

**RESEARCH ARTICLE**

10.1029/2019MS001895

**Special Section:**

Geophysical Fluid Dynamics Laboratory CMIP6 Models

Rich Gudgel retired at the end of March 2019.  
Shian-Jiann Lin retired in May 2019.

**Key Points:**

- Development and performance of the next generation GFDL seasonal to decadal prediction model is documented
- The response of this model to realistic radiative forcing changes is shown via a large ensemble of climate simulations for 1921–2100
- The influence of the Antarctic surface energy balance on the world ocean was crucial in model development as shown via sensitivity tests

**Supporting Information:**

- Supporting Information S1

**Correspondence to:**

T. L. Delworth,  
tom.delworth@noaa.gov

**Citation:**

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>

Received 11 SEP 2019

Accepted 29 JAN 2020

Accepted article online 31 JAN 2020

## SPEAR: The Next Generation GFDL Modeling System for Seasonal to Multidecadal Prediction and Projection

Thomas L. Delworth<sup>1</sup> , William F. Cooke<sup>1,2</sup> , Alistair Adcroft<sup>1,3</sup> , Mitchell Bushuk<sup>1,2</sup> , Jan-Huey Chen<sup>1,2</sup> , Krista A. Dunne<sup>4</sup> , Paul Ginoux<sup>1</sup> , Richard Gudgel<sup>1</sup> , Robert W. Hallberg<sup>1</sup> , Lucas Harris<sup>1</sup> , Matthew J. Harrison<sup>1</sup> , Nathaniel Johnson<sup>1,3</sup> , Sarah B. Kapnick<sup>1</sup> , Shian-Jiann Lin<sup>1</sup>, Feiyu Lu<sup>1,3</sup> , Sergey Malyshev<sup>1</sup> , Paul C. Milly<sup>4</sup> , Hiroyuki Murakami<sup>1,2</sup> , Vaishali Naik<sup>1</sup> , Salvatore Pascale<sup>1,3,5</sup> , David Paynter<sup>1</sup> , Anthony Rosati<sup>1,2</sup>, M.D. Schwarzkopf<sup>1</sup> , Elena Shevliakova<sup>1</sup>, Seth Underwood<sup>1</sup> , Andrew T. Wittenberg<sup>1</sup> , Baoqiang Xiang<sup>1,3</sup> , Xiaosong Yang<sup>1,2</sup> , Fanrong Zeng<sup>1</sup> , Honghai Zhang<sup>1,3,6</sup> , Liping Zhang<sup>1,2</sup> , and Ming Zhao<sup>1</sup>

<sup>1</sup>Geophysical Fluid Dynamics Laboratory, NOAA, Princeton, NJ, USA, <sup>2</sup>University Corporation for Atmospheric Research, Boulder, CO, USA, <sup>3</sup>Department of Geosciences, Princeton University, Princeton, NJ, USA, <sup>4</sup>Integrated Modeling and Prediction Division, U.S. Geological Survey, Princeton, NJ, USA, <sup>5</sup>Department of Earth System Science, Stanford University, Stanford, CA, USA, <sup>6</sup>Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA

**Abstract** We document the development and simulation characteristics of the next generation modeling system for seasonal to decadal prediction and projection at the Geophysical Fluid Dynamics Laboratory (GFDL). SPEAR (Seamless System for Prediction and EArth System Research) is built from component models recently developed at GFDL—the AM4 atmosphere model, MOM6 ocean code, LM4 land model, and SIS2 sea ice model. The SPEAR models are specifically designed with attributes needed for a prediction model for seasonal to decadal time scales, including the ability to run large ensembles of simulations with available computational resources. For computational speed SPEAR uses a coarse ocean resolution of approximately 1.0° (with tropical refinement). SPEAR can use differing atmospheric horizontal resolutions ranging from 1° to 0.25°. The higher atmospheric resolution facilitates improved simulation of regional climate and extremes. SPEAR is built from the same components as the GFDL CM4 and ESM4 models but with design choices geared toward seasonal to multidecadal physical climate prediction and projection. We document simulation characteristics for the time mean climate, aspects of internal variability, and the response to both idealized and realistic radiative forcing change. We describe in greater detail one focus of the model development process that was motivated by the importance of the Southern Ocean to the global climate system. We present sensitivity tests that document the influence of the Antarctic surface heat budget on Southern Ocean ventilation and deep global ocean circulation. These findings were also useful in the development processes for the GFDL CM4 and ESM4 models.

**Plain Language Summary** In this paper we describe the development and simulation characteristics of a new climate model that will be used for seasonal to multidecadal climate prediction and projection. The model combines a set of newly developed components that simulate the ocean, atmosphere, land, and sea ice. We document this model by assessing its performance in simulating the current climate and by showing the model's response to changing greenhouse gases and aerosols over the 20th and 21st centuries. We also show results from a set of sensitivity experiments that were an important part of the model development process. These sensitivity tests explore connections between the surface energy balance over Antarctica and the circulation of the deep ocean.

### 1. Introduction

In this manuscript we describe the development process and simulation characteristics of a new modeling system that will form the backbone for the next generation real-time prediction and projection efforts at the Geophysical Fluid Dynamics Laboratory (GFDL), extending seamlessly from seasonal to multidecadal time scales. Real-time experimental seasonal predictions have been conducted since the early 1990s at GFDL, with real-time predictions routinely distributed to institutions such as the International Research

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Institute for Climate and Society at Columbia University. Beginning in 2015, this ongoing activity contributed model predictions to the newly formed North American Multimodel Ensemble (Kirtman et al., 2014) seasonal prediction activity. Similarly, real-time decadal predictions have been performed routinely at GFDL since 2011 (Yang et al., 2013), with model predictions currently being made available through the U.K. Met Office as a WMO Lead Centre for Interannual to Decadal Prediction (<https://www.metoffice.gov.uk/research/climate/seasonal-to-decadal/long-range/wmolc-adcp>). These activities are part of a seamless seasonal to centennial climate prediction and projection effort at GFDL, as the same models that are used for seasonal and decadal prediction have also been used for century-scale climate change projections (Held et al., 2010, 2019; Knutson et al., 2006; Merlis et al., 2014).

The two coupled climate models currently used for the real-time experimental predictions are CM2.1 (Delworth et al., 2006) and FLOR (Vecchi et al., 2014). FLOR is derived from CM2.1 but uses a higher-resolution atmospheric component (50-km horizontal grid spacing in FLOR versus 200 km in CM2.1). Both coupled systems use an ocean model with a horizontal resolution of approximately 1°, with refined meridional resolution in the Tropics. The seasonal to decadal predictions are initialized using the output of an Ensemble Coupled Data Assimilation system (S. Zhang et al., 2007). The seasonal predictions are computed in real time each month, using all data from the previous month.

The fundamentals of the CM2.1 and FLOR models were developed more than 10 years ago. Since then there has been substantial component model development and improvement at GFDL. These new developments provide the rationale and basis for creating a new system for seasonal to multidecadal predictions and projections at GFDL that takes advantage of these advances. These developments include the adoption of a new dynamical core in the atmosphere with regional nesting capabilities (Harris & Lin, 2013), revised atmospheric physics (Zhao et al., 2018a, 2018b), the new MOM6 ocean code and SIS2 sea ice code (Adcroft et al., 2019), and a substantially enhanced land model. These new developments will enable new capabilities and advances. For example, the new dynamical core will allow for the use of variable resolution grids, thereby allowing future versions of SPEAR to use high spatial resolution in focused regions for the simulation and prediction of critical phenomena, such as hurricanes or other extremes. The MOM6 ocean code includes many improvements, such as the use of hybrid vertical coordinates, that may assist in the simulation of important oceanic processes potentially relevant to decadal prediction, such as overflows. A much improved representation of land hydrology will enable the simulation and prediction of extremes in river flows.

This manuscript describes the development process and simulation characteristics of this new seasonal to multidecadal prediction and projection system, which we call SPEAR (Seamless System for Prediction and EArth System Research). The SPEAR models share many components with the recently developed GFDL CM4 (Held et al., 2019) model but with configuration and physical parameterization choices in SPEAR geared toward physical climate prediction and projection on seasonal to multidecadal time scales. Table 1 places SPEAR within the context of a series of models developed at GFDL over the last 15 years.

In section 2 we describe the model components and the simulations conducted. We present in section 3 (and accompanying supporting information) a suite of diagnostics to illustrate model characteristics in control simulations, as well as the model response to idealized and realistic radiative perturbations. One aspect of the model development process that deserves special attention, which we focus on section 4, has to do with the role of the Southern Ocean (SO) in the global climate system and its influence on model bias and drift. We illustrate aspects of the development process that focus on simulating the SO and its global-scale climatic impacts through the design and analysis of sensitivity experiments that illustrate connections between continental climate over Antarctica, the SO, and model bias and drift. A summary is presented in section 5.

## 2. Model Description

We use a newly coupled ocean-atmosphere-land-sea ice modeling system (SPEAR) that was developed as a seamless climate prediction system capable of making predictions and projections from seasonal to centennial time scales. The Earth's climate system is seamless and unified, with short-term predictable variations, such as El Niño–Southern Oscillation (ENSO) or the Madden–Julian Oscillation (MJO), occurring at the same time as the system responds to changing greenhouse gases, aerosols, and other forms of radiative forcing changes. Further, changing radiative forcing can in turn alter the characteristics of internal variability

**Table 1**  
*Characterization and Chronology of Selected GFDL Climate Models Developed Since Approximately 2004*

Name	General description	Model grids	Primary use	Reference
Coupled model generation 4				
CM4	Coupled ocean-atmosphere model with advanced physics relative to CM3 (see below) in all components (atmosphere, land, ocean, sea ice); ocean uses hybrid vertical coordinate AM4, atmospheric component of CM4.	Atm & Land: 1°, 33 atm levels Ocn & Sea Ice: 0.25°, 75 ocn levels	Climate change	Held et al. (2019)
	OM4, ocean component of CM4			Zhao et al. (2018a)
				Adcroft et al. (2019)
SPEAR	Same atmosphere/land components as CM4, but ocean and ice have coarser horizontal resolution for lower computational cost; ocean uses hybrid vertical coordinate	Atm & Land: SPEAR_LO: 1°, 33 atm levels SPEAR_MED: 0.5°, 33 atm levels Ocn & Sea Ice: 1°, tropical refinement to 0.3°, 75 ocn levels	Seasonal to decadal prediction, multidecadal climate change, data assimilation	Delworth et al. (2020); (this paper)
ESM4.1	Similar to CM4 but incorporates comprehensive biogeochemical processes; relative to the atmosphere in CM4 and SPEAR, the ESM4 atmosphere has a more complete representation of atmospheric chemistry and a better resolved stratosphere	Atm & Land: 1°, 49 atm levels Ocn & Sea Ice: 0.5°, tropical refinement to 0.3°, 75 ocn levels	Climate change	
SHiELD	Very high resolution nonhydrostatic atmosphere model with integrated physics and dynamics	25, 13, or 3 km, uniform or refined grid; 63 or 91 atm levels	Weather to S2S prediction research	Early version (fvGFS): Zhou et al. (2019); Chen et al. (2019)
Coupled model generation 3				
CM3	Coupled ocean-atmosphere model; atmosphere component incorporates advanced physics, aerosol and chemistry representations relative to CM2.1, ocean has improved physics and numerics	Atm & Land: 2°, 49 atm levels Ocn & Sea Ice: 1°, tropical refinement to 0.3°, 50 ocn levels	Climate change	Atmosphere & land in Donner et al. (2011); Ocean & sea ice in Griffies et al. (2011)
HiRAM	Global Atmosphere Model similar to CM2.1 with higher-resolution atmosphere/land and updated convection scheme and microphysics	Atm & Land: 0.5°, 0.25° 32 or 63 atm levels—63 for nonhydrostatic	Weather to seasonal research	Zhao et al. (2009); Chen and Lin (2013)
Coupled model generation 2				
CM2.1	Coupled ocean-atmosphere model	Atm & Land: 2°, 24 atm levels Ocn & Sea Ice: 1°, tropical refinement to 0.3°, 50 ocn levels	Climate change, seasonal to decadal prediction, data assimilation	Delworth et al. (2006)
	AM2.1, Atmospheric component of CM2.1			Anderson et al. (2004)
CM2.5 CM2.6	High resolution versions of CM2.1 with advanced ocean numerics	Atm & Land: 0.5°, 24 atm levels Ocn & Sea Ice: CM2.5: 0.25° CM2.6: 0.1° 50 ocn levels	Climate change, process studies	Delworth et al. (2012)
ESM2M			Climate change	Dunne et al. (2012)

**Table 1** (continued)

Name	General description	Model grids	Primary use	Reference
ESM2G	Coupled ocean-atmosphere models, similar to CM2.1, but incorporating biogeochemical processes and improved ocean numerics and physics. ESM 2G uses isopycnal coordinates in the ocean; ESM 2M uses height based vertical coordinates	Atm & Land: 2°, 24 atm levels Ocn & Sea Ice: 1°, tropical refinement to 0.3°, 50 ocn levels		
FLOR	Coupled ocean-atmosphere model; similar to CM2.1 but with higher-resolution atmosphere/land and improved ocean numerics	Atm & Land: 0.5°, 24 atm levels Ocn & Sea Ice: 1°, tropical refinement to 0.3°, 50 ocn levels	Climate change, seasonal to decadal prediction	Vecchi et al. (2014)
HIFLOR	Similar to FLOR, but with higher resolution in atmosphere/land that helps with simulation of intense hurricanes	Atm & Land: 0.25°, 24 atm levels Ocn & Sea Ice: Same as FLOR	Climate change, seasonal prediction	Murakami et al. (2016)

Note. Models are categorized as parts of “Generations” of models, with Generation 2 being the oldest (listed at bottom of table) and Generation 4 being the most recently developed (at top of table).

of the climate system (Fang et al., 2014; Zhang & Delworth, 2016b). Modeling systems need to be designed to predict the behavior of the climate system as an initial value problem, the response of the climate system to radiative forcing changes, and the influence of radiative forcing changes on the characteristics of internal variability. This latest model is part of a tightly coupled new suite of models developed recently at GFDL (see Table 1). These models share the vast majority of their code and components but are designed with specific foci in mind. The SPEAR modeling system was designed with seasonal to multidecadal prediction and projection as its focus, with an emphasis on computational efficiency to facilitate large ensembles needed for predictions and projections.

## 2.1. Atmosphere/Land Formulation

The physics in the atmospheric and land components of SPEAR are identical to those of the recently developed GFDL AM4-LM4 model (Zhao et al., 2018a, 2018b), with one exception: The near-infrared isotropic reflectance parameter for cold snow over glacial surfaces has been increased from 0.58 to 0.68 (see Milly et al. (2014) for details on the formulation for snowpack in the land component). The increased albedo, as discussed in more detail in section 4, both improves the realism of the simulated albedo over Antarctica and improves aspects of the circulation of the World Ocean related to deepwater formation in the SO and global ocean temperature drift.

We use two versions of SPEAR that differ in the horizontal resolution for the atmosphere and land. SPEAR\_LO uses an atmosphere/land resolution of approximately 100 km, and SPEAR\_MED uses an atmosphere/land resolution of approximately 50 km. The 100-km resolution of SPEAR\_LO is identical to that in Zhao et al. (2018a, 2018b), with the exception of the albedo change noted above. The physics in the 50-km version of SPEAR are identical to those in the 100-km version. The dynamical core differs only in the time steps required for the difference in resolution. The atmospheric component uses 33 levels in the vertical. We present results in sections 2 and 3 (and supporting information) using both resolutions. For a more detailed discussion of the physics of the atmosphere and land components, as well as model performance when forced by observed sea surface temperatures (SST), the reader is referred to the detailed results in Zhao et al. (2018a, 2018b). We present results below (and in the supporting information) for model performance when coupled to a dynamic ocean. Both SPEAR\_LO and SPEAR\_MED are coupled to the same 1° ocean model as described in the next section.

## 2.2. Ocean/Sea Ice Formulation

The ocean and sea ice components are constructed from the new MOM6 ocean code (<https://github.com/NOAA-GFDL/MOM6>) and SIS2 sea ice code (<https://github.com/NOAA-GFDL/SIS2>). The ocean and sea ice components use a nominal horizontal resolution of 1° with refinement to 1/3° in the meridional direction in the Tropics. Since the model grid maintains an aspect ratio close to 1, the effective resolution increases poleward with the convergence of the meridians. The zonal grid spacing reduces from 110 km in the deep tropics to 46 km at 65° latitude. The meridional grid spacing is 40 km in the deep tropics (this equatorial refinement is important for representing tropical variability), increases to 105 km in the subtropics, and decreases to 46 km at 65° latitude in each hemisphere. In order to maintain approximately uniform grid spacing over the Arctic, a tripolar grid is used with poles over northern Eurasia, northern Canada, and Antarctica (Murray, 1996). North of 65°N, there is a nonuniform grid with cells of approximately equal area (approximately 46 km by 46 km) covering the polar cap. The ocean model has 75 layers in the vertical, with layer thickness as fine as 2 m near the surface, including 30 layers in the top 100 m. The vertical grid spacing becomes larger with depth, reaching 250 m below 5,000 m.

The ocean resolution employed in SPEAR is relatively coarse. This is a deliberate choice to facilitate the rapid and efficient execution of ensembles of simulations needed for climate prediction given available computer resources. Based on past experiences with FLOR, we also decided to employ higher resolution in the atmosphere (as fine as 50 km with planned refinement to 25 km) in order to simulate important phenomena such as tropical storms, other extremes, and regional hydroclimate. We plan to explore future versions of SPEAR with enhanced ocean resolution as resources permit.

The physics used in the ocean and sea ice components is very similar to that documented in Adcroft et al. (2019) (hereafter called OM4, the ocean component of CM4), including the use of a hybrid vertical coordinate in the ocean. The hybrid vertical coordinate is a function of height in the upper ocean, transitioning to isopycnal layers in the interior (see Adcroft et al., 2019, for further details). The use of the hybrid vertical coordinate provides a more realistic solution than model versions in our development process that used a purely level-based coordinate system; improvements include a reduced surface temperature bias and less subsurface temperature drift. Both models (SPEAR and OM4/CM4) use the same 75 vertical levels. There are, however, a few key differences reflecting the differences in horizontal resolution between OM4, which uses a 0.25° horizontal grid spacing, versus the 1° grid spacing used in the SPEAR model. First, the SPEAR model is not able to resolve oceanic mesoscale eddies and thus uses a parameterization of their effects. Spatially varying model lateral diffusivities are internally diagnosed based on a mesoscale energy budget. The production and dissipation of subgrid-scale mesoscale kinetic energy balanced by the release of resolved available potential energy are parameterized by an eddy overturning stream function (Jansen et al., 2015; Marshall & Adcroft, 2010). Additionally, predicted eddy velocity scales are combined with an eddy mixing length scale to calculate tracer mixing rates along neutral density surfaces. Second, similar to OM4, the SPEAR ocean component uses a parameterization of the effects of submesoscale eddies (Adcroft et al., 2019). Relative to OM4, different parameter coefficients were used in SPEAR based upon the results of sensitivity tests and the impact on SST bias. Third, the MOM6 code has the capability to represent flow through channels that are narrower than the grid spacing of the model. The locations where this functionality is needed are dependent on the model grid, and this was set up for SPEAR based on the grid resolution at 1°. These include regions with narrow channels such as occur in the Indonesian Throughflow, the Gibraltar Straits, and the Bosphorus Strait. This parameterization was tuned to produce flow through the channels similar in total transport to observed. Fourth, we found empirically an improvement in some aspects of the model solution by adding extra horizontal viscosity poleward of 50° in each hemisphere. Highly varying topography and a declining Rossby radius of deformation at high latitudes are challenges for a coarser-resolution model, and our experience was that the extra viscosity provided a better solution.

## 2.3. Observational Data

We use observational data sets for model assessment. For SST we use NOAA Extended Reconstructed Sea Surface Temperature version 4 (ERSSTv4) data (Huang et al., 2015) provided by NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at <https://www.esrl.noaa.gov/psd/>. For surface air (surface

air temperature over land and SST over the ocean) we used two data sets. The first (CRUTEM4) was developed at the Climatic Research Unit (Jones et al., 2012) and downloaded 26 April 2019, from NOAA's Earth System Research Laboratory (<https://www.esrl.noaa.gov/psd/data/gridded/data.crutem4.html>). The second (GISTEMP v4) is from the National Aeronautics and Space Administration's Goddard Institute for Space Sciences (Lennsen et al., 2019) and was downloaded on 30 July 2019 (<https://data.giss.nasa.gov/gistemp/>). For sea ice extent we used observational data based on passive microwave satellite information (Cavalieri et al., 1996) obtained from the National Snow and Ice Data Center derived from brightness temperature measured by satellite using the National Aeronautics and Space Administration Team algorithm. Atmospheric circulation features were compared to reanalyses (Dee et al., 2011; Rienecker et al., 2011). Precipitation observations are described in Schneider et al. (2017).

## 2.4. Simulations Conducted

### 2.4.1. Control Simulations

We conduct extended control simulations with atmospheric composition fixed at levels representative of either calendar year 1850 (a “preindustrial” control) or calendar year 2010 (a “modern” control). In each case the model ocean is initialized from an observed late twentieth century ocean climatology (Boyer et al., 2013). We conduct control simulations using both SPEAR\_LO and SPEAR\_MED. As shown below, in the 1850 control simulations the atmosphere and ocean cool over the first century as the system adjusts to 1850 radiative forcing. The initial conditions come from a late twentieth century observed state with a positive net radiative imbalance (Allan et al., 2014), whereas 1850 radiative conditions have a radiative imbalance closer to zero. This difference in radiative forcing would be expected to lead to cooler conditions for the 1850 control simulations than observed in recent decades. We show key aspects of the model solutions to provide an overall sense of model fidelity in simulating the climate system, including climate drifts. We also show additional aspects of the model solution in the supporting information. For both SPEAR\_LO and SPEAR\_MED the 2010 control simulations are 100 years in length, whereas the 1850 control simulations are 2,000 years in length.

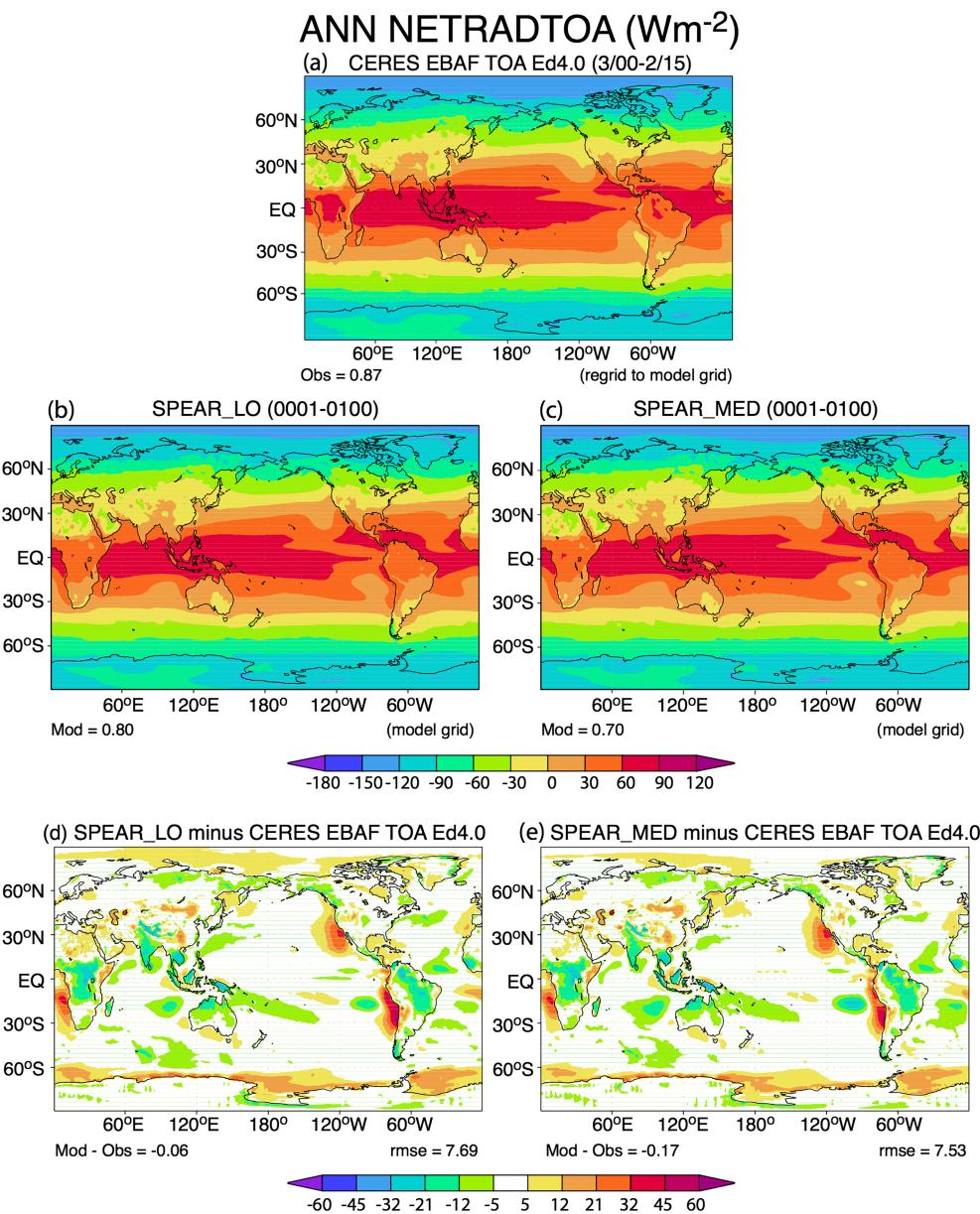
### 2.4.2. Idealized Climate Change Simulations

We also conduct idealized climate change experiments. Starting from arbitrary points in the 1850 control simulations, we increase atmospheric CO<sub>2</sub> at 1% per year until it reaches double the initial value after approximately 70 years and is held fixed thereafter until the end of the experiment at model year 200. We use these simulations to evaluate the Transient Climate Response for SPEAR\_LO and SPEAR\_MED.

### 2.4.3. Ensembles of Simulations With Historical and Future Projected Radiative Forcing Changes

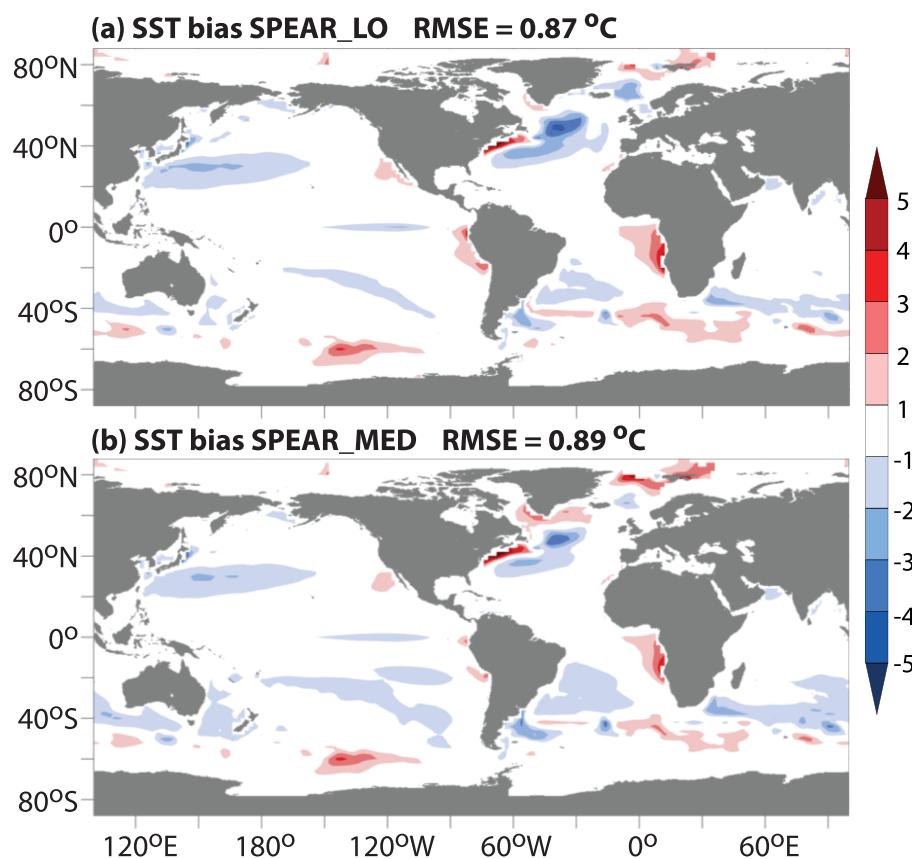
We also conduct 30-member ensembles of simulations with prescribed time-evolving changes in atmospheric composition and external radiative forcing. These simulations are initialized from conditions in the 1850 control simulations that are 20 years apart—new historical ensembles are started from the restart files from the control simulations at years 101, 121, 141, and every 20 years thereafter until year 681. All ensemble members share the same land model initial condition which was created by running a 300-year spin-up run to allow for land use transitions followed by a short 1-year run to create 1851 (SPEAR\_LO) and 1921 (SPEAR\_MED) initial conditions. Separate 30-member ensembles are conducted with SPEAR\_LO and SPEAR\_MED (each is initialized from its respective 1850 control simulation). The SPEAR\_LO simulations extend from calendar years 1851 to 2100, whereas the SPEAR\_MED simulations extend from calendar years 1921–2100 (shorter experiments were performed with SPEAR\_MED due to the larger computational expense of the higher-resolution model). The two 30-member ensembles comprise 12,900 simulated model years.

The ensemble simulations described here are driven by boundary conditions and natural and anthropogenic forcings developed in support of CMIP6 (Eyring et al., 2016) and available at the Earth System Grid Federation (<https://esgf-node.llnl.gov/search/input4mips/>). Historical forcings are applied for the time period 1850 to 2014, while projections for the Shared Socioeconomic Pathway 5–8.5 (SSP5–8.5) (Kriegler et al., 2017; Riahi et al., 2017) are applied for the 2015–2100 time period except where indicated. Here we provide brief descriptions of the forcings with references for more details.



**Figure 1.** Annual mean net radiative flux at the top of the atmosphere ( $\text{W m}^{-2}$ ) for (a) observations (Loeb et al., 2009), calculated as time mean from March 2000 to February 2015, (b) 100-year mean from SPEAR\_LO control simulation, and (c) 100-year mean from SPEAR\_MED control simulation. (d) Model biases calculated as SPEAR\_LO minus observations. (e) Model biases calculated as SPEAR\_MED minus observations.

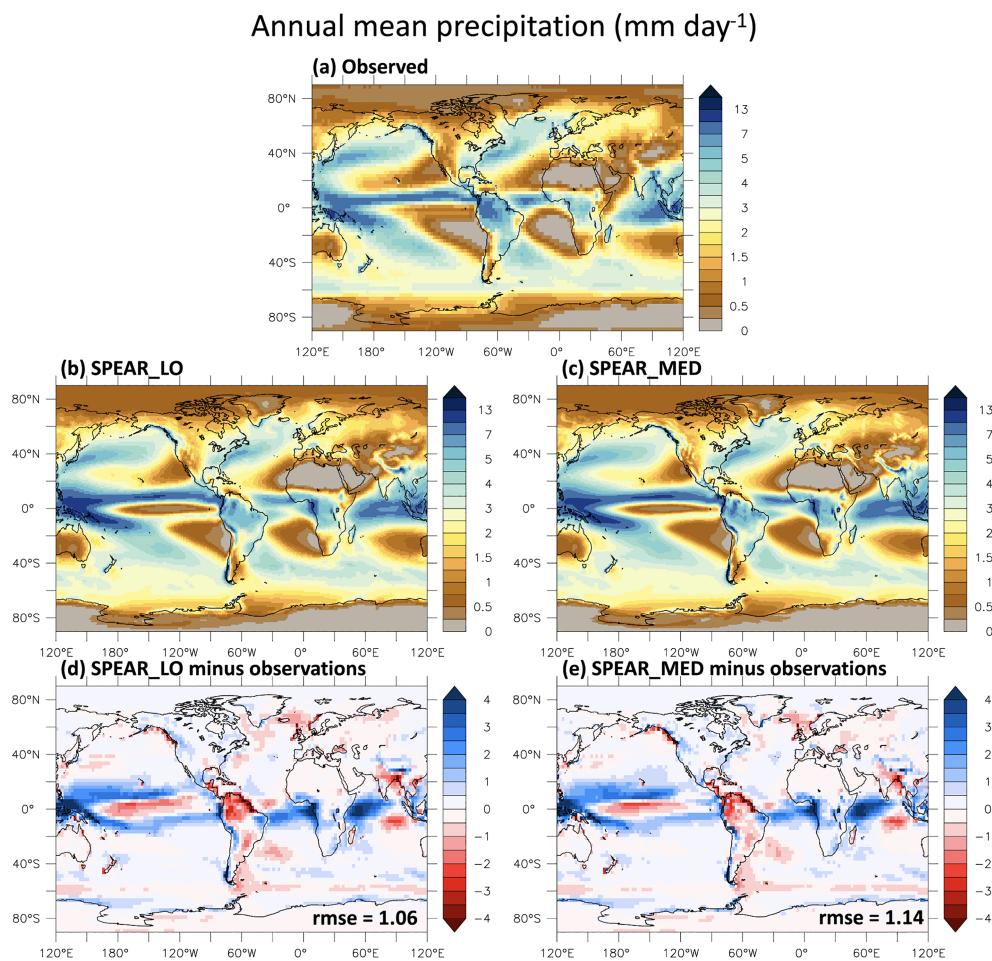
Historical reconstructions and future evolution of monthly solar irradiances are from Matthes et al. (2017) for the simulations over years 1851 to 2100. Global monthly mean concentrations of greenhouse gases, including carbon dioxide ( $\text{CO}_2$ ), methane ( $\text{CH}_4$ ), and nitrous oxide ( $\text{N}_2\text{O}$ ), and ozone-depleting substances (including CFC-11, CFC-12, CFC-113, and HCFC-22) are from Meinshausen et al. (2017) for the historical period and from Kriegler et al. (2017) for the future. Monthly mean ozone concentrations for the historical and future time period are from the CMIP6 forcing data set. Annually varying time series of monthly anthropogenic and biomass burning emissions of carbonaceous aerosols and sulfur dioxide ( $\text{SO}_2$ ) precursor to sulfate aerosols are from the Community Emissions Data System, CEDS (Hoesly et al., 2018) and the data set of van Marle et al. (2017), respectively, for the historical period and from Gidden et al. (2019) for SSP5-85.



**Figure 2.** Model bias in annual mean sea surface temperature (SST, units are in  $^{\circ}\text{C}$ ), calculated as annual mean SST in model minus annual mean of observed SST. For the model, means were computed over years 21–100 of a control simulation with atmospheric composition corresponding to calendar year 2010. For observations we use SST from ERSST (Huang et al., 2015) and compute means over years 1981–2010. (a) SPEAR\_LO minus observations. (b) SPEAR\_MED minus observations.

As with AM4 and CM4, we specify time series of stratospheric aerosol optical properties including contributions from both volcanoes and other natural and anthropogenic contributions (Thomason et al., 2018). For the period 2025–2100 we specify stratospheric aerosol properties as the time mean used over the 1850–2014 period. During the period 2015–2024 we linearly interpolate from the historical values in 2014 to the time mean values specified over the period 2025–2100. The contribution from continuously degassing and explosive volcanoes to tropospheric aerosols is treated in the same way as in AM3 (Donner et al., 2011) and AM4 (Zhao et al., 2018a).

Tropospheric aerosols in SPEAR are simulated from emissions with a simplified chemical mechanism as implemented in the AM4 model (Zhao et al., 2018a). A brief description of the representation of aerosols is provided here with more details given by Zhao et al. (2018a). Mass distribution of five aerosol types including sulfate, dust, black carbon, organic carbon, and sea salt is simulated in SPEAR. The size distribution for all aerosols is prescribed as lognormal except for dust and sea salt, which are discretized into five size bins from 0.1- to 10- $\mu\text{m}$  radius. Their concentrations are calculated based on their emissions (and precursor emissions), chemical production for sulfate and secondary organics, dry and wet (rainout and washout) deposition, transport by advection, and dry and wet convection. Chemical production of sulfate is from the oxidation of sulfur dioxide ( $\text{SO}_2$ ) in the gas phase by hydroxyl radical ( $\text{OH}$ ) and in the aqueous phase by ozone and hydrogen peroxide ( $\text{H}_2\text{O}_2$ ). Since ozone chemistry is not explicitly represented in the model, the monthly climatology of ozone and other oxidants ( $\text{OH}$ ,  $\text{HO}_2$ , and  $\text{NO}_3$ ) used by sulfate chemistry is prescribed based on a 20-year simulation of AM3 (Naik et al., 2013) that applied a comprehensive mechanism of ozone chemistry (Horowitz et al., 2003). Emissions of natural aerosols including dust, sea salt, and dimethyl

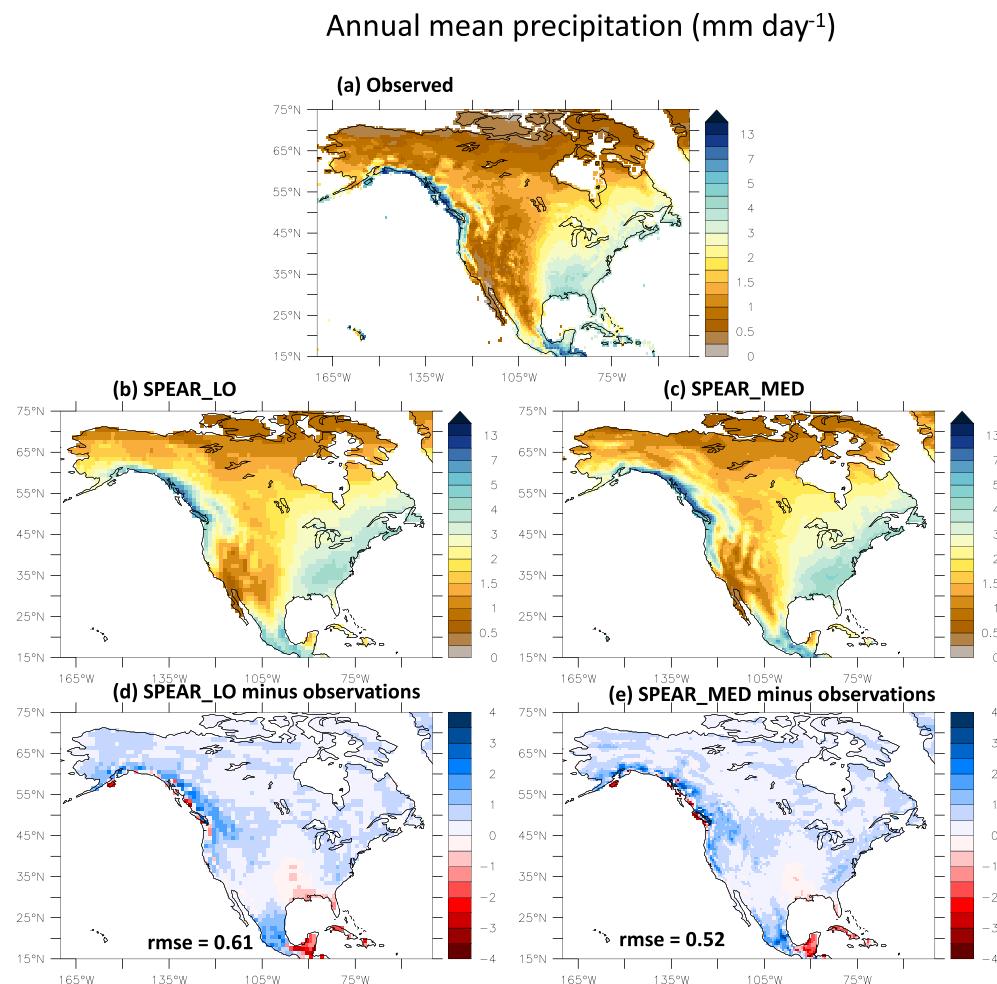


**Figure 3.** Time mean, annual mean precipitation (units are in  $\text{mm day}^{-1}$ ). (a) Observations from the Global Precipitation Climatology Project (Adler et al., 2003), calculated as the mean from 1979–2017. (b) Results from SPEAR\_LO, calculated as the time mean over years 21–100 of a 2010 control simulation. (c) Same as (b) using SPEAR\_MED, (d) precipitation bias, calculated as SPEAR\_LO minus observations, and (e) precipitation bias, calculated as SPEAR\_MED minus observations. Root-mean-square error (rmse) listed at bottom right of (d) and (e).

sulfide (a precursor for sulfate) are calculated interactively based on parameterizations (see Zhao et al., 2018a for more details).

Time-varying land surface properties are also specified. The model represents land surface heterogeneity as a mosaic of subgrid-scale tiles with distinct physical and ecological properties. Each tile has distinct energy and moisture balances for a vegetation-snow-soil column, biophysical properties, and exchanges of radiant and turbulent fluxes with the overlaying atmosphere. The land model predicts physical, biogeochemical, and ecological characteristics for each subgrid land surface tile from the top of the vegetation canopy to the bottom of the soil column including leaves and canopy air temperature, canopy-air-specific humidity, stomatal conductance, snow cover and depth, runoff, vertical distribution of soil moisture, ice, and temperature.

The land use history is prescribed using LUH2 reconstruction (<http://luh.umd.edu>) for each grid cell in terms of annual transition rates among four distinct land use types: undisturbed (natural), crops, pastures, and secondary vegetation. The transition rates are applied annually by splitting the areas undergoing land use change from existing subgrid tiles, applying appropriate conversion to those areas as necessary (e.g., clear cutting for natural-to-crop transition or logging for transitions describing forest harvesting), and merging the new land tiles into the existing tile structure. Croplands, pastures, and natural vegetation are each assigned one tile per grid cell, but there may exist several secondary vegetation tiles to keep track of the age



**Figure 4.** Same as Figure 3 but focused on North America.

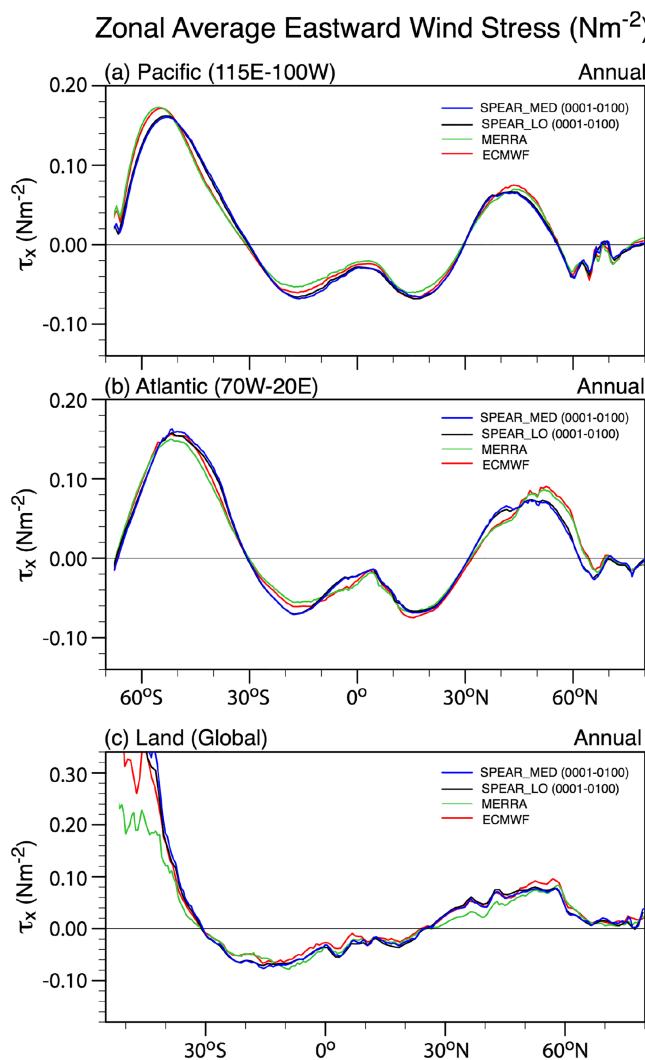
distribution of secondary lands. The transition process possibly increases the number of subgrid tiles, by creating areas of new secondary vegetation after agricultural abandonment or forest logging; secondary tiles in a grid cell with similar biological and physical characteristics are merged. All transitions and conversions conserve energy, water, and carbon mass.

The transition-based approach employed by the model eliminates one of the uncertainties of land use and land cover change implementation: The history of transitions among all land use type reflects paths of changes not just net effect. In particular, it allows the model to keep track of the age structure (and associated land surface properties) of secondary vegetation.

### 3. Simulation Characteristics

#### 3.1. Time Mean Climate

We show in Figure 1 the annual mean net radiation balance at the top of the atmosphere from observations and the model simulations, calculated over years 1–100 of the 2010 Control simulations for SPEAR\_LO and SPEAR\_MED. The overall agreement is encouraging, with global mean biases of less than  $0.2 \text{ W m}^{-2}$ . The largest errors are (a) positive biases off the west coast of continents in the tropics and subtropics (South America, North America, and Africa) where stratocumulus cloud decks are not well simulated, (b) positive biases off the coast of Antarctica, also associated with difficulties in simulating cloud cover, and (c) negative biases over some tropical continental locations. The global mean,



**Figure 5.** Annual mean, zonal mean zonal wind stress ( $\text{N m}^{-2}$ ) for (a) Pacific, (b) Atlantic, and (c) global land regions. Observations are from ERA-Interim reanalysis (Dee et al., 2011) and MERRA reanalysis (Rienecker et al., 2011).

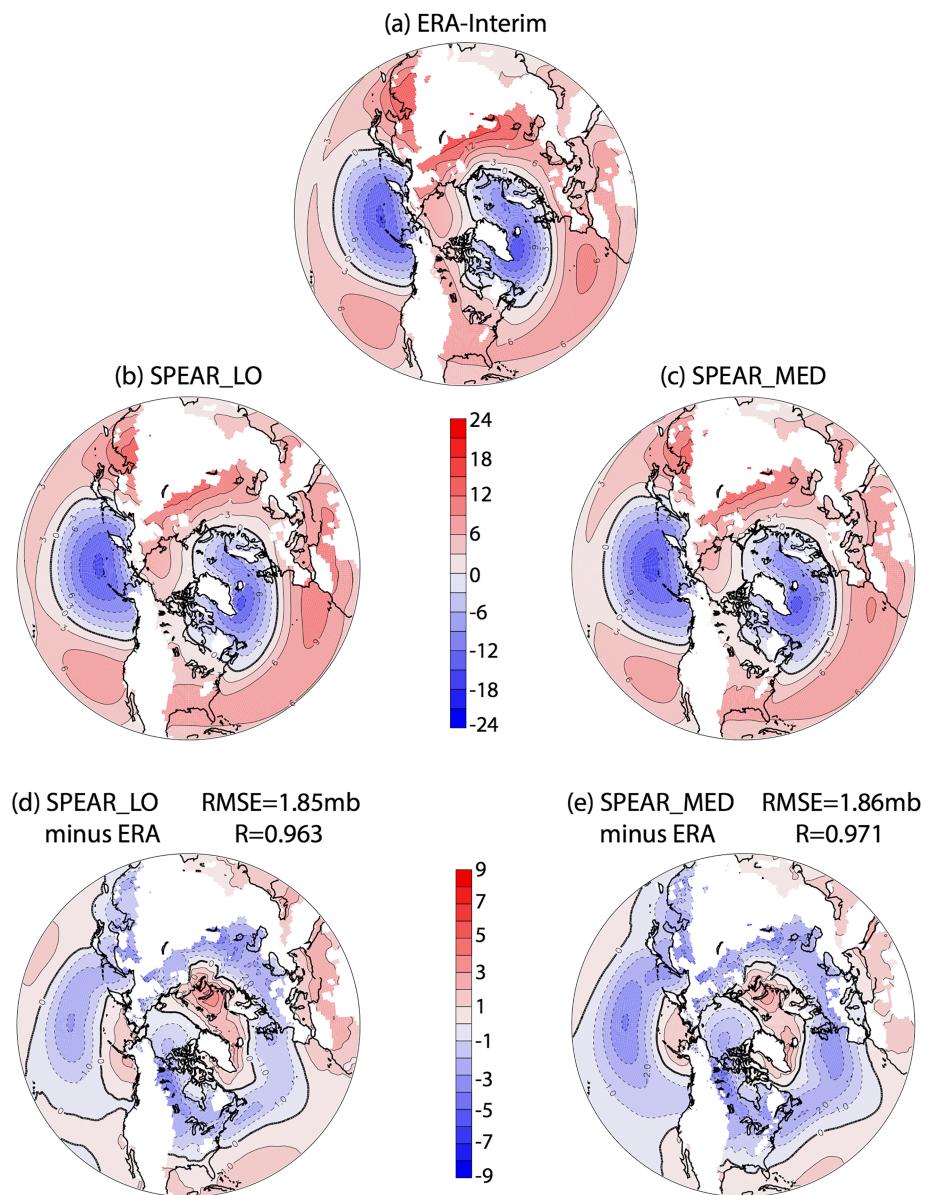
the simulation of orographically influenced precipitation over western North America, including regional-scale dry zones in the lee of mountainous areas.

The simulation of zonal winds is very important in driving ocean circulations, and simulated winds are shown in Figure 5. The top panel shows the zonal mean of the near-surface zonal wind stress over the Pacific Ocean sector; common to many models, the zonal winds in the middle and higher latitudes of the Southern Hemisphere are shifted equatorward in the models relative to their locations in observations-based reanalyses (Dee et al., 2011; Rienecker et al., 2011), and see also Figure 9.19 of Flato et al. (2013). This bias is not as readily apparent in the Atlantic sector (middle panel), although simulated zonal winds in the middle and higher latitudes of the Northern Hemisphere also have an equatorward shift relative to observations. Over continental regions (bottom panel) the locations and amplitudes of the simulated winds are in good agreement with observational estimates. A further perspective on simulated atmospheric circulation is shown in Figure 6 which shows sea level pressure for the months of December–February over the Northern Hemisphere. Both models have a tendency for southward displacements of the Icelandic and Aleutian Lows relative to observations, consistent with the zonal wind stress shown previously.

annual mean net radiation at the top of the atmosphere (averaged over years 1–100) is  $0.80 \text{ W m}^{-2}$  in SPEAR\_LO and  $0.70 \text{ W m}^{-2}$  in SPEAR\_MED, in good agreement with recent observational estimates (Johnson et al., 2016; Loeb et al., 2009, 2012). These positive imbalances reflect the effect of increased  $\text{CO}_2$  and other radiative forcing agents in calendar year 2010 relative to a longer-term steady state (the radiative balance associated with preindustrial conditions is discussed below).

Figure 2 shows the annual mean SST bias for SPEAR\_LO and SPEAR\_MED in their 2010 control simulations. The bias is calculated as the annual mean SST in the model (over years 21–100 of the 2010 control simulation) minus observed SST (Huang et al., 2015) over the period 1981–2010. The biases are relatively small and generally similar between the two model versions. Positive anomalies in the North Atlantic near the North American coast and negative anomalies to the east of Newfoundland are associated with difficulties in simulating the Gulf Stream and its separation from the North American coast. This is a typical bias in models that do not explicitly resolve oceanic mesoscale eddies; see Figure 9.2 of Flato et al. (2013). Positive biases off the west coast of Africa are associated with difficulties in simulating marine stratocumulus decks and regional atmospheric circulation, as reflected in the errors in the net radiation balance (see Figure 1). Both problems are long-standing biases in many climate models (see, e.g., Figure 9 of Zhao et al., 2018a). The bias pattern in SST over the rest of the global ocean is fairly modest, with a global root-mean-square error (rmse) of  $0.87^\circ\text{C}$  in SPEAR\_LO and  $0.89^\circ\text{C}$  in SPEAR\_MED. These are much reduced from the rmse values in the previous generation of GFDL prediction models:  $1.16^\circ\text{C}$  in CM2.1 and  $1.36^\circ\text{C}$  in FLOR.

The annual mean precipitation is shown in Figure 3 for both observations and models. While the overall patterns are realistic, there is still a tendency for a double Intertropical Convergence Zone in the tropical Pacific, along with excessive rainfall in the tropical Atlantic and the western tropical Indian Ocean. We examine in more detail the ability of the model to simulate regional precipitation patterns over North America in Figure 4. The patterns are realistic, but there is an overall wet bias, with more simulated precipitation than observed over North America. As expected, the finer resolution of SPEAR\_MED provides more details in

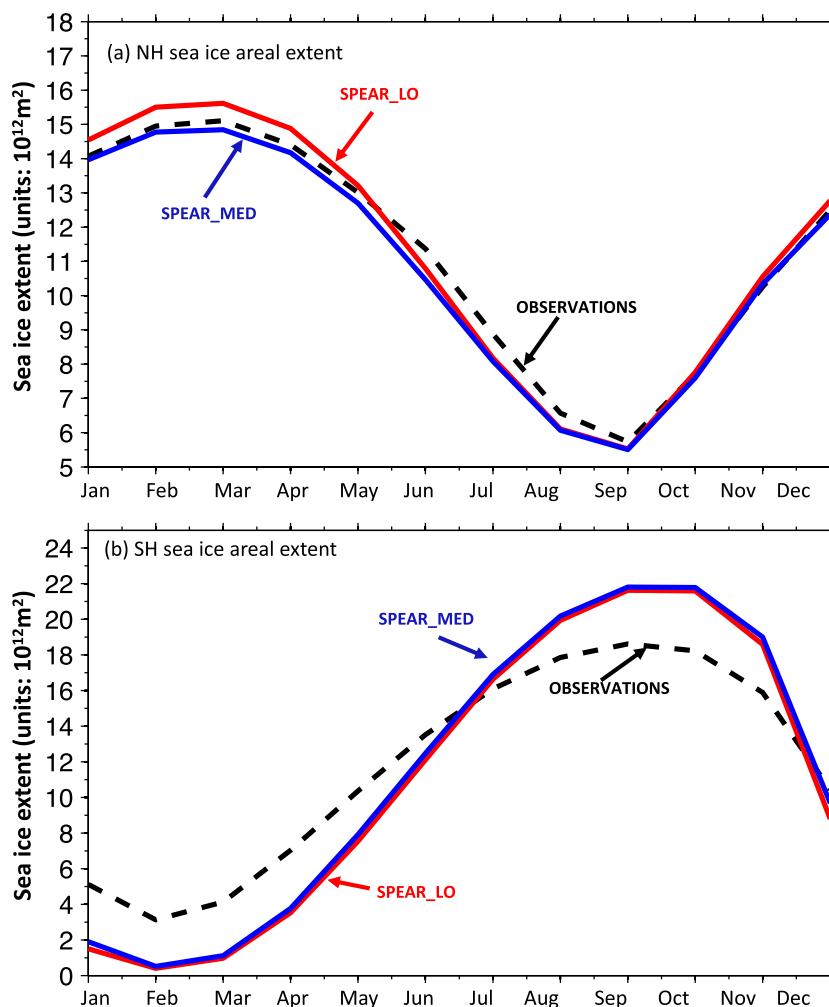


**Figure 6.** Long-term Northern Hemisphere seasonal mean (December–February) sea level pressure (SLP). Units are in hPa, expressed as SLP minus 1013.25 hPa. (a) Observations over the period 1980–2014 from the ERA-interim reanalysis (Dee et al., 2011). (b) SLP from SPEAR\_LO control simulation, calculated as the time mean over years 1–100 of a 2010 control simulation. (c) Same as (b) for SPEAR\_MED. White regions over continents denote areas where surface elevation is too high to compute sea level pressure from surface pressure. (d) Model bias, calculated as SPEAR\_LO minus observations. (e) Model bias calculated as SPEAR\_MED minus observations. Root mean square error (RMSE) and spatial correlations ( $R$ ) are shown in the text next to the figure.

The seasonal cycles of sea ice extent in observations and the model simulations are shown in Figure 7. Simulations in the Northern Hemisphere are fairly realistic. There is somewhat too much winter sea ice extent, especially in the Greenland Sea and Barents Sea regions (not shown). In the Southern Hemisphere the model simulates a larger than observed seasonal cycle of sea ice, with excessive extent in winter and summer values below observed. This bias may be related to a tendency for excessive absorbed shortwave radiation in summer, thereby reducing sea ice, or to a lack of dynamical thickening (ridging) processes in this model (Adcroft et al., 2019).

### 3.2. Time-Varying Climate From Internal Variability

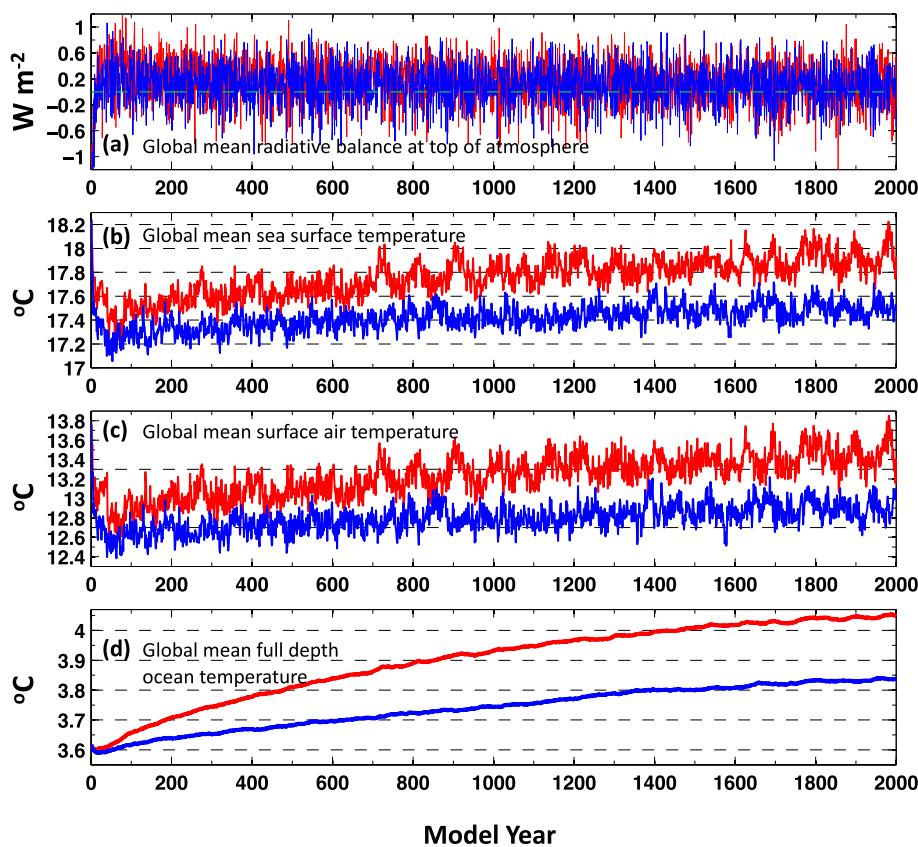
We also conduct 2000-year control simulations with atmospheric composition set at 1850 conditions. The lower concentrations of atmospheric greenhouse gases in 1850 relative to 2010 produce a simulated



**Figure 7.** Seasonal cycle of climatology of areal extent of sea ice in (a) the Northern Hemisphere and (b) the Southern Hemisphere. Model climatologies were calculated as time means over years 21–100 of a control simulation with 2010 radiative forcing. Black dashed line is for observations (Cavalieri et al., 1996, climatology for 1990–2017), red is SPEAR\_LO, and blue is SPEAR\_MED.

climate in 1850 that is both cooler, and closer to radiative balance, than the simulated 2010 climate. The long-term mean in global mean net radiation at the top of the atmosphere over years 501–1000 is  $0.18 \text{ W m}^{-2}$  in SPEAR\_LO and  $0.12 \text{ W m}^{-2}$  in SPEAR\_MED. The smaller imbalance in SPEAR\_MED appears to be related to differences in the cloud simulations between the two models and is a focus of current diagnostic analyses.

The models do undergo a long adjustment process, as the initial conditions from the ocean are representative of modern conditions, and the simulated climate system must move toward an equilibrium with atmospheric composition (and hence radiative forcing) consistent with the preindustrial period. We show in Figure 8a the time series of the net radiative balance at the top of the atmosphere for the 1850 control simulations of SPEAR\_LO and SPEAR\_MED. After an initial rapid adjustment, both models have a net positive bias that slowly declines over time, indicating that the climate system is gaining energy. This is a typical bias in many previous GFDL climate models. Even with 1850 radiative forcing (which should be less positive than today's climate), the ocean warms by moving heat from the near-surface layers into the ocean interior. Aspects of this process have been explored using an earlier coupled climate model with very high ocean resolution ( $0.1^\circ$ ) (Delworth et al., 2012) in which it was shown that an improved representation of the effects of ocean mesoscale eddies can ameliorate this problem. This persistent bias points to the importance of either explicitly resolving fine-scale processes in the ocean, including mesoscale eddies, or developing improved parameterizations of their impacts for use in coarser-resolution ocean models.

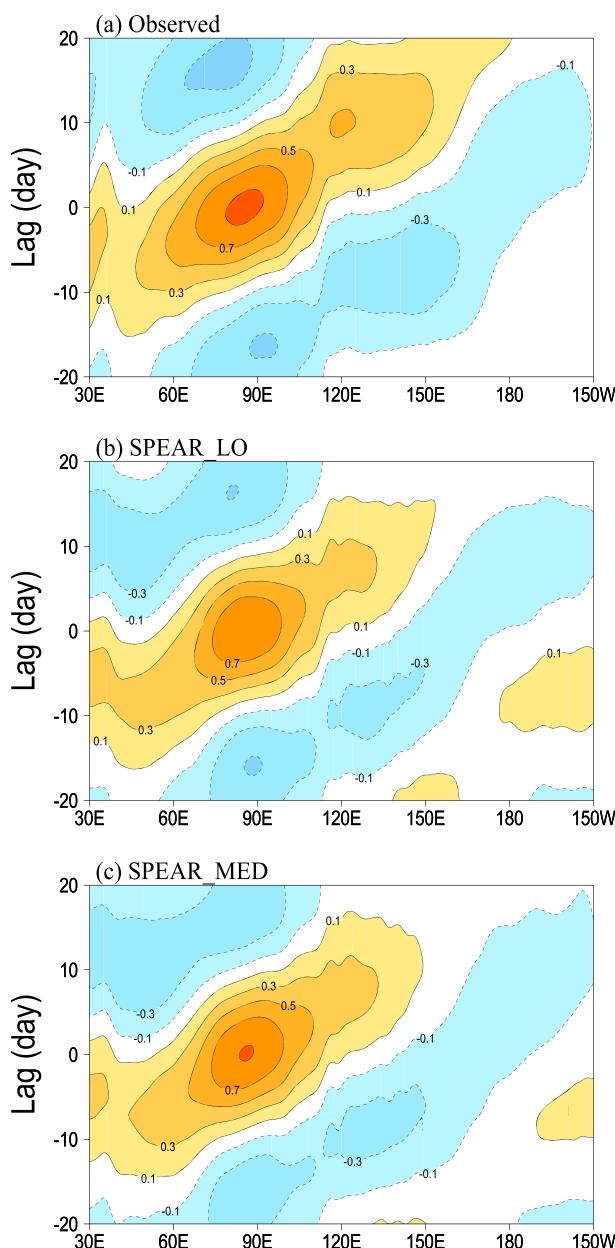


**Figure 8.** Time series of annual mean quantities from SPEAR\_LO (red lines) and SPEAR\_MED (blue lines). (a) Global mean net radiative balance at top of the atmosphere ( $\text{W m}^{-2}$ ), (b) global mean sea surface temperature ( $^{\circ}\text{C}$ ), (c) global mean surface air temperature ( $^{\circ}\text{C}$ ), (d) global mean full depth ocean temperature ( $^{\circ}\text{C}$ ).

We show in Figure 8b the time series of global mean SST, in Figure 8c the time series of global mean surface air temperature, and in Figure 8d the time series of global mean, depth mean ocean temperature. All time series show an initial cooling associated with the adjustment to 1850 radiative forcing. The time series have upward trends after the first 50 years, consistent with the small persistent net positive values of the top of the atmosphere net radiation. However, the smaller imbalance in SPEAR\_MED results in considerably less drift in all temperature measures. This smaller drift is clearest in the time series of global mean, volume mean ocean temperature (Figure 8d). While the drift continues even through year 2000, it is considerably smaller after the first 200 years or so, when the upper part of the ocean has equilibrated (not shown). The deeper layers take multiple millennia to come into full equilibration.

We next examine the model's ability to simulate a variety of phenomena relevant for seasonal to decadal predictions, starting with the MJO (Madden & Julian, 1971). As one measure of the ability of SPEAR to simulate the MJO, we show in Figure 9 the correlations of outgoing longwave radiation between values at each longitude, averaged over  $10^{\circ}\text{S}$  to  $10^{\circ}\text{N}$ , versus outgoing longwave radiation averaged over a reference box in the eastern Indian Ocean. The correlations are shown as a function of lag to illustrate the time behavior of the MJO. The tilt of the correlations upward and to the right denotes eastward propagation of the MJO signal. Both versions of SPEAR have a realistic MJO simulation, capturing the propagation features very well.

On seasonal time scales the ENSO is a dominant phenomenon. We show in Figure 10 the standard deviation of monthly SST anomalies in the tropical Pacific. The models broadly capture the spatial structure of SST variability in the tropical Pacific, although both models appear to have variability that is somewhat larger than observed and that extends further westward in the models than in observations. The



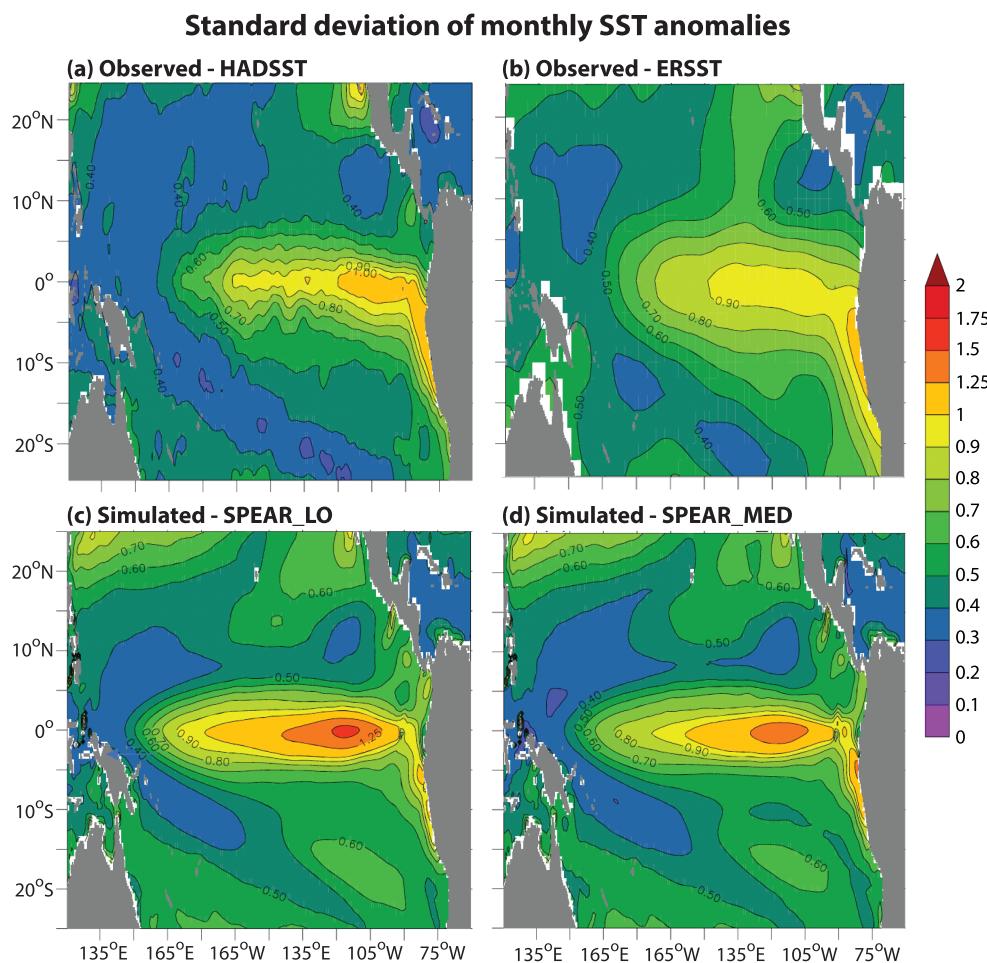
**Figure 9.** Characteristics of the MJO in observations (NOAA OLR, 1979–2018) and model simulations (years 1–100). OLR values are used from November–April and are first filtered to retain only time scales of 20–100 days. Each point shows the correlation for a given lag (y axis) and longitude (x axis) between OLR averaged over 10°S to 10°N versus OLR averaged over a reference box in the eastern Indian Ocean (10°S to 10°N, 70–100°E).

North Pacific. This issue is very common in fully coupled climate models, which is also seen in our previous FLOR model (Figure 14g). The North Pacific horseshoe-like pattern associated with the PDO also shows a connection with the tropical Pacific Ocean (Figures 14a, 14c, 14e, and 14g). The SST anomalies extend from the eastern North Pacific to the central tropical Pacific, suggesting strong teleconnections between the extratropics and the tropical ocean. Compared to observations, this connection is overall weaker in models. However, the SPEAR models have a better performance compared to the FLOR model and are closer to observations. This improvement is related to better simulations of wind, SST, and low cloud mean states over the eastern North Pacific in the SPEAR models.

higher-resolution SPEAR\_MED model appears to have a slightly better simulation than SPEAR\_LO by this metric. The time scale of ENSO variability is shown in Figure 11 by the spectra of SST averaged over the Niño 3 region the Pacific (150–90°W, 5°S to 5°N). Both models have spectral characteristics in relatively good agreement with observations, although the total variance in the models is larger than in observations. We show in Figure 12 the teleconnections of 200-hPa geopotential height averaged over December–February with the NINO3 index calculated both from a reanalysis (Figure 12a) and model simulations (Figures 12b–12d). Both SPEAR versions (Figures 12c and 12d) provide realistic simulations of the teleconnections, although there is a tendency for the Pacific and North American maximum teleconnections to be shifted westward relative to observations. The same is true for the previous generation GFDL prediction model (FLOR; see Table 1) as shown in Figure 12b.

On seasonal time scales one of the phenomena to predict is tropical cyclones and hurricanes. The existing GFDL prediction system has demonstrated skill (Murakami et al., 2016) in the prediction of seasonal tropical storm and hurricane activity. We show in Figure 13 the ability of SPEAR\_LO and SPEAR\_MED to simulate the frequency of tropical storms and hurricanes in various basins around the globe compared to observations (Knapp et al., 2010). While the spatial resolution of both models is insufficient to simulate major hurricanes, both models are able to simulate the broad-scale distributions of tropical storms in most ocean basins, with SPEAR\_MED able to simulate more intense storms than SPEAR\_LO. The spatial distribution of simulated storms is also more realistic in SPEAR\_MED than SPEAR\_LO.

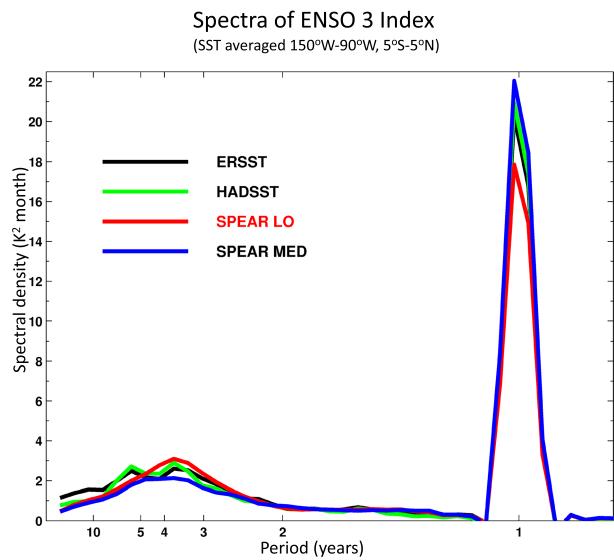
On decadal time scales the Pacific Decadal Oscillation (PDO, also referred to as Pacific Decadal Variability) is the dominant pattern of variability in the Pacific (Mantua et al., 1997). This low-frequency variability is associated with a large-scale change in the climate system and marine ecosystems (Cayan et al., 2001; Deser et al., 2004). We show in Figure 14 the spatial pattern and power spectrum of the PDO in observations and SPEAR, along with the older generation FLOR model (see Table 1). The SST anomaly associated with the PDO in observations is characterized by a horseshoe-like pattern over the North Pacific, with SST anomalies of one sign in the western and central North Pacific and SST anomalies of the opposite sign to the north, south, and east (Figure 14a). The observed pattern shows its maximum over the central North Pacific along with a secondary maximum in the Kuroshio-Oyashio-Extension region. The SPEAR models generally capture the observed horseshoe-like shape of the PDO over the North Pacific (Figures 14a, 14c, and 14e). However, the SST anomalies in the Kuroshio-Oyashio-Extension region are higher than those in the central



**Figure 10.** Standard deviation of monthly SST anomalies. (a) Observed, HADSST, years 1917–2016; (b) observed, ERSST, years 1915–2014; (c) simulated, SPEAR\_LO, calculated from 500-year segment of 1850 control simulation; and (d) simulated, SPEAR\_MED, calculated from 500-year segment of 1850 control simulation. Units are in °C. The model variability is larger than observed and extends further westward than observed. Both biases appear to be modestly reduced using the higher-resolution SPEAR\_MED model.

The spectrum of the observed has peaks around 15–25 years and 50–70 years (Figure 14b). Note that the spectral peaks do not exceed the 95% significance level because of the short time coverage of the observations. The models generally capture the observed 15- to 25-year peak, with peaks at 17, 25, and 20 years in SPEAR\_LO, SPEAR\_MED, and FLOR, respectively (Figures 14d, 14f, and 14h). The second time scale for the observed PDO, at around 50–70 years, is not captured by the GFDL models. We also examined the impact of the PDO on North American precipitation. Consistent with observations and the FLOR model (Zhang & Delworth, 2016a, 2016b), the PDO warm phase corresponds to wet conditions in the southwestern United States in both warm and cold seasons and vice versa. This is because the anomalous southerly winds associated with the Pacific-North-American teleconnection during the PDO warm phase enhance moisture transport from south to north and therefore leads to an excess of precipitation and vice versa.

One of the prime processes leading to decadal predictability is the Atlantic Meridional Overturning Circulation (AMOC) and its slow decadal variations. We show in Figure 15 the AMOC for both SPEAR\_LO and SPEAR\_MED, calculated in two ways. In the first method the volume flow on height surfaces in the ocean (depth space) is used for computing the overturning stream function. In the second method the volume flow in density layers is used to compute the stream function. The AMOC calculated in depth space can be problematic, as northward and southward flowing water parcels at the same depth would cancel in the calculation of a net meridional flow. In reality those water parcels could have very different temperature and density characteristics so that there is a net poleward heat transport. The differences



**Figure 11.** Spectra of NINO3 indices from observations and models. Model spectra are computed from 1850 control simulations (the first 200 years are discarded, and the spectra were computed over years 201–2000). Observational spectra are computed over the length of the available records (for ERSST used data from 1854–2017; for HADSST used data from 1870 to 2015).

dal variability (visually apparent from the time series in Figures 16a and 16b). The spectrum for SPEAR\_LO also has a peak around 100 years; this is very clear at 26°N but somewhat more muted at 45°N. Separate analyses (not shown) suggest that this lower-frequency peak is associated with centennial-scale variations in the formation of Antarctic Bottom Water (AABW) in the SO (Zhang et al., 2019) that are more pronounced in SPEAR\_LO than SPEAR\_MED. These AABW variations appear to influence the AMOC through remote teleconnections (Zhang & Delworth, 2016a).

For additional diagnostic analyses of the overall model simulation, please see Figures S1–S9 in the supporting information.

### 3.3. Idealized Climate Change

To characterize the climate sensitivity of the models, we have conducted idealized climate change experiments. Starting from year 201 of the 1850 control simulation for SPEAR\_LO and from year 101 of the 1850 control simulation for SPEAR\_MED, we conducted simulations in which atmospheric CO<sub>2</sub> increases at 1% per year until doubling and is held constant thereafter. The transient climate response, defined as the global mean temperature in a 20-year window centered at the time of doubling CO<sub>2</sub> minus the global mean temperature in the corresponding 20-year segment of the control simulation (see Box 12.2, (Collins et al., 2013)), is 1.65 K for SPEAR\_LO and 1.78 K for SPEAR\_MED. These values are in the middle of the likely range of distribution of transient climate response values (1–2.5 °C) as reported in IPCC AR5 (Collins et al., 2013).

### 3.4. Ensembles of Climate Change Simulations Over the Period 1851 to 2100

Although not the primary focus of this paper, we briefly document the simulation characteristics of the 30-member ensembles of the SPEAR\_LO and SPEAR\_MED models when driven by historical and projected changes in radiative forcing over the period 1851–2100 (for SPEAR\_LO) and 1921–2100 (for SPEAR\_MED). The design and radiative forcing for these simulations were discussed in section 2.4.3. These ensembles will be analyzed in greater detail in forthcoming papers, but we document here some basic aspects of their response to changing radiative forcing.

Shown in Figure 17 are the time series of global mean surface air temperature from SPEAR\_LO and SPEAR\_MED, along with two observational estimates (CRUTEM4, 1850–2018 and GISTEMP v4, 1880–2018). We show the 30-member ensemble mean from SPEAR\_LO and SPEAR\_MED as the thick

between the AMOC calculated in depth space and density space are most apparent at higher latitudes of the North Atlantic, where southward flowing cold denser water in the western Atlantic occurs at the same depth as northward flowing warm (light) waters in the eastern Atlantic, leading to a significant transport in density space but minimal transport calculated in depth space.

The primary observations to assess the realism of the simulated AMOC come from the RAPID array at 26.5°N (<http://www.rapid.ac.uk/rapid-moc/overview.php>) that has been in place since 2004. The long-term mean of the observed AMOC is  $17.0 \pm 4.4$  Sverdrups (1 Sverdrup = 1 Sv =  $10^6 \text{ m}^3 \text{ s}^{-1}$ ); the simulated values in z-space at 26.5°N are 15.9 Sv for both SPEAR\_LO and for SPEAR\_MED, reasonably close to the observations. The simulated values in density space (17.4 for SPEAR\_LO and 17.3 for SPEAR\_MED) are also consistent with the observations.

Time series for the AMOC can be calculated for each latitude as the maximum stream function value over the water column. We show in Figures 16a and 16b the time series of the AMOC at 26°N in both z-space and isopycnal coordinates. The AMOC remains stable over the 2,000 years of the simulations, with no long-term drifts. The AMOC does have substantial decadal and multidecadal variability, as shown by the spectra in Figures 16c and 16d at 26°N and 45°N, respectively. The spectra from both latitudes and both models have broad, statistically significant peaks on the 25- to 40-year time scale, showing the presence of substantial multidecadal variability (visually apparent from the time series in Figures 16a and 16b). The spectrum for SPEAR\_LO also has a peak around 100 years; this is very clear at 26°N but somewhat more muted at 45°N. Separate analyses (not shown) suggest that this lower-frequency peak is associated with centennial-scale variations in the formation of Antarctic Bottom Water (AABW) in the SO (Zhang et al., 2019) that are more pronounced in SPEAR\_LO than SPEAR\_MED. These AABW variations appear to influence the AMOC through remote teleconnections (Zhang & Delworth, 2016a).

For additional diagnostic analyses of the overall model simulation, please see Figures S1–S9 in the supporting information.

### 3.3. Idealized Climate Change

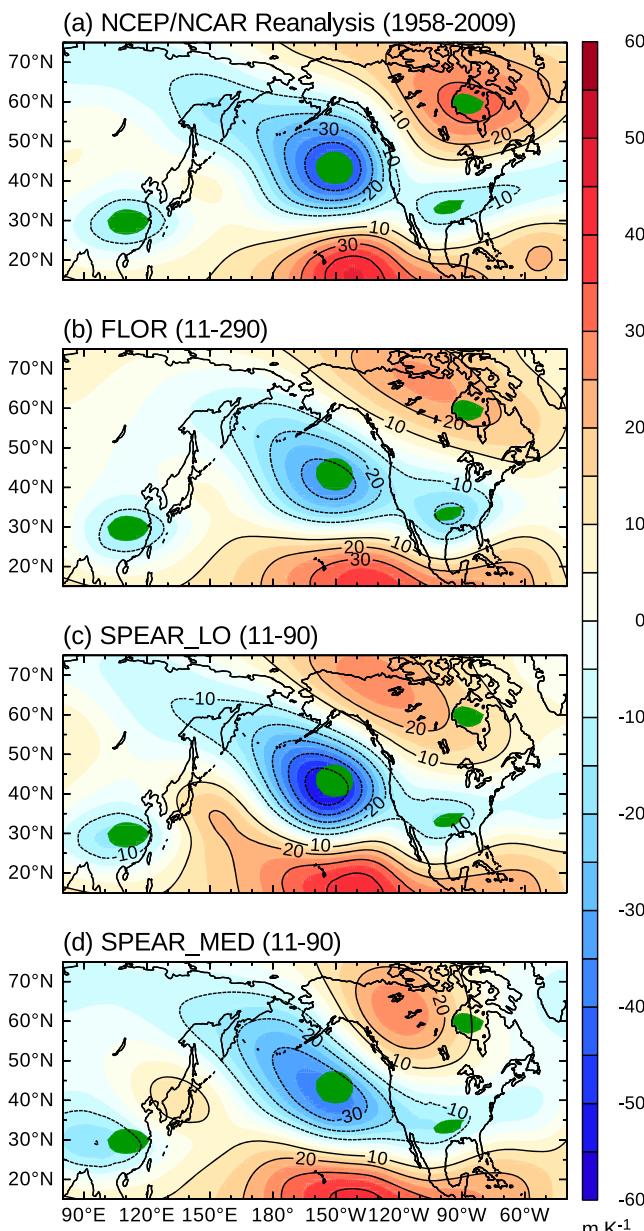
To characterize the climate sensitivity of the models, we have conducted idealized climate change experiments. Starting from year 201 of the 1850 control simulation for SPEAR\_LO and from year 101 of the 1850 control simulation for SPEAR\_MED, we conducted simulations in which atmospheric CO<sub>2</sub> increases at 1% per year until doubling and is held constant thereafter. The transient climate response, defined as the global mean temperature in a 20-year window centered at the time of doubling CO<sub>2</sub> minus the global mean temperature in the corresponding 20-year segment of the control simulation (see Box 12.2, (Collins et al., 2013)), is 1.65 K for SPEAR\_LO and 1.78 K for SPEAR\_MED. These values are in the middle of the likely range of distribution of transient climate response values (1–2.5 °C) as reported in IPCC AR5 (Collins et al., 2013).

### 3.4. Ensembles of Climate Change Simulations Over the Period 1851 to 2100

Although not the primary focus of this paper, we briefly document the simulation characteristics of the 30-member ensembles of the SPEAR\_LO and SPEAR\_MED models when driven by historical and projected changes in radiative forcing over the period 1851–2100 (for SPEAR\_LO) and 1921–2100 (for SPEAR\_MED). The design and radiative forcing for these simulations were discussed in section 2.4.3. These ensembles will be analyzed in greater detail in forthcoming papers, but we document here some basic aspects of their response to changing radiative forcing.

Shown in Figure 17 are the time series of global mean surface air temperature from SPEAR\_LO and SPEAR\_MED, along with two observational estimates (CRUTEM4, 1850–2018 and GISTEMP v4, 1880–2018). We show the 30-member ensemble mean from SPEAR\_LO and SPEAR\_MED as the thick

20yr-detrended DJF 200hPa height anomaly (m) regressed on 20yr-detrended DJF NINO3 SSTA (K)



**Figure 12.** Regression coefficients of 200 hPa geopotential height on NINO3 index for December–February. (a) Calculated from NCEP/NCAR reanalysis, (b) calculated from years 11–290 of a FLOR control simulation (see Table 1 for FLOR reference), (c) calculated from years 11–90 of SPEAR\_LO, and (d) calculated from years 11–90 of SPEAR\_MED. The green ellipses denote the centers of the teleconnection patterns from the reanalysis in (a).

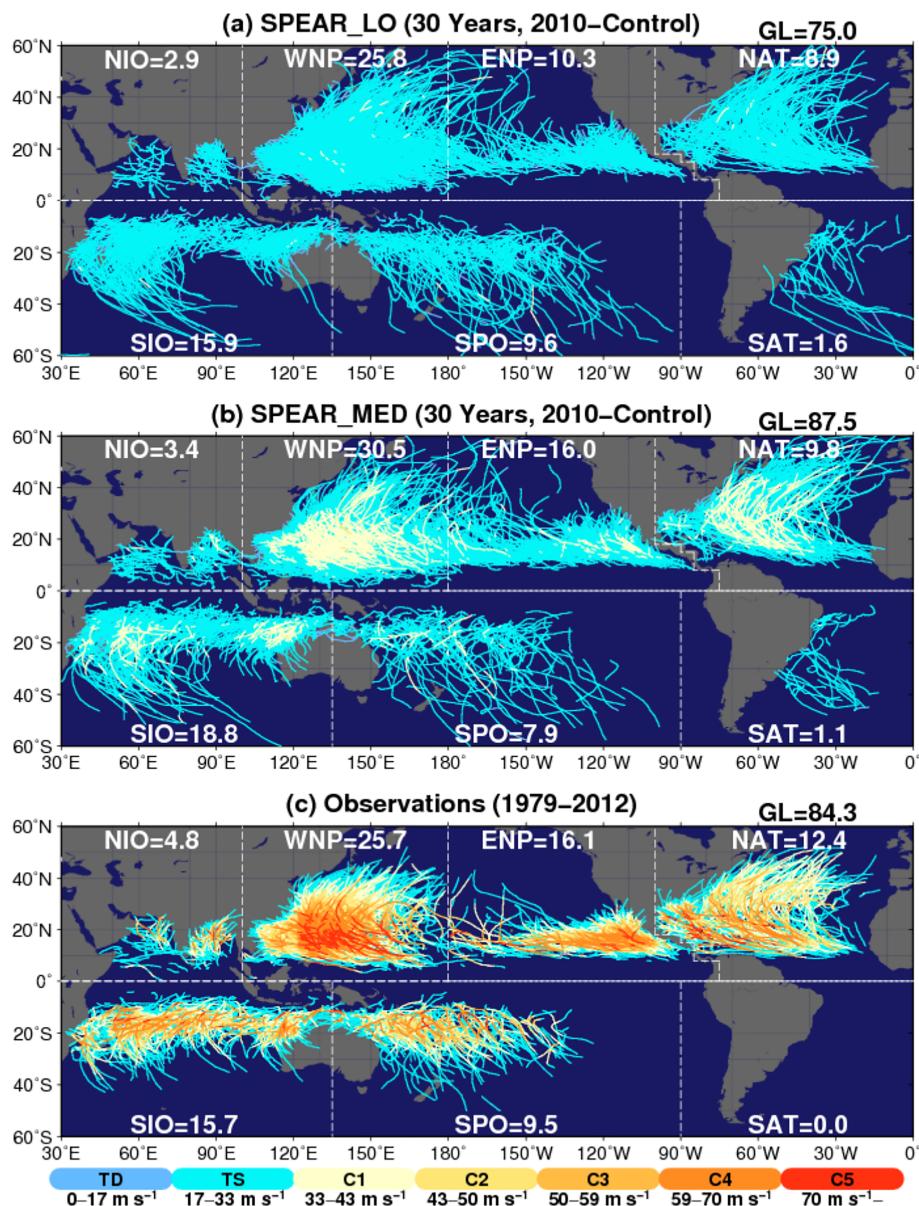
solid red and blue lines, respectively. The range of temperatures simulated across the ensemble members is denoted by the tan shading for SPEAR\_LO and the light blue shading for SPEAR\_MED. The models reproduce many of the features of the observed record, including trends over the last 80 years. The models do not reproduce the early twentieth century warming seen in the observations. The global mean surface air temperature increase by the year 2100, relative to 1961–1990, is 4.54 K for SPEAR\_LO and 4.78 K for SPEAR\_MED when forced with the SSP5-85 scenario (Kriegler et al., 2017) of future radiative forcing.

The patterns of changes in surface air temperature and annual mean precipitation are shown in Figure 18, calculated as the mean over the period 2041–2060 minus the period 1961–1990. Consistent with many previous model projections (Collins et al., 2013), warming is larger in the Northern Hemisphere than in the Southern Hemisphere (Figures 18a and 18b), with generally more warming over continental regions than oceanic regions. The muted warming over India and sub-Saharan Africa is related to increased rainfall (Figures 18c and 18d) and moister land surfaces, allowing any increase in the surface energy balance to be preferentially balanced by latent heat fluxes. The cooling south of Greenland is associated with a weakening of the AMOC (shown below in Figure 19) and associated reduction in oceanic meridional heat transport. The relative lack of warming in the Ross Sea and SO is due to relatively strong vertical mixing in the ocean, leading to an effectively larger heat capacity and slower rate of temperature increase.

The precipitation changes (Figures 18c and 18d) have many aspects that have been seen in many previous models using earlier versions of future radiative forcing scenarios. These changes include increased precipitation in the high latitudes of each hemisphere and in the deep tropics, with drying in many areas of the subtropics. The patterns of change in SPEAR\_LO and SPEAR\_MED are extremely similar, pointing both to the statistical robustness of the results and the lack of dependence of large-scale features on horizontal spatial resolution. There are increases in precipitation in the high latitudes of each hemisphere, with the exception of reductions in the subpolar gyre of the North Atlantic where AMOC weakening (shown below in Figure 19) leads to cooling and a reduction in precipitation. There is increased precipitation in the tropical Pacific, much of tropical Africa, and the Indian subcontinent.

The time series of the AMOC at 26°N are shown in Figure 19. The bold solid lines (red for SPEAR\_LO, blue for SPEAR\_MED) indicate the 30-member ensemble mean, with the model spread across the 30 members for SPEAR\_LO and SPEAR\_MED shown by the shading (tan for SPEAR\_LO, light blue for SPEAR\_MED). For reference we also show the annual mean AMOC (green lines and symbols) as observed over the period 2005–2017 from the RAPID array at 26°N in the Atlantic (data downloaded from [https://www.rapid.ac.uk/rapidmoc/rapid\\_data/datadl.php](https://www.rapid.ac.uk/rapidmoc/rapid_data/datadl.php) on 1 August 2019). During the period of the observations, the AMOC time series in both SPEAR\_LO and SPEAR\_MED are quite consistent with the observations, both in terms of their means and amplitude of variability. To the extent that the time series of the model ensemble means indicate the response of the AMOC to changing radiative forcing, the large variations in the observational record relative to the model ensemble mean suggest a strong role for internal variability. The increase in simulated AMOC prior to the 1980s is likely due to increasing anthropogenic aerosols (Delworth & Dixon, 2006). The increasing aerosols preferentially cool the Northern

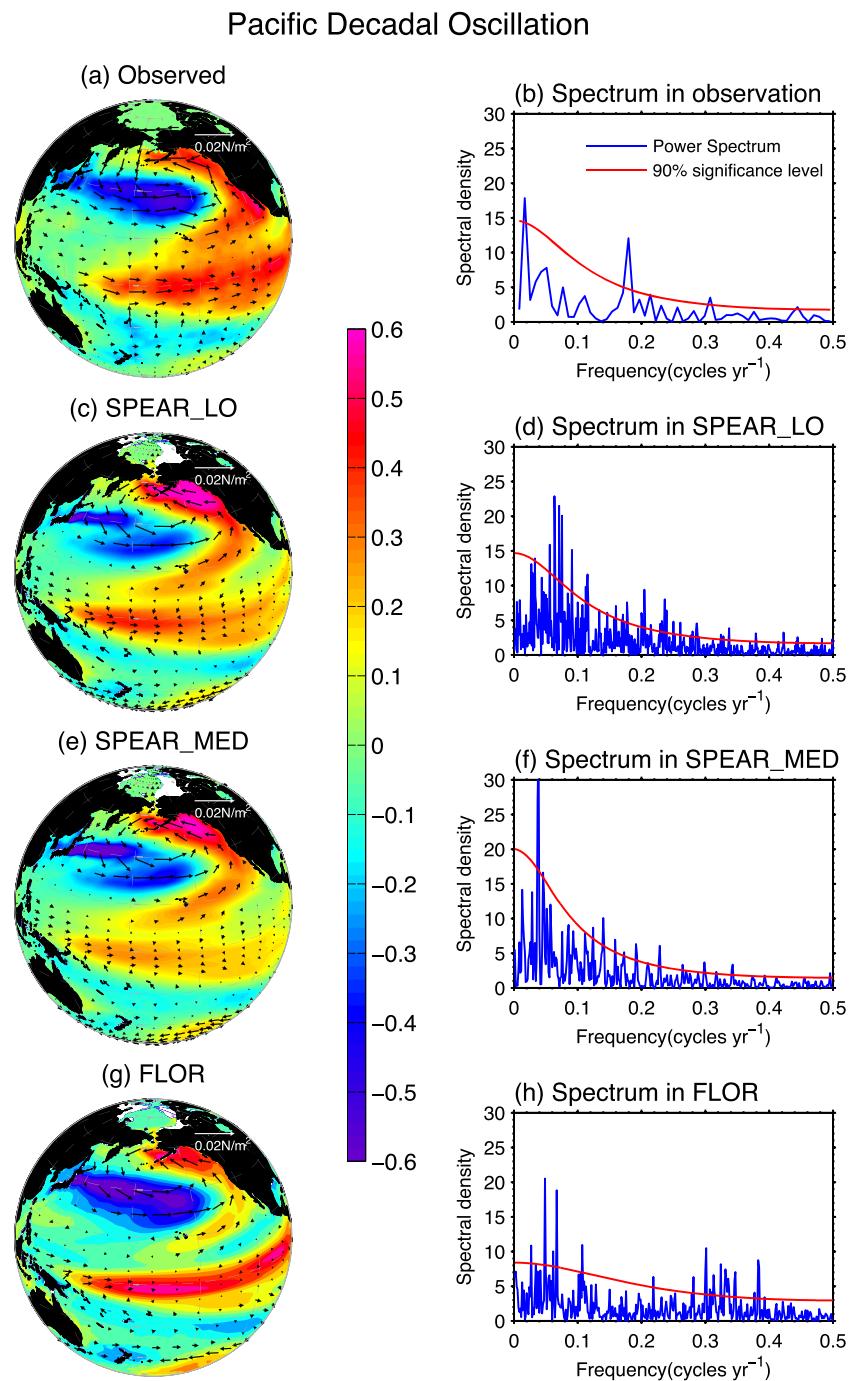
tent with the observations, both in terms of their means and amplitude of variability. To the extent that the time series of the model ensemble means indicate the response of the AMOC to changing radiative forcing, the large variations in the observational record relative to the model ensemble mean suggest a strong role for internal variability. The increase in simulated AMOC prior to the 1980s is likely due to increasing anthropogenic aerosols (Delworth & Dixon, 2006). The increasing aerosols preferentially cool the Northern



**Figure 13.** Global distribution of TC tracks during all seasons for 2010 control simulation by (a) SPEAR\_LO, (b) SPEAR\_MED, and (c) observations from 1979 to 2012. The numbers for each basin show the annual mean number of TCs. TC tracks are colored according to the intensities of the TCs as categorized by the Saffir-Simpson hurricane wind scale [e.g., tropical depression (TD), tropical storm (TS), and hurricane Categories 1 to 5 (C1–C5)].

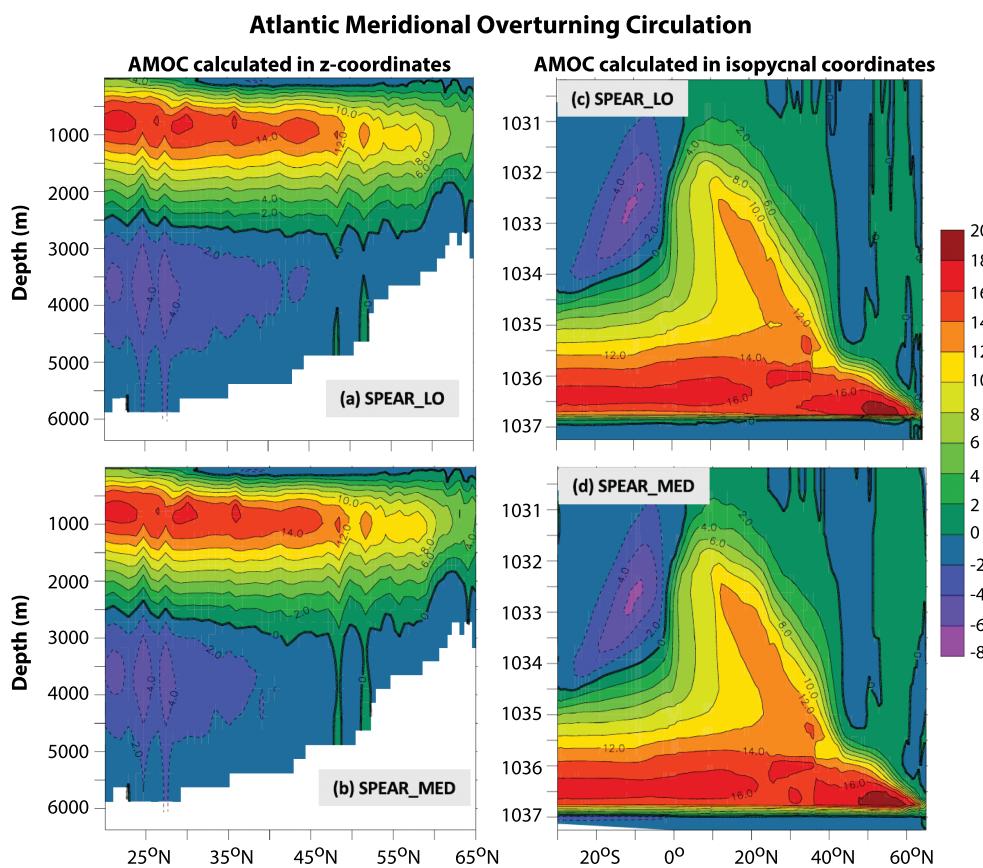
Hemisphere and reduce poleward water vapor transport, thereby cooling the near-surface waters in the higher latitudes of the North Atlantic and increasing salinity due to reduced poleward water vapor transport and runoff into the ocean. The increased upper ocean density leads to enhanced deepwater formation and an increase in the AMOC. The long-term decline after the year 2000 is due to both declining atmospheric aerosols and increasing greenhouse gases; together these two factors warm and freshen the near-surface waters, thus increasing vertical stability and reducing deepwater formation and the AMOC.

The time series of the areal extent of Arctic sea ice in September is shown in Figure 20 for SPEAR\_LO and SPEAR\_MED, as well as the observations. The observed values lie within the model spread for SPEAR\_LO and partially within the spread for SPEAR\_MED (the mean extent for SPEAR\_MED is slightly larger than SPEAR\_LO and the observations). Trends of sea ice extent in recent decades are well simulated by both



**Figure 14.** Shown are the Pacific Decadal Oscillation (PDO) spatial pattern (left panels) and power spectra of the PDO index (right panels) in observations (HadISST) (a, b), SPEAR\_LO (c, d), SPEAR\_MED (e, f), and FLOR (g, h) models. The PDO index is defined as the principal component time series of leading EOF over the North Pacific Ocean (20–60°N). The PDO spatial pattern (unit is in °C) is obtained by regressing the global SST onto the normalized PDO index. Note that the spectrum is presented in a variance-preserving form; therefore, the integral of the area under the spectrum curve equals to the variance of the PDO index (here is normalized and the unit is 1).

SPEAR\_LO and SPEAR\_MED. During the period 2007–2012 there appeared to be a more rapid decline of Arctic sea ice extent, but the observed trend over the last 5 years has moderated, so that the overall trend in both models agrees well with the observations. The first year in which the simulated Arctic in SPEAR\_LO has ice-free conditions in September (defined as areal extent less than  $1 \times 10^6 \text{ km}^2$ ) ranges



**Figure 15.** Time mean, annual mean of the Atlantic Meridional Circulation (AMOC) calculated over years 501–2000 of the 1850 control simulations. Units are in Sverdrups ( $10^6 \text{ m}^3 \text{ s}^{-1}$ ). Flow is clockwise around a maximum stream function value, with the strength of the flow proportional to the gradient of the stream function contours. Calculations in the left column are done in  $z$  coordinates, while those in the right column are done in isopycnal coordinates. For the left column the  $y$  axis is density with units of  $\text{kg m}^{-3}$ . (a) SPEAR\_LO in  $z$  coordinates, (b) SPEAR\_MED in  $z$  coordinate, (c) SPEAR\_LO in isopycnal coordinates, and (d) SPEAR\_MED in isopycnal coordinates.

from 2039 to 2060, with slightly later dates in SPEAR\_MED due to its small positive bias in extent over the instrumental period.

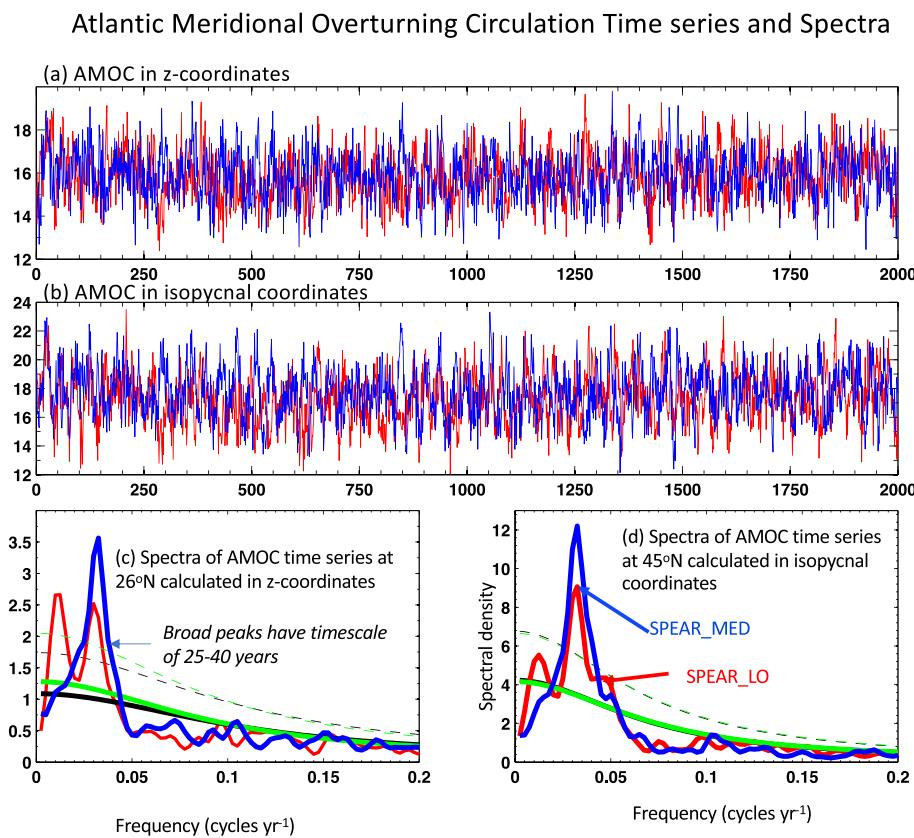
The above analysis represents a brief assessment of these ensembles and the performance of the SPEAR models in simulating climate change from the 19th through the 21st centuries. The output from these large ensembles will be examined in much greater detail in future studies, and output from these simulations will be made publicly available. In particular, the 50-km resolution of SPEAR\_MED makes this ensemble particularly useful for studies of regional hydroclimate and extremes.

#### 4. SO Simulation Characteristics

An important component of the model development process focused on the SO and its important role in the global climate system. Early versions of the model had little or no deepwater formation in the SO, creating substantial model drift along with episodic large events in which the SO was destabilized and vented large amounts of heat. These issues led to a focus during the model development process on factors impacting SO deepwater formation and climate drift. We illustrate this issue in this section through the use of sensitivity simulations involving albedo formulations over Antarctica.

##### 4.1. Issues in Simulating the SO

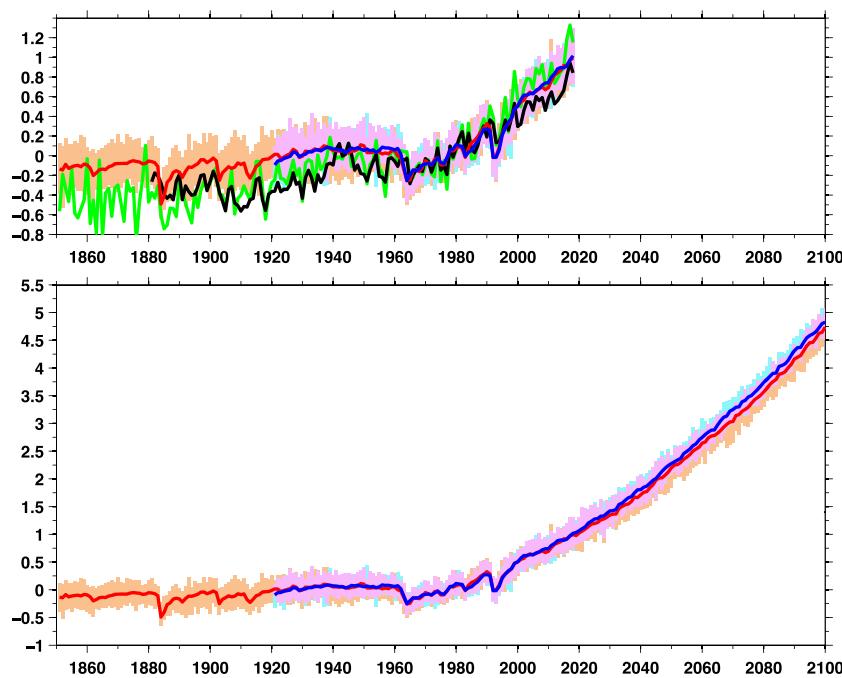
As the emission of CO<sub>2</sub> from industrial activity into the atmosphere continues, the amount of atmospheric CO<sub>2</sub> that enters the SO and subsequently moves to depth is an important factor determining how much CO<sub>2</sub>



**Figure 16.** (a) Time series of the AMOC index in the 1850 control simulations, calculated in  $z$  coordinates. Red is from SPEAR\_LO; blue is from SPEAR\_MED. The AMOC index for  $z$  coordinates is defined as the maximum stream function value at  $26^{\circ}\text{N}$  in the Atlantic. (b) Same as (a) but using isopycnal coordinates and defining the AMOC index as the maximum stream function value at  $45^{\circ}\text{N}$  in the Atlantic. (c) Spectra of the AMOC time series in (a). Thick red (blue) line is for SPEAR\_LO (SPEAR\_MED). The smooth thick black (green) line is the spectrum from a reference red noise process calculated from the lag-one autocorrelation in SPEAR\_LO (SPEAR\_MED). Dashed lines denote 95% confidence intervals. (d) Same as (c) but for AMOC indices calculated in isopycnal coordinates at  $45^{\circ}\text{N}$  as shown in (b).

remains in the atmosphere, with important implications for the characteristics of future climate change (Frölicher et al., 2015; Landschützer et al., 2015; Shi et al., 2018). Due to relatively low vertical stratification, near-surface water in the SO can move relatively easily to deeper layers of the ocean, thus facilitating the exchange of heat, fresh water, and  $\text{CO}_2$  between near-surface waters and the deep ocean. The characteristics of the upper ocean, including its variability, are tightly linked to the atmosphere via strong air-sea fluxes of heat, water, and momentum. In addition, model simulations suggest that the SO is characterized by long time scales of variability that may be predictable (Latif et al., 2017; Zhang et al., 2019) and that the characteristics of the variability depend on the model base state. Further, characteristics of the SO are a crucial factor influencing the melting rate of the seaward edge of land-based ice sheets, with substantial implications for sea level change. Simulating a realistic SO is thus an important metric for climate model development.

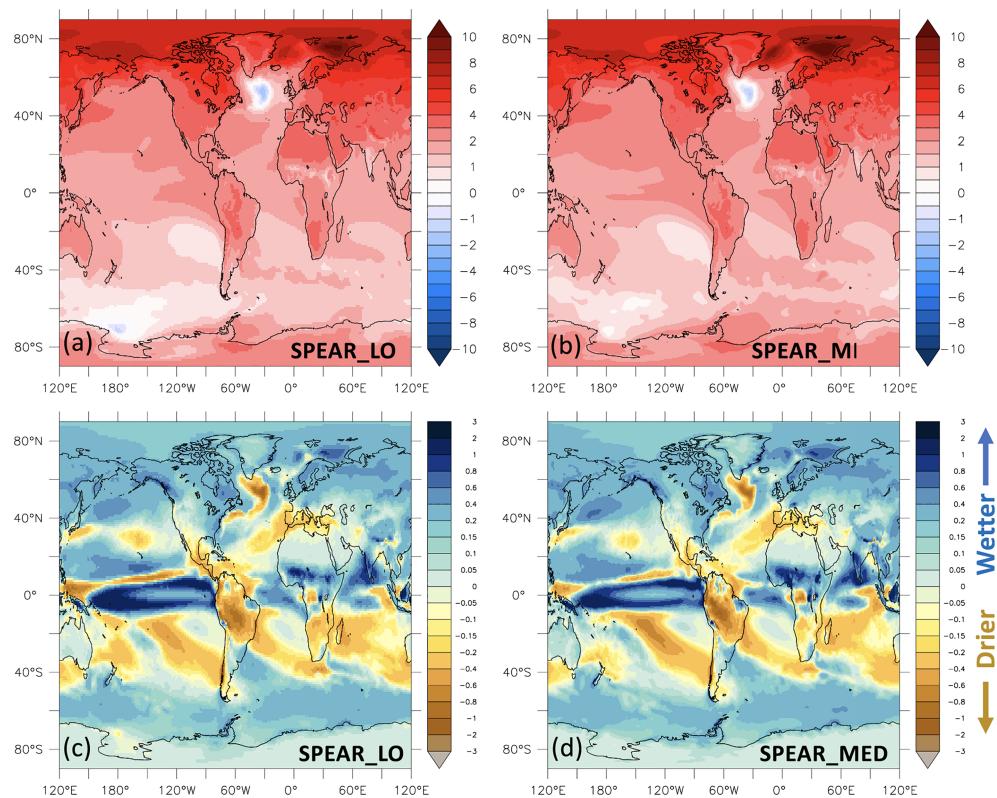
We characterize some aspects of circulation in the SO, including deepwater formation and ventilation, through the use of an ideal age tracer in the ocean. At the start of a simulation this tracer is set to zero everywhere in the world ocean. When a parcel of water is not in contact with the atmosphere, the age tracer grows linearly with time (over 1 year the age tracer increases by 1 year); however, whenever a water parcel moves back to the surface and is in contact with the atmosphere, the age tracer is once again set to zero. The age tracer is advected and mixed in the interior ocean similarly to a tracer such as salinity but has no influence on density or any other field and thus has no feedback on the ocean circulation. Relatively low values of age indicate regions where water parcels have been in recent contact with the atmosphere. Vigorous AABW



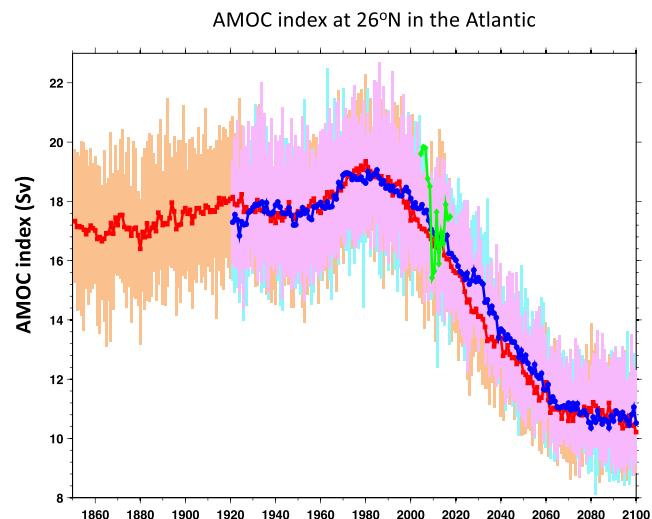
**Figure 17.** Time series of annual mean, global mean temperature computed using near-surface air temperature over land and sea surface temperature over ocean. All values expressed as departures from their respective time means over 1961–1990. (a) Observations in green (CRUTEM4) and black (GISTEM v4). Ensemble mean from SPEAR\_LO in red and from SPEAR\_MED in blue. The range of the 30 ensemble members for SPEAR\_LO is shown by the tan shading and for SPEAR\_MED by the light blue shading. The magenta shading indicates where the distributions for SPEAR\_LO and SPEAR\_MED overlap. (b) Model simulated values for SPEAR\_LO and SPEAR\_MED, shown from 1851 to 2100; note the different vertical axis in (b) compared to (a). Ensemble mean from SPEAR\_LO in red and for SPEAR\_MED in blue. Shading as in (a).

formation and ventilation should be characterized by low ages in the deeper layers of the SO, as water that has been recently at the surface moves to great depth. We show in Figure 21 the age tracer averaged over years 901–1000 in SPEAR\_LO (left column) and SPEAR\_MED (right column). Values at 2,000-m depth are shown in the top row, and values averaged from 4,000 m to the bottom of the ocean are shown in the bottom row. The maximum age values occur in the North Pacific, indicating that water at middle and lower depths of the North Pacific has not been in contact with the atmosphere for many centuries. Low age values at 2,000 m in the North Atlantic indicate the presence of recently ventilated North Atlantic Deep Water, moving southward in the western Atlantic from the subpolar North Atlantic to the South Atlantic. In the SO we see extensive areas with age values less than 200 years, indicating substantial regions with recently ventilated water. As we move to the deeper ocean below 4,000 m (bottom row), we see the continued presence of low age values in the SO, indicative that recently ventilated water is sinking into the deepest layers of the SO. In contrast, in the North Atlantic the age values below 4,000 m are considerably larger/older than they were at 2,000-m depth, indicating that the North Atlantic Deep Water formed in the subpolar gyre does not sink to the deepest layers of the North Atlantic. We also note that age in the deep North Pacific is smaller/younger than at 2,000 m, indicating the propagation of younger water masses resulting from AABW formation into the deepest layers of the North Pacific.

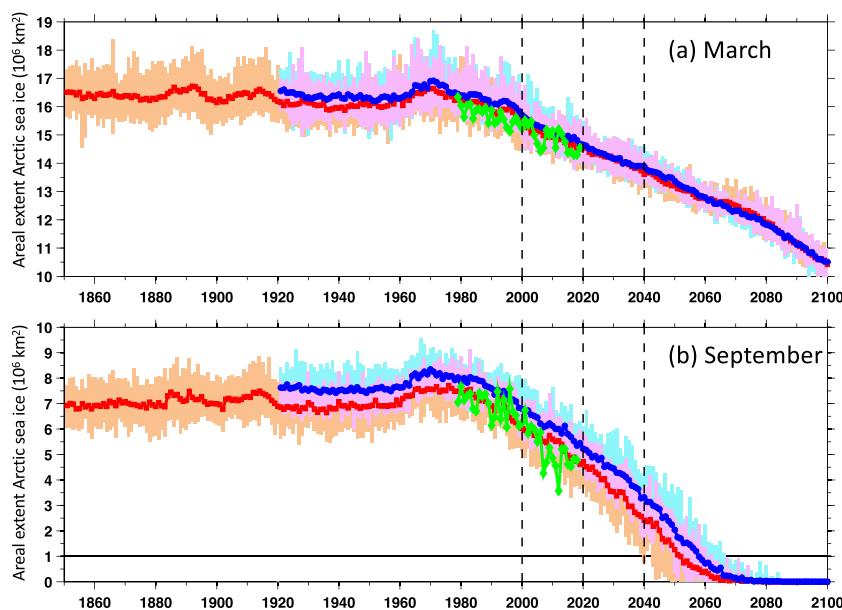
The age values shown above were for the final version of SPEAR after a number of sensitivity tests related to SO ventilation. In early stages of the SPEAR model development process, there was little deepwater formation and ventilation in the SO, leading to very low rates of AABW production, likely inconsistent with observations as inferred from oceanic tracers. The model has a warm fresh bias in the surface layers of the SO (associated with excess net shortwave radiation at the surface), thereby leading to vertical stratification that was larger than observed and inhibiting vertical exchange. Heat is transported in the subsurface ocean from the Atlantic Ocean into the SO; without sufficient vertical exchange, this heat accumulates in the subsurface ocean. After sufficient time, this subsurface accumulation of heat in the model tended to destabilize the



**Figure 18.** Differences in near-surface air temperature (top row) and precipitation (bottom row) calculated as 30-member ensemble mean, annual means for years 2041–2060 minus 1961–1990. Units are in °C for temperature and mm day<sup>-1</sup> for precipitation. The top row shows temperature differences for (a) SPEAR\_LO and (b) SPEAR\_MED. The bottom row shows precipitation differences for (c) SPEAR\_LO and (d) SPEAR\_MED.



**Figure 19.** Time series of an index of the Atlantic Meridional Overturning Circulation (AMOC) in the North Atlantic, computed using isopycnal coordinates (see Figures 12c and 12d). The index is computed each year as the maximum value of the overturning stream function at 26°N. Units are in Sverdrups (Sv), where 1 Sv =  $10^6 \text{ m}^3 \text{ s}^{-1}$ . The red line and symbols are the 30-member ensemble mean from SPEAR\_LO, while the blue line and symbols are the 30-member ensemble mean from SPEAR\_MED. The light tan (blue) shading shows the range of values each year across the ensemble members from SPEAR\_LO (SPEAR\_MED). The magenta shading indicates where the ranges of SPEAR\_LO and SPEAR\_MED overlap. The thick green line and symbols show the annual mean values from observations using the RAPID array at 26°N in the North Atlantic for 2005–2017.



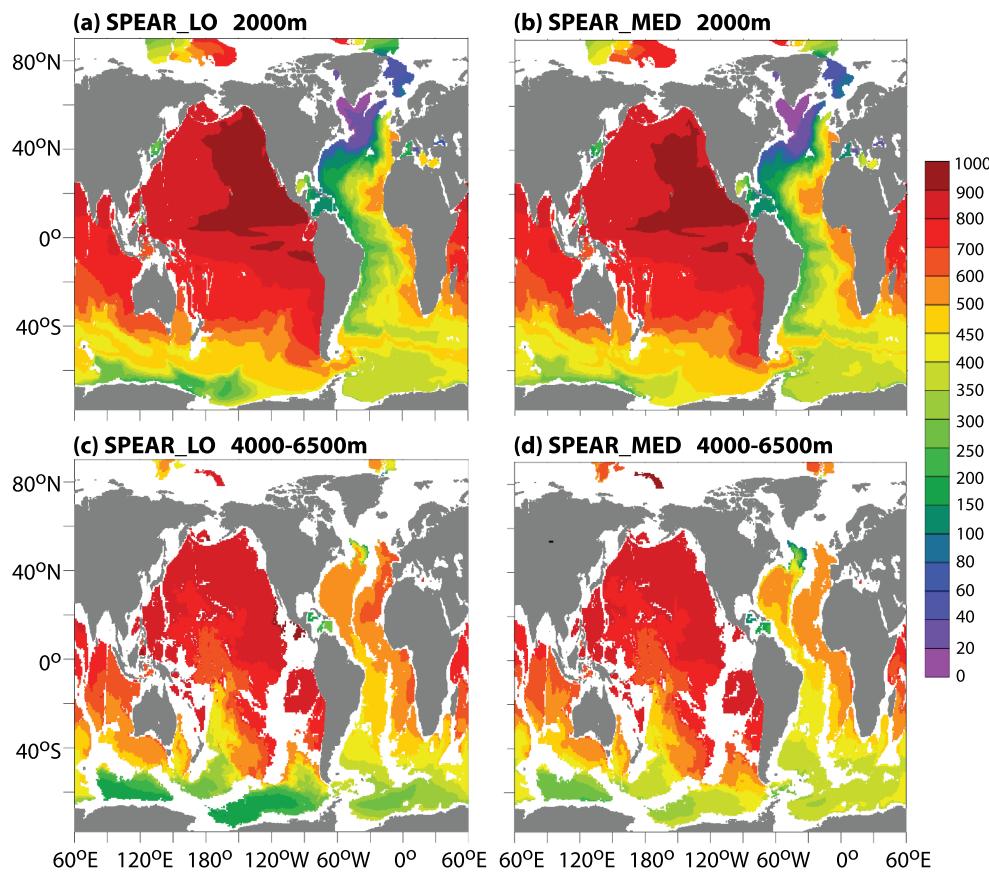
**Figure 20.** Time series of areal extent of Arctic sea ice; units are in millions of square kilometers. (a) March and (b) September. For each panel the red line and symbols are the 30-member ensemble mean from SPEAR\_LO, while the blue line and symbols are the 30-member ensemble mean from SPEAR\_MED. The tan (light blue) shading shows the range of values each year across the ensemble members from SPEAR\_LO (SPEAR\_MED). The magenta shading indicates where the distributions of SPEAR\_LO and SPEAR\_MED overlap. The three vertical dashed lines are for ease of reference to identify years 2000, 2020, and 2040. The thick green line and symbols show the observed sea ice extent for 1979 to 2018. Observed sea ice extent from Fetterer et al. (2017).

water column, eventually leading to abrupt releases of heat from the ocean through large, decadal-scale convective events that appeared inconsistent with the observational record.

In the model development process many sensitivity studies were done examining multiple processes related to SO deepwater formation, including tidal and other vertical mixing processes in the ocean. We found a crucial dependence of SO deepwater formation on the surface energy balance over Antarctica. The initial value specified for the near-infrared isotropic reflectance parameter for cold snow over glacial surfaces was 0.58 and resulted in albedos that were somewhat lower (~0.73) than suggested by observational estimates (Laine, 2008; Pirazzini, 2004; Wang & Zender, 2011; Wendler & Kelley, 1988) which range from 0.81 to 0.85. Increasing the reflectance parameter to 0.68 in the final version of SPEAR produced albedo values over Antarctica of ~0.81, in better agreement with the limited observational data sets. We found that these adjustments had a substantial impact on SO deepwater formation. We illustrate this aspect of the model development process by conducting a pair of idealized sensitivity studies in which we adjust this parameter over Antarctica and show how this impacted not only SO deepwater formation but characteristics of the global ocean and heat balance. Given the relative similarities between SPEAR\_LO and SPEAR\_MED, we focus on results from SPEAR\_LO.

#### 4.2. Sensitivity Tests

We conduct perturbation experiments using SPEAR\_LO in which we modify the surface energy balance over Antarctica so that the air flowing off the continent is systematically cooled or warmed relative to a control simulation. In these experiments we alter the albedo of snow on glacial surfaces over Antarctica. For experiment “BRIGHT” we increase the near-infrared isotropic reflectance parameter (the main determinant of near-infrared albedo) for cold snow over glacial surfaces south of 60°S from 0.68 to 0.83. This is done only for continental locations. Similarly, for experiment “DARK,” we reduce the value of this parameter from 0.68 to 0.53. In “CONTROL” we use a value of 0.68, as in SPEAR\_LO. Both cases have a parameter perturbation of 0.15 but of differing sign. We anticipate that (a) the BRIGHT (DARK) simulation will reflect more (less) radiation, and therefore will be colder (warmer) than the control simulation over Antarctica, and (b) the air flowing off the Antarctic continent in BRIGHT (DARK) will be colder (warmer) than the control, leading to changes in air-sea heat fluxes and potentially altering ventilation and deepwater formation rates.



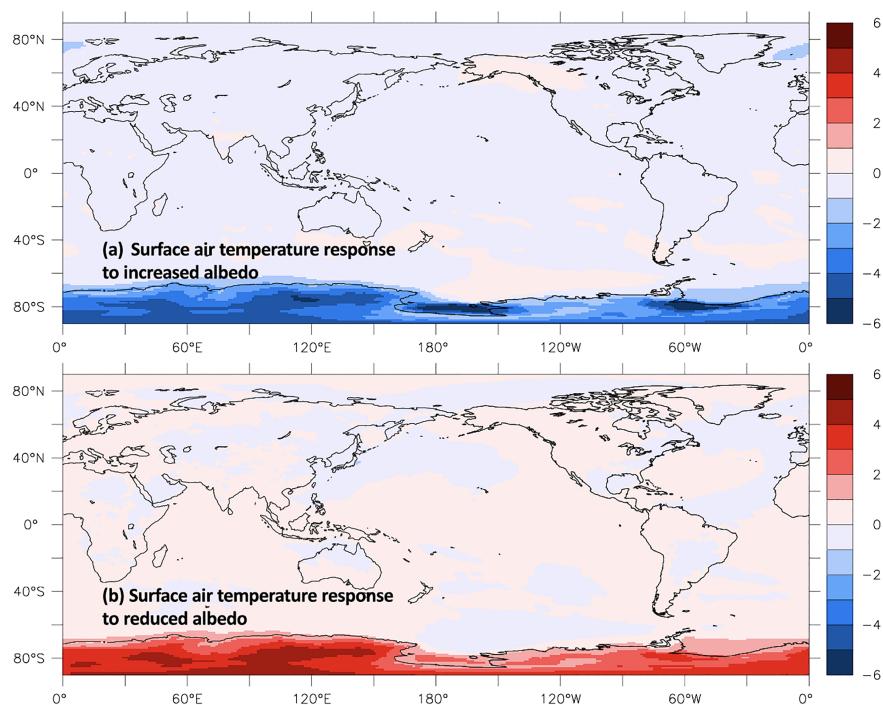
**Figure 21.** Idealized age tracer (units are in years) in SPEAR\_LO and SPEAR\_MED. The values of the age tracer are averaged over years 901–1000 of the 1850 control run. Lower age values indicate that a larger fraction of water at that grid point has recently been in contact with the atmosphere. (a) SPEAR\_LO at 2,000 m, (b) SPEAR\_MED at 2,000 m, (c) SPEAR\_LO averaged over 4,000–6,500 m, and (d) SPEAR\_MED averaged over 4,000–6,500 m.

We deliberately use relatively large perturbations to the reflectance parameter in order to create changes in surface air temperature over Antarctica that are clearly distinguishable from noise.

The sensitivity experiments are started from the same initial conditions as the original 1850 control simulation, and the simulations extend 500 years using 1850 atmospheric composition.

We show in Figure 22 the difference in near-surface air temperature between the perturbed albedo experiments and the control. The response to increased (decreased) albedo is in the top (bottom) panel. Each panel represents time mean differences over years 101–200, calculated as either BRIGHT minus control (top) or DARK minus control (bottom). The response is very stable and robust (not shown). It is clear that the response in near-surface air temperature is what was intended—the albedo changes have produced near-surface air temperature changes that are largely confined to the Antarctic continent, with some small changes over adjacent coastal locations. The responses have larger amplitude for BRIGHT than DARK.

Shown in Figure 23 are the responses in ocean-atmosphere heat fluxes, with negative (positive) values indicating enhanced (reduced) transfer of heat from the ocean to the atmosphere. In the BRIGHT case (top panel) the ocean to atmosphere heat flux is increased, with the opposite response for the DARK case (middle panel). The colder air flowing off the continent extracts more heat from the upper ocean, consistent with an increased air-sea temperature gradient. In addition, the increased heat flux cools the upper ocean, leading to weakened oceanic stratification and more mixing of warmer subsurface waters to the surface layers. This in turn further increases the air-sea heat fluxes. The changes in zonal mean fluxes (bottom panel) can be in excess of  $10 \text{ W m}^{-2}$ , with larger amplitudes for the cooling (BRIGHT) case. This asymmetry (amplitude of changes in heat flux are larger for increase in albedo than for decrease in albedo) is related to the fact that cooling can destabilize the column and tap into larger heat reservoirs at depth. These larger reservoirs



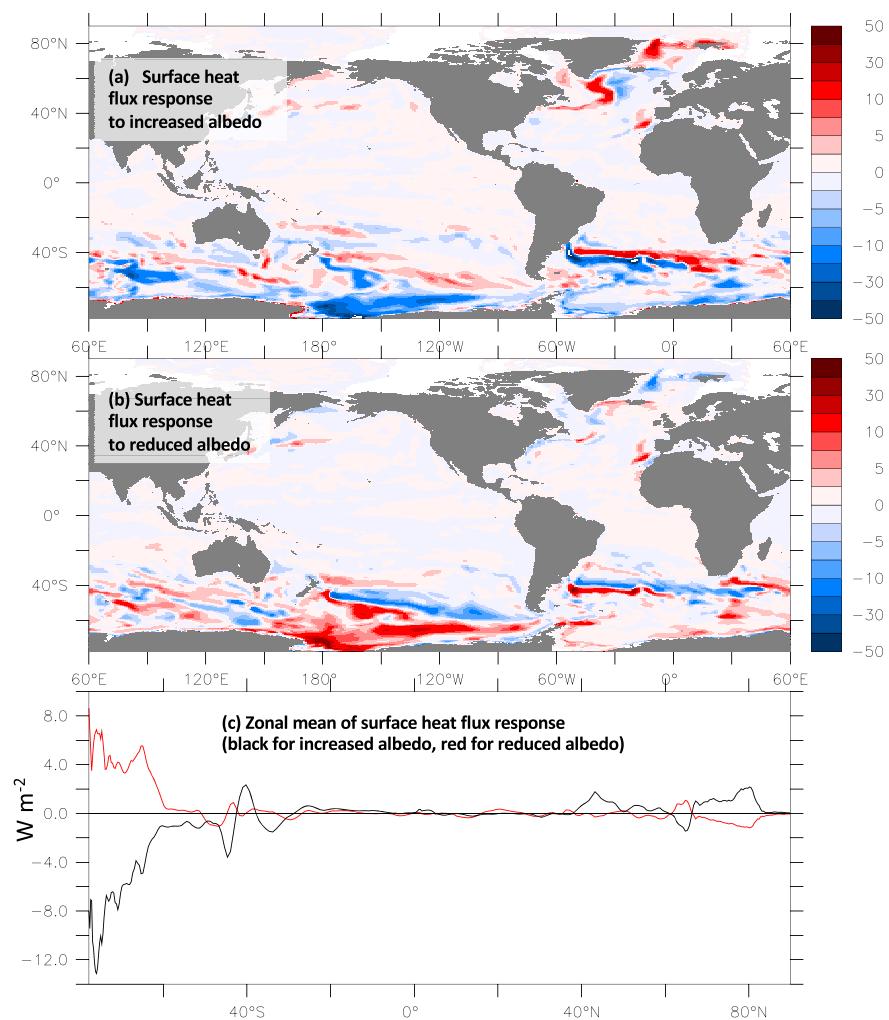
**Figure 22.** Surface air temperature response to imposed albedo changes, calculated as a time mean over years 101–200 of the perturbed simulation minus years 101–200 of the control simulation. Units are in °C. (a) Response to increased surface albedo over Antarctica. (b) Response to reduced surface albedo over Antarctica.

provide a large potential for the amplitude of the heat flux changes. In contrast, reduced albedo tends to increase oceanic stratification through warming of the near-surface atmosphere, thereby weakening mixing between the near-surface ocean and deeper layers. The feedback process that can amplify a response to the increased albedo case (tapping into deep ocean heat reservoirs by reducing vertical stability) is not present for the reduced albedo/warming case, and the amplitude of the response is smaller.

The changes in surface heat fluxes have a profound impact on the interior ocean. We show in Figure 24 cross sections of changes in temperature, salinity, and density in response to the albedo changes over Antarctica, calculated as the time mean over years 401–500. The left (right) column shows BRIGHT (DARK) minus control. The top row shows changes in zonal mean temperature. The increased albedo (Figure 24a) leads to a deep cooling of the SO, with that cooling extending at depth well into the Northern Hemisphere. This is consistent with both an enhanced rate of AABW formation and AABW that is colder than in the control simulation. The reduced albedo case (Figure 24d) shows changes of opposite sign but smaller in both amplitude and spatial extent. The reduced ventilation and AABW formation indicate that less heat is released from deeper layers of the SO, leading to subsurface warming. There is a reduction of the (cold) AABW penetration into the rest of the world ocean.

The middle row of Figure 24 shows changes in ocean salinity. The increased albedo case (Figure 24b) shows increased salinity in the near-surface layers around the coast of Antarctica. This is associated with an increased rate of sea ice formation and brine rejection associated with the cooling and is also consistent with increased vertical mixing. There is not a strong global-scale signal of salinity change. For the reduced albedo case (Figure 24e) there is a stronger signal of reduced salinity in the near-surface and increased salinity at depth. The reduced ventilation and AABW formation indicate that the warmer, saltier water at depth is not effectively mixed with the colder, fresher waters in the near-surface layers.

The bottom row of Figure 24 shows changes in zonal mean density. For the increased albedo case (Figure 24c) there is a very clear signal of increase in density for the entire SO and the deepest layers of the rest of the world ocean. This is consistent with enhanced AABW formation and AABW that is colder than in the control simulation. Comparing the temperature changes to the salinity changes, it is clear that

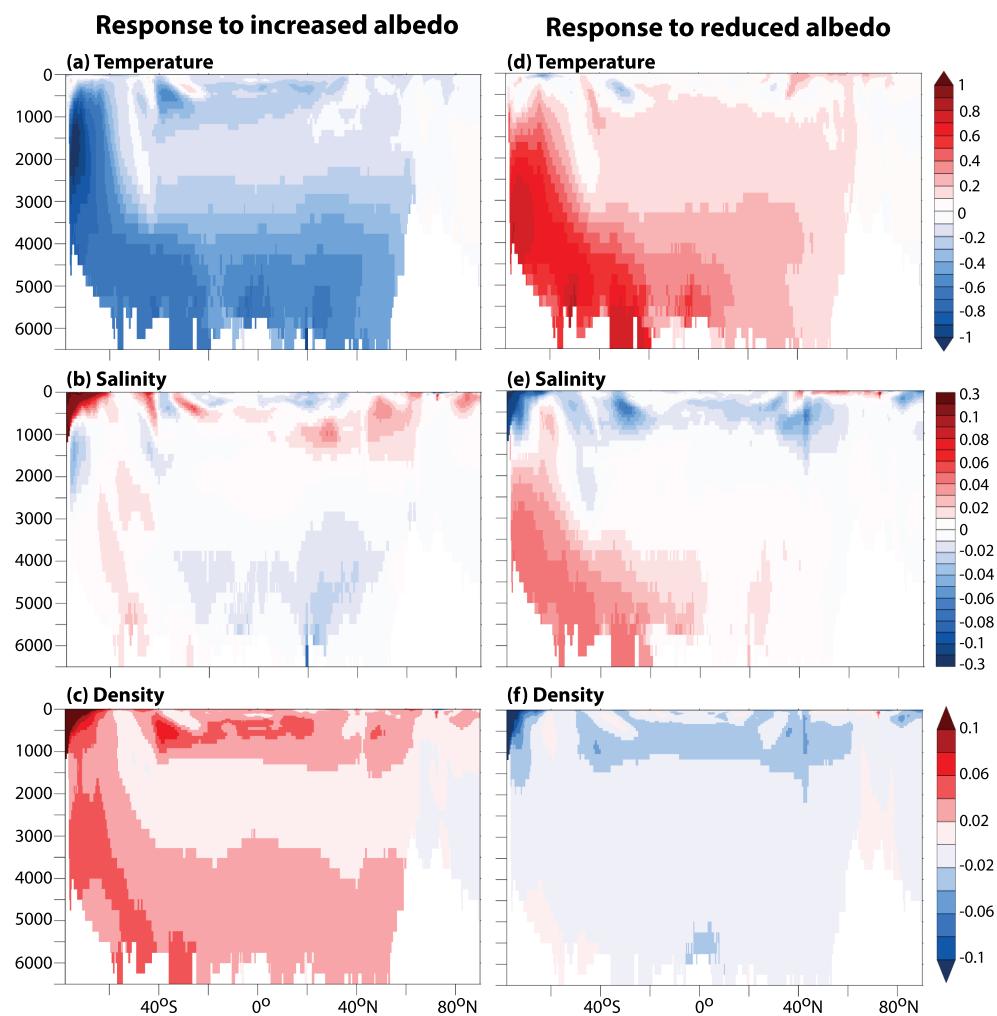


**Figure 23.** (a) Surface net heat flux response to increased surface albedo over Antarctica, calculated over years 101–500 as the net heat flux in the increased albedo run minus net heat flux in control. For all panels negative values indicate an enhanced ocean to atmosphere heat flux (blue shading in (a) and (b)). Units are in  $\text{W m}^{-2}$ . (b) Same as (a) for reduced albedo case. (c) Zonal mean of surface heat flux response ( $\text{W m}^{-2}$ ) for increased albedo case (black) and reduced albedo (red).

most of the increase in density is a result of colder water rather than more saline water, with the exception of the near-surface layers off the coast of Antarctica. There is also an increase in density in the thermocline extending from approximately 50°S to 50°N. This may be related to colder Antarctic Intermediate Water (AIW) than in the control simulation, associated with the enhanced air-sea heat fluxes in the higher latitudes of the SO.

For the reduced albedo case (Figure 24f), there is a reduction of density in the upper ocean at high latitudes of the SO, associated largely with reduced salinity. The reduced rate of AABW formation reduces the mixing of more saline deeper ocean water with the fresher upper ocean, thereby effectively reducing salinity and density. There is also a reduction of salinity in the thermocline, extending from approximately 50°S to 50°N. This appears to be linked to fresher AIW in the reduced albedo simulation than in the control. Since the near-surface waters in the SO are the source waters for AIW, a freshening in the SO, associated with reduced albedo over Antarctica, is then communicated to the global thermocline via altered characteristics of AIW.

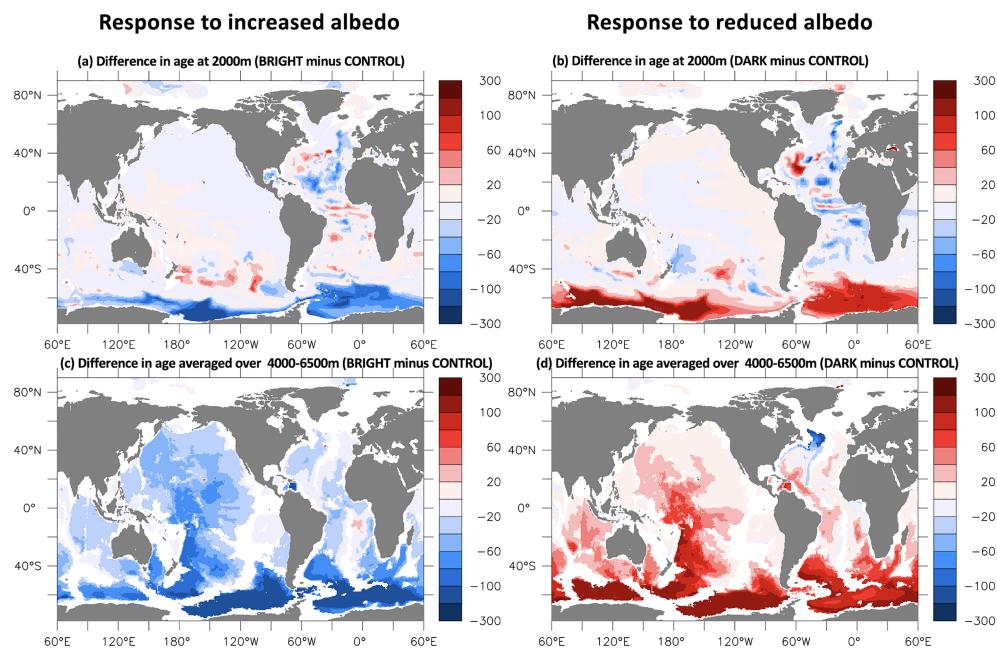
Figure 25 shows the changes in age averaged over years 491–500, illustrating a long-term adjustment of the system to the altered conditions over Antarctica. The left (right) column shows the response for the BRIGHT (DARK) case minus the control. The top row shows results at 2,000 m, while the bottom row shows results



**Figure 24.** Zonal mean (over all longitudes) of the oceanic responses to changes in albedo over Antarctica, calculated as perturbed experiment minus control for years 401–500 of the perturbed experiments. Left column shows responses to increased albedo; right column shows responses to reduced albedo. (a, d) Temperature, units are in °C. (b, e) Salinity, units are in psu. (c, f) Density, units are in  $\text{kg m}^{-3}$ .

averaged from 4,000 m to the bottom of the ocean. At 2,000 m there is reduced (increased) age in the SO for the BRIGHT (DARK) case. In both cases there are changes in the Atlantic, but they lack a well-defined spatial structure. In the deep ocean (4,000 m to the bottom of the ocean) the changes have extended northward out of the SO, extending through much of the Northern Hemisphere. The age reductions for BRIGHT extend well into the North Pacific, much further than for the DARK case. These changes are consistent with the cross sections of temperature shown in Figure 24 and the generally weaker response in the DARK case. This difference in propagation of the age change signal is consistent with the larger amplitude of heat flux change in the increased albedo case relative to the reduced heat flux case. There is also a notable signal of younger water at depth in the subpolar North Atlantic.

Stream functions of the global mean overturning circulation in density space ( $\text{GMOC}_\rho$ ) are illustrated in Figure 26. To calculate the stream function, we first compute the integral over all longitudes of the poleward volume transport within each density layer and then calculate the indefinite integral in the vertical of that term. Flow is counterclockwise around a minimum in stream function, with the strength of the flow proportional to the gradient of the stream function. We interpret the minimum value of the  $\text{GMOC}_\rho$  over the domain 90°S to 55°S as a measure of the strength of AABW formation—more negative values indicate a stronger  $\text{GMOC}_\rho$ . The three panels show stream functions for DARK (top), CONTROL (middle), and BRIGHT (bottom). Moving from lowest to highest albedo, the stream function minimum in the SO both

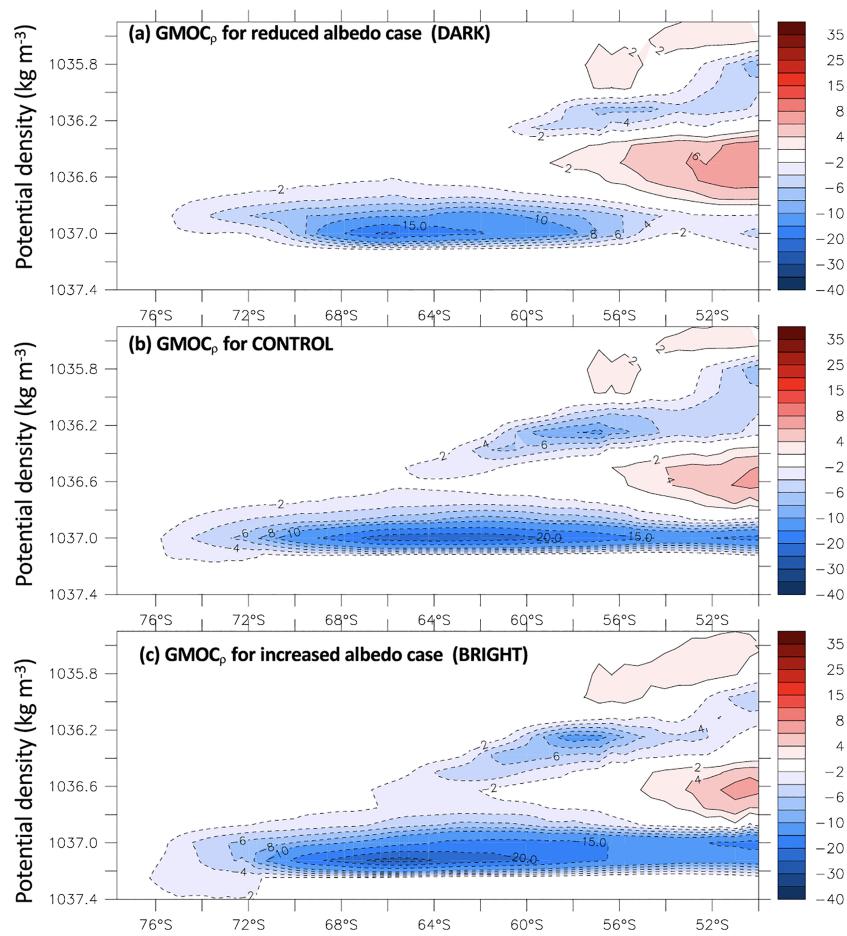


**Figure 25.** Differences in ocean age, averaged over years 491–500, in response to albedo changes over Antarctica. Units are in years. Differences are calculated as age in perturbation experiment minus age in control simulation. Blue indicates younger water in the perturbed experiment. Left column shows responses to increased albedo; right column shows responses to reduced albedo. (a, b) Difference in age at 2,000 m. (c, d) Difference in age averaged over 4,000–6,500 m.

strengthens and moves to greater values in density space, indicating that AABW is becoming denser. This is consistent with the changes in air-sea fluxes and density changes shown previously. The bottom cell (indicated by negative stream function values at depth) extends progressively farther northward as the albedo increases, showing the stronger impact of AABW on world ocean circulation. There is a suggestion of a weaker North Atlantic overturning circulation as the AABW strengthens and moves northward, potentially consistent with AMOC reconstructions from periods of glacial maxima (Böhm et al., 2015).

An important feature of the global ocean circulation is the Antarctic Circumpolar Current (ACC). This zonal current is related to the meridional gradient of density between the coast of Antarctica and the coast of South America. A stronger (weaker) gradient, with denser water along the coast of Antarctica, is associated with a stronger (weaker) zonal current. Since the albedo change over Antarctica alters this meridional density gradient (see Figure 24, bottom panels), we would expect an impact on the ACC. Figure 27a shows the time series of the ACC, calculated as the annual mean volume transport through the Drake Passage. The three curves differ sharply with the strongest (weakest) ACC associated with the BRIGHT (DARK) case. The variation is substantial, from approximately 110 Sv ( $1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$ ) in the reduced albedo case to almost 180 Sv in the increased albedo case. This substantial sensitivity points to the important influence of the energy balance over Antarctica on the ACC. An intriguing feature is the marked near-centennial-scale variability in the control simulation that appears weaker in the perturbed simulations. This variability has resemblance to variability seen in other versions of GFDL models and that has been hypothesized to play a role in the interpretation of recent decadal trends in sea ice and SST in the SO (Zhang et al., 2019).

As we have seen from zonal mean temperature and age, the responses to the albedo changes over Antarctica propagate through the world ocean. We show in Figure 27b the time series of the global mean net radiative balance at the top of the atmosphere. There is a clear dependence on Antarctic albedo, with the reduced (increased) albedo case having the most (least) positive imbalance at the top of the atmosphere. These differences persist throughout the 500 years of the simulations. Since the world ocean is the largest reservoir of heat in the climate system, the net radiative imbalance at the top of the atmosphere is usually linked with changes in the global mean, volume mean ocean temperature. We show time series of global mean, volume mean ocean temperature in Figure 27c. There is a wide range in long-term ocean heating or cooling,



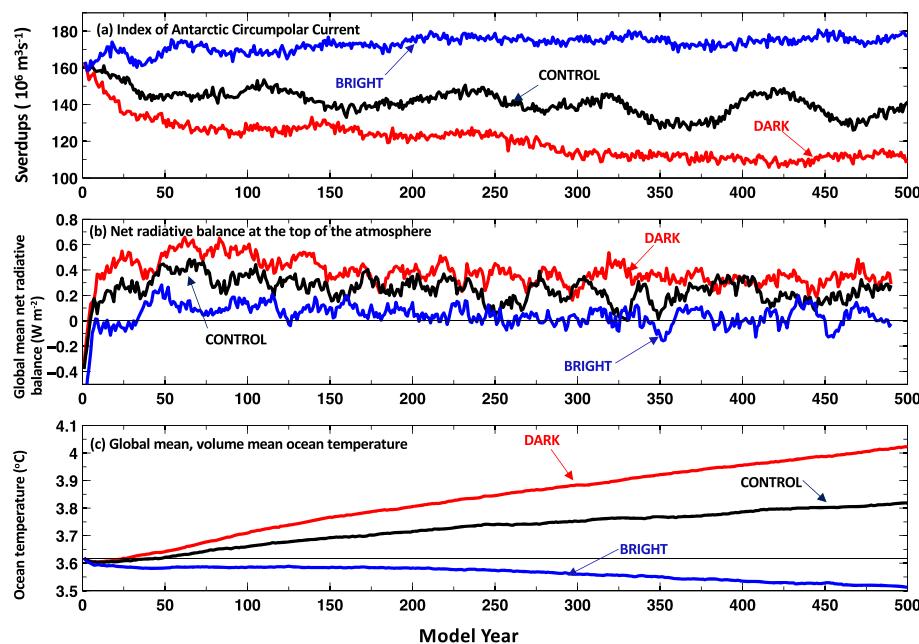
**Figure 26.** Global mean overturning circulation ( $\text{GMOC}_p$ ) in isopycnal coordinates for three cases: (a) reduced albedo (dark), (b) control, and (c) increased albedo (bright). Vertical axis is density; horizontal axis is latitude. Flow is clockwise around a maximum in stream function.

dependent on Antarctic albedo. From one perspective this is not surprising since Antarctic albedo contributes to global albedo and hence to global net radiative balance. However, it is not simply the direct impact of local albedo changes over Antarctica that is crucial; the related changes in ocean heat uptake, associated with flux-driven changes in SO deepwater formation, play a fundamental role in the long-term changes in global mean ocean temperature.

## 5. Summary and Discussion

We have documented the development and simulation characteristics of a newly developed coupled ocean-atmosphere-land-sea ice model (“SPEAR”). This model is intended for studies of seasonal to multidecadal climate variability, predictability, and change, as well as experimental real-time seasonal to decadal prediction. This new modeling system incorporates many recent developments at GFDL, including improvements in the atmosphere, ocean, land and sea ice components, and the capability for variable resolution grids. The biases and simulation characteristics of the SPEAR model are significantly improved relative to corresponding characteristics in the previous generation GFDL seasonal to decadal prediction models: CM2.1 (Delworth et al., 2006) and FLOR (Vecchi et al., 2014). We also present the design characteristics and simulation characteristics of new 30-member ensembles of climate change simulations using SPEAR\_LO and SPEAR\_MED, extending from 1851 to 2100 using observed and projected changes in atmospheric composition and radiative forcing.

We also examined the sensitivity of aspects of the SO and global climate system to the surface energy balance over Antarctica. Part of the motivation for these sensitivity experiments came from the early stages of the



**Figure 27.** Time series of various annual mean quantities for simulations DARK, CONTROL, and BRIGHT. (a) Strength of Antarctic Circumpolar Current (ACC) as measured by zonal eastward volume flow over the depth of the ocean at  $72^{\circ}\text{W}$  between the coast of Antarctica and South America in the control and two albedo perturbation experiments. Units are in Sverdups ( $1 \text{ Sv} = 10^6 \text{ m}^3 \text{s}^{-1}$ ). (b) Global mean net radiative balance at the top of the atmosphere for the control and two albedo perturbation simulations (units are in  $\text{W m}^{-2}$ ). (c) Global mean, volume mean ocean temperature (units are in  $^{\circ}\text{C}$ ).

SPEAR development process during which the model produced little or no deepwater in the SO. This appeared to be related to relative warmth and freshness in the upper ocean around Antarctica and led to relatively strong stratification in the SO. This deficiency contributed to substantial long-term drifts in the world ocean; the reduced ventilation led to a buildup of heat in the interior of the SO from the transport of warmer waters from the Atlantic into the SO.

Multiple sensitivity tests were conducted to explore this bias involving ocean mixing and surface energy balance terms. We found that changing the heat balance over Antarctica by altering aspects of the snow albedo formulation had a strong impact on SO deepwater formation. We explored this in more detail by conducting additional sensitivity tests in which we altered terms in the snow albedo formulation. When albedo is increased, surface air temperature is reduced over Antarctica relative to a control simulation. The cooled air flows out over the SO and, due to an enhanced ocean-atmosphere temperature gradient, extracts more heat from the ocean than in a comparable control simulation. This cools the near-surface ocean, increasing upper ocean density through both cooling and enhanced sea ice formation and brine rejection, increases ventilation and deepwater formation, and leads to the formation of colder, denser AABW that spreads at depth through the world ocean. More locally in the SO, the increased ocean density near the coast of Antarctica increases the meridional density gradient between Antarctica and South America, increasing the strength of the Antarctic Circumpolar Current (ACC) as measured by volume flow through the Drake Passage. The deep cooling of the world ocean leads to a substantial reduction in global ocean heat uptake in long control simulations, as well as reduced overall model drift. Generally, opposite changes occur for the case of reduced albedo (“DARK”), although the amplitude of the impacts is somewhat smaller.

One of the additional intriguing findings in this study is that this model is able to simulate dense water formation through shelf processes that lead to dense water flowing down along the continental slope into the deepest layers of the SO (see supporting information Figure S10). This process occurs in addition to open ocean convection in the SO. Formation of AABW from shelf waters has been a challenge for coarse-resolution models, and we find this development to be encouraging. This may be the result of improved numerical algorithms and/or the use of hybrid coordinates in MOM6 and remains the subject of ongoing exploration. The SO is difficult to model due to the importance of small-scale processes, such as mesoscale

eddies, small-scale mixing processes, topographic variations, and the difficulty in simulating clouds and the surface radiation balance over the SO and Antarctica.

This study does not include the effects of freshwater from melting of Antarctic land ice (e.g., ice shelf and ice sheet) that could contribute to surface freshening and increased stratification. Inclusion of that process (Bronselaer et al., 2018) may well lead to reduced AABW formation but is beyond the scope of the present study.

The SPEAR model will be the primary tool for GFDL participation in a number of national and international prediction and attribution activities over the next several years. These include the North American MultiModel Ensemble for seasonal prediction (Kirtman et al., 2014) which uses models from five modeling centers across North America (<https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/north-american-multi-model-ensemble>). Real-time seasonal predictions are conducted once per month. SPEAR will also be used for decadal prediction activities contributing to CMIP6 and with ongoing predictions being made available through the U.K. Met Office as a WMO Lead Centre for Interannual to Decadal Prediction (<https://www.metoffice.gov.uk/research/climate/seasonal-to-decadal/long-range/wmolc-adcp>). In addition, ensembles of SPEAR at high resolution are being used for the prediction and projection of changes in climate extremes.

The current version of the atmospheric component of SPEAR has 33 vertical levels. Work is underway exploring the simulation characteristics of a version of this model with 65 levels. This version has a substantially improved simulation of the climatological characteristics of sudden stratospheric warmings and their connection to near-surface climate, but the computational cost of the model is also doubled. This version of SPEAR will be used for exploration of the impact of the stratosphere on seasonal to decadal climate variability and predictability.

## Acknowledgments

We thank Drs. Leo Donner and John Dunne, as well as two anonymous reviewers, for providing valuable comments on preliminary versions of this manuscript. Financial support for this work was provided through base funding from the National Oceanic and Atmospheric Administration to the Geophysical Fluid Dynamics Laboratory. Model output relevant to this paper can be found online ([ftp://data1.gfdl.noaa.gov/users/Tom.Delworth/SPEAR\\_Documentation\\_paper/SPEAR/](ftp://data1.gfdl.noaa.gov/users/Tom.Delworth/SPEAR_Documentation_paper/SPEAR/)).

## 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(10), 3167–3211. <https://doi.org/10.1029/2019MS001726>
- Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., et al. (2003). The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present). *Journal of Hydrometeorology*, 4, 1147–1167.
- Allan, R. P., Liu, C., Loeb, N. G., Palmer, M. D., Roberts, M., Smith, D., & Vidale, P.-L. (2014). Changes in global net radiative imbalance 1985–2012. *Geophysical Research Letters*, 41(15), 5588–5597. <https://doi.org/10.1002/2014GL060962>
- Anderson, J. L., Balaji, V., Broccoli, A. J., Cooke, W. F., Delworth, T. L., Dixon, K. W., et al. (2004). The new GFDL global atmosphere and land model AM2-LM2: Evaluation with prescribed SST simulations. *Journal of Climate*, 17(24), 4641–4673. <https://doi.org/10.1175/JCLI-3223.1>
- Böhm, E., Lippold, J., Gutjahr, M., Frank, M., Blaser, P., Antz, B., et al. (2015). Strong and deep Atlantic meridional overturning circulation during the last glacial cycle. *Nature*, 517(7532), 73–76. <https://doi.org/10.1038/nature14059>
- Boyer, T. P., Antonov, J. I., Baranova, O. K., Garcia, H. E., Johnson, D. R., Mishonov, A. V., et al. (2013). World ocean database 2013. <https://doi.org/10.7289/v5nz85mt>
- Bronselaer, B., Winton, M., Griffies, S. M., Hurlin, W. J., Rodgers, K. B., Sergienko, O. V., et al. (2018). Change in future climate due to Antarctic meltwater. *Nature*, 564(7734), 53–58. <https://doi.org/10.1038/s41586-018-0712-z>
- Cavalieri, D. J., Parkinson, C. L., Gloersen, P., & Zwally, H. J. (1996). *Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS passive microwave data, version 1*. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. Boulder, Colorado USA: NASA National Snow and Ice Data Center Distributed Active Archive Center. <https://doi.org/doi.org/10.5067/8GQ8LZQVLOVL>
- Cayan, D. R., Dettinger, M. D., Kammerdiener, S. A., Caprio, J. M., & Peterson, D. H. (2001). Changes in the onset of spring in the western United States. *Bulletin of the American Meteorological Society*, 82(3), 399–415. [https://doi.org/10.1175/1520-0477\(2001\)082<0399:CITOOS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<0399:CITOOS>2.3.CO;2)
- Chen, J.-H., & Lin, S.-J. (2013). Seasonal predictions of tropical cyclones using a 25-km-resolution general circulation model. *Journal of Climate*, 26(2), 380–398. <https://doi.org/10.1175/JCLI-D-12-00061.1>
- Chen, J.-H., Lin, S.-J., Zhou, L., Chen, X., Rees, S., Bender, M., & Morin, M. (2019). Evaluation of tropical cyclone forecasts in the next generation global prediction system. *Monthly Weather Review*, 147(9), 3409–3428. <https://doi.org/10.1175/MWR-D-18-0227.1>
- Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Gao, X., et al. (2013). Long-term climate change: Projections, commitments and irreversibility. In T. F. Stocker 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* (pp. 1029–1136). Cambridge: Cambridge University Press.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597. <https://doi.org/10.1002/qj.828>
- 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. Retrieved from. <http://journals.ametsoc.org/doi/abs/10.1175/JCLI3629.1>

- Delworth, T. L., & Dixon, K. W. (2006). Have anthropogenic aerosols delayed a greenhouse gas-induced weakening of the North Atlantic thermohaline circulation? *Geophysical Research Letters*, 33, L02606. <https://doi.org/10.1029/2005GL024980>
- Delworth, T. L., Rosati, A., Anderson, W., Adcroft, A. J., Balaji, V., Benson, R., et al. (2012). Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *Journal of Climate*, 25(8), 2755–2781. <https://doi.org/10.1175/JCLI-D-11-00316.1>
- Deser, C., Phillips, A. S., & Hurrell, J. W. (2004). Pacific interdecadal climate variability: Linkages between the tropics and the North Pacific during boreal winter since 1900. *Journal of Climate*, 17(16), 3109–3124. [https://doi.org/10.1175/1520-0442\(2004\)017<3109:PICVLB>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<3109:PICVLB>2.0.CO;2)
- Donner, L. J., Wyman, B. L., Hemler, R. S., Horowitz, L. W., Ming, Y., Zhao, M., et al. (2011). The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3. *Journal of Climate*, 24(13), 3484–3519. <https://doi.org/10.1175/2011JCLI3955.1>
- Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., et al. (2012). GFDL's ESM 2 global coupled climate-carbon earth system models. Part I: Physical formulation and baseline simulation characteristics. *Journal of Climate*, 25(19), 6646–6665. <https://doi.org/10.1175/JCLI-D-11-00560.1>
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model inter-comparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- Fang, C., Wu, L., & Zhang, X. (2014). The impact of global warming on the pacific decadal oscillation and the possible mechanism. *Advances in Atmospheric Sciences*, 31(1), 118–130. <https://doi.org/10.1007/s00376-013-2260-7>
- Fetterer, F., Knowles, K., Meier, W. N., Savoie, M., & Windnagel, A. K. (2017). Updated daily. Sea Ice Index, Version 3. Boulder, CO: NSIDC: National Snow and Ice Data Center. <https://doi.org/10.7265/N5K072F8>. Data accessed Aug 5, 2019.
- Flato, G., Marotzke, J., Abiodun, B., Bracconot, P., Chou, S. C., Collins, W., et al. (2013). Evaluation of climate models. In T. F. Stocker 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* (Chap. 9, pp. 741–866). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.
- Frölicher, T. L., Sarmiento, J. L., Paynter, D. J., Dunne, J. P., Krasting, J. P., & Winton, M. (2015). Dominance of the Southern Ocean in anthropogenic carbon and heat uptake in CMIP5 models. *Journal of Climate*, 28(2), 862–886. <https://doi.org/10.1175/JCLI-D-14-00117.1>
- Gidden, M. J., Riahi, K., Smith, S. J., Fujimori, S., Luderer, G., Kriegler, E., et al. (2019). Global emissions pathways under different socioeconomic scenarios for use in CMIP6: A dataset of harmonized emissions trajectories through the end of the century. *Geoscientific Model Development*, 12(4), 1443–1475. <https://doi.org/10.5194/gmd-12-1443-2019>
- Griffies, S. M., Winton, M., Donner, L. J., Horowitz, L. W., Downes, S. M., Farneti, R., et al. (2011). The GFDL CM3 coupled climate model: Characteristics of the ocean and sea ice simulations. *Journal of Climate*, 24(13), 3520–3544. <https://doi.org/10.1175/2011JCLI3964.1>
- Harris, L. M., & Lin, S.-J. (2013). A two-way nested global-regional dynamical core on the cubed-sphere grid. *Monthly Weather Review*, 141(1), 283–306. <https://doi.org/10.1175/MWR-D-11-00201.1>
- 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(11), 3691–3727. <https://doi.org/10.1029/2019MS001829>
- Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., & Vallis, G. K. (2010). Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *Journal of Climate*, 23(9), 2418–2427. <https://doi.org/10.1175/2009JCLI3466.1>
- Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkänen, T., et al. (2018). Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geoscientific Model Development*, 11(1), 369–408. <https://doi.org/10.5194/gmd-11-369-2018>
- Horowitz, L. W., Walters, S., Mauzerall, D. L., Emmons, L. K., Rasch, P. J., Granier, C., et al. (2003). A global simulation of tropospheric ozone and related tracers: Description and evaluation of MOZART, version 2: MOZART-2 description and evaluation. *Journal of Geophysical Research*, 108(D24), 4784. <https://doi.org/10.1029/2002JD002853>
- Huang, B., Banzon, V. F., Freeman, E., Lawrimore, J., Liu, W., Peterson, T. C., et al. (2015). Extended Reconstructed Sea Surface Temperature Version 4 (ERSST.v4). Part I: Upgrades and intercomparisons. *Journal of Climate*, 28(3), 911–930. <https://doi.org/10.1175/JCLI-D-14-00006.1>
- Jansen, M. F., Held, I. M., Adcroft, A., & Hallberg, R. (2015). Energy budget-based backscatter in an eddy permitting primitive equation model. *Ocean Modelling*, 94, 15–26. <https://doi.org/10.1016/j.ocemod.2015.07.015>
- Johnson, G. C., Lyman, J. M., & Loeb, N. G. (2016). Improving estimates of Earth's energy imbalance. *Nature Climate Change*, 6(7), 639–640. <https://doi.org/10.1038/nclimate3043>
- Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., & Morice, C. P. (2012). Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010: Land-surface temperature variations. *Journal of Geophysical Research*, 117, D05127. <https://doi.org/10.1029/2011JD017139>
- Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., et al. (2014). The North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, 95(4), 585–601. <https://doi.org/10.1175/BAMS-D-12-00050.1>
- Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010). The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone data. *Bulletin of the American Meteorological Society*, 91(3), 363–376. <https://doi.org/10.1175/2009BAMS2755.1>
- Knutson, T. R., Delworth, T. L., Dixon, K. W., Held, I. M., Lu, J., Ramaswamy, V., et al. (2006). Assessment of twentieth-century regional surface temperature trends using the GFDL CM2 coupled models. *Journal of Climate*, 19(9), 1624–1651. <https://doi.org/10.1175/JCLI3709.1>
- 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>
- Laine, V. (2008). Antarctic ice sheet and sea ice regional albedo and temperature change, 1981–2000, from AVHRR polar pathfinder data. *Remote Sensing of Environment*, 112, 646–667.
- Landschützer, P., Gruber, N., Haumann, F. A., Rödenbeck, C., Bakker, D. C. E., van Heuven, S., et al. (2015). The reinvigoration of the Southern Ocean carbon sink. *Science*, 349(6253), 1221–1224. <https://doi.org/10.1126/science.aab2620>
- Latif, M., Martin, T., Reintges, A., & Park, W. (2017). Southern Ocean decadal variability and predictability. *Current Climate Change Reports*, 3(3), 163–173. <https://doi.org/10.1007/s40641-017-0068-8>

- Lennsen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., & Zyss, D. (2019). Improvements in the GISTEMP uncertainty model. *Journal of Geophysical Research-Atmospheres*, 124, 6307–6326. <https://doi.org/10.1029/2018JD029522>
- Loeb, N. G., Lyman, J. M., Johnson, G. C., Allan, R. P., Doelling, D. R., Wong, T., et al. (2012). Observed changes in top-of-the-atmosphere radiation and upper-ocean heating consistent within uncertainty. *Nature Geoscience*, 5(2), 110–113. <https://doi.org/10.1038/ngeo1375>
- Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., et al. (2009). Toward optimal closure of the Earth's top-of-atmosphere radiation budget. *Journal of Climate*, 22(3), 748–766. <https://doi.org/10.1175/2008JCLI2637.1>
- Madden, R. A., & Julian, P. R. (1971). Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *Journal of the Atmospheric Sciences*, 28, 702–708.
- 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, 1069–1079.
- Marshall, D. P., & Adcroft, A. J. (2010). Parameterization of ocean eddies: Potential vorticity mixing, energetics and Arnold's first stability theorem. *Ocean Modelling*, 32(3–4), 188–204. <https://doi.org/10.1016/j.ocemod.2010.02.001>
- Matthes, K., Funke, B., Andersson, M. E., Barnard, L., Beer, J., Charbonneau, P., et al. (2017). Solar forcing for CMIP6 (v3.2). *Geoscientific Model Development*, 10(6), 2247–2302. <https://doi.org/10.5194/gmd-10-2247-2017>
- Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., et al. (2017). Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geoscientific Model Development*, 10(5), 2057–2116. <https://doi.org/10.5194/gmd-10-2057-2017>
- Merlis, T. M., Held, I. M., Stenchikov, G. L., Zeng, F., & Horowitz, L. W. (2014). Constraining transient climate sensitivity using coupled climate model simulations of volcanic eruptions. *Journal of Climate*, 27(20), 7781–7795. <https://doi.org/10.1175/JCLI-D-14-00214.1>
- Milly, P. C. D., Malyshev, S. L., Sheviakova, E., Dunne, K. A., Findell, K. L., Gleeson, T., et al. (2014). An enhanced model of land water and energy for global hydrologic and Earth-system studies. *Journal of Hydrometeorology*, 15(5), 1739–1761. <https://doi.org/10.1175/JHM-D-13-0162.1>
- Murakami, H., Villarini, G., Vecchi, G. A., Zhang, W., & Gudgel, R. (2016). Statistical-dynamical seasonal forecast of North Atlantic and U.S. landfalling tropical cyclones using the high-resolution GFDL FLOR coupled model. *Monthly Weather Review*, 144(6), 2101–2123. <https://doi.org/10.1175/MWR-D-15-0308.1>
- Murray, R. J. (1996). Explicit generation of orthogonal grids for ocean models. *Journal of Computational Physics*, 126, 251–273.
- Naik, V., Horowitz, L. W., Fiore, A. M., Ginoux, P., Mao, J., Aghedo, A. M., & Levy, H. (2013). Impact of preindustrial to present-day changes in short-lived pollutant emissions on atmospheric composition and climate forcing: Impact of short-lived pollutant emission. *Journal of Geophysical Research-Atmospheres*, 118(14), 8086–8110. <https://doi.org/10.1002/jgrd.50608>
- Pirazzini, R. (2004). Surface albedo measurements over Antarctic sites in summer. *Journal of Geophysical Research*, 109, D20118. <https://doi.org/10.1029/2004JD004617>
- 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>
- Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., et al. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of Climate*, 24(14), 3624–3648. <https://doi.org/10.1175/JCLI-D-11-00015.1>
- Schneider, U., Finger, P., Meyer-Christoffer, A., Rustemeier, E., Ziese, M., & Becker, A. (2017). Evaluating the hydrological cycle over land using the newly-corrected precipitation climatology from the Global Precipitation Climatology Centre (GPCC). *Atmosphere*, 8(12), 52. <https://doi.org/10.3390/atmos8030052>
- Shi, J.-R., Xie, S.-P., & Talley, L. D. (2018). Evolving relative importance of the Southern Ocean and North Atlantic in anthropogenic ocean heat uptake. *Journal of Climate*, 31(18), 7459–7479. <https://doi.org/10.1175/JCLI-D-18-0170.1>
- Thomason, L. W., Ernest, N., Millán, L., Rieger, L., Bourassa, A., Vernier, J.-P., et al. (2018). A global space-based stratospheric aerosol climatology: 1979–2016. *Earth System Science Data*, 10(1), 469–492. <https://doi.org/10.5194/essd-10-469-2018>
- van Marle, M. J. E., Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A.-L., Field, R. D., et al. (2017). Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fire models (1750–2015). *Geoscientific Model Development*, 10(9), 3329–3357. <https://doi.org/10.5194/gmd-10-3329-2017>
- Vecchi, G. A., Delworth, T., Gudgel, R., Kapnick, S., Rosati, A., Wittenberg, A. T., et al. (2014). On the seasonal forecasting of regional tropical cyclone activity. *Journal of Climate*, 27(21), 7994–8016. <https://doi.org/10.1175/JCLI-D-14-00158.1>
- Wang, X., & Zender, C. (2011). Arctic and Antarctic diurnal and seasonal variations of snow albedo from multiyear Baseline Surface Radiation Network measurements. *Journal of Geophysical Research*, 116(F3), F03008. <https://doi.org/10.1029/2010JF001864>
- Wendler, G., & Kelley, J. (1988). On the albedo of snow in Antarctica: A contribution to I.A.G.O. *Journal of Glaciology*, 34, 19–25.
- 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>
- 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. (2016a). Impact of the Antarctic bottom water formation on the Weddell Gyre and its northward propagation characteristics in GFDL CM2.1 model: AABW and Weddell gyre characteristics. *Journal of Geophysical Research, Oceans*, 121(8), 5825–5846. <https://doi.org/10.1002/2016JC011790>
- Zhang, L., & Delworth, T. L. (2016b). Simulated response of the Pacific Decadal Oscillation to climate change. *Journal of Climate*, 29(16), 5999–6018. <https://doi.org/10.1175/JCLI-D-15-0690.1>
- Zhang, S., Harrison, M. J., Rosati, A., & Wittenberg, A. (2007). System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Monthly Weather Review*, 135(10), 3541–3564. <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(3), 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(3), 735–769. <https://doi.org/10.1002/2017MS001209>

- Zhao, M., Held, I. M., Lin, S.-J., & Vecchi, G. A. (2009). Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *Journal of Climate*, 22(24), 6653–6678. <https://doi.org/10.1175/2009JCLI3049.1>
- Zhou, L., Lin, S.-J., Chen, J.-H., Harris, L. M., Chen, X., & Rees, S. L. (2019). Toward convective-scale prediction within the next generation global prediction system. *Bulletin of the American Meteorological Society*, 100(7), 1225–1243. <https://doi.org/10.1175/BAMS-D-17-0246.1>