JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2019MS001829

Special Section:
Geophysical Fluid Dynamics Laboratory CMIP6 Models

Structure and Performance of GFDL's CM4.0 Climate Model

I. M. Held[1], H. Guo[2], A. Adcroft[2], J. P. Dunne[2], L. W. Horowitz[1,2], J. Krasting[2], E. Shevliakova[2], M. Winton[2], M. Zhao[2], M. Bushuk[3], A. T. Wittenberg[2], B. Wyman[2], B. Xiang[3], R. Zhang[1,2], W. Anderson[2], V. Balaji[4], L. Donner[2,5], K. Dunne[5], J. Durachta[2], P. P. G. Gauthier[4], P. Ginoux[2], J.-C. Golaz[6], S. M. Griffies[1,2], R. Hallberg[1,2], L. Harris[2], M. Harrison[2], W. Hurlin[2], J. John[2], P. Lin[4], S.-J. Lin[2], S. Malyshev[2], R. Menzel[3], P. C. D. Milly[2,5], Y. Ming[2], V. Naik[2], D. Paynter[2], F. Paulot[2], V. Ramaswamy[1,2], B. Reichl[2], T. Robinson[7], A. Rosati[4], C. Seman[2], L. G. Silvers[4], S. Underwood[2], and N. Zadeh[7]

[1]Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ, USA, [2]GFDL/NOAA, Princeton, NJ, USA, [3]University Corporation for Atmospheric Research/GFDL, Princeton, NJ, USA, [4]Cooperative Institute for Modeling the Earth System, Princeton University, Princeton, NJ, USA, [5]U.S. Geological Survey/GFDL, Princeton, NJ, USA, [6]Lawrence Livermore National Laboratory, Livermore, CA, USA, [7]SAIC, Science Applications International Corporation, Reston, VA, USA

Key Points:
- A team at GFDL has developed a new model of the physical climate system referred to as CM4.0
- Strengths of model include ENSO simulation and small biases in TOA fluxes, precipitation, Arctic sea ice extent, and sea surface temperature
- Problematic aspects include large variability in Southern Ocean and historical simulation with little warming prior to 1990

Correspondence to:
H. Guo,
huan.guo@noaa.gov

Citation:
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*. https://3691–3727. https://doi.org/10.1029/2019MS001829

Received 25 JUL 2019
Accepted 14 OCT 2019
Accepted article online 30 OCT 2019
Published online 21 NOV 2019

Abstract We describe the Geophysical Fluid Dynamics Laboratory's CM4.0 physical climate model, with emphasis on those aspects that may be of particular importance to users of this model and its simulations. The model is built with the AM4.0/LM4.0 atmosphere/land model and OM4.0 ocean model. Topics include the rationale for key choices made in the model formulation, the stability as well as drift of the preindustrial control simulation, and comparison of key aspects of the historical simulations with observations from recent decades. Notable achievements include the relatively small biases in seasonal spatial patterns of top-of-atmosphere fluxes, surface temperature, and precipitation; reduced double Intertropical Convergence Zone bias; dramatically improved representation of ocean boundary currents; a high-quality simulation of climatological Arctic sea ice extent and its recent decline; and excellent simulation of the El Niño-Southern Oscillation spectrum and structure. Areas of concern include inadequate deep convection in the Nordic Seas; an inaccurate Antarctic sea ice simulation; precipitation and wind composites still affected by the equatorial cold tongue bias; muted variability in the Atlantic Meridional Overturning Circulation; strong 100 year quasiperiodicity in Southern Ocean ventilation; and a lack of historical warming before 1990 and too rapid warming thereafter due to high climate sensitivity and strong aerosol forcing, in contrast to the observational record. Overall, CM4.0 scores very well in its fidelity against observations compared to the Coupled Model Intercomparison Project Phase 5 generation in terms of both mean state and modes of variability and should prove a valuable new addition for analysis across a broad array of applications.

Plain Language Summary The Geophysical Fluid Dynamics Laboratory (GFDL) of the National Oceanic and Atmospheric Administration participates along with a number of model centers around the world in constructing state-of-the-art climate models for use in studies for climate change and prediction. GFDL's latest multipurpose atmosphere-ocean coupled climate model, CM4.0, is described here. It consists of GFDL's latest atmosphere and land models at about 100 km horizontal resolution and ocean and sea ice models at roughly 25 km horizontal resolution. A handful of standard experiments have been conducted with CM4.0 for participation in the Coupled Model Intercomparison Project Phase 6, an archive of climate model results utilized by the Intergovernmental Panel on Climate Change and the climate research community more generally. The model results have been extensively evaluated against observations. This paper makes the case that CM4.0 ranks high among state-of-the-art coupled climate models by many measures of bias in the simulated climatology and in its ability to capture modes of climate variability such as the El Niño-Southern Oscillation and Madden-Julian Oscillation. The paper also discusses some potential weaknesses, including unrealistically large internal variability in the Southern Ocean and insufficient warming before 1990 in the simulation of the twentieth century.

1. Introduction

This paper provides descriptive documentation of some of the critical elements of the CM4.0 coupled model developed at the National Oceanic and Atmospheric Administration's Geophysical Fluid Dynamics Laboratory (GFDL), results from which are being submitted to the Coupled Model Intercomparison Project Phase 6 (CMIP6) archive. As part of this description, we motivate some of the choices made when this project was initiated and along the development path. We highlight some of the strengths and limitations of the resulting model that might be of particular interest to those analyzing its climate simulations and utilizing it to address a variety of key issues in climate science.

Data from the CM4.0 simulations deposited in the CMIP6 archive have been tagged with the identifier doi.org/10.22033/ESGF/CMIP6.1402 for citation purposes. Model source code is available online (https://doi.org/10.5281/zenodo.3339397). We also provide the links to CM4.0 model inputs and outputs, selected analysis scripts, and supporting data, among others (https://data1.gfdl.noaa.gov/nomads/forms/cm4/).

We begin in section 2 with a brief description of how this effort relates to other coupled climate models constructed at GFDL and the components from which CM4.0 is constructed. Several of the key new model components have been documented in detail elsewhere, specifically the atmospheric component in Zhao et al. (2018a, 2018b) and the ocean/sea ice component in Adcroft et al. (2019). In section 3, some of the central issues faced during the development of this model are introduced. Section 4 then describes the preindustrial control (piControl) simulation, including drift and the Southern Ocean-induced low-frequency variability (section 4.1), as well as the Atlantic Meridional Overturning Circulation (AMOC) and other aspects of the Atlantic simulation (section 4.2). The historical simulation with CMIP6 recommended forcings from 1850–2014 is described in section 5. The time evolution of temperature in the historical simulations is presented in section 5.1. The transient climate response (TCR) and our estimate of effective and equilibrium climate sensitivity are mentioned, but the idealized 1%/year and abrupt quadrupling CO_2 simulations on which these estimates are based are addressed in a companion paper on the climate sensitivity of CM4.0. A necessarily selective presentation of the quality of the simulation of the climate of recent decades in the historical simulation is provided in section 5.2, focusing on temperature, precipitation, and top-of-atmosphere (TOA) fluxes, comparing these to the CMIP5 ensemble of models, to other GFDL coupled models and/or to Atmospheric Model Intercomparison Project (AMIP) simulations. We return to the AMOC simulation in the historical run in section 5.3, the sea ice simulation in section 5.4, and biases in atmosphere and oceans in the tropical basins in section 5.5, providing background for the discussion of the El Niño-Southern Oscillation (ENSO) simulation in section 5.6. Some concluding remarks and lessons learned are included in section 6.

2. Model Configuration and Relationships With Other Past and Ongoing Modeling Efforts at GFDL

Following on earlier pioneering efforts, global coupled model development activity at GFDL was reorganized on a more lab-wide basis in the early 2000s, building on a common software infrastructure (the Flexible Modeling System, Balaji, 2012). This effort resulted in the CM2.1 model (Delworth et al., 2006) from which results were provided for CMIP3. CM2.1 later evolved into higher-resolution versions CM2.5 and CM2.6 (Delworth et al., 2012), as well as models with high resolution only in the atmosphere (Vecchi et al., 2014; Murakami et al., 2015). The higher-resolution atmosphere in these coupled models utilizes the cubed sphere version of the finite volume atmospheric dynamical core (Putman & Lin, 2007).

The atmospheric components of all of these models have prescribed aerosols and ozone concentrations. The ocean component is based on MOM4 (Griffies et al., 2004), using a z-like vertical coordinate and including a mesoscale eddy closure of the Gent-McWilliams form (Gent et al., 1995), except in CM2.5 and CM2.6, in which the mesoscale closure is removed (Griffies et al., 2015). An Earth System Model (ESM2M) with closed atmosphere/land/ocean carbon cycle was built on top of CM2.1, while an alternative ESM (denoted ESM2G) was constructed using the atmospheric component of CM2.1 and an ocean model with an isopycnal vertical coordinate (Dunne et al., 2012). An ESM based on CM2.6, with high resolution in atmosphere and ocean, has also been analyzed (Saba et al., 2016).

The CM3 model that provided results for CMIP5 (Donner et al., 2011) was a departure from the CM2 series in the atmosphere (AM3), with a new convective closure, with the addition of a substantial tropospheric

and stratospheric chemical mechanism allowing for the treatment of ozone as a prognostic variable and the ability to run from emissions of aerosol precursors, and with more resolution in the stratosphere and a higher model top. This model also incorporated the cubed sphere dynamical core. It included the laboratory's first attempt at simulating indirect aerosol effects, which emerged as being very large. Modified versions of CM3, with weaker indirect aerosol cooling (Golaz et al., 2013), were found to provide better simulations of the trajectory of twentieth century warming. The ocean and sea ice components of CM3 are very close to those in CM2.1 (Griffies et al., 2011). LM3, the land component of ESM2M, ESM2G, and CM3 (Milly et al., 2014), was substantially modified from that in previous models, in its hydrology, and especially in its interactive vegetation (Shevliakova et al., 2009).

A higher-resolution atmospheric model (HiRAM) was also developed to explore the potential of a simpler convective closure unifying shallow and deep convection but without the higher top and the chemistry/aerosol package of AM3. AMIP and future snapshot simulations from HiRAM at both 50 and 25 km resolution were provided to the CMIP5 archive.

As described in Zhao et al. (2018a, 2018b), the AM4.0 model was primarily designed to build upon the results from AM3 and HiRAM. New features of this model include substantial improvements to the radiation code; a topographic gravity wave drag with a novel treatment of anisotropy; a "light" chemistry mechanism using AM3's as a starting point, with the capability of simulating aerosols from emissions but with prescribed ozone and other oxidants; aerosol and cloud microphysics modules similar to AM3's but with modifications that affect the indirect aerosol effect; and a new "double-plume" convective closure for shallow and deep convection. The atmospheric model's dynamical core is the cube-sphere finite volume (FV3) hydrostatic version (Putman & Lin, 2007) modified from the version in AM3 to improve treatment of orography and computation efficiency.

The atmospheric component of CM4.0 is identical to the atmospheric model AM4.0 documented by Zhao et al. (2018a, 2018b) with no retuning. AM4.0 has C96 resolution, that is, there are 96×96 grid boxes on each of the six cube faces that cover the sphere or roughly 100 km resolution. It has 33 vertical levels topped at 1 hPa with limited stratospheric resolution that would be commonly categorized as "low top."

CM4.0 uses a new ocean code (MOM6) with a variety of innovations including a generalized vertical coordinate, new treatment of the bottom boundary, and a new surface boundary layer scheme. The ocean code is documented in Adcroft et al. (2019) along with the detailed formulation used in OM4.0, the oceanic component of CM4.0. OM4.0 has a nominal horizontal grid spacing of 0.25° with a tri-polar grid following Murray (1996). A noteworthy feature is that the ocean model is run with no advective or diffusive mesoscale eddy closure, although it does include a submesoscale stirring parameterization based on that in Fox-Kemper et al. (2008) that provides a vertical buoyancy flux in the mixed layer, to which the model is sensitive. The model has 75 degrees of freedom in the vertical utilizing hybrid vertical coordinates similar to those in Chassignet et al. (2003). The upper water column is z coordinate with 2 m grid spacing near the surface, increasing to roughly 20 m at 200 m depth. The coordinate transitions to isopycnal surfaces (referenced to 2,000 dbar) in the upper thermocline and remains isopycnal throughout the abyssal ocean. The simulations referred to here as OM4.0 follow the interannual forcing Coordinated Ocean-Ice Reference Experiment 2 (CORE2) protocol (Danabasoglu et al., 2014) in which an atmospheric reanalysis drives the ice-ocean model.

CM4.0 also incorporates a new sea ice model, SIS2.0, which is a significant upgrade from the SIS model used in previous generations of GFDL climate models (Delworth et al., 2006). SIS2.0 has five thickness categories, and the vertical thermodynamics uses four ice layers and one snow layer. A radiative transfer scheme is incorporated for simulation of the vertical profile of shortwave absorption. The model dynamics use a C-grid stencil as does the ocean. More details about SIS2.0 can be found in Adcroft et al. (2019).

While the land model utilized has much in common with the model LM4.0 used in the AM4.0 documentation papers, there are significant differences, and the land component of CM4.0 is referred to as LM4.0.1. In particular, there are differences in settings related to albedos associated with snow masking of vegetation as well as the albedo of snow on glaciers. The model here has interactive vegetation similar to that used in LM3. The land model has a tiling structure in which different tiles interact with the same atmospheric column. The piControl, 1%, and 4X simulations share a version of the land model in which crop, pasture, natural vegetation, and glacial tiles are included. The historical runs include secondary-growth forest tiles as well. This difference is relevant for the spin-up procedure for the historical simulation in the coupled model, but

the most important effects are on the carbon cycle over land, and we leave detailed discussion of this aspect of the model for the ESM4.0 documentation to follow.

We include a carbon cycle on land and in the ocean that is purely diagnostic (all CM4.0 simulations use prescribed atmospheric CO_2 concentrations). The land component is similar to that in ESM2M (Shevliakova et al., 2009). The ocean biogeochemistry component, referred to as Biology Light Iron Nutrient and Gas model, will be documented elsewhere. We ensure the purely diagnostic character of the model's carbon cycle by prescribing ocean color—eliminating the interaction between ocean biology and the depth of penetration of solar radiation. However, given the lack of developmental focus and resulting problematic aspects of the land carbon simulation, we de-emphasize the carbon simulation here in favor of the more comprehensive and better calibrated version incorporated into the Earth system version of this model, ESM4.0.

3. Summary of Key Focus Areas Along Development Path

This paper does not attempt to provide detailed documentation of the development path for this model, but we do provide a qualitative guide to some of the main issues considered, confining the description of quantitative results for the final configuration of CM4.0 in the sections to follow.

1. As discussed in Adcroft et al. (2019), the choice of 0.25° resolution in the ocean results in a much improved representation of boundary currents throughout the global ocean as well as eddy formation from the tropics through midlatitudes but is presumed inadequate for baroclinic eddy formation in higher latitudes. We experimented with a mesoscale closure that sets in at the point at which the first deformation radius is unresolved, as described by Hallberg (2013), and also an energy backscatter scheme following Jansen et al. (2015), to improve the simulation of marginally resolved eddy fields, both of which show some promise, but issues with the resulting simulations prevented the incorporation of either of these approaches.
2. The implementation of a hybrid vertical coordinate in OM4.0, which reduces spurious cross-isopycnal mixing, impacts bottom water formation simulations in particular but still leaves room for further optimization, as discussed in Adcroft et al. (2019). The development of OM4.0 was intended in part to produce a relatively clean baseline whose performance would be useful as a guide for the generation of future versions with optimized subgrid closures and specification of the hybrid vertical coordinate.
3. Our original intention was to create a model that was more evenly balanced computationally between ocean and atmosphere, using a C192 atmospheric component (roughly 50 km resolution) rather than C96, with C96 providing a development prototype for this higher-resolution version. However, development of the C192 coupled model proved to be too computationally intensive given the resources available, and initial indications were that increasing horizontal resolution only, leaving all physics packages and vertical resolution unchanged, did not result in major reductions in most coupled model biases. We chose instead to use the C192 version of CM4.0, which as expected does improve simulations of tropical cyclones, the probability density function of rainfall intensity more generally, and climatology in regions of complex topography, as a contribution to HighResMIP (Haarsma et al., 2016).
4. While several forays were made in the development of CM4.0 to leverage simulations using the 0.5° ocean being developed for ESM4.0, the climate biases were sufficiently different in the two models, for example, in ocean drifts and in the ENSO simulation, that the process effectively evolved into distinct development branches for the two models.
5. While the quality of the mean tropical climate was relatively steady over the full model development path, the characteristics of the ENSO simulation varied considerably, from being dominated by higher, quasi-biennial frequencies to having lower frequencies and larger amplitudes. Quantification of these ENSO characteristics required at least 100 year simulations. As such long simulations were rarely generated, clean comparisons for attribution of individual parameter sensitivities were not performed. As such, we did not tune explicitly for ENSO quality, given uncertainty as to the key controlling modeling choices and being unclear on how to do so efficiently.
6. As we discuss in section 4.1, a key problematic feature in late-stage development of CM4.0 was the appearance of quasiperiodic superpolynyas in the Southern subpolar ocean with roughly 100 year period. This kind of variability is present in other models at GFDL and elsewhere (e.g., Dufour et al., 2017; Martin et al., 2013; Zhang et al., 2017), but the CM4.0 version is very strong, likely unrealistically so. Following experience in a complementary development effort by GFDL colleagues coupling the AM4.0 atmosphere to a 1° ocean model, we attempted to ameliorate the strength of these events by increasing the near-infrared

albedo of glaciers and icecaps when snow covered. This served to decrease ocean heat uptake and delay the formation of the first superpolynyas during the piControl simulation, by reducing the stratification in the upper subpolar oceans through invigorated convection on the margins of Antarctica, preventing the heat buildup that results in more explosive release in the open ocean. (This change could also be motivated by the desire to reduce the bias in TOA reflected shortwave flux over Antarctica.) But superpolynyas still form eventually.

7. Land model development focused on concerns regarding slow trends toward unrealistically extensive boreal forests in the preindustrial control simulation, which was addressed with modifications to vegetation masking of snow albedos, as well as on the aforementioned glacial albedos.
8. As discussed in Zhao et al. (2018b), the Cess sensitivity of AM4.0/LM4.0, computed by increasing sea surface temperature (SST) uniformly by 2 K and examining the change in the net TOA flux (Cess et al., 1990, 1996), is relatively modest (0.57 K/Wm^2). However, throughout the CM4.0 development, it was clear that this coupled model had higher than expected TCR and high equilibrium sensitivity, given this modest Cess sensitivity.
9. The aerosol forcing in CM4.0 is consistent with that discussed using long (150 year) AMIP runs in Zhao et al. (2018b) and will be analyzed in more detail elsewhere. There was concern during the development process that CM4.0's historical simulation would have undesirable features consistent with too large a TCR and too strong aerosol cooling, and we did perform a few preliminary historical simulations as the model developed with this issue in mind. Zhao et al. (2018b) describe some choices in the convection/cloud scheme in AM4.0 in which alternatives were favored that reduced the Cess sensitivity but did not impact the overall quality of the AMIP model, choices based in part on early qualitative indications of the coupled model's high sensitivity. But, in fact, these turned out to have little effect on the coupled model's sensitivity. The reduction in the strength of the negative aerosol forcing as compared to AM3 is described in Zhao et al. (2018b) along with a discussion of the multifaceted explanation for this difference based on analysis of the long AMIP runs. Relevant factors include changes in wet removal, especially by snow, and from improvements in the numerical algorithm for computing aerosol activation.
10. No alternative ocean or sea ice formulations were incorporated with the specific goal of trying to modify TCR. While uncertainties in sea ice albedos provide some room for a reduction in albedo feedback, the high quality of the simulation of the seasonal cycle of Arctic sea ice provided little motivation for this kind of optimization.
11. AM4.0/CM4.0 are the first GFDL models that incorporate the orographic drag scheme described in Garner (2005, 2018). Use of this scheme is motivated by its natural representation of the effects of anisotropic subgrid scale orography (the resulting parameterized mountain torque is related to the low level winds through a nondiagonal tensor, the form of this tensor being determined by a single precomputation using high-resolution orography). CM4.0 is sensitive to a number of features of this scheme, sensitivity that is much less apparent in uncoupled atmosphere/land simulations, in a manner that we have not fully analyzed, including factors that control the subgrid drag simulated over Greenland.

4. The Preindustrial Control

4.1. Mean Temperature Evolution, Energy Balance, and Centennial-Scale Variability

Preliminary preindustrial control (piControl) simulations were almost exclusively less than 100 years in length. The only run extended for more than 500 years, before the model was finalized, was terminated to adjust glacial albedos in an attempt to ameliorate or at least postpone the superpolynya behavior, as mentioned in section 3.

The piControl simulation designated as the CMIP6 Diagnostic, Evaluation and Characterization of Klima (DECK) (Eyring et al., 2016) contribution for this model consists of Years 151–650 from the run for which the evolution of global mean SST is displayed in Figure 1. The ocean initial condition was taken from temperatures and salinities in the World Ocean Atlas (Locarnini et al., 2013; Zweng et al., 2013). The first 150 years are a period of rapid adjustment of the upper ocean and are discarded for the analysis described here. In the following, Years 151–650 of this run are referred to as piControl Years 1–500. Also displayed in Figure 1 are the observed SSTs over the 1870–2014 period, simply to provide some context for the drift and variability of the control. These observed SSTs are placed in the figure to correspond to the branching time of the DECK historical simulation from the piControl.

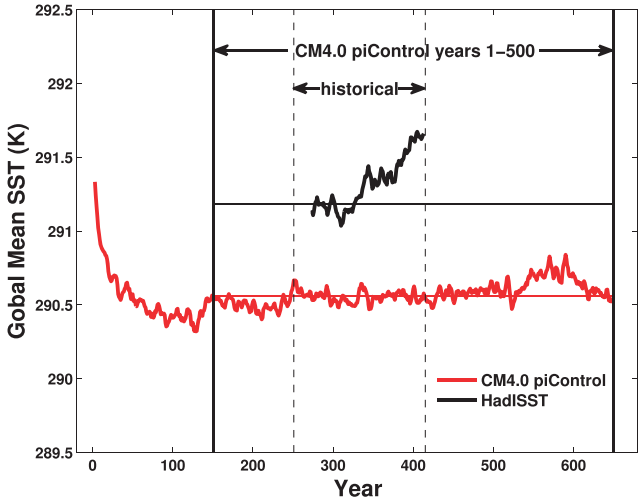

Figure 1. Global mean sea surface temperature (SST) using 5-year running averages from the CM4.0 piControl and from the Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST) (Rayner et al., 2003). The horizontal black and red lines show the SST averages over preindustrial period 1880–1900 for HadISST and over 281–300 for CM4.0 piControl, respectively, indicating about 0.62 K cold bias in the CM4.0 piControl. The solid and dashed vertical lines are the 500 year extent of the CM4.0 piControl placed in the CMIP6 archive as well as the time span of the CM4.0 historical run for the DECK.

The model is too computationally intensive to run multiple millennial simulations or integrate to a true equilibrium, as part of model development (cf. Stouffer, 2004). The characterization of CM4.0's fully equilibrated state is left for future study. While the model has only weak SST trends over the period that serves as a control for the historical coupled run, there is a slow warming trend interrupted most noticeably by a warm event near the end of the piControl. As described below, this feature emanates from the Southern Ocean.

The preindustrial simulation is biased cold in globally averaged SSTs by roughly 0.6 K, comparing the 1880–1900 periods in Figure 1. This global mean temperature bias increased in the development process when snow-covered glacial albedos were increased with the goal of producing a piControl simulation with less dramatic fluctuation in Southern Ocean deep water formation. No retuning of the atmosphere's TOA energy balance was attempted to balance this cooling. As mentioned in Zhao et al. (2018a), the global mean surface air temperature is colder than observed by roughly 0.3 K when run in AMIP mode, so the surface air temperature cold bias is presumably larger than the SST bias. (Observations for this early period are not

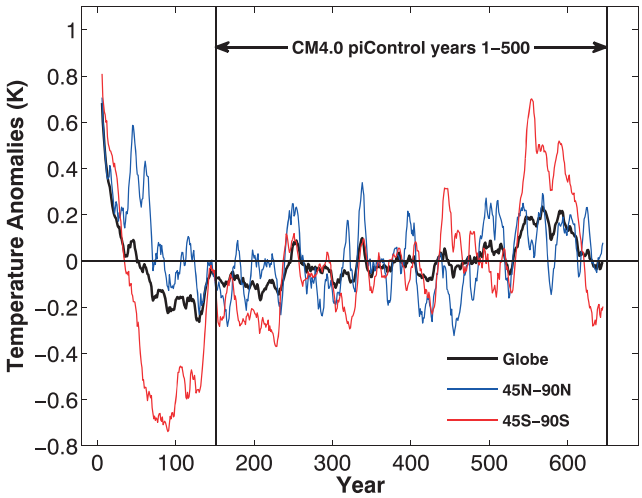

Figure 2. Surface air temperature anomalies using 11 year running averages in the CM4.0 piControl with respect to the average over piControl Years 1–500 for the globe (black), north of 45°N (blue), and south of 45°S (red).

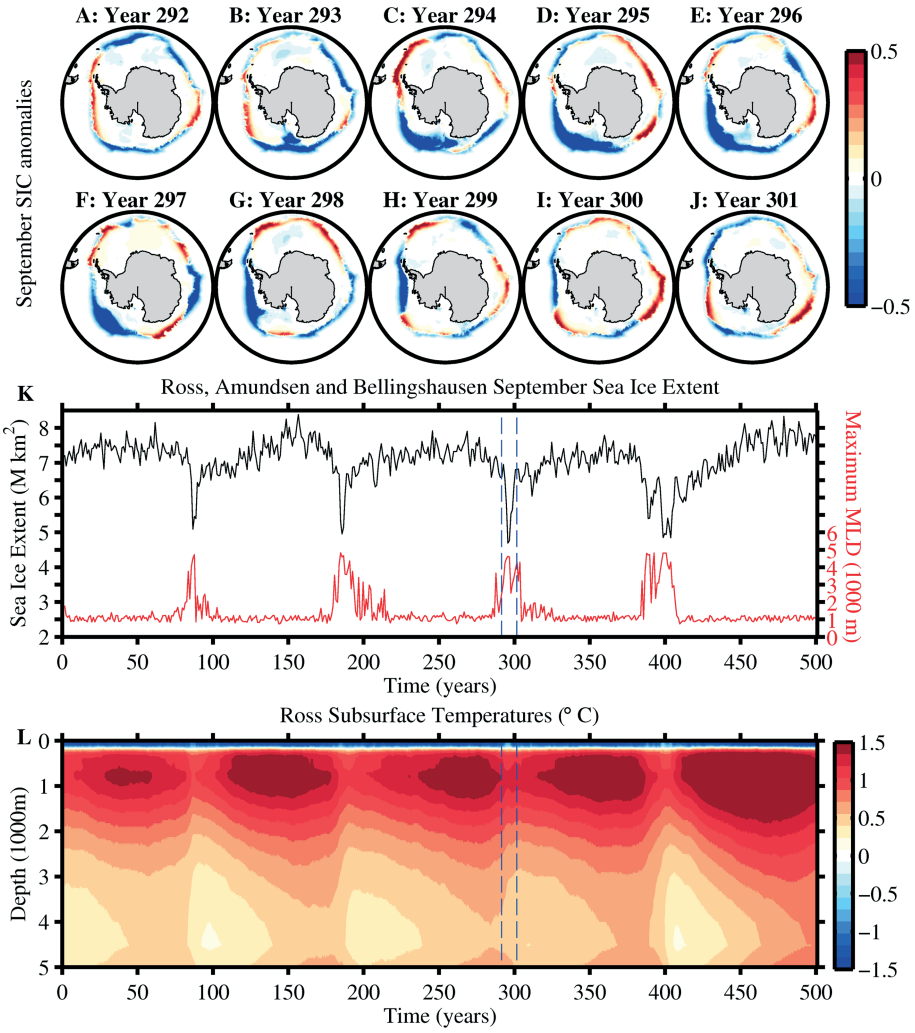

Figure 3. Ross Sea polynya events in the CM4.0 piControl experiment. (a–j) The 10 year sequence of September sea ice concentration (SIC) anomalies (relative to the model climatology) for the polynya event spanning Years 292–301. (k) Time series of September sea ice extent and maximum mixed layer depth computed over the Ross, Amundsen, and Bellingshausen sector (160–300° E). (l) Hovmöller plot of winter (JAS) Ross Sea (160–230° E and 60–90° S) ocean temperature. Dashed blue lines indicate the timing of the polynya event shown in (a)–(j).

adequate to quantify the cold bias in surface air temperature. As decribed below, the bias in recent decades of the historical simulation is sensitive to the time period chosen.)

Figure 2 shows the piControl surface air temperature anomalies (over land and ocean) for the global mean and for means of the two extratropical domains north and south of 45°. The figure illustrates how the global fluctuations are primarily a consequence of much larger fluctuations in the extratropical Southern Hemisphere (SH), and that these fluctuations have relatively little effect on the extratropical Northern Hemisphere (NH).

Information on the superpolynyas that are a central ingredient of this SH variability is shown in Figure 3. The time series of sea ice extent (SIE) and maximum mixed layer depth (MLD) in the middle panel show that the period of these convective events is roughly 100 years in this model and that they penetrate close to the ocean bottom. The geographical pattern of the winter sea ice anomalies over a 10 year period shown in the upper part of Figure 3 illustrates the magnitude of a typical event as seen through the sea ice simulation and indicates the predominate Ross Sea location. The time evolution of the Ross Sea temperature profile (Figure 3L) shows the effect of the polynyas in redistributing heat to the deep ocean and the overall warming trend at all levels over the full 500 years, interrupted by decadal cooling periods during the polynyas. These results reinforce the expectation that the simulation is not equilibrated in the subpolar Southern Ocean.

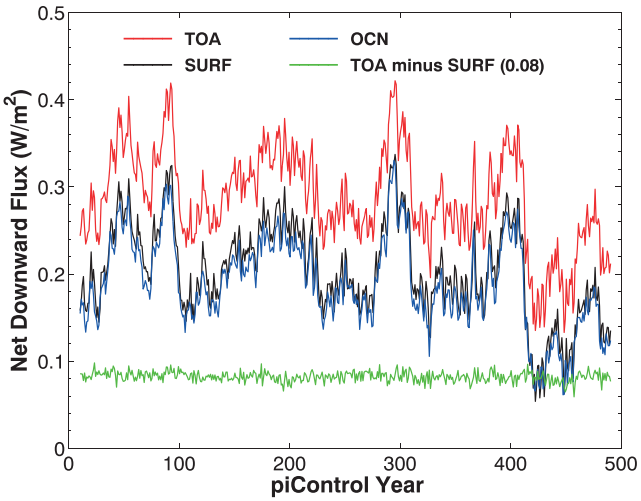

Figure 4. Globally averaged heat fluxes using 20 year running averages at top-of-atmosphere (TOA), surface (SURF), ocean (OCN), and TOA minus surface (the green curve, representing an artificial energy sink in the model atmosphere of about 0.08 W/m^2) in the CM4.0 piControl.

Because none of our simulations are integrated to equilibrium, we have no information on the variability in the final statistically steady state. A change in glacial albedos near the final stages of model development (see section 3) had a significant effect on the characteristics of the model's polynyas. While CM4.0's polynyas are generated in the Ross Sea, before the albedo change the simulations were dominated by Weddell Sea polynyas of comparable or even larger amplitude. All observed open-ocean polynyas have occurred in the Weddell Sea. (For analysis of recent observations of off-shore Antarctic polynyas and their potential role in climate variability, see Campbell et al., 2019.) Once again, because of the limited length of our integrations, we are not confident that the effects of this albedo change on variability are indicative of the behavior in the final equilibrated state.

The net downward radiative imbalance at the top of the model atmosphere over the piControl simulation is provided in Figure 4. Also shown are the net downward energy flux out of the atmosphere at the surface, computed from the atmospheric model, the difference between these two, as well as the net flux into the ocean. The latter (the blue line in Figure 4) is obtained by first creating a time series for total heat content and then differentiating in time.

There is an artificial energy sink in the model atmosphere of 0.08 W/m^2 in the piControl simulation. Analysis of this sink indicates that it is associated with an inconsistency in the energies conserved by the dynamical core and the atmospheric model physics. The atmospheric model dissipates kinetic energy through its numerics, which does not allow a direct computation of this energy sink, so it is balanced by a uniformly distributed energy-conservation fix, but in such a way as to include in this fix the effects of the energetic inconsistency between the model physics and dynamics. The specifics of this fix were designed in AMIP mode and leave this 0.08 W/m^2 energy sink in the piControl coupled run. This energy sink is nearly constant over the piControl and would need to be retuned in the coupled model to remove this remaining inconsistency. The energy imbalance felt by the model is given by the net flux at the surface, either out of the atmosphere at the surface or into the ocean (the latter distinction, as seen in Figure 4, is very small). By this measure, the model is out of balance by 0.22 W/m^2 averaged over the 500 year piControl, with a relatively sharp drop during the large polynya-induced event near piControl Year 400.

4.2. North Atlantic Ocean State and AMOC

The CM4.0 model has an improved simulation of some key characteristics of the North Atlantic Ocean circulation compared with previous GFDL models and many CMIP3 and CMIP5 climate models. For example, in the CM4.0 piControl simulation, the deep western boundary current (DWBC) in the middle low latitude North Atlantic has a realistic strength (Figure 5) and coherency relative to the CM2.5 simulation of Delworth et al. (2012), which uses a similar ocean grid resolution as CM4.0 yet with a z-like vertical coordinate. Climate models with coarser-resolution ocean components often have weak DWBC flow, with much of the deep return flow trapped unrealistically to the mid-ocean ridge (Zhang & Vallis, 2007). CM2.5 also has

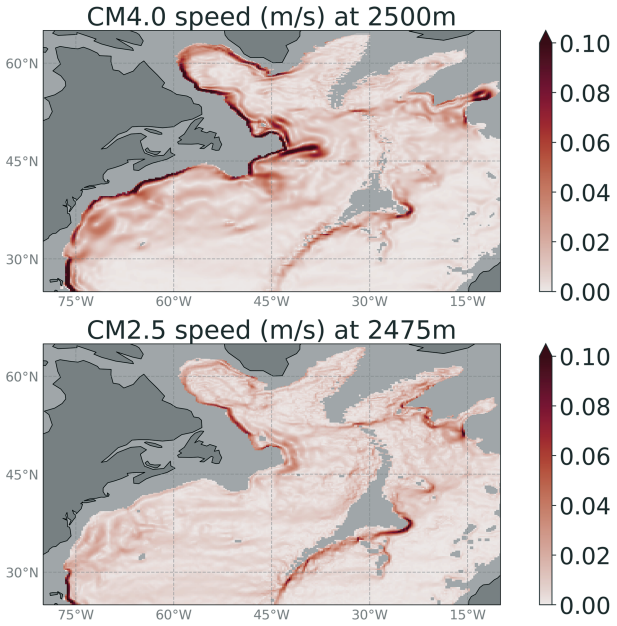

Figure 5. Climatological North Atlantic ocean current speed at 2,500 m for CM4.0 (upper panel) and 2,475 m (closest model level to 2,500 m) in CM2.5 (lower panel), computed using the 40 year time-mean horizontal velocity computed over Years 161–200 from preindustrial control simulations. The light gray is the 2,500 m bottom topography. Note the stronger and more coherent boundary circulation in CM4.0, whereas the CM2.5 current fades away south of the Grand Banks. This difference between the boundary currents is reflected at all depths.

this weakness (Zhang et al., 2011). However, as shown in Adcroft et al. (2019), the hybrid vertical ocean coordinate employed in CM4.0, which reduces to an isopycnal coordinate in the low and midlatitude pycnocline and deeper ocean, reduces spurious numerical mixing. It also has an improved treatment of interaction of flow with topography due to its continuous representation of the bottom boundary. We conjecture that these properties of the vertical coordinate contribute to a deeper and more coherent DWBC in CM4.0 relative to CM2.5 and other comparable models (Wang et al., 2015).

The simulated mean North Atlantic upper ocean circulation in CM4.0 also exhibits a more realistic Gulf Stream separation and path in the open ocean, as revealed by the comparison of the observed mean dynamic topography with the simulated mean sea surface height (Figure 6). This result is consistent with previous studies showing that a relatively strong DWBC and associated bottom vortex stretching keep the Gulf Stream path downstream of Cape Hatteras separated from the North American coast (Yeager & Jochum, 2009; Zhang & Vallis, 2007). After passing the Grand Banks, a significant part of the simulated Gulf Stream

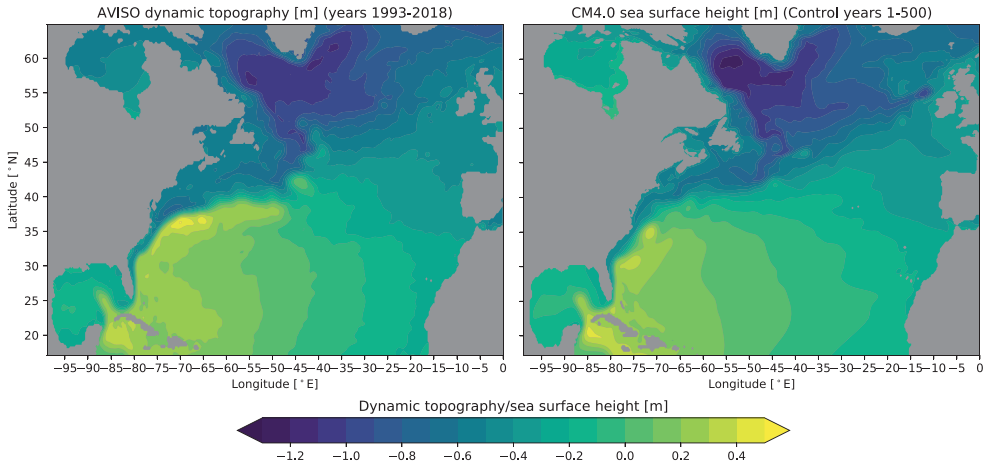

Figure 6. Time-mean North Atlantic dynamic topography from the AVISO satellite analysis and the CM4.0 simulation.

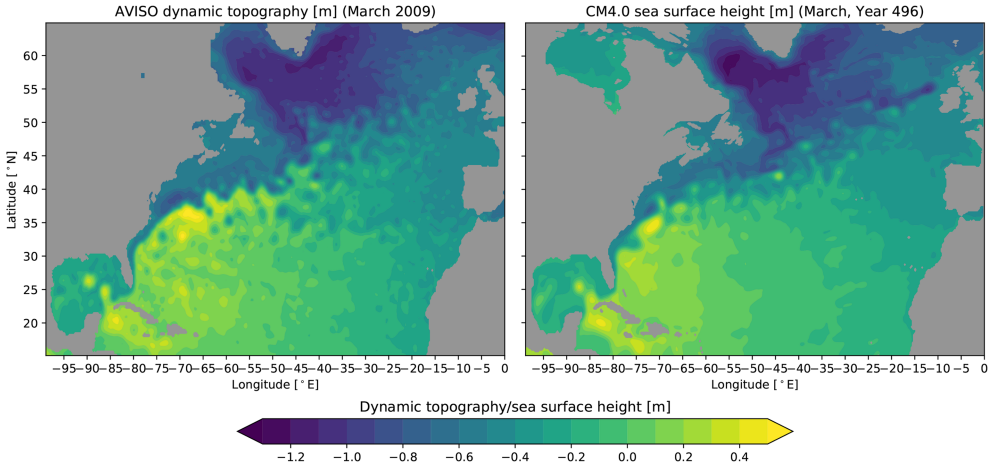

Figure 7. Monthly mean North Atlantic dynamic topography from the AVISO satellite analysis versus the CM4.0 simulation. We chose months with comparable realizations of the Gulf of Mexico loop current.

in CM4.0 turns realistically toward the northeast as the North Atlantic Current. In contrast, in many previous GFDL models, such as ESM2M (Dunne et al., 2012), CM3 (Griffies et al., 2011), and CM2.5 (Delworth et al., 2012), each with a weaker DWBC, the entire North Atlantic Current shifts unrealistically eastward, resulting in a severe cold SST bias east of Newfoundland (Talandier et al., 2014; Zhang et al., 2011). Even with the improvements in CM4.0, there remains some poleward bias in the Gulf Stream path downstream of Cape Hatteras (Figure 6), yet there is only a relatively weak cold bias east of Newfoundland, as can be seen below in Figure 13.

Relatively fine horizontal ocean resolution (~25 km) in CM4.0 as compared to many CMIP models also leads to improvements in regional currents, a good example being the shape and strength of the Gulf of Mexico loop current. As seen in the climatological mean sea surface height in Figure 6 and monthly mean snapshot in Figure 7, CM4.0 simulates the observed Ω-shaped mean pathway of the loop current, reflecting the shedding of warm core eddies into the Gulf of Mexico, a feature that is generally misrepresented in coarse-resolution climate models (Delworth et al., 2012). The monthly mean snapshot in Figure 7 shows an anticyclonic warm core eddy with enhanced sea surface height that is detaching from the loop current, while an older fading eddy can be seen in the western Gulf. The presence of a realistic loop current and detached warm core eddies are important for more realistic simulations of tropical storm intensities and cyclone/ocean interactions in the Gulf of Mexico. The monthly mean snapshot in Figure 7 more generally provides a qualitative indication of the mesoscale energy level in the model as compared to observations.

CM4.0's AMOC, averaged over the piControl period, is displayed in Figure 8. The maximum strength is ~16.7 Sv at 26°N, as calculated from the mean stream function integrated down from the surface. This result

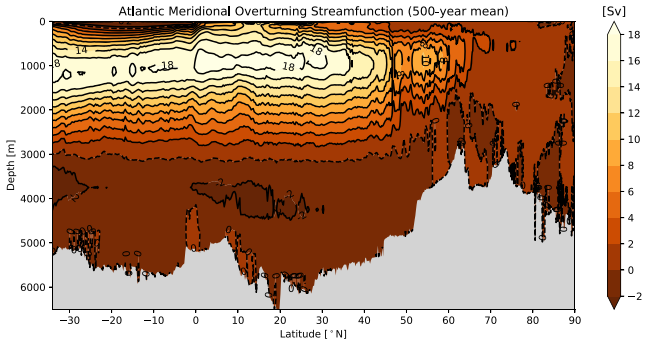

Figure 8. Climatological residual mean Atlantic meridional overturning stream function from the CM4.0 500 year piControl simulation. The stream function is integrated from the bottom, so the simulated AMOC strength relative to the surface is defined as the difference between the maximum value (near 1,000 m) and the surface value (on the order of 1 Sv due to freshwater transport over the North Atlantic).

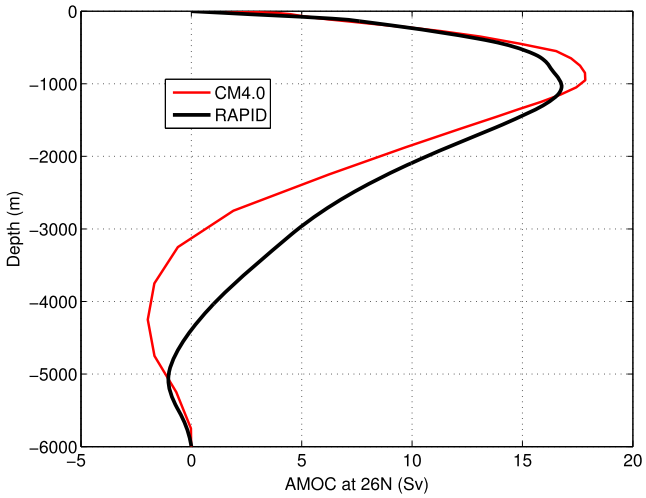

Figure 9. Vertical profiles of the time-averaged circulation at 26° N from the RAPID observation (2004–2015) (Cunningham et al., 2007; McCarthy et al., 2015; Smeed et al., 2018) and from the CM4.0 500 year piControl simulation. Note here only the Eulerian mean component of the simulated AMOC is shown for the comparison with the RAPID data, similar to the approach used in Evans et al. (2017).

is very similar to the ~16.8 Sv measured by the RAPID array over the period 2004–2015 (Cunningham et al., 2007; Smeed et al., 2018) (As shown in Figure 23 as part of the discussion of the historical simulations, the simulated AMOC strength in the period of the RAPID measurements is ~1 Sv lower than this mean piControl value.). The simulated AMOC penetration depth as measured by the zero crossing in the stream function is too shallow (~3,000 m) compared with that directly observed (~4,300 m) at 26°N (McCarthy et al., 2015) (Figure 9). Consistently, the mean piControl northward Atlantic heat transport at 26°N is ~1 PW, which is weaker than the direct estimate of ~1.2 PW for the period of 2004–2013 (Johns et al., 2011). The shallower than observed AMOC and weaker than observed northward Atlantic heat transport found in CM4.0 historical and piControl simulations reflect a common deficiency in both ocean-only and coupled simulations of the deep-penetrating Nordic Sea overflow (Adcroft et al., 2019; Danabasoglu et al., 2010, 2014; Wang et al., 2015; Zhang et al., 2011).

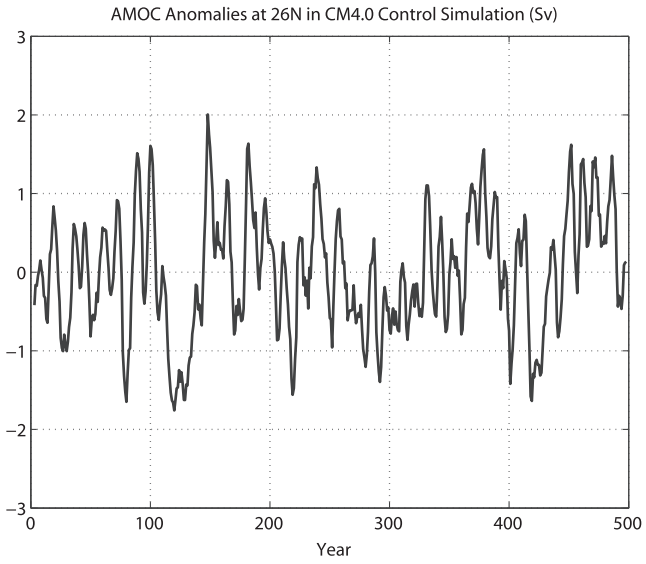

Figure 10. Time series of AMOC anomalies at 26°N from CM4.0 500 year piControl simulations (5 year smoothed), computed from the maximum stream function obtained by integrating down from the surface. The vertical range is set for comparison to the historical AMOC variability show in Figure 23.

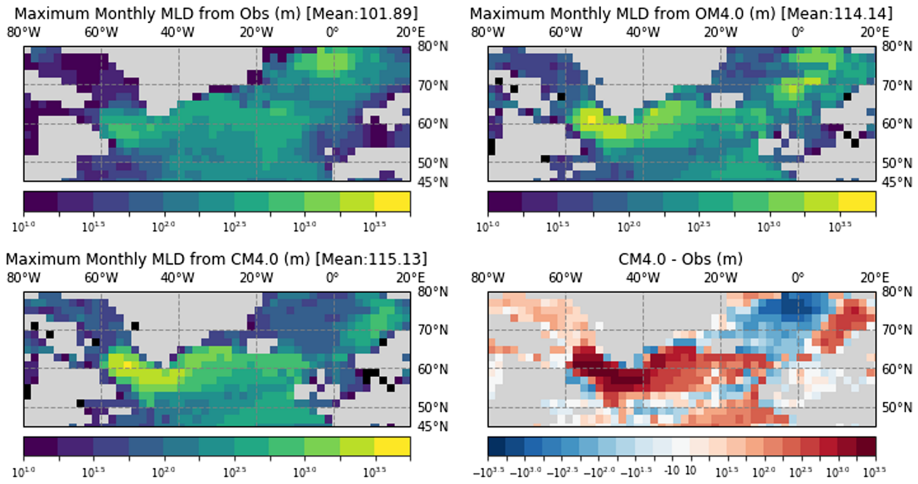

Figure 11. Climatological maximum of monthly mixed layer depth in the North Atlantic and Nordic Sea using a criterion of 0.03kgm^{-3} for the density difference across the mixed layer. (upper left panel) Observational estimates from de Boyer Montégut et al. (2004) updated to include observations through 2015; (upper right) OM4.0 ocean/sea ice simulation averaged over 1987–2007 (Adcroft et al., 2019); (lower left) CM4.0 simulation averaged over the final 100 years of simulation; (lower right) CM4.0 minus observations where red indicates deeper mixed layer depths relative to the observations in CM4.0 and blue indicates shallower mixed layer depths in CM4.0.

The time series in piControl of AMOC strength (anomalies from the 500 year mean) is provided in Figure 10. The statistically stable behavior of the AMOC is consistent with the stability of the Northern extratropical temperatures seen earlier in Figure 2. The multidecadal AMOC variability in the CM4.0 piControl simulation has a modest amplitude of 0.72 Sv (standard deviation of 10 year low-pass filtered AMOC strength at 26°N), which is weaker than that simulated in previous GFDL models with coarser horizontal ocean resolution (1.26 Sv in ESM2G, 1.09 Sv in ESM2M, and 1.07 Sv in CM3) and is in the lower half of the spread among the CMIP5 models (Yan et al., 2018). It is also substantially weaker than the variability 1.4 Sv inferred indirectly from observations by Yan et al. (2018).

In piControl, the maximum MLD in the Labrador Sea is very deep (more than 3,000 m), reaching the bottom in the central Labrador Sea basin, and with deep convection extending over a much broader area than that observed. This excessive MLD is consistent between OM4.0 and CM4.0 (Figure 11) and reflects weak vertical stratification in the deep Labrador Sea. This weak stratification is associated with the absence of a realistic dense Nordic Sea overflow that generally provides stratification to the deep Labrador Sea (e.g., Yeager & Danabasoglu, 2012; Zhang et al., 2011). It is also associated with the absence of realistic high-latitude mesoscale eddies (or a parameterization of their effects; Adcroft et al., 2019) that act to restratify convective regions (e.g., Marshall & Schott, 1999).

In contrast, the maximum MLD in the Greenland Sea in CM4.0 is very shallow (Figure 11). This shallow MLD in CM4.0 (lower left panel) is not seen in OM4.0 (upper right panel), though neither shows the region of maximum MLD consistent with that portrayed in the observations. One possible reason for the lack of the Greenland Sea deep convection in CM4.0 is that the warm salty upper North Atlantic water entering the Nordic Seas moves primarily along the eastern boundary of the Norwegian Sea and is not mixed into the interior Greenland Sea. We conjecture that the muted multidecadal AMOC variability in CM4.0 is very likely linked to the very strong Labrador Sea deep convection (which, being so strong, is resistant to change), the muted Nordic Sea overflow, and the lack of Greenland Sea deep convection.

All of these North Atlantic modeling biases (the excessive maximum MLD in the Labrador Sea, the lack of the Greenland Sea deep convection, weak Nordic Sea overflow and shallow AMOC, and weak multidecadal AMOC variability) are also present in the historical integrations with CM4.0. They are also present in CM2.5 simulations, a model with the same horizontal resolution as CM4.0 (Delworth et al., 2012; Zhang et al., 2011). In a perturbed experiment using CM2.5, a deepened bathymetry south of the Denmark Strait leads to a stronger and deeper-penetrating Nordic Sea overflow and thus deeper AMOC, stronger northward Atlantic heat transport, and reduced Labrador Sea MLD (Zhang et al., 2011). Although the Nordic Sea overflow instantaneously increases after the artificial bathymetry deepening, it gradually weakens afterward due to

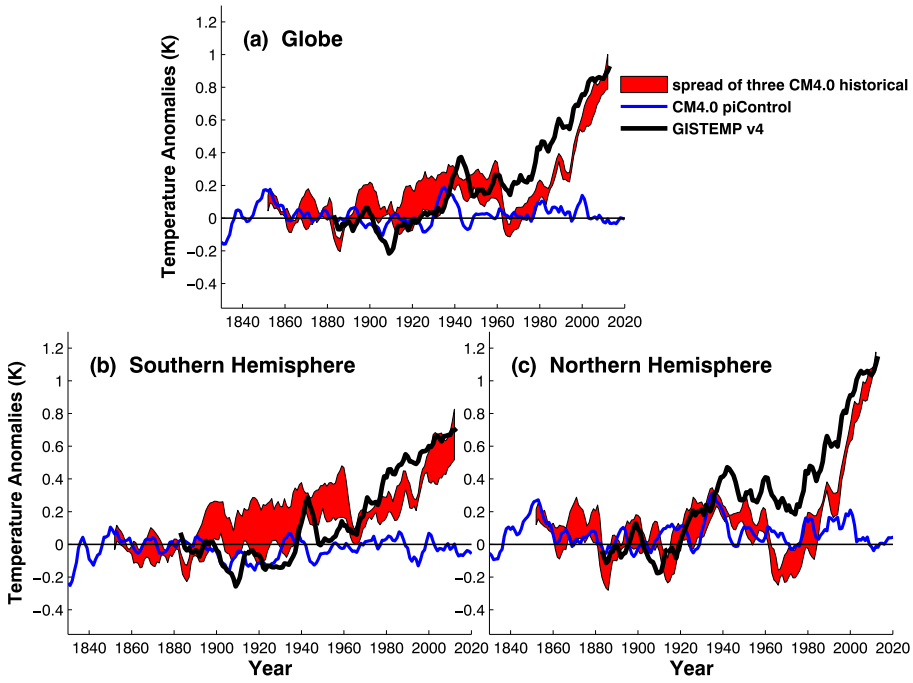

Figure 12. Time series of surface air temperature over land/sea ice and sea surface temperature over open ocean anomalies from 1880–1900 using 5 year running averages over globe, Southern Hemisphere, and Northern Hemisphere. The spread of three CM4.0 historical runs are shown in red, the CM4.0 piControl in blue, and the NASA Goddard Institute for Space Studies Surface Temperature product version 4 (GISTEMP v4) in black (GISTEMP-Team, 2019; Lenssen et al., 2019).

the lack of sustained Greenland Sea deep convection in CM2.5 (Zhang et al., 2011). These long-standing biases, especially in models with a relatively high horizontal ocean resolution, are important areas to focus on for future model development.

5. Historical Simulation (1850–2014)

5.1. Temperature Evolution

The historical simulations that have been branched off the CM4.0 piControl run utilize the identical atmospheric and ocean models, but the land model has additional tiles in each grid box representing secondary growth of forest. The land in the historical run was initialized by first executing a branch from piControl of several centuries duration with potential vegetation (no land use), followed by a bridge run from 1750 to 1849 with full land use transitions to allow the forests to adjust. (The piControl run includes an estimate of midnineteenth century land use but no land use transitions.) The ocean, sea ice, and atmosphere initial conditions were then reset to the conditions at the branch-off time of piControl (e.g., piControl Year 101) to minimize initial differences between piControl and the historical run due to the ocean and sea ice states. In addition to land use, the historical simulations include the time-evolving greenhouse gas concentrations (including ozone for radiation calculations), aerosol precursor emissions, and solar irradiance consistent with CMIP6 specifications (documented at http://goo.gl/r8up31). For chemistry, time-varying ozone and other oxidant concentrations, such as OH, are prescribed based on the historical simulations with CM3 conducted in support of CMIP5 (John et al., 2012).

To assess robustness in the results, we analyze three ensemble members of the historical simulations. They are identical except for the times at which the ocean, sea ice, and atmosphere initial conditions are spun off the piControl simulation (i.e., piControl Years 101, 140, and 182). It is noted that the uneven spacing of start times was the result of technical problems with availability of restarts.

The time evolution of global mean surface air temperature and NH and SH means in these three realizations are displayed in Figure 12. While the overall warming from the late nineteenth century to 2014 agrees quite well with observations, there are strong hints from the NH time series that the balance of factors generating this overall temperature change has unrealistic features. The NH evolution reflects the behavior commonly

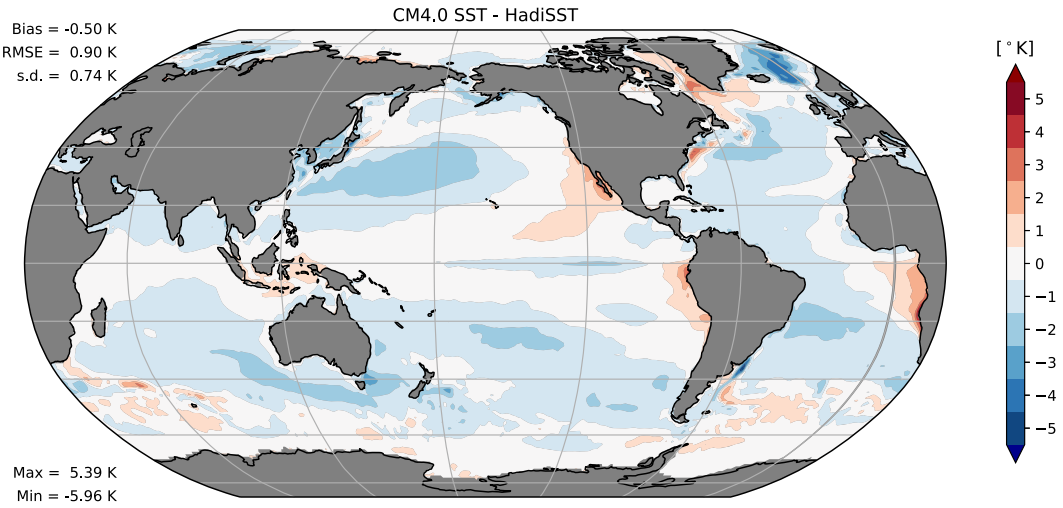

Figure 13. Sea surface temperature bias (K) in CM4.0, averaged over three CM4.0 historical simulations and the Years 2005–2014, compared to the HadISST data set (Rayner et al., 2003) for the same time period.

seen in models with relatively strong aerosol forcing and relatively high transient climate sensitivity: The NH aerosol forcing is able to counteract the greenhouse warming until about 1990, after which the reduction in aerosol forcing and the rapid increase in greenhouse forcing combine to produce rapid warming. While we cannot be definitive based on such a small ensemble, the model's modest NH multidecadal variability implies that the odds of consistency with observations in the NH is low. The early twentieth century warming in the global mean emanates primarily from the SH, unlike the observations. The model's large SH multidecadal variability is once again due to quasiperiodic superpolynyas in the subpolar SH oceans.

The model's TCR is roughly 2.05 K. TCR is the globally averaged temperature increase at the time of CO_2 doubling in idealized experiments where CO_2 is increased at 1%/year. It is estimated here from the warming averaged over Years 60–80. CM4.0's effective sensitivity is 3.2 K, computed from the TCR and the TOA energy imbalance, H, at the time of doubling in the 1%/year simulation ($T_{\text{eff}} = TCR * F/(F - H)$), where F is the estimated radiative forcing for doubling of CO_2). Our estimate of equilibrium sensitivity is about 5.0 K, obtained from an extended abrupt CO_2 quadrupling simulation, extrapolating on a Gregory et al. (2004) plot using Years 51–300. Further research is required, but our working hypothesis is that middle and high-latitude cloud feedbacks interacting with snow and sea ice feedbacks are the primary cause of this high sensitivity. In particular, the large difference between the model's effective and equilibrium sensitivities is due to a high heat-uptake efficacy, as discussed in an upcoming paper on the climate sensitivity of CM4.0. Given that much of the heat uptake is localized in subpolar latitudes, heat uptake efficacy is expected to be sensitive to high-latitude feedbacks.

5.2. Mean Climate Biases

The SST bias pattern averaged over the three historical runs and over the last 10 years of each run (2005–2014) is displayed in Figure 13, showing the bias from observations over this same period. Compared to many previous GFDL models, in which the bias is often of the opposite signs in the two subpolar regions, here the bias is relatively balanced between the two hemispheres (admittedly, this result needs to be tempered by the fact that the Southern Ocean is far from fully equilibrated). The equatorial cold bias in the Pacific is still present but muted compared to many other models. The overall cold bias is concentrated in the centers and the poleward margins of the subtropical highs, with warm anomalies marking coastal subtropical upwelling regions. The root mean square (RMS) bias is 0.90 K, averaged over the three historical simulations. Removing the global mean bias, the standard deviation is 0.74 K. The global mean and therefore the RMS bias is sensitive to the time period chosen for this comparison because the mean cold bias decreases in time over this period, as seen in Figure 12; the standard deviation is somewhat less sensitive

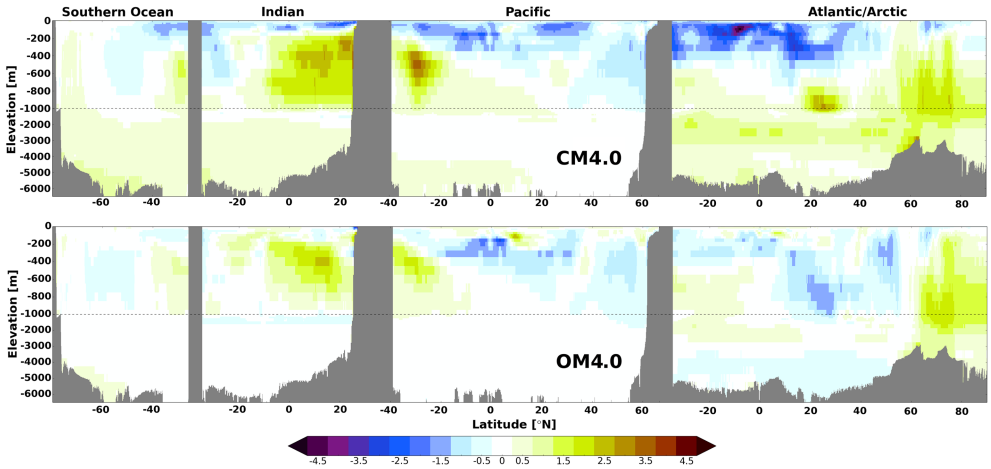

Figure 14. Zonal average ocean temperature biases relative to World Ocean Atlas 2005 for CM4.0 (upper row, 1990–2009 average) and OM4.0 (lower row, 1988–2007 average) historical experiments.

to the averaging period. Given the larger warming trend near the end of the historical runs as compared to observations, this sensitivity to averaging period needs to be kept in mind when using these metrics to evaluate the model.

Figure 14 compares the zonal mean ocean temperature biases by basin at present day in the CM4.0 historical experiment with those of a reanalysis-forced experiment using only its ocean/sea ice components, OM4.0 (Adcroft et al., 2019). Although the ocean/ice states of CM4.0 and OM4.0 are initialized identically, at present day the CM4.0 and OM4.0 runs are roughly 400 and 300 years from initialization, respectively, so the comparison of drift must be qualitative. The difference in ocean surface forcing between the coupled and reanalysis-forced runs appears to account for CM4.0's Antarctic bottom water layer warm bias, which has been eliminated in the OM4.0 run, although surface salinity restoring to climatology, applied in OM4.0, makes these differences difficult to interpret. CM4.0's North Atlantic Deepwater warm bias is also substantially reduced with reanalysis forcing. The surface layer is much warmer and closer to observations in the reanalysis-forced run, a direct impact of the strong coupling to the imposed, observed, surface air temperatures. Below the surface layer and above about 1 km depth, the two runs have strikingly similar biases including (1) a cold bias in the Southern Ocean, (2) a warm bias in the Indian Ocean except in its southernmost part, (3) a pattern of subsurface warm bias in the southernmost Pacific accompanied by a surface cold bias that expands downward moving north, finally reaching 1 km depth at the northern boundary, (4) a cold-over-warm bias in the South Atlantic, and (5) a warm bias in the Nordic Seas and Arctic Ocean. To the extent that the reanalysis-based surface forcing is reliable, this similarity suggests that a significant component of these middepth biases in CM4.0 have their origin in the ocean component rather than in biases in surface forcing.

As described in Zhao et al. (2018a), the strategy in the development of the atmospheric model in AMIP mode was to focus intently on minimizing biases in the TOA energy fluxes, not simply in global means but in the spatial patterns, the result being RMS biases in all seasons that are smaller than in any AMIP simulations in the CMIP5 archive. It is encouraging that CM4.0 maintains the quality of these fluxes. Figure 15 shows the RMS biases in the net TOA fluxes, for the average over time period 1980–2014, comparing against the coupled models in the CMIP5 database. CM4.0 has the smallest RMS bias in all seasons and the global mean. Also shown is the AMIP simulation with AM4.0 atmosphere, quantifying the modest degradation in this field upon coupling. The very small spread among the three-member CM4.0 ensemble supports the adequacy of a small ensemble (or a single run) for this comparison.

Analogous figures for longwave and shortwave TOA fluxes separately are also provided in Figure 15, which indicates that this quality of agreement extends to these individual components of the net flux. The longwave flux has a stronger effect on the quality of the simulation in AMIP mode than does the shortwave, affecting the atmospheric heating directly, but the spatial pattern of the shortwave flux is particularly important to ensure that tuning in AMIP mode is consistent with small drifts when coupled.

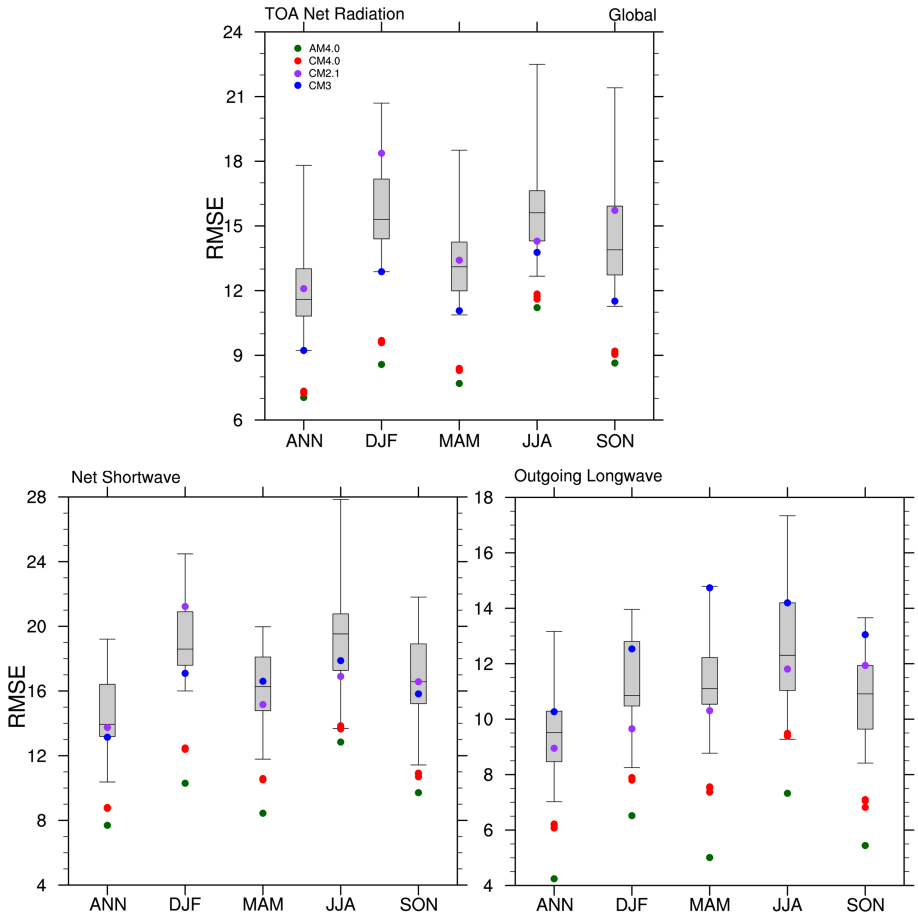

Figure 15. Root mean square errors (RMSE) in net, net shortwave, and outgoing longwave radiation (in W/m^2) at top-of-atmosphere (TOA) for the annual mean and the individual seasons, in three historical CM4.0 simulations (red dots), in GFDL's CM3 model (blue) and CM2.1 model (purple), in the AMIP simulation with the AM4.0 (green), and in the set of the CMIP5 coupled models (box and whiskers showing the full spread, the 25–75% range and the median). Averages are over the Years 1980–2014, and biases are with respect to the CERES-EBAF-Ed2.8 data set (Loeb et al., 2009; Smith et al., 2011).

The pattern of biases in outgoing longwave and net absorbed shortwave in CM4.0 is compared with CM2.1 and CM3 in Figures 16 and 17. The improvement in the shortwave over the Southern Ocean is the most dramatic regional improvement. The equatorial Pacific biases still have a similar structure but are muted compared to previous GFDL models. The shortwave absorption bias is relatively small over the Amazon in CM4.0, while the outgoing longwave radiation is biased high. Since the Amazon has a dry bias (Figure 20), the shortwave properties of the cloud field produced per unit precipitation are actually deficient in this region. Overall, low shortwave and longwave biases away from tropical convection are still present, as in previous models, but with reduced amplitude. The improvement in the TOA flux pattern, which is quite large in this global RMS measure, does not appear to have a single explanation. Rather, we conjecture that it is a result of systematic effort while developing the atmospheric model to maintain a realistic TOA radiation field while other aspects of the model, including the boundary layer and convection modules, were adjusted for other reasons.

We suggest that the high-quality TOA flux simulation is a key ingredient in maintaining a high-quality precipitation simulation in the coupled model. Figure 18 is a box-whisker plot for precipitation analogous to that for the TOA fluxes in Figure 15. While the improvement over the CMIP5 ensemble is less dramatic than for the TOA flux, CM4.0 would still rank as producing the best annual mean, MAM and JJA precipitation field by this measure, and close to the best for the SON and DJF seasons. This RMS measure for precipitation is typically dominated by the tropics, and in part, the improvement in this metric is due to reduced biases toward excessive precipitation south of the equator.

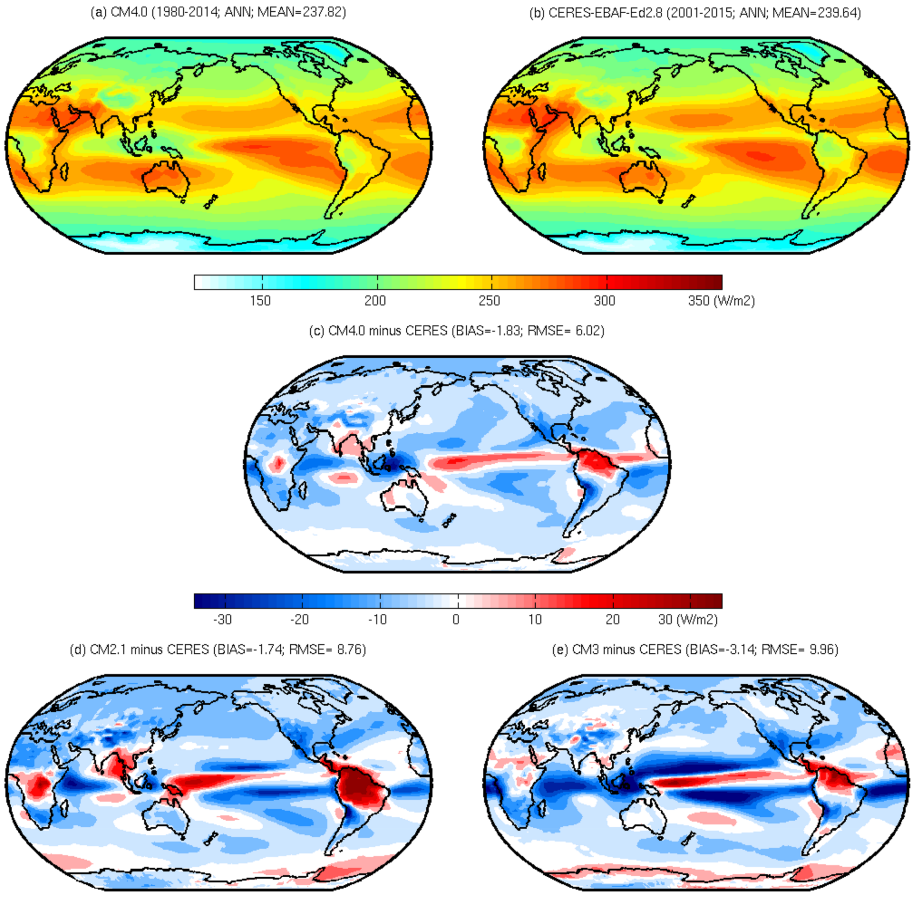

Figure 16. Annual mean outgoing longwave radiation, (a) CM4.0 averaged over three historical runs; (b) CERES EBAF Ed2.8; (c) CM4.0 minus CERES EBAF Ed2.8; (d) as in (c) but for CM2.1; (e) same as (c) but for CM3. Note that the time periods differ for the model and observational data, due to unavailability of CERES EBAF data prior to March 2000, but the specific averaging period is unimportant for these climatological averages.

This point is made more explicitly in Figure 19 which shows the latitudinal structure of annual mean precipitation in the eastern equatorial Pacific. The AM4.0 AMIP simulation is close to the observations, with little hint of a double Intertropical Convergence Zone (ITCZ) bias by this measure. The three CM4.0 runs show modest deterioration, lying in the least biased quantile of the CMIP5 distribution, and superior than previous versions of GFDL models (purple, brown, and blue).

The global pattern of the annual mean precipitation biases is shown in Figure 20, which also provides a comparison of the biases in the AMIP and fully coupled CM4.0 model. A double ITCZ/dry equator bias in the Pacific is still visible in the coupled model. The AMIP simulation has unrealistically large precipitation in the western equatorial Pacific, especially in the vicinity of the Philippines (arising from a summertime bias), but this bias is partly ameliorated in the coupled model, behavior seen in other models, including superparamterized models (Stan et al., 2010). In contrast, a significant deficiency is the dry bias that forms over the Amazon in CM4.0, due to a dry season of too long duration, that is not present in the AMIP run. On the other hand, tropical African rainfall is hardly affected by coupling. A dry bias in central North America in the AMIP is also weakened somewhat in the coupled model.

Understanding these differences between AMIP and coupled runs, due in part to the SST biases that develop but also potentially in part due to the effects of coupling on convection and transients disturbances, presents important challenges, each in its own right. For example, exploratory simulations aimed at understanding the dry Amazon bias in CM4.0 indicate that the coupled model SST biases in the Atlantic are a main source of this dry bias. In a preliminary version of the coupled model that was part of the development process, annual mean Amazon rainfall is roughly 40% lower than in the corresponding AMIP simulation, while in a

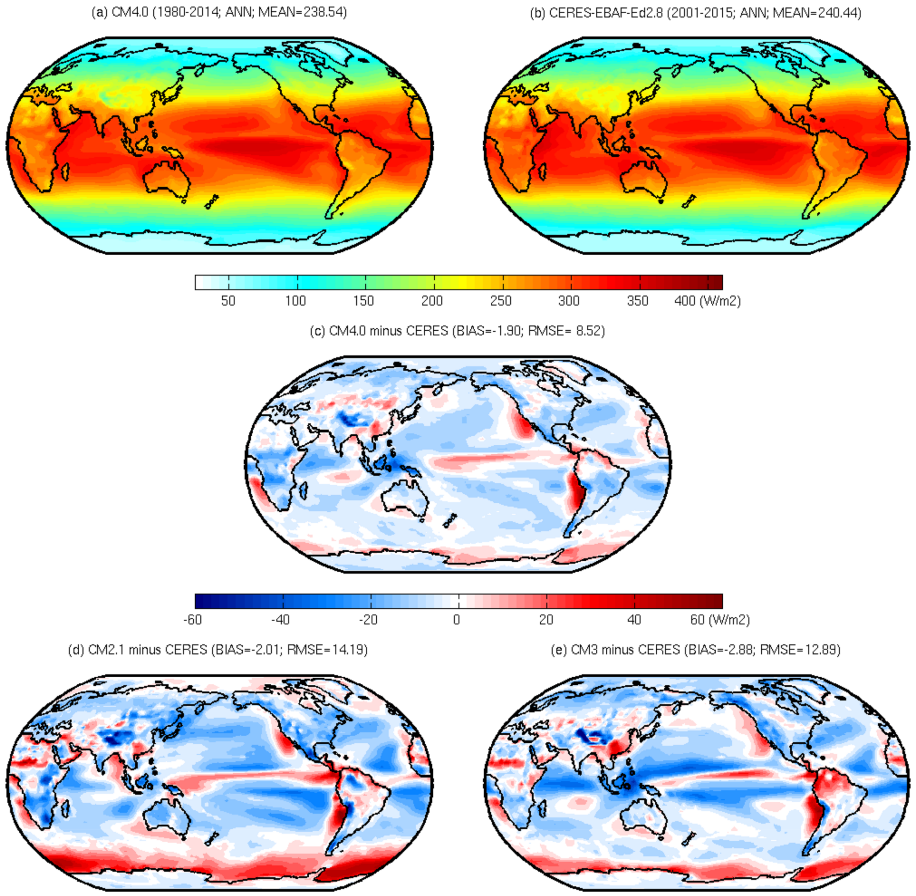

Figure 17. As in 16 but for the TOA net shortwave flux.

coupled model in which the SSTs in the tropical Atlantic are constrained by observations, the underestimate is reduced to roughly 15%.

The annual and zonal mean temperature biases in the coupled and AMIP simulations in the latitude-pressure plane are shown in Figure 21. In the AMIP simulation, there is a weak cold bias, less than or of the order of 1 K, but with hardly any of the cold bias near and immediately above the extratropical tropopause seen in many other models. However, there is a larger cold bias below the tropical tropopause. The sensitivity of this feature to modeling choices in the convection scheme is described by Zhao et al. (2018b). When coupled, the cold bias increases by roughly a factor of 2, in large part because the coupled model's tropical SSTs are biased cold but also due to an equatorward contraction of the midlatitude circulation. The latter is likely in part due to the tropical cold bias itself and in part to the pattern of extratropical SST biases.

Consistent with the development of these biases in temperature gradient within the troposphere, significant zonal wind biases develop in CM4.0 that are not present in the corresponding AMIP simulation (Figure 22). Most noticeable is an equatorward shift in the jets, in thermal wind balance with the temperature biases, but which are also accompanied by a surface wind signature, most clearly in the SH, with magnitude roughly 1 m/s near the surface. The bias toward too strong trade winds in AM4.0 is present in CM4.0 as well.

During the development process, the sensitivity of wind and temperature biases in northern subpolar to midlatitudes in the coupled model, including the lower stratospheric biases centered at 40°N, were found to be sensitive to the gravity wave drag formulation. Because these biases were not as well-defined in AMIP simulations, they evidently involve feedbacks with sea ice and oceans which were difficult to address during development and will need to be analyzed further in future studies.

The stratospheric simulation in CM4.0 is otherwise very similar to that in the AMIP simulation. This similarity encompasses the frequency of sudden warmings, the strength of the surface pressure response to sudden

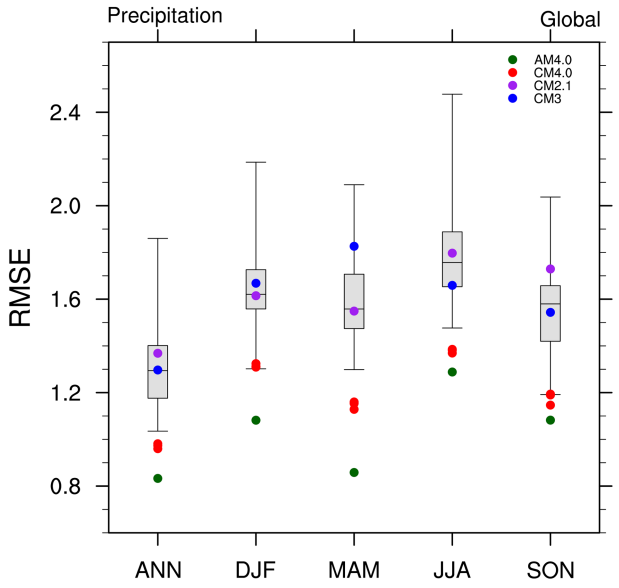

Figure 18. Root mean square errors (RMSE) in precipitation (in mm/day) in three historical CM4.0 simulations (red dots), in GFDL's CM3 model (blue) and CM2.1 model (purple), in the AMIP simulation with the AM4.0 (green), and in the set of the CMIP5 coupled models (box and whiskers showing the full spread, the 25–75% range and the median). Averages are over the Years 1980–2014, and biases are with respect to the GPCP v2.3 data set (Adler et al., 2003, 2016).

warmings, and the stratospheric and surface response to the ozone hole (recall that ozone is specified in CM4.0), all of which are discussed in Zhao et al. (2018b) in the AMIP context.

5.3. North Atlantic Salinity and AMOC

Figure 23 shows the time series of AMOC anomalies from the CM4.0 historical simulations. Over the full period of these simulations, the model produces a trend toward strengthening AMOC. In all of the ensemble members, the trend consists of a strong increase from 1940 to 1980 and a compensating reduction from 1980 to present with a peak around 1980. In comparison, the AMOC in the piControl simulation is statistically stationary with no significant trend over the 500 year period (Figure 10). We thus conclude that the model's AMOC evolution in the historical period is predominately forced. Based on previous studies (e.g.,

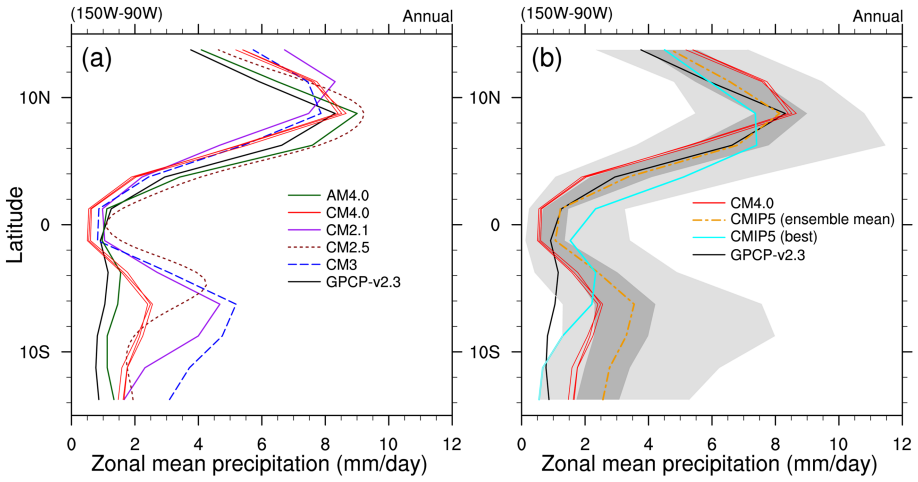

Figure 19. Annual zonal mean precipitation in the tropical Eastern Pacific averaged over longitudes 150–90°W as a function of latitude. Observations (GPCP v2.3), AM4.0, three CM4.0 realizations averaged over the Years 1980–2014, and three earlier GFDL models, CM2.1, CM2.5, and CM3, are shown in (a); and the CMIP5 ensemble mean, the best CMIP5 model (here referring to the CMIP5 model with smallest root mean square error over the region [14°S:1°4N, 150–90°W], which is MICRO-ESM-CHEM model shown in cyan; Watanabe et al., 2011), the full spread and the 25th to 75th percentile in CMIP5 coupled models in light and dark gray in (b).

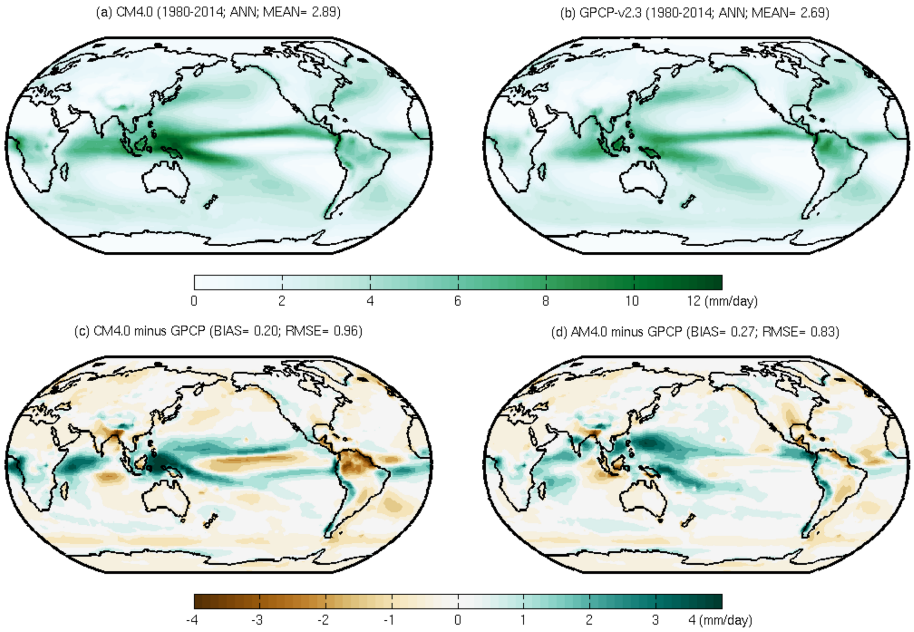

Figure 20. Annual mean precipitation (mm/day), (a) CM4.0 averaged over three historical runs; (b) observations (GPCP v2.3); (c) difference between CM4.0 and GPCP v2.3; (d) difference between AM4.0 for 1980-2014 and GPCP v2.3.

Delworth & Dixon, 2006), the increase in strength prior to 1980 is likely due to aerosol forcing, while the following decrease is forced by the combination of a reduction in aerosol forcing and an increase in greenhouse forcing. It is of interest that the strength of the aerosol forcing relative to greenhouse gas forcing in changing the AMOC strength is greater than for changing NH mean surface temperature. That is, aerosol cooling prevents but does not reverse the greenhouse gas-induced warming (Figure 12), whereas it does reverse the greenhouse gas-induced AMOC weakening.

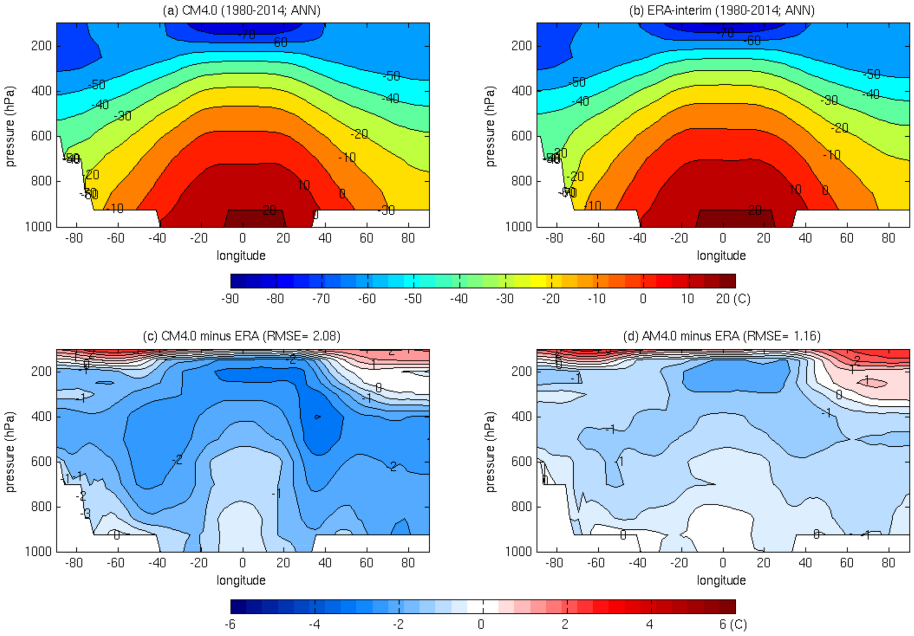

Figure 21. Latitude-pressure plots of annual zonal mean temperature (°C) for (a) three-member ensemble mean of CM4.0, (b) ERA reanalysis (Dee et al., 2011), (c) difference between CM4.0 and ERA, and (d) difference between AM4.0 for 1980–2014 and ERA.

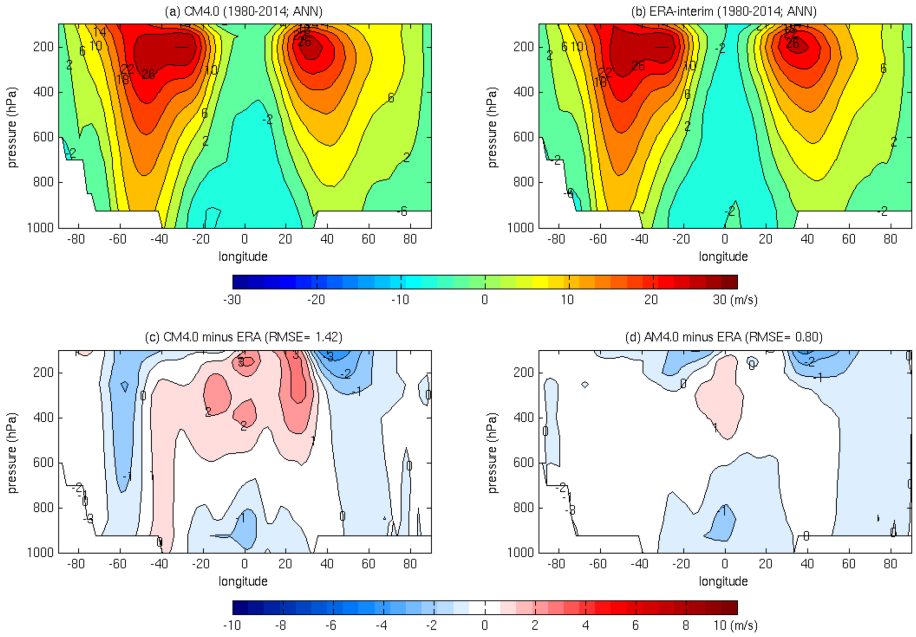

Figure 22. As in 21 but for the annual zonal mean wind.

This picture of forced AMOC changes in the historical simulations is consistent with the simulated sea surface salinity (SSS) changes averaged over the subpolar North Atlantic region (Figure 24). Here, we also see an increasing salinity trend from 1940 to 1980 and a decreasing trend from 1980 to present, with a peak around 1980. A close linkage between forced AMOC changes and subpolar North Atlantic SSS has been found previously in CMIP models and has been attributed to anthropogenic aerosols. That is, increasing anthropogenic aerosol precursor emissions lead to the positive trends in subpolar North Atlantic SSS and vice versa. This salinity response is most evident in models with strong indirect aerosol effects (Menary et al., 2013; Zhang et al., 2013). The anthropogenic aerosols affect SSS mainly through changes in surface evaporation, that is, an enhanced anthropogenic aerosols lead to an increase in surface evaporation and thus an increase in SSS over the subpolar North Atlantic (Menary et al., 2013). There is a coupled positive feedback between SSS

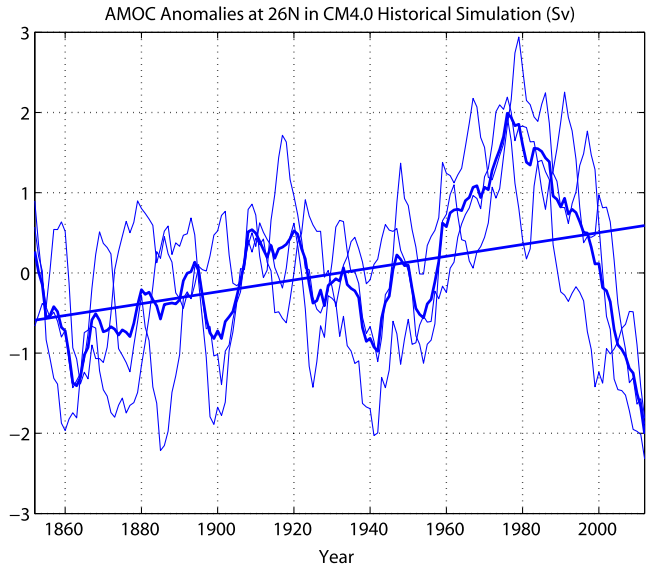

Figure 23. Time series of AMOC anomalies at 26°N from CM4.0 historical simulations (5 year smoothed). (solid blue lines) Ensemble mean anomalies and long-term trend. (thin blue lines) Anomalies from three individual ensemble members. (Recall that the three ensemble members use initial conditions from the piControl spaced by about 40 years).

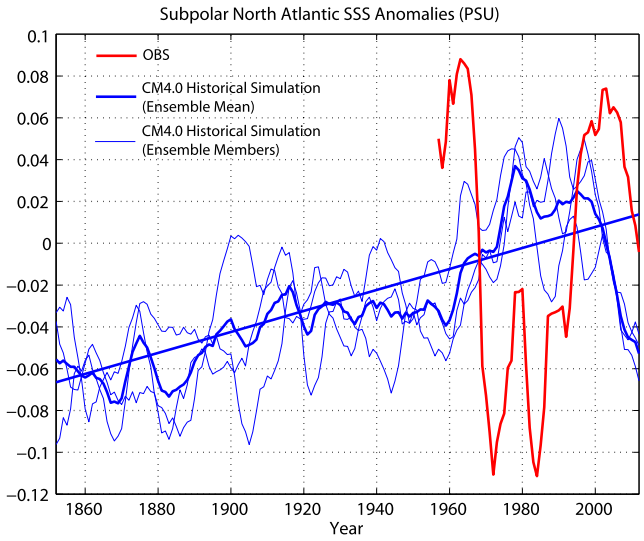

Figure 24. Time series of the sea surface salinity (SSS) anomalies averaged over the subpolar North Atlantic domain (60°W–0°E, 50-65°N) from CM4.0 historical simulations and observations (5 year smoothed). (solid blue lines) Ensemble mean anomalies and long-term trend. (thin blue lines) Anomalies from three individual ensemble members. (red solid line) Observed anomalies. The observed SSS data are 5 year averages of the objectively analyzed ocean salinity anomalies since 1957 in Boyer et al. (2005). Both observed and simulated SSS anomalies are relative to the 1957–2012 mean.

and AMOC changes. The increasing/decreasing AMOC trends in the twentieth century modulate SSS over the subpolar North Atlantic and vice versa; the increasing/decreasing trends in SSS over the deep water formation sites in the subpolar North Atlantic reinforce AMOC changes (Menary et al., 2013). The observed multidecadal subpolar North Atlantic SSS variations are very different from those found in these CM4.0 historical simulations over the second half of the twentieth century, with observations showing a negative phase during 1970s and 1980s and positive phase during 1960s and post-1990 (Figure 24). Based on earlier studies, we suggest that the observed multidecadal changes are dominated by internal variability (Zhang, 2017; Zhang et al., 2013, 2019). This discrepancy between CM4.0 from the observational records reinforces the hypothesis that CM4.0 has insufficient multidecadal variability. The lack of spread in the model realizations in CM4.0 suggests that it is incapable of generating a subpolar SSS evolution consistent with observations.

While the CM4.0 variability in subpolar Atlantic salinity is problematic, the model's simulation of the mean North Atlantic salinity is greatly improved over other GFDL models. Figure 25 compares the CM4.0 historical simulation with that of CM3 over recent decades. The effects of the improved open ocean circulation (e.g., more realistic North Atlantic Current pathway) described earlier are evident in the extratropical salinities, and some pathological behaviors in the North Sea are clearly improved. But especially encouraging is the simulation of the low salinity ribbon tightly constrained to the western boundary, realistically present down to Cape Hatteras and beyond, which is related to the more realistic Gulf Stream separation. The Labrador Sea is too saline consistent with the picture that there is insufficient mixing of these coastal low salinities into the center of the Labrador Sea, consistent with the steadiness of deep convection in that basin.

5.4. Sea Ice

The observed magnitude and timing of the seasonal cycle of Pan-Arctic SIE are quite well represented by CM4.0, whereas Pan-Antarctic SIE displays an amplified seasonal cycle with too little summer ice and too much winter ice (Figure 26). The RMS difference between the modeled and observed monthly seasonal cycle of Pan-Antarctic SIE is 2.32×10^6 km^2, which is substantial but still lower than the CMIP5 median RMS bias (3.42) (Shu et al., 2015) and also represents a large improvement relative to earlier-generation GFDL models CM2.1 and CM3 (Griffies et al., 2011; RMS values of 4.01 and 5.82, respectively). The Pan-Arctic SIE climatology has a lower RMS error of 0.59×10^6 km^2. This value is also lower than the CMIP5 median RMS error (1.45) (Shu et al., 2015), is improved relative to CM2.1 (1.68), and slightly degraded relative to CM3 (0.41).

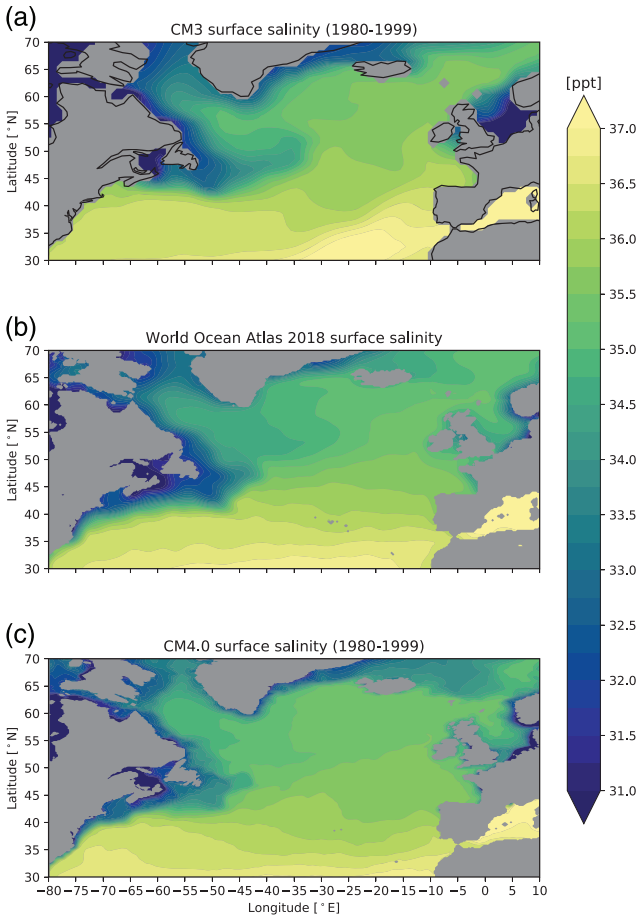

Figure 25. North Atlantic sea surface salinity from (a) GFDL's previous generation CM3 model historical simulation averaged over Years 1980–1999 (ftp://nomads.gfdl.noaa.gov/CMIP5/output1/NOAA-GFDL/GFDL-CM3/historical/mon/ocean/Omon/r1i1p1/v20110601/so/; Donner et al., 2011; Griffies et al., 2011), (b) World Ocean Atlas 0.25° climatology (https://www.nodc.noaa.gov/OC5/woa18/woa18data.html; Zweng et al., 2018), and (C) GFDL-CM4.0 historical simulation averaged over Years 1980–1999.

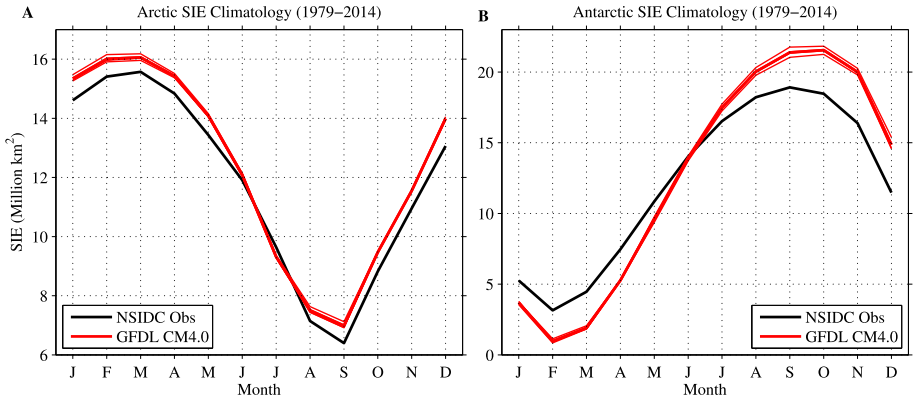

Figure 26. Pan-Arctic (a) and Pan-Antarctic (b) sea ice extent (SIE) climatologies (million km^2) computed over 1979–2014 from satellite observations (black), the three CM4.0 historical simulations (red), and the CM4.0 historical ensemble mean (thick red). The observed SIE climatology is computed using monthly-averaged passive microwave satellite sea ice concentration observations from the National Snow and Ice Data Center (NSIDC) processed using the NASA Team Algorithm (Data Set ID: NSIDC-0051; Cavalieri et al., 1996). Pan-Arctic and Pan-Antarctic SIE are defined as the areal sum of all grid points whose sea ice concentration (SIC) exceeds 15% in the Northern and Southern Hemispheres, respectively.

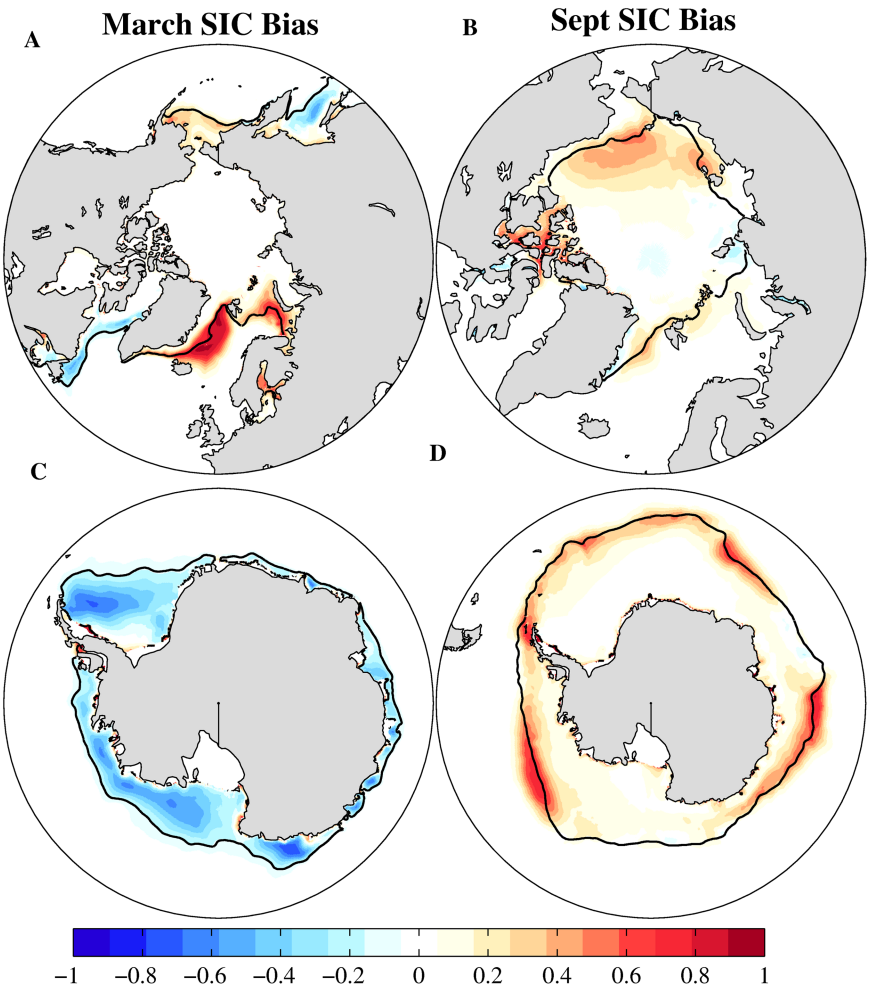

Figure 27. Climatological Arctic (a, b) and Antarctic (c, d) sea ice concentration (SIC) biases (model minus observations), computed over Years 1979–2014. The black contour indicates the observed climatological sea ice edge (15% SIC contour), and the observed data are monthly averaged passive microwave satellite sea ice concentration observations from NSIDC processed using the NASA Team Algorithm (Data Set ID: NSIDC-0051; Cavalieri et al., 1996).

The CM4.0 Antarctic summer SIE bias primarily results from negative summer sea ice concentration (SIC) biases in the Weddell, Ross, Amundsen, and Bellingshausen seas, and the positive winter SIE bias primarily results from positive winter SIC biases in the Amundsen and Bellingshausen, West Pacific, and Weddell Sea sectors (see Figures 27c and 27d). Summer Arctic SIC displays positive biases in the Canadian Arctic Archipelago, Beaufort, Chukchi, and East Siberian seas, whereas the winter ice cover is characterized by positive biases in the Barents, Greenland-Iceland-Norwegian, and Bering seas and negative biases in the Labrador Sea and Sea of Okhotsk (see Figures 27a and 27b).

CM4.0 simulates declines of Pan-Arctic SIE in all months of the year, with trend magnitudes that are in reasonable agreement with satellite-observed trends (see Figure 28). Computed over Years 1979–2014, the ensemble mean CM4.0 Pan-Arctic SIE trends are -0.66 and -0.28×10^6 km^2 per decade in September and March, respectively, which are slightly weaker than the observed trends of -0.87 in September and -0.39 in March. The CM4.0 Pan-Arctic SIE trends are comparable to the CMIP5 multimodel ensemble mean trends of -0.57×10^6 km^2 per decade in September and -0.30×10^6 km^2 per decade in March (Massonnet et al., 2012; Shu et al., 2015; Stroeve et al., 2012). The three ensemble members make it clear that internal variability in CM4.0 can make only a minor contribution to the trend, consistent with its weak AMOC variability. If, as suggested in some studies, internal variability has contributed a nonnegligible fraction to the observed trend, it is possible that the model's externally forced decline is too large despite agreeing with observations

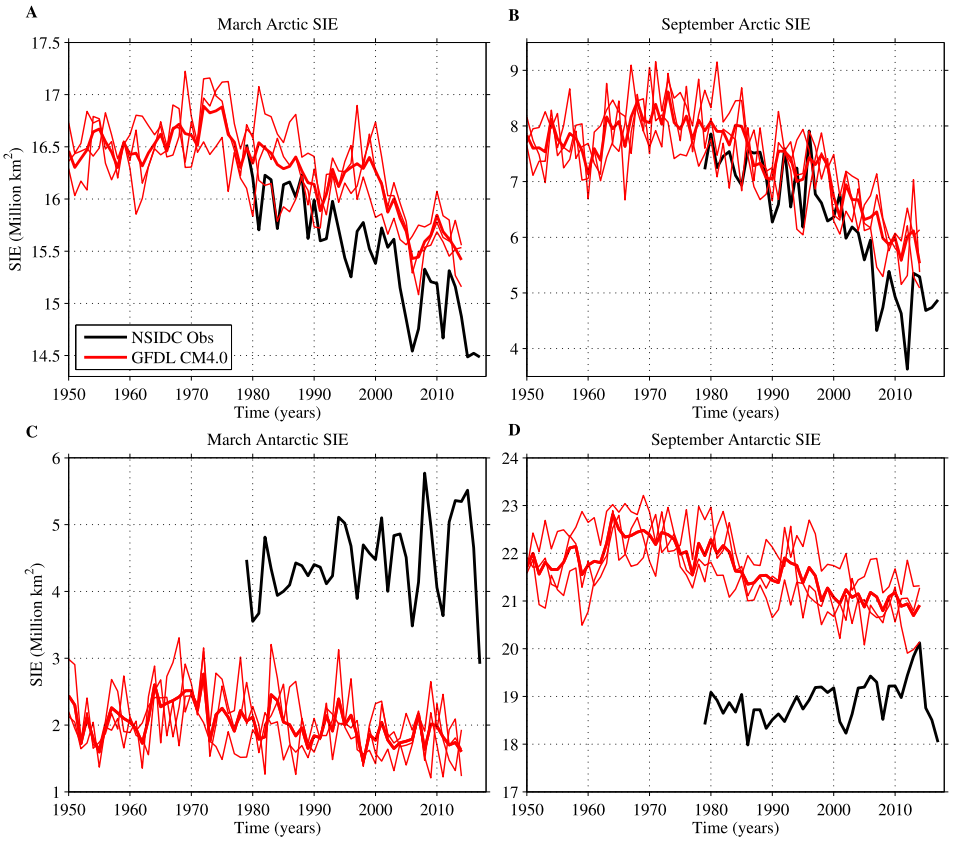

Figure 28. Time series of March and September Pan-Arctic (a,b) and Pan-Antarctic (c,d) sea ice extent (SIE; million km^2) from NSIDC satellite observations (black; Data Set ID: NSIDC-0051; Cavalieri et al., 1996), the three CM4.0 historical simulations (red), and the CM4.0 ensemble mean (thick red).

in isolation. In this regard, a more satisfying metric would be consistent with both the SSS subpolar Atlantic evolution, as a proxy for AMOC strength, as well as the Arctic sea ice trends.

Similar to most CMIP5 models, CM4.0 simulates negative trends for Pan-Antarctic SIE and fails to capture the slightly positive observed SIE trends of $+0.21 \times 10^6$ km^2 per decade in March and $+0.24 \times 10^6$ km^2 per decade in September. The simulated CM4.0 trends are -0.34 and -0.09×10^6 km^2 per decade in September and March, respectively, which can be compared to the CMIP5 multimodel ensemble mean trends of -0.40 in September and -0.25 in March (Turner et al., 2013). The CM4.0 historical simulations do not include the effects of freshwater input from melting Antarctic land ice, which may be an important factor in capturing observed Antarctic sea ice trends (Bronselaer et al., 2018).

CM4.0 has a clear bias in its spatial pattern of Arctic sea ice thickness (SIT) (see Figures 29a and 29b). CM4.0 has its thickest ice in the Beaufort, Chukchi, and East Siberian seas, lacks the observed East-to-West SIT gradient, and generally resembles a 90° westward rotation of the observed SIT pattern. The CM4.0 ice velocities are biased highly relative to satellite-observed velocities. The model generally displays the key climatological circulation features of Arctic sea ice, such as an anticyclonic circulation in the Beaufort Sea and a transpolar drift with export through the Fram Strait; however, these features are less clearly defined than in observations. The simulated ice velocities generally promote ice advection into the Beaufort and Chukchi Seas and have less advection toward the north coast of Greenland than observations, consistent with the model's SIT bias pattern. These differences are robust across the three historical ensemble members. The drift and SIT model biases are suggestive of a shifted Beaufort high position. Indeed, we find that CM4.0 has a wintertime Beaufort high that is shifted toward the Siberian coastline, favoring ice advection into the Beaufort Sea (see Figures 29d and 29e). As mentioned above, this Arctic wind field is sensitive to the precise formulation of orographic gravity wave drag in CM4.0.

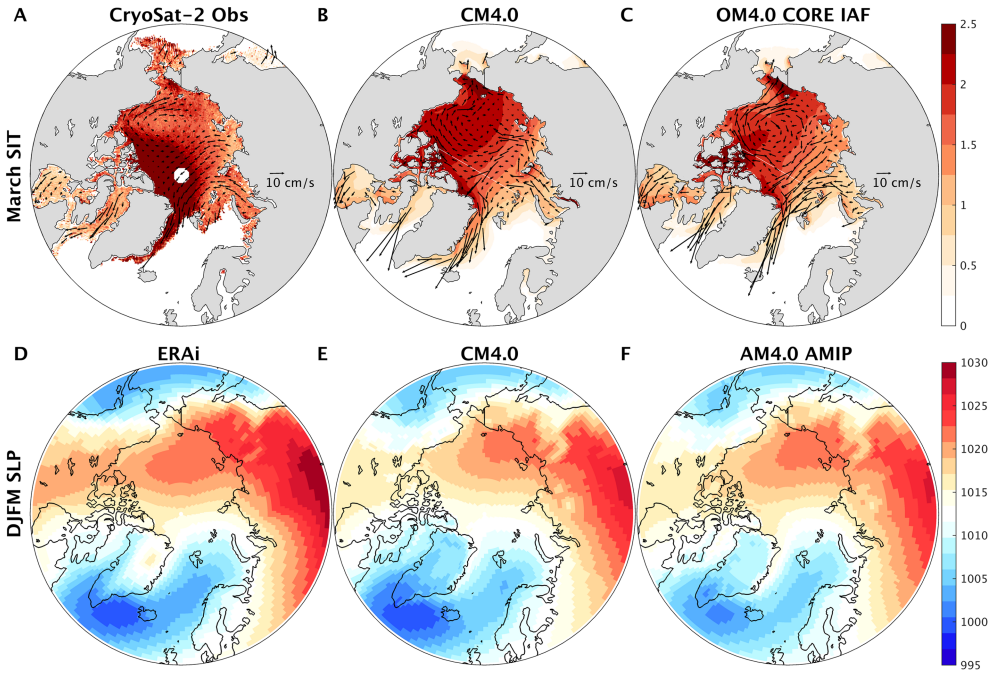

Figure 29. March Arctic sea ice thickness climatology (m) from CryoSat-2 Observations (a), CM4.0 historical simulations (b), and an OM4.0 CORE-forced simulation (c). Observed and simulated climatological sea ice velocities (DJFM; cm/s) are plotted as vectors in (a)–(c). Also shown are Arctic sea level pressure (SLP) climatologies (hPa) for DJFM computed using Era-Interim reanalysis (d), CM4.0 (e), and an AM4.0 AMIP simulation (f). The observational data come from the Alfred-Wegener-Institute CryoSat-2 SIT product, which provides monthly winter season data spanning 2010–2011 to 2017–2018 (Ricker et al., 2014), the low-resolution daily sea ice drift product of the EUMETSAT Ocean and Sea Ice Satellite Application Facility (Lavergne et al., 2010), and Era-Interim reanalysis (Dee et al., 2011). The CM4.0 climatologies are computed over the final 10 years of the historical simulation (2005–2014), observed climatologies using Years 2011–2018, and OM4.0 climatologies over Years 1998–2007. SLP climatologies are computed over Years 2005–2014.

However, the Arctic sea level pressure bias is unable to fully account for the biased SIT pattern. We find that an OM4.0 ice-ocean simulation forced by CORE interannual atmospheric forcing (Adcroft et al., 2019) also displays a biased SIT pattern despite the reanalysis-based winds used to force this run (see Figure 29c). The sea ice drift patterns of the CORE-forced run have an improved representation of the Beaufort Gyre and transpolar drift stream; however, the velocities retain a notable high bias relative to the satellite-observed product. This simulation displays a tongue of thick ice extending from the Canadian Arctic Archipelago to the Siberian coastline and resembles a muted version of the bias pattern of CM4.0. These findings suggest that sea ice dynamics are likely an important contributor to the model's SIT bias pattern. Additional sensitivity studies exploring these SIT and drift biases are ongoing.

5.5. Tropical Oceans

Figure 30 shows CM4.0's annual-mean climatological biases over the tropical oceans, averaged over the three historical simulations during 1980–2014. The SST panel (Figure 30a) confirms with more detail the broad picture described above, with cold SST biases approaching 2 K in the eastern equatorial Pacific cold tongue (which are reduced by roughly 25% if we remove the cold global mean SST bias) and warm SST biases along the western coasts of South America and Africa (which are enhanced by removing the global mean bias). These equatorial and coastal SST biases are opposed by the time-mean net surface heat flux biases (Figure 30b), strongly suggesting that they are generated by emergent oceanic biases in transport, upwelling, or mixing. Those emergent ocean biases, in turn, may be partly attributable to local atmospheric biases (e.g., near the Pacific coast of South America, insufficient atmospheric resolution leading to weak southerly winds and oceanic upwelling) or to air-sea interactions (e.g., in the Pacific near the equator, excessive easterly trade winds driving an enhanced thermocline slope and upwelling, thereby intensifying the cold tongue and in turn the trade winds).

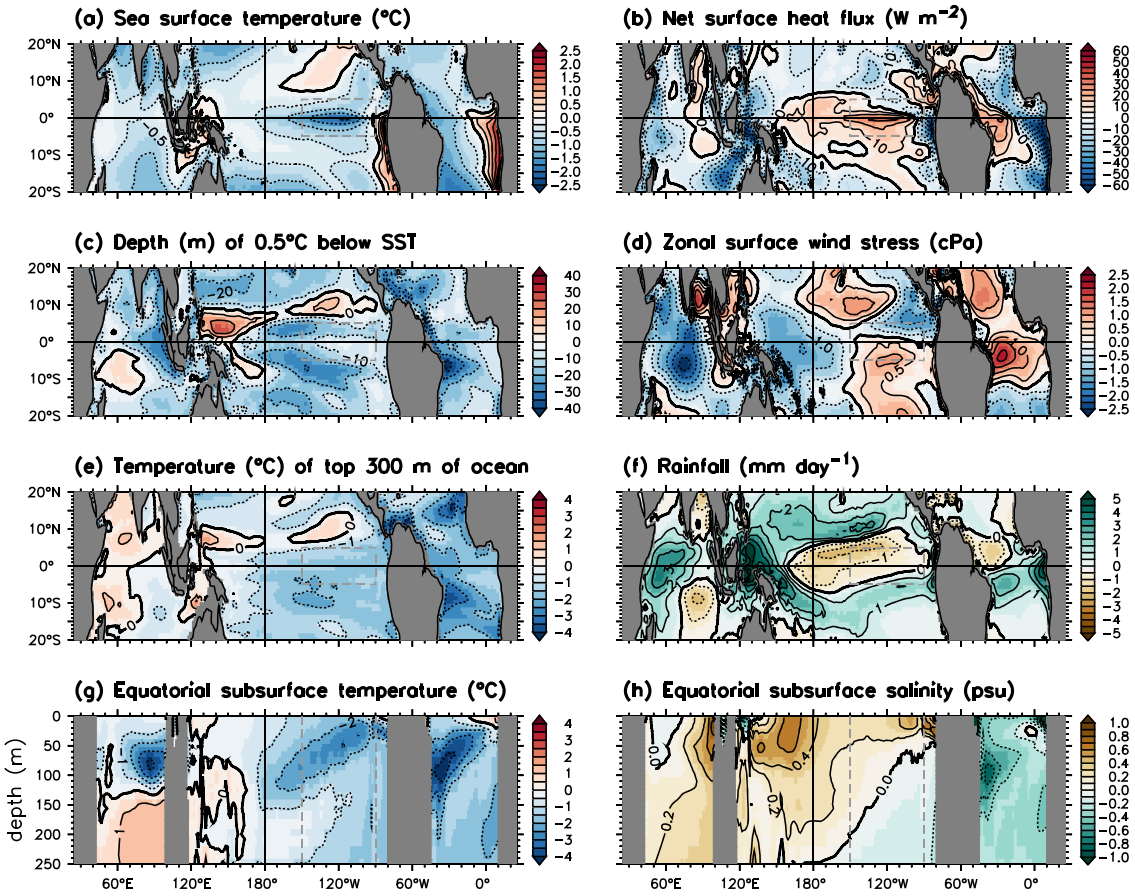

Figure 30. Annual-mean (1980–2014) climatological biases of CM4.0 relative to observational reanalyses. (a) SST (°C). (b) Net surface heat flux (W/m^2); positive values heat the ocean. (c) Depth (m) at which the time-mean ocean temperature falls 0.5 K below the time-mean SST. (d) Zonal component of the surface wind stress (cPa); positive values exert a westerly (i.e. eastward) stress on the ocean. (e) Vertically averaged temperature of the top 300 m of the ocean (shallower regions are masked out). (f) Surface rainfall (mm/day). (g) Subsurface temperature (°C) interpolated to the equator. (h) Subsurface salinity (psu) interpolated to the equator. Shading is incremented every half contour. CM4.0 represents a mean of three historical runs with CMIP6 historical forcings. The observational reference data sets for the 1980–2014 period are the ECMWF Interim reanalysis (Dee et al., 2011) for (a), (b), and (d); the ORA-S4 reanalysis (Balmaseda et al., 2013) for (c), (e), (g), and (h); and the GPCP.v2.3 data set (Adler et al., 2003, 2016) for (f). Dashed gray box is the Niño-3 region (150–90°W, 5°S–5°N).

Figure 30c shows CM4.0's bias in the climatological isothermal layer depth (ILD; defined here as the depth at which the ocean temperature is 0.5 K below the time-mean SST). In the tropics, the ILD is roughly comparable to a MLD (diagnosed from a density difference from the surface), except in regions like the Indo-Pacific warm pool where salinity contributes to near-surface density stratification. The ILD is shown here since it highlights the amount of near-surface ocean heat that is available to affect the atmosphere when the ocean is perturbed, for example, by ENSO. CM4.0's ILD is shallower than observed over most of the tropical oceans; suggesting insufficient vertical mixing in the surface ocean as discussed in Adcroft et al. (2019). Over most of the tropics, CM4.0's shallow ILD is linked to a shallower-than-observed thermocline, indicated by cold biases in the vertical-mean temperature over the top 300 m of the ocean (Figure 30e) and cold biases at thermocline depths along the equator in all three basins (Figure 30g). Biases in the equatorial wind stress (Figure 30d) also induce biases in the zonal tilt of the equatorial thermocline: Excessive easterlies produce too *much* upward tilt of the thermocline toward the eastern equatorial Pacific and Indian Oceans, and insufficient easterlies produce too *little* upward tilt of the thermocline toward the eastern equatorial Atlantic. As a result, CM4.0 produces excessive time-mean near-surface thermal stratification in the equatorial Atlantic Ocean and in the eastern equatorial Pacific and Indian Oceans, regions that play key roles in Earth's seasonal-to-interannual climate variations.

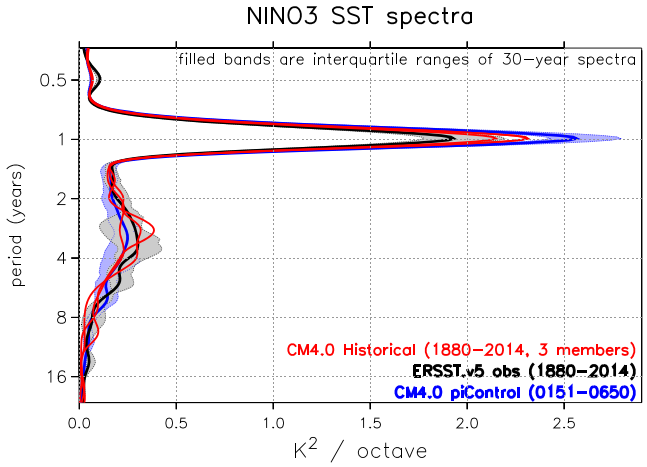

Figure 31. Power spectra of SST averaged over the Niño-3 region (150–90°W, 5°S–5°N), following Figure 2 of Wittenberg (2009). Abscissa indicates the time-averaged spectral power (K^2 per octave) from a Morlet wave number 6 wavelet analysis. The area to the left of each curve represents the total spectral power within a given range of periods on the ordinate. Red lines are 1880–2014 time-mean spectra for the three CM4.0 historical ensemble members. Black curve is the 1880–2014 time-mean spectrum from the ERSST.v5 observational reanalysis (Huang et al., 2017); gray band indicates the interquartile range of ERSST.v5 sliding 30 year subspectra during 1880–2014. Blue curve is the CM4.0 preindustrial control run time-mean spectrum during Simulated Years 151–650; light blue band is the interquartile range of sliding 30 year spectra during that 500 year epoch.

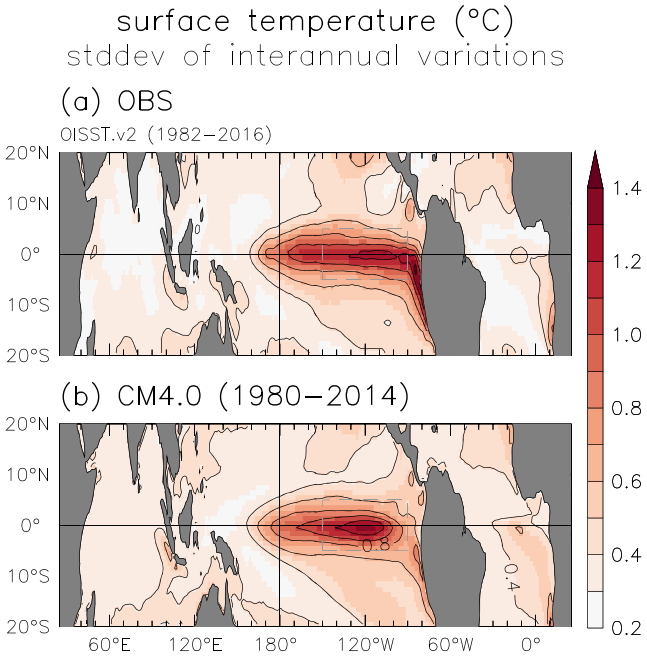

Figure 32. Standard deviation of tropical SST anomalies (K), after subtracting a 12 month climatology and applying a 9 month triangle smoother that transmits (25, 50, 75)% of the spectral amplitude at periods of (8, 11, 17) months, following Figure 17 of Delworth et al. (2012). (a) Observations from the OISST.v2 reanalysis (1982–2016; Reynolds et al., 2002). (b) Ensemble mean of the three 1980–2014 standard deviations computed from the three CM4.0 historical runs. All data sets are averaged onto a 1.25° lon × 1° lat grid before computing the temporal standard deviation. Dashed gray box is the Niño-3 region (150–90°W, 5°S–5°N).

CM4.0's wind stress biases (Figure 30d) are closely linked to its rainfall biases shown in Figure 30f, which indicate excessive rainfall in the ITCZ, south Pacific convergence zone (SPCZ), and near the Maritime Continent. CM4.0's eastern equatorial Pacific cold SST bias (Figure 30a) also displaces the Pacific ITCZ and SPCZ poleward of their observed positions, giving rise to a dry bias over the cold tongue region. Although much suppressed in the eastern Pacific compared to that in other models, as documented above, a double-ITCZ bias still appears in the southeast tropical Pacific and southwest tropical Atlantic, associated with excessive convection south of the equator during boreal spring. These rainfall biases are linked to biases in cloudiness (not shown), reducing surface shortwave heating and net ocean heat uptake (Figure 30b) near the Maritime Continent, ITCZ/SPCZ, eastern equatorial Atlantic, and western Indian Ocean. The rainfall biases also drive surface wind stress biases (Figure 30d); for example, the excessive westward rain gradients in the western equatorial Pacific and Indian Oceans result in excessive easterly trade winds in those regions. The excessive rainfall over the Maritime Continent and near-equatorial Africa also leads to reduced surface salinity and intensified near-surface haloclines in those regions (Figure 30f).

Although there are relatively straightforward connections between these biases in many cases, their ultimate cause remains a subject of investigation (Ray et al., 2018a, 2018b; Wittenberg et al., 2018). Also unclear are the precise connections between these mean biases and variability on various time scales. It is encouraging that CM4.0's simulation of the Madden-Julian Oscillation (MJO), for example, is realistic despite these mean biases, as outlined in Zhao et al. (2018a), where the fact that the MJO in this model is strongly affected by ocean coupling is described. (The coupled model utilized in Zhao et al., 2018a, to make this point is slightly different from CM4.0, but its MJO simulation is practically identical to that of CM4.0, and we refer to that paper for the characteristics of the MJO in this model.)

5.6. ENSO

Figure 31 shows power spectra for SST averaged over the Niño-3 region (150–90°W, 5°S–5°N), from an observational reconstruction and from the CM4.0 historical and preindustrial simulations. The observed time-mean spectrum (black curve) shows an energetic annual peak, a weak semiannual peak, and a broad interannual peak spanning 2–8 years. The broad interannual signature arises both from the interdecadal modulation of the ENSO's dominant period (Wittenberg, 2009; Wittenberg et al., 2014) and from the multiscale nature of ENSO's episodic and skewed SST anomalies (SSTAs)—which are characterized by strong, brief El Niño peaks but weaker, more prolonged La Niña events. The strong multidecadal modulation of ENSO is evident from the gray band in Figure 31, which shows the interquartile range of spectral curves diagnosed from 30-year (sliding) epochs during 1880–2014.

The red curves in Figure 31 show the three CM4.0 simulations driven by historical forcings over 1880–2014, which closely resemble the observed spectrum. The realism of the equatorial Pacific SST spectrum is a remarkable improvement over most previous coupled global climate models including those previously produced by GFDL (Delworth et al., 2012; Vecchi et al., 2014; Wittenberg et al., 2006). CM4.0's historical runs do appear to generate slightly less power than observed at 5–6 year time scales, where all three red curves are weaker than the black curve, although more ensemble members would be needed to confirm this. CM4.0 also produces a slightly stronger seasonal cycle than observed (mainly west of the Niño-3 region, not shown), although this is still an improvement relative to other GFDL models.

CM4.0 also appears to produce a reasonable multidecadal modulation of ENSO. The blue band of 30-year spectra in Figure 31, based on a 500-year CM4.0 piControl run, compares well with the gray band based on observations. Comparing CM4.0's historical (red) and piControl (blue) curves suggests that the historical forcings have hardly affected the time-mean spectrum of Niño-3 SST, apart from a slight weakening of the annual cycle. The historical forcings do appear to affect the higher moments of simulated ENSO SSTAs: Using the time series analyzed in Figure 31, the monthly mean Niño-3 SSTAs for (*piControl 151-650*; three historical members 1880-2014; *ERSST.v5 obs 1880-2014*) have skewness of (*0.1*; −0.1, 0.0, 0.0; *0.5*) and excess kurtosis of (*0.3*; −0.2, −0.1, −0.1; *0.6*). In other words, the Niño-3 SSTAs in CM4.0 are more normally distributed than the warm-skewed and heavy tailed observations (Newman et al., 2018), and the Niño-3 SSTAs in CM4.0 are more normally distributed under historical forcings than preindustrial.

Figure 32 shows a map of the standard deviation of tropical interannual SSTAs, for the observations and CM4.0 historical simulations. CM4.0's SSTA variability is quite realistic in both amplitude and pattern,

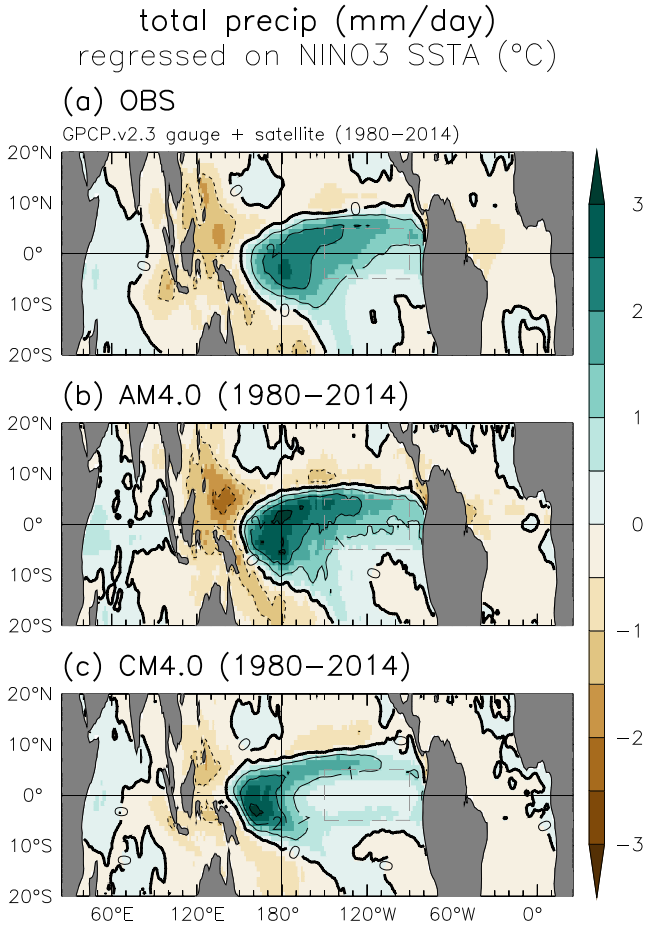

Figure 33. Linear regression of monthly rainfall anomalies onto monthly Niño-3-averaged SST anomalies ($mm \cdot day^{-1} \cdot K^{-1}$), for all months during 1980–2014. (a) Observations from GPCP.v2.3 (Adler et al., 2003, 2016) for rainfall and ERSST.v4 (Huang et al., 2015) for SST. (b) Regression computed from the AM4.0 component forced by observed SSTs during 1980–2014. (c) Ensemble mean of the three 1980–2014 regressions computed from the three CM4.0 historical runs. All data sets are averaged onto a 1.25° lon × 1° lat grid before computing the temporal regression. Dashed gray box is the Niño-3 region (150–90°W, 5°S–5°N).

except for underestimating the variability along the South American coast. CM4.0 also shows quite realistic patterns of SSTA variability in the tropical Indian and Atlantic Oceans, though slightly stronger than observed.

Figure 33 shows tropical maritime rainfall anomalies regressed onto Niño-3 SSTAs, for the observations and the AM4.0 and CM4.0 historical simulations. The observations indicate an eastward and equatorward shift of tropical Pacific rainfall during El Niño, with drying near the Maritime Continent and over the tropical Atlantic. AM4.0 (which is forced by observed SSTs during 1980–2014) shows a similar rainfall pattern, though with a slightly stronger response consistent with AM4.0's stronger *climatological* rainfall features (Figure 11 of Zhao et al., 2018a). CM4.0's rainfall response is also quite realistic, although compared to the observations and AM4.0, it is displaced farther off-equator, 10–20° farther west and shows a greater rainfall increase in the southeastern equatorial Pacific double ITCZ region along 5°S. CM4.0's biases in the ENSO rainfall response pattern do not appear to arise from biases in the SSTA pattern (Figure 32), which is quite realistic; rather, the response biases are reminiscent of CM4.0's climatological rainfall biases (Figure 30f), linked to CM4.0's cold SST bias in the eastern equatorial Pacific (Figure 30a) that inhibits local atmospheric deep convection.

Figure 34 shows the ENSO response of the zonal component of the surface wind stress. AM4.0 shows a realistic wind stress response, though with some meridional narrowing of the zone of westerly wind stress anomalies and enhanced cyclonic curl on the poleward flanks of this zone. In contrast, CM4.0's wind stress

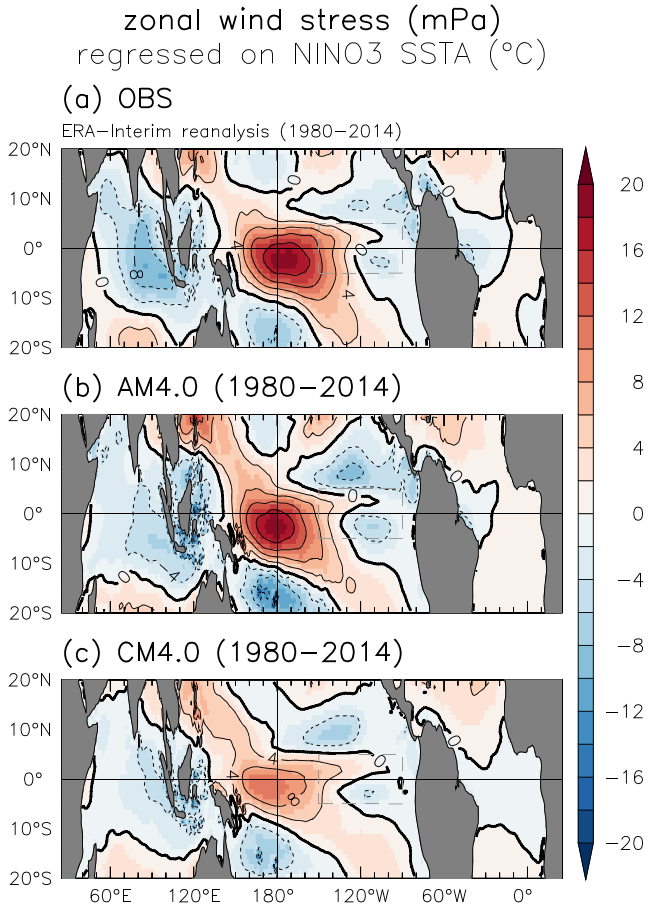

Figure 34. As in Figure 33 but for zonal wind stress anomalies regressed onto Niño-3 SST anomalies (mPa/K). Positive values indicate anomalous westerly (i.e., eastward) stress on the ocean during El Niño. Observational estimates are from the ERA-Interim reanalysis (Dee et al., 2011) for both wind stress and SST.

response is substantially weaker and narrower and shifted slightly farther west consistent with CM4.0's westward displaced rainfall anomalies (Figure 33). CM4.0's meridionally narrower and westward displaced wind stress response—which would be expected to enhance the Sverdrup discharge of equatorial oceanic heat, hastening the turnabout of El Niño into La Niña (Capotondi et al., 2006; Wittenberg, 2002)—may be partly responsible for CM4.0's reduced ENSO spectral variance at time scales of 5–6 years (Figure 31).

Figure 35 shows the ENSO response of the net surface heat flux into the ocean. Similar regressions for the individual surface heat flux components (shortwave, longwave, latent, and sensible) and cloud cover (total, low, middle, and high) were also examined to support this discussion, though for brevity are not displayed here. The observed pattern generally indicates a damping of the ENSO-driven SSTAs, with enhanced cooling over the Pacific equatorial cold tongue during El Niño. This cooling is driven mainly by enhanced evaporation in the eastern equatorial Pacific (associated with the warm SSTAs) and enhanced shading of the surface by deep convective clouds in the central and slightly off-equatorial Pacific (on the equatorward flanks of the climatological ITCZ and SPCZ and on the eastern flank of the climatological Maritime Continent convective zone). The observations also show enhanced heating of the tropical Indian and Atlantic Oceans during El Niño, due to reduced evaporation and cloudiness over those basins. AM4 produces a realistic net surface heat flux response, though with a stronger cloud shading response over all three basins, again probably linked to AM4.0's climatological biases in deep convection (Figure 11 of Zhao et al., 2018a). CM4.0, however, produces a much weaker-than-observed damping of El Niño SSTAs, with the cloud shading response displaced too far poleward and west of the evaporative cooling response in the cold tongue.

The weakness of CM4.0's ENSO wind stress coupling and surface heat flux damping relative to observations (Figures 34 and 35) seems clearly linked to the coupled model's climatological biases over the tropical

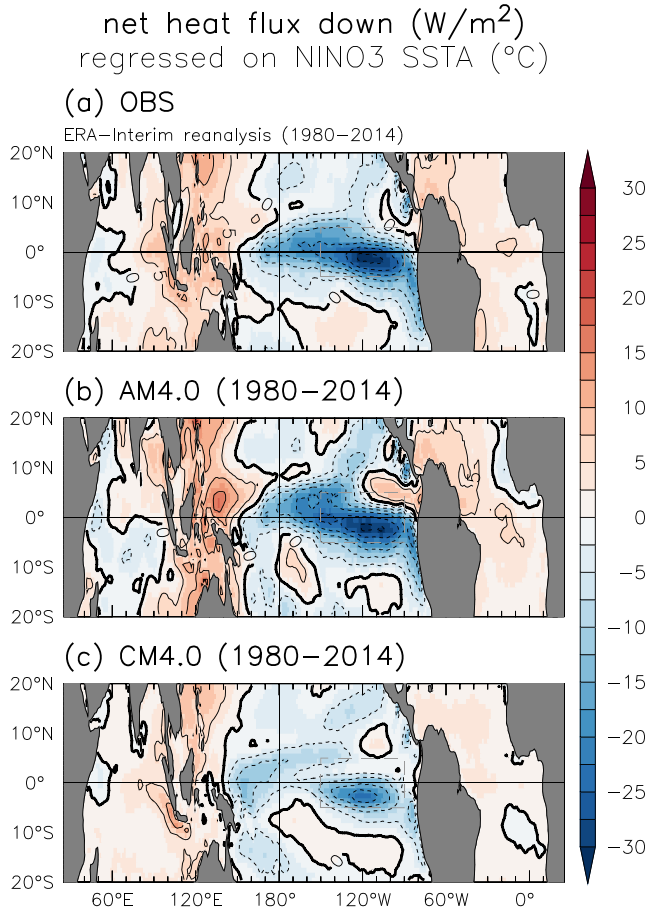

Figure 35. As in Figure 34 but for net surface heat flux anomalies regressed onto Niño-3 SST anomalies ($W \cdot m^{-2} \cdot K^{-1}$). Positive values indicate anomalous heating of the ocean during El Niño.

Pacific (Figure 30)—in particular the excessive equatorial cold tongue and resulting poleward and westward displaced atmospheric convective zones. These are all common problems in coupled global climate models, typically leading to a compensation of feedbacks in which the ENSO-weakening effect of an attenuated wind stress response is countered by the ENSO-strengthening effect of too little heat flux damping (Bayr et al., 2018). That the SST-forced AM4.0 shows that such a realistic atmospheric response to El Niño suggests that correcting CM4.0's climatological SST biases could substantially improve the coupled model's ENSO response patterns and balance of feedbacks and possibly its sensitivities to external forcings. The possibility must also be admitted that the quality of CM4.0's ENSO spectrum hides problems in these compensating stress and heat flux feedbacks.

6. Concluding Remarks

The hope is that the new CM4.0 model described herein will be of value not only for the strengths of various aspects of its climate simulations but also by serving as a starting point for additional model development, given the limitations of these same simulations. Some key conclusions or areas for further study suggested by this project include the following.

The CM4.0 coupled simulation supports an approach to atmospheric model development in which the spatial pattern of TOA fluxes is used as a prime metric when optimizing the atmosphere/land model in AMIP mode. This approach builds on numerous recent studies demonstrating that tropical precipitation, the ITCZ in particular, responds significantly to perturbations in the energy balance outside of the deep tropics (Broccoli et al., 2006; Chiang & Bitz, 2005; Kang et al., 2008). The AM4.0 TOA flux RMS errors are substantially smaller than in any other AMIP model in the CMIP5 archive in all seasons (Zhao et al., 2018a), and CM4.0's RMS biases in TOA fluxes remain smaller than all CMIP5 coupled models by the same metric.

The tuning of AM4.0's cloud scheme focused on these TOA fluxes rather than on cloud process level constraints on water/ice content or droplet size. While the results do not guarantee accurate cloud feedbacks to climate change, the ability to fit seasonal changes is a relevant constraint. Future development will need to focus on satisfying observational constraints on process level cloud variables while minimizing any loss of realism in the TOA fluxes.

The RMS precipitation biases in CM4.0 are smaller in the annual mean and in most seasons than in other GFDL and CMIP5 coupled models, helped by the relatively high quality of precipitation in AM4.0 compared to other AMIP models plus the modest SST biases that develop in CM4.0, the latter being related in part to the realism of TOA fluxes. The double ITCZ/equatorial cold tongue bias is still present to some extent but is much smaller than in GFDL's CM2.1 and CM3 models that provided data to CMIP3 and CMIP5. A problematic aspect of the CM4.0 precipitation simulation is the dry bias in Amazon rainfall that is present in coupled mode but not in the AMIP simulation, due in part to biases in tropical Atlantic SSTs.

While an ocean model resolution of 0.25° is somewhat awkward in its inability to permit realistic mesoscale eddy generation in subpolar latitudes and in dealing with marginally eddy-permitting models more generally, we find that this ocean resolution is very promising in many respects. Favorable features include the better-resolved basin boundaries and bathymetry and the quality of the eddies generated in low and middle latitudes, each of which helps in improving simulations of phenomena such as Gulf Stream separation and the Gulf of Mexico loop current, the DWBCs involved in the AMOC return flow, and the North Atlantic salinity distribution. As discussed in Adcroft et al. (2019), there is substantial room for optimization of the ocean simulation at this resolution, involving the closure for submesoscale mixed layer eddies, the surface boundary layer scheme, the generalized vertical coordinate, and backscatter-based approaches for scale-aware mesoscale eddy closures.

The strength of AMOC in CM4.0 is realistic despite misrepresentation of regional patterns of subpolar deep convection in the North Atlantic. However, it is too shallow, and relatedly, its heat transport per unit mass transport is weak, perhaps by 20%. There is very little drift in this strength over the 500 year piControl simulation. But the internal low-frequency variability of the model's AMOC is weaker than in observational estimates as well as many other models. Despite this rather weak unforced variability, the AMOC responds strongly to external forcing, strengthening due to aerosol cooling and weakening in response to greenhouse warming.

The high quality of the ENSO simulation, the frequency spectrum, and SSTA structure in particular suggest that questions concerning the response of ENSO to warming in this model will be worth pursuing. Combined with the MJO simulation described in Zhao et al. (2018a), we suggest that this model may be valuable in studies of subseasonal-to-seasonal tropical variability and predictability more generally. Yet the deterioration in the magnitude of the wind stress regressions against ENSO when moving from an AMIP to a coupled simulation suggests that biases in the base state of CM4.0 in the tropical Pacific still play a significant role in distorting ENSO dynamics.

Attribution studies of the historical era with this model will have to deal with several problematic features of the simulations. One of these is the large centennial scale variability in the SH, produced by superpolynyas forming in the Ross Sea. These do not affect the NH extratropics significantly so attribution studies focusing in that region may not have to deal with this potentially unrealistic feature. The presence of these extreme polynyas is likely due to some combination of unresolved eddy effects, small-scale vertical mixing deficiencies near the Antarctic coast, and biases in surface buoyancy forcing, resulting in too much open ocean as opposed to coastal bottom water formation. This is a central issue for future model development.

Another important difficulty affecting attribution studies is apparent in the historical simulation, which has an unrealistic temporal evolution, with little warming until roughly 1990 with rapid warming thereafter. We view this as a signature of a transient climate sensitivity and aerosol forcing that are both too large. The rather small interdecadal internal variability in the model suggests that a large model ensemble would indicate formal inconsistency with the observations with high probability.

The relatively high fidelity simulation of the extent of Arctic sea ice in CM4.0 relative to CMIP5 models is impressive, both regarding the climatological seasonal cycle and the trend in recent decades. Whether this recent trend should be captured accurately in a model with little multidecadal variability, given evidence for some contribution from the latter to the observed decline (Day et al., 2012; Ding et al., 2019; Li et al., 2017,

2018; Mahajan et al., 2011; Swart et al., 2015; Zhang, 2015; Zhang & Knutson, 2013), is a key question. In this regard we are particularly concerned with the unrealistic evolution of SSS in the subpolar North Atlantic in the model's historical simulation. The simulation of Arctic SIT also suggests deficiencies in the sea ice model's rheology or thermodynamics. The sea ice simulation around Antarctica has numerous issues—most notably too little summer sea ice and large centennial-scale variability associated with Ross Sea superpolynya events.

A variety of sensitivity studies with this coupled model have effectively been performed during the development process. Interesting sensitivities include that of subpolar Northern latitude winds and temperatures to subgrid orographic drag (especially over Greenland), temperature sensitivity to masking of snow albedos by vegetation, and overall bias responses to both horizontal and vertical atmospheric resolution, to the inclusion of energy conserving backscatter, and to specifics of the oceanic hybrid vertical coordinate. Most of these need to be reexamined with the finalized model and with longer time integrations, so we postpone discussion of these sensitivities, instead planning to return to some of them in future publications.

A model such as CM4.0 is meant to be an all-purpose climate simulator. As such, results from a variety of studies by the climate science community at large need to accumulate to determine the true value of the model for addressing various problems of interest. Substantial further work will also be required before evaluating the value of CM4.0 in providing a platform for future model improvements.

Acknowledgments
We thank Ron Stouffer and Tom Knutsen for very helpful reviews of an earlier draft. We thank Tom Delworth for consultations concerning the sensitivity of the modeled Antarctic superpolynyas to ice sheet albedo and for providing the CM2.5 data used in Figure 19. We acknowledge the World Climate Research Program which, through its Working Group on Coupled Modeling coordinated and promoting CMIP5 and CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agents who support CMIP5/6 and ESGF. We thank the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the IPCC Date Archive at the Lawrence Livermore National Laboratory/Department of Energy (LNL/DOE) for collecting the CMIP5 data. LLNL is operated by Lawrence Laboratory National Security and LLC for DOE. M. Zhao, J.-C. Golaz, B. Xiang, and Y. Ming acknowledge partial support by NOAA's Climate Program Office (CPO) Climate Variability and predictability (CVP) Program (GC14-252) through a CPO CVP funded proposal entitled "Process Level Investigation of the Role of Convection and Cloud Parameterization in Tropical Pacific Bias in GFDL Next Generation Global Climate Models". Work at LLNL was performed under Contract DE-AC52-07NA27344 from DOE/LLNL. Many GFDL colleagues and collaborators not included on the author list but who have played invaluable roles in the model workflow and publication of the data are acknowledged through doi.org/10.22033/ESGF/CMIP6.1402 website. The CM4.0 source code is available online (https://doi.org/10.5281/zenodo.3339397). Data available from authors have been deposited in the CMIP6 archive tagged with the identifier (doi.org/10.22033/ESGF/CMIP6.1402).

References

Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bishuk, 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*. 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.

Adler, RF, Sapiano, M., Huffman, G., Bolvin, D., Gu, G., Wang, J., et al. (2016). The new version 2.3 of the Global Precipitation Climatology Project (GPCP) monthly analysis product, Internal Report. College Park, MD: Earth System Science Interdisciplinary Center, Univ. of Maryland.

Balaji, V. (2012). The flexible modeling system. In *Earth system modelling-volume 3* (pp. 33–41). Berlin, Heidelberg: Springer.

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

Bayr, T., Latif, M., Dommenget, D., Wengel, C., Harlaß, J, & Park, W. (2018). Mean-state dependence of ENSO atmospheric feedbacks in climate models. *Climate Dynamics*, *50*, 3171–3194. https://doi.org/10.1007/s00382-017-3799-2

Boyer, T. P., Levitus, S., Antonov, J. I., Locarnini, R. A., & Garcia, H. E. (2005). Linear trends in salinity for the world ocean, 1955–1998. *Geophysical Research Letters*, *32*, L01604. https://doi.org/10.1029/2004GL021791

Broccoli, A. J., Dahl, K. A., & Stouffer, R. J. (2006). Response of the ITCZ to Northern Hemisphere cooling. *Geophysical Research Letters*, *33*, L01702. https://doi.org/10.1029/2005GL024546

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.

Campbell, E. C., Wilson, E. A., Moore, G. W. K., Riser, S. C., Brayton, C. E., Mazloff, M. R., & Talley, L. D. (2019). Antarctic offshore polynyas linked to Southern Hemisphere climate anomalies. *Nature*, *570*, 1.

Capotondi, A., Wittenberg, A., & Masina, S. (2006). Spatial and temporal structure of tropical Pacific interannual variability in 20th century coupled simulations. *Ocean Modelling*, *15*, 274–298. https://doi.org/10.1016/j.ocemod.2006.02.004

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. NASA DAAC at the Natl. Snow and Ice Data Cent. https://doi.org/10.5067/8GQ8LZQVL0VL

Cess, R. D., Potter, G. L., Blanchet, J. P., Boer, G. J., Del Genio, A. D., DéQué, M., et al. (1990). Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models. *Journal of Geophysical Research*, *95*(D10), 16,601–16,615.

Cess, R. D., Zhang, M. H., Ingram, W. J., Potter, G. L., Alekseev, V., Barker, H. W., et al. (1996). Cloud feedback in atmospheric general circulation models: An update. *Journal of Geophysical Research*, *101*(D8), 12,791–12,794.

Chassignet, E. P., Smith, L. T., Halliwell, G. R., & Bleck, R. (2003). North Atlantic simulations with the Hybrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference pressure, and thermobaricity. *Journal of Physical Oceanography*, *33*(12), 2504–2526. https://doi.org/10.1175/1520-0485(2003)033<2504:NASWTH>2.0.CO;2

Chiang, J. C., & Bitz, C. M. (2005). Influence of high latitude ice cover on the marine Intertropical Convergence Zone. *Climate Dynamics*, *25*(5), 477–496.

Cunningham, S. A., Kanzow, T., Rayner, D., Baringer, M. O., Johns, W. E., Marotzke, J., et al. (2007). Temporal variability of the Atlantic Meridional Overturning Circulation at 26.5°N. *Science*, *317*(5840), 935–938.

Danabasoglu, G., Large, W. G., & Briegleb, B. P. (2010). Climate impacts of parameterized Nordic Sea overflows. *Journal of Geophysical Research*, *115*, C11005. https://doi.org/10.1029/2010JC006243

Danabasoglu, G., Yeager, S. G., Bailey, D., Behrens, E., Bentsen, M., Bi, D., et al. (2014). North Atlantic simulations in Coordinated Ocean-Ice Reference Experiments Phase II (CORE-II). Part I: Mean states. *Ocean Modelling*, *73*, 76–107.

Day, J. J., Hargreaves, J. C., Annan, J. D., & Abe-Ouchi, A. (2012). Sources of multi-decadal variability in Arctic sea ice extent. *Environmental Research Letters*, *7*(3), 034011.

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research*, *109*, C12003. https://doi.org/10.1029/2004JC002378

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*, 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.

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*, 2755–2781. https://doi.org/10.1175/JCLI-D-11-00316.1

Ding, Q., Schweiger, A., L'Heureux, M., Steig, E. J., Battisti, D. S., Johnson, N. C., et al. (2019). Fingerprints of internal drivers of Arctic sea ice loss in observations and model simulations. *Nature Geoscience*, *12*(1), 28.

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.

Dufour, C. O., Morrison, A. K., Griffies, S. M., Frenger, I., Zanowski, H., & Winton, M. (2017). Preconditioning of the Weddell Sea polynya by the ocean mesoscale and dense water overflows. *Journal of Climate*, *30*(19), 7719–7737. https://doi.org/10.1175/JCLI-D-16-0586.1

Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., et al. (2012). GFDL's ESM2 global coupled climate–carbon earth system models. Part I: Physical formulation and baseline simulation characteristics. *Journal of Climate*, *25*(19), 6646–6665.

Evans, D. G., Toole, J., Forget, G., Zika, J. D., Naveira Garabato, A. C., Nurser, A. J. G., & Yu, L. (2017). Recent wind-driven variability in Atlantic Water mass distribution and meridional overturning circulation. *Journal of Physical Oceanography*, *47*(3), 633–647.

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

Fox-Kemper, B., Ferrari, R., & Hallberg, R. (2008). Parameterization of mixed layer eddies. Part I: Theory and diagnosis. *Journal of Physical Oceanography*, *38*(6), 1145–1165.

GISTEMP-Team (2019). GISS Surface Temperature Analysis (GISTEMP), version 4. NASA Goddard Institute for Space Studies. Dataset accessed 2016-09-27 at https://data.giss.nasa.gov/gistemp/

Garner, S. T. (2005). A topographic drag closure built on an analytical base flux. *Journal of the Atmospheric Sciences*, *62*(7), 2302–2315.

Garner, S. T. (2018). Ground-truth model evaluation of subgrid orographic base-flux parameterization. *Journal of the Atmospheric Sciences*, *75*(10), 3653–3670.

Gent, P. R., Willebrand, J., McDougall, T. J., & McWilliams, J. C. (1995). Parameterizing eddy-induced tracer transports in ocean circulation models. *Journal of Physical Oceanography*, *25*, 463–474.

Golaz, J.-C., Horowitz, L. W., & Levy, H. (2013). Cloud tuning in a coupled climate model: Impact on 20th century warming. *Geophysical Research Letters*, *40*, 2246–2251. https://doi.org/10.1002/grl.50232

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., et al. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, *31*, L03205. https://doi.org/10.1029/2003GL018747

Griffies, S. M., Harrison, M. J., Pacanowski, R. C., & Rosati, A. (2004). A technical guide to MOM4, *5*, 371.

Griffies, S. M., Winton, M., Anderson, W. G., Benson, R., Delworth, T. L., Dufour, C. O., et al. (2015). Impacts on ocean heat from transient mesoscale eddies in a hierarchy of climate models. *Journal of Climate*, *28*(3), 952–977.

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.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., et al. (2016). High resolution model intercomparison project (HighResMIP v1. 0) for CMIP6. *Geoscientific Model Development*, *9*(1), 4185–4208.

Hallberg, R. (2013). Using a resolution function to regulate parameterizations of oceanic mesoscale eddy effects. *Ocean Modelling*, *72*, 92–103.

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*, 911–930. https://doi.org/10.1175/JCLI-D-14-00006.1

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017). Extended Reconstructed Sea Surface Temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *Journal of Climate*, *30*, 8179–8205.

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.

John, J. G., Fiore, A. M., Naik, V., Horowitz, L. M., & Dunne, J. P. (2012). Climate versus emission drivers of methane lifetime against loss by tropospheric oh from 1860–2100. *Atmospheric Chemistry and Physics*, *12*, 12,021–12,036. https://doi.org/10.5194/acp-12-12021-2012

Johns, W. E., Baringer, M. O., Beal, L. M., Cunningham, S. A., Kanzow, T., Bryden, H. L., et al. (2011). Continuous, array-based estimates of Atlantic Ocean heat transport at 26.5°N. *Journal of Climate*, *24*(10), 2429–2449.

Kang, S. M., Held, I. M., Frierson, D. M. W., & Zhao, M. (2008). The response of the ITCZ to extratropical thermal forcing: Idealized slab-ocean experiments with a GCM. *Journal of Climate*, *21*(14), 3521–3532.

Lavergne, T., Eastwood, S., Teffah, Z., Schyberg, H., & Breivik, L.-A. (2010). Sea ice motion from low-resolution satellite sensors: An alternative method and its validation in the Arctic. *Journal of Geophysical Research*, *115*, C10032. https://doi.org/10.1029/2009JC005958

Lenssen, N., Schmidt, G., Hansen, J., Menne, M., 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

Li, D., Zhang, R., & Knutson, T. R. (2017). On the discrepancy between observed and CMIP5 multi-model simulated Barents Sea winter sea ice decline. *Nature Communications*, *8*, 14991.

Li, D., Zhang, R., & Knutson, T. R. (2018). Comparison of mechanisms for low-frequency variability of summer Arctic sea ice in three coupled models. *Journal of Climate*, *31*(3), 1205–1226.

Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H. E., Baranova, O. K., et al. (2013). World ocean atlas 2013. Volume 1, temperature.

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.

Mahajan, S., Zhang, R., & Delworth, T. L. (2011). Impact of the Atlantic Meridional Overturning Circulation (AMOC) on Arctic surface air temperature and sea ice variability. *Journal of Climate*, *24*(24), 6573–6581.

Marshall, J., & Schott, F. (1999). Open-ocean convection: Observations, theory, and models. *Reviews of Geophysics*, *37*, 1–64.
Martin, T., Park, W., & Latif, M. (2013). Multi-centennial variability controlled by Southern Ocean convection in the Kiel Climate Model. *Climate Dynamics*, *40*(7-8), 2005–2022.
Massonnet, F., Fichefet, T., Goosse, H., Bitz, C. M., Philippon-Berthier, G., Holland, M. M., & Barriat, P.-Y. (2012). Constraining projections of summer Arctic sea ice. *The Cryosphere*, *6*(6), 1383–1394.
McCarthy, G. D., Smeed, D. A., Johns, W. E., Frajka-Williams, E., Moat, B. I., Rayner, D., et al. (2015). Measuring the Atlantic Meridional Overturning Circulation at 26° N. *Progress in Oceanography*, *130*, 91–111.
Menary, M. B., Roberts, C. D., Palmer, M. D., Halloran, P. R., Jackson, L., Wood, R. A., et al. (2013). Mechanisms of aerosol-forced amoc variability in a state of the art climate model. *Journal of Geophysical Research: Oceans*, *118*, 2087–2096. https://doi.org/10.1002/jgrc.20178
Milly, P. C., Malyshev, S. L., Shevliakova, 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.
Murakami, H., Vecchi, G. A., Underwood, S., Delworth, T. L., Wittenberg, A. T., Anderson, W. G., et al. (2015). Simulation and prediction of category 4 and 5 hurricanes in the high-resolution GFDL HiFLOR coupled climate model. *Journal of Climate*, *28*(23), 9058–9079.
Murray, R. J. (1996). Explicit generation of orthogonal grids for ocean models. *Journal of Computational Physics*, *126*, 251–273.
Newman, M., Wittenberg, A. T., Cheng, L., Compo, G. P., & Smith, C. A. (2018). The extreme 2015/16 El Niño, in the context of historical climate variability and change in "Explaining extreme events of 2016 from a climate perspective". *Bulletin of the American Meteorological Society*, *99*(1), S16–S20. https://doi.org/10.1175/BAMS-D-17-0116.1
Putman, W. M., & Lin, S.-J. (2007). Finite-volume transport on various cubed-sphere grids. *Journal of Computational Physics*, *227*(1), 55–78.
Ray, S., Wittenberg, A. T., Griffies, S. M., & Zeng, F. (2018a). Understanding the equatorial Pacific cold tongue time-mean heat budget, part I: Diagnostic framework. *Journal of Climate*, *31*, 9965–9985. https://doi.org/10.1175/JCLI-D-18-0152.1
Ray, S., Wittenberg, A. T., Griffies, S. M., & Zeng, F. (2018b). Understanding the equatorial Pacific cold tongue time-mean heat budget, Part II: Evaluation of the GFDL-FLOR coupled GCM. *Journal of Climate*, *31*, 9987–10011. https://doi.org/10.1175/JCLI-D-18-0153.1
Rayner, N. A. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002JD002670
Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., & Wang, W. (2002). An improved in situ and satellite SST analysis for climate. *Journal of Climate*, *15*, 1609–1625.
Ricker, R., Hendricks, S., Helm, V., Skourup, H., & Davidson, M. (2014). Sensitivity of CryoSat-2 Arctic sea-ice freeboard and thickness on radar-waveform interpretation. *The Cryosphere*, *8*(4), 1607–1622.
Saba, V. S., Griffies, S. M., Anderson, W. G., Winton, M., Alexander, M. A., Delworth, T. L., et al. (2016). Enhanced warming of the Northwest Atlantic Ocean under climate change. *Journal of Geophysical Research: Oceans*, *121*, 118–132. https://doi.org/10.1002/2015JC011346
Shevliakova, E., Pacala, S. W., Malyshev, S., Hurtt, G. C., Milly, P. C. D., Caspersen, J. P., et al. (2009). Carbon cycling under 300 years of land use change: Importance of the secondary vegetation sink. *Global Biogeochemical Cycles*, *23*, GB2022. https://doi.org/10.1029/2007GB003176
Shu, Q., Song, Z., & Qiao, F. (2015). Assessment of sea ice simulations in the CMIP5 models. *The Cryosphere*, *9*(1), 399–409.
Smeed, D. A., Josey, S. A., Beaulieu, C., Johns, W. E., Moat, B. I., Frajka-Williams, E., et al. (2018). The North Atlantic Ocean is in a state of reduced overturning. *Geophysical Research Letters*, *45*, 1527–1533. https://doi.org/10.1002/2017GL076350
Smith, G. L., Priestley, K. J., Loeb, N. G., Wielicki, B. A., Charlock, T. P., Minnis, P., et al. (2011). Clouds and Earth Radiant Energy System (CERES), a review: Past, present and future. *Advances in Space Research*, *48*(2), 254–263.
Stan, C., Khairoutdinov, M., DeMott, C. A., Krishnamurthy, V., Straus, D. M., Randall, D. A., et al. (2010). An ocean-atmosphere climate simulation with an embedded cloud resolving model. *Geophysical Research Letters*, *37*, L01702. https://doi.org/10.1029/2009GL040822
Stouffer, R. J. (2004). Time scales of climate response. *Journal of Climate*, *17*(1), 209–217.
Stroeve, J. C., Kattsov, V., Barrett, A., Serreze, M., Pavlova, T., Holland, M., & Meier, W. N. (2012). Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations. *Geophysical Research Letters*, *39*, L16502. https://doi.org/10.1029/2012GL052676
Swart, N. C., Fyfe, J. C., Hawkins, E., Kay, J. E., & Jahn, A. (2015). Influence of internal variability on arctic sea-ice trends. *Nature Climate Change*, *5*, 86–89.
Talandier, C., Deshayes, J., Treguier, A.-M., Capet, X., Benshila, R., Debreu, L., et al. (2014). Improvements of simulated Western North Atlantic current system and impacts on the AMOC. *Ocean Modelling*, *76*, 1–19.
Turner, J., Bracegirdle, T. J., Phillips, T., Marshall, G. J., & Hosking, J. S. (2013). An initial assessment of Antarctic sea ice extent in the CMIP5 models. *Journal of Climate*, *26*(5), 1473–1484.
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*, 7994–8016. https://doi.org/10.1175/JCLI-D-14-00158.1
Wang, H., Legg, S. A., & Hallberg, R. W. (2015). Representations of the Nordic Seas overflows and their large scale climate impact in coupled models. *Ocean Modelling*, *86*, 76–92.
Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., et al. (2011). MIROC-ESM 2010: Model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development*, *4*, 845–872. https://doi.org/10.5194/gmd-4-845-2011
Wittenberg, A. T. (2002). *ENSO response to altered climates* (PhD Thesis). Princeton, NJ: Princeton University. https://extranet.gfdl.noaa.gov/~atw/research/thesis/
Wittenberg, A. T. (2009). Are historical records sufficient to constrain ENSO simulations?. *Geophysical Research Letters*, *36*, L14709. https://doi.org/10.1029/2009GL038710
Wittenberg, A. T., Rosati, A., Delworth, T. L., Vecchi, G. A., & Zeng, F. (2014). ENSO modulation: Is it decadally predictable? *Journal of Climate*, *27*, 2667–2681. https://doi.org/10.1175/JCLI-D-13-00577.1
Wittenberg, A. T., Rosati, A., Lau, N.-C., & Ploshay, J. J. (2006). GFDL's CM2 global coupled climate models. Part III: Tropical Pacific climate and ENSO. *Journal of Climate*, *19*, 698–722. https://doi.org/10.1175/JCLI3631.1
Wittenberg, A. T., Vecchi, G. A., Delworth, T. L., Rosati, A., Anderson, W., Cooke, W. F., et al. (2018). Improved simulations of tropical Pacific annual-mean climate in the GFDL FLOR and HiFLOR coupled GCMs. *Journal of Advances in Modeling Earth Systems*, *10*, 3176–3220. https://doi.org/10.1029/2018MS001372
Yan, X., Zhang, R., & Knutson, T. R. (2018). Underestimated AMOC variability and implications for AMV and predictability in CMIP models. *Geophysical Research Letters*, *45*, 4319–4328. https://doi.org/10.1029/2018GL077378
Yeager, S. G., & Danabasoglu, G. (2012). Sensitivity of Atlantic Meridional Overturning Circulation variability to parameterized Nordic Sea overflows in CCSM4. *Journal of Climate*, *25*, 2077–2103.
Yeager, S. G., & Jochum, M. (2009). The connection between Labrador Sea buoyancy loss, deep western boundary current strength, and Gulf Stream path in an ocean circulation model. *Ocean Modelling*, *30*(2-3), 207–224.

Zhang, R. (2015). Mechanisms for low-frequency variability of summer Arctic sea ice extent. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(15), 4570–4575.

Zhang, R. (2017). On the persistence and coherence of subpolar sea surface temperature and salinity anomalies associated with the Atlantic multidecadal variability. *Geophysical Research Letters*, *44*, 7865–7875. https://doi.org/10.1002/2017GL074342

Zhang, L., Delworth, T. L., & Jia, L. (2017). Diagnosis of decadal predictability of Southern Ocean sea surface temperature in the GFDL CM2.1 model. *Journal of Climate*, *30*(16), 6309–6328.

Zhang, R., Delworth, T. L., Rosati, A., Anderson, W. G., Dixon, K. W., Lee, H.-C., & Zeng, F. (2011). Sensitivity of the North Atlantic Ocean circulation to an abrupt change in the Nordic Sea overflow in a high resolution global coupled climate model. *Journal of Geophysical Research*, *116*, C12024. https://doi.org/10.1029/2011JC007240

Zhang, R., Delworth, T. L., Sutton, R., Hodson, D. R., Dixon, K. W., Held, I. M., et al. (2013). Have aerosols caused the observed Atlantic multidecadal variability? *Journal of the Atmospheric Sciences*, *70*(4), 1135–1144.

Zhang, R., & Knutson, T. R. (2013). The role of global climate change in the extreme low summer Arctic sea ice extent in 2012 in "Explaining Extreme events of 2012 from a climate perspective". *Bulletin of the American Meteorological Society*, *94*(9), S23–S26.

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

Zhang, R., & Vallis, G. K. (2007). The role of bottom vortex stretching on the path of the North Atlantic western boundary current and on the northern recirculation gyre. *Journal of Physical Oceanography*, *37*(8), 2053–2080.

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

Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., et al. (2018b). The GFDL global atmosphere and land model AM4.0/LM4.0: 2. Model description, sensitivity studies, and tuning strategies. *Journal of Advances in Modeling Earth Systems*, *10*, 735–769. https://doi.org/10.1002/2017MS001209

Zweng, M. M., Reagan, J. R., Antonov, J. I., Locarnini, R. A., Mishonov, A. V., Boyer, T. P., et al. (2013). World ocean atlas 2013. Volume 2, salinity.

Zweng, M. M., Reagan, J. R., Seidov, D., Boyer, T. P., Locarnini, R. A., Garcia, H. E., & Mishonov, A. V. (2018). World ocean atlas 2018. Volume 2, salinity.