

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

I. M. Held⁶, H. Guo¹, A. Adcroft¹, J. P. Dunne¹, L. W. Horowitz^{1,6}, J. Krasting¹, E. Shevliakova¹, M. Winton¹, M. Zhao¹, M. Bushuk⁴, A. T. Wittenberg¹, B. Wyman¹, B. Xiang⁴, R. Zhang^{1,6}, W. Anderson¹, V. Balaji², L. Donner^{1,6}, K. Dunne³, J. Durachta¹, P.P.G. Gauthier², P. Ginoux¹, J.-C. Golaz⁷, S. M. Griffies^{1,6}, R. Hallberg^{1,6}, L. Harris¹, M. Harrison¹, W. Hurlin¹, J. John¹, P. Lin², S.-J. Lin¹, S. Malyshev¹, R. Menzel⁴, P.C.D. Milly^{1,3}, Y. Ming¹, V. Naik¹, D. Paynter¹, F. Paulot¹, V. Ramaswamy^{1,6}, B. Reichl¹, T. Robinson⁵, A. Rosati², C. Seman¹, L. Silvers², S. Underwood¹, N. Zadeh⁵

¹GFDL/NOAA, Princeton, New Jersey

²Cooperative Institute for Modeling the Earth System, Princeton University, Princeton, New Jersey

³U.S. Geological Survey/GFDL, Princeton, New Jersey

⁴University Corporation for Atmospheric Research/GFDL, Princeton, New Jersey

⁵SAIC, Science Applications International Corporation

⁶Princeton University Program in Atmospheric and Oceanic Sciences, Princeton, NJ, USA

⁷Lawrence Livermore National Laboratory, Livermore, CA

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.

Corresponding author: Huan Guo, huan.guo@noaa.gov

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination, and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2019MS001829

24 **Abstract**

25 We describe GFDL's CM4.0 physical climate model, with emphasis on those aspects
 26 that may be of particular importance to users of this model and its simulations. The model
 27 is built with the AM4.0/LM4.0 atmosphere/land model and OM4.0 ocean model. Topics in-
 28 clude the rationale for key choices made in the model formulation, the stability as well as
 29 drift of the pre-industrial control simulation, and comparison of key aspects of the historical
 30 simulations with observations from recent decades. Notable achievements include the rela-
 31 tively small biases in seasonal spatial patterns of top-of-atmosphere fluxes, surface temper-
 32 ature, and precipitation; reduced double Intertropical Convergence Zone bias; dramatically
 33 improved representation of ocean boundary currents; a high quality simulation of climato-
 34 logical Arctic sea ice extent and its recent decline; and excellent simulation of the El Niño-
 35 Southern Oscillation spectrum and structure. Areas of concern include inadequate deep con-
 36 vention in the Nordic Seas; an inaccurate Antarctic sea ice simulation; precipitation and wind
 37 composites still affected by the equatorial cold tongue bias; muted variability in the Atlantic
 38 Meridional Overturning Circulation; strong 100 year quasi-periodicity in Southern Ocean
 39 ventilation; and a lack of historical warming before 1990 and too rapid warming thereafter
 40 due to high climate sensitivity and strong aerosol forcing, in contrast to the observational
 41 record. Overall, CM4.0 scores very well in its fidelity against observations compared to the
 42 Coupled Model Intercomparison Project Phase 5 generation in terms of both mean state and
 43 modes of variability and should prove a valuable new addition for analysis across a broad
 44 array of applications.

45 **Plain Language Summary**

46 The Geophysical Fluid Dynamics Laboratory (GFDL) of the National Oceanic and
 47 Atmospheric Administration (NOAA) participates along with a number of model centers
 48 around the world in constructing state-of-the-art climate models for use in studies for cli-
 49 mate change and prediction. GFDL's latest multi-purpose atmosphere-ocean coupled climate
 50 model, CM4.0, is described here. It consists of GFDL's latest atmosphere and land mod-
 51 els at about 100 km horizontal resolution, and ocean and sea ice models at roughly 25 km
 52 horizontal resolution. A handful of standard experiments have been conducted with CM4.0
 53 for participation in the Coupled Model Intercomparison Project Phase 6, an archive of cli-
 54 mate model results utilized by the Intergovernmental Panel on Climate Change and the cli-
 55 mate research community more generally. The model results have been extensively evaluated
 56 against observations. This paper makes the case that CM4.0 ranks high among state-of-the-
 57 art coupled climate models by many measures of bias in the simulated climatology and in
 58 its ability to capture modes of climate variability such as the El Niño-Southern Oscillation
 59 and Madden-Julian Oscillation. The paper also discusses some potential weaknesses, includ-
 60 ing unrealistically large internal variability in the Southern Ocean and insufficient warming
 61 before 1990 in the simulation of the 20th century.

62 **1 Introduction**

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

71 Data from the CM4.0 simulations deposited in the CMIP6 archive have been tagged
 72 with the identifier doi.org/10.22033/ESGF/CMIP6.1402 for citation purposes. Model source

73 code is available at <https://doi.org/10.5281/zenodo.3339397>. We also provide the links to
 74 CM4.0 model inputs and outputs, selected analysis scripts and supporting data, among others
 75 at <https://data1.gfdl.noaa.gov/nomads/forms/cm4/>.

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

98 2 Model configuration and relationships with other past and ongoing modeling 99 efforts at GFDL

100 Following on earlier pioneering efforts, global coupled model development activity at
 101 GFDL was reorganized on a more lab-wide basis in the early 2000's, building on a common
 102 software infrastructure (the Flexible Modeling System, *Balaji* [2012]). This effort resulted
 103 in the CM2.1 model [*Delworth et al.*, 2006] from which results were provided for CMIP3.
 104 CM2.1 later evolved into higher resolution versions CM2.5 and CM2.6 [*Delworth et al.*,
 105 2012], as well as models with high resolution only in the atmosphere: the Forecast-oriented
 106 Low Ocean Resolution (FLOR) model [*Vecchi et al.*, 2014], and the High-Resolution Forecast-
 107 oriented Low Ocean Resolution (HiFLOR) model [*Murakami et al.*, 2015]. The higher res-
 108 olution atmosphere in these coupled models utilizes the cubed sphere version of the finite
 109 volume atmospheric dynamical core [*Putman and Lin*, 2007].

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

119 The CM3 model that provided results for CMIP5 [*Donner et al.*, 2011] was a departure
 120 from the CM2 series in the atmosphere (AM3), with a new convective closure, with the ad-
 121 dition of a substantial tropospheric and stratospheric chemical mechanism allowing for the
 122 treatment of ozone as a prognostic variable and the ability to run from emissions of aerosol
 123 precursors, and with more resolution in the stratosphere and a higher model top. This model

124 also incorporated the cubed sphere dynamical core. It included the laboratory's first attempt
 125 at simulating indirect aerosol effects, which emerged as being very large. Modified versions
 126 of CM3, with weaker indirect aerosol cooling [Golaz *et al.*, 2013] were found to provide bet-
 127 ter simulations of the trajectory of 20th century warming. The ocean and sea ice components
 128 of CM3 are very close to those in CM2.1 [Griffies *et al.*, 2011]. LM3, the land component of
 129 ESM2M, ESM2G, and CM3 [Milly *et al.*, 2014], was substantially modified from that in pre-
 130 vious models, in its hydrology and especially in its interactive vegetation [Shevliakova *et al.*,
 131 2009].

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

136 As described in Zhao *et al.* [2018a,b], the AM4.0 model was primarily designed to
 137 build upon the results from AM3 and HiRAM. New features of this model include: substan-
 138 tial improvements to the radiation code; a topographic gravity wave drag with a novel treat-
 139 ment of anisotropy; a "light" chemistry mechanism using AM3's as a starting point, with the
 140 capability of simulating aerosols from emissions but with prescribed ozone and other oxi-
 141 dants; aerosol and cloud microphysics modules similar to AM3's but with modifications that
 142 affect the indirect aerosol effect; and a new "double-plume" convective closure for shallow
 143 and deep convection. The atmospheric model's dynamical core is the cube-sphere finite vol-
 144 ume (FV3) hydrostatic version [Putman and Lin, 2007] modified from the version in AM3 to
 145 improve treatment of orography and computation efficiency.

146 The atmospheric component of CM4.0 is identical to the atmospheric model AM4.0
 147 documented by Zhao *et al.* [2018a,b] with no retuning. AM4.0 has C96 resolution, i.e., there
 148 are 96x96 grid boxes on each of the 6 cube faces that cover the sphere, or roughly 100km
 149 resolution. It has 33 vertical levels topped at 1 hPa with limited stratospheric resolution that
 150 would be commonly categorized as "low top".

151 CM4.0 uses a new ocean code (MOM6) with a variety of innovations including a gen-
 152 eralized vertical coordinate, new treatment of the bottom boundary, and a new surface bound-
 153 ary layer scheme. The ocean code is documented in Adcroft *et al.* [2019] along with the de-
 154 tailed formulation used in OM4.0, the oceanic component of CM4.0. OM4.0 has a nominal
 155 horizontal grid spacing of 0.25 degrees with a tri-polar grid following Murray [1996]. A
 156 noteworthy feature is that the ocean model is run with no advective or diffusive mesoscale
 157 eddy closure, although it does include a sub-mesoscale stirring parameterization based on
 158 that in Fox-Kemper *et al.* [2008] that provides a vertical buoyancy flux in the mixed layer,
 159 to which the model is sensitive. The model has 75 degrees of freedom in the vertical utiliz-
 160 ing hybrid vertical coordinates similar to those in Chassignet *et al.* [2003]. The upper water
 161 column is z-coordinate with 2m grid spacing near the surface, increasing to roughly 20m at
 162 200 meter depth. The coordinate transitions to isopycnal surfaces (referenced to 2000 dbar)
 163 in the upper thermocline and remains isopycnal throughout the abyssal ocean. The simula-
 164 tions referred to here as OM4.0 follow the inter-annual forcing CORE2 (Coordinated Ocean-
 165 Ice Reference Experiment) protocol [Danabasoglu *et al.*, 2014] in which an atmospheric
 166 reanalysis drives the ice-ocean model.

167 CM4.0 also incorporates a new sea ice model, SIS2.0, which is a significant upgrade
 168 from the SIS model used in previous generations of GFDL climate models [Delworth *et al.*,
 169 2006]. SIS2.0 has five thickness categories and the vertical thermodynamics uses four ice
 170 layers and one snow layer. A radiative transfer scheme is incorporated for simulation of the
 171 vertical profile of shortwave absorption. The model dynamics use a C-grid stencil as does
 172 the ocean. More details about SIS2.0 can be found in Adcroft *et al.* [2019].

173 While the land model utilized has much in common with the model LM4.0 used in
 174 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 CO₂ concentrations). The land component is similar to that in ESM2M [Shevliakova *et al.*, 2009]. The ocean biogeochemistry component, referred to as BLING (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.

- 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 mid-latitudes, 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.
- 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 sub-grid closures and specification of the hybrid vertical coordinate.
- 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 func-

tion (pdf) 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 below in Section 4.1, a key problematic feature in late-stage development of CM4.0 was the appearance of quasi-periodic super-polynyas in the Southern sub-polar 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]*; *Zhang et al. [2017]*; *Martin et al. [2013]* but the CM4.0 version is very strong, likely unrealistically so. Following ongoing experience 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 super-polynyas during the piControl simulation, by reducing the stratification in the upper sub-polar 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 super-polynyas still form eventually.
 7. Land model development focused on concerns regarding slow trends towards unrealistically extensive boreal forests in the pre-industrial 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 (SSTs) uniformly by 2K and examining the change in the net TOA flux [*Cess and Coauthors, 1990, 1996*], is relatively modest ($0.57K/Wm^{-2}$). However, throughout the CM4.0 development it was clear that this coupled model had higher than expected transient climate response (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 sen-

sitivity. 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 multi-faceted 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 sub-grid scale orography (the resulting parameterized mountain torque is related to the low level winds through a non-diagonal 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 sub-grid drag simulated over Greenland.

4 The pre-industrial control

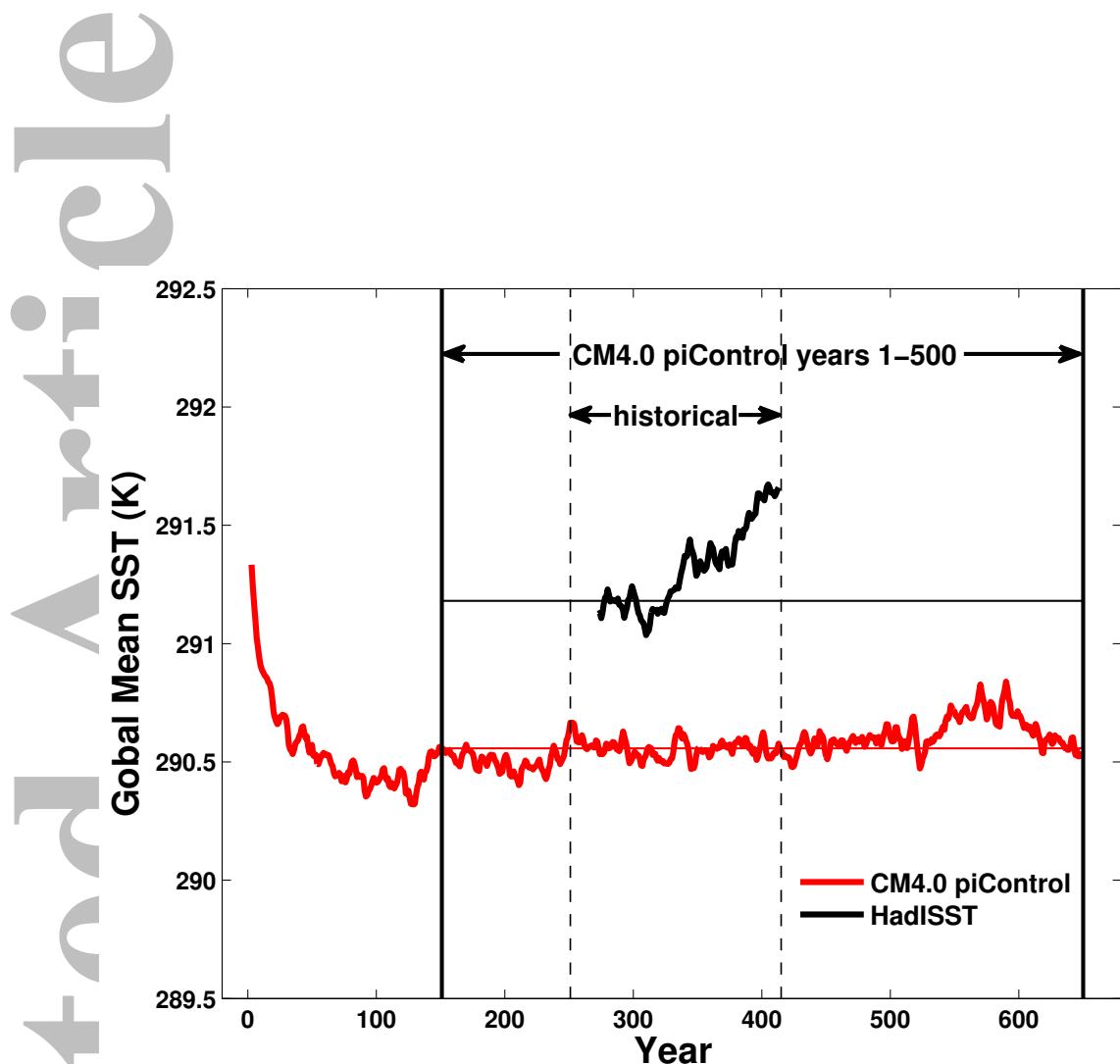
4.1 Mean temperature evolution, energy balance, and centennial-scale variability

Preliminary pre-industrial 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 super-polynya behavior, as mentioned in Sec.3.

The piControl simulation designated as the CMIP6 DECK (Diagnostic, Evaluation and Characterization of Klima [*Eyring et al.*, 2016]) contribution for this model consists of years 151-650 from the run for which the evolution of global mean sea surface temperature (SSTs) is displayed in Fig. 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 Fig. 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.

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 pre-industrial simulation is biased cold in globally averaged SSTs by roughly 0.6 K, comparing the 1880-1900 periods in Fig. 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 top-of-atmosphere energy balance was attempted to balance this cooling. As mentioned in *Zhao et al.* [2018a], the global mean



308 **Figure 1.** Global mean sea surface temperature (SST) using 5-year running averages from the CM4.0 pi-
 309 Control and from the Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST) [Rayner et al.,
 310 2003]. The horizontal black and red lines show the SST averages over pre-industrial period 1880-1900 for
 311 HadISST and over 281-300 for CM4.0 piControl, respectively, indicating about 0.62 K cold bias in the CM4.0
 312 piControl. The solid and dashed vertical lines are the 500 year extent of the CM4.0 piControl placed in the
 313 CMIP6 archive as well as the time span of the CM4.0 historical run for the DECK

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 the SST bias. (Observations for this early period are not adequate to quantify the cold bias in surface air temperature. As described below, the bias in recent decades of the historical simulation is sensitive to the time period chosen.)

Fig. 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, and that these fluctuations have relatively little effect on the extratropical Northern Hemisphere.

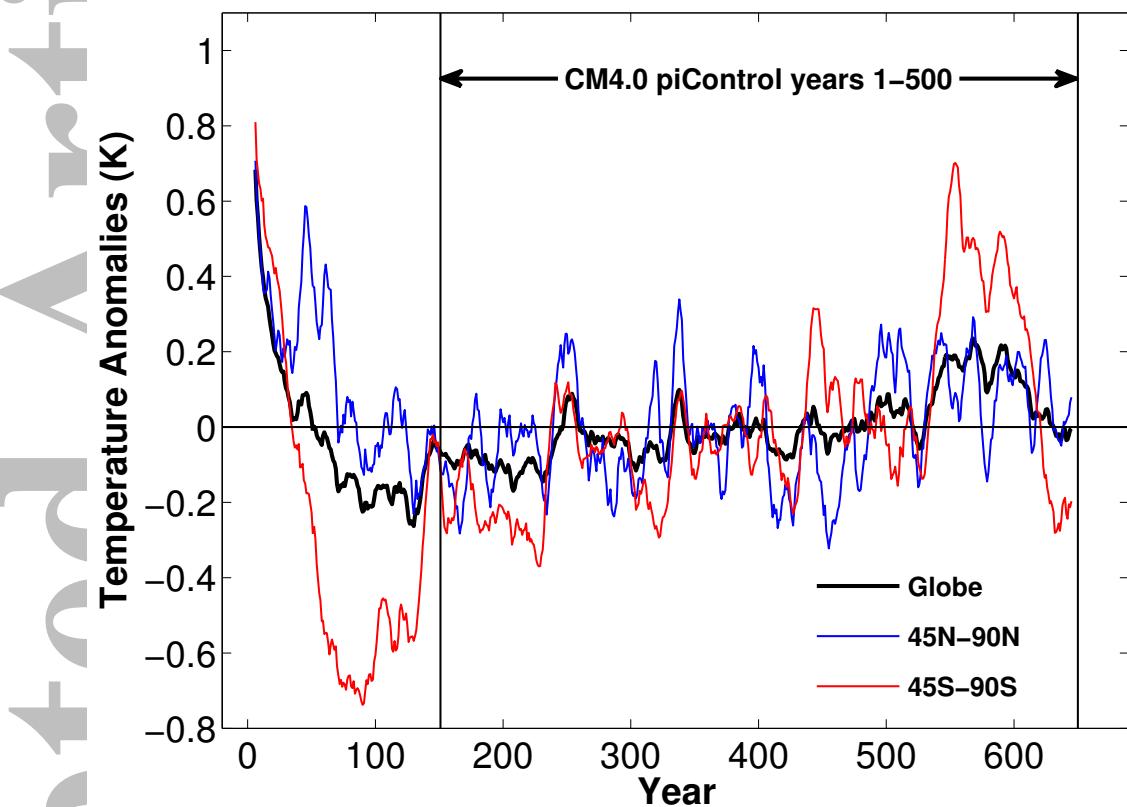


Figure 2. Surface air temperature anomalies using 11-year running averages in the CM4.0 piControl with respect to the average over piControl years 1-500 for the globe (black), north of 45°N (blue), and south of 45°S (red).

Information on the super-polynyas that are a central ingredient of this Southern Hemisphere variability are shown in Fig. 3. The time series of sea ice extent and maximum mixed layer depth 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 Fig. 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 (Fig. 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. Because none of our simulations are integrated

350 to equilibrium, we have no information on the variability in the final statistically steady state.
 351 A change in glacial albedos near the final stages of model development (see Sec. 3) had a
 352 significant effect on the characteristics of the model's polynyas. While CM4.0's polynyas are
 353 generated in the Ross Sea, before the albedo change the simulations were dominated by Wed-
 354 dell Sea polynyas of comparable or even larger amplitude. All observed open-ocean polynyas
 355 have occurred in the Weddell Sea. (For analysis of recent observations of off-shore Antarc-
 356 tic polynyas and their potential role in climate variability, see *Campbell et al.* [2019]). Once
 357 again, because of the limited length of our integrations, we are not confident that the effects
 358 of this albedo change on variability are indicative of the behavior in the final equilibrated
 359 state.

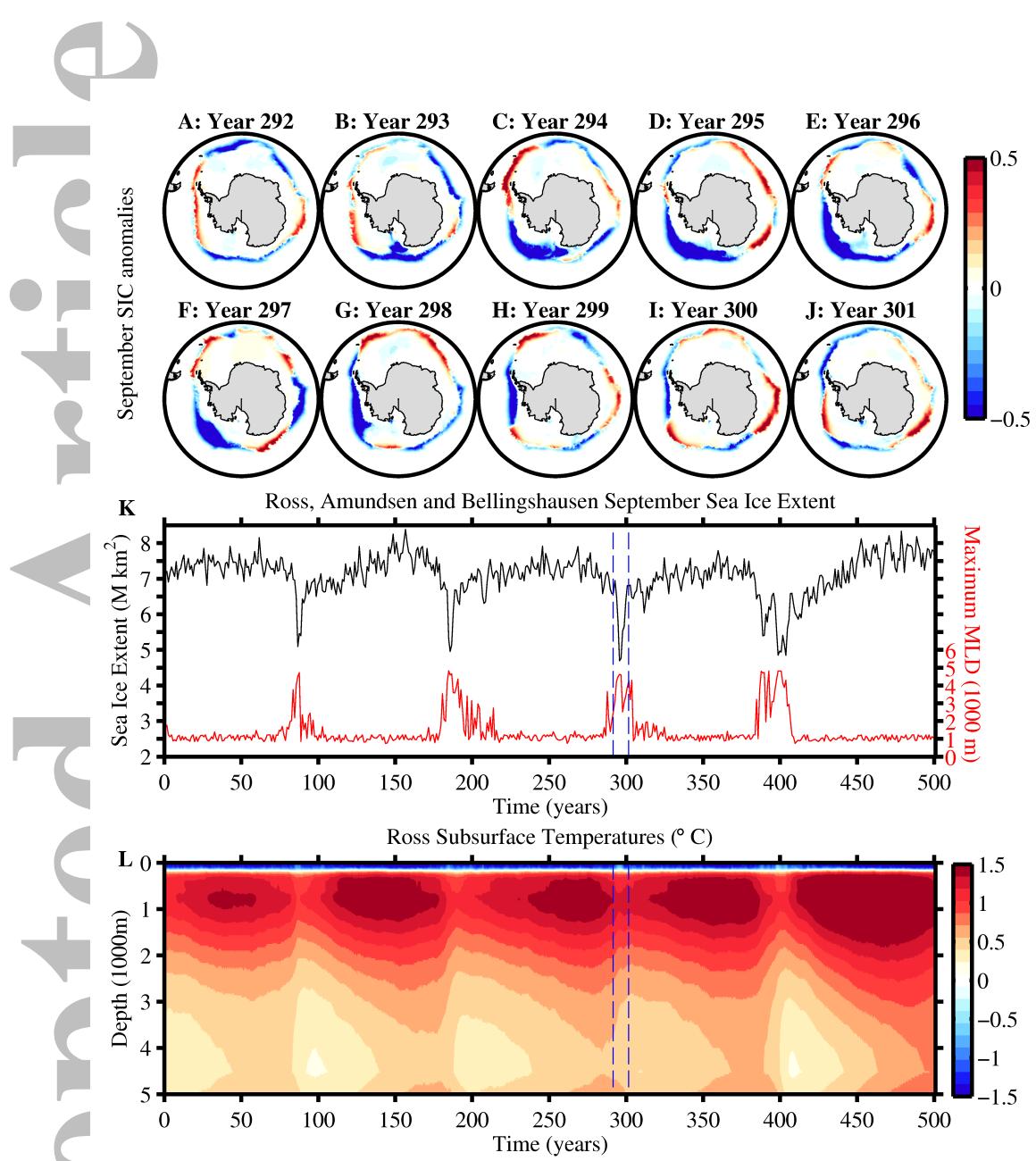
360 The net downward radiative imbalance at the top of the model atmosphere over the
 361 piControl simulation is provided in Fig. 4. Also shown are the net downward energy flux
 362 out of the atmosphere at the surface, computed from the atmospheric model, the difference
 363 between these two, as well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

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

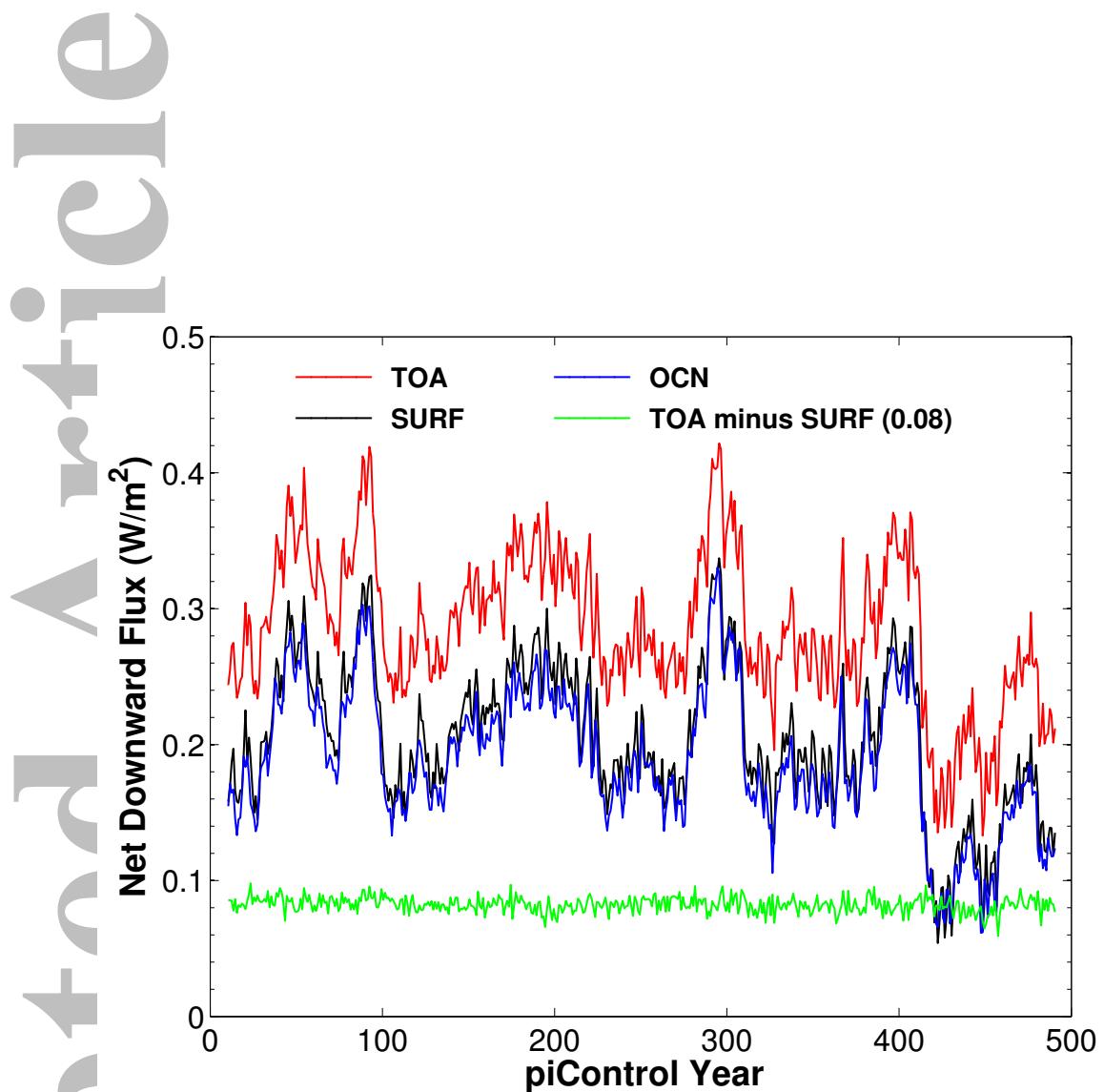
388 4.2 North Atlantic Ocean state and AMOC

389 The CM4.0 model has an improved simulation of some key characteristics of the North
 390 Atlantic Ocean circulation compared with previous GFDL models and many CMIP3 and
 391 CMIP5 climate models. For example, in the CM4.0 piControl simulation, the deep west-
 392 ern boundary current (DWBC) in the mid-low latitude North Atlantic has a realistic strength
 393 (Fig. 5) and coherency relative to the CM2.5 simulation of *Delworth et al.* [2012], which
 394 uses a similar ocean grid resolution as CM4.0 yet with a z-like vertical coordinate. Climate
 395 models with coarser resolution ocean components often have weak DWBC flow, with much
 396 of the deep return flow trapped unrealistically to the mid-ocean ridge [*Zhang and Vallis*,
 397 2007]. CM2.5 also has this weakness [*Zhang et al.*, 2011]. However, as shown in *Adcroft
 398 et al.* [2019], the hybrid vertical ocean coordinate employed in CM4.0, which reduces to
 399 an isopycnal coordinate in the low and midlatitude pycnocline and deeper ocean, reduces
 400 spurious numerical mixing. It also has an improved treatment of interaction of flow with to-
 401 topography due to its continuous representation of the bottom boundary. We conjecture that
 402 these properties of the vertical coordinate contribute to a deeper and more coherent DWBC
 403 in CM4.0 relative to CM2.5 and other comparable models [*Wang et al.*, 2015].

410 The simulated mean North Atlantic upper ocean circulation in CM4.0 also exhibits a
 411 more realistic Gulf Stream separation and path in the open ocean, as revealed by the com-
 412 parison of the observed mean dynamic topography with the simulated mean sea surface
 413 height (Fig. 6). This result is consistent with previous studies showing that a relatively strong
 414 DWBC and associated bottom vortex stretching keeps the Gulf Stream path downstream
 415 of Cape Hatteras separated from the North American coast [*Zhang and Vallis*, 2007; *Yea-*

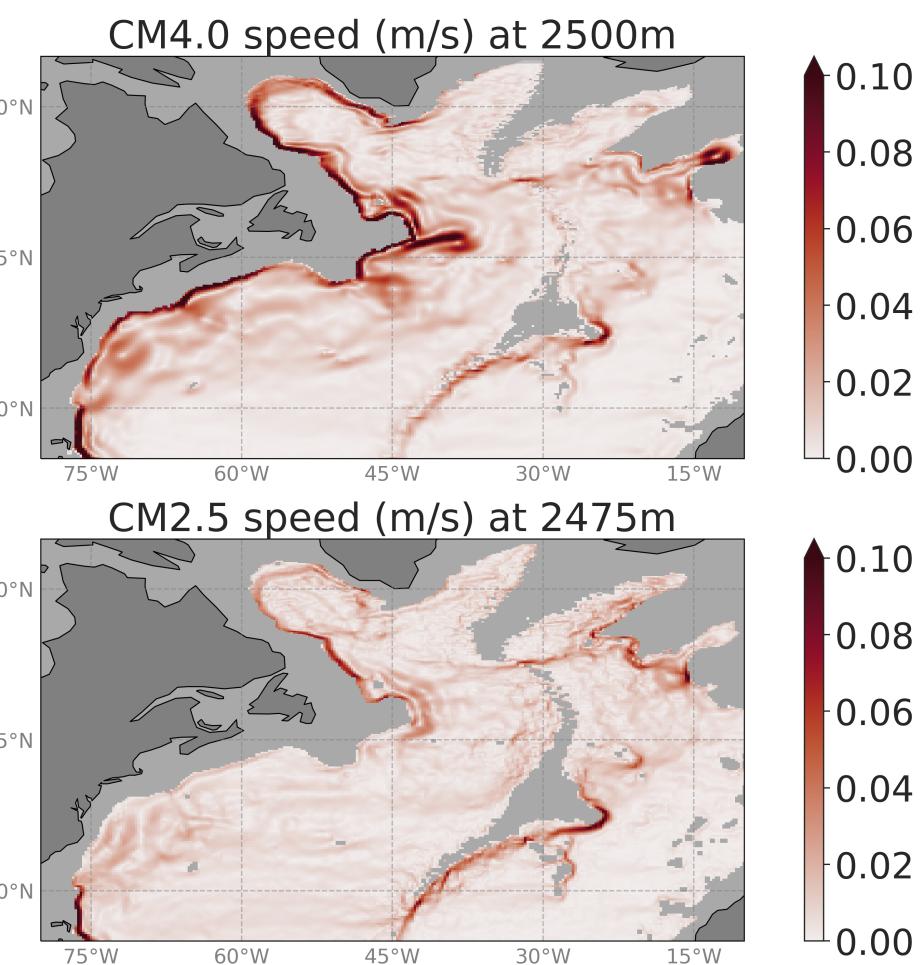


360 **Figure 3.** Ross Sea polynya events in the CM4.0 piControl experiment. Panels A–J: 10-year sequence of
 361 September sea-ice concentration (SIC) anomalies (relative to the model climatology) for the polynya event
 362 spanning years 292–301. Panel K: time series of September sea-ice extent and maximum mixed-layer depth
 363 computed over the Ross, Amundsen, and Bellingshausen sector (160–300°E). Panel L: Hovmöller plot of
 364 winter (JAS) Ross Sea (160–230°E and 60–90°S) ocean temperature. Dashed blue lines indicate the timing of
 365 the polynya event shown in panels A–J.



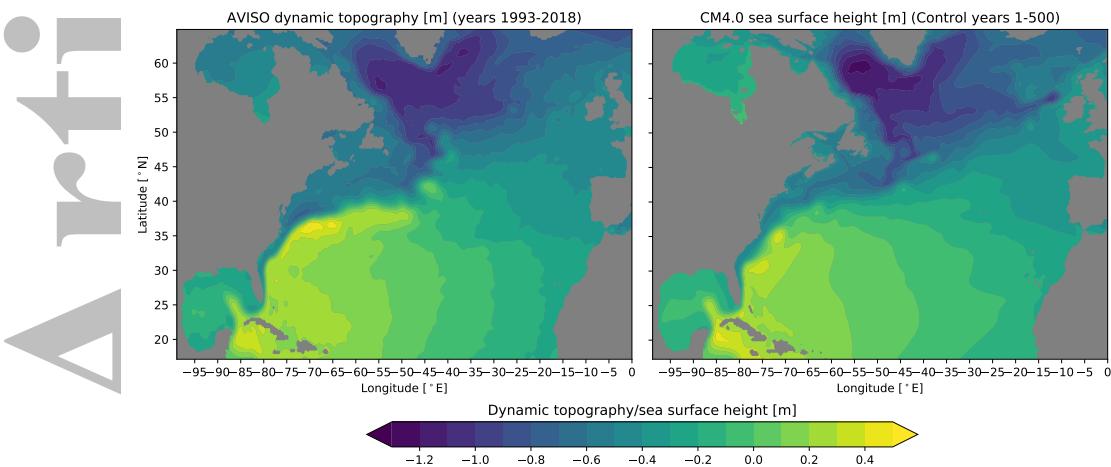
366 **Figure 4.** Globally averaged heat fluxes using 20-year running averages at top-of-atmosphere (TOA), sur-
367 face (SURF), ocean (OCN), and TOA minus surface (the green curve, representing an artificial energy sink
368 in the model atmosphere of about 0.08 W/m^2) in the CM4.0 piControl.

Accepted Article



404 **Figure 5.** Climatological North Atlantic ocean current speed at 2500 m for CM4.0 (upper panel) and
405 2475 m (closest model level to 2500 m) in CM2.5 (lower panel), computed using the 40-year time mean hor-
406 izontal velocity computed over years 161-200 from pre-industrial control simulations. The light gray is the
407 2500 m bottom topography. Note the stronger and more coherent boundary circulation in CM4.0, whereas
408 the CM2.5 current fades away south of the Grand Banks. This difference between the boundary currents is
409 reflected at all depths.

416 After passing the Grand Banks, a significant part of the simulated
 417 Gulf Stream in CM4.0 turns realistically toward the northeast as the North Atlantic Cur-
 418 rent. In contrast, in many previous GFDL models, such as ESM2M [Dunne *et al.*, 2012],
 419 CM3 [Griffies *et al.*, 2011], and CM2.5 [Delworth *et al.*, 2012], each with a weaker DWBC,
 420 the entire North Atlantic Current shifts unrealistically eastward, resulting in a severe cold
 421 SST bias east of Newfoundland [Zhang *et al.*, 2011; Talandier *et al.*, 2014]. Even with the
 422 improvements in CM4.0, there remains some poleward bias in the Gulf Stream path down-
 423 stream of Cape Hatteras (Fig. 6), yet there is only a relatively weak cold bias east of New-
 424 foundland, as can be seen below in Fig. 13.

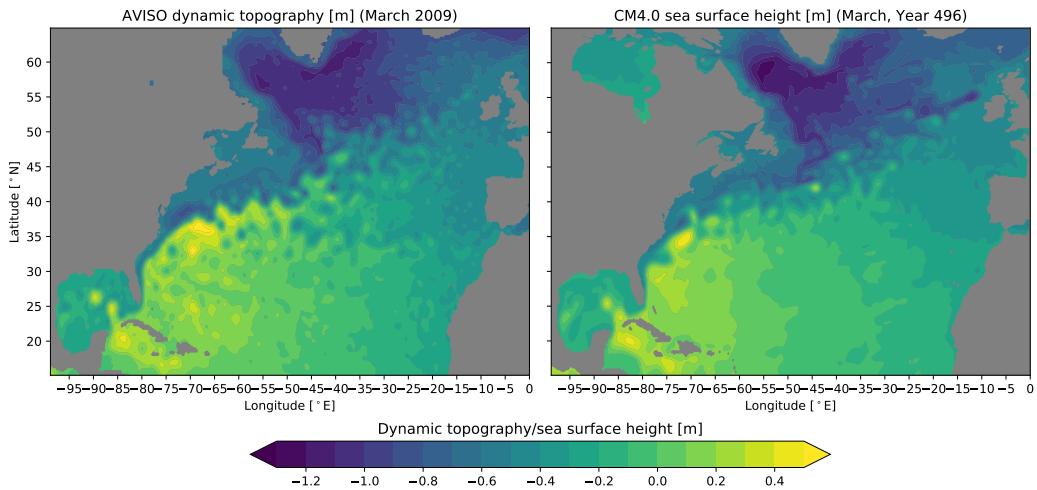


425 **Figure 6.** Time mean North Atlantic dynamic topography from the AVISO satellite analysis and the CM4.0
 426 simulation.

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

442 CM4.0's Atlantic meridional overturning circulation (AMOC), averaged over the pi-
 443 Control period, is displayed in Fig. 8. The maximum strength is ~16.7 Sv at 26°N, as cal-
 444 culated from the mean streamfunction integrated down from the surface. This result is very
 445 similar to the ~16.8 Sv measured by the RAPID array over the period 2004-2015 [Cunning-
 446 ham *et al.*, 2007; Smeed *et al.*, 2018]¹. The simulated AMOC penetration depth as measured
 447 by the zero crossing in the streamfunction is too shallow (~3000 m) compared with that di-

¹ As shown below in Fig. 23 as part of the discussion of the historical simulations, the simulated AMOC strength in the period of the RAPID measurements is ~1Sv lower than this mean piControl value.



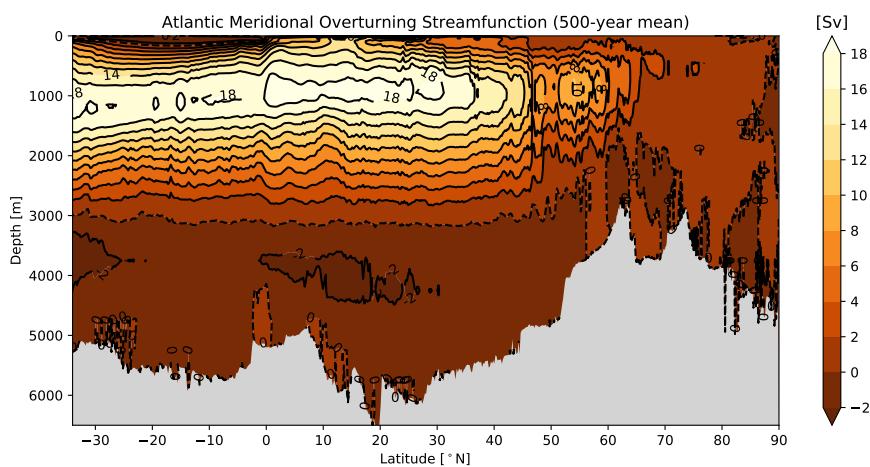
440 **Figure 7.** Monthly mean North Atlantic dynamic topography from the AVISO satellite analysis versus the
 441 CM4.0 simulation. We chose months with comparable realizations of the Gulf of Mexico loop current.

448 rectly observed (~ 4300 m) at 26°N [McCarthy *et al.*, 2015] (Fig. 9). Consistently, the mean
 449 piControl northward Atlantic heat transport at 26°N is ~ 1 PW, which is weaker than the di-
 450 rect estimate of ~ 1.2 PW for the period of 2004–2013 [Johns *et al.*, 2011]. The shallower
 451 than observed AMOC and weaker than observed northward Atlantic heat transport found in
 452 CM4.0 historical and piControl simulations, reflect a common deficiency in both ocean-only
 453 and coupled simulations of the deep-penetrating Nordic Sea overflow [Danabasoglu *et al.*,
 454 2010; Zhang *et al.*, 2011; Danabasoglu *et al.*, 2014; Wang *et al.*, 2015; Adcroft *et al.*, 2019].

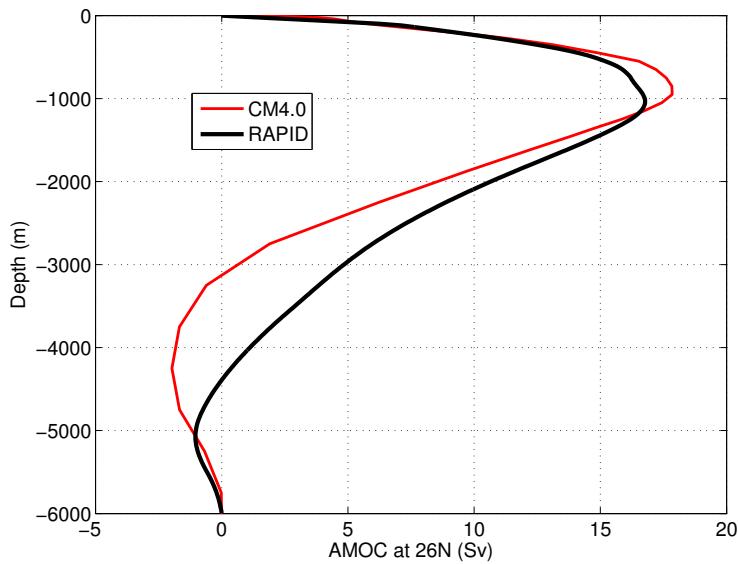
463 The time series in piControl of AMOC strength (anomalies from the 500 year mean)
 464 is provided in Fig. 10. The statistically stable behavior of the AMOC is consistent with
 465 the stability of the Northern extratropical temperatures seen earlier in Fig. 2. The multi-
 466 decadal AMOC variability in the CM4.0 piControl simulation has a modest amplitude of
 467 0.72 Sv (standard deviation of 10 year low pass filtered AMOC strength at 26°N), which is
 468 weaker than that simulated in previous GFDL models with coarser horizontal ocean resolu-
 469 tion (1.26 Sv in ESM2G, 1.09 Sv in ESM2M, 1.07 Sv in CM3) and is in the lower half of the
 470 spread among the CMIP5 models [Yan *et al.*, 2018]. It is also substantially weaker than the
 471 variability 1.4 Sv inferred indirectly from observations by Yan *et al.* [2018].

475 In piControl, the maximum mixed layer depth (MLD) in the Labrador Sea is very deep
 476 (more than 3000m), reaching the bottom in the central Labrador Sea basin, and with deep
 477 convection extending over a much broader area than that observed. This excessive mixed
 478 layer depth is consistent between OM4.0 and CM4.0 (Fig. 11) and reflects weak vertical
 479 stratification in the deep Labrador Sea. This weak stratification is associated with the ab-
 480 sence of a realistic dense Nordic Sea overflow that generally provides stratification to the
 481 deep Labrador Sea (e.g., Zhang *et al.* [2011]; Yeager and G.Danabasoglu [2012]). It is also
 482 associated with the absence of realistic high latitude mesoscale eddies (or a parameterization
 483 of their effects [Adcroft *et al.*, 2019]) that act to re-stratify convective regions (e.g., Marshall
 484 and Schott [1999]).

492 In contrast, the maximum MLD in the Greenland Sea in CM4.0 is very shallow (Fig.
 493 11). This shallow MLD in CM4.0 (lower left panel) is not seen in OM4.0 (upper right panel),
 494 though neither shows the region of maximum MLD consistent with that portrayed in the ob-
 495 servations. One possible reason for the lack of the Greenland Sea deep convection in CM4.0



455 **Figure 8.** Climatological residual mean Atlantic meridional overturning stream function from the CM4.0
 456 500-year piControl simulation. The stream function is integrated from the bottom, so the simulated AMOC
 457 strength relative to the surface is defined as the difference between the maximum value (near 1000m) and the
 458 surface value (on the order of 1Sv due to freshwater transport over the North Atlantic).



459 **Figure 9.** Vertical profiles of the time averaged circulation at 26°N from the RAPID observation (2004–
 460 2015) [Cunningham et al., 2007; McCarthy et al., 2015; Smeed et al., 2018] and from the CM4.0 500-year
 461 piControl simulation. Note here only the Eulerian mean component of the simulated AMOC is shown for the
 462 comparison with the RAPID data, similar to the approach used in Evans et al. [2017].

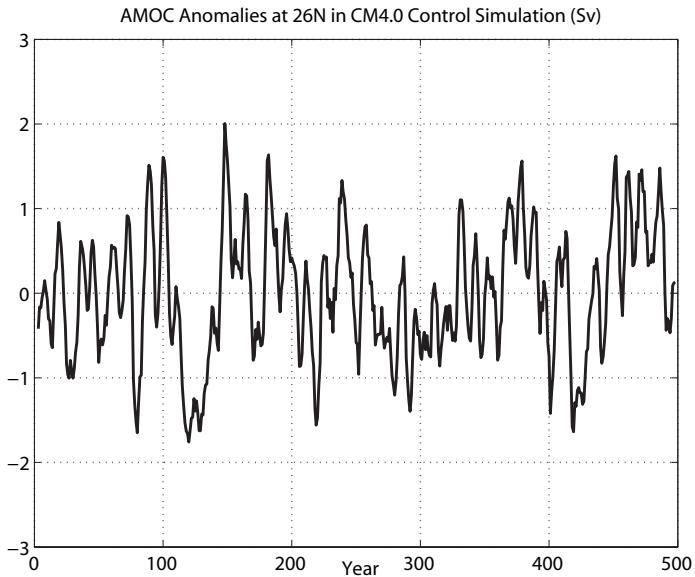


Figure 10. Time series of AMOC anomalies at 26°N from CM4.0 500-year piControl simulations (5-year smoothed), computed from the maximum streamfunction obtained by integrating down from the surface. The vertical range is set for comparison to the historical AMOC variability show in Fig. 23.

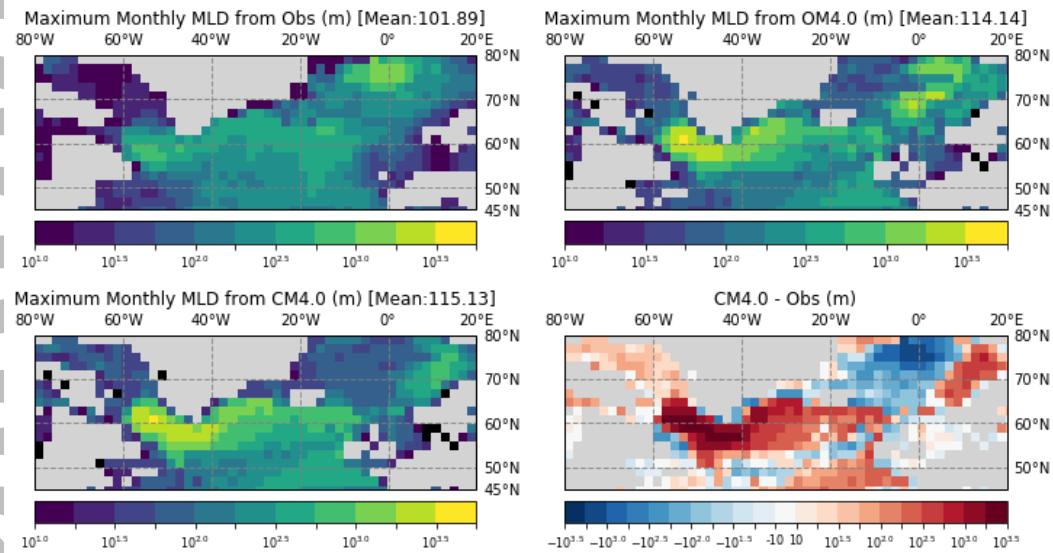


Figure 11. Climatological maximum of monthly mixed layer depth in the North Atlantic and Nordic Sea using a criterion of 0.03 kg m^{-3} for the density difference across the mixed layer. Upper left panel: observational estimates from *de Boyer Montégut et al.* [2004] updated to include observations through 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.

496 is that the warm salty upper North Atlantic water entering the Nordic Seas moves primarily
497 along the eastern boundary of the Norwegian Sea and is not mixed into the interior Green-
498 land Sea. We conjecture that the muted multidecadal AMOC variability in CM4.0 is very
499 likely linked to the very strong Labrador Sea deep convection (which, being so strong, is re-
500 sistant to change), the muted Nordic Sea overflow, and the lack of Greenland Sea deep con-
501 vection.

502 All of these North Atlantic modeling biases (the excessive maximum MLD in the
503 Labrador Sea, the lack of the Greenland Sea deep convection, weak Nordic Sea overflow and
504 shallow AMOC, and weak multidecadal AMOC variability) are also present in the histori-
505 cal integrations with CM4.0. They are also present in CM2.5 simulations, a model with the
506 same horizontal resolution as CM4.0 [Zhang *et al.*, 2011; Delworth *et al.*, 2012]. In a per-
507 turbed experiment using CM2.5, a deepened bathymetry south of the Denmark Strait leads to
508 a stronger and deeper-penetrating Nordic Sea overflow, and thus deeper AMOC, stronger
509 northward Atlantic heat transport, and reduced Labrador Sea MLD [Zhang *et al.*, 2011].
510 Although the Nordic Sea overflow instantaneously increases after the artificial bathymetry
511 deepening, it gradually weakens afterwards due to the lack of sustained Greenland Sea deep
512 convection in CM2.5 [Zhang *et al.*, 2011]. These long-standing biases, especially in models
513 with a relatively high horizontal ocean resolution, are important areas to focus on for future
514 model development.

515 **5 Historical simulation (1850-2014)**516 **5.1 Temperature evolution**

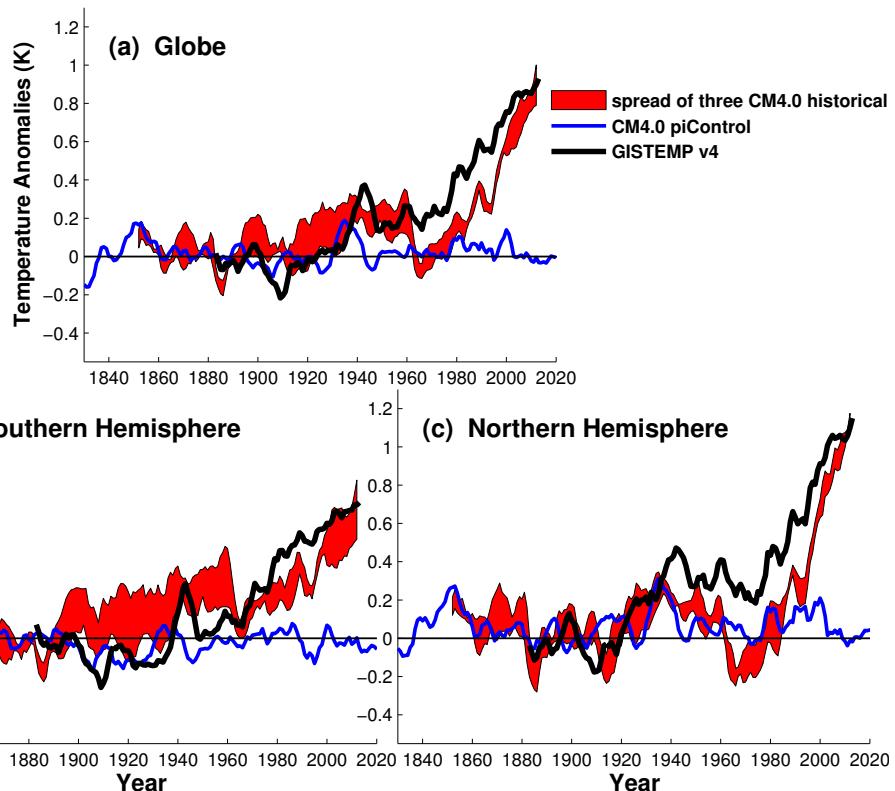
517 The historical simulations that have been branched off the CM4.0 piControl run utilize
 518 the identical atmospheric and ocean models, but the land model has additional tiles in each
 519 grid box representing secondary growth of forest. The land in the historical run was initial-
 520 ized by first executing a branch from piControl of several centuries duration with potential
 521 vegetation (no land use), followed by a bridge run from 1750 to 1849 with full land use tran-
 522 sitions to allow the forests to adjust. (The piControl run includes an estimate of mid-19th
 523 century land use but no land use transitions.) The ocean, sea ice, and atmosphere initial con-
 524 ditions were then reset to the conditions at the branch-off time of piControl (e.g., piControl
 525 year 101) to minimize initial differences between piControl and the historical run due to the
 526 ocean and sea ice states. In addition to land use, the historical simulations include the time
 527 evolving greenhouse gas concentrations (including ozone for radiation calculations), aerosol
 528 precursor emissions and solar irradiance consistent with CMIP6 specifications (documented
 529 at <http://goo.gl/r8up31>). For chemistry, time-varying ozone and other oxidant concentrations,
 530 such as OH, are prescribed based on the historical simulations with CM3 conducted in sup-
 531 port of CMIP5 [John *et al.*, 2012].

532 To assess robustness in the results, we analyze three ensemble members of the histor-
 533 ical simulations. They are identical except for the times at which the ocean, sea ice, and at-
 534 mosphere initial conditions are spun off the piControl simulation (i.e., piControl years 101,
 535 140, and 182). It is noted that the uneven spacing of start times was the result of technical
 536 problems with availability of restarts.

537 The time evolution of global mean surface air temperature, and Northern Hemisphere
 538 (NH) and Southern Hemisphere (SH) means, in these three realizations are displayed in
 539 Fig. 12. While the overall warming from the late 19th century to 2014 agrees quite well
 540 with observations, there are strong hints from the NH time series that the balance of fac-
 541 tors generating this overall temperature change has unrealistic features. The NH evolution
 542 reflects the behavior commonly seen in models with relatively strong aerosol forcing and
 543 relatively high transient climate sensitivity: the NH aerosol forcing is able to counteract the
 544 greenhouse warming until about 1990, after which the reduction in aerosol forcing and the
 545 rapid increase in greenhouse forcing combine to produce rapid warming. While we cannot
 546 be definitive based on such a small ensemble, the model's modest NH multi-decadal vari-
 547 ability implies that the odds of consistency with observations in the NH is low. The early
 548 20th century warming in the global mean emanates primarily from the SH, unlike the obser-
 549 vations. The model's large SH multi-decadal variability is once again due to quasi-periodic
 550 super-polynyas in the subpolar SH oceans.

551 As analyzed in an upcoming paper on the climate sensitivity of CM4.0, the model's
 552 transient climate response (TCR) is roughly 2.05K. TCR is the globally averaged tempera-
 553 ture increase at the time of CO₂ doubling (about 70 years) in idealized experiments where
 554 CO₂ is increased at 1% per year. Sometimes, it is estimated from the warming averaged over
 555 years 60-80. CM4.0's effective sensitivity is 3.2K, computed from the TCR and the TOA en-
 556 ergy imbalance, H , at the time of doubling in the 1%/year simulation ($T_{eff} = TCR * F / (F - H)$), where F is the estimated radiative forcing for doubling of CO₂). Our estimate of equi-
 557 librium sensitivity is about 5.0K, obtained from an extended abrupt CO₂ quadrupling sim-
 558 ulation, extrapolating on a *Gregory et al.* [2004] plot using years 51-300. Further research
 559 is required, but our working hypothesis is that middle and high latitude cloud feedbacks in-
 560 teracting with snow and sea ice feedbacks are the primary cause of this high sensitivity. In
 561 particular, the large difference between the model's effective and equilibrium sensitivities is
 562 due to a high heat-uptake efficacy, as discussed in an upcoming paper on the climate sensi-
 563 tivity of CM4.0. Given that much of the heat uptake is localized in subpolar latitudes, heat
 564 uptake efficacy is expected to be sensitive to high latitude feedbacks.

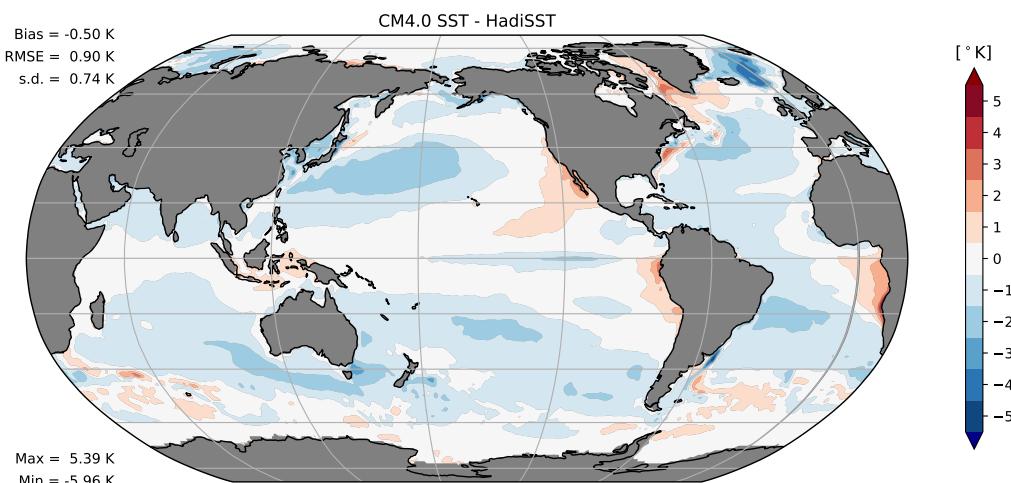
Article
Accepted
for
Publication



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

571 **5.2 Mean climate biases**

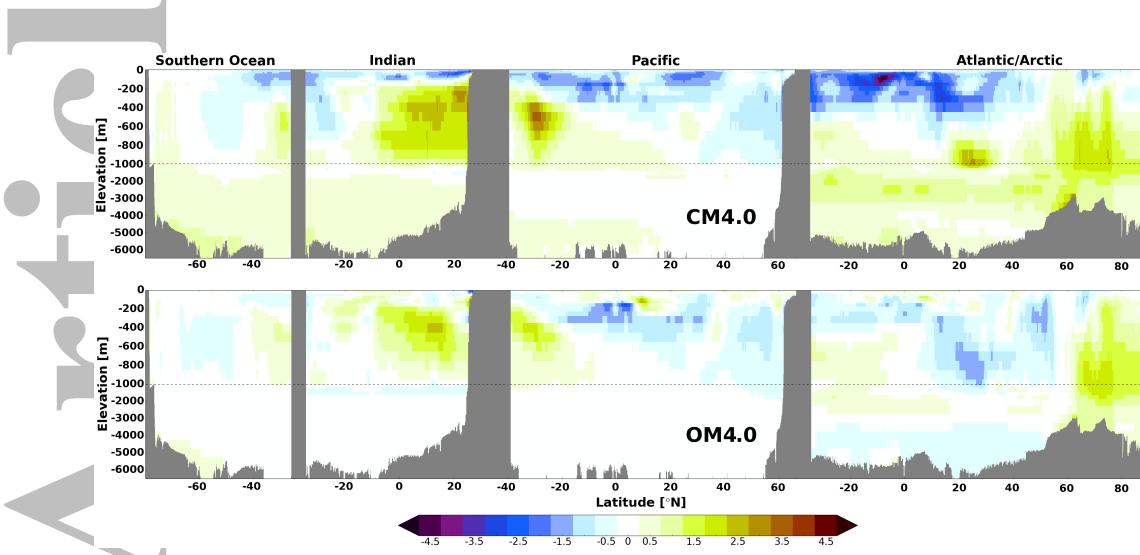
572 The SST bias pattern averaged over the three historical runs and over the last 10 years
 573 of each run (2005-2014), is displayed in Fig.13, computing the bias from observations over
 574 this same period. Compared to many previous GFDL models, in which the bias is often
 575 of the opposite signs in the two subpolar regions, here the bias is relatively balanced be-
 576 tween the two hemispheres (Admittedly, this result needs to be tempered by the fact that the
 577 Southern Ocean is far from fully equilibrated.) The equatorial cold bias in the Pacific is still
 578 present but muted compared to many other models. The overall cold bias is concentrated in
 579 the centers and the poleward margins of the subtropical highs, with warm anomalies marking
 580 coastal subtropical upwelling regions. The root mean square (RMS) bias is 0.90 K, averaged
 581 over the three historical simulations. Removing the global mean bias, the standard deviation
 582 is 0.74 K. The global mean and therefore the RMS bias is sensitive to the time period chosen
 583 for this comparison because the mean cold bias decreases in time over this period, as seen
 584 in Fig. 12; the standard deviation is somewhat less sensitive to the averaging period. Given
 585 the larger warming trend near the end of the historical runs as compared to observations, this
 586 sensitivity to averaging period needs to be kept in mind when using these metrics to evaluate
 587 the model.



588 **Figure 13.** Sea surface temperature bias ($^{\circ}\text{K}$) in CM4.0, averaged over three CM4.0 historical simulations
 589 and the years 2005-2014, compared to the HadISST dataset [Rayner et al., 2003] for the same period.

590 Fig. 14 compares the zonal mean ocean temperature biases by basin at present-day
 591 in the CM4.0 historical experiment with those of a reanalysis-forced experiment using only
 592 its ocean/sea ice components, OM4.0 [Adcroft et al., 2019]. Although the ocean/ice states
 593 of CM4.0 and OM4.0 are initialized identically, at present-day the CM4.0 and OM4.0 runs
 594 are roughly 400 years and 300 years from initialization, respectively, so the comparison of
 595 drift must be qualitative. The difference in ocean surface forcing between the coupled and
 596 reanalysis-forced runs appears to account for CM4.0's Antarctic bottom water layer warm
 597 bias, which has been eliminated in the OM4.0 run; although surface salinity restoring to cli-
 598 matology, applied in OM4.0, makes these differences difficult to interpret. CM4.0's North
 599 Atlantic Deepwater warm bias is also substantially reduced with reanalysis forcing. The sur-
 600 face layer is much warmer and closer to observations in the reanalysis-forced run, a direct
 601 impact of the strong coupling to the imposed, observed, surface air temperatures. Below the
 602 surface layer and above about 1 km depth, the two runs have strikingly similar biases includ-
 603 ing: (1) a cold bias in the Southern Ocean, (2) a warm bias in the Indian Ocean except in
 604 its southernmost part, (3) a pattern of subsurface warm bias in the southernmost Pacific ac-

605 accompanied by a surface cold bias that expands downward moving north, finally reaching 1
 606 km depth at the northern boundary, (4) a cold-over-warm bias in the South Atlantic, and (5)
 607 a warm bias in the Nordic Seas and Arctic Ocean. To the extent that the reanalysis-based
 608 surface forcing is reliable, this similarity suggests that a significant component of these mid-
 609 depth biases in CM4.0 have their origin in the ocean component rather than in biases in sur-
 610 face forcing.

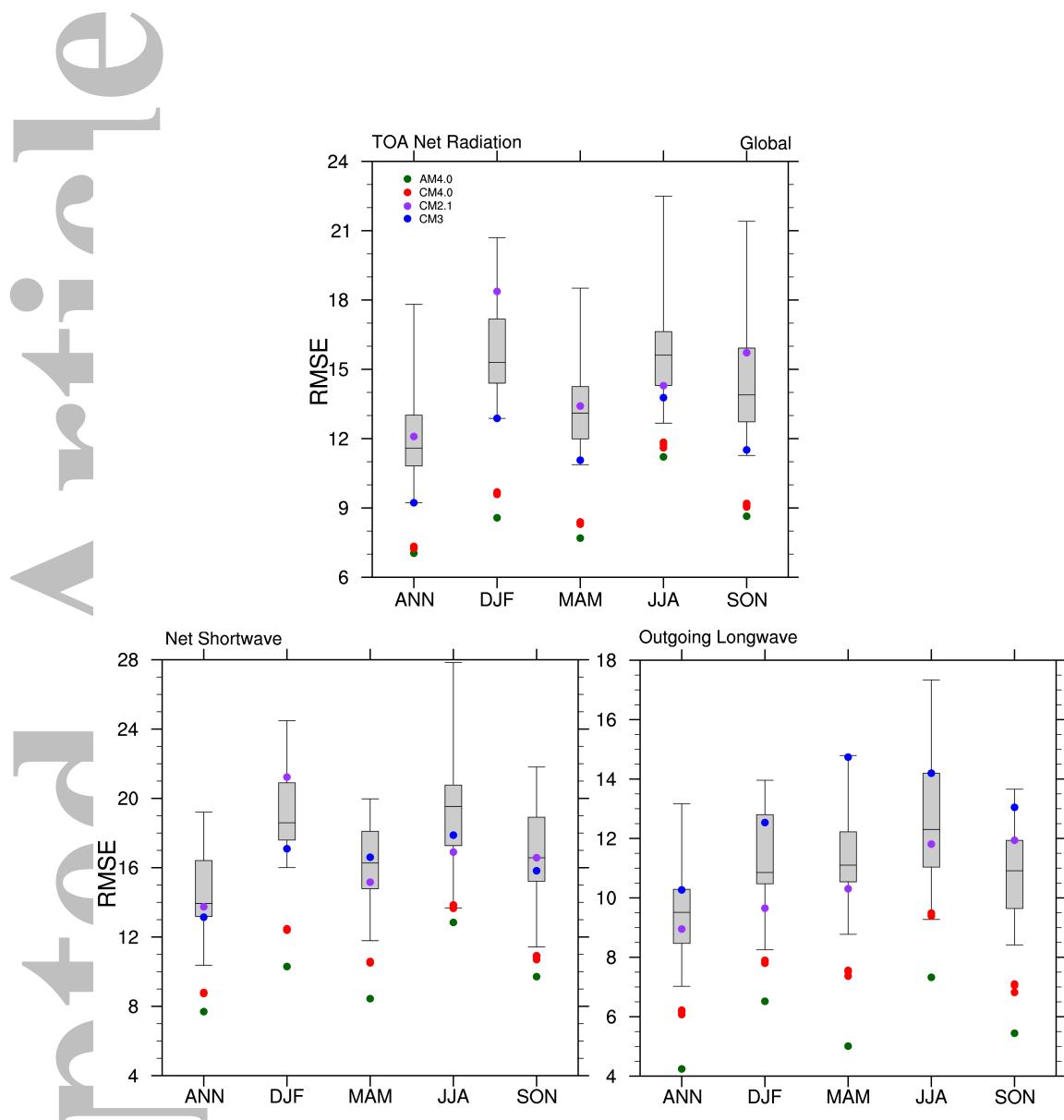


611 **Figure 14.** Zonal average ocean temperature biases relative to World Ocean Atlas 2005 for CM4.0 (upper
 612 row, 1990–2009 average) and OM4.0 (lower row, 1988–2007 average) historical experiments.

613 As described in *Zhao et al. [2018a]*, the strategy in the development of the atmospheric
 614 model in AMIP mode was to focus intently on minimizing biases in the top-of-atmosphere
 615 (TOA) energy fluxes, not simply in global means but in the spatial patterns, the result be-
 616 ing root mean square (RMS) biases in all seasons that are smaller than in any AMIP simu-
 617 lations in the CMIP5 archive. It is encouraging that CM4.0 maintains the quality of these
 618 fluxes. Fig. 15 shows the RMS biases in the net TOA fluxes, for the average over time period
 619 1980–2014, comparing against the coupled models in the CMIP5 database. CM4.0 has the
 620 smallest RMS bias in all seasons and the global mean. Also shown is the AMIP simulation
 621 with AM4.0 atmosphere, quantifying the modest degradation in this field upon coupling. The
 622 very small spread among the 3-member CM4.0 ensemble supports the adequacy of a small
 623 ensemble (or a single run) for this comparison.

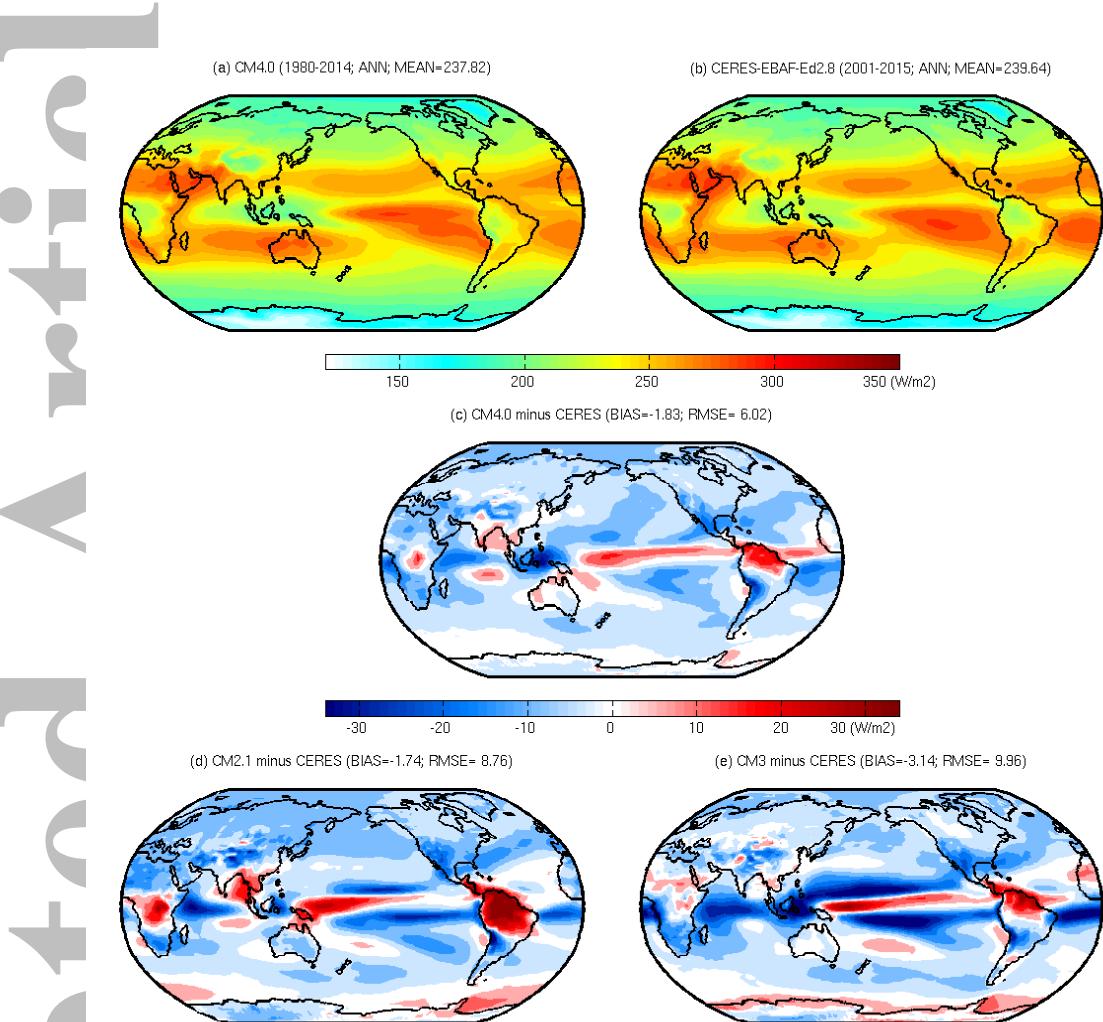
624 Analogous figures for longwave and shortwave TOA fluxes separately are also pro-
 625 vided in Fig. 15, which indicates that this quality of agreement extends to these individual
 626 components of the net flux. The longwave flux has a stronger effect on the quality of the sim-
 627 ulation in AMIP mode than does the shortwave, affecting the atmospheric heating directly,
 628 but the spatial pattern of the shortwave flux is particularly important to ensure that tuning in
 629 AMIP mode is consistent with small drifts when coupled.

636 The pattern of biases in outgoing longwave and net absorbed shortwave in CM4.0 are
 637 compared with CM2.1 and CM3 in Figs. 16 and 17. The improvement in the shortwave over
 638 the Southern Ocean is the most dramatic regional improvement. The equatorial Pacific bi-
 639 ases still have a similar structure but are muted compared to previous GFDL models. The
 640 shortwave absorption bias is relatively small over the Amazon in CM4.0, while the outgo-
 641 ing longwave radiation (OLR) is biased high. Since the Amazon has a dry bias (Fig. 20), the
 642 shortwave properties of the cloud field produced per unit precipitation are actually deficient
 643 in this region. Overall, low shortwave and longwave biases away from tropical convection
 644 are still present, as in previous models, but with reduced amplitude. The improvement in the



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

645 TOA flux pattern, which is quite large in this global RMS measure, does not appear to have
 646 a single explanation. Rather, we conjecture that it is a result of systematic effort while develop-
 647 ing the atmospheric model to maintain a realistic TOA radiation field while other aspects
 648 of the model, including the boundary layer and convection modules, were adjusted for other
 649 reasons.

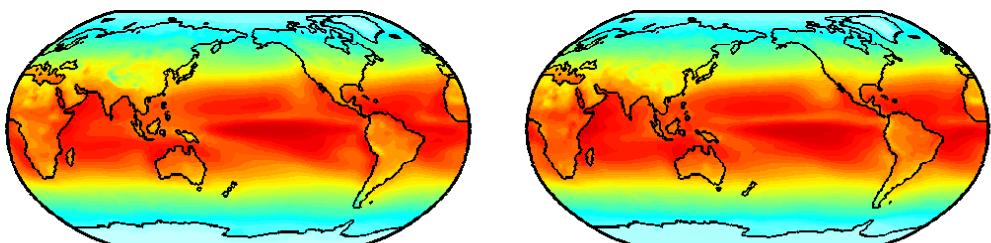


650 **Figure 16.** Annual mean outgoing longwave radiation, (a) CM4.0 averaged over 3 historical runs; (b)
 651 CERES EBAF Ed2.8; (c) CM4.0 minus CERES EBAF Ed2.8; (d) as in (c) but for CM2.1; (e) same as (c)
 652 but for CM3. Note that the time periods differ for the model and observational data, due to unavailability of
 653 CERES EBAF data prior to March 2000, but the specific averaging period is unimportant for these climato-
 654 logical averages.

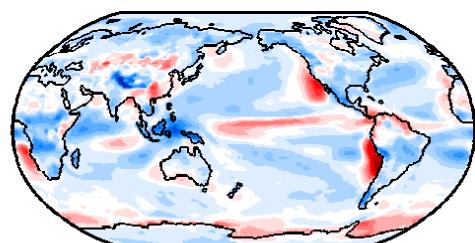
656 We suggest that the high quality TOA flux simulation is a key ingredient in maintaining
 657 a high quality precipitation simulation in the coupled model. Fig. 18 is a box-whisker plot
 658 for precipitation analogous to that for the TOA fluxes in Fig. 15. While the improvement
 659 over the CMIP5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

Accepted Article

(a) CM4.0 (1980-2014; ANN; MEAN=238.54) (b) CERES-EBAF-Ed2.8 (2001-2015; ANN; MEAN=240.44)



(c) CM4.0 minus CERES (BIAS=-1.90; RMSE= 8.52)



(d) CM2.1 minus CERES (BIAS=-2.01; RMSE=14.19)

(e) CM3 minus CERES (BIAS=-2.88; RMSE=12.89)

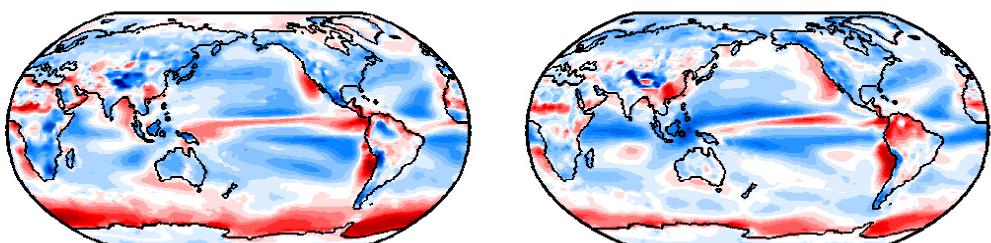
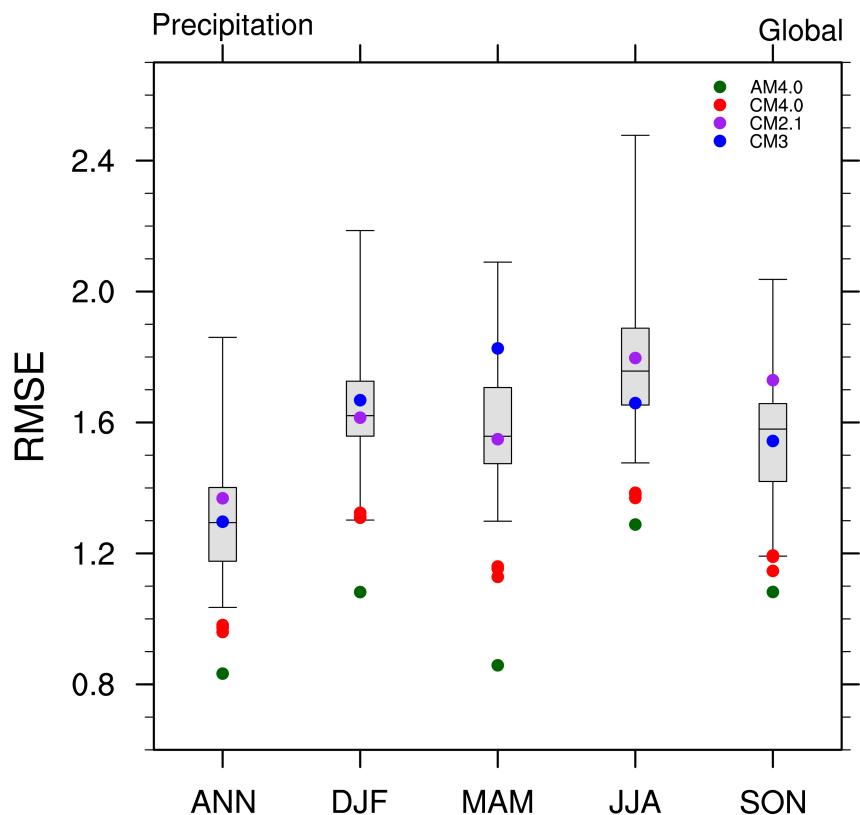


Figure 17. As in 16 but for the TOA net shortwave flux

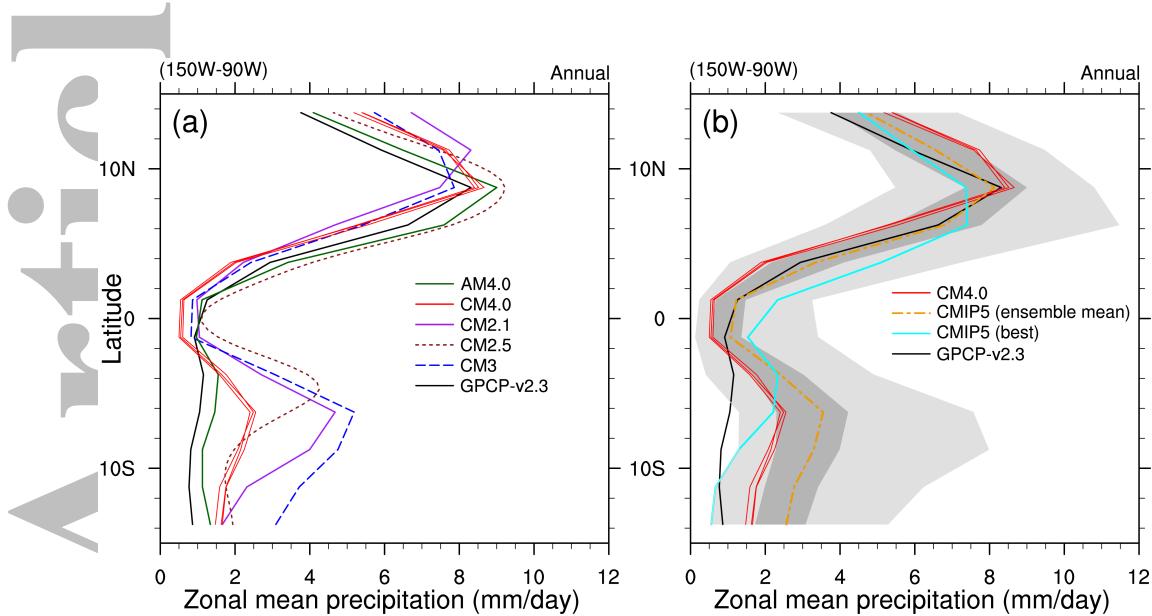
655

Accepted Article



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

669 This point is made more explicitly in Fig. 19 which shows the latitudinal structure of
 670 annual mean precipitation in the eastern equatorial Pacific. The AM4.0 AMIP simulation is
 671 close to the observations, with little hint of a double ITCZ bias by this measure. The three
 672 CM4.0 runs show modest deterioration, lying in the least biased quantile of the CMP5 distri-
 673 bution, and superior than previous versions of GFDL models (purple, brown, and blue).

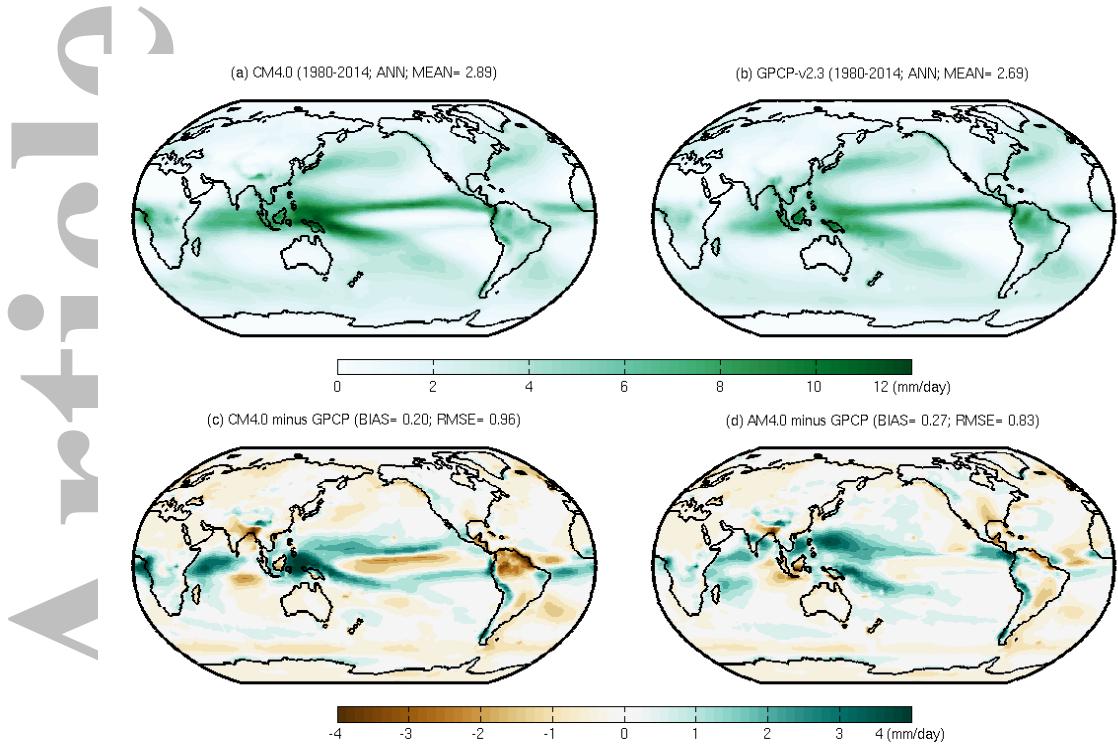


674 **Figure 19.** Annual zonal mean precipitation in the tropical Eastern Pacific averaged over longitudes 150W-
 675 90W as a function of latitude. Observations (GPCP v2.3), AM4.0, three CM4.0 realizations averaged over the
 676 years 1980-2014, and three earlier GFDL models, CM2.1, CM2.5, and CM3 are shown in (a); and the CMIP5
 677 ensemble mean, the best CMIP5 model (here referring to the CMIP5 model with smallest root mean square
 678 error over the region [14S:14N, 150W-90W], which is MICRO-ESM-CHEM model shown in cyan [Watanabe
 679 *et al.*, 2011]), the full spread and the 25th to 75th percentile in CMIP5 coupled models in light and dark grey
 680 in (b).

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

692 Understanding these differences between AMIP and coupled runs, due in part to the
 693 SST biases that develop, but also potentially in part due to the effects of coupling on con-
 694vection and transients disturbances, present important challenges, each in its own right. For
 695 example, exploratory simulations aimed at understanding the dry Amazon bias in CM4.0 in-
 696 dicate that the coupled model SST biases in the Atlantic are a main source of this dry bias. In
 697 a preliminary version of the coupled model that was part of the development process, annual
 698 mean Amazon rainfall is roughly 40% lower than in the corresponding AMIP simulation,

699 while in a coupled model in which the SSTs in the tropical Atlantic are constrained by obser-
 700 vations, the underestimate is reduced to roughly 15%.



701 **Figure 20.** Annual mean precipitation bias (mm/day) for three-member ensemble mean of CM4.0. Obser-
 702 vations (GPCP v2.3), CM4.0, and AM4.0 are averaged over years 1980-2014.

703 The annual and zonal mean temperature biases in the coupled and AMIP simulations
 704 in the latitude-pressure plane are shown in Fig. 21. In the AMIP simulation, there is a weak
 705 cold bias, less than or of the order of 1K, but with hardly any of the cold bias near and im-
 706 mediately above the extratropical tropopause seen in many other models. However, there
 707 is a larger cold bias below the tropical tropopause. The sensitivity of this feature to model-
 708 ing choices in the convection scheme is described by *Zhao et al.* [2018b]. When coupled,
 709 the cold bias increases by roughly a factor of two, in large part because the coupled model's
 710 tropical SSTs are biased cold, but also due to an equatorward contraction of the midlatitude
 711 circulation. The latter is likely in part due to the tropical cold bias itself and in part to the
 712 pattern of extratropical SST biases.

713 Consistent with the development of these biases in temperature gradient within the
 714 troposphere, significant zonal wind biases develop in CM4.0 that are not present in the corre-
 715 sponding AMIP simulation (Fig. 22). Most noticeable is an equatorward shift in the jets, in
 716 thermal wind balance with the temperature biases, but which are also accompanied by a sur-
 717 face wind signature, most clearly in the Southern Hemisphere, with magnitude roughly 1 m/s
 718 near the surface. The bias towards too strong trade winds in AM4.0 is present in CM4.0 as
 719 well.

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

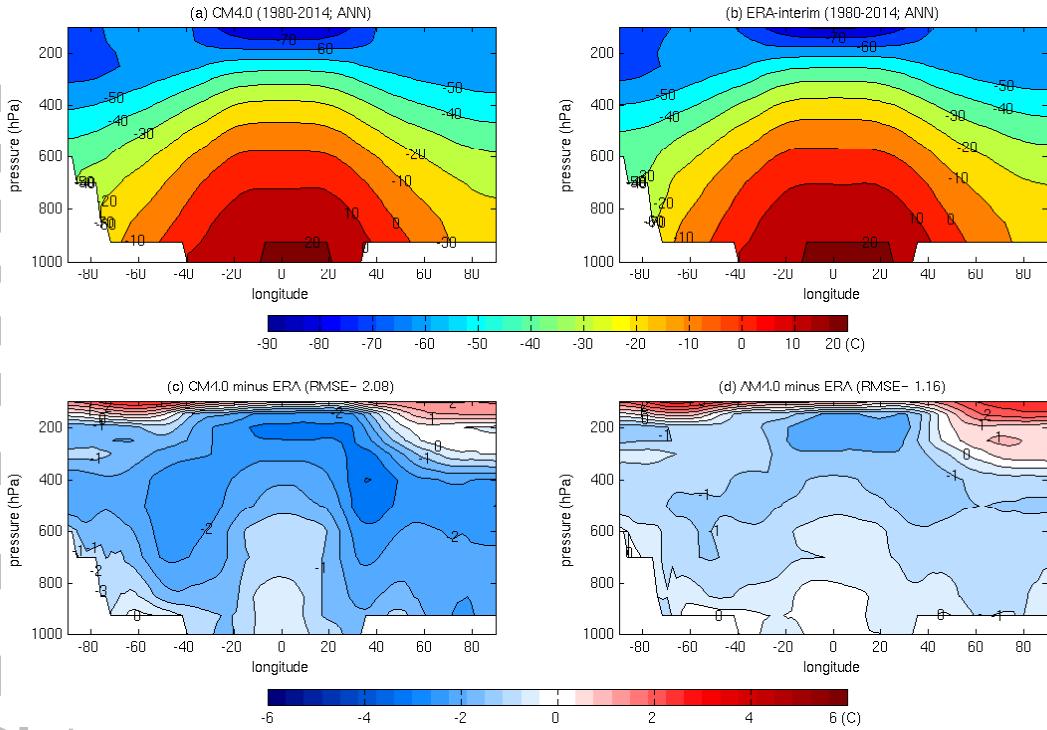


Figure 21. Latitude-pressure plots of annual zonal mean temperature ($^{\circ}\text{C}$) for (a) three-member ensemble mean of CM4.0, (b) ERA reanalysis [Dee et al., 2011], (c) difference between CM4.0 and ERA, and (d) difference between AM4.0 for 1980-2014 and ERA.

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 warmings, as well as 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

Fig. 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 (Fig. 10). We thus conclude that the model's AMOC evolution in the historical period is predominately forced. Based on previous studies (e.g. Delworth and 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 Northern Hemisphere mean surface temperature. That is, aerosol cooling prevents but does not reverse the greenhouse gas induced warming (Fig. 12) whereas it does reverse the greenhouse gas induced AMOC weakening.

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

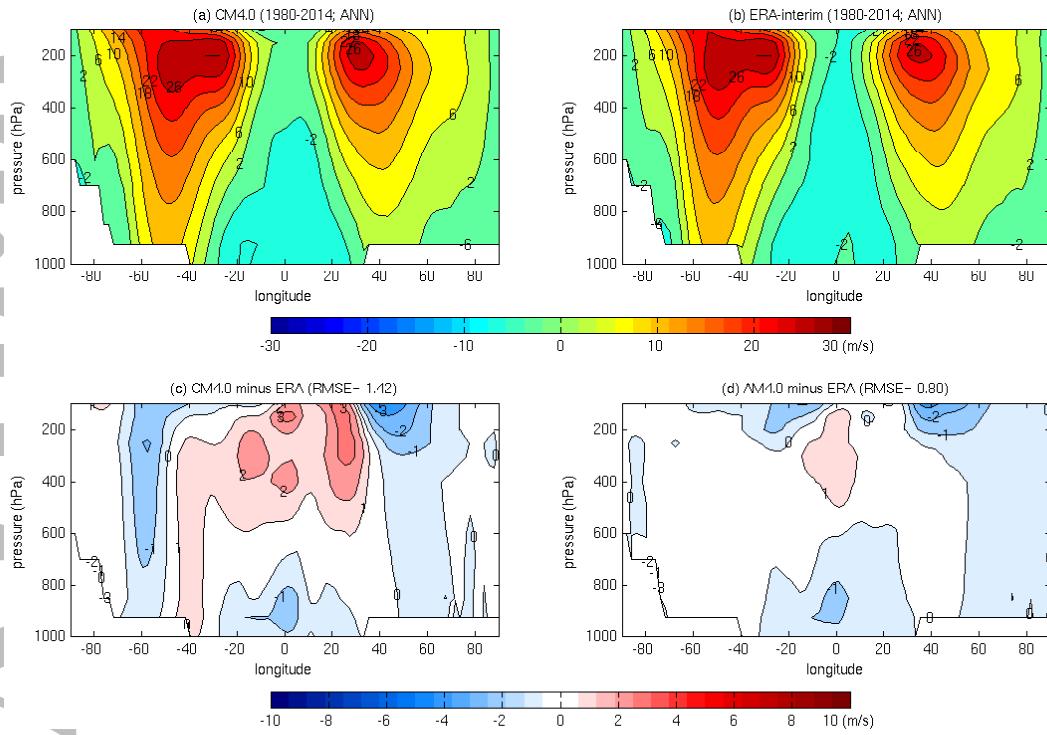


Figure 22. As in 21 but for the annual zonal mean wind.

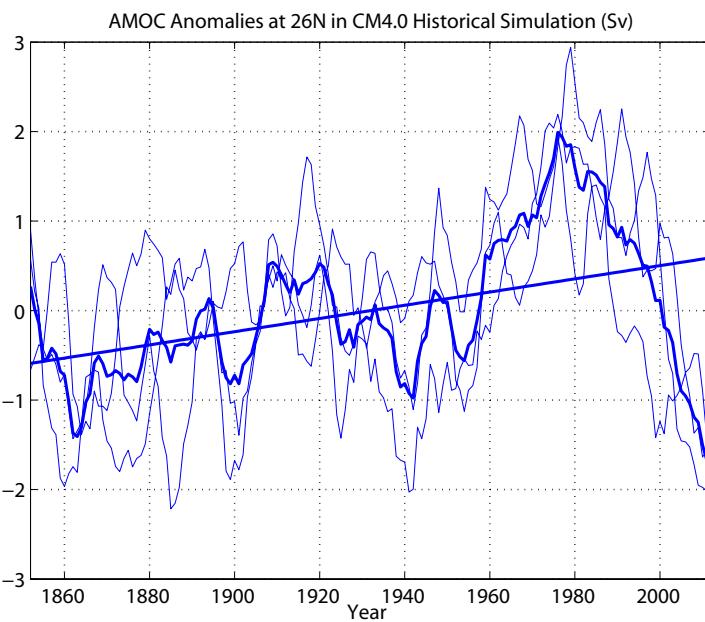
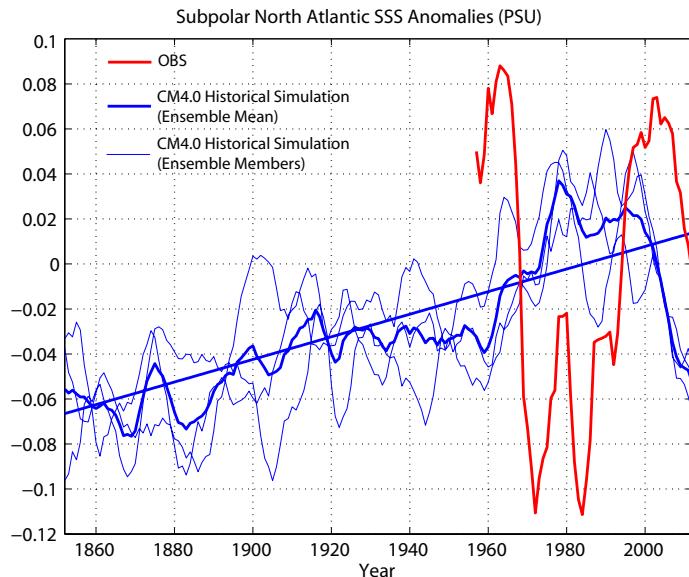


Figure 23. Time series of AMOC anomalies at 26°N from CM4.0 historical simulations (5-year smoothed). Solid blue lines: ensemble mean anomalies and long-term trend. Thin blue lines: anomalies from three individual ensemble members. (Recall that the three ensemble members use initial conditions from the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



778 **Figure 24.** Time series of the sea surface salinity (SSS) anomalies averaged over the subpolar North
 779 Atlantic domain (60°W - 0°E , 50°N - 65°N) from CM4.0 historical simulations and observations (5-year
 780 smoothed). Solid blue lines: ensemble mean anomalies and long-term trend. Thin blue lines: anomalies
 781 from three individual ensemble members. Red solid line: observed anomalies. The observed SSS data are
 782 5-year averages of the objectively analyzed ocean salinity anomalies since 1957 in Boyer *et al.* [2005]. Both
 783 observed and simulated SSS anomalies are relative to the 1957-2012 mean.

756 region (Fig. 24). Here, we also see an increasing salinity trend from 1940 to 1980 and a
 757 decreasing trend from 1980 to present, with a peak around 1980. A close linkage between
 758 forced AMOC changes and subpolar North Atlantic SSS has been found previously in CMIP
 759 models and has been attributed to anthropogenic aerosols. That is, increasing anthropogenic
 760 aerosol precursor emissions lead to the positive trends in subpolar North Atlantic SSS, and
 761 vice versa. This salinity response is most evident in models with strong indirect aerosol
 762 effects [Zhang *et al.*, 2013; Menary *et al.*, 2013]. The anthropogenic aerosols affect SSS
 763 mainly through changes in surface evaporation, i.e., an enhanced anthropogenic aerosols lead
 764 to an increase in surface evaporation and thus an increase in SSS over the subpolar North At-
 765 lantic [Menary *et al.*, 2013]. There is a coupled positive feedback between SSS and AMOC
 766 changes. The increasing/decreasing AMOC trends in the 20th century modulate SSS over
 767 the subpolar North Atlantic, and vice versa, the increasing/decreasing trends in SSS over the
 768 deep water formation sites in the subpolar North Atlantic reinforce AMOC changes [Menary
 769 *et al.*, 2013]. The observed multidecadal subpolar North Atlantic SSS variations are very
 770 different from those found in these CM4.0 historical simulations over the second half of the
 771 20th century, with observations showing a negative phase during 1970s and 1980s and pos-
 772 itive phase during 1960s and post-1990 (Fig. 24). Based on earlier studies, we suggest that
 773 the observed multidecadal changes are dominated by internal variability [Zhang *et al.*, 2013;
 774 Zhang, 2017; Zhang *et al.*, 2019]. This discrepancy between CM4.0 from the observational
 775 records reinforces the hypothesis that CM4.0 has insufficient multidecadal variability. The
 776 lack of spread in the model realizations in CM4.0 suggests that it is incapable of generating a
 777 subpolar SSS evolution consistent with observations.

784 While the CM4.0 variability in subpolar Atlantic salinity is problematic, the model's
 785 simulation of the mean North Atlantic salinity is greatly improved over other GFDL mod-
 786 els. Fig. 25 compares the CM4.0 historical simulation with that of CM3 over recent decades.

787 The effects of the improved open ocean circulation (e.g. more realistic North Atlantic Cur-
 788 rent pathway) described earlier is evident in the extra-tropical salinities, and some patholog-
 789 ical behavior in the North Sea is clearly improved. But especially encouraging is the sim-
 790 ulation of the low salinity ribbon tightly constrained to the western boundary, realistically
 791 present down to Cape Hatteras and beyond, which is related to the more realistic Gulf Stream
 792 separation. The Labrador Sea is too saline consistent with the picture that there is insufficient
 793 mixing of these coastal low salinities into the center of the Labrador Sea, consistent with the
 794 steadiness of deep convection in that basin.

800 5.4 Sea Ice

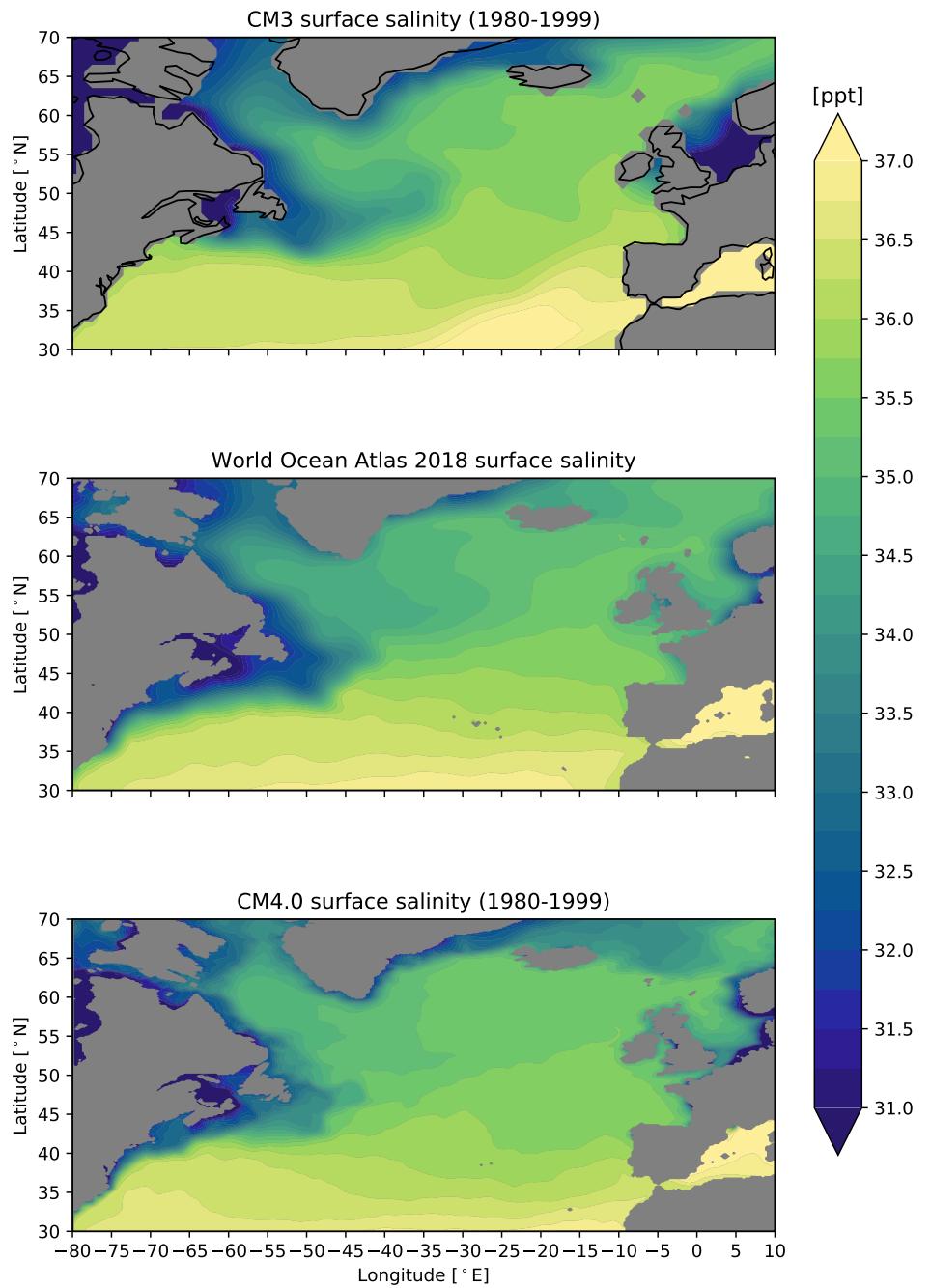
801 The observed magnitude and timing of the seasonal cycle of Pan-Arctic sea-ice ex-
 802 tent (SIE) are quite well represented by CM4.0, whereas Pan-Antarctic SIE displays an am-
 803 plified seasonal cycle with too little summer ice and too much winter ice (Fig. 26). The
 804 RMS difference between the modeled and observed monthly seasonal cycle of Pan-Antarctic
 805 SIE is 2.32 million km², which is substantial but still lower than the CMIP5 median RMS
 806 bias (3.42) [Shu *et al.*, 2015] and also represents a large improvement relative to earlier-
 807 generation GFDL models CM2.1 and CM3 (Griffies *et al.* [2011] – RMS values of 4.01 and
 808 5.82, respectively). The Pan-Arctic SIE climatology has a lower RMS error of 0.59 million
 809 km². This value is also lower than the CMIP5 median RMS error (1.45) [Shu *et al.*, 2015], is
 810 improved relative to CM2.1 (1.68), and slightly degraded relative to CM3 (0.41).

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

819 CM4.0 simulates declines of Pan-Arctic SIE in all months of the year, with trend mag-
 820 nitudes that are in reasonable agreement with satellite-observed trends (see Fig. 28). Com-
 821 puted over years 1979–2014, the ensemble mean CM4.0 Pan-Arctic SIE trends are -0.66 mil-
 822 lion and -0.28 million km²/decade in September and March, respectively, which are slightly
 823 weaker than the observed trends of -0.87 in September and -0.39 in March. The CM4.0 Pan-
 824 Arctic SIE trends are comparable to the CMIP5 multi-model ensemble mean trends of -0.57
 825 million km²/decade in September and -0.30 million km²/decade in March [Stroeve *et al.*,
 826 2012; Massonnet *et al.*, 2012; Shu *et al.*, 2015]. The three ensemble members make it clear
 827 that internal variability in CM4.0 can make only a minor contribution to the trend, consis-
 828 tent with its weak AMOC variability. If, as suggested in some studies, internal variability
 829 has contributed a non-negligible fraction to the observed trend, it is possible that the model's
 830 externally forced decline is too large despite agreeing with observations in isolation. In this
 831 regard, a more satisfying metric would be consistency with both the SSS subpolar Atlantic
 832 evolution, as a proxy for AMOC strength, as well as the Arctic sea ice trends.

833 Similar to most CMIP5 models, CM4.0 simulates negative trends for Pan-Antarctic
 834 SIE and fails to capture the slightly positive observed SIE trends of +0.21 million km²/decade
 835 in March and +0.24 million km²/decade in September. The simulated CM4.0 trends are -
 836 0.34 and -0.09 million km²/decade in September and March, respectively, which can be com-
 837 compared to the CMIP5 multi-model ensemble mean trends of -0.40 in September and -0.25 in
 838 March [Turner *et al.*, 2013]. The CM4.0 historical simulations do not include the effects of
 839 freshwater input from melting Antarctic land ice, which may be an important factor in cap-
 840 turing observed Antarctic sea ice trends [Bronseelaer *et al.*, 2018].

841 CM4.0 has a clear bias in its spatial pattern of Arctic sea ice thickness (SIT) (see Figs. 29a,b).
 842 CM4.0 has its thickest ice in the Beaufort, Chukchi, and East Siberian Seas, lacks the ob-



795 **Figure 25.** North Atlantic sea surface salinity from (A) GFDL's previous generation CM3 model historical
 796 simulation averaged over years 1980-1999 (<ftp://nomads.gfdl.noaa.gov/CMIP5/output1/NOAA-GFDL/GFDL->
 797 CM3/historical/mon/ocean/Omon/r1i1p1/v20110601/so/; *Donner et al. [2011]; Griffies et al. [2011]*) (B)
 798 World Ocean Atlas 0.25 degree 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.
 799

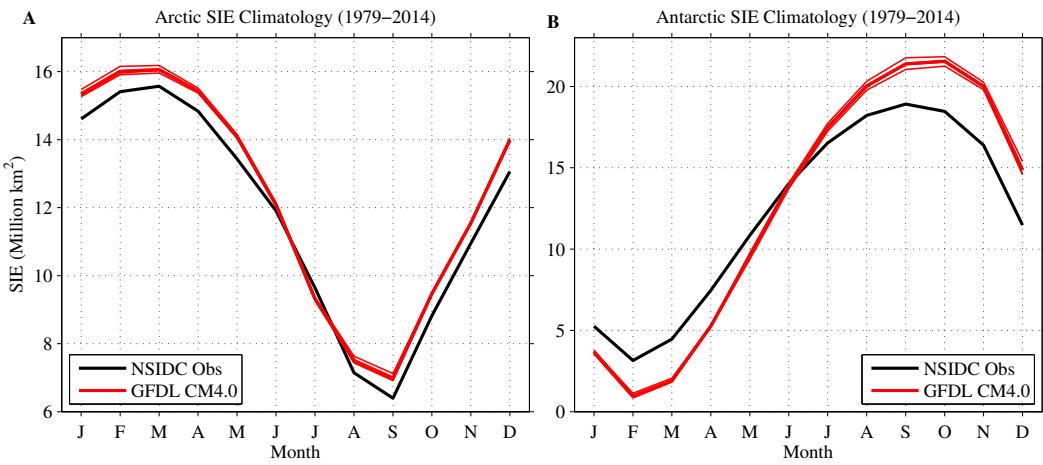
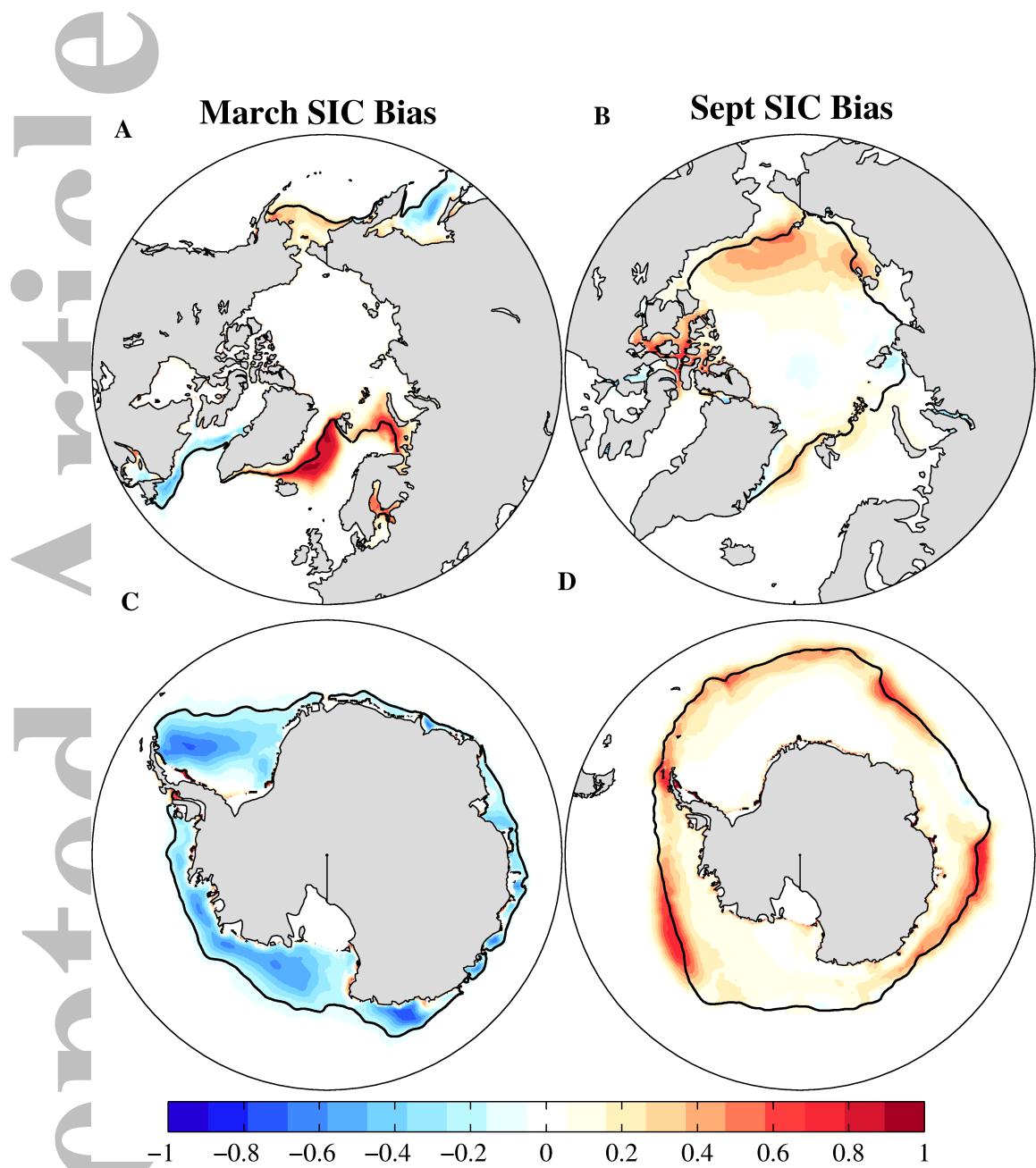


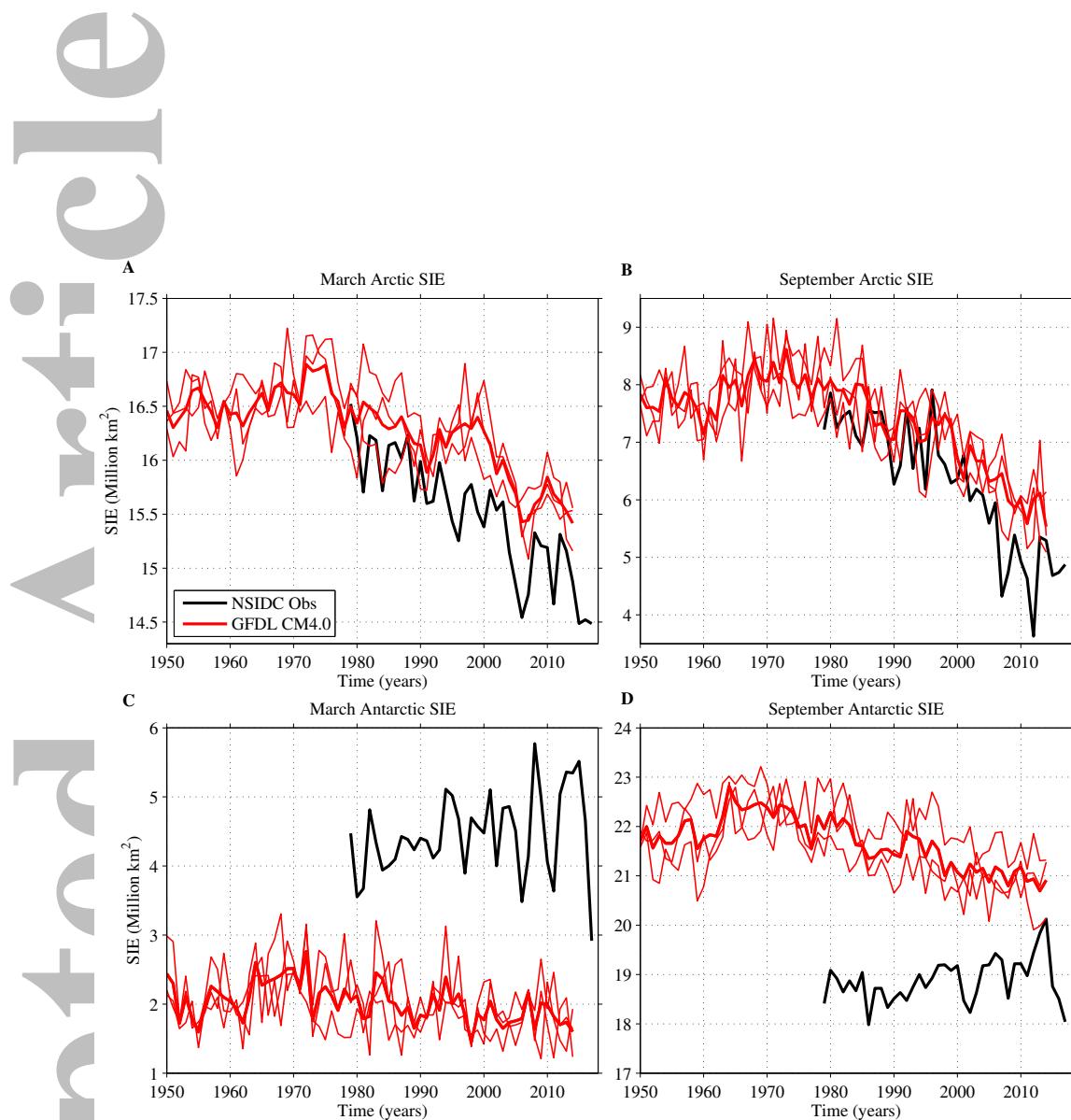
Figure 26. Pan-Arctic (A) and Pan-Antarctic (B) sea-ice extent (SIE) climatologies (million km²) computed over 1979–2014 from satellite observations (black), the three CM4.0 historical simulations (red), and the CM4.0 historical ensemble mean (thick red). The observed SIE climatology is computed using 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.

served East-to-West SIT gradient, and generally resembles a 90° westward rotation of the observed SIT pattern. The CM4.0 ice velocities are biased high relative to satellite-observed velocities. The model generally displays the key climatological circulation features of Arctic sea ice, such as an anti-cyclonic 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 towards 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 towards the Siberian coastline, favoring ice advection into the Beaufort Sea (see Fig. 29d,e). As mentioned above, this Arctic wind field is sensitive to the precise formulation of orographic gravity wave drag in CM4.0.

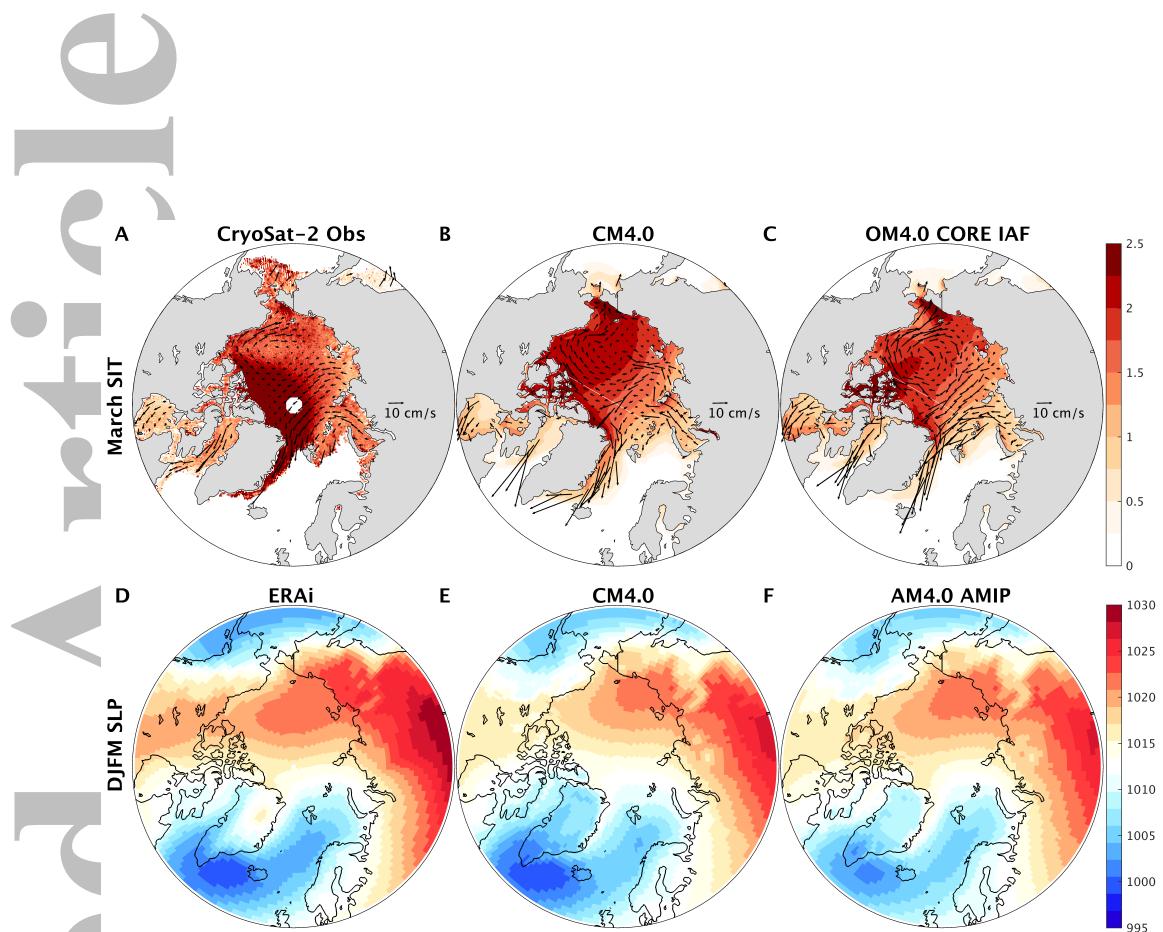
However, the Arctic sea level pressure (SLP) bias is unable to fully account for the biased SIT pattern. We find that an OM4.0 ice-ocean simulation forced by CORE (Coordinated Ocean-Ice Reference Experiment) interannual atmospheric forcing [Adcroft *et al.*, 2019] also displays a biased SIT pattern despite the reanalysis-based winds used to force this run (see Fig. 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.



826 **Figure 27.** Climatological Arctic (A, B) and Antarctic (C, D) sea-ice concentration (SIC) biases (model
 827 minus observations), computed over years 1979–2014. The black contour indicates the observed climatolog-
 828 ical sea-ice edge (15% SIC contour) and the observed data is monthly-averaged passive microwave satellite
 829 sea-ice concentration observations from NSIDC processed using the NASA Team Algorithm (data set ID:
 830 NSIDC-0051, Cavalieri *et al.* [1996]).



845 **Figure 28.** Time series of March and September Pan-Arctic (A,B) and Pan-Antarctic (C,D) sea-ice extent
 846 (SIE, million km^2) from NSIDC satellite observations (black; data set ID: NSIDC-0051, Cavalieri *et al.*
 847 [1996]), the three CM4.0 historical simulations (red), and the CM4.0 ensemble mean (thick red).



856 **Figure 29.** March Arctic sea-ice thickness climatology (m) from CryoSat-2 Observations (A), CM4.0
 857 historical simulations (B), and an OM4.0 CORE-forced simulation (C). Observed and simulated climatolog-
 858 ical sea-ice velocities (DJFM; cm/s) are plotted as vectors in Panels A–C. Also shown are Arctic sea level
 859 pressure (SLP) climatologies (hPa) for DJFM computed using Era-Interim reanalysis (D), CM4.0 (E), and an
 860 AM4.0 AMIP simulation (F). The observational data come from the Alfred-Wegener-Institute CryoSat-2 SIT
 861 product, which provides monthly winter season data spanning 2010–11 to 2017–18 [Ricker *et al.*, 2014], the
 862 low resolution daily sea ice drift product of the EUMETSAT Ocean and Sea Ice Satellite Application Facility
 863 (OSISAF, Lavergne *et al.* [2010]), and Era-Interim reanalysis [Dee *et al.*, 2011]. The CM4.0 climatologies are
 864 computed over the final ten years of the historical simulation (2005–2014), observed climatologies using years
 865 2011–2018, and OM4.0 climatologies over years 1998–2007. SLP climatologies are computed over years
 866 2005–2014.

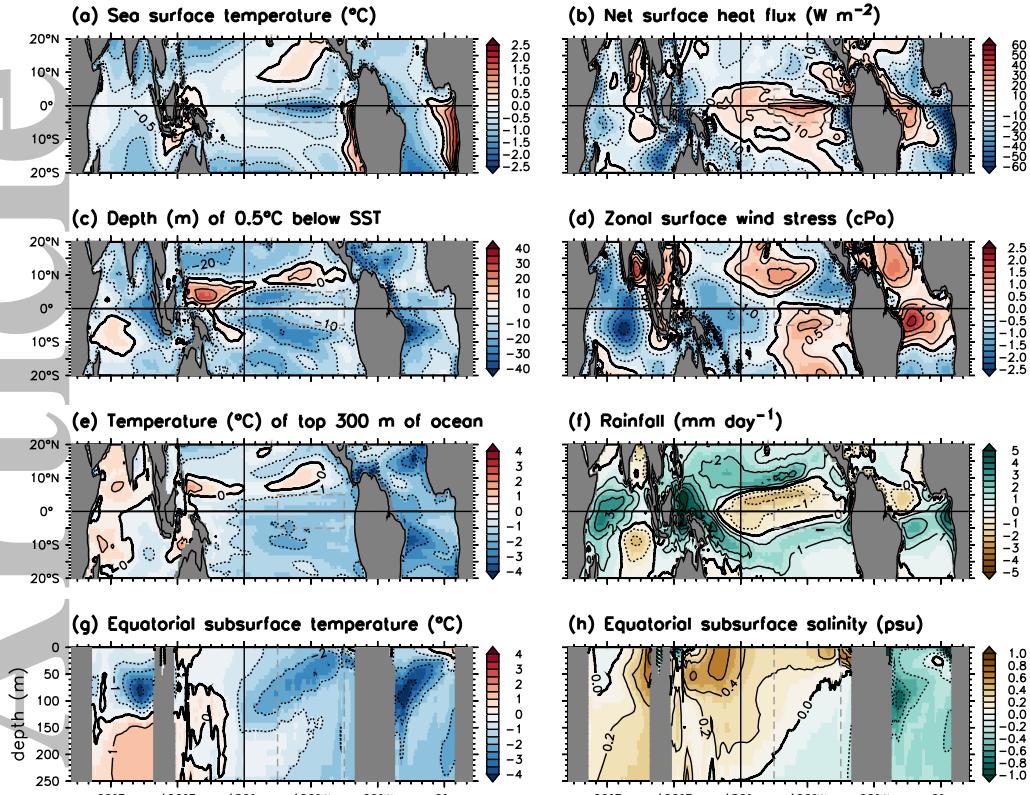
893 **5.5 Tropical oceans**

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

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

927 CM4.0's wind stress biases (Fig. 30d) are closely linked to its rainfall biases shown
 928 in Fig. 30f, which indicate excessive rainfall in the ITCZ (inter-tropical convergence zone),
 929 SPCZ (south Pacific convergence zone), and near the Maritime Continent. CM4.0's east-
 930 ern equatorial Pacific cold SST bias (Fig. 30a) also displaces the Pacific ITCZ and SPCZ
 931 poleward of their observed positions, giving rise to a dry bias over the cold tongue region.
 932 Although much suppressed in the eastern Pacific compared to that in other models, as doc-
 933 umented above, a double-ITCZ bias still appears in the southeast tropical Pacific and south-
 934 west tropical Atlantic, associated with excessive convection south of the equator during bo-
 935 real spring. These rainfall biases are linked to biases in cloudiness (not shown), reducing
 936 surface shortwave heating and net ocean heat uptake (Fig. 30b) near the Maritime Continent,
 937 ITCZ/SPCZ, eastern equatorial Atlantic, and western Indian Ocean. The rainfall biases also
 938 drive surface wind stress biases (Fig. 30d); for example, the excessive westward rain gra-
 939 dients in the western equatorial Pacific and Indian Oceans result in excessive easterly trade
 940 winds in those regions. The excessive rainfall over the Maritime Continent and near equa-
 941 torial Africa also lead to reduced surface salinity and intensified near-surface haloclines in
 942 those regions (Fig. 30f).

943 Although there are relatively straightforward connections between these biases in many
 944 cases, their ultimate cause remains a subject of investigation [*Wittenberg et al.*, 2018; *Ray*
 945 *et al.*, 2018a,b]. Also unclear are the precise connections between these mean biases and



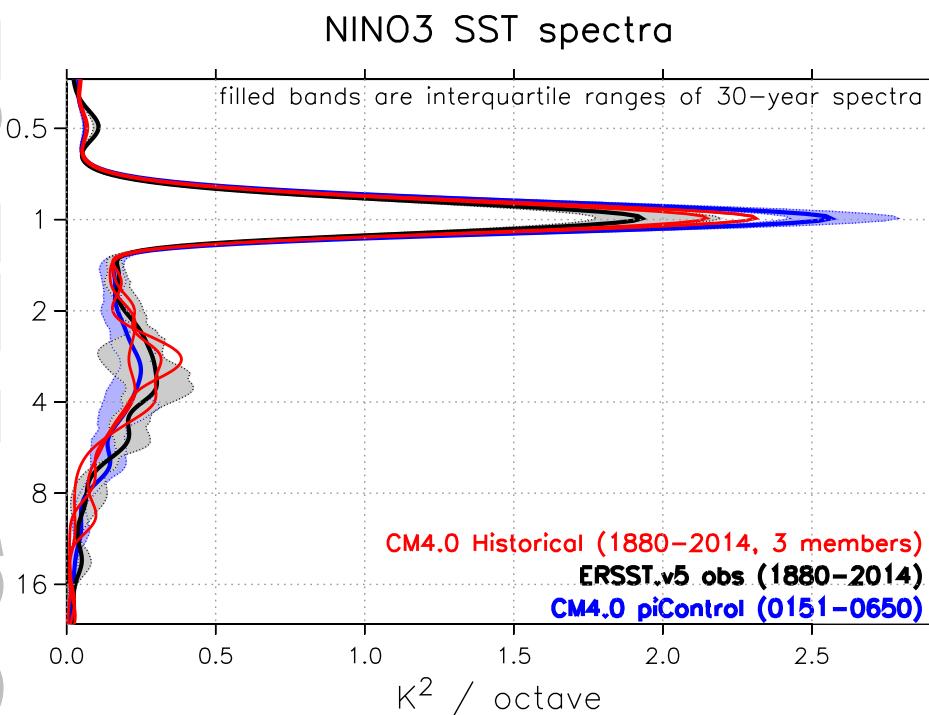
952 **Figure 30.** Annual-mean (1980–2014) climatological biases of CM4.0 relative to observational reanalyses.
 953 (a) SST ($^{\circ}\text{C}$). (b) Net surface heat flux (W m^{-2}); positive values heat the ocean. (c) Depth (m) at which the
 954 time-mean ocean temperature falls 0.5 K below the time-mean SST. (d) Zonal component of the surface wind
 955 stress (cPa); positive values exert a westerly (i.e. eastward) stress on the ocean. (e) Vertically-averaged tem-
 956 perature of the top 300 m of the ocean (shallower regions are masked out). (f) Surface rainfall (mm day^{-1}).
 957 (g) Subsurface temperature ($^{\circ}\text{C}$) interpolated to the equator. (h) Subsurface salinity (psu) interpolated to the
 958 equator. Shading is incremented every half contour. CM4.0 represents a mean of three historical runs with
 959 CMIP6 historical forcings. The observational reference datasets for the 1980–2014 period are: the ECMWF
 960 Interim reanalysis [Dee et al., 2011] for panels a,b,d; the ORA-S4 reanalysis [Balmaseda et al., 2013] for
 961 panels c,e,g,h; and the GPCP.v2.3 dataset [Adler et al., 2003, 2016] for panel f. Dashed gray box is the Niño-3
 962 region (150°W – 90°W , 5°S – 5°N).

946 variability on various time scales. It is encouraging that CM4.0's simulation of the Madden-
 947 Julian Oscillation (MJO), for example, is realistic despite these mean biases, as outlined in
 948 Zhao et al. [2018a], where the fact that the MJO in this model is strongly affected by ocean
 949 coupling is described. (The coupled model utilized in Zhao et al. [2018a] to make this point
 950 is slightly different from CM4.0, but its MJO simulation is practically identical to that of
 951 CM4.0, and we refer to that paper for the characteristics of the MJO in this model.)

963 5.6 ENSO

973 Fig. 31 shows power spectra for SST averaged over the Niño-3 region (150°W – 90°W ,
 974 5°S – 5°N), from an observational reconstruction and from the CM4.0 historical and pre-
 975 industrial simulations. The observed time-mean spectrum (black curve) shows an energetic

Accepted Article



964 **Figure 31.** Power spectra of SST averaged over the Niño-3 region (150°W – 90°W , 5°S – 5°N), following
 965 Fig. 2 of Wittenberg [2009]. Abscissa indicates the time-averaged spectral power ($K^2 \text{ octave}^{-1}$) from a Morlet
 966 wavenumber 6 wavelet analysis. The area to the left of each curve represents the total spectral power within
 967 a given range of periods on the ordinate. Red lines are 1880–2014 time-mean spectra for the three CM4.0
 968 historical ensemble members. Black curve is the 1880–2014 time-mean spectrum from the ERSST.v5 ob-
 969 servational reanalysis [Huang *et al.*, 2017]; gray band indicates the interquartile range of ERSST.v5 sliding
 970 30-year subspectra during 1880–2014. Blue curve is the CM4.0 pre-industrial control run time-mean spec-
 971 trum during simulated years 151–650; light blue band is the interquartile range of sliding 30-year spectra
 972 during that 500-year epoch.

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 Fig. 31, which shows the interquartile range of spectral curves diagnosed from 30-year (sliding) epochs during 1880–2014.

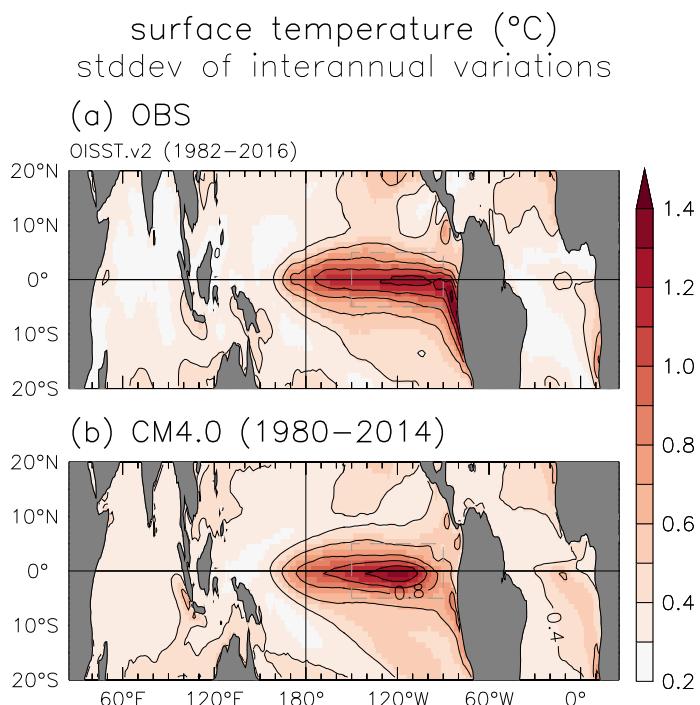
The red curves in Fig. 31 show the three CM4.0 simulations driven by historical forcings over 1880–2014, which closely resemble the observed spectrum. The realism of the equatorial Pacific SST spectrum is a remarkable improvement over most previous coupled global climate models (GCMs) including those previously produced by GFDL [Wittenberg *et al.*, 2006; Delworth *et al.*, 2012; Vecchi *et al.*, 2014]. 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 Fig. 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 Fig. 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 pre-industrial.

Fig. 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, 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.

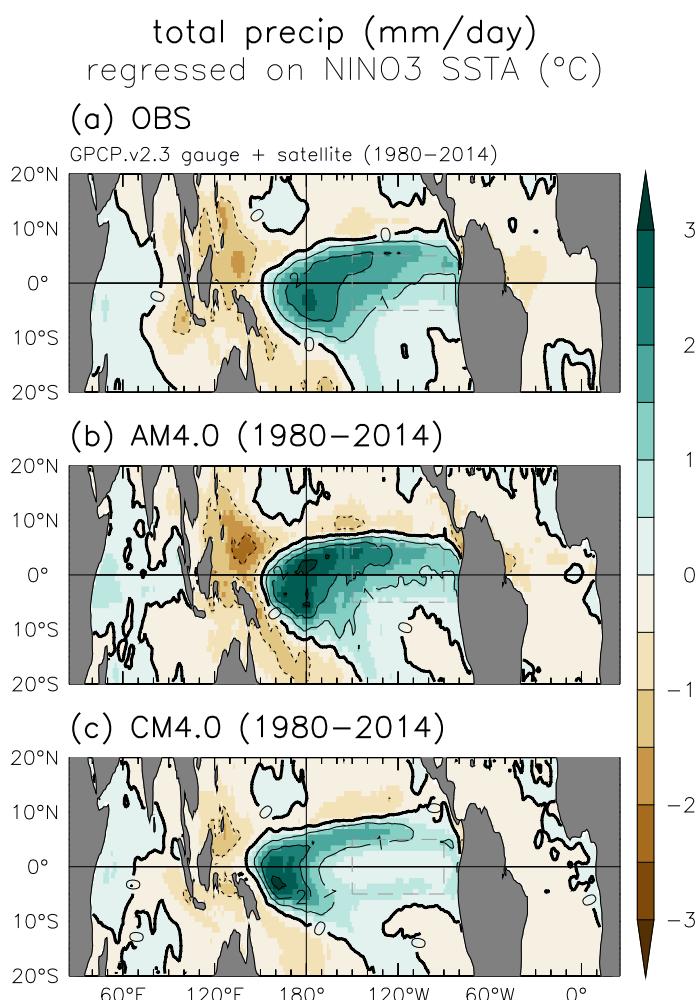
Fig. 33 shows tropical maritime rainfall anomalies regressed onto Niño-3 SSTAs, for the observations and the AM4.0 and CM4.0 historical simulations. The observations indicate an eastward and equatorward shift of tropical Pacific rainfall during El Niño, with drying near the Maritime Continent and over 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 [Fig. 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 (Fig. 32), which is quite realistic; rather, the response biases are reminiscent of CM4.0’s climatological rainfall biases (Fig. 30f), linked to CM4.0’s cold SST bias in the eastern equatorial Pacific (Fig. 30a) that inhibits local atmospheric deep convection.

Fig. 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



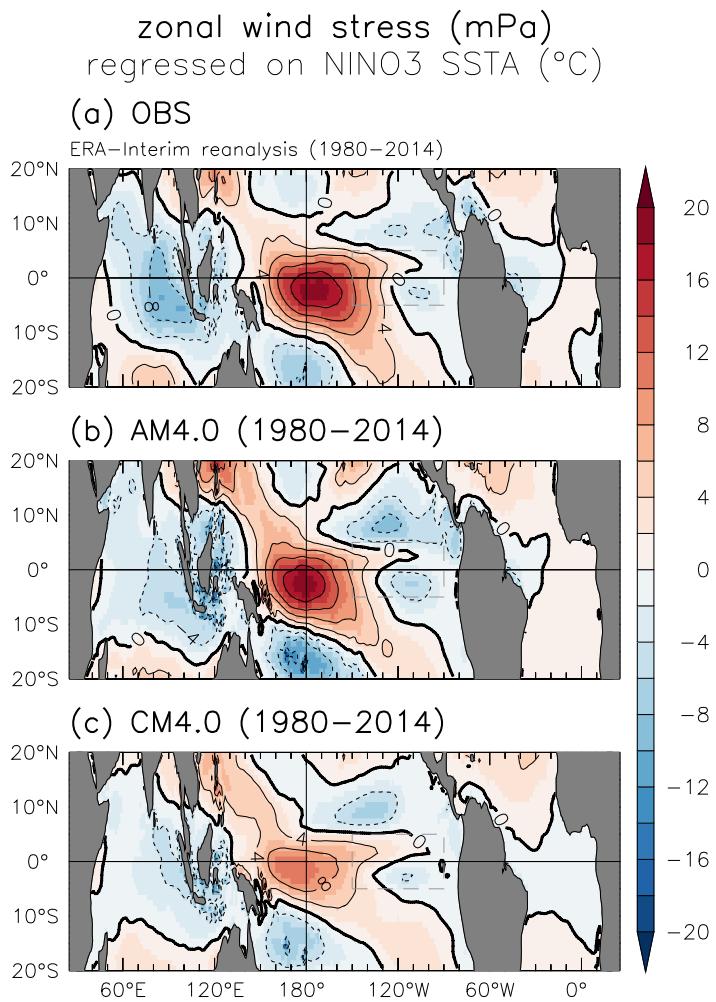
1006 **Figure 32.** Standard deviation of tropical SST anomalies (K), after subtracting a 12-month climatology and
1007 applying a 9-month triangle smoother that transmits (25, 50, 75)% of the spectral amplitude at periods of (8,
1008 11, 17) months, following Fig. 17 of Delworth *et al.* [2012]. (a) Observations from the OISST.v2 reanalysis
1009 [1982–2016; Reynolds *et al.*, 2002]. (b) Ensemble mean of the three 1980–2014 standard deviations com-
1010 puted from the three CM4.0 historical runs. All datasets are averaged onto a 1.25° lon \times 1° lat grid before
1011 computing the temporal standard deviation. Dashed gray box is the Niño-3 region (150°W – 90°W , 5°S – 5°N).

Accepted Article



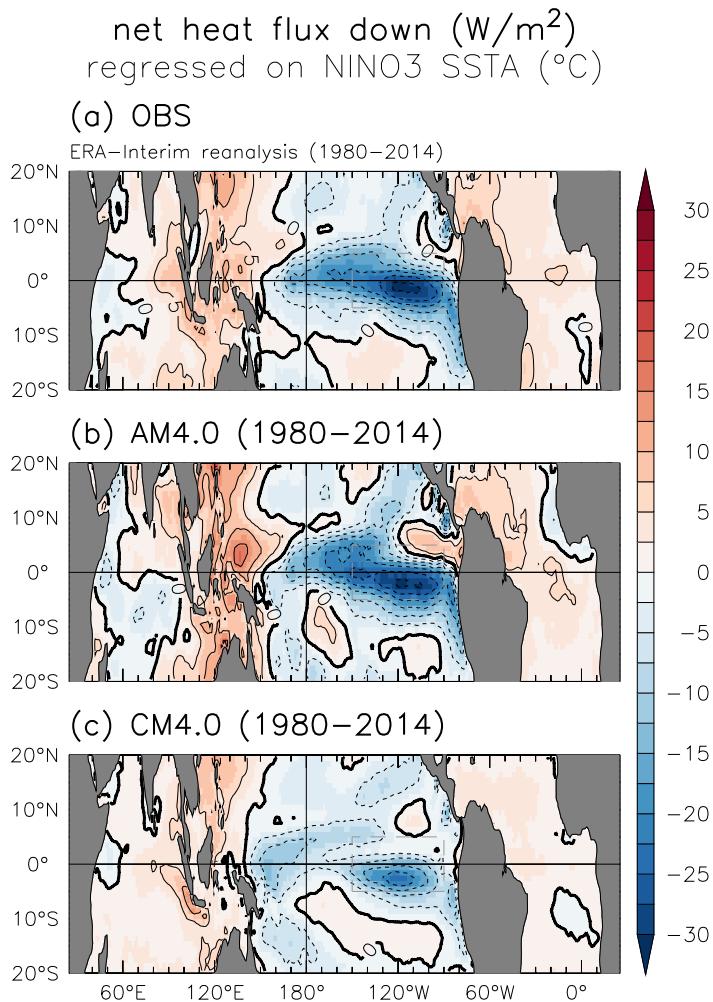
1017 **Figure 33.** Linear regression of monthly rainfall anomalies onto monthly Niño-3-averaged SST anomalies
 1018 ($\text{mm day}^{-1} \text{ K}^{-1}$), for all months during 1980–2014. (a) Observations from GPCP.v2.3 [Adler *et al.*, 2003,
 1019 2016] for rainfall, and ERSST.v4 [Huang *et al.*, 2015] for SST. (b) Regression computed from the AM4.0
 1020 component forced by observed SSTs during 1980–2014. (c) Ensemble mean of the three 1980–2014 regres-
 1021 sions computed from the three CM4.0 historical runs. All datasets are averaged onto a $1.25^\circ \text{ lon} \times 1^\circ \text{ lat}$ grid
 1022 before computing the temporal regression. Dashed gray box is the Niño-3 region (150°W – 90°W , 5°S – 5°N).

Accepted Article



1037 **Figure 34.** As in Fig. 33, but for zonal wind stress anomalies regressed onto Niño-3 SST anomalies
 1038 (mPa K^{-1}). Positive values indicate anomalous westerly (i.e. eastward) stress on the ocean during El Niño.
 1039 Observational estimates are from the ERA-Interim reanalysis [Dee *et al.*, 2011] for both wind stress and SST.

Accepted Article



1050 **Figure 35.** As in Fig. 34, but for net surface heat flux anomalies regressed onto Niño-3 SST anomalies
 1051 ($\text{W m}^{-2} \text{ K}^{-1}$). Positive values indicate anomalous heating of the ocean during El Niño.

1043 of this zone. In contrast, CM4.0's wind stress response is substantially weaker and narrower,
 1044 and shifted slightly farther west consistent with CM4.0's westward-displaced rainfall anomalies
 1045 (Fig. 33). CM4.0's meridionally narrower and westward-displaced wind stress response
 1046 — which would be expected to enhance the Sverdrup discharge of equatorial oceanic heat,
 1047 hastening the turnaround of El Niño into La Niña [Wittenberg, 2002; Capotondi *et al.*, 2006]—
 1048 may be partly responsible for CM4.0's reduced ENSO spectral variance at time scales of 5–6
 1049 years (Fig. 31).

1052 Fig. 35 shows the ENSO response of the net surface heat flux into the ocean. Similar
 1053 regressions for the individual surface heat flux components (shortwave, longwave, latent,
 1054 sensible) and cloud cover (total, low, middle, high) were also examined to support this dis-
 1055 cussion, though for brevity are not displayed here. The observed pattern generally indicates a
 1056 damping of the ENSO-driven SSTAs, with enhanced cooling over the Pacific equatorial cold
 1057 tongue during El Niño. This cooling is driven mainly by enhanced evaporation in the eastern
 1058 equatorial Pacific (associated with the warm SSTAs) and enhanced shading of the surface by
 1059 deep convective clouds in the central and slightly off-equatorial Pacific (on the equatorward
 1060 flanks of the climatological ITCZ and SPCZ, and on the eastern flank of the climatological
 1061 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 [Fig. 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 (Figs. 34, 35) seems clearly linked to the coupled model's climatological biases over the tropical Pacific (Fig. 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 (GCMs), 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 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 [*Chiang and Bitz*, 2005; *Broccoli et al.*, 2006; *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 relative 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.

1112 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 mid-latitudes, each of which help in improving simulations of phenomena such as Gulf Stream separation and the Gulf of Mexico loop current, the deep western boundary currents 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, as well as backscatter-based approaches for scale-aware mesoscale eddy closures.

1124 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, relatively, 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.

1132 The high quality of the ENSO simulation, the frequency spectrum and SST anomaly 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.

1140 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 Southern Hemisphere, produced by super-polynyas forming in the Ross Sea. These do not affect the Northern Hemisphere 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.

1149 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.

1155 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 [*Mahajan et al.*, 2011; *Day et al.*, 2012; *Zhang and Knutson*, 2013; *Swart et al.*, 2015; *Zhang*, 2015; *Li et al.*, 2017, 2018; *Ding et al.*, 2019] is a key question. In this regard we are particularly concerned with the unrealistic evolution of sea surface salinity in the subpolar North Atlantic in the model's historical simulation. The simulation of Arctic sea ice thickness also suggests deficiencies in the sea ice model's rheology or thermodynam-

1164 ics. The sea ice simulation around Antarctica has numerous issues - most notably too little
 1165 summer sea ice and large centennial-scale variability associated with Ross Sea super-polynya
 1166 events.

1167 A variety of sensitivity studies with this coupled model have effectively been per-
 1168 formed during the development process. Interesting sensitivities include that of subpolar
 1169 Northern latitude winds and temperatures to sub-grid orographic drag (especially over Green-
 1170 land), temperature sensitivity to masking of snow albedos by vegetation, and overall bias re-
 1171 sponses to both horizontal and vertical atmospheric resolution, to the inclusion of energy
 1172 conserving backscatter and to specifics of the oceanic hybrid vertical coordinate. Most of
 1173 these need to be re-examined with the finalized model and with longer time integrations, so
 1174 we postpone discussion of these sensitivities, instead planning to return to some of them in
 1175 future publications.

1176 A model such as CM4.0 is meant to be an all-purpose climate simulator. As such, re-
 1177 sults from a variety of studies by the climate science community at large need to accumulate
 1178 to determine the true value of the model for addressing various problems of interest. Sub-
 1179 stantial further work will also be required before evaluating the value of CM4.0 in providing
 1180 a platform for future model improvements.

1181 Acknowledgments

1182 We thank Ron Stouffer and Tom Knutson for very helpful reviews of an earlier draft.
 1183 We thank Tom Delworth for consultations concerning the sensitivity of the modelled Antarc-
 1184 tic super-polynyas to ice sheet albedo, and for providing the CM2.5 data used in Figure 19.
 1185 We acknowledge the World Climate Research Program which, through its Working Group
 1186 on Coupled Modelling coordinated and promoting CMIP5 and CMIP6. We thank the cli-
 1187 mate modeling groups for producing and making available their model output, the Earth Sys-
 1188 tem Grid Federation (ESGF) for archiving the data and providing access, and the mutiple
 1189 funding agents who support CMIP5/6 and ESGF. We thank the Program for Climate Model
 1190 Diagnosis and Intercomparison (PCMDI) and the IPCC Data Archive at the Lawrence Liver-
 1191 more National Laboratory/Department of Energy (LNL/DOE) for collecting the CMIP5 data.
 1192 LLNL is operated by Lawrence Laboratory National Security. LLC, for DOE. M. Zhao, J.-
 1193 C. Golaz, B. Xiang, and Y. Ming acknowledge partial support by NOAA's Climate Program
 1194 Office (CPO) Climate Variability and predictability (CVP) Program (GC14-252) through a
 1195 CPO CVP funded proposal entitled "Process Level Investigation of the Role of Convection
 1196 and Cloud Parameterization in Tropical Pacific Bias in GFDL Next Generation Global Cli-
 1197 mate Models". Work at LLNL was performed under Contract DE-AC52-07NA27344 from
 1198 DOE/LLNL. Many GFDL colleagues and collaborators not included on the author list but
 1199 who have played invaluable roles in the model workflow and publication of the data are ac-
 1200 knowledged at doi.org/10.22033/ESGF/CMIP6.1402. The CM4.0 source code is available at
 1201 <https://doi.org/10.5281/zenodo.3339397>. Data available from authors have been deposited in
 1202 the CMIP6 archive tagged with the identifier doi.org/10.22033/ESGF/CMIP6.1402.

1203 References

- 1204 Adcroft, A., W. Anderson, V. Balaji, C. Blanton, M. Bishuk, et al. (2019), The GFDL global
 1205 ocean and sea ice model OM4.0: Model description and simulation features, *Journal of*
 1206 *Advances in Modeling Earth Systems*, doi:10.1029/2019MS001726.
 1207 Adler, R. F., G. J. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schnei-
 1208 der, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, and P. Arkin (2003), The Version 2
 1209 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-
 1210 Present), *J. Hydrometeor.*, 4, 1147–1167.
 1211 Adler, R. F., M. Sapiano, G. Huffman, D. Bolvin, G. Gu, J. Wang, E. Nelkin, P. Xie, L. Chiu,
 1212 R. Ferraro, U. Schneider, and A. Becker (2016), The new version 2.3 of the Global Pre-

- 1213 precipitation Climatology Project (GPCP) monthly analysis product, *Internal Report*, Earth
 1214 System Science Interdisciplinary Center, Univ. of Maryland, College Park, MD.
- 1215 Balaji, V. (2012), The Flexible Modeling System, in *Earth System Modelling-Volume 3*, pp.
 1216 33–41, Springer.
- 1217 Balmaseda, M. A., K. Mogensen, and A. T. Weaver (2013), Evaluation of the ECMWF
 1218 ocean reanalysis system ORAS4, *Quart. J. Roy. Meteor. Soc.*, 139, 1132–1161,
 1219 doi:10.1002/qj.2063.
- 1220 Bayr, T., M. Latif, D. Dommenget, C. Wengel, J. Harlaß, and W. Park (2018), Mean-state
 1221 dependence of ENSO atmospheric feedbacks in climate models, *Climate Dyn.*, 50, 3171–
 1222 3194, doi:10.1007/s00382-017-3799-2.
- 1223 Boyer, T. P., S. Levitus, J. I. Antonov, R. A. Locarnini, and H. E. Garcia (2005), Linear
 1224 trends in salinity for the World Ocean, 1955–1998, *Geophysical Research Letters*, 32(1),
 1225 doi:10.1029/2004GL021791.
- 1226 Broccoli, A. J., K. A. Dahl, and R. J. Stouffer (2006), Response of the ITCZ to Northern
 1227 Hemisphere cooling, *Geophysical Research Letters*, 33(1).
- 1228 Bronselaer, B., M. Winton, S. M. Griffies, W. J. Hurlin, K. B. Rodgers, O. V. Sergienko, R. J.
 1229 Stouffer, and J. L. Russell (2018), Change in future climate due to Antarctic meltwater,
 1230 *Nature*, 564(7734), 53.
- 1231 Campbell, E. C., E. A. Wilson, G. K. Moore, S. C. Riser, C. E. Brayton, M. R. Mazloff, and
 1232 L. D. Talley (2019), Antarctic offshore polynyas linked to southern hemisphere climate
 1233 anomalies, *Nature*, p. 1.
- 1234 Capotondi, A., A. Wittenberg, and S. Masina (2006), Spatial and temporal structure of trop-
 1235 ical Pacific interannual variability in 20th century coupled simulations, *Ocean Modelling*,
 1236 15, 274–298, doi:10.1016/j.ocemod.2006.02.004.
- 1237 Cavalieri, D. J., C. L. Parkinson, P. Gloersen, and H. J. Zwally (1996), Sea ice concentrations
 1238 from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 1,
 1239 *NASA DAAC at the Natl. Snow and Ice Data Cent.*, doi:10.5067/8GQ8LZQVL0VL.
- 1240 Cess, R. D., and Coauthors (1990), Intercomparison and interpretation of climate feedback
 1241 processes in 19 Atmospheric General Circulation Models, *Journal of Geophysical Re-*
 1242 *search*, 95(D10), 16,601–16,615.
- 1243 Cess, R. D., and Coauthors (1996), Cloud feedback in atmospheric general circulation mod-
 1244 els: An update, *Journal of Geophysical Research*, 101(D8), 12,791–12,794.
- 1245 Chassignet, E. P., L. T. Smith, G. R. Halliwell, and R. Bleck (2003), North Atlantic Simula-
 1246 tions with the Hybrid Coordinate Ocean Model (HYCOM): Impact of the Vertical Coordi-
 1247 nate Choice, Reference Pressure, and Thermobaricity, *Journal of Physical Oceanography*,
 1248 33(12), 2504–2526, doi:10.1175/1520-0485(2003)033<2504:NASWTH>2.0.CO;2.
- 1249 Chiang, J. C., and C. M. Bitz (2005), Influence of high latitude ice cover on the marine In-
 1250 tertropical Convergence Zone, *Climate Dynamics*, 25(5), 477–496.
- 1251 Cunningham, S. A., T. Kanzow, D. Rayner, M. O. Baringer, W. E. Johns, J. Marotzke, H. R.
 1252 Longworth, E. M. Grant, J. J.-M. Hirschi, L. M. Beal, et al. (2007), Temporal variability of
 1253 the Atlantic meridional overturning circulation at 26.5 N, *Science*, 317(5840), 935–938.
- 1254 Danabasoglu, G., W. G. Large, and B. P. Briegleb (2010), Climate impacts of parameterized
 1255 Nordic Sea overflows, *Journal of Geophysical Research: Oceans*, 115(C11).
- 1256 Danabasoglu, G., S. G. Yeager, D. Bailey, E. Behrens, M. Bentsen, D. Bi, A. Biastoch,
 1257 C. Böning, A. Bozec, V. M. Canuto, et al. (2014), North Atlantic simulations in coor-
 1258 dinated ocean-ice reference experiments phase II (CORE-II). Part I: mean states, *Ocean*
 1259 *Modelling*, 73, 76–107.
- 1260 Day, J., J. Hargreaves, J. Annan, and A. Abe-Ouchi (2012), Sources of multi-decadal vari-
 1261 ability in Arctic sea ice extent, *Environmental Research Letters*, 7(3), 034,011.
- 1262 de Boyer Montégut, C., G. Madec, A. S. Fischer, A. Lazar, and D. Iudicone (2004), Mixed
 1263 layer depth over the global ocean: An examination of profile data and a profile-based cli-
 1264 matology, *Journal of Geophysical Research: Oceans*, 109(C12003).
- 1265 Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae,
 1266 M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg,

- 1267 J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger,
 1268 S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Källberg, M. Köhler, M. Matricardi,
 1269 A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay,
 1270 C. Tavolato, J.-N. Thépaut, and F. Vitart (2011), The ERA-Interim reanalysis: Configura-
 1271 tion and performance of the data assimilation system, *Quart. J. Roy. Meteor. Soc.*, 137,
 1272 553–597, doi:10.1002/qj.828.
- 1273 Delworth, T. L., and K. W. Dixon (2006), Have anthropogenic aerosols delayed a greenhouse
 1274 gas-induced weakening of the North Atlantic thermohaline circulation?, *Geophysical Re-
 1275 search Letters*, 33(2).
- 1276 Delworth, T. L., A. J. Broccoli, A. Rosati, R. J. Stouffer, V. Balaji, J. A. Beesley, W. F.
 1277 Cooke, K. W. Dixon, J. Dunne, K. Dunne, et al. (2006), GFDL's CM2 global coupled
 1278 climate models. Part I: Formulation and simulation characteristics, *Journal of Climate*,
 1279 19(5), 643–674.
- 1280 Delworth, T. L., A. Rosati, W. Anderson, A. J. Adcroft, V. Balaji, R. Benson, K. Dixon,
 1281 S. M. Griffies, H.-C. Lee, R. C. Pacanowski, G. A. Vecchi, A. T. Wittenberg, F. Zeng,
 1282 and R. Zhang (2012), Simulated climate and climate change in the GFDL CM2.5 high-
 1283 resolution coupled climate model, *J. Climate*, 25, 2755–2781, doi:10.1175/JCLI-D-11-
 1284 00316.1.
- 1285 Ding, Q., A. Schweiger, M. LâŽHeureux, E. J. Steig, D. S. Battisti, N. C. Johnson,
 1286 E. Blanchard-Wrigglesworth, S. Po-Chedley, Q. Zhang, K. Harnos, et al. (2019), Finger-
 1287 prints of internal drivers of Arctic sea ice loss in observations and model simulations, *Nature
 1288 Geoscience*, 12(1), 28.
- 1289 Donner, L. J., B. L. Wyman, R. S. Hemler, L. W. Horowitz, Y. Ming, M. Zhao, J.-C. Golaz,
 1290 P. Ginoux, S.-J. Lin, M. D. Schwarzkopf, et al. (2011), The dynamical core, physical pa-
 1291 rameterizations, and basic simulation characteristics of the atmospheric component AM3
 1292 of the GFDL global coupled model CM3, *Journal of Climate*, 24(13), 3484–3519.
- 1293 Dufour, C. O., A. K. Morrison, S. M. Griffies, I. Frenger, H. Zanowski, and M. Winton
 1294 (2017), Preconditioning of the Weddell Sea Polynya by the Ocean Mesoscale and Dense
 1295 Water Overflows, *Journal of Climate*, 30(19), 7719–7737, doi:10.1175/JCLI-D-16-0586.1.
- 1296 Dunne, J. P., J. G. John, A. J. Adcroft, S. M. Griffies, R. W. Hallberg, E. Shevliakova, R. J.
 1297 Stouffer, W. Cooke, K. A. Dunne, M. J. Harrison, et al. (2012), GFDL's ESM2 global cou-
 1298 pled climate–carbon earth system models. Part I: Physical formulation and baseline simu-
 1299 lation characteristics, *Journal of Climate*, 25(19), 6646–6665.
- 1300 Evans, D. G., J. Toole, G. Forget, J. D. Zika, A. C. Naveira Garabato, A. J. G. Nurser, and
 1301 L. Yu (2017), Recent wind-driven variability in atlantic water mass distribution and merid-
 1302 ional overturning circulation, *Journal of Physical Oceanography*, 47(3), 633–647.
- 1303 Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Tay-
 1304 lor (2016), Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6)
 1305 experimental design and organization, *Geoscientific Model Development*, 9, 1937–1958,
 1306 doi:10.5194/gmd-9-1937-2016.
- 1307 Fox-Kemper, B., R. Ferrari, and R. Hallberg (2008), Parameterization of mixed layer eddies.
 1308 Part I: Theory and diagnosis, *Journal of Physical Oceanography*, 38(6), 1145–1165.
- 1309 Garner, S. T. (2005), A topographic drag closure built on an analytical base flux, *Journal of
 1310 the atmospheric sciences*, 62(7), 2302–2315.
- 1311 Garner, S. T. (2018), Ground-truth model evaluation of subgrid orographic base-flux parame-
 1312 terization, *Journal of the Atmospheric Sciences*, 75(10), 3653–3670.
- 1313 Gent, P. R., J. Willebrand, T. J. McDougall, and J. C. McWilliams (1995), Parameterizing
 1314 eddy-induced tracer transports in ocean circulation models, *Journal of Physical Oceanog-
 1315 raphy*, 25, 463–474.
- 1316 GISTEMP-Team (2019), GISS Surface Temperature Analysis (GISTEMP), version
 1317 4. NASA Goddard Institute for Space Studies, *Dataset accessed 2016-09-27 at
 1318 https://data.giss.nasa.gov/gistemp/*.
- 1319 Golaz, J.-C., L. W. Horowitz, and H. Levy (2013), Cloud tuning in a coupled climate model:
 1320 Impact on 20th century warming, *Geophysical Research Letters*, 40(10), 2246–2251.

- 1321 Gregory, J., W. Ingram, M. Palmer, G. Jones, P. Stott, R. Thorpe, J. Lowe, T. Johns, and
 1322 K. Williams (2004), A new method for diagnosing radiative forcing and climate sensi-
 1323 tivity, *Geophysical Research Letters*, 31(3).
- 1324 Griffies, S. M., M. J. Harrison, R. C. Pacanowski, and A. Rosati (2004), A technical guide to
 1325 MOM4, *GFDL Ocean Group Tech. Rep*, 5, 371.
- 1326 Griffies, S. M., M. Winton, L. J. Donner, L. W. Horowitz, S. M. Downes, R. Farneti,
 1327 A. Gnanadesikan, W. J. Hurlin, H.-C. Lee, Z. Liang, et al. (2011), The GFDL CM3 cou-
 1328 pled climate model: characteristics of the ocean and sea ice simulations, *Journal of Cli-
 1329 mate*, 24(13), 3520–3544.
- 1330 Griffies, S. M., M. Winton, W. G. Anderson, R. Benson, T. L. Delworth, C. O. Dufour, J. P.
 1331 Dunne, P. Goddard, A. K. Morrison, A. Rosati, A. T. Wittenberg, J. Yin, and R. Zhang
 1332 (2015), Impacts on Ocean Heat from Transient Mesoscale Eddies in a Hierarchy of Cli-
 1333 mate Models, *Journal of Climate*, 28(3), 952–977, doi:10.1175/JCLI-D-14-00353.1.
- 1334 Haarsma, R. J., M. J. Roberts, P. L. Vidale, C. A. Senior, A. Bellucci, Q. Bao, P. Chang,
 1335 S. Corti, N. S. Fučkar, V. Guemas, et al. (2016), High resolution model intercomparison
 1336 project (HighResMIP v1. 0) for CMIP6, *Geoscientific Model Development*, 9(1), 4185–
 1337 4208.
- 1338 Hallberg, R. (2013), Using a resolution function to regulate parameterizations of oceanic
 1339 mesoscale eddy effects, *Ocean Modelling*, 72, 92–103.
- 1340 Huang, B., V. F. Banzon, E. Freeman, J. Lawrimore, W. Liu, T. C. Peterson, T. M. Smith,
 1341 P. W. Thorne, S. D. Woodruff, and H.-M. Zhang (2015), Extended Reconstructed Sea Sur-
 1342 face Temperature version 4 (ERSST.v4): Part I. Upgrades and intercomparisons, *J. Cli-
 1343 mate*, 28, 911–930, doi:10.1175/JCLI-D-14-00006.1.
- 1344 Huang, B., P. W. Thorne, V. F. Banzon, T. Boyer, G. Chepurin, J. H. Lawrimore, M. J.
 1345 Menne, T. M. Smith, R. S. Vose, and H. Zhang (2017), Extended Reconstructed Sea Sur-
 1346 face Temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons, *J.
 1347 Climate*, 30, 8179–8205.
- 1348 Jansen, M. F., I. M. Held, A. Adcroft, and R. Hallberg (2015), Energy budget-based
 1349 backscatter in an eddy permitting primitive equation model, *Ocean Modelling*, 94, 15–26.
- 1350 John, J. G., A. M. Fiore, V. Naik, L. M. Horowitz, and J. P. Dunne (2012), Climate versus
 1351 emission drivers of methane lifetime against loss by tropospheric oh from 1860-2100, *At-
 1352 mospheric Chemistry and Physics*, 12, 12,021–12,036, doi:10.5194/acp-12-12021-2012.
- 1353 Johns, W. E., M. O. Baringer, L. M. Beal, S. Cunningham, T. Kanzow, H. L. Bryden,
 1354 J. Hirschi, J. Marotzke, C. Meinen, B. Shaw, et al. (2011), Continuous, array-based esti-
 1355 mates of Atlantic Ocean heat transport at 26.5 N, *Journal of Climate*, 24(10), 2429–2449.
- 1356 Kang, S. M., I. M. Held, D. M. Frierson, and M. Zhao (2008), The response of the ITCZ to
 1357 extratropical thermal forcing: Idealized slab-ocean experiments with a GCM, *Journal of
 1358 Climate*, 21(14), 3521–3532.
- 1359 Lavergne, T., S. Eastwood, Z. Teffah, H. Schyberg, and L.-A. Breivik (2010), Sea ice mo-
 1360 tion from low-resolution satellite sensors: An alternative method and its validation in the
 1361 Arctic, *Journal of Geophysical Research: Oceans*, 115(C10).
- 1362 Lenssen, N., G. Schmidt, J. Hansen, M. Menne, A. Persin, R. Ruedy, and D. Zyss (2019),
 1363 Improvements in the GISTEMP Uncertainty Model, *Journal of Geophysical Research:
 1364 Atmospheres*, 124, doi:<https://doi.org/10.1029/2018JD029522>.
- 1365 Li, D., R. Zhang, and T. R. Knutson (2017), On the discrepancy between observed and
 1366 CMIP5 multi-model simulated Barents Sea winter sea ice decline, *Nature Communica-
 1367 tions*, 8.
- 1368 Li, D., R. Zhang, and T. R. Knutson (2018), Comparison of mechanisms for low-frequency
 1369 variability of summer Arctic sea ice in three coupled models, *Journal of Climate*, 31(3),
 1370 1205–1226.
- 1371 Locarnini, R. A., A. V. Mishonov, J. I. Antonov, T. P. Boyer, H. E. Garcia, O. K. Baranova,
 1372 M. M. Zweng, C. R. Paver, J. R. Reagan, D. R. Johnson, et al. (2013), World ocean atlas
 1373 2013. Volume 1, Temperature.

- 1374 Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-
 1375 Smith, and T. Wong (2009), Toward optimal closure of the Earth's top-of-atmosphere radi-
 1376 ation budget, *Journal of Climate*, 22(3), 748–766.
- 1377 Mahajan, S., R. Zhang, and T. L. Delworth (2011), Impact of the Atlantic meridional over-
 1378 turning circulation (AMOC) on Arctic surface air temperature and sea ice variability,
 1379 *Journal of Climate*, 24(24), 6573–6581.
- 1380 Marshall, J., and F. Schott (1999), Open-ocean convection: observations, theory, and models,
 1381 *Reviews of Geophysics*, 37, 1–64.
- 1382 Martin, T., W. Park, and M. Latif (2013), Multi-centennial variability controlled by Southern
 1383 Ocean convection in the Kiel Climate Model, *Climate dynamics*, 40(7-8), 2005–2022.
- 1384 Massonnet, F., T. Fichefet, H. Goosse, C. M. Bitz, G. Philippon-Berthier, M. M. Hol-
 1385 land, and P.-Y. Barriat (2012), Constraining projections of summer Arctic sea ice, *The
 1386 Cryosphere*, 6(6), 1383–1394.
- 1387 McCarthy, G., D. Smeed, W. Johns, E. Frajka-Williams, B. Moat, D. Rayner, M. Baringer,
 1388 C. Meinen, J. Collins, and H. Bryden (2015), Measuring the Atlantic meridional overturning
 1389 circulation at 26 N, *Progress in Oceanography*, 130, 91–111.
- 1390 Menary, M. B., C. D. Roberts, M. D. Palmer, P. R. Halloran, L. Jackson, R. A. Wood, W. A.
 1391 Müller, D. Matei, and S. K. Lee (2013), Mechanisms of aerosol-forced amoc variability in
 1392 a state of the art climate model, *Journal of Geophysical Research: Oceans*, 118(4), 2087–
 1393 2096.
- 1394 Milly, P. C., S. L. Malyshev, E. Shevliakova, K. A. Dunne, K. L. Findell, T. Gleeson,
 1395 Z. Liang, P. Phillipps, R. J. Stouffer, and S. Swenson (2014), An enhanced model of land
 1396 water and energy for global hydrologic and earth-system studies, *Journal of Hydrometeorology*,
 1397 15(5), 1739–1761.
- 1398 Murakami, H., G. A. Vecchi, S. Underwood, T. L. Delworth, A. T. Wittenberg, W. G. An-
 1399 derson, J.-H. Chen, R. G. Gudgel, L. M. Harris, S.-J. Lin, et al. (2015), Simulation and
 1400 prediction of category 4 and 5 hurricanes in the high-resolution GFDL HiFLOR coupled
 1401 climate model, *Journal of Climate*, 28(23), 9058–9079.
- 1402 Murray, R. (1996), Explicit generation of orthogonal grids for ocean models, *Journal of
 1403 Computational Physics*, 126, 251–273.
- 1404 Newman, M., A. T. Wittenberg, L. Cheng, G. P. Compo, and C. A. Smith (2018), The ex-
 1405 treme 2015/16 El Niño, in the context of historical climate variability and change [in "Ex-
 1406 plaining extreme events of 2016 from a climate perspective"], *Bull. Amer. Meteor. Soc.*,
 1407 99(1), S16–S20, doi:10.1175/BAMS-D-17-0116.1.
- 1408 Putman, W. M., and S.-J. Lin (2007), Finite-volume transport on various cubed-sphere grids,
 1409 *Journal of Computational Physics*, 227(1), 55–78.
- 1410 Ray, S., A. T. Wittenberg, S. M. Griffies, and F. Zeng (2018a), Understanding the equatorial
 1411 Pacific cold tongue time-mean heat budget, Part I: Diagnostic framework, *J. Climate*, 31,
 1412 9965–9985, doi:10.1175/JCLI-D-18-0152.1.
- 1413 Ray, S., A. T. Wittenberg, S. M. Griffies, and F. Zeng (2018b), Understanding the equatorial
 1414 Pacific cold tongue time-mean heat budget, Part II: Evaluation of the GFDL-FLOR cou-
 1415 pled GCM, *J. Climate*, 31, 9987–10,011, doi:10.1175/JCLI-D-18-0153.1.
- 1416 Rayner, N., D. E. Parker, E. Horton, C. K. Folland, L. V. Alexander, D. Rowell, E. Kent, and
 1417 A. Kaplan (2003), Global analyses of sea surface temperature, sea ice, and night marine
 1418 air temperature since the late nineteenth century, *Journal of Geophysical Research: Atmo-
 1419 spheres*, 108(D14).
- 1420 Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang (2002), An im-
 1421 proved in situ and satellite SST analysis for climate, *J. Climate*, 15, 1609–1625.
- 1422 Ricker, R., S. Hendricks, V. Helm, H. Skourup, and M. Davidson (2014), Sensitivity of
 1423 CryoSat-2 Arctic sea-ice freeboard and thickness on radar-waveform interpretation, *The
 1424 Cryosphere*, 8(4), 1607–1622.
- 1425 Saba, V. S., S. M. Griffies, W. G. Anderson, M. Winton, M. A. Alexander, T. L. Delworth,
 1426 J. A. Hare, M. J. Harrison, A. Rosati, G. A. Vecchi, et al. (2016), Enhanced warming of
 1427 the Northwest Atlantic Ocean under climate change, *Journal of Geophysical Research:*

- 1428 *Oceans*, 121(1), 118–132.
- 1429 Shevliakova, E., S. W. Pacala, S. Malyshev, G. C. Hurtt, P. Milly, J. P. Caspersen, L. T. Sent-
- 1430 man, J. P. Fisk, C. Wirth, and C. Crevoisier (2009), Carbon cycling under 300 years of
- 1431 land use change: Importance of the secondary vegetation sink, *Global Biogeochemical*
- 1432 *Cycles*, 23(2).
- 1433 Shu, Q., Z. Song, and F. Qiao (2015), Assessment of sea ice simulations in the CMIP5 mod-
- 1434 els, *The Cryosphere*, 9(1), 399–409.
- 1435 Smeed, D., S. Josey, C. Beaulieu, W. Johns, B. Moat, E. Frajka-Williams, D. Rayner,
- 1436 C. Meinen, M. Baringer, H. Bryden, et al. (2018), The North Atlantic Ocean is in a state
- 1437 of reduced overturning, *Geophysical Research Letters*, 45(3), 1527–1533.
- 1438 Smith, G., K. Priestley, N. Loeb, B. Wielicki, T. Charlock, P. Minnis, D. Doelling, and
- 1439 D. Rutan (2011), Clouds and Earth Radiant Energy System (CERES), a review: Past,
- 1440 present and future, *Advances in Space Research*, 48(2), 254–263.
- 1441 Stan, C., M. Khairoutdinov, C. A. DeMott, V. Krishnamurthy, D. M. Straus, D. A. Randall,
- 1442 J. L. Kinter, and J. Shukla (2010), An ocean-atmosphere climate simulation with an em-
- 1443 bedded cloud resolving model, *Geophysical Research Letters*, 37(1).
- 1444 Stouffer, R. J. (2004), Time scales of climate response, *Journal of Climate*, 17(1), 209–217.
- 1445 Stroeve, J. C., V. Kattsov, A. Barrett, M. Serreze, T. Pavlova, M. Holland, and W. N. Meier
- 1446 (2012), Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations, *Geophys.*
- 1447 *Res. Lett.*, 39(16), doi:10.1029/2012GL052676.
- 1448 Swart, N. C., J. C. Fyfe, E. Hawkins, J. E. Kay, and A. Jahn (2015), Influence of internal
- 1449 variability on arctic sea-ice trends, *Nature Climate Change*, 5, 86–89.
- 1450 Talandier, C., J. Deshayes, A.-M. Treguier, X. Capet, R. Benshila, L. Debreu, R. Dussin, J.-
- 1451 M. Molines, and G. Madec (2014), Improvements of simulated Western North Atlantic
- 1452 current system and impacts on the AMOC, *Ocean Modelling*, 76, 1–19.
- 1453 Turner, J., T. J. Bracegirdle, T. Phillips, G. J. Marshall, and J. S. Hosking (2013), An initial
- 1454 assessment of Antarctic sea ice extent in the CMIP5 models, *Journal of Climate*, 26(5),
- 1455 1473–1484.
- 1456 Vecchi, G. A., T. Delworth, R. Gudgel, S. Kapnick, A. Rosati, A. T. Wittenberg, F. Zeng,
- 1457 W. Anderson, V. Balaji, K. Dixon, L. Jia, H. Kim, L. Krishnamurthy, R. Msadek, W. F.
- 1458 Stern, S. D. Underwood, G. Villarini, X. Yang, and S. Zhang (2014), On the seasonal fore-
- 1459 casting of regional tropical cyclone activity, *J. Climate*, 27, 7994–8016, doi:10.1175/JCLI-
- 1460 D-14-00158.1.
- 1461 Wang, H., S. A. Legg, and R. W. Hallberg (2015), Representations of the Nordic Seas over-
- 1462 flows and their large scale climate impact in coupled models, *Ocean Modelling*, 86, 76–92.
- 1463 Watanabe, S., T. Hajima, K. Sudo, T. Nagashima, T. Takemura, H. Okajima, T. Nozawa,
- 1464 H. Kawase, M. Abe, T. Yokohata, T. Ise, H. Sato, E. Kato, K. Takata, S. Emori, and
- 1465 M. Kawamiya (2011), MIROC-ESM 2010: model description and basic results of CMIP5-
- 1466 20c3m experiments, *Geoscientific Model Development*, 4, 845–872, doi:doi:10.5194/gmd-
- 1467 4-845-2011.
- 1468 Wittenberg, A. T. (2002), ENSO response to altered climates, Ph.D. thesis, Princeton Univer-
- 1469 sity, 475pp. Available at <http://www.gfdl.noaa.gov/~atw/research/thesis>.
- 1470 Wittenberg, A. T. (2009), Are historical records sufficient to constrain ENSO simulations?,
- 1471 *Geophys. Res. Lett.*, 36, L14,709, doi:10.1029/2009GL038710.
- 1472 Wittenberg, A. T., A. Rosati, N.-C. Lau, and J. J. Poshay (2006), GFDL's CM2 global cou-
- 1473 pled climate models. Part III: Tropical Pacific climate and ENSO, *J. Climate*, 19, 698–722,
- 1474 doi:10.1175/JCLI3631.1.
- 1475 Wittenberg, A. T., A. Rosati, T. L. Delworth, G. A. Vecchi, and F. Zeng (2014), ENSO mod-
- 1476 ulation: Is it decadally predictable?, *J. Climate*, 27, 2667–2681, doi:10.1175/JCLI-D-13-
- 1477 00577.1.
- 1478 Wittenberg, A. T., G. A. Vecchi, T. L. Delworth, A. Rosati, W. Anderson, W. F. Cooke,
- 1479 S. Underwood, F. Zeng, S. M. Griffies, and S. Ray (2018), Improved simulations of tropi-
- 1480 cal Pacific annual-mean climate in the GFDL FLOR and HiFLOR coupled GCMs, *J. Adv.*
- 1481 *Model. Earth Syst.*, 10, 3176–3220, doi:10.1029/2018MS001372.

- 1482 Yan, X., R. Zhang, and T. R. Knutson (2018), Underestimated AMOC variability and impli-
 1483 cations for AMV and predictability in CMIP models, *Geophysical Research Letters*, 45(9),
 1484 4319–4328.
- 1485 Yeager, S. G., and G. Danabasoglu (2012), Sensitivity of Atlantic meridional overturning cir-
 1486 culation variability to parameterized Nordic Sea overflows in CCSM4, *Journal of Climate*,
 1487 25, 2077–2103.
- 1488 Yeager, S. G., and M. Jochum (2009), The connection between Labrador Sea buoyancy loss,
 1489 deep western boundary current strength, and Gulf Stream path in an ocean circulation
 1490 model, *Ocean Modelling*, 30(2-3), 207–224.
- 1491 Zhang, L., T. L. Delworth, and L. Jia (2017), Diagnosis of decadal predictability of Southern
 1492 Ocean sea surface temperature in the GFDL CM2.1 model, *Journal of Climate*, 30(16),
 1493 6309–6328.
- 1494 Zhang, R. (2015), Mechanisms for low-frequency variability of summer Arctic sea ice extent,
 1495 *Proc. Natl. Acad. Sci.*, 112(15), 4570–4575.
- 1496 Zhang, R. (2017), On the persistence and coherence of subpolar sea surface temperature and
 1497 salinity anomalies associated with the Atlantic multidecadal variability, *Geophysical Re-
 1498 search Letters*, 44(15), 7865–7875.
- 1499 Zhang, R., and T. R. Knutson (2013), The role of global climate change in the extreme low
 1500 summer Arctic sea ice extent in 2012 [in “Explaining Extreme events of 2012 from a cli-
 1501 mate perspective”], *Bulletin of the American Meteorological Society*, 94(9), S23–S26.
- 1502 Zhang, R., and G. K. Vallis (2007), The role of bottom vortex stretching on the path of the
 1503 North Atlantic western boundary current and on the northern recirculation gyre, *Journal
 1504 of Physical Oceanography*, 37(8), 2053–2080.
- 1505 Zhang, R., T. L. Delworth, A. Rosati, W. G. Anderson, K. W. Dixon, H.-C. Lee, and F. Zeng
 1506 (2011), Sensitivity of the North Atlantic Ocean circulation to an abrupt change in the
 1507 Nordic Sea overflow in a high resolution global coupled climate model, *Journal of Geo-
 1508 physical Research: Oceans*, 116(C12).
- 1509 Zhang, R., T. L. Delworth, R. Sutton, D. R. Hodson, K. W. Dixon, I. M. Held, Y. Kush-
 1510 nir, J. Marshall, Y. Ming, R. Msadek, J. Robson, A. J. Rosati, M. Ting, and G. A. Vecchi
 1511 (2013), Have aerosols caused the observed Atlantic multidecadal variability?, *Journal of
 1512 the Atmospheric Sciences*, 70(4), 1135–1144.
- 1513 Zhang, R., R. Sutton, G. Danabasoglu, Y.-O. Kwon, R. March, S. G. yeager, D. E. Amrhein,
 1514 and L. C. M (2019), A Review of the Role of the Atlantic Meridional Overturning Circu-
 1515 lation in Atlantic Multidecadal Variability and Associated Climate Impacts, *Reviews of
 1516 Geophysics*, 57.
- 1517 Zhao, M., J.-C. Golaz, I. Held, H. Guo, et al. (2018a), The GFDL global atmosphere and
 1518 land model AM4. 0/LM4. 0: 1. Simulation characteristics with prescribed SSTs, *Journal
 1519 of Advances in Modeling Earth Systems*, 10(3), 691–734.
- 1520 Zhao, M., J.-C. Golaz, I. Held, H. Guo, et al. (2018b), The GFDL global atmosphere and
 1521 land model AM4. 0/LM4. 0: 2. Model description, sensitivity studies, and tuning strate-
 1522 gies, *Journal of Advances in Modeling Earth Systems*, 10(3), 735–769.
- 1523 Zweng, M. M., J. R. Reagan, J. I. Antonov, R. A. Locarnini, A. V. Mishonov, T. P. Boyer,
 1524 H. E. Garcia, O. K. Baranova, D. R. Johnson, D. Seidov, et al. (2013), World ocean atlas
 1525 2013. Volume 2, Salinity.
- 1526 Zweng, M. M., J. R. Reagan, D. Seidov, T. P. Boyer, R. A. Locarnini, H. E. Garcia, A. V.
 1527 Mishonov, et al. (2018), World ocean atlas 2018. Volume 2, Salinity.