, M., Bi, D., et al. (2016). North Atlantic simulations in coordinated ocean-ice reference experiments phase II (CORE-II). Part II: Inter-annual to decadal variability. *Ocean Modeling*, *97*, 65–90. https://doi.org/10.1016/j.ocemod.2015.11.007

DelSole, T., & Tippett, M. K. (2009a). Average predictability time. Part I: Theory. *Journal of Atmospheric Sciences*, *66*, 1172–1187. https://doi.org/10.1175/2008jas2868.1

DelSole, T., & Tippett, M. K. (2009b). Average predictability time. Part II: Seamless diagnosis of predictability on multiple time scales. *Journal of Atmospheric Sciences*, *66*, 1188–1204. https://doi.org/10.1175/2008jas2869.1

Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A., et al. (2006). GFDL's CM2 Global Coupled Climate Models. Part I: Formulation and Simulation Characteristics. *Journal of Climate*, *19*(5), 643–674.

Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J.-H., Dunne, K. A., et al. (2020). SPEAR: The next generation GFDL modeling system for seasonal to multidecadal prediction and projection. *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS001895. https://doi.org/10.1029/2019MS001895

Delworth, T. L., & Zeng, F. (2016). The impact of the north Atlantic oscillation on climate through its influence on the atlantic meridional overturning circulation. *Journal of Climate*, *29*, 941–962. https://doi.org/10.1175/JCLI-D-15-0396.1

Delworth, T. L., Zeng, F., Vecchi, G. A., Yang, X., Zhang, L., & Zhang, R. (2016). The North Atlantic Oscillation as a driver of rapid climate change in the Northern Hemisphere. *Nature Geoscience*, *9*, 509–512. https://doi.org/10.1038/ngeo2738

Delworth, T. L., Zeng, F., Zhang, L., Zhang, R., Vecchi, G. A., & Yang, X. (2017). The central role of ocean dynamics in connecting the north Atlantic oscillation to the extratropical component of the Atlantic multidecadal oscillation. *Journal of Climate*, *30*, 3789–3805. https://doi.org/10.1175/JCLI-D-16-0358.1

Eden, C., & Willebrand, J. (2001). Mechanism of Interannual to Decadal Variability of the North Atlantic Circulation. *Journal of Climate*, *14*, 2266–2280. https://doi.org/10.1175/1520-0442(2001)014<2266:MOITDV>2.0.CO;2

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757. https://doi.org/10.1002/qj.49712555417

Gelaro, R., McCarty, W., MaxSuárez, J. M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, *30*, 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Goddard, L., Kumar, A., Solomon, A., Smith, D., Boer, G., Gonzalez, P., et al. (2013). A verification framework for interannual-to-decadal predictions experiments. *Climate Dynamics*, *40*, 245–272. https://doi.org/10.1007/s00382-012-1481-2

Good, S. A., Martin, M. J., & Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, *118*, 6704–6716. https://doi.org/10.1002/2013JC009067

Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, *7*, 109. https://doi.org/10.1038/s41597-020-0453-3

Held, I. M., Delworth, T., Lu, J., Findell, K. L., & Knutson, T. (2005). Simulation of Sahel drought in the 20th and 21st centuries. *Proceedings of the National Academy of Sciences*, *102*(50), 17891–17896. https://doi.org/10.1073/pnas.0509057102

Held, I. M., Guo, H., Adcroft, A., Dunne, J. P., Horowitz, L. W., Krasting, J., et al. (2019). Structure and performance of GFDL's CM4.0 climate model. *Journal of Advances in Modeling Earth Systems*, *11*, 3691–3727. https://doi.org/10.1029/2019MS001829

Huang, B., Thorne, P. W., Banzon, V. F., Tim Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017). Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *Journal of Climate*, *30*, 8179–8205. https://doi.org/10.1175/JCLI-D-16-0836.1

Hurrell, J. W., & Deser, C. (2009). North Atlantic climate variability: The role of the North Atlantic Oscillation. *Journal of Marine Systems*, *78*(1), 28–41. https://doi.org/10.1016/j.jmarsys.2008.11.026

Ingleby, N. B. (1995). Assimilation of station level pressure and errors in station height. *Weather and Forecasting*, *10*, 172–182. https://doi.org/10.1175/1520-0434(1995)010<0172:aoslpa>2.0.co;2

Japan Meteorological Agency/Japan. (2013). *JRA-55: Japanese 55-year Reanalysis, Daily 3-Hourly and 6-Hourly Data*. Dataset: Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D6HH6H41

Jia, L., Yang, X., Vecchi, G. A., Gudgel, R. G., Delworth, T. L., Rosati, A., et al. (2015). Improved seasonal prediction of temperature and precipitation over land in a high-resolution GFDL Climate Model. *Journal of Climate*, *28*(5), 2044–2062. https://doi.org/10.1175/jcli-d-14-00112.1

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Gandin, L., Iredell, M., et al. (1996). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, *77*, 437–472. https://doi.org/10.1175/1520-0477(1996)077<0437:tnyrp>2.0.co;2

Karspeck, A. R., Stammer, D., Köhl, A., Danabasoglu, G., Alonso Balmaseda, M., Smith, D. M., et al. (2017). Comparison of the Atlantic meridional overturning circulation between 1960 and 2007 in six ocean reanalysis products. *Climate Dynamics*, *49*, 957–982. https://doi.org/10.1007/s00382-015-2787-7

Kataoka, T., Tatebe, H., Koyama, H., Mochizuki, T., Ogochi, K., Naoe, H., et al. (2020). Seasonal to decadal predictions with MIROC6: Description and basic evaluation. *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS002035. https://doi.org/10.1029/2019MS002035

Kim, H.-M., Webster, P. J., & Curry, J. A. (2012). Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts. *Geophysical Research Letters*, *39*, L10701. https://doi.org/10.1029/2012GL051644

Kirtman, B., Power, S. B., Adedoyin, J. A., Boer, G. J., Bojariu, R., Camilloni, I., et al. (2013). Near-term Climate Change: Projections and Predictability. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, et al. (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 Reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, *93*(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Kriegler, E., Bauer, N., Popp, A., Humpenöder, F., Leimbach, M., Strefler, J., et al. (2017). Fossil-fueled development (SSP5): An energy and resource intensive scenario for the 21st century. *Global Environmental Change*, *42*, 297–315. https://doi.org/10.1016/j.gloenvcha.2016.05.015

Lu, F., Harrison, M. J., Rosati, A., Delworth, T. L., Yang, X., Cooke, W. F., et al. (2020). GFDL's SPEAR seasonal prediction system: Initialization and ocean tendency adjustment (OTA) for coupled model predictions. *Journal of Advances in Modeling Earth Systems*, *12*, e2020MS002149. https://doi.org/10.1029/2020MS002149

Mantua, N. J., & Hare, S. R. (2002). The Pacific Decadal Oscillation. *Journal of Oceanography*, *58*, 35–44. https://doi.org/10.1023/A:1015820616384

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997). A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production. *Bulletin of the American Meteorological Society*, *78*(6), 1069–1080. https://doi.org/10.1175/1520-0477(1997)078<1069:apicow>2.0.co;2

Marshall, G. J. (2003). Trends in the Southern Annular Mode from observations and reanalyses. *Journal of Climate*, *16*, 4134–4143. https://doi.org/10.1175/1520-0442(2003)016<4134:titsam>2.0.co;2

Meehl, G. A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., et al. (2014). Decadal climate prediction: An update from the trenches. *Bulletin of the American Meteorological Society*, *95*, 243–267. https://doi.org/10.1175/BAMS-D-12-00241.1

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., et al. (2009). Decadal prediction. *Bulletin of the American Meteorological Society*, *90*(10), 1467–1486. https://doi.org/10.1175/2009bams2778.1

Merryfield, W. J., Lee, W., Boer, G. J., Kharin, V. V., Scinocca, J. F., Flato, G. M., et al. (2013). The Canadian Seasonal to Interannual Prediction System. Part I: Models and Initialization. *Monthly Weather Review*, *141*(8), 2910–2945. https://doi.org/10.1175/mwr-d-12-00216.1

Moat, B. I., Frajka-Williams, E., Smeed, D., Rayner, D., Sanchez-Franks, A., Johns, W. E., et al. (2020). *Atlantic meridional overturning circulation observed by the RAPID-MOCHA-WBTS (RAPID-Meridional Overturning Circulation and Heatflux Array-Western Boundary Time Series) array at 26N from 2004 to 2018 (v2018.2)*. NERC, UK: British Oceanographic Data Centre, National Oceanography Centre.

Msadek, R., Delworth, T. L., Rosati, A., Anderson, W. G., Vecchi, G. A., Chang, Y.-S., et al. (2014). Predicting a Decadal Shift in North Atlantic Climate Variability Using the GFDL Forecast System. *Journal of Climate*, *27*(17). https://doi.org/10.1175/JCLI-D-13-00476.1

Newman, L., Heil, P., Trebilco, R., Katsumata, K., Scinocca, J. F., van Wijk, E., et al. (2019). Delivering sustained, coordinated and integrated observations of the Southern Ocean for global impact. *Frontiers in Marine Science*. https://doi.org/10.3389/fmars.2019.00433

Newman, M. (2013). An Empirical Benchmark for Decadal Forecasts of Global Surface Temperature Anomalies. *Journal of Climate*, *26*, 5260–5269. https://doi.org/10.1175/JCLI-D-12-00590.1

O'Kane, T. J., Sandery, P. A., Kitsios, V., Sakov, P., Chamberlain, M. A., Collier, M. A., et al. (2021). CAFE60v1: A 60-Year Large Ensemble Climate Reanalysis. Part I: System Design, Model Configuration, and Data Assimilation. *Journal of Climate*, *34*(13), 5153–5169.

O'Kane, T. J., Sandery, P. A., Kitsios, V., Sakov, P., Chamberlain, M. A., Squire, D. T., et al. (2021). CAFE60v1: A 60-Year Large Ensemble Climate Reanalysis. Part II: Evaluation. *Journal of Climate*, *34*(13), 5171–5194.

Pohlmann, H., Müller, W. A., Bittner, M., Hettrich, S., Modali, K., Pankatz, K., & Marotzke, J. (2019). Realistic quasi-biennial oscillation variability in historical and decadal hindcast simulations using CMIP6 forcing. *Geophysical Research Letters*, *46*, 14118–14125. https://doi.org/10.1029/2019GL084878

Pohlmann, H., Smith, D. M., Balmaseda, M. A., Keenlyside, N. S., Masina, S., Matei, D., et al. (2013). Predictability of the mid-latitude Atlantic meridional overturning circulation in a multi-model system. *Climate Dynamics*, *41*, 775–785. https://doi.org/10.1007/s00382-013-1663-6

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., et al. (2017). The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, *42*, 153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Robertson, F. R., Roberts, J. B., Bosilovich, M. G., Bentamy, A., Clayson, C. A., Fennig, K., et al. (2020). Uncertainties in Ocean Latent Heat Flux Variations over Recent Decades in Satellite-Based Estimates and Reduced Observation Reanalyses. *Journal of Climate*, *33*, 8415–8437. https://doi.org/10.1175/JCLI-D-19-0954.1

Robson, J., Sutton, R., Lohmann, K., Smith, D., & Palmer, M. D. (2012). Causes of the rapid warming of the North Atlantic Ocean in the Mid-1990s. *Journal of Climate*, *25*(12), 4116–4134. https://doi.org/10.1175/jcli-d-11-00443.1

Sheen, K., Smith, D., Dunstone, N., Eade, R., Rowell, D. P., & Vellinga, M. (2017). Skilful prediction of Sahel summer rainfall on inter-annual and multi-year timescales. *Nature Communications*, *8*, 14966. https://doi.org/10.1038/ncomms14966

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., et al. (2019). Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quarterly Journal of the Royal Meteorological Society*, *145*, 2876–2908. https://doi.org/10.1002/qj.3598

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

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

Smith, D. M., Eade, R., Scaife, A. A., Caron, L.-P., Danabasoglu, G., DelSole, T., et al. (2019). Robust skill of decadal climate predictions. *npj Climate and Atmospheric Science*, *2*, 13. https://doi.org/10.1038/s41612-019-0071-y

Smith, D. M., Scaife, A. A., Boer, G. J., Caian, M., Doblas-Reyes, F. J., Guemas, V., et al. (2013). Real-time multi-model decadal climate predictions. *Climate Dynamics*, *41*, 2875–2888. https://doi.org/10.1007/s00382-012-1600-0

Sospedra-Alfonso, R., & Boer, G. J. (2020). Assessing the impact of initialization on decadal prediction skill. *Geophysical Research Letters*, *47*, e2019GL086361. https://doi.org/10.1029/2019GL086361

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of the CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*, 485–498. https://doi.org/10.1175/bams-d-11-00094.1

Timmreck, C., Pohlmann, H., Illing, S., & Kadow, C. (2016). The impact of stratospheric volcanic aerosol on decadal-scale climate predictions. *Geophysical Research Letters*, *43*, 834–842. https://doi.org/10.1002/2015GL067431

Trenberth, K. E., & Hurrell, J. W. (1994). Decadal atmosphere-ocean variations in the Pacific. *Climate Dynamics*, *9*, 303–319. https://doi.org/10.1007/BF00204745

Wang, Y., Counillon, F., Bethke, I., Keenlyside, N., Bocquet, M., & Shen, M.-L. (2017). Optimising assimilation of hydrographic profiles into isopycnal ocean models with ensemble data assimilation. *Ocean Modelling*, *114*, 33–44. https://doi.org/10.1016/j.ocemod.2017.04.007

Whitaker, J. S., & Hamill, T. M. (2012). Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, *140*, 3078–3089. https://doi.org/10.1175/mwr-d-11-00276.1

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., et al. (2017). The Met Office Global Coupled model 3.0 and 3.1 (GC3.0 and GC3.1) configurations. *Journal of Advances in Modeling Earth Systems*, *10*, 357–380. https://doi.org/10.1002/2017MS001115

Wu, B., Chen, X., Song, F., Sun, Y., & Zhou, T. (2015). Initialized decadal predictions by LASG/IAP Climate System Model FGOALS-s2: Evaluations of strengths and weaknesses. *Advances in Meteorology*, *12*. https://doi.org/10.1155/2015/904826

Yang, X., Rosati, A., Zhang, S., Delworth, T. L., Gudgel, R. G., Zhang, R., et al. (2013). A Predictable AMO-like pattern in the GFDL fully coupled ensemble initialization and decadal forecasting system. *Journal of Climate*, *26*(2), 650–661. https://doi.org/10.1175/JCLI-D-12-00231.1

Yashayaev, I., & Loder, J. W. (2017). Further intensification of deep convection in the Labrador Sea in 2016. *Geophysical Research Letters*, *44*, 1429–1438. https://doi.org/10.1002/2016GL071668

Yeager, S. (2015). Topographic coupling of the Atlantic overturning and gyre circulations. *Journal of Physical Oceanography*, *45*, 1258–1284. https://doi.org/10.1175/JPO-D-14-0100.1

Yeager, S., & Danabasoglu, G. (2014). The origins of late-twentieth-century variations in the large-scale North Atlantic circulation. *Journal of Climate*, *27*, 3222–3247. https://doi.org/10.1175/JCLI-D-13-00125.1

Yeager, S., Karspeck, A., Danabasoglu, G., Tribbia, J., & Teng, H. (2012). A decadal prediction case study: Late twentieth-century north Atlantic ocean heat content. *Journal of Climate*, *25*(15), 5173–5189. https://doi.org/10.1175/jcli-d-11-00595.1

Yeager, S. G., Danabasoglu, G., Rosenbloom, N. A., Strand, W., Bates, S. C., Meehl, G. A., et al. (2018). Predicting Near-Term Changes in the Earth System: A large ensemble of initialized decadal prediction simulations using the community earth system model. *Bulletin of the American Meteorological Society*, *99*, 1867–1886. https://doi.org/10.1175/BAMS-D-17-0098.1

Yeager, S. G., Karspeck, A. R., & Danabasoglu, G. (2015). Predicted slowdown in the rate of Atlantic sea ice loss. *Geophysical Research Letters*, *42*(10). https://doi.org/10.1002/2015GL065364

Yeager, S. G., & Robson, J. I. (2017). Recent progress in understanding and predicting atlantic decadal climate variability. *Current Climate Change Report*, *3*, 112–127. https://doi.org/10.1007/s40641-017-0064-z

Zhang, J., & Zhang, R. (2015). On the evolution of Atlantic meridional overturning circulation fingerprint and implications for decadal predictability in the North Atlantic. *Geophysical Research Letters*, *42*, 5419–5426. https://doi.org/10.1002/2015GL064596

Zhang, L., Delworth, T. L., Cooke, W., Goosse, H., Bushuk, M., Morioka, Y., & Yang, X. (2021). The dependence of internal multidecadal variability in the southern ocean on the ocean background mean state. *Journal of Climate*, *34*(3), 1061–1080. https://doi.org/10.1175/jcli-d-20-0049.1

Zhang, L., Delworth, T. L., Cooke, W., & Yang, X. (2019). Natural variability of Southern Ocean convection as a driver of observed climate trends. *Nature Climate Change*, *9*(1), 59–65. https://doi.org/10.1038/s41558-018-0350-3

Zhang, L., Delworth, T. L., Yang, X., Gudgel, R. G., Jia, L., Vecchi, G. A., et al. (2017). Estimating Decadal Predictability for the Southern Ocean Using the GFDL CM2.1 Model. *Journal of Climate*, *30*, 5187–5203. https://doi.org/10.1175/JCLI-D-16-0840.1

Zhang, R., & Delworth, T. L. (2006). Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes. *Geophysical Research Letters*, *33*, L17712. https://doi.org/10.1029/2006GL026267

Zhang, R., Sutton, R., Danabasoglu, G., Kwon, Y.-O., Marsh, R., Yeager, S. G., et al. (2019). A review of the role of the Atlantic Meridional overturning circulation in Atlantic multidecadal variability and associated climate impacts. *Reviews of Geophysics*, *57*, 316–375. https://doi.org/10.1029/2019RG000644

Zhang, S., Harrison, M. J., Rosati, A., & Wittenberg, A. T. (2007). System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Monthly Weather Review*, *135*(10). https://doi.org/10.1175/MWR3466.1

Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., et al. (2018a). The GFDL global atmosphere and land model AM4.0/LM4.0: 1. Simulation characteristics with prescribed SSTs. *Journal of Advances in Modeling Earth Systems*, *10*, 691–734. https://doi.org/10.1002/2017MS001208

Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., et al. (2018b). The GFDL global atmosphere and land model AM4.0/LM4.0: 2. Model description, sensitivity studies, and tuning strategies. *Journal of Advances in Modeling Earth Systems*, *10*, 735–769. https://doi.org/10.1002/2017ms001209