

The GFDL Earth System Model version 4.1 (GFDL-ESM4.1): Overall coupled model description and simulation characteristics

J. P. Dunne¹, L. W. Horowitz¹, A. J. Adcroft², P. Ginoux¹, I. M. Held², J. G. John¹, J. P. Krasting¹, S. Malyshev¹, V. Naik¹, F. Paulot¹, E. Shevliakova¹, C. A. Stock¹, N. Zadeh³, V. Balaji², C. Blanton³, K. A. Dunne⁴, C. Dupuis³, J. Durachta¹, R. Dussin⁶, P. P. G. Gauthier⁵, S. M. Griffies^{1,2}, H. Guo^{1,6}, R. W. Hallberg¹, M. Harrison¹, J. He², W. Hurlin¹, C. McHugh³, R. Menzel⁶, P. C. D. Milly⁴, S. Nikonov², D. J. Paynter¹, J. Ploshay¹, A. Radhakrishnan², K. Rand³, B. G. Reichl¹, T. Robinson³, D. M. Schwarzkopf¹, L. T. Sentman¹, S. Underwood¹, H. Vahlenkamp⁶, M. Winton¹, A. T. Wittenberg¹, B. Wyman¹, Y. Zeng⁷, and M. Zhao¹

¹NOAA Geophysical Fluid Dynamics Laboratory, Princeton NJ USA

²Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton NJ USA

³SAIC/GFDL Princeton NJ, USA

⁴U. S. Geological Survey/GFDL, Princeton NJ USA

⁵Princeton Environmental Institute, Princeton University, Princeton NJ 08540

⁶UCAR/GFDL Princeton NJ USA

⁷Cooperative Institute for Modeling the Earth System, Princeton University, Princeton NJ 08540

Corresponding author: John Dunne (John.Dunne@noaa.gov)

Key Points:

A new coupled chemistry-carbon-climate Earth system model has been developed at the Geophysical Fluid Dynamics Laboratory.

This model unifies component advances in chemistry, carbon, and ecosystem comprehensiveness within a single coupled climate framework.

This model features much improved climate mean patterns and variability from previous chemistry and carbon coupled models.

Abstract

We describe the baseline coupled model configuration and simulation characteristics of GFDL's Earth System Model version 4.1 (ESM4.1) which builds on component and coupled model developments at GFDL over 2013-2018 for coupled carbon-chemistry-climate simulation contributing to the sixth phase of the Coupled Model Intercomparison Project. In contrast with GFDL's CM4.0 development effort which focuses on ocean resolution for physical climate, ESM4.1 focuses on comprehensiveness of Earth system interactions. ESM4.1 features doubled horizontal resolution of both atmosphere (2° to 1°) and ocean (1° to ½°) relative to GFDL's previous-generation coupled ESM2-carbon and CM3-chemistry models. ESM4.1 brings together key representational advances in CM4.0 dynamics and physics along with those in

aerosols and their precursor emissions, land ecosystem vegetation and canopy competition and multi-day fire; ocean ecological and biogeochemical interactions, comprehensive land-atmosphere-ocean cycling of CO₂, dust and iron, and interactive ocean-atmosphere nitrogen cycling described in detail across this volume of JAMES and presented here in terms of the overall coupling and resulting fidelity. ESM4.1 provides much improved fidelity in CO₂ and chemistry over ESM2 and CM3, captures most of CM4.0's baseline simulations characteristics, and notably improves on CM4.0 in: 1) Southern Ocean mode and intermediate water ventilation, 2) Southern Ocean aerosols, and 3) reduced spurious ocean heat uptake. ESM4.1 has reduced transient and equilibrium climate sensitivity compared to CM4.0. Fidelity concerns include: 1) moderate degradation in sea surface temperature biases, 2) degradation in aerosols in some regions, and 3) strong centennial scale climate modulation by Southern Ocean convection.

Plain Language Summary

GFDL has developed a coupled chemistry-carbon-climate Earth System Model (ESM4.1) as part of its 4th generation coupled model development activities with model results contributed publicly to the 6th phase of the Coupled Model Intercomparison Project. With similar computational expense as GFDL's first coupled model CM4.0, ESM4.1 focuses on chemistry and ecosystem comprehensiveness rather than the ocean resolution-focus of CM4.0. With fidelity near to that of CM4.0, ESM4.1 features much improved representation of climate mean patterns and variability from previous GFDL ESMs as well as comprehensive couplings for chemistry, carbon, and dust.

1 Introduction

The fifth phase of the Coupled Model Intercomparison Project (CMIP5) [Taylor *et al.*, 2012] provided an unprecedented opportunity for not only the international provision and intercomparison of coupled carbon-climate models to significantly decrease uncertainty in known coupled carbon-climate feedbacks [Arora *et al.*, 2013], but it also served to bring to the forefront an array of formerly underappreciated sources of uncertainty including soil and nitrogen processes [Todd-Brown *et al.*, 2013; Randerson *et al.*, 2015], land use interactions [Brovkin *et al.*, 2013], and the role of the Southern Ocean on heat and carbon uptake [Frölicher *et al.*, 2015]. Further, idealized scenarios of coupled carbon climate feedbacks were used to explore earth system evolution under scenarios of long term commitment to CO₂ emissions [Krasting *et al.*, 2016; Palter *et al.*, 2018] and CO₂ reversal [Boucher *et al.*, 2012; John *et al.*, 2015] that were previously only possible with simplified Earth System Models of Intermediate Complexity (EMICS). As such, GFDL-ESM2 series models, ESM2M and ESM2G [Dunne *et al.*, 2012; 2013], served as important contributors to these intercomparisons bringing new robustness to the characterization of coupled carbon climate feedbacks with land [Shevliakova *et al.*, 2014] and ocean [Frölicher *et al.*, 2014, Resplandy *et al.*, 2018] and across the Earth as a system [Randerson *et al.*, 2015]. Additionally, these models were found profoundly useful in serving a broad range of ocean ecological and biogeochemical impact studies [e.g. Bopp *et al.*,

2013] relating to acidification [Resplandy *et al.*, 2013; Gehlen *et al.*, 2014], hypoxia, and living marine resource implications [Blanchard *et al.*, 2012; Lotze *et al.*, 2019].

Similarly in the area of the coupled chemistry-climate system, several groups including GFDL also contributed models with comprehensive atmospheric chemistry including interactive ozone over the full atmospheric domain [CM3; Donner *et al.*, 2011] and were able to detail the strong lingering uncertainties in robust characterization of coupled chemistry-climate interactions such as the aerosol precursor emission rates and efficiencies as cloud condensation nuclei and aerosol indirect effects [Golaz *et al.*, 2011]. This model was found to have a high overall climate sensitivity to both aerosols and greenhouse gases with moderate cooling in the latter half of the 20th century followed by rapid warming around the turn of the 21st century [e.g. Golaz *et al.*, 2013]. Considerable effort has gone into attributing the sources of this model behavior [Golaz *et al.*, 2013; Zhao, 2014; Zhao *et al.*, 2016]. These models were recently formally compared in the separate AeroCOM and ACCMIP intercomparison [e.g. Naik *et al.*, 2013; Shindell *et al.*, 2013], and have broad applications beyond chemistry-climate implications to air quality, including detection and attribution of extremes in ozone [Lin *et al.*, 2012], and particulate loading [Westervelt *et al.*, 2016] and changes in aerosol forcing [Paulot *et al.*, 2018].

There have been considerable developments at GFDL in both coupled carbon-climate and coupled chemistry-climate modeling to advance understanding of the changing climate system in recent years. Physical developments include the ability to draw on larger high performance computing resources to increase resolution, and to improve dynamical and physical parameterizations in atmosphere [Zhao *et al.*, 2018], ocean and sea ice [Adcroft *et al.*, 2019], as well as land and overall coupling [Held *et al.*, 2019]. On land, GFDL developments on a broad spectrum of biogeochemical and ecological fronts have occurred including the key role of mineral protection in soils [Sulman *et al.*, 2014], the role of pasture and grazing in land use, the heterogeneities involved in single-daily, multi-day and crown fire [Rabin *et al.* 2015; Ward *et al.*, 2018], the dynamic role of land vegetation and hydrology in dust emissions [Evans *et al.*, 2016], the role of competition in the development of resource allocation and biodiversity in forest canopy [Purves *et al.*, 2008; Weng *et al.*, 2015]. Ocean biogeochemical developments at GFDL include enhanced representations of plankton ecology and food web dynamics controlling the flow of carbon and energy between phytoplankton and fish [Stock *et al.*, 2014, 2017, 2019], more robust characterization of controls on the biological pump [Laufkötter *et al.*, 2017] and revised carbon chemistry [Orr *et al.*, 2015]. Atmospheric chemistry efforts at GFDL have resulted in implementation of reduced nitrogen cycling [Paulot *et al.*, 2016], improved representation of the seasonal cycle in sulfate [Paulot *et al.*, 2017], and improved representation of ozone precursors, sulfur dioxide and carbonaceous aerosols [Naik *et al.*, 2013].

Over this same time, the vision of the Coupled Model Intercomparison Project evolved from being centered on a few key idealized, historical and future projection scenarios in CMIP5 to a

broad and nested structure of model intercomparison in CMIP6 [Eyring *et al.*, 2016]. CMIP6 explicitly couples the general baseline intercomparison in the Diagnostic, Evaluation and Characterization of Klima (DECK) to a suite of over 20 endorsed model intercomparison projects covering independent and complementary scientific questions. The range of questions and topics considered span response to forcing, systematic bias, variability, predictability, and future scenarios as well as regional and mechanistic foci. As such, GFDL chose to change its strategy of participation from contributing to each of a limited set of scientific questions with separate GFDL models to maximize the applicability of a single model to the broadest suite of endorsed intercomparisons. ESM4.1 was thus developed with the intention of having sufficient comprehensiveness, numerical efficiency, and fidelity to participate as a state of the art contribution to intercomparisons of projection scenarios (ScenarioMIP; [O'Neill *et al.*, 2016]), the coupled climate carbon cycle (C4MIP; [Jones *et al.*, 2016]), atmospheric chemistry and aerosols (AerChemMIP; [Collins *et al.*, 2017]), land use (LUMIP; [Lawrence *et al.*, 2016]), ocean biogeochemistry (OMIP-BGC; [Orr *et al.*, 2017]), carbon dioxide removal (CDRMIP; [Keller *et al.*, 2018]), radiative forcing (RFMIP; [Pincus *et al.*, 2016]), and detection and attribution (DAMIP; [Gillett *et al.*, 2016]) among others.

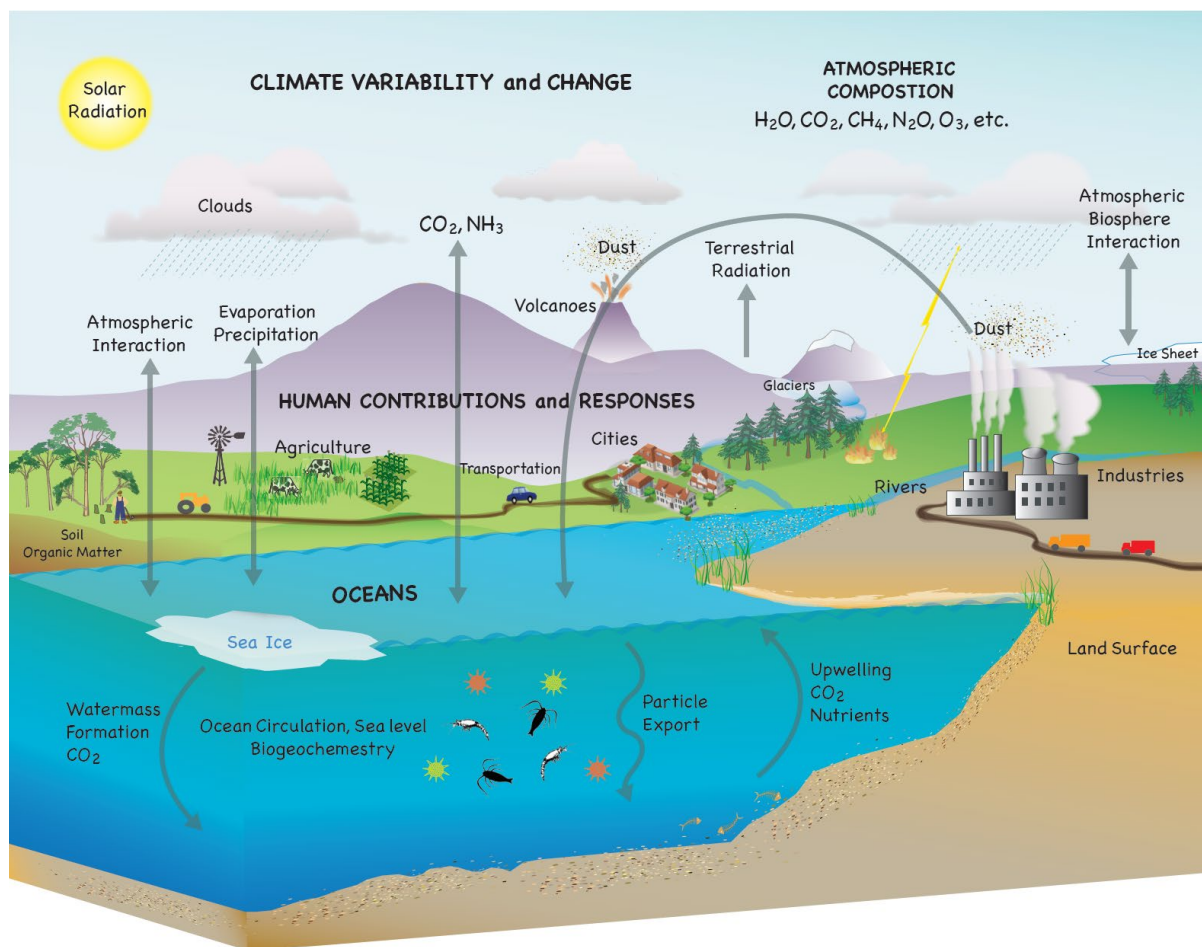


Figure 1: Schematic of major natural and anthropogenic processes and influences on the climate system including CO₂, dust, iron, and nitrogen interactions between Earth system components.

The Earth System Model (ESM) tool described here is provided in schematic form in Figure 1. The atmospheric component of the ESMs includes physical features such as aerosols (both natural and anthropogenic), cloud physics, and precipitation. The land component includes terrestrial hydrology and ecology to simulate dynamic reservoirs of surface snow and carbon pools as well as soil temperature, liquid and frozen water, and carbon and runoff through streams, lakes, and rivers. The ice-ocean component includes water fluxes, currents, interactive sea ice dynamics, iceberg transport of freshwater, vertical and lateral mixing as well as marine biogeochemistry and ecology. ESMs respond to a diverse suite of emissions and other forcing including greenhouse gas and aerosol chemistry, radiation and physics interactions and variations of land surface albedo due to both natural vegetation changes and land use history and include representation of impacts and feedbacks.

This work describes the development of GFDL-ESM4.1 from its contributory components and focuses on baseline simulation characteristics and idealized and historical climate response to forcing with particular emphasis on comparison with GFDL's CM4.0 [Held *et al.*, 2019; Winton *et al.*, 2020] and GFDL's previous generation models of coupled carbon-climate [Dunne *et al.*, 2012; 2013] and coupled chemistry climate (CM3; [Donner *et al.*, 2010]).

2 Methods

2.1 Component composition in contrast to CM4.0, CM3 and ESM2

A design criterion for ESM4.1 was that the model's computational expense be roughly equivalent to that of GFDL's fourth generation coupled physical climate model [CM4.0; Held *et al.*, 2019], but with emphasis on Earth system comprehensiveness rather than ocean horizontal resolution. Both models build from a basis of GFDL's AM4.0 atmospheric model [Zhao *et al.*, 2018] using the finite volume version 3 [FV3; Lin, 2004] dynamical core with critical developments in parameterization of moist convection, clouds, radiation, topographical drag and other physical parameterizations. Both models use GFDL's open source Modular Ocean Model version 6 [MOM6; Adcroft *et al.*, 2019] configured with 75 vertical hybrid coordinate layers within the Arbitrary-Lagrangian-Eulerian algorithm [Adcroft and Hallberg, 2006]. Both models use a new version of GFDL's Sea Ice Simulator [SIS2; Adcroft *et al.*, 2018]. The computational cost of both models is approximately 11K cpu-hours per simulation year on NOAA's Gaea supercomputer (Cray XE6) for a little over 7 simulation years per day on 3400 cores. This is about 12 times the cost of CM3 and 36 times the cost of ESM2M and ESM2G per simulation year.

The contrasts between these models are summarized in Table 1. CM4.0 [Held *et al.*, 2019] emphasizes $\frac{1}{4}^\circ$ horizontal ocean resolution [Adcroft *et al.*, 2019] with only a highly simplified carbon cycle of Biogeochemistry with Light, Irradiance, Nutrient and Gas version 2 [BLINGv2; Dunne *et al.*, in review] and an atmosphere with ~ 100 km horizontal resolution and 33 vertical layers and interactive aerosols but externally forced ozone and other chemistry [Zhao *et al.*, 2018) with approximately three quarters of its computational demand on the ocean. In contrast, ESM4.1 is configured to allocate approximately 2/3 of the computational burden on the atmosphere through the combination of increased number of atmospheric prognostic tracers (5 water tracers and a radon tracer in both models, plus 82 chemistry tracers in ESM4.1 compared to 21 chemistry tracers in CM4.0) and increased number of layers to better resolve the stratosphere (49 compared to 33 in CM4.0), and a more advanced land [Shevliakova *et al.*, in review] and ocean biogeochemical models [Stock *et al.*, in review]. ESM4.1 accomplishes this improved atmospheric, land, and ocean biogeochemical comprehensiveness by sacrificing horizontal ocean resolution from $\frac{1}{4}^\circ$ to $\frac{1}{2}^\circ$.

Table 1: Summary comparison of component level differences in ESM4.1 from CM4.0 along with a brief statement of motivation. In some cases, only one or a few parameters were changed. In others, whole components were changed as discussed as in the text and references provided.

Component	CM4.0	ESM4.1	Motivation
Atmospheric Dynamics and Physics (<i>Horowitz et al.</i> , in review)			
Vertical grid	33 levels to 100 Pa (~ 45 km)	49 levels to 1 Pa (~ 80 km)	Representation of stratosphere dynamics and chemistry
Convective Gravity Wave Drag	$bt_0=0.005$, $bt_{nh}=0.0025$, $bt_{sh}=-0.0002.5$	$bt_0=0.004$, $bt_{nh}=0.001$, $bt_{sh}=-0.0005$	Account for higher top
Ice fall speed scaling	0.9	0.85	Longwave Radiative tuning
Tiedke cloud erosion scaling	6.9 hours	5.6 hours	Shortwave Radiative tuning
Atmospheric Aerosols and Chemistry (<i>Horowitz et al.</i> , in review)			
Ozone chemistry	prescribed	Comprehensive with 55 gas phase + 6 ideal tracers including reactive nitrogen and sulfur cycles	Ozone, nitrogen, and other chemistry-climate interactions and air quality applications

Dust emissions	Based on surface wind with fixed source areas	Based on surface wind, snow fraction, soil ice, surface LAI and SAI bareness, and land use as simulated for each LM4.1 tile	Internal consistency between components for variability, comprehensiveness
Sulfate	Simple SO ₄ with prescribed oxidants	Part of comprehensive chemistry	Comprehensiveness and fidelity
Sea Salt emissions	<i>Monahan et al. [1986], Martensson et al. [2003]</i>	<i>Monahan et al. [1986], Jaeglé et al. [2011]</i>	Address high aerosol biases in Southern Ocean
Aerosol Deposition	Different schemes for dust, sea salt, Black Carbon, and SO ₄	Consistent scheme for dust, sea salt, Black Carbon, and SO ₄	Internal consistency
Ocean (<i>Adcroft et al., 2019; Stock et al., in review</i>)			
Horizontal resolution	1/4° Mercator with enhanced polar resolution (1440x1080)	1/2° Mercator with enhanced equatorial and polar resolution (720x576)	Computational efficiency
Mesoscale eddies	No parameterization	Mesoscale eddy kinetic energy parameterization	Preserve thermocline stratification at lower resolution
Mixed layer eddy parameterization	Frontal length = 500	Front length = 200	Preserve near-surface stratification and SST
Biogeochemistry	BLINGv2 (6 tracers) for C, Alk, P, O ₂ with steady state ecosystem	COBALTv2 (33 tracers) for C, N, P, Alk, Si, Fe, Lith, O ₂ with prognostic ecosystem	Application to ecosystems and biogeochemistry
Shortwave penetration	Prescribed from satellite climatology	Interactive	Representation of bio-climate feedback
Atmospheric CO ₂ exchange	Externally prescribed atmospheric CO ₂	Either interactive or restored to externally prescribed value	Coupled carbon-climate interactions and feedbacks
Other Atmospheric exchange	Iron deposition prescribed from climatology	Interactive deposition of dust, iron, NO ₃ , NH ₃ plus NH ₃ gas exchange	Internal consistency between components for variability, comprehensiveness

Land (<i>Shevliakova et al.</i> , 2020)			
Snow/Soil/ Hydrology	Lower albedo of snow on glacier	Higher albedo of snow on glacier	Stimulate ocean heat loss through Antarctic coastal polynya
Stomatal Conductance	Leung, 1995	Wolfe et al., 2016	Mechanistic drought response
Radiation	Single leaf	Canopy dynamics	Internal consistency and comprehensiveness
Soil Carbon	Bulk	Vertically resolved	Improved soil microbial dynamics
Land Ecosystem	Single leaf canopy dynamics	Perfect Plasticity Approximation canopy dynamics	Application to ecosystems and biogeochemistry/ comprehensiveness
Fire	Annual	Daily, including crown/canopy and multi- day fires	Application to variability and comprehensiveness
Dust	Emission based on surface wind with fixed source areas	Emission based on surface wind, snow fraction, soil ice, surface LAI and SAI bareness, and land use as simulated for each LM4.1 tile	Internal consistency between components for variability, comprehensiveness
Atmospheric CO ₂ exchange	Externally prescribed	Either interactive or restored to externally prescribed value	Coupled carbon-climate interactions and feedbacks

2.2 Atmospheric chemistry

The differences in the ESM4.1 atmospheric component (AM4.1) from the CM4.0 atmospheric component (AM4.0) described in *Zhao et al.* [2018] are described in detail in *Horowitz et al.* [in review] including interactive land-atmosphere ocean interactions with dust and iron, and interactive atmosphere-ocean interactions with nitrogen. As such, we only highlight the substantive differences here. AM4.1 includes interactive tropospheric and stratospheric gas-phase and aerosol chemistry. The bulk aerosol scheme, including 18 transported aerosol tracers, is similar to that in AM4.0 [*Zhao et al.*, 2018], with the following updates: (1) ammonium and nitrate aerosols are treated explicitly, with ISORROPIA [*Fountoukis and Nenes*, 2007] used to

simulate the sulfate–nitrate–ammonia thermodynamic equilibrium, as described by *Paulot et al.* [2016]; (2) oxidation of sulfur dioxide and dimethyl sulfide to produce sulfate aerosol is driven by the gas-phase oxidant concentrations (OH, H₂O₂, and O₃) and cloud pH simulated by the online chemistry scheme [*Paulot et al.*, 2016, 2017], and (3) the rate of aging of black and organic carbon aerosols from hydrophobic to hydrophilic forms varies with calculated concentrations of hydroxyl radical (OH), as described by *Liu et al.* [2011].

Sources of secondary organic aerosols (SOA) include an anthropogenic source from oxidation of the simulated C₄H₁₀ hydrocarbon tracer by hydroxyl radical (with 10% per-carbon yield) and a biogenic pseudo-emission assuming a 10% per-carbon yield from emissions of biogenic volatile organic compounds (BVOCs; isoprene and monoterpenes) from vegetation, consistent with the *Dentener et al.* [2006] assumption of a 15% yield (by mass) from terpenes (and none from isoprene) and more recent studies of *Pai et al.*, [2020] that derived a 3% yield for isoprene and 10% for terpenes using a simple scheme and *Bates and Jacob* [2019] that required a 13% yield for isoprene using a more complex scheme to give a similar global burden. The BVOC emissions are calculated online in the AM4.1 using the Model of Emissions of Gases and Aerosols from Nature [MEGAN; *Guenther et al.*, 2006, 2012], as a function of simulated air temperature and shortwave radiative fluxes.

Unlike AM4.0, the AM4.1 model has an online representation of gas-phase tropospheric and stratospheric chemistry. The tropospheric chemistry includes reactions of the NO_x-HO_x-O_x-CO-CH₄ system and oxidation schemes for other non-methane volatile organic compounds. The stratospheric chemistry accounts for the major ozone loss cycles (O_x, HO_x, NO_x, ClO_x, and BrO_x) and heterogeneous reactions on liquid and solid stratospheric aerosols as in *Austin et al.* [2013]. The base chemical mechanism is updated from that in AM3 [*Naik et al.*, 2013], using gas-phase and heterogeneous chemistry updates from *Mao et al.* (2013a, b), similar to the configuration described by *Schnell et al.* (2018). The combined tropospheric and stratospheric chemistry scheme includes 56 prognostic (transported) tracers and 36 diagnostic (non-transported) chemical tracers, with 43 photolysis reactions, 190 gas-phase kinetic reactions, and 15 heterogeneous reactions. The chemical system is solved using an implicit Euler backward method with Newton-Raphson iteration, as in *Horowitz et al.*, [2003; *Naik et al.*, 2013]. Photolysis rates are calculated interactively using the FAST-JX version 7.1 code, as described by *Li et al.* [2016], accounting for the radiative effects of simulated aerosols and clouds.

2.3 Land hydrology and ecosystems

Land hydrology and ecosystem dynamics are represented on the same 1° grid as AM4.1 by the GFDL Land Model version 4.1 (LM4.1; [*Shevliakova et al.*, in review]) which builds off of the previous generation land model through updated soil types based on the Harmonized World Soil Database (HWSD) [*Nachtergaele et al.*, 2014], improved radiative properties for vegetation, soil,

and snow, and updated hydrology in LM4.0 [Held *et al.*, 2019]. In addition, LM4.1 includes a much more sophisticated treatment than LM4.0 of radiation within its multistory canopy and revised transpiration and stomatal conductance based on Wolfe *et al.* [2016] rather than the Leuning [1995] approach used in LM4.0. Soil carbon dynamics and biogeochemistry represented through the Carbon, Organisms, Rhizosphere, and Protection in the Soil Environment (CORPSE) model [Sulman *et al.*, 2014; 2019]. LM4.1 also includes a new fire model that calculates ignites fires daily and includes both multi-day and crown fires (Ward *et al.*, 2018) rather than the annual figures used in LM4.0. Vegetation dynamics are represented through the Perfect Plasticity Approximation (PPA; Purves *et al.* 2008; Weng *et al.*, 2015] rather than the simple, broad leaf approximation used in LM4.0. Like LM4.0, there are 6 live carbon pools in LM4.1 representing leaves, fine roots, heartwood, sapwood, seeds, and non-structural carbon (i.e. sugars). Litter is broken into leaf and coarse wood categories as well into fast and slow timescale partitions. Whereas LM4.0 included only a single bulk soil carbon pool, each of the 20 vertical soil levels in LM4.1 also represents separate fast and slow soil carbon pools along with two carbon storage pools associated with soil microbes and microbial products. Like LM4.0, LM4.1 includes 6 plant functional types in representing C3 grass, C4 grass, tropical trees, temperate deciduous trees, and cold evergreen trees. Unlike some of the variants of LM3 (e.g. Sulman *et al.*, 2019], LM4.1 does not include an interactive nitrogen cycle. As none of the GFDL land models formally represent long-term permafrost cycling, final tuning of land carbon in LM4.1 included setting a maximum lifetime for soil at 500 years to prevent the potential for accumulation of recalcitrant material under cold, dry conditions and achieve land carbon equilibration. Land use in LM4.1 occurs via annual wood harvesting which is divided into secondary tiles by age, crop planting and harvesting at prescribed times of year, and daily grazing on pastures (5% per day), as well as rangelands (vegetation taller than 3 m not being grazed) that were not included in the previous generation models. Further details can be found in Shevliakova *et al.* [in review].

2.4 Interactive dust

Dust emissions from land are a dynamic function of vegetation cover, soil moisture and wind speed after Evans *et al.* [2016] with retuning of emissions and wet scavenging factors. These factors have been tuned by comparing surface concentration with observations at maritime sites worldwide [Savoie and Prospero, 1989; Zuidema *et al.*, 2019] to align the regression between model and observed values - the emission and wet scavenging factors modulated to control high values near the source areas and low values in remote regions, respectively against simulation of the first decade of the 21st century for the natural sources. To constrain emission from pasture and cropland, which have different sets of parameters, additional decade-long runs were performed to simulate the transition of wet to dry Sahel by the end of the 1960s. This transition produced a sharp increase of dust from Africa from long term, high quality reference observations in Barbados [Prospero and Lamb, 2003]. Source areas are calculated as the

negative exponential of the Leaf Area Index (LAI) and ten times the Stem Area Index (SAI), both calculated online by LM4.1. Emitted dust is partitioned into five atmospheric tracers based on size with each tracer receiving a fixed percentage (radius bins 0.1–1 μm - 5%, 1–2 μm - 15%, 2–3 μm - 30%, 3–6 μm - 27%, and 6–10 μm - 23%). Each atmospheric dust tracer is transported by advection, vertical diffusion, and gravitational settling and both absorbs and scatters radiation. Dust affects cloud properties in remote regions where sulfate number concentration is low [Donner *et al.*, 1996] but is not chemically active. Atmospheric dust is removed through gravitational settling, dry deposition at the surface from turbulence, and wet deposition from both in-cloud scavenging and below-cloud scavenging.

2.5 Ocean and sea ice physics

Like CM4.0, the ESM4.1 ocean physical model is based on MOM6 and includes nominal resolution of $1/2^\circ$ horizontally (720 zonal points, 576 meridional points) on an Arakawa C-grid with a baseline Mercator projection but enhanced meridional resolution between 30°N – 30°S down to 0.26° between 5°N and 5°S (169 grid points between 30°S – 30°N) with a tripolar grid in the arctic north of 65°N with a nominal resolution of 0.21° (118 grid points north of 65°N) and fixed resolution of 0.19° south of 68°S (54 grid points between 80°S – 68°S). The Sea Ice Simulator version 2 with the same Arakawa C-grid. ESM4.1 uses the same 75 vertical hybrid depth-isopycnal layers as CM4.0. A detailed comparison of these ocean and sea ice model configurations, parameterizations, and fidelity is provided in Adcroft *et al.* [2019].

2.6 Ocean biogeochemistry and ecosystems

The ocean biogeochemical component of ESM4.1 is version 2 of the Carbon, Ocean Biogeochemistry and Lower Trophics (COBALTv2) model [Stock *et al.*, 2014, in review]. COBALTv2 uses 33 tracers to represent carbon, alkalinity, oxygen, nitrogen, phosphorus, iron, silica, calcite and lithogenic mineral cycling within the ocean. The primary structural difference between COBALTv2 and biogeochemical dynamics from GFDL’s past ESM2 series of models (TOPAZ, [Dunne *et al.*, 2013]) is that COBALTv2 includes an enhanced representation of plankton food web dynamics to resolve the flow of energy from phytoplankton to fish [Stock *et al.*, 2014b, 2017] and enhance the model’s capacity to resolve linkages between food webs and biogeochemical cycles. COBALTv2 explicitly includes small, large (split into diatoms and non-diatoms), and diazotrophic phytoplankton groups, three zooplankton groups, bacteria, and three labilities of dissolved organic matter. Other updates relative to TOPAZ include adding temperature-dependence to sinking organic matter remineralization [Laufkötter *et al.*, 2017], the addition of semi-labile dissolved organic material, and a growing number of dynamical linkages to other ESM components including interactive dust and associated iron, NO_3 and NH_3 wet and dry deposition and NH_3 gas exchange in addition to CO_2 gas exchange (Table 1). Notable additions include reduced inorganic nitrogen (ammonia/ammonium) exchanges across the air-sea

interface [Paulot *et al.*, 2015, in review], dust and iron supplies linked to terrestrial and atmospheric conditions [Evans *et al.*, 2016], and iron sources from icebergs [Laufkötter *et al.*, 2018]. The most noteworthy simplification relative to TOPAZ is COBALTv2's reliance on fixed nitrogen to phosphorus ratios for plankton function types.

Carbonate chemistry and thermodynamic calculations in COBALTv2 were upgraded to use the open-source Model of the Ocean Carbonate SYstem version 2.0 (Mocsy, [Orr *et al.* 2015]). Improvements relative to the OCMIP2 formulation used in previous generation GFDL models include an improved solver for computing the hydrogen ion concentration that performed better than the older iterative-style solver, particularly in high latitude regions where freshwater input from rivers and sea ice melt led to unrealistically high values of pH seen in previous generation models. Mocsy was configured to use the Lee *et al.* [2010] formulation for total boron, Millero [2010] for the K1 and K2 solubility coefficients, and the Dickson and Riley [1979] formulation for Kf.

2.7 Coupling

TheESM4.1 atmosphere, land, sea ice and main coupler run on one set of processors while the ocean run concurrently on a second set of processors with lagged flux coupling to the ocean handled through the sea ice. The main dynamical atmosphere and land time step is 30 minutes with radiation, hydrology, CO₂ and dust emission coupling at each time step through an implicit solver. Within the atmosphere, the radiation time step is 3 hours. Radiation, hydrology, and tracer coupling with the sea ice and ocean is performed once an hour through the land-atmosphere-sea-ice-ocean coupler with fluxes accumulated and applied at the ocean tracer time step of 2 hours. Wet and dry atmospheric deposition of dust and associated iron, NO₃ and NH₃ are provided from the atmosphere to the ocean every land-atmosphere-sea-ice-ocean coupling time step with gas exchange of CO₂ and NH₃ handled similarly (O₂ flux is also handled in ocean, but with a fixed atmospheric concentration). River runoff of liquid water is delivered from the land to the ocean at every land-atmosphere-sea-ice-ocean coupling time step while frozen water is collected and delivered to the ocean as individual calving icebergs after Adcroft *et al.* [2019].

2.8 Conservation

As in CM4.0 [Held *et al.*, 2019], ESM4.1 conserves tracers in atmosphere, land, sea ice and ocean to numerical precision. The limits of this conservation is best seen in the case of total salt. In the preindustrial control simulation (piControl), the ESM4.1 ocean is initialized with $4.7967518803517374 \times 10^{19}$ kg of salt. The total amount of salt in sea ice varies the ocean inventory on the order of 10^{10} kg. Total drift in ocean salt content over the 900 year piControl through accumulated errors in the numerics made for an addition of 4×10^{13} kg overall or 4×10^{10} kg yr⁻¹. While this is a large amount of salt in the absolute sense, it is only one part in a million

overall, or one part in a billion per year which we consider acceptable for such millennial scale simulations. Also as in CM4.0, ESM4.1 has an artificial energy sink in the model atmosphere of 0.082W m^{-2} in the piControl. This calculation is based on the long term difference between the net incoming radiative flux at the top of the atmosphere (shortwave down minus shortwave up minus outgoing longwave radiation) and the surface (shortwave and longwave down minus shortwave and longwave up radiation minus sensible, evaporation and frozen precipitation heat fluxes). Analysis of this sink indicates that it is associated with the atmospheric numerics dissipating kinetic energy and is balanced by a uniformly distributed energy- conservation fix as described in Held et al. [2019].

2.9 Logistics of ESM4.1 development

As is typical in high performance computing applications, throughput limitations were a key factor throughout development. Using 3384 cores, the ESM4.1 preindustrial control simulation (piControl) with the extensive production diagnostic suite (output file size 221 Gb per simulated year) was able to simulate 6.9 model years per day with continuous integration at a computational cost of 12K cpu-hours per simulated year. At a higher count of 4896 cores, the same configuration was able to simulate 8.4 model years per day with continuous integration at a computational cost of 14K cpu-hours per simulated year. Considerable variability in throughput was found as a function of diagnostics, with a configuration using a reduced diagnostics suite (46 Gb per simulated year) affording throughputs over 10 model years per day while higher diagnostic loading during the historical simulation (188 Gb per simulated year) dropped the 4896 core configuration to 7.5 model years per day with continuous integration. With a computational budget of roughly 5M cpu-hours per month, this effort could afford approximately 400 years of total allocated ESM4.1 integration per month—roughly two continuous streams of integration. As such, this ultimate ESM4.1 configuration was not a suitable platform to support independent efforts to explore and revise all the different Earth system dimensions of the ESM4.1 development effort.

Instead, this effort required a high degree of synchrony between collaborators as they each worked on component-level and coupled prototypes. While the starting points for GFDL's 4th generation coupled carbon-chemistry-climate effort began from past efforts in coupled carbon-climate (ESM2) and chemistry-climate (CM3), the dynamical and physical cores of this effort were largely redesigned in an ongoing effort when ESM4.1 development began. The early stages of component development towards ESM4.1 leveraged the main laboratory AM4.0/OM4.0/CM4.0 effort. For example, as discussed in *Held et al.* [2019], a series of early prototypes for the 33-layer, fixed-ozone AM4.0 atmosphere were run over a $\frac{1}{2}$ degree ocean. This approach held the advantage of a much lower computational cost than CM4.0 (approximately one third), suitable to a diverse array of development experiments, and was key to paving the way to subsequent ESM4.1 development.

In September 2017, as GFDL's CM4.0 development effort was coming to closure, an Earth System working group was formed to coordinate ESM4.1 development activities. To efficiently and effectively leverage these ongoing developments of the OM4 and CM4 physical models, ESM4.1 development progressed on three initially separate tracks in ocean, atmosphere, and land before being incrementally merged. For ocean physical tuning and implementation of the COBALTv2 ocean biogeochemistry [Stock *et al.*, in review], development was focused on the prototype coupled AM4.0 run over a $\frac{1}{2}$ degree ocean described above. Separately, a prototype 49-layer full chemistry AM4.1 AMIP configuration [Horowitz *et al.*, in review] was developed based on AM4.0 [Zhao *et al.*, 2018a, 2018b] and later coupled to the maturing $\frac{1}{2}$ degree ocean. Both of the aforementioned intermediate configurations were coupled to the CM4.0 land model, LM4.0. Development of the ESM4.1 land model, LM4.1 [Shevliakova *et al.*, in review], was initially accomplished in stand-alone land mode. As the effort matured, an AMIP configuration with the AM4.0 atmosphere was configured for further land development and tuning of the dynamic land-atmosphere dust interactions. The intermediate dust deposition fields from this AMIP configuration were also used in COBALTv2 development. Only at the later stages was the LM4.1 land component merged into ESM4.1 and a brief effort in dust emissions retuning in the full model undertaken. As these separate, intermediate prototype configurations in atmosphere, ocean and land were the only development platforms until the ultimate configuration of ESM4.1 became available for initial testing in March of 2018, limited opportunities for testing were available for before official DECK simulations were performed starting in February 2019.

2.10 ESM4.1 coupled model development highlights

Several aspects of coupled model biases required intensive development from the ESM2M/2G, CM3, and CM4.0 base models and are highlighted below:

Atmosphere radiative balance: Starting from AM4.0 whose Top-Of-Atmosphere (TOA) radiative fluxes (outgoing longwave radiation and shortwave absorption) were tuned toward the observed values primarily through tuning the parameters related to cloud processes (Zhao *et al.*, 2018b), AM4.1 was initially found to have excessive reflection from convective cloudiness over sub-Saharan Africa, North Indian Ocean, and the western tropical Pacific observed top-of-atmosphere shortwave and longwave radiative fluxes in AM4.1. This is likely related to the higher top and increased model layers as well as the ozone and aerosol changes. Ice fall speeds in ESM4.1 are scaled from those in Heymsfield and Donner [1990] by a factor of 0.85 - reduced from the factor of 0.9 used in CM4.0 - to increase the simulated ice water path and decrease the outgoing longwave radiation. Note that these are much closer to the originally derived values than the factor of 1.5 used in CM3. The cloud erosion timescale in convectively active regions

was reduced by 20% relative to CM4.0 from 6.9 hours to 5.6 hours to decrease cloud cover and increase absorbed shortwave radiation.

Aerosol emission and deposition: The overall goal in this effort was to reconsolidate diverse representations of specific aspects of atmospheric aerosols that grew from the last decade of research starting from the original coupled chemistry-climate model (CM3; *Donner et al.*, 2010]. As described in *Horowitz et al.* [in review], ESM4.1 development updated the representation of dry deposition from that in CM3 and CM4.0 to use the same representation for all aerosol tracers. Additionally, the representation of sea salt, found to have a strong high bias in CM4.0 using the temperature dependence of *Martensson et al.* [2003], was revised to the stronger temperature-dependent function of *Jaeglé et al.* [2011] which takes into account newer constraints from in-situ and remotely sensed observations and improves comparison against surface observations in the Southern Ocean and satellite based marine aerosol optical depth (AOD). The wind-generated aerosols, such as dust and sea salt, are treated differently in ESM4.1 than in CM4.0. While dust has fixed sources (unvegetated bare surfaces or sparsely vegetated surfaces) based on satellite observations from 2000 in CM4.0, the sources evolve in ESM4.1 following the vegetation and land use changes simulated by the dynamic land vegetation model LM4.1. To achieve consistency between dust and the land model, emissions are parameterized individually in each tile of LM4.1 and then passed to the atmosphere by flux exchange (the land model receiving gravitational settling flux from the atmosphere). This dynamic interaction of dust between land and atmosphere models improves dust comparison with observations [*Evans et al.*, 2016], which consequently improves the representation of dust interactions with radiation [*Horowitz et al.*, in review]. Dust and iron are also provided from wet and dry deposition to the ocean biogeochemistry. As described above, the ESM4.1 use of a stronger dependency of sea salt emission on sea surface temperature than CM4.0 results in a shift in sea-salt production to the lower latitudes in agreement with observations but also significant cloud reduction and warming over Southern Ocean.

Sea ice-ocean coupling: A key limiting computational factor in GFDL's previous generation models has been the need for very small coupled ocean and ice time steps (down to 5 minutes for some models) relative to those necessary to run the component models to prevent coupled atmosphere-ice-ocean instability. A key numerical advance in coupling this ocean model to the sea ice and atmosphere is the ability for long tracer time steps (2 hours) relative to both dynamical time steps (30 minutes) and coupling time steps with the sea-ice and atmosphere models (1 hour). This allowed for the small coupling time steps required to avoid atmosphere-ice-ocean instabilities without the associated tracer time step becoming burdensome.

Mid-latitude SST: One of the grand challenges in climate modeling is the tendency for mid-latitude zonal winds to fall equatorward [*Lee et al.*, 2013] through a coupled feedback of cold SST generating low clouds that further cool SST to further shift the area of high pressure and

associated zonal winds equatorward. Key to this fidelity is the representation of eddy restratification that is at and below the grid scale in a $\frac{1}{2}$ ° ocean model such as ESM4.1 and must be parameterized. Despite similar global mean biases and root mean square error (RMSE) to World Ocean Atlas [2013] as in the $\frac{1}{4}$ ° ocean in CM4.0, the $\frac{1}{2}$ ° ocean in ESM4.1 was found to experience strong mid latitude SST cold biases and associated strong ocean heat uptake in both hemispheres unless the mixing parameterization associated with restratification by mesoscale eddies and enhanced submesoscale mixed layer eddies was applied. As described in *Adcroft et al.* [2019], application of the Mesoscale Eddy Kinetic Energy (MEKE) formulation of *Jansen et al.* [2015; 2019] and the vertical structure function of *Jansen et al.* [2019] to the $\frac{1}{2}$ ° ocean in ESM4.1 was necessary to reproduce the good SST fidelity of CM4.0 and to reduce spurious ocean heat uptake to minimal values. A notable trade off was found in that while the mesoscale parameterization notably raised midlatitude SST and reduced ocean heat uptake as desired, it was also generated overly cold SSTs in the central subpolar North Atlantic. Midlatitude SST was further raised to match those in CM4.0 by boosting the submesoscale restratification parameterization relative to the strength used in CM4.0 [Adcroft et al., 2019].

Southern Ocean SST and mixed layer biases: Another common bias among CMIP6 models [Russell et al., 2006; Beadling et al., 2019] is warm SST and shallow mixed layer biases in the Southern Ocean which has important implications for representation of the seasonal cycle in $p\text{CO}_2$ and associated carbon fluxes. For both CM4.0 and ESM4.1, incorporation of the *Reichl and Hallberg* [2018] surface boundary layer parameterization brought strong improvement to both summer and winter mixed layers. However, this improvement was partially offset by correcting the high atmospheric sea salt bias discussed above in the atmosphere parameterization section, where the reduction in sea salt over the Southern Ocean had a warming effect. In ESM4.1, the SST bias was reduced by increasing the epipycnal diffusivity from an initial assumption of zero to a value consistent with those diagnosed from MEKE calculations. In contrast to previous generation GFDL models which used a constant epipycnal diffusivity of $600 \text{ m}^2 \text{ s}^{-1}$, ESM4.1 was able to lower the parameterized epipycnal diffusivity to a nominal $200 \text{ m}^2 \text{ s}^{-1}$ that reduced Southern Ocean warm SST biases.

Southern Ocean ventilation: Representation of Southern Ocean ventilation is critical for representation of ocean heat and carbon uptake [e.g. *Frölicher et al.*, 2015] as well as global ocean biogeochemistry [e.g. *Sarmiento et al.*, 2004]. Initial versions of ESM4.1 suffered from a lack of ventilation of Sub-Antarctic Mode Waters (SAMW) and Antarctic Intermediate Waters (AIW), as well as an almost complete lack of Antarctic Bottom Water (AABW) production. Since sufficiently dense surface waters are important to deep water mass formation, these biases were attributed to a lack of surface southward salt flux and warm SST biases. Considerable attention was paid in contrasting the role of mesoscale epipycnal diffusion, epipycnal stirring, and viscosity on Southern Ocean ventilation wherein the ideal age tracer was cross-calibrated against CFC age in a single 40 year run. This calibrated age was then used in a series of 40 year

test simulations to reproduce ventilation patterns in both the mode and intermediate waters of the Southern Ocean as well as the AABW circulation. Similar to the experience developing the equivalent 1° MOM6 ocean in the Seamless Prediction for EArth system (SPEAR) effort [Delworth *et al.*, 2020], extremely high levels of viscosity (up to 2000 m² s⁻¹) in ESM4.1 were necessary along the Antarctic coast to maintain the propagation of ventilated waters away from Antarctica, with the viscosity implemented as a strong function of latitude to avoid high values in mid latitudes which would have otherwise interfered with the propagation of North Atlantic Deep Water as described in Adcroft *et al.*, [2019]. Note that after development was complete, it was discovered that the tidal component of vertical mixing had not been recalibrated from the parameterization in CM4.0 and constituted 1.5 TW of mixing rather than the intended 1.0 TW. The implications of this high bias in tidal mixing are still being investigated.

Sea ice albedos: Representation of sea ice concentration and extent is a key contributor to the polar radiation budget and driver of ocean circulation and heat and carbon uptake as well as serving ecosystem and societal roles. The combined switch to ½ ocean and reduced high latitude sea salt emissions in ESM4.1 resulted in significant sub-polar warming relative to CM4.0. To counter the associated reduction in both Northern Hemisphere and Southern Hemisphere sea ice relative to CM4.0, the values of sea ice, snow and pond optical properties were raised from values of 1 to 1.5 standard deviations of observational uncertainty on their surface albedo influence [Briegleb and Light, 2007].

Land CO₂ cycling: One of the primary applications of coupled carbon-climate models is the application to the regional patterns in atmospheric CO₂ seasonal and interannual variability. Model development foci specifically targeted towards improving comparison with CO₂ observations included 1) the switch from an annual fire model which prevented exploration of the important role of fire in seasonal and interannual modes of variability such as the El Niño / Southern Oscillation, 2) increasing Amazon precipitation, cloud-cover, and soil moisture in the to reduce the dry season and enhance carbon uptake, and 3) Improving the seasonal cycle in boreal climates to delay and enhance summer CO₂ uptake.

Land albedos: As described in Santanello *et al.* [2018], the empirical tuning of land albedos is a common coupled model development phase once atmospheric development has been completed. Extensive efforts to improve boreal climate were made with respect to the role of snow on land albedo through representation of high latitude larches and to assess the sensitivity of the spring transition to snow masking depth. Snow-on-glacier near infra-red albedo: As described in Delworth *et al.* [2020], Southern Ocean heat fluxes in the MOM6 OM4 class of models is highly sensitive to the parameterization of the albedo of snow on glacier in the near infra-red region. As shown by Flanner *et al.* [2007], the albedo of snow on glaciers in the near infrared is a strong function of wavelength, with values near 1.0 in the visible falling to values of 0.9 at 800 nm, 0.75-0.8 at 1.1 µm, and 0.55 at 1.3 µm. As was observed in the case of SPEAR [Delworth *et al.*,

A prototype configuration of ESM4.1 was initialized for the ocean from World Ocean Atlas 2005 for potential temperature and salinity. Ocean biogeochemistry was initialized from World Ocean Atlas 2013 for nutrients and oxygen and GLODAP v2 for dissolved organic carbon and alkalinity. Land for the piControl was first initialized with potential vegetation (no land use) from a “cold start” for all glacier-free land cells with initial vegetation population density of 0.1 m^{-2} , vegetation height of 0.5 m for trees and 0.0025 m for grasses, and soil carbon density of 0 kgC m^{-2} , zero snow cover, and zero soil moisture and run offline for several centuries forced by output from a CM4.0 piControl. Near the end, soil carbon spinup was accelerated by substituting the long-term equilibrium soil carbon distribution for carbon pools derived from the 100-year average distribution of litter, sapwood and wood carbon pools, and the average decomposition function. This was followed by a one year bridge run applying land use transitions to the 1850 state (area fractions) of crops and pastures (no secondary forest) for an estimate of mid-19th century land use but no land use transitions. Sea ice was initialized to zero. The atmosphere state was from previous piControl simulations of development versions of ESM4.1. This prototype ESM4.1 configuration was run for over 800 years with 1850 piControl radiative forcing designated as the CMIP6 DECK [Diagnostic, Evaluation and Characterization of Klima; Eyring *et al.*, 2016]). Over this simulation, a series of changes to land were brought in beginning at year 401 as various technical issues were discovered and addressed, including specification of a maximum residence time for soil carbon to prevent long term accumulation. The final land configuration was spun up to carbon cycle equilibration at year 800 of this prototype run. The final ESM4.1 configuration including the near infrared albedos of snow on glaciers described above was then reset to year 401 of the prototype for all components except land, and the land set to the spin-up from year 800, for year 1 of the ESM4.1 piControl simulation.

Similar to the approach used for CM4.0, the historical simulations that were branched off the ESM4.1 piControl run utilize the identical atmospheric and ocean models, but the land model has additional tiles in each grid box representing secondary growth of forest. The land in the historical run was initialized by first executing a branch from prototype piControl of 100 years duration, followed by a 1700-1750 land-only run starting from potential vegetation (no land use) state, and 1750-1849 coupled bridge run, both with full land use transitions to allow the secondary forests to spin-up to realistic state of re-growth. The ocean, sea ice, and atmosphere initial conditions were then reset to the conditions at the branch-off time of piControl (e.g., piControl year 101) to minimize initial differences between piControl and the historical run due to the ocean and sea ice states. In addition to land use, the historical simulations include the time evolving greenhouse gas concentrations (including ozone for radiation calculations), aerosol precursor emissions and solar irradiance consistent with CMIP6 specifications (documented at <http://goo.gl/r8up31>). Two additional ensemble members for the historical simulation were branched at year 151 and 201 of the piControl.

Model results are available for community analysis via the Earth System Grid Federation (<https://esgf-node.llnl.gov/projects/esgf-llnl/>) and GFDL's public data server (<https://data1.gfdl.noaa.gov/>). Many GFDL colleagues and collaborators not included on the author list but who have played invaluable roles in the model workflow and publication of the data are acknowledged through <https://doi.org/10.22033/ESGF/CMIP6.1407> website. [Krasting *et al.*, 2018].

4 Results

4.1 Pre-industrial prototype spinup and piControl simulation

In this section we describe the global-scale evolution of ESM4.1 from its initialization with ocean observations, through a preliminary version spinup for 400 years and continued simulation for 900 years in the final version forced by piControl 1850 radiative forcing and fixed 1850 land use. Figure 2 depicts the temporal evolution of key variables in the prototype spin up (left side of Figure 2) and final ESM4.1 piControl (right side of Figure 2). Global average net radiation at the top of the atmosphere (Figure 2A) is strongly negative for the first few years of the spinup (average -0.63 W m^{-2} over first decade) before becoming positive at about 0.37 W m^{-2} over the next two hundred years and settling in at 0.23 W m^{-2} over the next two hundred years in the prototype. The ESM4.1 piControl had a reduced value of 0.09 W m^{-2} with its final and spun up land settings. Global average 2-meter air temperature drops initially a few tenths of a degree before rising over the next few hundred years back to approximately $13.45^{\circ}\text{C} \pm 0.13^{\circ}\text{C}$. Upper 700 m ocean temperature (Figure 2C) closely follows the 2-m air temperature. The volume-averaged global ocean (Figure 2D) cooled briefly in the first decade before warming 0.076°C per century in the prototype. The final ESM4.1 configuration warms at a much lower 0.017°C over the first 300 years and 0.013°C over the last 300 years, illustrating the slow approach to equilibrium. Centennial scale variability in this model is highlighted by quasi-regular periodicity in accumulation of heat from the north into the Ross Sea (Figure 2E, black) that is released to the atmosphere through reduction in sea ice (Figure 2F, black). Less regular is the variability in Weddell Sea heat content (Figure 2E, red) associated with strong reductions in sea ice extent (Figure 2F, red) similar to that observed in CM4.0 [Held *et al.*, 2019]. Maximum North Atlantic Meridional Overturning Circulation (Figure 2G) in ESM4.1 is strong ($21.6 \pm 1.4 \text{ Sv}$) with substantial interannual and longer time scale variability. Time-integrated air-sea CO_2 fluxes (Figure 2H) were found to quickly meet the C4MIP equilibrium requirements - drift less than 10 PgC per century [Jones *et al.*, 2016]. The spinup gained 5 PgC within the first few years in response to surface cooling followed by a loss of 10 PgC over the first century in response to adjustments in ventilation before falling into a long-term gain of 15 PgC over the next 300 years. The final piControl configuration accumulated approximately 10 PgC over the first 400 years and another 10 PgC over the next 500 years punctuated by multidecadal losses on the order of 5

PgC during periods of Ross Sea open ice (Figure 2F; black) and heat loss (Figure 2E; black) and gains during quiescent periods.

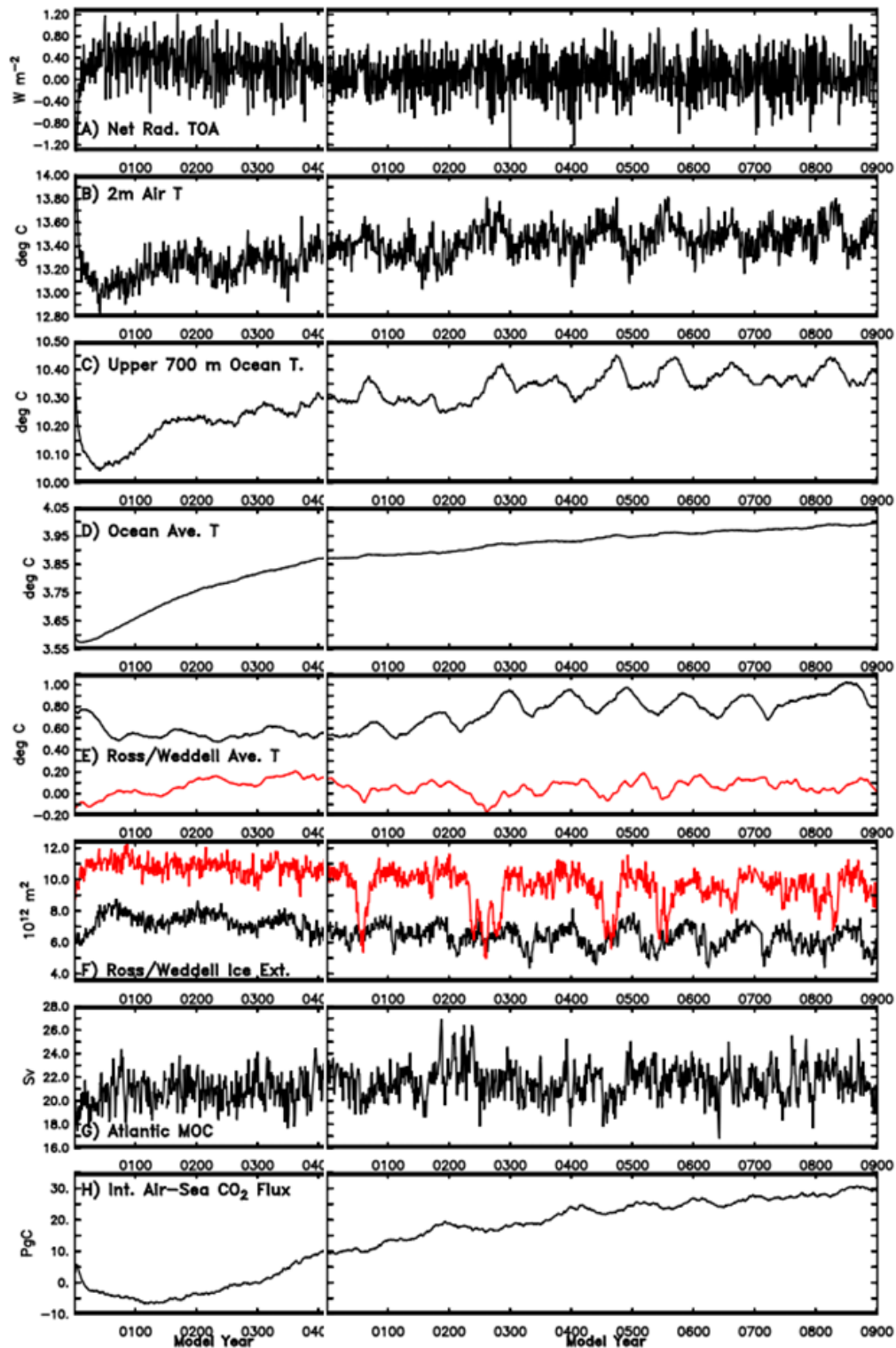


Figure 2: Time series for the 400 year ESM4.1 prototype spin and 900 year ESM4.1 piControl for global average net radiation at the top of the atmosphere (A; W m^{-2}), global average air temperature at 2 m (B; $^{\circ}\text{C}$), upper 700 m average temperature (C; $^{\circ}\text{C}$), global ocean volume average temperature (D; $^{\circ}\text{C}$), Ross sea (160 $^{\circ}\text{E}$ -60 $^{\circ}\text{W}$, 90 $^{\circ}\text{S}$ -60 $^{\circ}\text{S}$; black) and Weddell Sea (60 $^{\circ}\text{W}$ -80 $^{\circ}\text{E}$, 90 $^{\circ}\text{S}$ -60 $^{\circ}\text{S}$; red) volume average temperature (E; $^{\circ}\text{C}$), Ross sea (160 $^{\circ}\text{E}$ -60 $^{\circ}\text{W}$, 90 $^{\circ}\text{S}$ -60 $^{\circ}\text{S}$; black) and Weddell Sea (60 $^{\circ}\text{W}$ -80 $^{\circ}\text{E}$, 90 $^{\circ}\text{S}$ -60 $^{\circ}\text{S}$; red) sea ice extent (F; 10^{12} m^2), Maximum North Atlantic Meridional Overturning (G; Sv); and time-integrated air-sea carbon flux positive into the ocean (H; PgC yr^{-1}).

4.2 Simulation characteristics compared to present-day observations

In this section, we describe the baseline simulation characteristics of ESM4.1 paying particular attention to comparability with GFDL's CM4.0 [Held *et al.*, 2019] and to other previous GFDL and CMIP5 models where possible. We focus our analysis on coupled model physical climate properties and carbon trajectories and refer readers to component model papers for detailed discussion of individual component formulations and behaviors for atmospheric chemistry [Horowitz *et al.*, in review], land [Shevliakova *et al.*, in review] and ocean biogeochemistry and ecosystems [Stock *et al.*, in review]. We compare GFDL's models with recent observational datasets to a standard Atmospheric Model Intercomparison Project [AMIP; Eyring *et al.*, 2016] model period of 1980-2014 for atmospheric variables, and 1990-2010 for oceanic variables to be roughly consistent with the major observational period for ocean products.

4.3 Patterns of shortwave and longwave radiation in comparison with satellite estimates:

Net radiation at the top of the atmosphere in ESM4.1 shows a slightly larger negative bias (-0.29 versus -0.08 W m^{-2} in CM4.0) and RMSE (7.60 versus 7.28 W m^{-2} in CM4.0) comparison to CERES EBAF Terra Edition 4.0 ([Loeb *et al.*, 2018]; <https://ceres.larc.nasa.gov/>; Figure 3). Comparison of net shortwave absorption at the top of the atmosphere using CERES EBAF edition 4.0 gives an overall improvement in bias (-1.55 versus -2.43 W m^{-2} in CM4.0) but a slight degradation in RMSE (8.98 versus 8.91 in CM4; Figure 4). Much of this difference stems from differences in tropical clouds associated with the Intertropical Convergence Zone (ITCZ) and the increased near infrared albedo of snow over glacier in ESM4.1 over CM4.0. Both CM4.0 and ESM4.1 exhibit strongly reduced biases in the Southern Ocean though there remains a moderate lack of cloud cover (and as we show later, associated lack of summer sea ice) over the Ross and Weddell Seas. One of the key biases in all three GFDL models continues to be the lack of ability to maintain stratus clouds over the ocean in the California, Peru, and Benguela Eastern Boundary Current regions. While most other strong biases were ameliorated and considerable effort was applied during the development of AM4.0 (Zhao *et al.*, 2018b), these areas remain the source of considerable remaining local bias. As described in Zhao *et al.* [2018a], this effort was focused on the implementation of the Cloud Layers Unified By Binormals scheme [Gao *et al.*,

2015] which demonstrated encouraging results in developing stratus clouds that helped prevent ocean surface warming. This approach was ultimately deferred due poor coupling in mid-latitudes where the propensity to develop stratus clouds induced a strong feedback of ocean cooling and equatorial constriction of the mid-latitude winds.

Similarly, Outgoing Longwave Radiation (OLR) gives an overall improvement in bias (-1.26 versus -2.35 W m^{-2} in CM4.0) but a slight degradation in RMSE (6.44 versus 6.24 in CM4.0) associated with the ITCZ (Figure 5). As discussed in Zhao et al. [2018b], these improvements in OLR are associated in the combined improvement in the precipitation climatology and associate high cloud cover and OLR response to precipitation. The improved OLR response is primarily due to convection scheme, in particular, the convective detrainment, which may be understood as the amount of condensate detrained from deep convective clouds per unit surface precipitation. This process exerts a strong control on the amount of high clouds and condensate in the deep tropics and is responsible for the strong improvement between CM3 and CM4.0/ESM4.1.

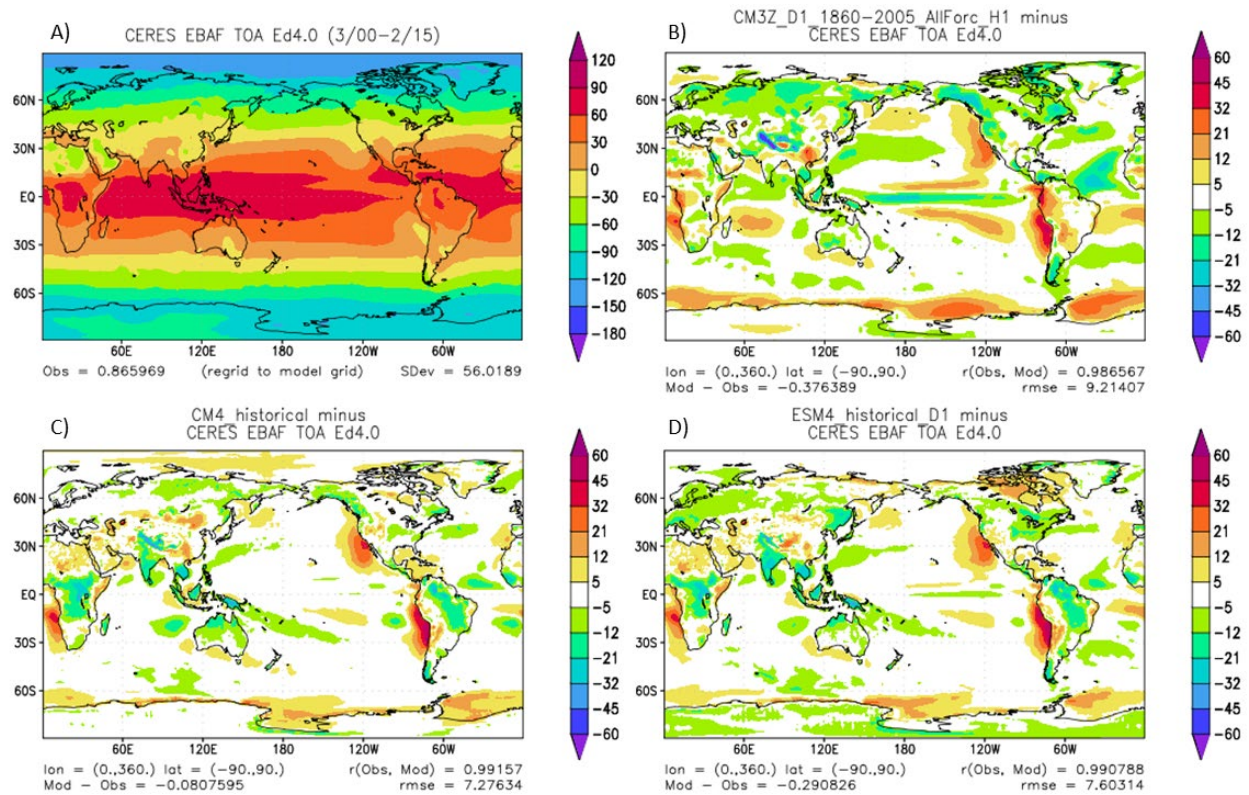


Figure 3: Maps of net downward radiation at the top of the atmosphere (W m^{-2}) from CERES EBAF 4.0 (A), and bias in CM3 (B; 1980-2004 of historical run), CM4.0 (C; 1980-2014 of historical run), and ESM4.1 (D; 1980-2014 of historical run) compared to CERES.

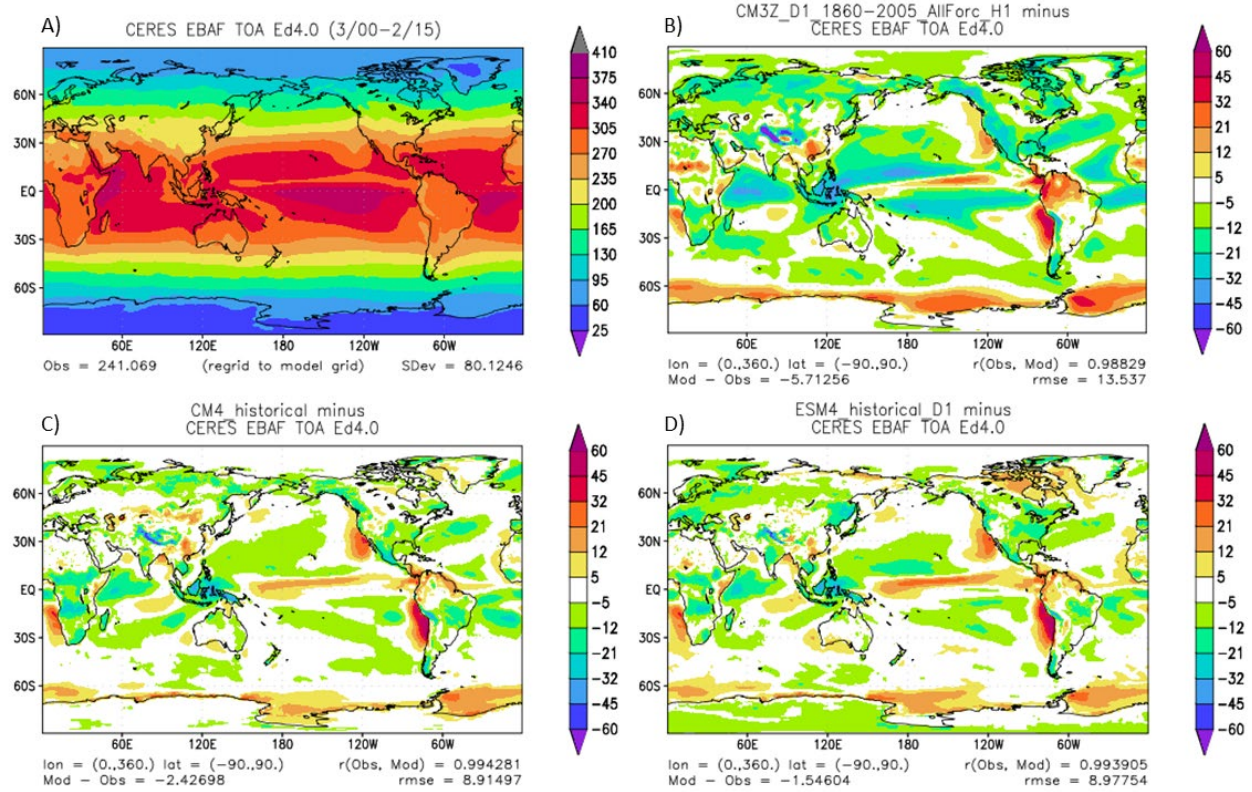


Figure 4: Shortwave absorption as estimated at the top of the atmosphere (W m^{-2}) from CERES EBAF 4.0 (A), and bias in CM3 (B; 1980-2004 of historical run), CM4.0 (C; 1980-2014 of historical run), and ESM4.1 (D; 1980-2014 of historical run) compared to CERES.

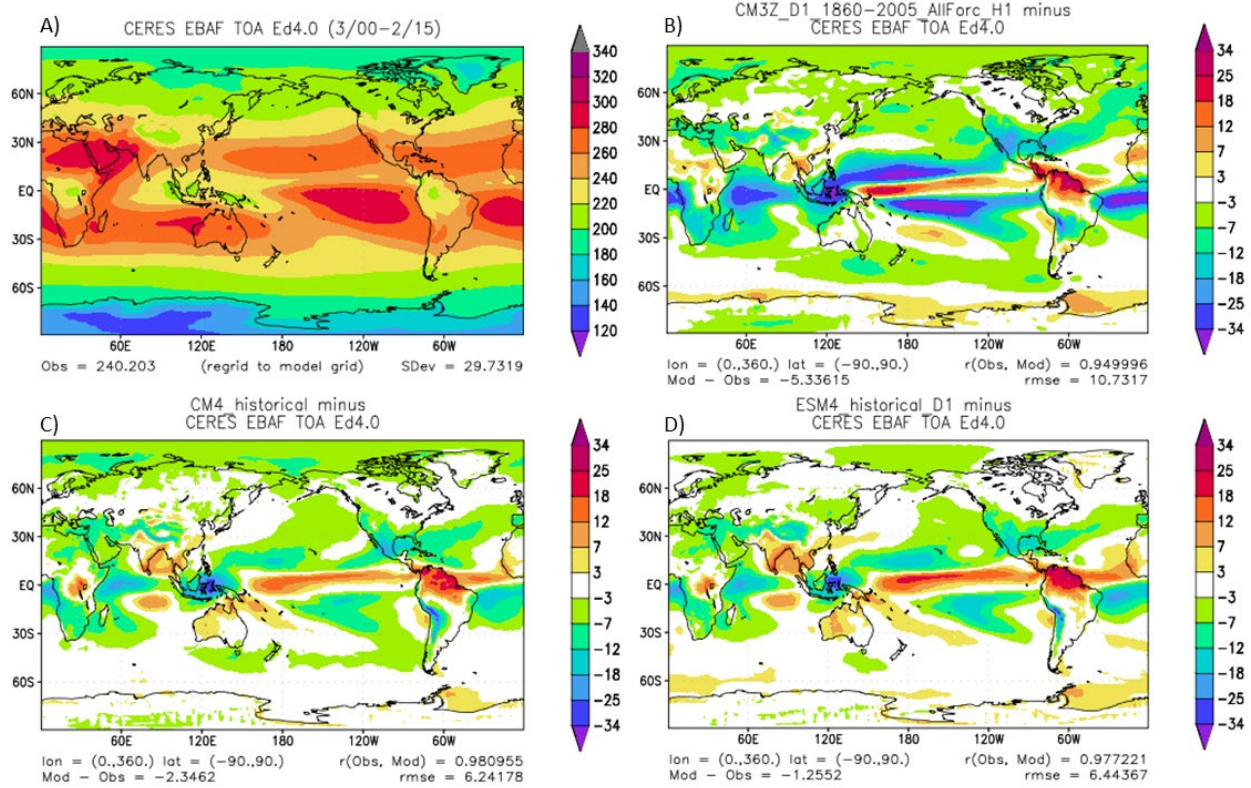


Figure 5: Outgoing longwave radiation at the top of the atmosphere (W m⁻²) from CERES EBAF 4.0 (A), and bias in CM3 (B; 1980–2004 of historical run), CM4.0 (C; 1980–2014 of historical run), and ESM4.1 (D; 1980–2014 of historical run) compared to CERES.

4.4 Aerosol comparison with aerosol surface concentration and satellite measurements:

Aerosol differences with CM4.0 and ESM4.1 (Figure 6) illustrate a combination of advantages and disadvantages compared to observations from measurements collected continuously by the University of Miami at 28 stations, mostly on islands, over the previous 2 to 5 (in the case of the Barbados station) decades [Savoie and Prospero, 1989; Zuidema et al., 2019]. The more comprehensive treatment of SO₄ in ESM4.1 including interactions with nitrogen (see Horowitz et al., [in review] for details) compares slightly better to aerosol surface concentration measurements relative to the simple treatment in CM4.0 both in terms of correlation coefficient (0.93 versus 0.89 in CM4.0) and RMSE (0.20 versus 0.23 in CM4.0). Major differences include an overall decrease of SO₄ over North America and increase over central Europe, India and China. Dust distributions in ESM4.1, which are prognostically-derived from vegetation dynamics rather than specified externally in CM4.0, illustrate a significant degradation in fidelity with correlation coefficient (0.78 versus 0.90 in CM4.0) and RMSE (0.49 versus 0.35 in CM4.0). Most of this difference is associated with an increase in dust sources emanating from North and South America where the presence of fire in ESM4.1 leads to bare area which becomes a dust source instead of the fixed dust sources associated with dry lake beds in CM4.0. While ESM4.1

underestimates dust in some of the highest dust regions, it overestimates dust in larger-area low dust and overall aerosol regions over the ocean which may make ESM4.1 overly sensitive to changes in dust emissions overall. In terms of sea salt, ESM4.1 exhibits a large improvement with the *Jaeglé et al.* [2011] scheme which incorporates a stronger function of temperature over the previous implementation in CM4.0 with much higher correlation coefficients (0.79 versus 0.50 in CM4.0) and lower RMSE (0.36 versus 0.50 in CM4.0).

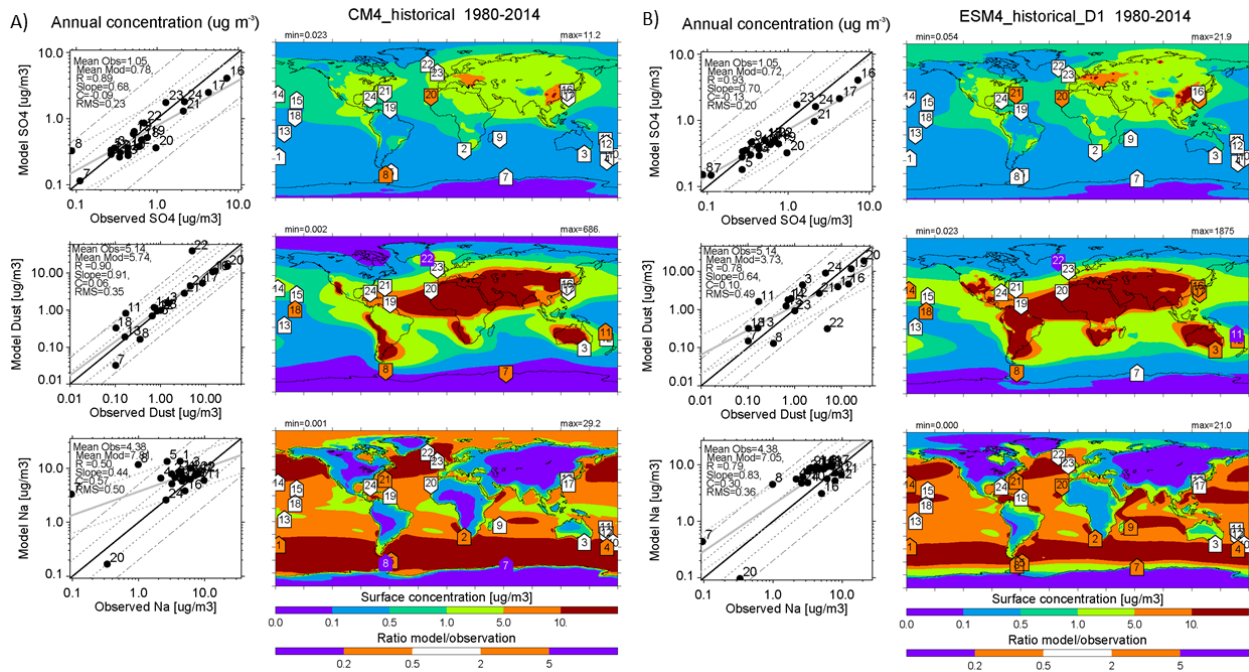


Figure 6: Comparison of CM4.0 (A) and ESM4.1 (B) for sulfate (top), dust (middle), and sea salt (bottom) aerosols ($\mu\text{g m}^{-3}$) for the period of 1980-2014 with aerosol surface concentration measurements from measurements collected continuously by the University of Miami at 28 stations, mostly on islands, over the previous 2 to 5 (in the case of the Barbados station) decades (Savoie and Prospero, 1989; Zuidema et al., 2019) shown in the right panels as boxes with the ratio of model to observations and in the left panels with log-model versus log-observations with associated statistics for mean observed (Mean Obs), simulated (Mean Mod), correlation coefficient (R), slope (Slope - 1.0 for a perfect comparison), intercept (C - 0.0 for a perfect comparison), and Root Mean Square Error (RMS) of the linear regression provided in the upper left corner of each panel.

Overall, AOD exhibits a slight increase from 0.15 to 0.16 in better agreement with the satellite data product from MISR ([*Kahn et al.*, 2009]; https://eosweb.larc.nasa.gov/project/misr/cgas_table; Figure 7). Patterns in ESM4.1 (Figure 8C) largely resemble biases in CM4.0 (Figure 8B). The largest differences between the two are related to high AOD regions over Brazil, West Africa, India and China in ESM4.1 that are not captured in CM4.0.

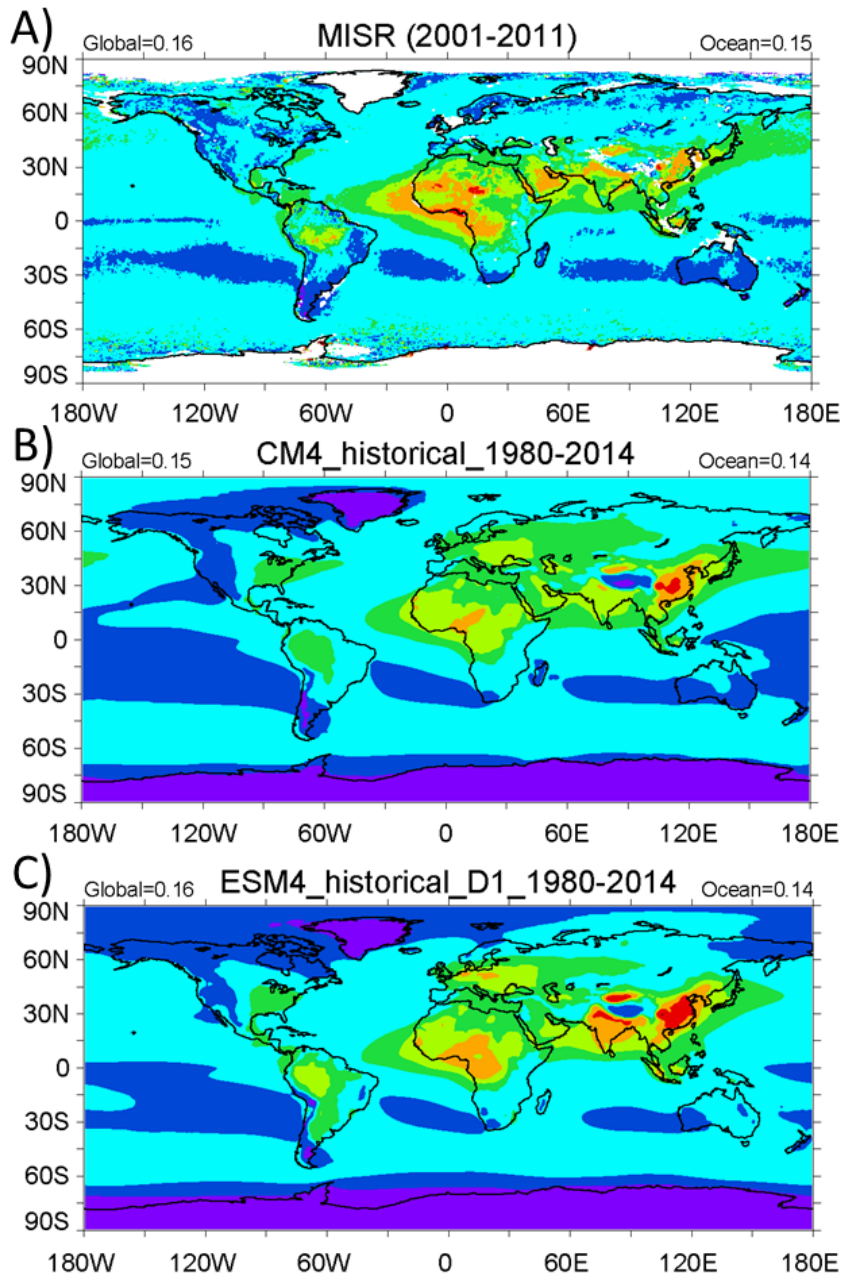


Figure 7: Comparison of aerosol optical depth (AOD) from MISR (A; Kahn et al., 2009), CM4.0 (B); and ESM4.1 (C).

4.5 Wind comparison

Upper atmosphere zonal winds in ESM4.1 demonstrate very different bias patterns and a slight overall degradation relative to CM4.0 compared to European Center for Medium Range Weather Forecasting Re-Analysis 40 [ERA40; Uppala *et al.*, 2005; <http://apps.ecmwf.int/datasets/data/era40-daily/levtype5sfc/>]. While upper level tropical winds in ERA40 and CM4.0 are uniformly easterly, those in ESM4.1 include westerly wind jets centered on the equator at 30 hPa and 5 hPa (Figure 8). ESM4.1 South polar winds exhibit considerable improvement relative to CM4.0 while northern subpolar and polar winds ESM4.1 biases switch from being overly strong to being overly weak. ESM4.1 also exhibits much lower variability than CM4.0 in the upper atmosphere above 10 hPa agreeing with ECMWF reanalysis at southern subpolar latitudes, while CM4.0 agrees better with ECMWF at northern subpolar latitudes (not shown). The biases in ESM4.1 may be related to the gravity wave scheme working on a finer stratospheric grid than was initially parameterized for CM4.0. Meridional velocity and vertical velocity (ω) patterns look similar between the two models except that ESM4.1 has a slightly weak southern arm of the equatorial overturning streamfunction relative to observations and CM4.0 (not shown). Like CM4.0, ESM4.1 does not exhibit a significant quasi-biennial oscillation even with the increased number of vertical layers from 33 to 49.

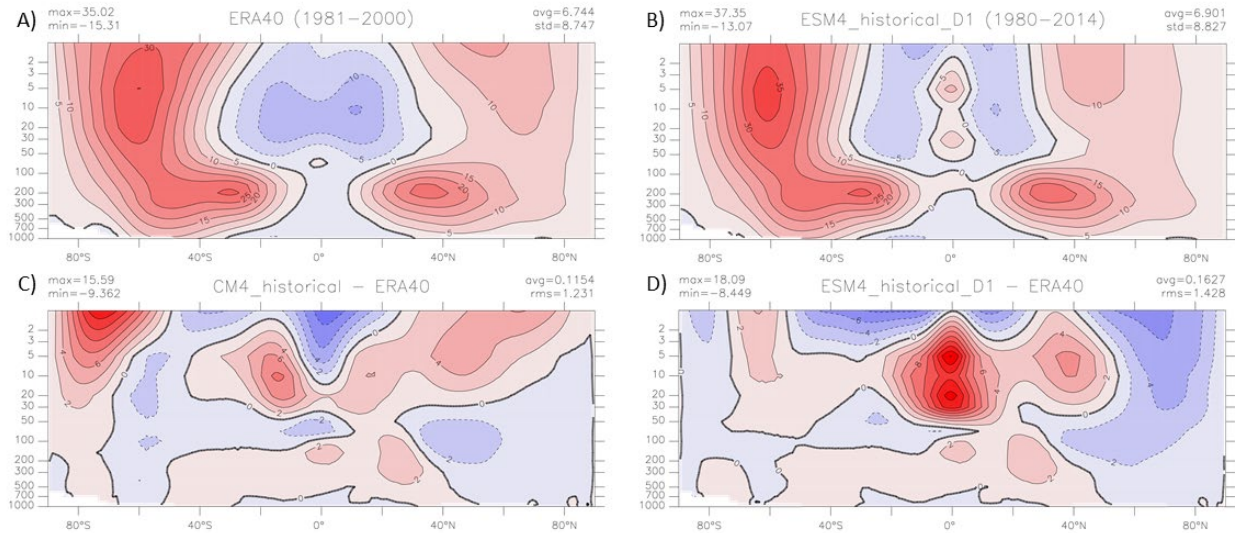


Figure 8: Comparison of meridional and vertical (hPa) patterns in zonally averaged ERA-40 zonal wind (A) with ESM4.1 (B), bias in CM4.0 compared to ERA (C), and bias in ESM4.1 compared to ERA (D).

At the surface, zonal wind stress (Figure 9) in ESM4.1 (red) is more consistent with ERA-40 (blue) and Modern-Era Retrospective Analysis for Research and Applications (MERRA; [Rienecker *et al.*, 2011]; <https://gmao.gsfc.nasa.gov/reanalysis/MERRA/>; green) than CM4.0 (black). ESM4.1 also exhibits a stronger and more poleward peak zonal wind stress than CM4.0 over the midlatitude ocean, particularly in the Pacific Sector.

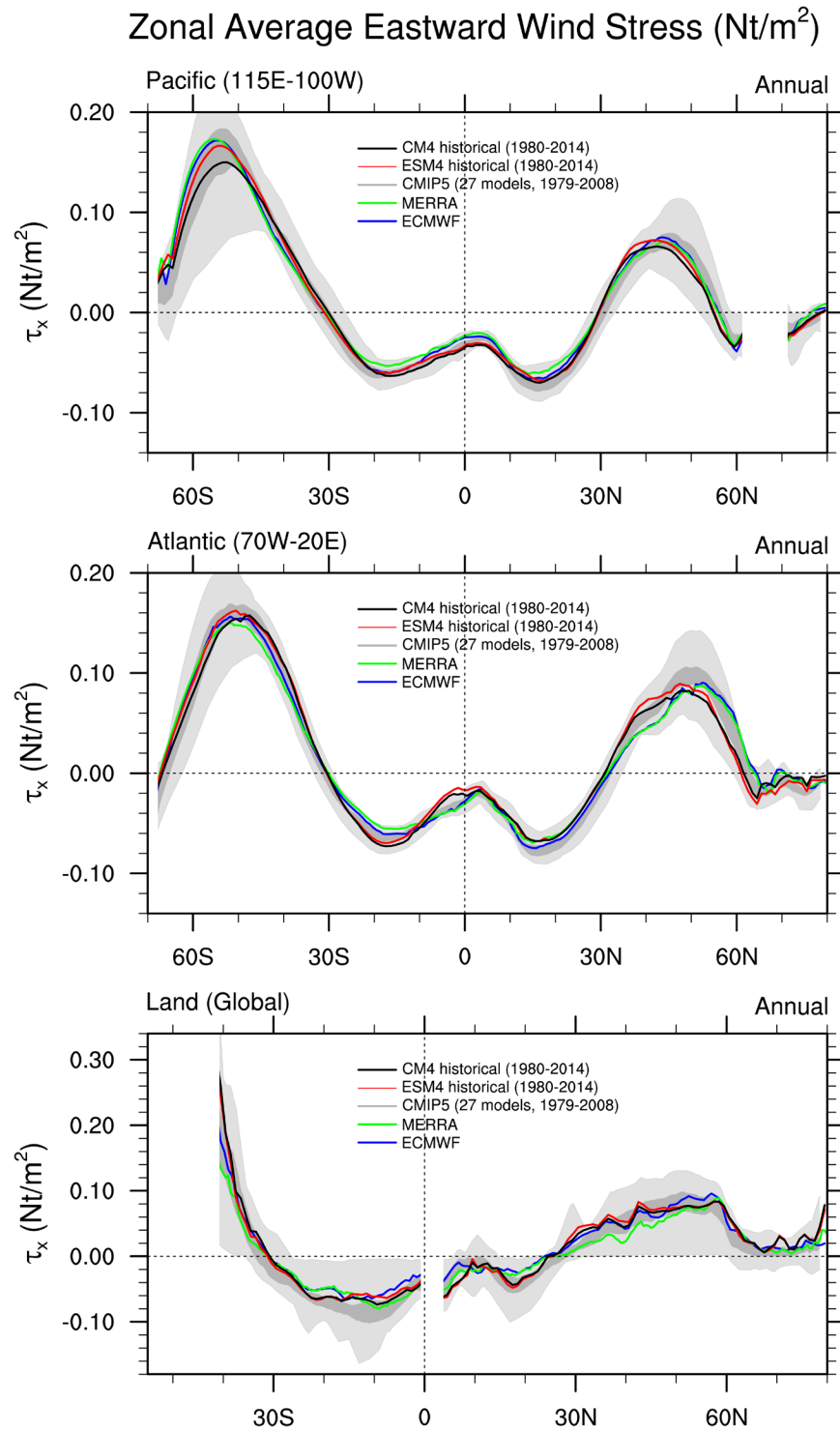


Figure 9: Comparison of zonal eastward wind stress (Nt m^{-2}) on the Pacific Ocean (A), Atlantic Ocean (B), and Land surface (C) from CM4.0 (black), ESM4.1 (red), CMIP5 (dark grey shading – standard deviation about multi-model mean; light grey shading – all 27 models), and reanalysis from MERRA (green) and ECMWF (blue).

4.6 Sea Level Pressure

As shown in Figure 10 (upper panels), ESM4.1 exhibits an overall good comparability in sea level pressure patterns compared to ERA-Interim [Dee *et al.*, 2011] with a correlation coefficient ($r=0.93$) nearly as high as that for CM4.0 ($r=0.95$). However, ESM4.1 also exhibits moderate degradation relative to CM4.0 with a high pressure bias in the Arctic east of Greenland (Figure 10, lower panels), exacerbating the bias and associated RMSE in CM4.0 (1.33 versus 1.16 in CM4.0) and RMSE over the northern hemisphere more generally (0.92 versus 0.73 in CM4.0; not shown). We suspect that these differences are driven primarily by the interactive ozone in ESM4.1 versus prescribed ozone in CM4.0 and the higher snow on glacier albedos than those used in CM4.0. ESM4.1 also exhibits higher dust and lower sea salt than CM4.1 in this region (Figure 6). Similarly slight increases in RMSE are exhibited in the southern hemisphere as well (2.96 versus 2.69 in CM4.0; not shown).

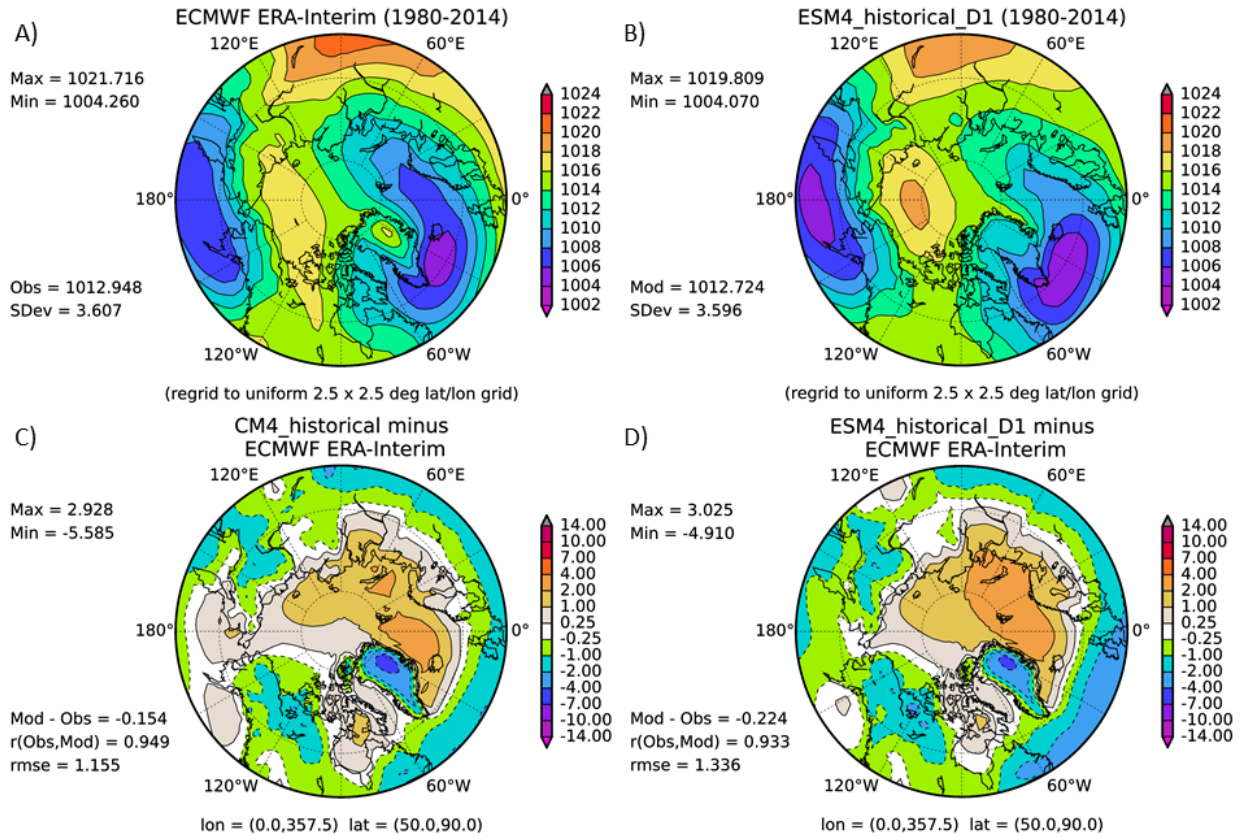


Figure 10: Comparison of ECMWF ERA-Interim Reanalysis for Arctic winter (DJF) season sea level pressure (A) with ESM4.1 (B), bias in CM4.0 (C), and bias in ESM4.1 (D).

4.7 Surface air temperature and humidity

Comparison of 2 meter air temperature (Figure 11, upper panels) illustrates the marked improvement in ESM4.1 relative to CM3 and even CM4.0 compared to ERA Interim ([Dee *et al.*, 2011]; <https://www.ecmwf.int/en/research/climate-reanalysis/era-interim>) in RMSE (1.31°C in ESM4.1 versus 1.78°C in CM3 and 1.55°C in CM4.0), which can be attributed to an improvement in pattern in the case of CM3 and reduction in mean bias in the case of CM4 (-0.52°C in ESM4 versus -0.71°C in CM3 and -1.11°C in CM4.0). However, strong positive biases in ESM4.1 air temperature are associated with deep water formation regions in the Ross Sea, Weddell Sea, and Labrador Sea associated with deep biases in wintertime mixed layer depths and low biases in summer sea ice as shown in 4.9 *Surface Ocean*. Overall biases over the Southern Ocean are markedly reduced relative to CM3.

Comparison of 2 m relative humidity (Figure 11 lower panels) also illustrates marked improvement in ESM4.1 compared to CM3 but moderate degradation compared to CM4.0 in both RMSE (7.7% in ESM4.1 versus 9.2% in CM3 and 7.1% in CM4.0) and bias (3.7% in ESM4.1 versus 4.1% in CM3 and 3.3% in CM4.0). Regionally, this degradation relative to CM4.0 is expressed as high biases over Siberia and North American continents while other areas appear to retain the improvements gained from CM3 to CM4.0. Initial analysis has found this bias to be restricted to the spring season during the onset of the growing season through enhancement of relative humidity and associated low cloud cover as relative humidity is used for parameterizing cloud amounts. We suspect that differences in snow cover, vegetation and soil moisture in are driving increases in relative humidity and generating more low clouds under conditions of a stable planetary boundary layer when surface temperature is much colder than the atmosphere. Further description of the ESM4.1 land component and its fidelity can be found in *Shevliakova et al.* [in review]. This mechanism may also be related to the general propensity of climate models to produce excessive low cloud amounts in the Arctic (e.g. *Vavrus and Waliser*, 2008).

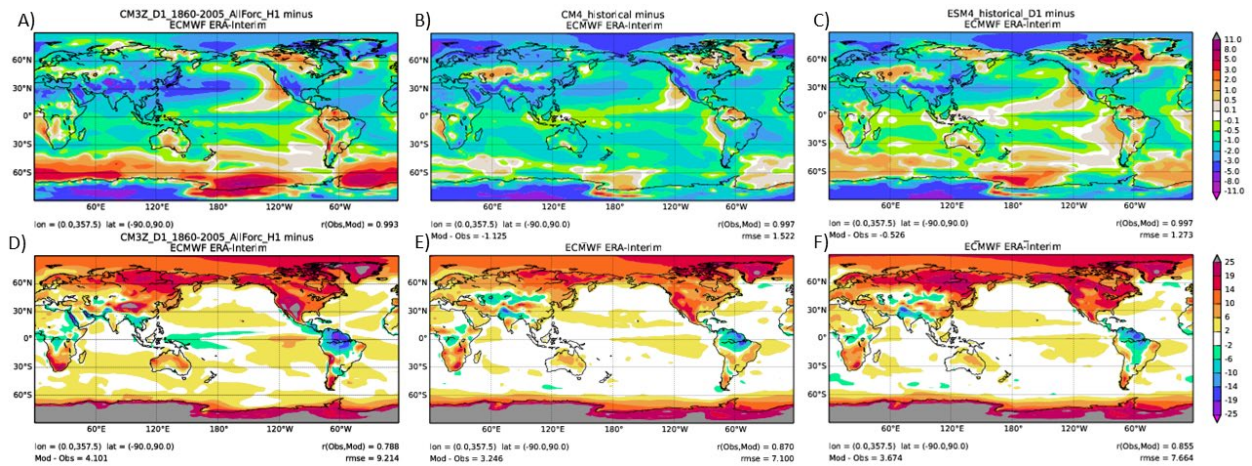


Figure 11: Comparison of 2 m bias in air temperature (upper panels) and relative humidity (lower panels) in GFDL's previous generation CM3 (A), CM4.0 (B) and ESM4.1 (C) against ECMWF ERA Interim reanalysis along with statistical metrics

4.8 Precipitation

Comparison of model precipitation in comparison to the Global Precipitation Climatology Project version 2.3 (GPCPv2.3; <http://www.esrl.noaa.gov/psd/data/gridded/data.gpcp.html> [Adler *et al.*, 2003, 2018]) is provided in Figure 12. Precipitation over ocean in ESM4.1 exhibits very similar overall patterns but a stronger double ITCZ bias than CM4.0 expressed both as Pacific and Atlantic south equatorial maxima are higher in ESM4.1 than CM4.0 (Figure 12). ESM4.1 and CM4.0 have similar precipitation patterns over land with substantial improvements in both the wet bias over the US west and high mountain Asia and dry bias in the Amazon exhibited in CM3. ESM4.1 demonstrates a much stronger overall tropical precipitation and specifically double ITCZ bias than does CM4.0. Irrespective of these lingering issues with representation of the ITCZ, this generation of model demonstrates considerable improvement in the long standing model biases over previous generation GFDL models with significantly lower RMSE in both CM4.0 (0.96 mm d⁻¹) and ESM4.1 (1.02 mm d⁻¹) compared to CM3 (1.29 mm d⁻¹).

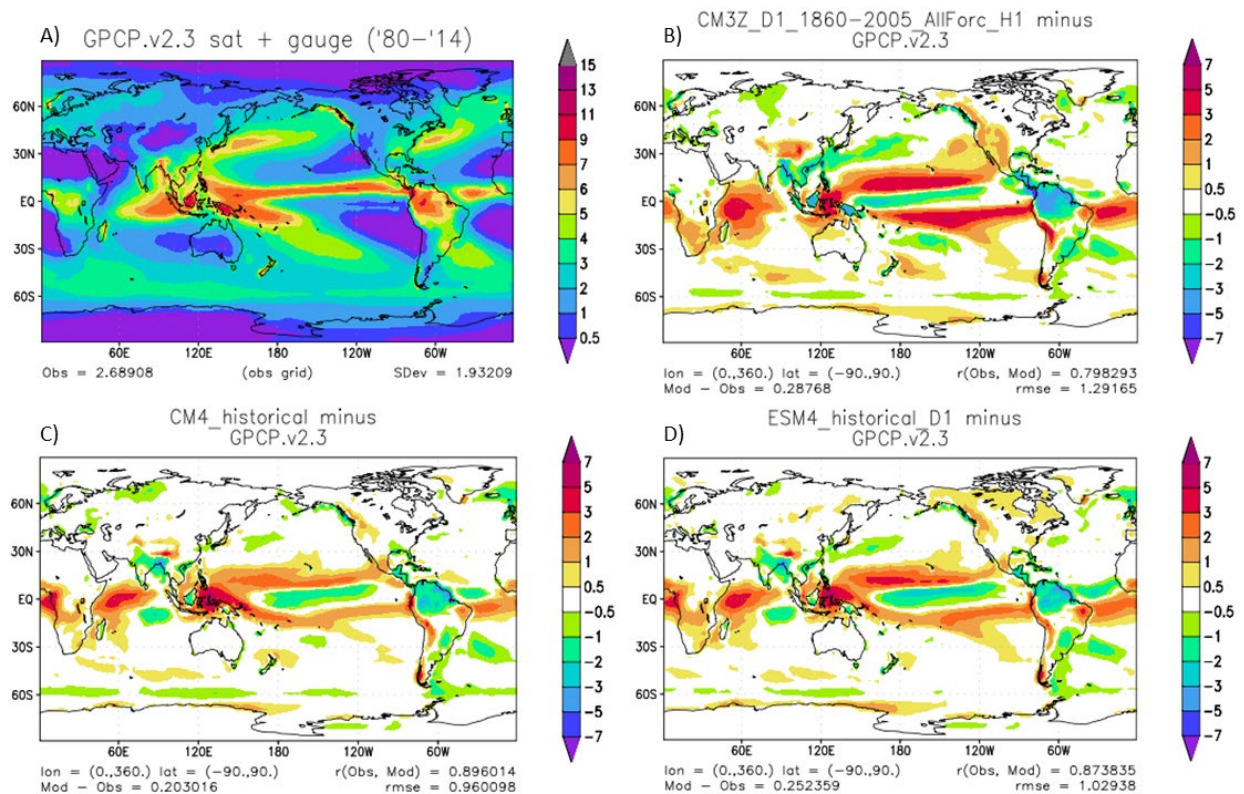


Figure 12: Comparison of precipitation from GPCPv2.3 (A; Adler et al., 2003, 2018), and bias in CM3 (B; 1980-2004 of historical run), CM4.0 (C; 1980-2014 of historical run), and ESM4.1 (D; 1980-2014 of historical run) compared to GPCPv2.3.

To further highlight this improvement across generations of models, we focus in on the Amazon region which has remained one of the main recalcitrant biases in climate models [e.g. *Richter and Xie*, 2008]. For this comparison we focus on ESM2M rather than CM3 because of the important implications of Amazon precipitation in representation of the carbon cycle in ESM2M [Dunne et al., 2013]. In the Northern region (Figure 13, left panel, solid), minimum precipitation during the Amazon dry season has increased from 0.38 mm d^{-1} in GFDL's previous generation ESM2M to 2.0 mm d^{-1} in CM4.0 and ESM4.1 as the number of consecutive months of precipitation below 4.0 mm d^{-1} declined from 9 in ESM2M to 3 in CM4.0 and ESM4.1. Interannual variability in the June-August onset of the dry season in the Northern Amazon as also increased considerably (Figure 13, right panel, solid). In the Southern Amazon, minimum precipitation (Figure 13, left panel, dash) has also increased considerably from 0.01 mm d^{-1} in ESM2M to 0.14 mm d^{-1} in CM4.0 and 0.12 mm d^{-1} in ESM4.1 as the number of consecutive months of precipitation below 1.0 mm d^{-1} declined from 5 in ESM2M to 4 in CM4.0 and ESM4.1. Interannual variability in the August-September termination of the dry season in the Northern Amazon has also increased considerably (Figure 13, right panel, dash).

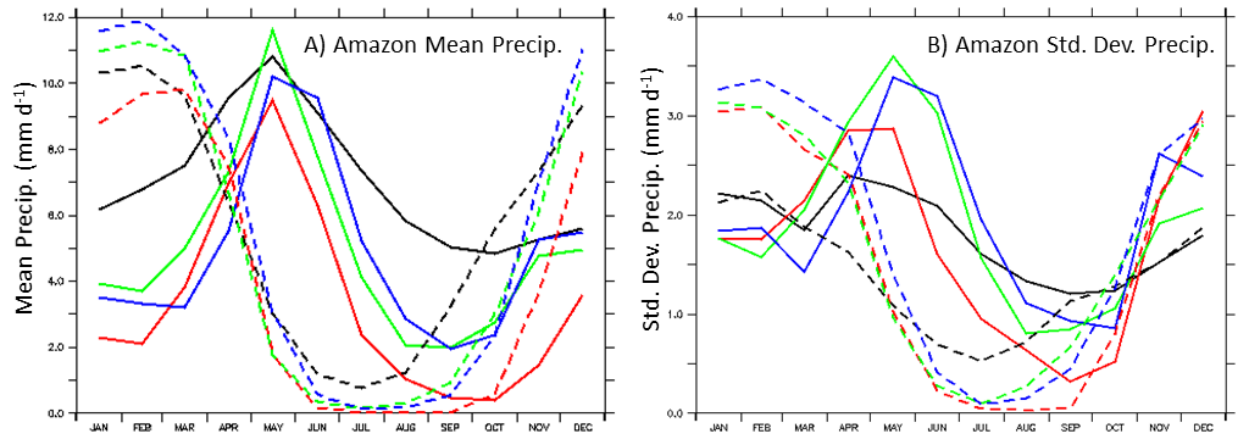


Figure 13: Regional precipitation (mm d^{-1}) climatology over the Northern (70°W - 55°W , 5°S - 5°N ; solid) and Southern (70°W - 50°W , 15°S - 5°S ; dash) Amazon based on definitions from Yin et al. (2013) from GPCPv2.3 (black; Adler et al., 2003, 2018). ESM2M (red), CM4.0 (green), and ESM4.1 (blue).

4.9 Surface Ocean

Surface ocean patterns of bias in sea surface temperature compared to World Ocean Atlas 2013 [WOA13; Locarnini et al., 2013] are shown in Figure 14 for ESM2M, CM3, CM4.0, and

ESM4.1 highlighting the substantial improvement in root mean square error (RMS) across each model generation with ESM4.1 (RMS=0.68°C) exhibiting a 45% reduction in root mean square error relative to ESM2M (RMS=1.24°C). While CM4.0 exhibits a moderate cold bias of approximately -0.45°C compared to WOA13, ESM4.1 exhibits only a slight cold bias of -0.07°C leading to an overall smaller error (RMS=0.68°C) compared to CM4.0 (RMS=0.84°C). While the absolute differences with observations between the models reflect their average temperature bias (CM4.0 appearing largely blue in Figure 14C, for example), relative patterns are very similar across all GFDL models with a tendency to exhibit cold biases in the subtropical gyres associated with a lack of eddy restratification [Adcroft *et al.*, 2019] and equatorially-constricted mid latitude winds (Figure 9), and warm biases in eastern boundary current regions associated with a lack of cloud cover and upwelling-favorable winds (Figure 4), the Southern Ocean associated with lack of cloud cover (Figure 4), and the Labrador Sea associated with shallow overturning [Adcroft *et al.*, 2019] (and see below). Major differences between ESM4.1 and CM4.0 include cold biases in CM4.0 along 40°S and in the Norwegian Sea that are largely absent in ESM4.1 and an extreme cold bias in ESM4.1 subpolar North Atlantic that is relatively mild in CM4.0. Much of these differences relate to the use of a mesoscale eddy formulation in ESM4.1 that is absent in CM4.0 with its higher resolution. As described in Adcroft *et al.* [2019], the improvement in subpolar North Atlantic temperature distributions in the CM4.0 ocean are associated with good representation of the deep western boundary current at the CM4.0 $\frac{1}{4}^\circ$ resolution without the mesoscale eddy parameterization that becomes diffuse and follows the mid-Atlantic Ridge at the ESM4.1 $\frac{1}{2}^\circ$ resolution without the mesoscale eddy parameterization. In the Southern Ocean, we attribute much of the relative warmth of ESM4.1 to the lower sea salt emissions in the Jaeglé *et al.* [2011] scheme.

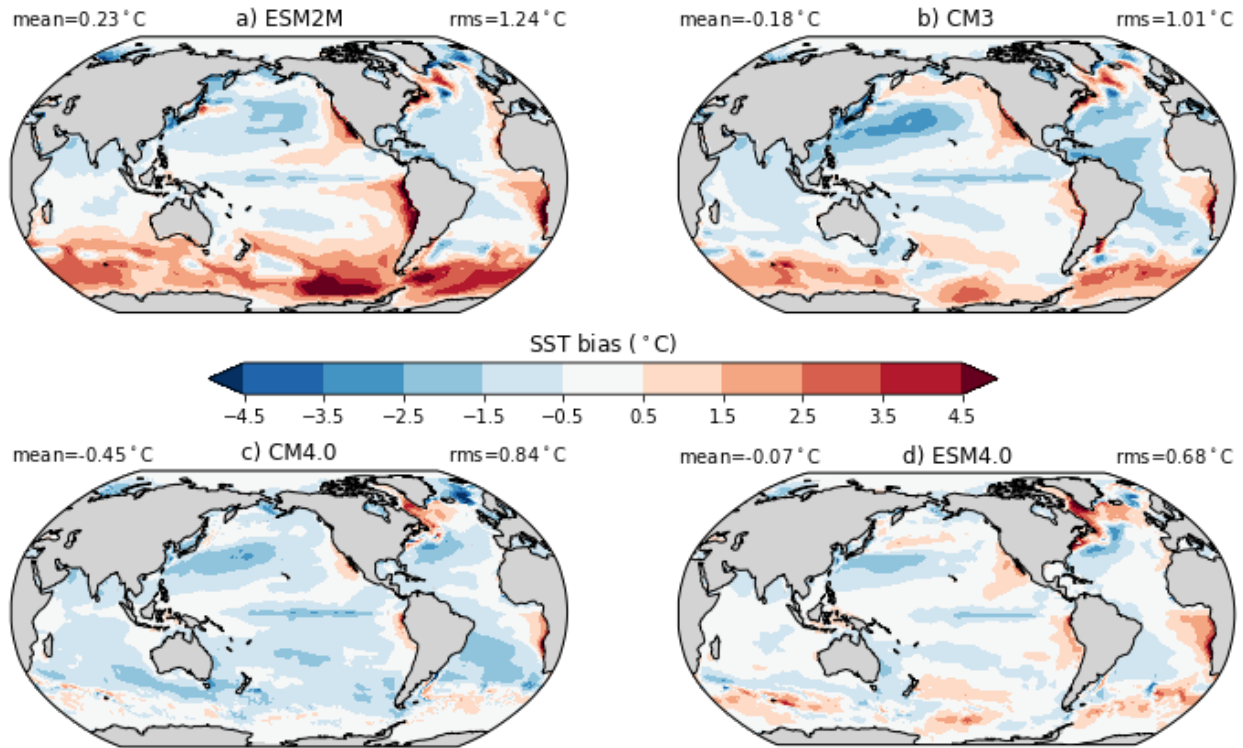


Figure 14: Bias in sea surface temperature (°C) in ESM2M (A), CM3 (B), CM4.0 (C) and ESM4.1 (D) compared to World Ocean Atlas 2013 (Locarnini et al., 2013)

Surface ocean patterns of bias in sea surface salinity compared to World Ocean Atlas 2013 [WOA13; Zweng *et al.*, 2013; Figure 15] for ESM2M, CM3, CM4.0, and ESM4.1 similarly highlight substantial improvement across each model generation with CM4.0 and ESM4.1 (RMSE=0.47 ppt) exhibiting a 41% reduction in root mean square error relative to ESM2M (RMSE=0.80 ppt). Most of this improvement is associated with reductions in low salinity biases in the Indian and North and South Pacific midlatitude Oceans and high salinity bias in the Eastern North Pacific associated with improvements in the sea surface temperature (Figure 14), wind (Figure 9) and precipitation (Figure 12) improvements. Recalcitrant biases include the Arctic high bias and equatorial and Southeast Atlantic low bias. The Arctic high bias is associated with relatively vigorous mixing compared to observations typical of this class of model [e.g. Zhang and Steele, 2007], and the South Atlantic bias associated with a lack of upwelling-favorable winds in the Benguela system as shown in radiation (Figure 4) and temperature (Figure 15) biases. Other biases in sea surface salinity shared between CM4.0 and ESM4.1 include low biases in Oceania associated with the precipitation bias in that region (Figure 12) and high biases in the off of the Northeast Coast of the United States associated with a weak source of fresh water via the Labrador Current and Bay of Bengal (which unfortunately is worse in CM4.0 and ESM4.1 than previously) associated with a lack of precipitation (Figure 12) and river runoff (not shown) and overly deep mixing in that region. Differences between CM4.0

and ESM4.1 include a CM4.0 low bias in the Norwegian Sea absent in ESM4.1, a stronger high bias in the Tropical Pacific in CM4.0 and stronger high bias off the US East Coast in ESM4.1.

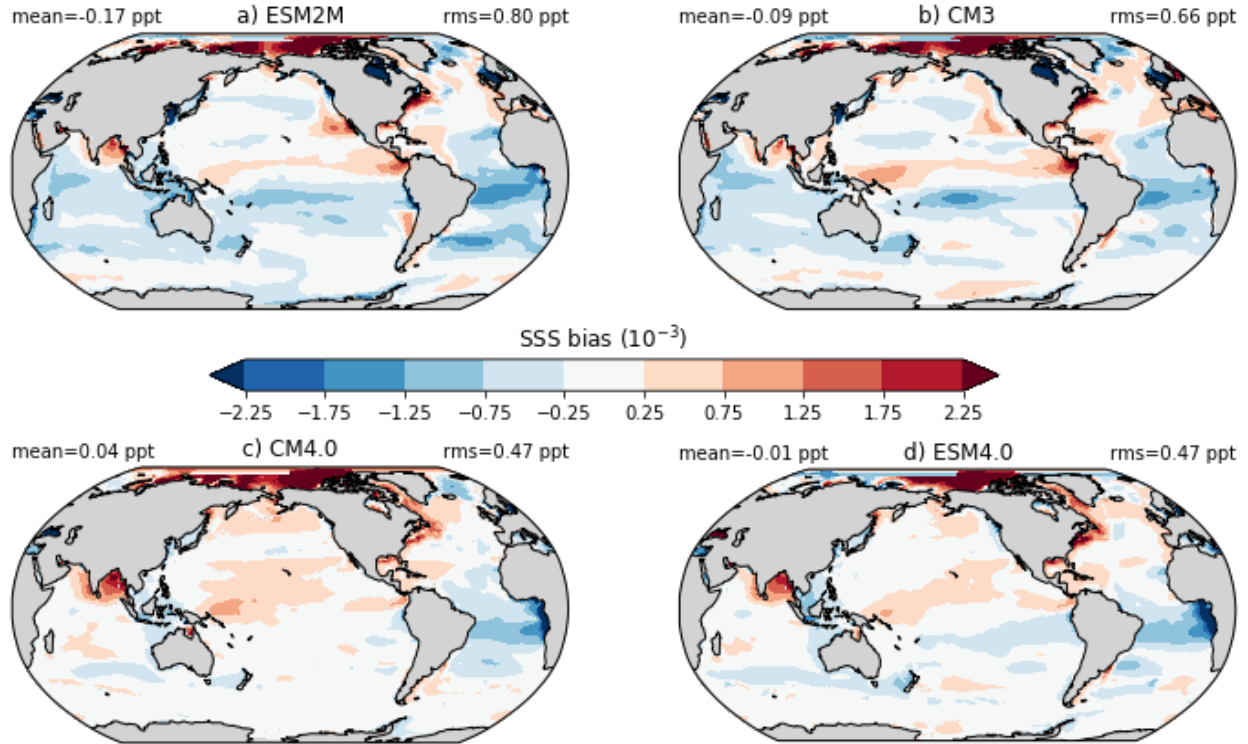


Figure 15: Bias in sea surface salinity (ppt) in ESM2M (A), CM3 (B), CM4.0 (C) and ESM4.1 (D) compared to World Ocean Atlas 2013 (Zweng et al., 2013).

Monthly minimum mixed layer depths (Figure 16) compared to World Ocean Atlas 2013 (upper left) in CM4.0 (lower left, $r^2 = 0.70$) and ESM4.1 (lower right, $r^2 = 0.68$) overall exhibit considerably more regional pattern variance representation relative to the previous generation Earth System Model, ESM2M (upper right; $r^2 = 0.49$). The primary sources of this improved fidelity are the combination of relatively fine (2 m) resolution of the upper layers (compared to 10 m in ESM2M) and representation of wind driven turbulent mixing in the Southern Ocean [Adcroft et al., 2019]. However, both CM4.0 and even more so ESM4.1 exhibit a shallow bias relative to the prominent mostly subtropical bands of deeper (15-25 m) mixing in WOA13 (Figure 16; upper left). ESM2M exhibits a shallow bias relative to WOA13 in the tropical and subtropical periphery of the equatorial upwelling regions (Figure 16, upper right), ESM4.1 exhibits little evidence of such a pattern (Figure 16, lower right) while CM4.0 exhibits some evidence of deeper mixed layers. We interpret these differences largely to the stronger settings for the frontal length scale in the submesoscale mixed layer baroclinic eddy restratification scheme [Fox-Kemper et al., 2011] chosen for ESM4.1 (200 m compared to 500 m in CM4.0). This setting represents the dominant lateral length scale of the unstable mixed layer baroclinic

eddies. As described with sensitivity experiments in *Adcroft et al.*, [2019], in the absence of a theory and identification of strong sensitivity of SST biases to this parameter, a frontal length the most efficient parameter was optimized to reduce SST biases with lower frontal length values increasing both near surface restratification and overall SST in mid-latitude areas of strong meridional SST gradients. In the Southern Ocean, both CM4.0 and ESM4.1 provide relatively deep mixed layers relative to ESM2M and consistent with observations with those in CM4.0 slightly deeper than those in ESM4.1.

While a more detailed analysis of sea ice is provided below, we note here that minimum sea ice extent from the National Snow and Ice Data Center (NSIDC) processed using the NASA Team Algorithm (Data Set ID: NSIDC-0051; *Cavalieri et al.*, [1996]; black line, all panels in Figure 16) for years 1978-2011 exhibits good correspondence with all models in the Arctic (black dashed lines). However, there is considerable improvement in minimum sea ice extent in the Southern Ocean, which was nearly completely melted in ESM2M. This improvement is highlighted in the Weddell Sea where CM4.0 compares well with NSIDC and ESM4.1 exhibits a low (poleward) bias in sea ice extent.

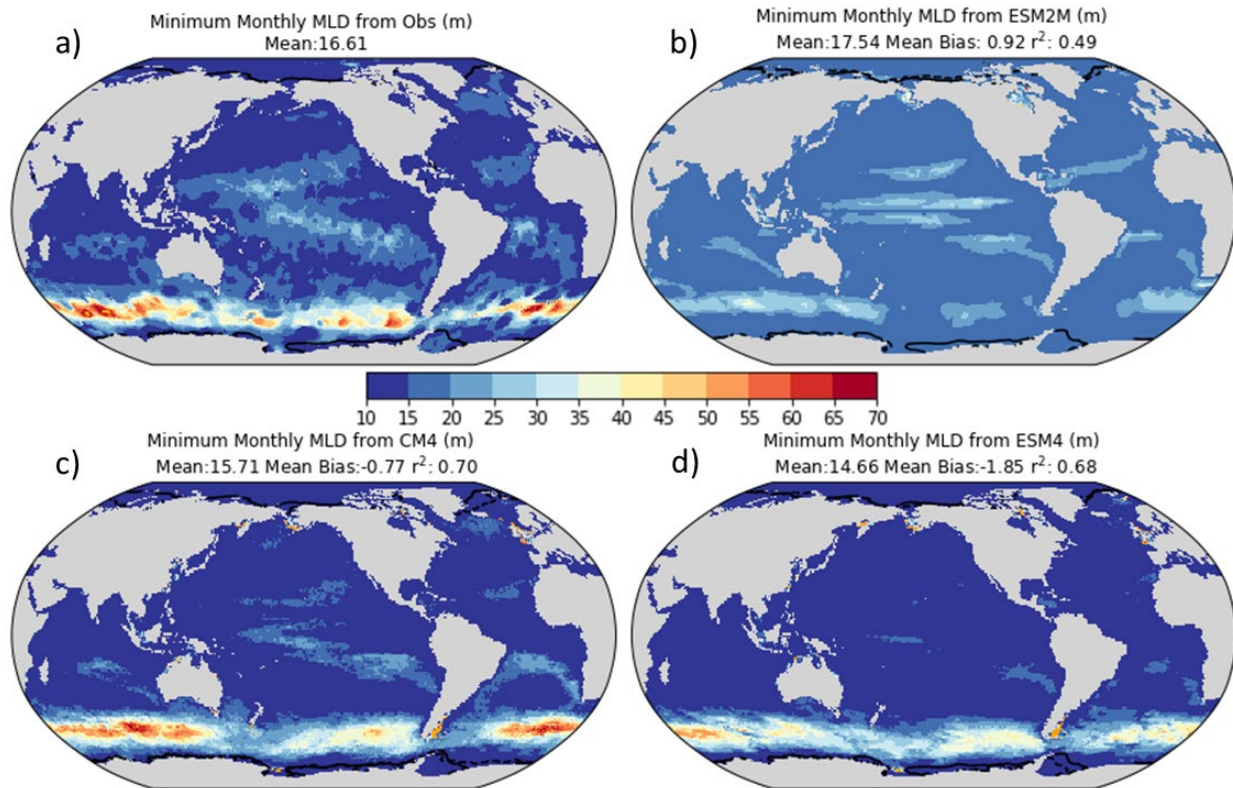


Figure 16: Monthly minimum mixed layer depths (m) calculated from the 0.03 density threshold relative to the upper 10 m in World Ocean Atlas 2013 (a; Locarnini et al., 2013; Zweng et al., 2013), ESM2M (b), CM4.0 (c), and ESM4.1 (d). Black lines correspond to minimum ice extent in NSIDC (solid) and model (dash).

Monthly maximum mixed layer depths (Figure 17) in WOA13 (upper left) compared to CM4.0 (lower left; $r^2 = 0.54$) and ESM4.1 (lower right; $r^2 = 0.51$) exhibit more than double the regional pattern variance representation relative to the previous generation ESM2M (upper right; $r^2 = 0.24$). The primary sources of this improved fidelity are improved representation of the relatively poleward onset of convective deep mixing in the North Atlantic, Southern Ocean, and North Pacific [Adcroft *et al.*, 2019]. However, both CM4.0 and even more so ESM4.1 exhibit a shallow bias relative to the WOA13 in the tropics. In the Southern Ocean, WOA13 exhibits a somewhat continuous line of deep mixing stretching from the central Indian Ocean at about 35°S to the tip of South America at about 57°S but not the Atlantic or Western Indian basins. ESM4.1 exhibits much of this structure except that it is much less a continuous ribbon than a series of deep mixing cores including an equatorward arm of deep mixing in the Central Pacific near 35°S and two centers of deep mixing in the western (~50°S) and eastern (~35°S) Atlantic Basins. Deep Southern Ocean mixing in CM4.0 (Figure 17, lower left) is meridionally broader and deeper than those in ESM4.1 or observations. ESM4.1 exhibits Western Pacific and Atlantic ribbons of relatively deep mixing at Northern midlatitudes compared to observations (Figure 17, lower right). Much like observations, deep mixing in the subpolar and polar North Atlantic in ESM4.1 is constricted to a line between the British Isles, South of Iceland and the southern tip of Greenland with an area of deep mixing in the central Norwegian Sea, while deep mixing in CM4.0 extends well into the subpolar gyre. Both ESM4.1 and CM4.0 exhibit deep mixing along the East coast of Siberia that does not appear in observations.

Comparison of maximum mixed layer depths and maximum sea ice extent in NSIDC observations (black line, all panels in Figure 17) demonstrate key features associated with the long-term evolution of interior ocean properties and resulting climate modulation in these models. In the Arctic, all models (dashed lines) exhibit fairly good correspondence with observations (solid lines). Regional exceptions are seen for ESM2M and CM4.0 which exhibit an overestimation of sea ice extent in the Barents Sea associated with a lack of deep convection and North Pacific associated with an overall relatively cold climate state. Both Arctic sea ice biases are substantially improved in both CM4.0 and ESM4.1 highlighted in Figure 17 through the linkage between maximum sea ice extent in the Southern Ocean. Whereas ESM2M exhibited permanent deep convection in the Weddell Sea and a strong associated low bias in sea ice extent, the low bias and associated deep convection center is removed in CM4.0 and ESM4.1 with CM4.0 exhibiting overly extensive maximum sea ice extent. The mechanisms involved and implications of the ability of this generation of model to sustain phases of stability and implications for multidecadal and centennial scale variability has been described previously for a prototype ocean under AM2 atmosphere [Zhang *et al.*, 2018], for SPEAR [Delworth *et al.*, 2019] for CM4.0 [Held *et al.*, 2019] and is described in more detail in the context of ESM4.1.

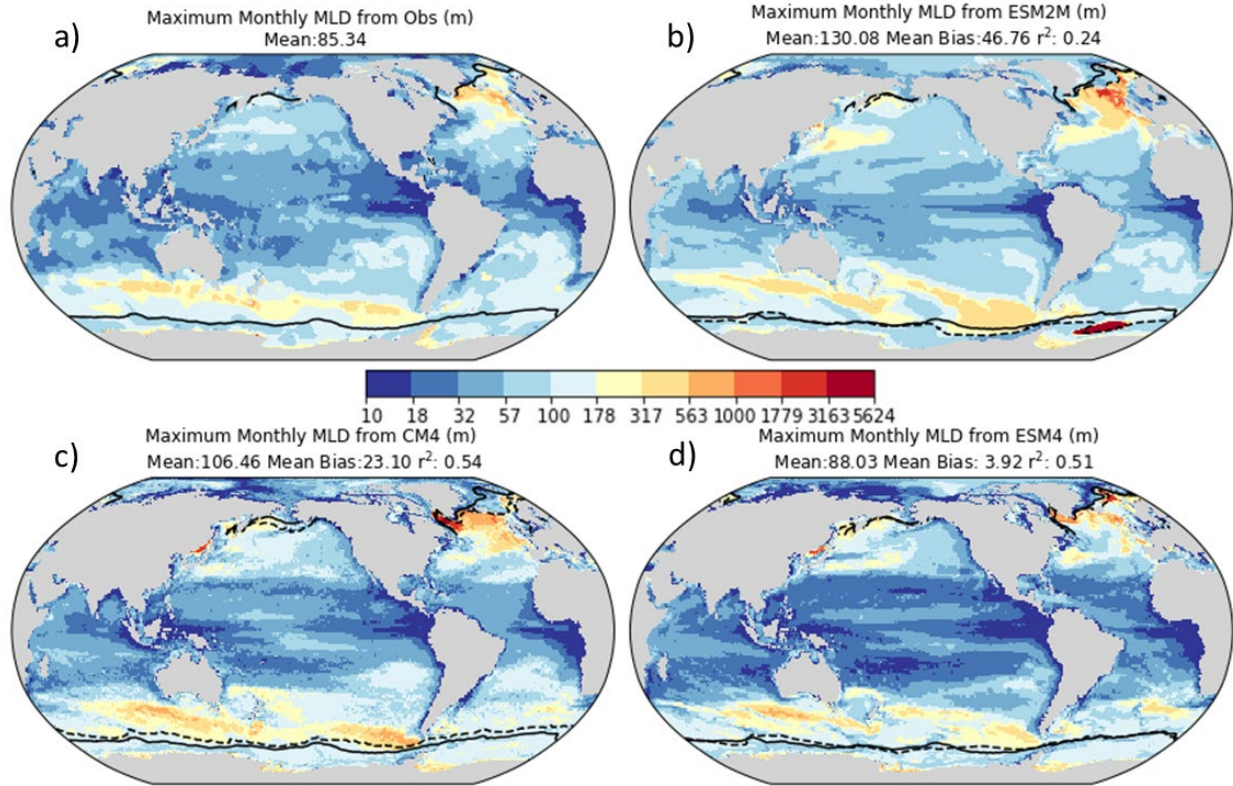


Figure 17: Monthly maximum mixed layer depths (m) calculated from the 0.03 density threshold relative to the upper 10 m in World Ocean Atlas 2013 (a; Locarnini et al., 2013; Zweng et al., 2013), ESM2M (b), CM4.0 (c), and ESM4.1 (d). Black lines correspond to maximum ice extent in NSIDC (solid) and model (dash).

Ocean meridional heat transport ESM4.1 (not shown) is nearly identical to that described for CM4.0 in *Held et al.* [2019]. In fact, global ocean meridional heat transport out of the tropics at 24°N (1.60 PW) and 19°S (0.93 PW) in ESM4.1 is stronger and compares slightly better to observations [*Ganachaud and Wunsch*, 2003; *Trenberth and Caron*, 2001] at both 24°N (1.8-1.9 PW) and 19°S (0.8-1.1 PW) than CM4.0 (1.54 PW at 24°N and 0.80 PW at 19°S). This result is somewhat surprising given the higher ocean resolution and subsequent ability to maintain boundary currents in CM4.0.

As discussed above, Northern Hemisphere sea ice distributions in Figure 18 from ESM4.1 (Figure 18B, C) compare well with observations from National Snow and Ice Data Center (NSIDC) processed using the NASA Team Algorithm (A; %; data set ID: NSIDC-0051, *Cavalieri et al.* [1996]) overall, with extremely good correspondence in Northern Hemisphere sea ice area seasonal cycle (Figure 18D) with a relative surplus of sea ice in the Norwegian and Barents Sea roughly compensating for a lack of sea ice in the Labrador Sea compared to observations. ESM4.1 suffers from a slight low bias in sea ice extent (Figure 18E), however,

indicating a relative consolidation of sea ice and lack of areas affected by partial ice compared to observations.

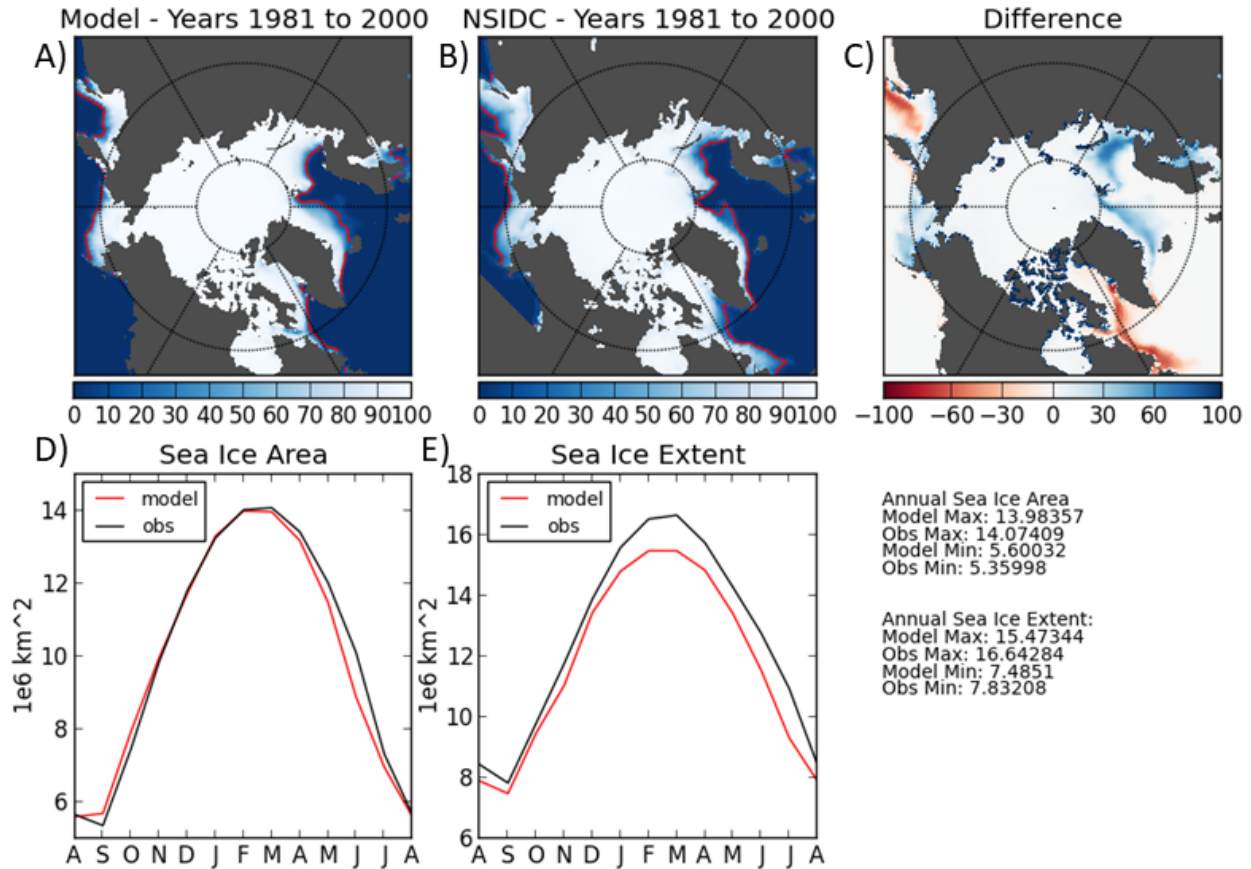


Figure 18: Comparison of Northern Hemisphere sea ice between National Snow and Ice Data Center (NSIDC) processed using the NASA Team Algorithm (A; %; data set ID: NSIDC-0051, Cavalieri et al. [1996]), ESM4.1 (B; %), the difference (C; %) as well as the seasonal cycle of sea ice area (D; 1e6 km²) and sea ice extent (E; 1e6 km²) in NSIDC (black) and ESM4.1 (red) for the period of 1981-2000. Also shown are statistics for Annual Sea Ice Area and Extent Model Maximum, Observations Maximum, Model Minimum and Observations Minimum in units of 1e6 km².

Southern Hemisphere sea ice distributions in ESM4.1 (Figure 19B, C) compare well with observations from NSIDC (Figure 19A) on the annual average but with an enhanced seasonal cycle of slightly too much maximum ice and too little summer ice in both area (Figure 19D) and extent (Figure 19E), with a relative surplus of sea ice seaward of West Antarctica and Wilkes Land roughly compensating for a lack of sea ice seaward of the Ross Sea and the Scott Mountains (~60°E) compared to observations.

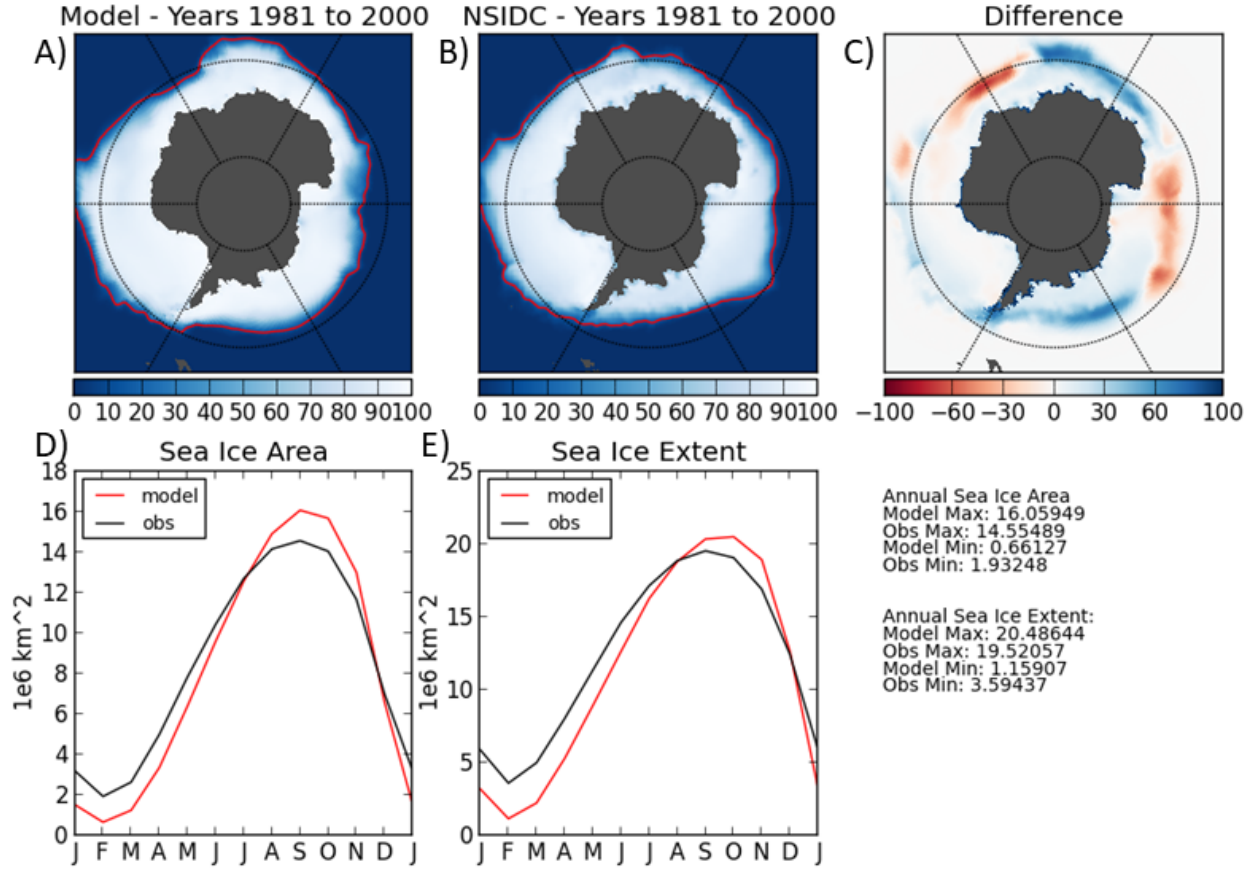


Figure 19: Comparison of Southern Hemisphere sea ice between National Snow and Ice Data Center (NSIDC) processed using the NASA Team Algorithm (A; %; data set ID: NSIDC-0051, Cavalieri et al. [1996]), ESM4.1 (B; %), the difference (C; %) as well as the seasonal cycle of sea ice area (D; 1e6 km²) and sea ice extent (E; 1e6 km²) in NSIDC (black) and ESM4.1 (red) for the period of 1981-2000. Also shown are statistics for Annual Sea Ice Area and Extent Model Maximum, Observations Maximum, Model Minimum and Observations Minimum in units of 1e6 km².

4.10 Interior Ocean Properties

As shown in Figure 2C, ESM4.1 exhibits a global average ocean warming. This drift in ESM4.1 ocean interior temperature (and salinity) reflect a diverse suite of processes and time scales over its 400 years of spinup and 100 years of piControl integration before the historical 1850-2015 simulation was launched and are more difficult to interpret in the transient. This warming is far from spatially uniform, however, being focused at the base of the thermocline and deep Atlantic and Indian basins as shown in Figure 20 (upper panel). The upper thermocline, rather, has a cold bias, most notably in the Pacific. This combination of cold over warm, or “diffuse thermocline” bias [Stockdale et al., 1998] is a long standing challenge in climate modeling relating to the lack

of restratification process to work against wind, tide and heat flux driven buoyancy supply into the ocean interior [Griffies *et al.*, 2015]. The salinity bias in ESM4.1 (Figure 20, lower panel) develops as overly warm, salty North Atlantic Deep Water leaving a deficit of salt in the Pacific thermocline. Strong, shallow overflows of the Red Sea in the Indian Ocean at 200-400 m and Mediterranean Sea in the Atlantic Ocean at 600-1000 lead to loci of warm, salty water suggesting that the overflow scheme overly enhanced shallow mixing. Overall, we see strong improvement in the representation of interior ocean temperature across model generations. Comparison of mean bias, explained variance, and RMSE for years 501-600 of preindustrial spinup are progressively improved from ESM2M to CM4.0 and ESM4.1 over the upper 5000 m with bias reductions from 0.27°C in ESM2M to 0.12°C and 0.14°C in CM4.0 and ESM4.1, respectively, r^2 increases from 0.84 in ESM2M to 0.88 and 0.92 in CM4.0 and ESM4.1, respectively, and RMSE reductions from 0.71°C in ESM2M to 0.63°C and 0.54°C in CM4.0 and ESM4.1, respectively.

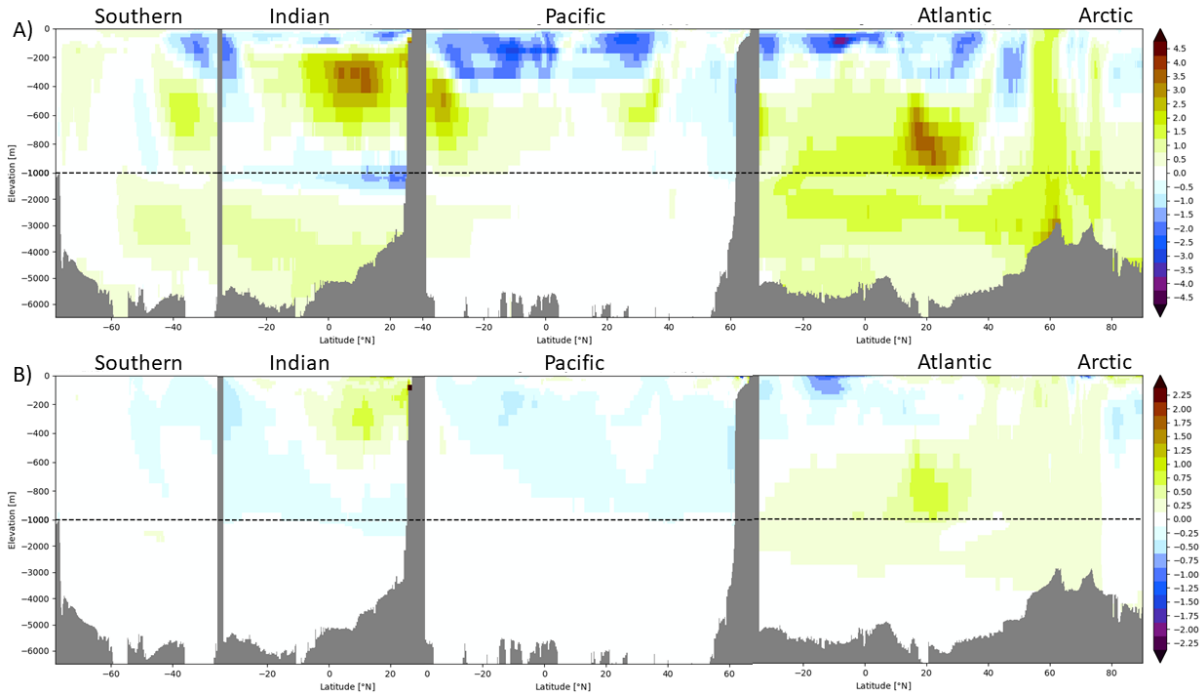


Figure 20: Basin zonal mean bias in ESM4.1 temperature (A; °C) and salinity (B; ppt) versus latitude compared to World Ocean Atlas, 2013. Note split axis to better resolve the upper 1000 m.

Meridional Overturning in GFDL's CM4.0 and ESM4.1 share several improvements relative to the previous generation ESM2M and ESM2G models [Dunne *et al.*, 2012] but also several deficiencies. As described in Adcroft *et al.* [2019] and Held *et al.* [2019], Southward Atlantic Meridional Overturning Circulation (AMOC) at 24°N (Figure 21A) in CM4.0 (dark blue) and ESM4.1 (light blue) begins shallower than observations (black) and previous modeling efforts (ESM2M, red, and ESM2G, green) and terminates at 3000 m with northward flow below as did

ESM2M. AMOC exiting in the South Atlantic at 30°S (Figure 21B), however, deepens in CM4.0 and ESM4.1 to agree with observational products with bottom southward circulation in ESM4.1 extending deeper akin to ESM2G than in CM4.0 akin to ESM2M. In the North Pacific at 24°N (Figure 21C), both ESM4.1 and CM4.0 are much more consistent with observations than previous generation models in exhibiting considerably less northward supply of deep water upwelled through the North Pacific thermocline before advecting Southward within the thermocline. As shown in Figure 21D, both CM4.0 and ESM4.1 exhibit relatively weak Antarctic Bottom Water transport into the Southern Indo-Pacific below 4000 m with much of that water leaving the Indo-Pacific between 3000-4000 and a secondary transport of water into the Indo-Pacific between 1500-2500 m upwelling through the tropical Indo-Pacific thermocline before exiting.

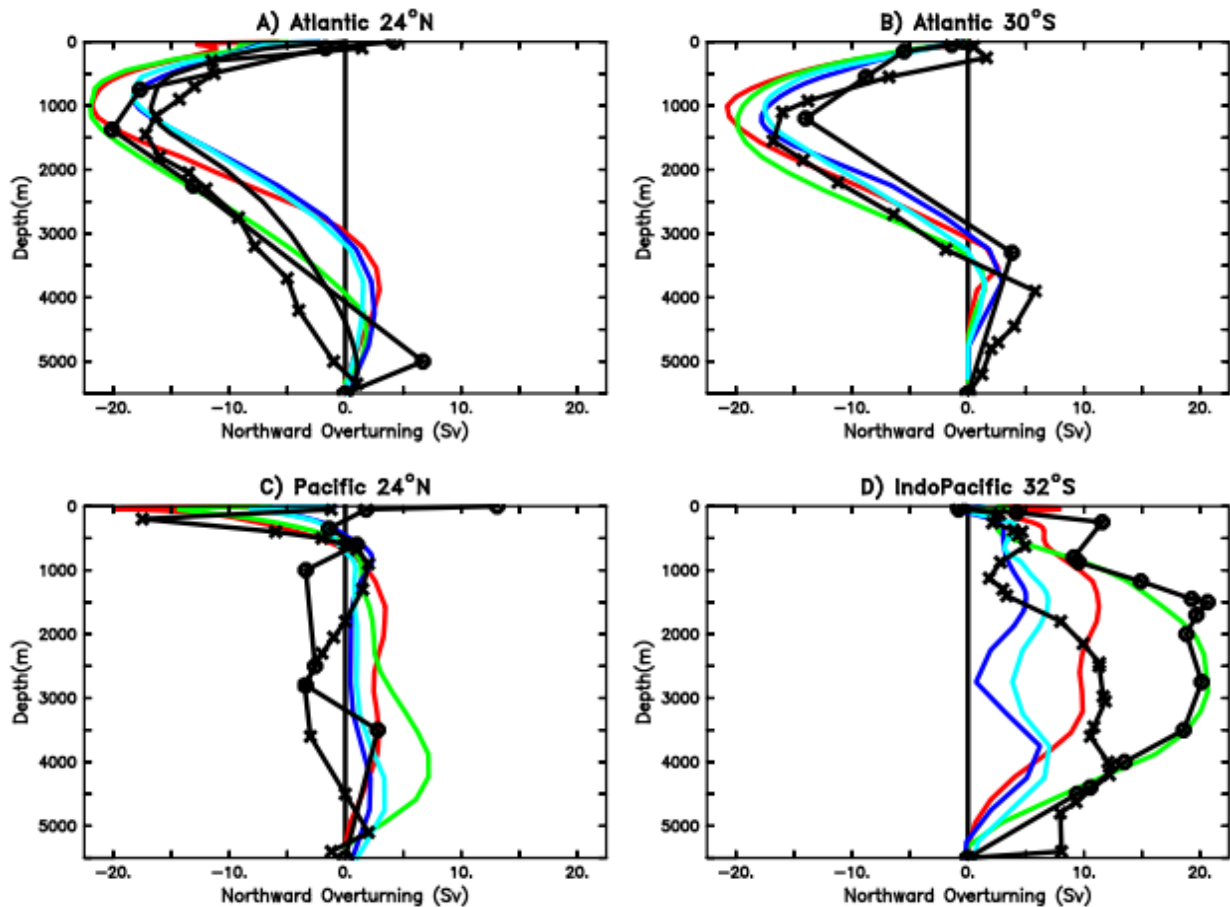


Figure 21: Meridional Overturning comparison with observational estimates taken from Dunne et al. (2012; their Figure 8) of depth distributions of the overturning streamfunction (Sv) at various zonal bands for the North Atlantic at 24°N (A), South Atlantic at 30°S (B), Pacific at 24°N (C), and Southern Indo-Pacific at 32°S (D) in ESM2M (red) and ESM2G (green) compared to observationally based estimates from Talley et al. (2003; open circles), Ganachaud and Wunsch [2003; crosses] and RAPID array at 26.5°N over the period 2004-2015 [Cunningham et

al., 2007; *Smeed et al.*, 2018]. Also shown are results from Historical simulation of CM4.0 (dark blue) and ESM4.1 (light blue). The overturning in the models is computed just from the Eulerian mean transport to remove effects from parameterized mesoscale and submesoscale eddies and integrated from the bottom, so the simulated strength relative to the surface is defined as the difference between the maximum value and the surface value (on the order of 1 Sv due to freshwater transport in the Atlantic).

ESM4.1 and CM4.0 have notable differences in their oceanic inventory of CFC-11, an inert tracer commonly used to assess ventilation and part of the OMIP protocol [*Orr et al.*, 2016]. The longitudinally and depth-integrated inventory of CFC-11 was calculated for year 1994 from both models to compare with GLODAPv1 gridded observations. Total column inventories (0-5000m) of CFC-11 from CM4.0 compare better against observations than ESM4.1, but much of the uptake in the midlatitudes in CM4.0 is associated with an overly high bias in uptake below the thermocline region (700-2000 m), whereas CFC-11 uptake at this depth range is greatly improved in ESM4.1. The role of deep convection in tracer uptake is evident when considering the deep inventory (2000-5000 m) in both models. One ESM4.1 historical ensemble member (initialized from year 201 of the piControl simulation; green line in Figure 22) experiences a large open-ocean polynya after year 2000 and this full-column sustained convection leads to better agreement with the GLODAPv1 inventory of CFC-11. In contrast, the other ESM4.1 ensemble members (initialized from year 101 and 151 of the piControl simulation; blue and orange lines, respectively in Figure 22) experience timing of polynya activity around 1905 and 1950, respectively, and ventilate less in the deep ocean. CM4.0, with its much shorter spin-up period than ESM4.1, has had less time to develop a vertical structure that is conducive large polynya activity and has the least amount of CFC-11 in the deep Southern Ocean. With respect to the length of the spin-up period, the mean sea surface state of ESM4.1 is warmer than CM4.0, and thus its surface solubility with respect to CFC-11 is reduced. This effect is consistent with generally less CFC-11 uptake in ESM4.1 compared with CM4.0 at all depth ranges and latitudes, particularly in the near-surface tropical ocean.

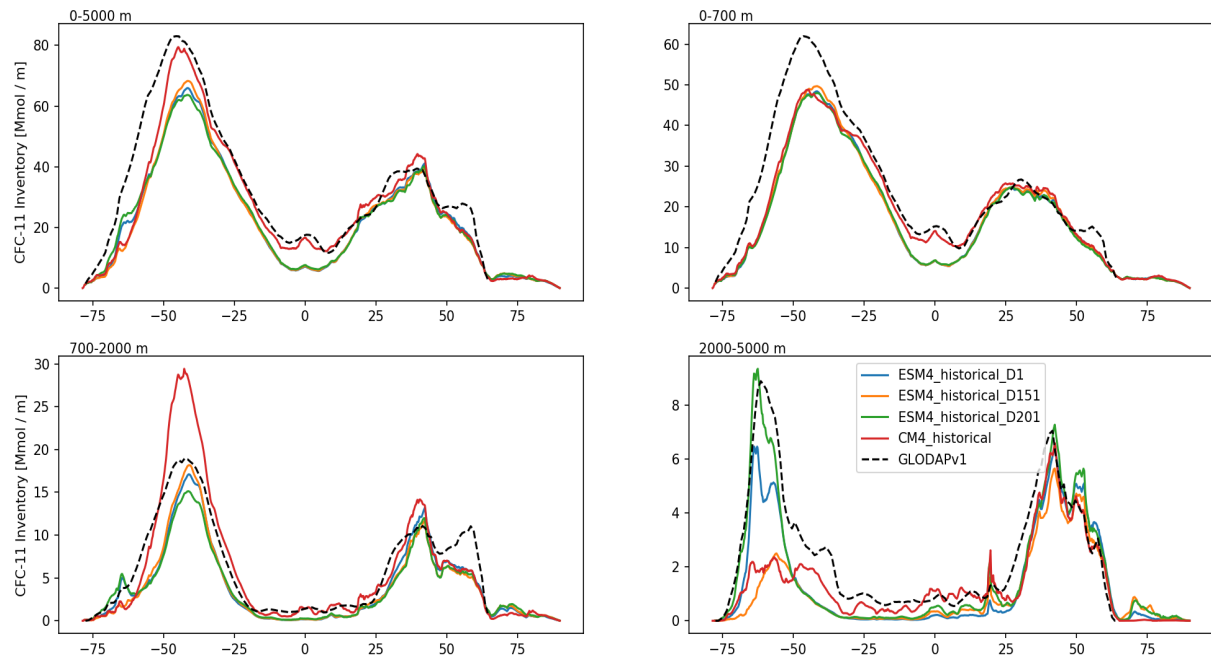


Figure 22: Comparison of ESM4.1 ensemble members initialized from different years (blue – 101, orange – year 151, green - 201) and CM4.0 (red) inventories of CFC-11 with the GLODAPv1 data product (Key et al., 2004; black dash) for 0-5000 m (upper left) as well as the thermocline (0-700 m; upper right), interior (700-2000 m; lower left) and abyss (2000-5000 m; lower right).

4.11 Biogeochemistry and Ecosystems

Ecosystem and biogeochemistry comparison with observations is treated elsewhere for the land [Shevliakova *et al.*, in review] and ocean [Stock *et al.*, in review] components of ESM4.1. However, we provide a brief treatment of ocean biogeochemical properties in comparison with the CM4.0 in Figure 23 in the Tropical Indo-Pacific (solid) and Atlantic (dash). Note that the historical simulation for CM4.0 began after 250 years of spinup from observations while that for ESM.1 began after 500 years, so ESM4.1 has had more opportunity to accumulate bias on its path to equilibrium. Overall, ESM4.1 compares well with observations of PO_4 (Figure 23A), O_2 (Figure 23B), dissolved inorganic carbon (DIC; Figure 23C) and excess alkalinity (Alk-DIC; Figure 23D) and illustrates biases similar to CM4.0 with good representation of the upper nutricline/oxycline. In the tropical Atlantic, ESM4.1 exhibits both an underestimation of PO_4 and O_2 in the main nutricline/oxycline and mid-depth (1000-3000 m) biases consistent with a strong, shallow AMOC (Figure 21A). In the Indo-Pacific, the main nutricline/oxycline are reproduced. In both the Atlantic and Indo-Pacific, ESM4.1 exhibits the signature of excess remineralization in the abyss in all four variables.

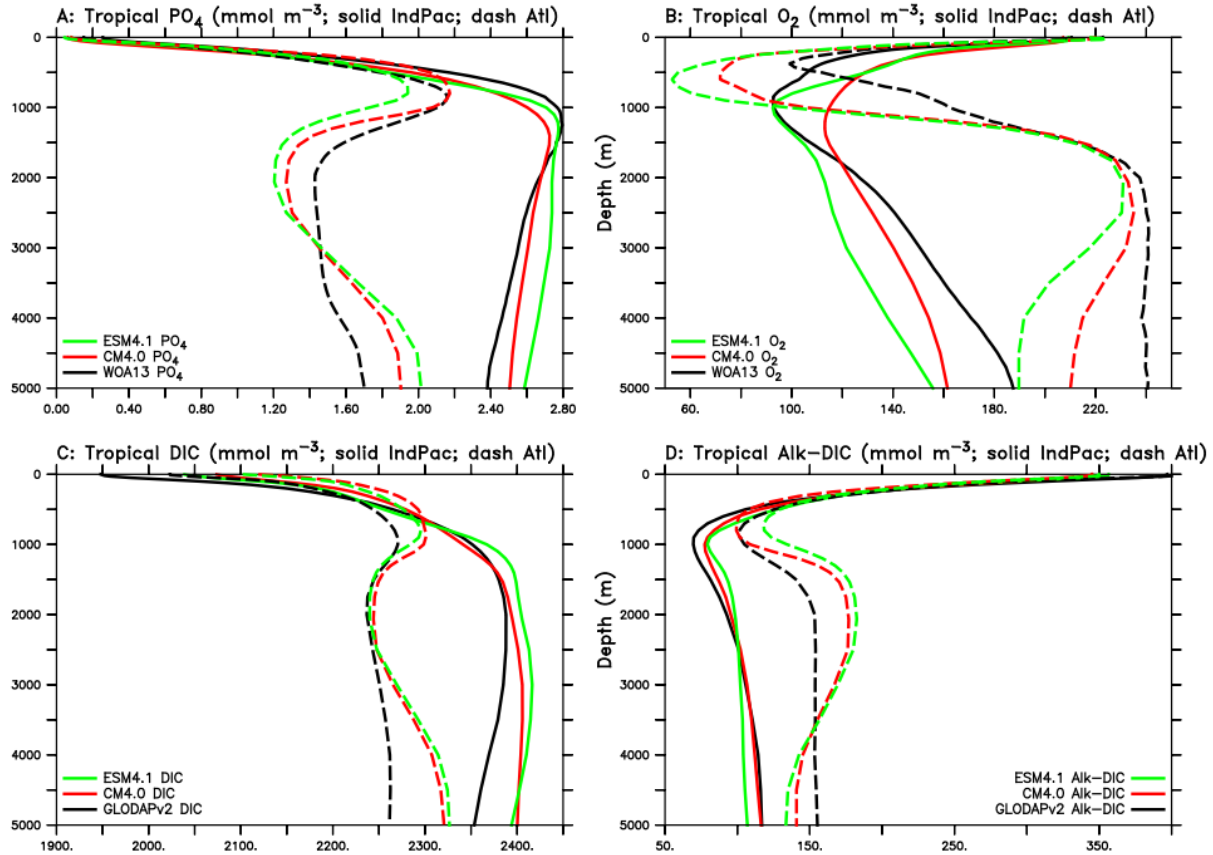


Figure 23: Tropical (30°S-30°N) basin average concentrations (μM) for PO_4 (A), O_2 (B), dissolved inorganic carbon (DIC; C) and Alkalinity minus DIC (D). World Ocean Atlas 2013 PO_4 and O_2 and GLODAPv2 DIC and Alkalinity are shown in black, and CM4.0 in red, and ESM4.1 in green for the Indo-Pacific (solid) and Atlantic (dash) averaged over 1990-2009 of the historical simulation.

Analysis of the terrestrial biosphere through the Köppen climate types [Köppen, 1923; Lohmann *et al.*, 1993; Gnanadesikan and Stouffer, 2006] is shown in Figure 24 for observations from New *et al.* [1999] (Figure 24a) and ESM4.1 (Figure 24b). Areas of disagreement between Köppen climate types highlights the strong improvements made in the representation of regional climates over previous generation models. Whereas Gnanadesikan and Stouffer [2006] found a misclassification by type (e.g. Arid or Semi-arid simulated in ESM4.1 as Humid Continental such as in Western North America; Figure 24c) of 30.8% in CM2.1 (not shown), this misclassification has dropped to 17.1% in ESM4.1. Moreover, misclassification by subtype (e.g. Tropical Rainforest simulated in ESM4.1 as Tropical Savanna such as in the Amazon; Figure 24d) dropped from 25.9% to 15.1% such that correct classification by both type and subtype increased from less than half (43.3%) in GFDL's second generation models to slightly better than two thirds (67.8%) in the fourth generation. Some of this improvement can be attributed to the representation of topography such as the Andes providing stark climate differences on either side of the range (Figure 24b). Major areas of bias include the under-representation of Tropical

Rainforest (Af) and monsoon (Am) in the Amazon and South Asia. This under-representation of wet areas has important implications for the magnitude and timing of carbon fluxes. Other major areas of bias include under-representation of Semi-Arid (Bs) and desert (Bw) areas in the Western North America, Central Europe, South America, southern Africa, and central Asia. This under-representation of deserts has important implications for the prognostic representation of dust emissions.

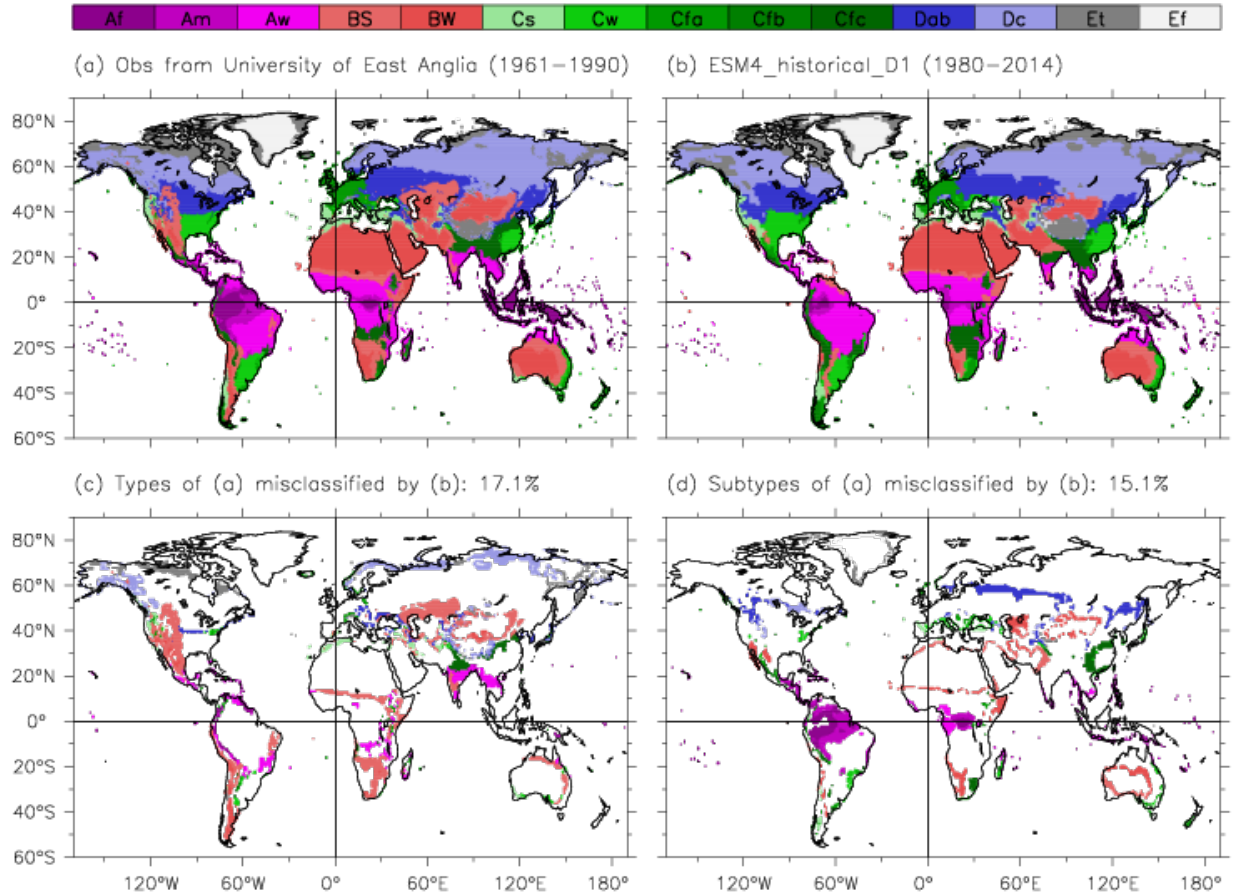


Figure 24: Köppen [1923] climate types as described in Gnanadesikan and Stouffer (2006) for observations from the University of East Anglia Climate Research Unit half-degree monthly climatology (A; UEA CRU05; New et al., 1999) and ESM4.1 (b) including tropical rainforest (Af), monsoon (Am), and savanna (As), semiarid (Bs) and desert (Bw), dry summer (Cs), dry-winter oceanic (Cw), humid subtropical (Cfa), oceanic (Cfb), subpolar oceanic (Cfc), humid continental (Dab), subarctic and boreal (Dc), tundra (Et), and ice cap (Ef) climates based on monthly climatology of temperature and precipitation. Areas of major type misclassifications (c) where model and observations type disagree (e.g. Bs where observations give As), and sub-type misclassifications (d) where model and observations capital letter agree but subtype disagree (e.g. Am where observations give Af) are also shown.

4.12 Tropical Oceans, El Niño / Southern Oscillation (ENSO) and teleconnections

Based on a historical ensemble run of ESM4.1 (3 members, 1880-2014), the annual-mean climatological biases of ESM4.1 (Figure 25) appear similar to those in CM4.0. ESM4.1 has warmer tropical SSTs but a cooler upper-ocean than CM4.0 — giving shallower tropical mixed layers, which were already shallower than observed in CM4.0. Compared to CM4.0, ESM4.1 also has somewhat larger tropical climatological rainfall biases (poleward-shifted ITCZ and double ITCZ), weaker climatological zonal surface wind stress off-equator, and more cyclonic wind stress curl over the tropical Pacific. This enhanced cyclonic curl acts to shoal the equatorial Pacific thermocline via poleward Sverdrup transport [Wittenberg *et al.*, 2018], exacerbating the equatorial Pacific subsurface cold bias in ESM4.1 relative to CM4.0.

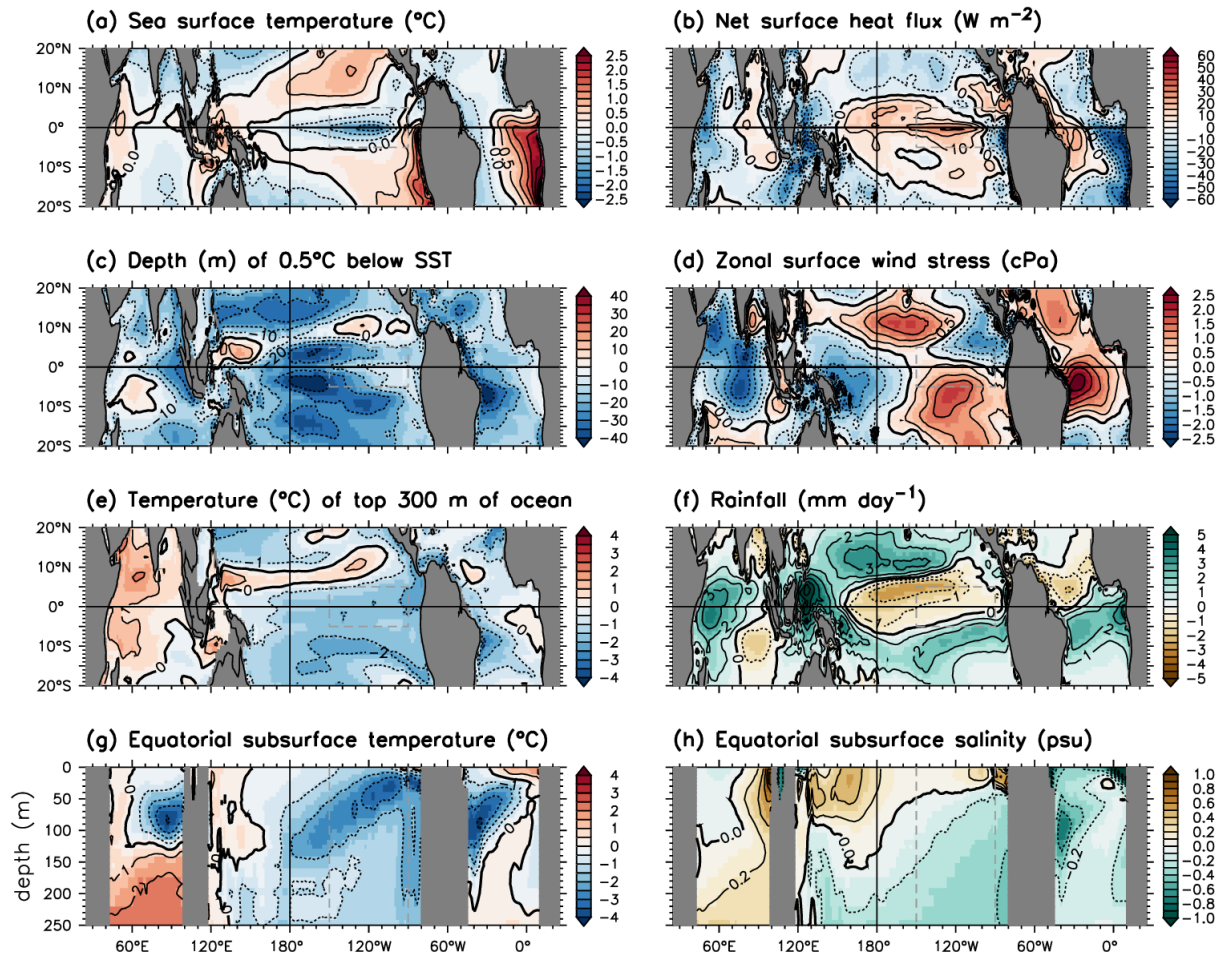


Figure 25: Annual-mean (1980-2014) climatological biases of ESM4.1 relative to observational reanalyses. (a) Sea surface temperature (SST, °C). (b) Net surface heat flux (W m^{-2}); positive values heat the ocean. (c) Depth (m) at which the time-mean ocean temperature falls 0.5 K below the time-mean SST. (d) Zonal component of the surface wind stress (cPa); positive values exert a westerly (i.e. eastward) stress on the ocean. (e) Vertically-averaged temperature of the top 300 m of the ocean (shallower regions are masked out). (f) Surface rainfall (mm day^{-1}). (g) Subsurface temperature (°C) interpolated to the equator. (h) Subsurface salinity (psu)

interpolated to the equator. Shading is incremented every half contour. ESM4.1 represents a mean of three historical runs with CMIP6 historical forcing. The observational reference datasets for the 1980-2014 period are: the ECMWF Interim reanalysis (Dee et al., 2011) for panels a,b,d; the ORA-S4 reanalysis (Balmaseda et al., 2013) for 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 region (150°W–90°W, 5°S–5°N).

As shown in *Held et al. (2019)*, GFDL’s CM4.0 provides a high standard for simulation of the El Niño / Southern Oscillation (ENSO) and its teleconnections. Based on the historical ensemble and the 700-year piControl run, ESM4.1 exhibits a Niño-3 SST spectrum nearly as good as that in CM4.0, with a slightly stronger and longer-period ENSO than observations (Figure 26). The ENSO period in both observations and CM4.0 peaks at about 3.3 years, while the period in ESM4.1 peaks at about 4 years. Like CM4.0, ESM4.1 shows a realistically broad spectral peak, arising from both the episodic character and the interdecadal modulation of ENSO [*Larkin and Harrison, 2002; Kessler, 2002; Wittenberg, 2009; Wittenberg et al., 2014; Atwood et al., 2017*]. Also like CM4.0, ESM4.1 shows little clear change in the ENSO spectrum between the piControl run (driven by 1850 forcing) and the historical epoch (1880-2014).

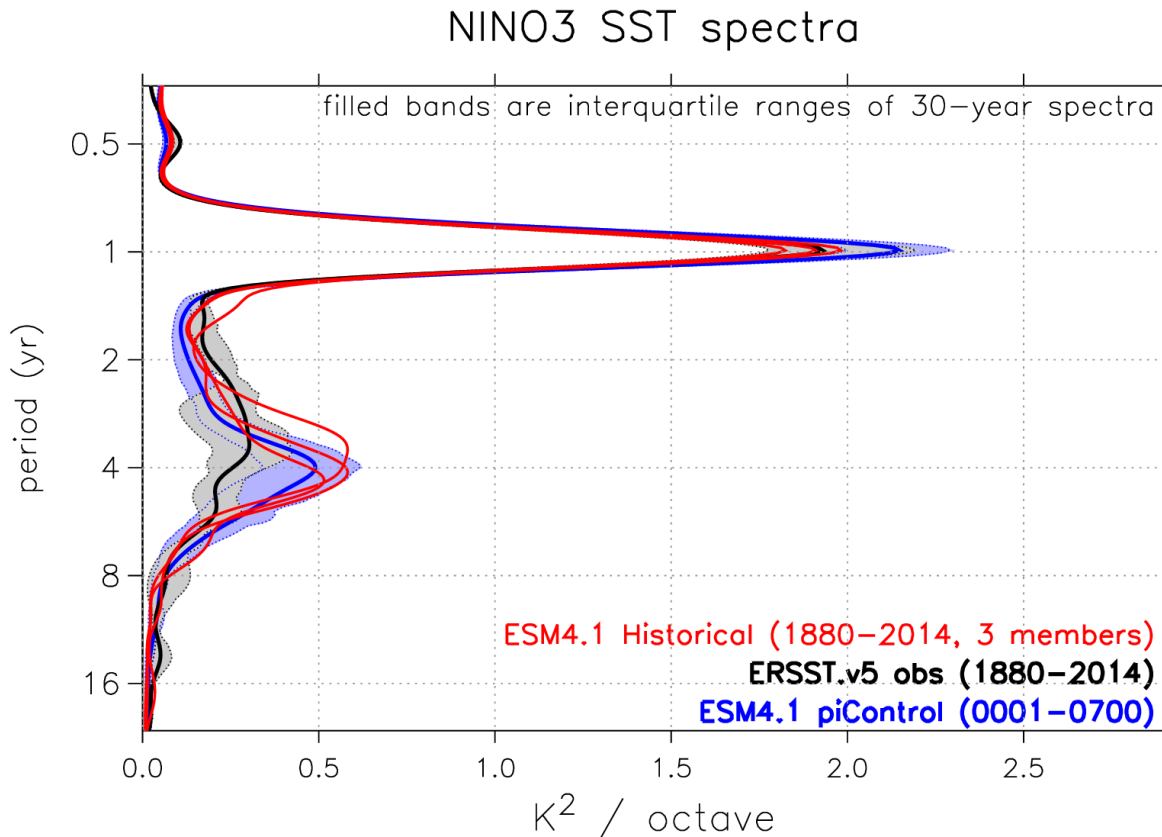


Figure 26: Power spectra of SST averaged over the Niño-3 region (150°W–90°W, 5°S–5°N), following Fig. 2 of Wittenberg (2009). Abscissa indicates the time-averaged spectral power (K^2

octave⁻¹) from a Morlet wavenumber 6 wavelet analysis. The area to the left of each curve represents the total spectral power within a given range of periods on the ordinate. Red lines are 1880–2014 time-mean spectra for the three ESM4.1 historical ensemble members. Black curve is the 1880–2014 time-mean spectrum from the ERSST.v5 observational reanalysis [Huang *et al.*, 2017]; gray band indicates the interquartile range of ERSST.v5 sliding 30-year sub-spectra during 1880–2014. Blue curve is the ESM4.1 piControl run time-mean spectrum during simulated years 1-700; light blue band is the interquartile range of sliding 30-year sub-spectra during that 700-year epoch.

The longer ENSO period in ESM4.1 relative to CM4.0 may be linked to ESM4.1's meridionally-broader tropical Pacific westerly wind stress anomalies during El Niño (Figure 27), which would weaken the off-equatorial cyclonic curl anomalies during El Niño, in turn leading to slower poleward discharge of equatorial Pacific heat and a slower ENSO cycle [Kirtman, 1997; Capotondi *et al.*, 2006]. ESM4.1's slightly stronger ENSO relative to observations may be linked to (1) its weaker damping from surface fluxes, due to less convective cloud and associated shortwave damping of SST anomalies near the equator; and (2) its stronger thermal stratification in the equatorial Pacific cold tongue region (Figure 25c,g), which may boost the local sensitivity of SST to fluctuations in upwelling and thermocline depth.

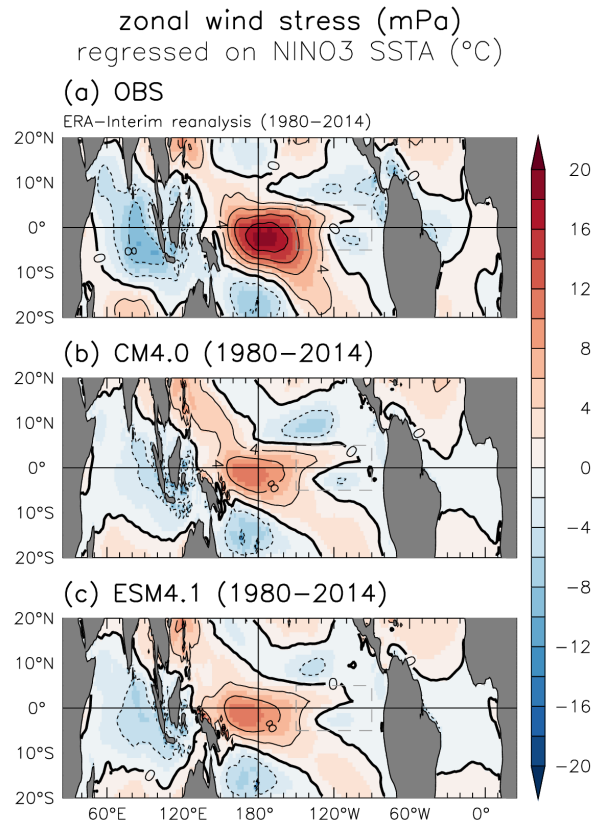


Figure 27: Linear regression of monthly zonal surface wind stress anomalies onto monthly Niño-3-averaged SST anomalies (mPa K⁻¹) for all months during 1980-2014. Positive values indicate

anomalous westerly (i.e. eastward) stress on the ocean during El Niño. (a) Observational estimate based on the ERA-Interim reanalysis [Dee *et al.*, 2011]. (b) Ensemble mean of the three 1980-2014 regressions computed from the three CM4.0 historical runs. (c) As in (b) but for the three ESM4.1 historical runs. All datasets are averaged onto a $1.25^\circ \text{lon.} \times 1^\circ \text{lat.}$ grid before computing the temporal regression. Dashed gray box is the Niño-3 region (150°W – 90°W , 5°S – 5°N).

The pattern of ENSO SST anomalies in ESM4.1 (not shown) is very similar to that in CM4.0, although the peak in eastern equatorial Pacific is slightly stronger and farther west in ESM4.1. The tropical Pacific rainfall sensitivity to ENSO (Figure 28), which drives many of ENSO's remote teleconnections, is slightly weaker and farther west in ESM4.1 than CM4.0, slightly worsening the bias relative to observations. Overall, ESM4.1 provides a reasonable simulation of ENSO, compared to CM4.0 and earlier (e.g. CMIP3 and CMIP5) models [Wittenberg *et al.*, 2006; Dunne *et al.*, 2012a; Delworth *et al.*, 2012; Vecchi *et al.*, 2014; Guilyardi *et al.*, 2012; Chen *et al.*, 2017; Graham *et al.*, 2017].

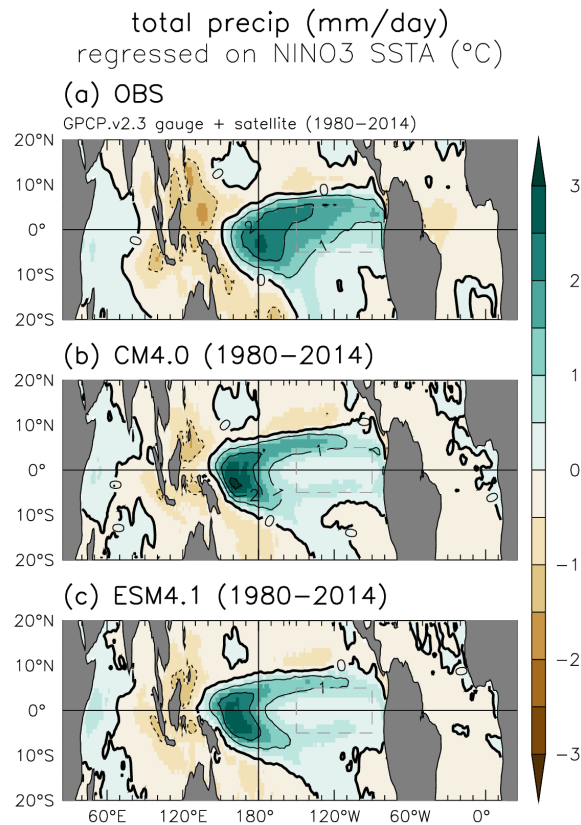


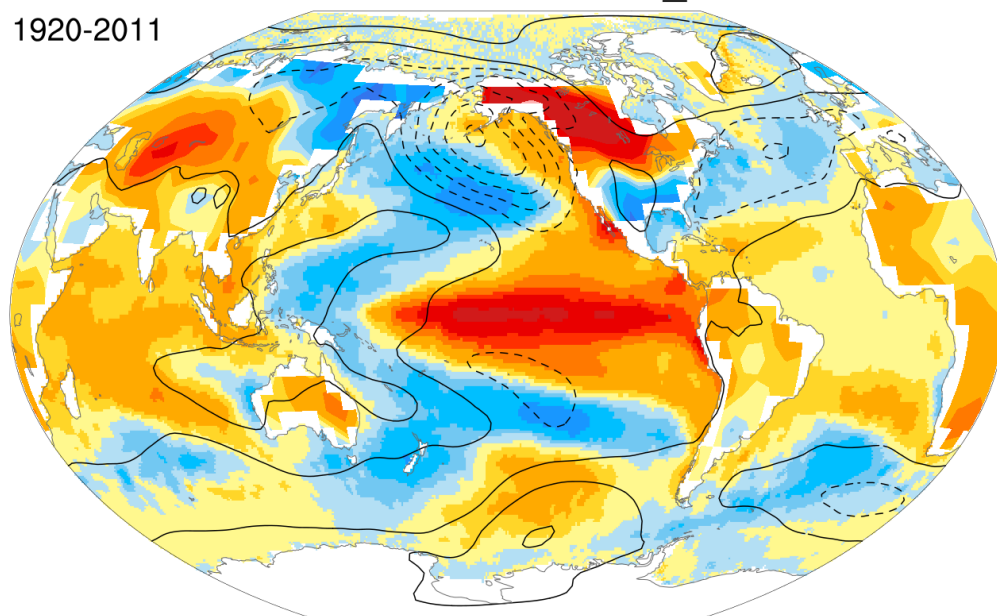
Figure 28: As in Figure 27, but for monthly rainfall anomalies regressed onto Niño-3 SST anomalies ($\text{mm day}^{-1} \text{K}^{-1}$). Observations in (a) are from GPCP.v2.3 (Adler *et al.*, 2003, 2016) for rainfall, and ERSST.v4 (Huang *et al.*, 2015) for SST.

A composite map of the winter time temperature and sea level pressure response to the previous ENSO is shown for observations from HadISST [Rayner et al., 2003], NOAA's Merged Land-Ocean Surface Temperature Analysis (MLOST) [Vose et al., 2012], and the 20th Century Reanalysis Project version 2 [Compo et al., 2011] and ESM4.1 in Figure 29. ESM4.1 has a strong fidelity in the global temperature teleconnection pattern over all oceans as well as North and South Africa, and Australia and captures the observed pattern of low pressure center in the Eastern North Pacific and high pressure center in the South Central Pacific extremely well.

nino3.4 TS,TAS,PSL Spatial Composite (DJF⁺₁)

HadISST / MLOST / 20thC_ReanV2

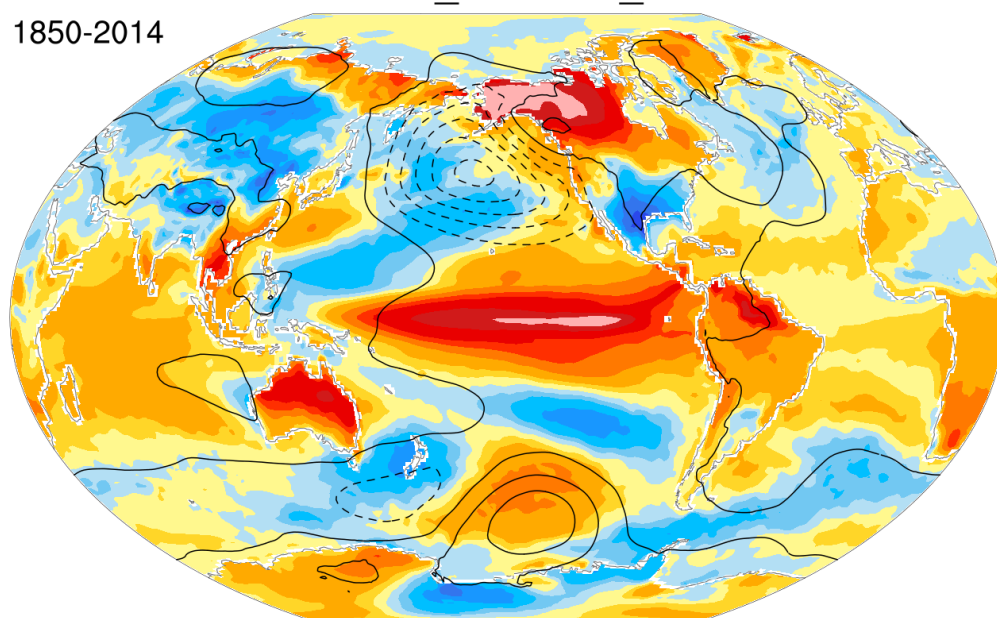
1920-2011



CONTOUR FROM -16 TO 16 BY 2

ESM4_historical_D1

1850-2014



CONTOUR FROM -16 TO 16 BY 2



Figure 29: Multi-parameter comparison of sea level pressure (contours) and composite surface and surface air temperature (shading). The normalized December nino3.4 time series is used to

composite all years greater than 1 standard deviation (El Niño) and all years less than -1 standard deviation (La Niña). The December nino3.4 time series is based on the December values of the monthly nino3.4 time series smoothed with a 3-point binomial filter after Deser et al. [2012]. Observations are based on HadISST (Rayner et al., 2003), NOAA's Merged Land-Ocean Surface Temperature Analysis (MLOST), and the 20th Century Reanalysis Project version 2 [Compo et al., 2011]. This figure was created using NCAR Climate Variability and Diagnostics Package (CVDP) Version 3.7.0 (Phillips et al., 2014).

4.13 North Atlantic Variability

Significant climate variability in the areas surrounding the North Atlantic is associated with a sea surface temperature pattern called the Atlantic Multidecadal Oscillation (AMO) [Kerr, 2000]. Previous work has delved into the representation of AMO in previous GFDL models and its connection to Atlantic meridional overturning variability [e.g. Delworth et al. 2017]. In Figure 30, we present a comparison of the pattern of AMO in HadISST and ESM4.1 based on [Trenberth and Shea, 2006] illustrating the strong pattern correspondence between observations (Figure 30A) and the ESM4.1 historical (Figure 30B) and piControl (Figure 30C) simulations with peaks in the Irminger Sea and Eastern tropical North Atlantic. The AMO pattern in observations exhibits strong multidecadal periodicity (Figure 30D) out to 50 years (note the record length is only 92 years). While the ESM4.1 piControl (Figure 30F) exhibits a similarly strong multidecadal periodicity, that in the ESM4.1 historical simulation is moderately weaker and includes an additional roughly 5 year period of oscillation. Given its absence in the longer control simulation, we cannot speak to the robustness of this feature in the present analysis.

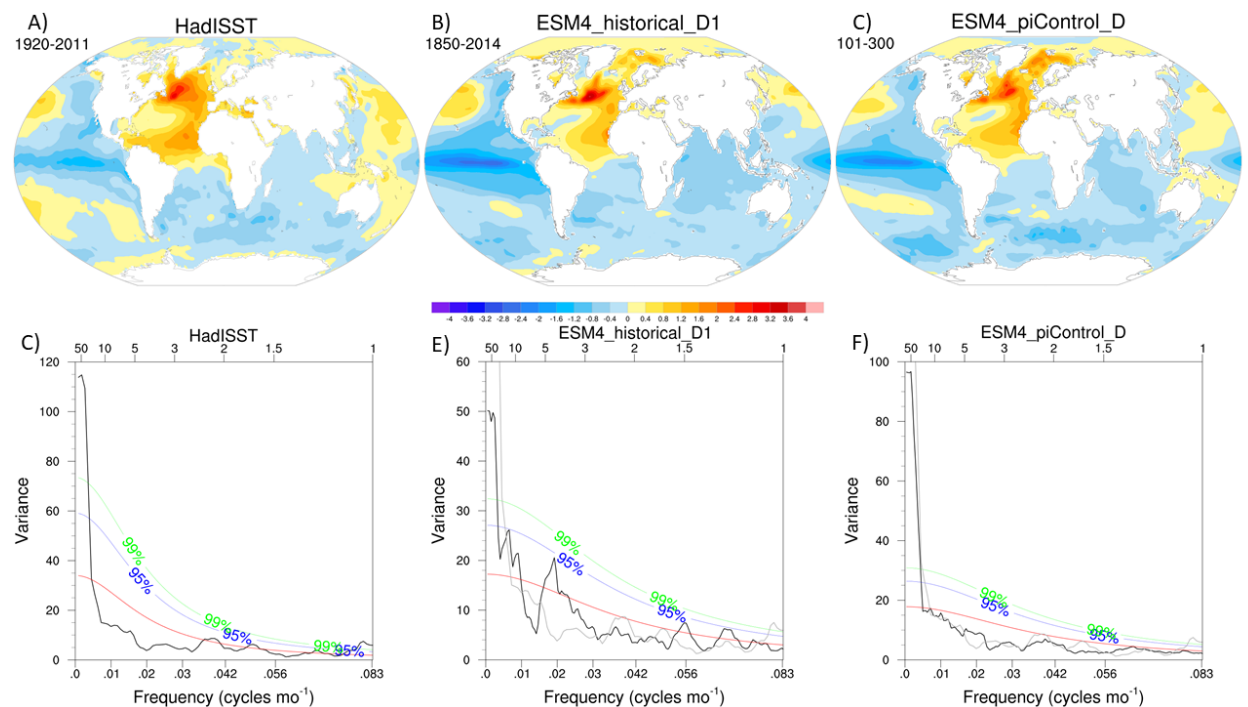


Figure 30: Comparison of the pattern of the Atlantic Meridional Oscillation (AMO) in HadISST from monthly index time series defined as North Atlantic (0:60°N, 80°W:0°E) SST anomalies minus global (60°S:60°N) SST anomalies. Pattern created by regressing global SST anomalies onto index time series and smoothing with a 9-point spatial filter. Low pass-filtered time series (black curve) is based on a 61-month running mean based on [Trenberth and Shea, 2006]. For power spectra presented for HadISST (B) and ESM4.1 (D), are displayed in variance-conserving form as the best-fit first-order Markov red noise spectrum (red curve) and its 95% (blue curve) and 99% (green curve) confidence bounds are shown on each panel. Top X-axis shows the period (in years), and the bottom X-axis shows the frequency (cycles/mo). For ESM4.1 (D) the observational spectrum is overlaid in gray. This figure was created using NCAR Climate Variability and Diagnostics Package (CVDP) Version 3.7.0 (Phillips et al., 2014).

4.14 Centennial scale Southern Ocean variability

One of the most intriguing features of this generation of GFDL models is the manifestation of strong Southern Ocean variability on centennial timescales. This variability was first documented in a MOM6 ocean coupled to the previous generation AM2 atmosphere [Zhang et al., 2018] and further described in the Seamless Prediction of the EArth system (SPEAR; [Delworth et al., 2020]). The Polar Southern Ocean (Weddell and Ross Sea) region is somewhat unique in the world ocean in being salinity stratified rather than temperature stratified and subject to rapid heat loss to the atmosphere under conditions of subsurface warming and/or surface salinification and/or surface cooling. As such, its dynamical instabilities are a topic of considerable research interest as they relate to ocean circulation and climate in general but are still not well understood or confidently modeled. This historical challenge and recent progress is nicely illustrated in the previously weak representation of Southern ocean maximum sea ice extent and overly active Weddell Sea deep convection in ESM2M shown above (Figure 17; upper right) that has been largely corrected in CM4.0 and ESM4.1 with sea ice extent extending beyond what was previously the convection region in the climatological mean (Figure 17; lower panels). As shown in this section, the implications of improving the surface representation on the climatological mean result in enhanced variability under large Southern Ocean convection events with centennial timescales. This is a complex topic we contextualize further below but remains an area of active research.

To overview the complex interactions we see in ESM4.1 Southern Ocean climate interactions, we present a 7-member composite centennial cycle of the Southern Polar region over the first 700 years for vertical and sea ice metrics in Figure 31A, and heat and salt metrics in Figure 31B as normalized anomalies for the area south of 65°S. The Southern Ocean oscillations shown for ESM4.1 in Figure 2D and 2E have been synchronized at model cycle year 60 to coincide with the minimum in total heat content for each cycle (black line in Figure 31B). A 13-year box car smoother has been applied to focus on the centennial scale signal. We find that the metrics of

vertical convection in Figure 31A associated with deepening Mixed Layer Depth (MLD; black), heat flux across the ocean interface (red), and log average vertical diffusivity at 2000 m ($\log(K_{Z2000})$); light blue vary synchronously with each other on multidecadal timescales. Moving from an initial state of high ice coverage (negative sea ice extent anomaly; green) and negative ice convergence anomaly (i.e. strong ice export north of 65°S), we see that decreases in sea ice coverage (expressed as negative sea ice extent anomaly; green) initially lag metrics of convection as convergence of sea ice (dark blue) from the southern part of the region initially modulates sea ice coverage (green) for the first few years of deep convection onset around year 30. We find similar delays in recovery of sea ice extent and even longer delays in the recovery of sea ice converge and associated transport over the next two decades following convection termination (years 60-80).

The flux-associated metrics shown in Figure 31A have an integrated effect on column average heat content (Figure 31B; black) which drives synchronous variability in sea surface height (SSH) and overall ACC strength with the ACC spinning up from a low of 170 Sv at peak heat content which compares well to the *Donohue et al.* [2016] observed estimate 173 ± 9 Sv, and associated high SSH to a high of 200 Sv near the time of minimum heat content and low SSH with a lag/adjustment timescale of a few years. Increased ACC strength is associated with increased advective heat supply to the region with a lag/adjustment timescale of another few years (Figure 30A; black with crosses) for a total lag between heat content minimum and advective heat supply of approximately a decade. The salinity stratification over the top 700 m (Figure 28B, red) leads column average heat content (Figure 30B, black) by approximately several years while total salt content (Figure 30B; green) lags total heat content (Figure 31B, black) by approximately a decade. Notably, log maximum vertical diffusivity at 300 m (Figure 31B; light blue) associated with shallow convection in the Southernmost Ross and Weddell Seas exhibits its lowest values only upon the termination of convection lasting approximately two decades.

We describe how the overall cycle of the regions south of 65°S with the following 6 stages: 1) Destratification (Decades 1-2): relatively low heat loss to the atmosphere which is not compensated by low meridional heat supply leads to reduction in heat content while high ice coverage and sea ice export reject brine at the surface, increase total salt, and weaken the halocline. 2) Onset of Convection (Decade 3): The weakening thermal inversion overcomes weakening halocline to stimulate strong convection and further loss of heat in a rapidly growing positive feedback as the loss in sea ice promotes further heat loss. Lower associated SSH stimulates ACC increases meridional heat supply to replace heat as halocline weakening continues from convection even as freshwater export via sea ice weakens. 3) Strong Convection (Decade 4-5): Strong vertical heat loss melts ice and exposes open water to more heat loss as sea ice converges and freshwater accumulates. 4) Convection termination (Decade 5): As upper water temperatures cool to favor sea ice formation, deep and then shallow convection

progressively shut off while continued sea ice convergence helps reestablish the halocline from above while continued strong meridional advection brings heat and salt to the interior. 5) Continuance of Meridional heat and salt supply (Decades 6-7): With no convection, sea ice reforms and atmospheric heat loss diminishes, but continued low SSH and high ACC strength continues to supply meridional heat and rebuild both the interior thermal instability and halocline stability. 6) Increased ice export and reduced meridional heat supply (Decades 8-9): As the halocline continues to recover and sea ice extent and export continue to strengthen, MLD shoals and atmospheric heat loss and advective heat gain decline.

The first three stages are thus characterized by the large-scale influence of the stabilizing halocline and un-stabilizing thermocline and total heat content declining. Inversely, the last three stages are characterized by the large-scale influence of the stabilizing halocline and un-stabilizing thermocline and total heat content increasing. Initially, total salt increases as brine rejection and sea ice export stimulate only shallow convection but eventually deep convection. Freshwater supply from ice melt eventually and cooling surface waters eventually terminate deep convection, but recovery of ice export delays termination of convection and cooler polar waters and associated advection of warm waters continue to supply heat and salt to the interior as the halocline until the meridional gradients across 65°S stabilize. In highlighting the important role of sea ice export, these results are consistent with the freshwater budget assessment for the Weddell Sea by *Timmerman et al.* [2001]. While the above discussion describes these mechanisms from a budget perspective, a fully mechanistic assessment of the dynamics awaits future study. The quantitative interaction of buoyancy drivers of convection and combined roles of circulation, sea ice interactions and climate variability are all intriguing areas of future studies. Overall, the impact is a centennial cycle in the release some of the excess heat supplied meridionally from the overly shallow (Figure 21) and consequently warm (Figure 20) AMOC.

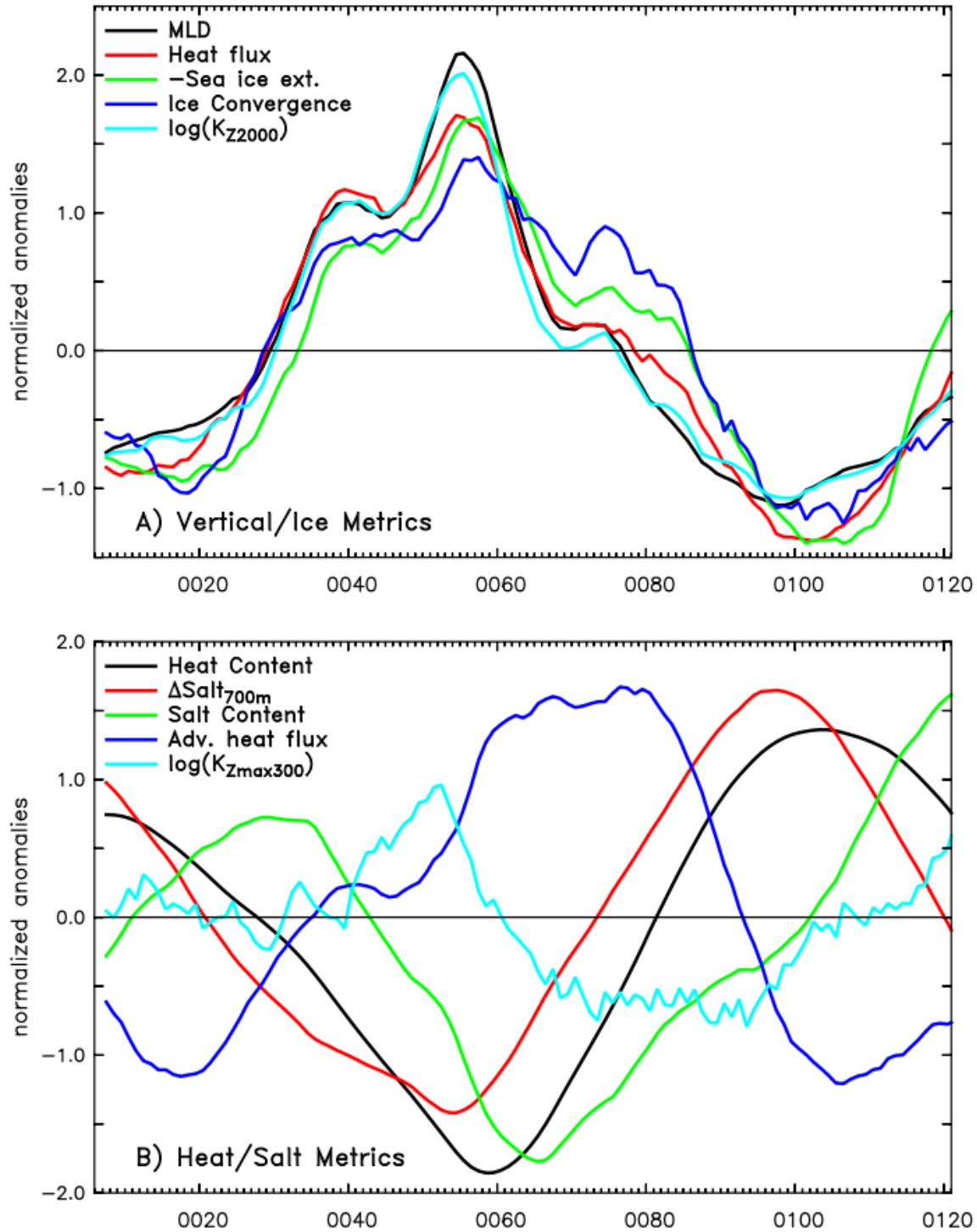


Figure 31: 7-member composite of normalized anomalies for the Southern Ocean Polar (>65°S) region. Vertical/Ice metrics (A) include Mixed Layer Depth by the 0.03 density difference criterion (MLD; black; deep MLD is positive), heat flux across the atmosphere-ocean surface (red; positive up), negative sea ice extent (green; low sea ice is positive), sea ice convergence (dark blue; low sea ice export across 65°S is positive), and log average vertical diffusivity at

2000 m (light blue). Heat/Salt metrics (B) include total heat content south of 65°S (black), the salinity difference over the upper 700 m (red; high being stable low salinity over high salinity) total salt content south of 65°S (green), advective heat flux into the region across 65°S (dark blue; positive South), log maximum vertical diffusivity between 76°S-65°S (light blue). All variables are presented as 13 year smoothed anomalies from the average normalized by the standard deviation and are unitless.

4.15 Response to idealized and historical forcing

Considerable interest and effort has gone into the analysis of climate sensitivity in CMIP6 models and the causes of relatively high sensitivity in some of these models [e.g. *Zelinka et al.* 2020]. While CM4.0 and ESM4.1 have many aspects in common, their sensitivity to CO₂ greenhouse gas forcing are markedly different. Overall ESM4.1 has a 20% lower transient climate sensitivity (1.6°C) relative to CM4.0 (2.1°C) at CO₂ doubling (average of years 61-80). The equilibrium climate sensitivity (ECS) of ESM4.1 (3.1°C) is 36% lower than that of CM4.0 (5.0°C) based on the 300-year abrupt 4xCO₂ integration method of *Winton et al.* [accepted]. There is a diversity in methods and approaches to estimating ECS, however, and each gives slightly different values. ECS estimates using the 150-year integration method of Gregory et al. [2004; Andrews et al. 2012] for ESM4.1 (2.7°C) are similarly 32% lower than for CM4.0 (3.9°C). ECS estimate by slab ocean model method, 3.3°C is slightly lower than the ECS value of 3.4°C estimated from a long (>800 year) continuation of the abrupt 4xCO₂ integration. Thus, while there exists a considerable range in ECS estimates for ESM4.1 of 2.7°C-3.4°C, all are considerably lower than those for CM4.0 (3.9°C-5.0°C)

Preliminary investigation into the causes for this lower climate sensitivity in ESM4.1 compared to CM4.0 have indicated at least 6 drivers which we plan to quantify individually in future work. As was discussed above, ESM4.1 includes a revised representation of sea salt emissions of *Jaeglé et al.* [2011] that increases the temperature sensitivity of sea salt emissions constituting a negative climate feedback of approximately 0.5°C and associated with an AOD change that is about 2.5 times stronger in ESM4.1 than in CM4.0 as described in *Paulot et al.* [2020]. The SOA source from BVOC emissions, which depend strongly on temperature, lead to an additional AOD increase that is about 5 times stronger in ESM4.1 than in CM4.0. Dust emission are similarly responsive to warming, with an AOD increase about 1.7 times in ESM4.1 stronger in ESM4.1 than in CM4.0. Together, these three negative aerosol climate feedbacks act roughly equally on AOD. A fourth negative climate feedback in ESM4.1 not present in CM4.0 is stratospheric ozone which increases by 11% above 20 hPa and 18% above 10hPa as the stratosphere cools and serves to reduce stratospheric cooling by 43% relative to CM4.0. Another difference comes from the explicit representation of atmospheric CO₂ with an annual restoring term in ESM4.1 compared to the externally driven concentration in CM4.0 leading to an initial 3% drop in atmospheric CO₂ due to ocean and land uptake but falls within 0.2% of the target

CO₂ concentration by year 80. A final consideration of the differences in ESM4.1 relative to CM4.0 relates to their differing ocean heat uptake. While the ESM4.1 ocean takes up 5% less heat than CM4.0 at CO₂ doubling in the transient 1% CO₂ per year experiment, ESM4.1 takes up 22% more heat per degree of warming than CM4.0. Similarly, averaged over years 131-150 under abrupt 4 times CO₂ experiment, the ESM4.1 ocean takes up 13% less heat overall but 19% more heat per degree of warming than CM4. However, the higher ratio of transient to equilibrium sensitivity in ESM4.1 ($0.52=1.63/3.16$) than in CM4.0 ($0.41=2.05/5.0$) suggests less long-term ocean cooling influence and so does not explain ESM4's lower equilibrium sensitivity.

Much like was shown for CM4.0 [Guo *et al.*, 2018; Held *et al.*, 2019], forcing of ESM4.1 with historical greenhouse gases, aerosol emissions gives largely consistent trends compared to observations from HadCRUT4 ([Morice *et al.*, 2012]; <https://crudata.uea.ac.uk/cru/data/temperature/>) on the global scale (Figure 32A) and extremely good correspondence in the tropics (Figure 32B) across the 3 member ensemble. Similar to CM4.0, ESM4.1 exhibits relative warmth before 1940, and too much warming after 2000 compared to the 1961-1990 reference rather than the pause or “hiatus” in warming in the early 2000's seen in observations and typical of previous models [e.g. Meehl *et al.*, 2014]. All of these apparent model biases were also typical of CMIP5 models [Flato *et al.*, 2013]. Overall, squared correlation coefficients illustrate that both CM4.0 and ESM4.1 are able to capture 68-82% of the global annual variance in temperature. Extratropical comparisons in both the northern (Figure 32C) and Southern (Figure 32D) hemispheres lead to an apparent spurious 0.1-0.2 warming around the turn of the 20th century. The causes of this early warming signal are still under investigation but initial indications from volcanic forcing sensitivity studies in CM4.0 suggest an important role for volcano aerosol interactions [Winton *et al.*, accepted]. Northern Hemisphere warming in ESM4.1 adheres much closer to observations than CM4.0 after 1990. The situation is reversed in the Southern Hemisphere however, as one ensemble member in ESM4.1 initialized in year 101 (r1; see discussion of Figure 22) exhibits a strong Southern Ocean polynya event around year 1905 while another initialized in year 201 (r3; see discussion of Figure 22) exhibits a strong Southern Ocean polynya event after 2000 that provides a large ensemble spread over this time period. Note that r1, the ensemble member with large polynya around year 1905, has the lowest r^2 in both the Southern (0.68 compared to 0.72 and 0.77) and Northern (0.58 compared to 0.75 and 0.78) demonstrating the important role of centennial scale internal variability in impacting the comparability with the observational record - the inference to the real world, following from Zhang *et al.* [2018] being that even though there are no observations in the Southern Ocean from the early part of the observational activity, that such a large polynya event, if it had occurred, would have been captured in the Northern extratropical record and thus is not likely to have occurred at that time.

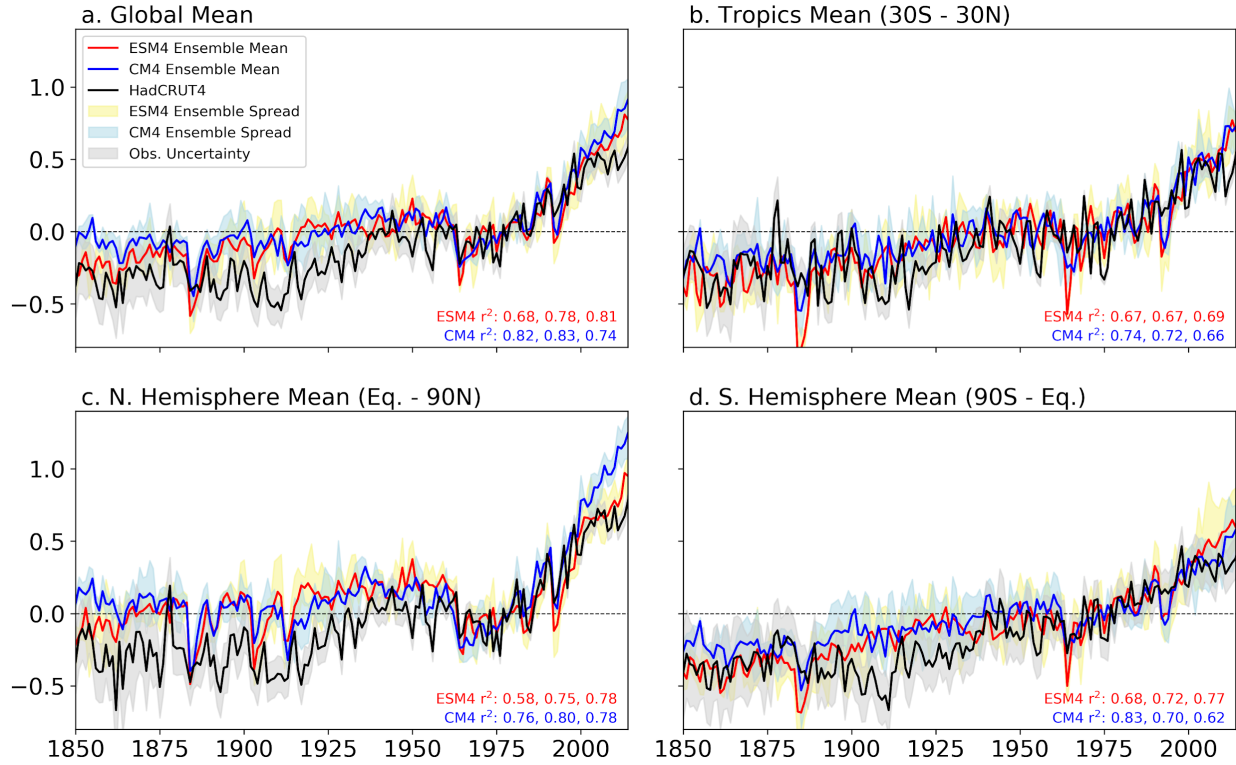


Figure 32: Comparison of near surface (2m) temperature during the 1850-2014 historical period with HadCRUT4 (Morice et al., 2012) for the globe (a), tropics (30°S-30°N; b), Northern Hemisphere (0-90°N; c) and Southern Hemisphere (0-90°S; d) referenced to 1961-1990. Also shown in each panel is the correlation coefficients for ensemble members r1, r2, and r3, with HadCRUT4, respectively.

4.16 Atmospheric CO₂ variability

One of the primary goals of ESM4.1 development was to improve the representation of coupled carbon-climate interactions above the previous generation model. While the ESM4.1 component-level carbon cycles in land (Shevliakova et al., in review) and ocean (Stock et al., in review) are described separately, here we present a description of the overall coupled behavior of the carbon system with an atmospheric focus. Comparison of coupled carbon-climate feedbacks as part of the C4MIP project is also provided in Arora et al. [2019] wherein all feedback parameters for ESM4.1 were found to lie within a standard deviation of the multi-model mean. Here, we first consider the ability of ESM4.1 to stably simulate atmospheric CO₂ under preindustrial 1850 piControl forcing. We find that over the 500 year piControl run, ESM4.1 atmospheric CO₂ receives a net supply of CO₂ from the land of 0.027 PgC yr⁻¹ and a net loss to the ocean of 0.031 PgC yr⁻¹ for a net loss of 0.0042 PgC yr⁻¹ corresponding to an average decline at the surface of 0.0025 ppm yr⁻¹ or approximately 0.41 ppm over the 165 years of the historical simulation. As shown in Figure 33, this drift in ESM4.1 atmospheric CO₂ (upper panel, cyan line) is extremely small compared to the anthropogenic CO₂ signal, much smaller than the

interannual standard deviation (1.1 ppm) and a reassuring demonstration that ESM4.1 is able to represent a stable CO₂ cycle in the absence of anthropogenic forcing.

Comparison of simulations forced by historical CO₂ emissions including land use demonstrate substantial improvement in ESM4.1 comparability to observations up through 1970 relative to ESM2M and ESM2G but point to lingering challenges in both the CMIP6 generation of coupled carbon-climate model and the experimental design. As has been well documented (e.g. *Hoffman et al.*, 2013), many of the CMIP5 class of models had difficulty representing the historical CO₂ trend, with GFDL's ESM2M (Figure 33, red line) and ESM2G (Figure 33, green line) both deviating from CO₂ concentration forcing from observations (Figure 33, black line) in the early 20th century with a developing high bias (Figure 33, lower panel). While this bias grew quickly after 1940 and stabilized at about 20-25 ppm by 1960 in ESM2M (red) and ESM2G (green), it grows more slowly in ESM4.1 up to that level of bias by 2005 and continues to grow over the following decade to approximately 30 ppm by 2014 (note that this decade was not part of the CMIP5 historical protocol used for ESM2M and ESM2G).

The time evolution of total ESM4.1 biosphere carbon uptake and its partitioning into land and ocean for both the concentration and emission simulations is shown in Figure 33c. While the ESM4.1 ocean (dark and light blue) began taking up carbon from the beginning of the simulation and increased in parallel with the atmospheric CO₂ increase, the ESM4.1 land steadily emitted approximately 0.8 PgC yr⁻¹ for the first century of the simulation before stabilizing around year 1960 forced by emissions (red) and year 1980 forced by concentrations (magenta). In both land and ocean, the increased atmospheric CO₂ led to a substantial (15-20 PgC) positive offset in net uptake. The latitudinal structure of land and ocean uptake integrated over the historical period is shown in Figure 33d as an integral from the South to North Poles. Ocean uptake over 165 years in the control simulation (black) illustrates the important role of the ocean circulation and biogeochemistry on natural ocean carbon fluxes with 0.24 PgC yr⁻¹ emitted south of 50°S, uptake between 20°S-50°S of 0.64 PgC yr⁻¹, tropical emission of 1.1 PgC yr⁻¹, uptake between 20°N-50°N of 0.49 PgC yr⁻¹, and uptake north of 50°N of 0.20 PgC yr⁻¹. Land uptake in the control (red) is stable demonstrating satisfactory equilibrium has been achieved under the limitation that the explicit role of rivers in transporting carbon from land to ocean is ignored (though the ocean receives carbon from rivers and buries it in sediments internally).

Regional variations in carbon cycling are illustrated as a function of latitude in Figure 33d. Ocean uptake in the historical concentration (dark blue) and emission (light blue) simulations referenced to the control illustrate the important role of the Southern Ocean south of 40°S (21% of ocean area) taking up 37% of anthropogenic CO₂ and south of 30°S (30% of ocean area) taking up 44%, similar to the 43±3% found with CMIP5 models (*Frölicher et al.* [2015]). The role of the Northern Oceans is also highlighted with 35N-60°N (10% of ocean area) taking up 16%. In contrast, ESM4.1 land emissions are primarily driven by the land use forcing rather

than the increase in CO₂ with integrated net land biosphere emission in the tropics between 22°S and 8°N (21% of land area) emit 29% of the total and the narrow boreal band of 37°N-48°N emitting 30% of the total. There is also a net uptake in polar areas north of 60°N more easily attributable to climate and carbon changes rather than land use that gain 12.6 PgC in the emission (red) and 8.3 PgC in the concentration (magenta) simulations (See *Arora et al.* 2019 for discussion of the separate responses of ESM carbon cycles to climate and CO₂ change as part of C4MIP).

Potential causes of these biases including uncertainties in land use forcing and model biases were discussed in *Hoffman et al.* [2013] for the CMIP5 suite. While explanation for the land remain under investigation, subsequent analysis of the CMIP5 suite has demonstrated that biases in ocean uptake can be attributed to the experimental design rather than a bias in ocean chemistry or dynamics. As discussed in *Stock et al.* [in review], about 20-36 PgC (9-17 ppm) of this difference is explained by low biases in ocean carbon uptake compared to observational estimates. *Bronse laer et al.* [2017] demonstrate that this ocean bias can be attributed to the limitations of the CMIP5 and CMIP6 experimental design - whereas the models are forced to air-sea equilibrium at 1850 (pCO₂ = 284 ppm), the real world experiences a ΔpCO₂ of approximately 7-10 ppm based on atmospheric CO₂ increase over the previous century (pCO₂ ~277 at 1765) that result in additional cumulative ocean uptake of approximately 17-30 PgC including both pre-1850 uptake and continued uptake associated with that increase partial pressure difference that is not representable in the CMIP6 experimental design.

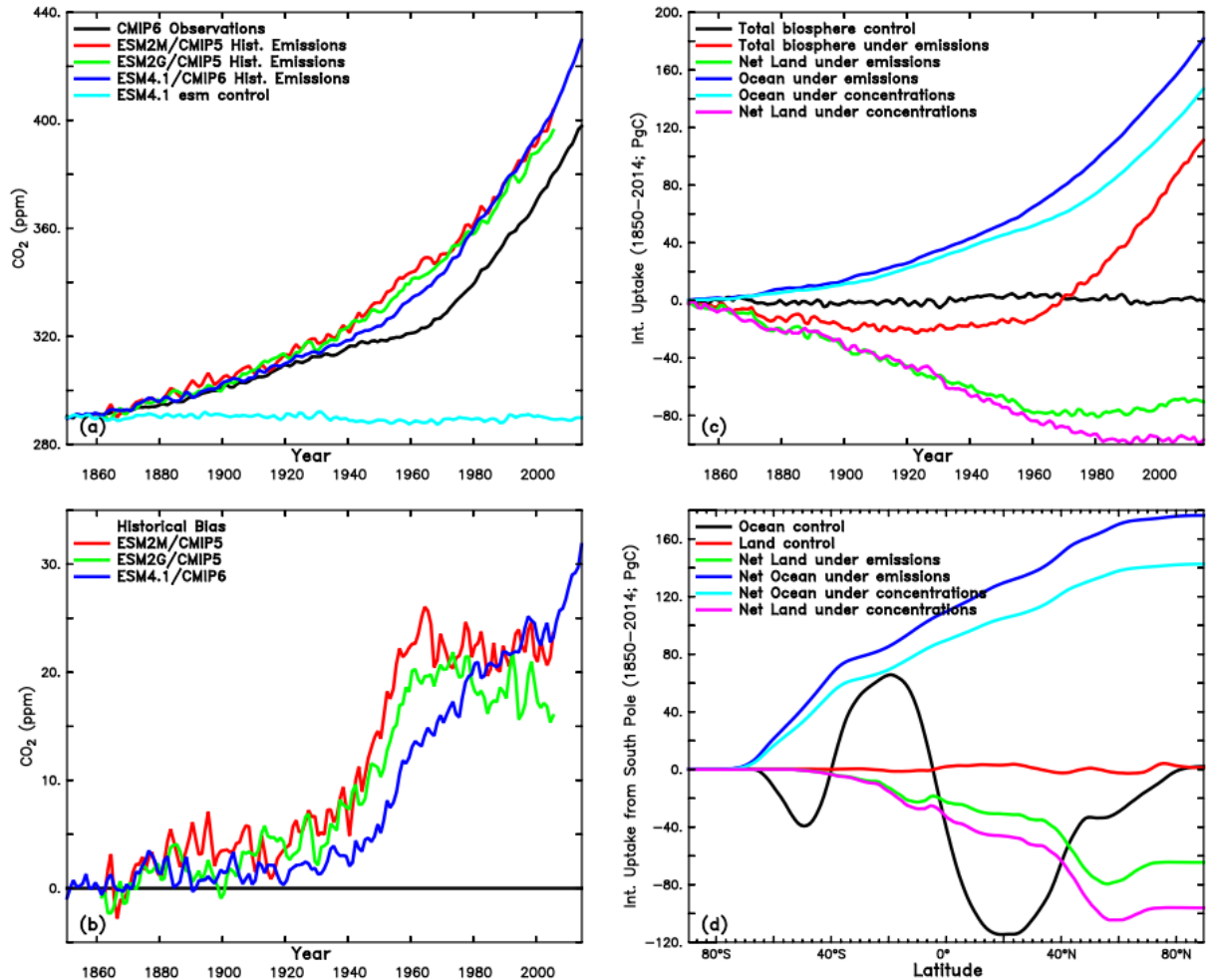


Figure 33: Comparison of global average historical surface concentrations of CO₂ (a; ppm) from the ESM4.1 standard CO₂-concentration-forced simulation (black) along with CMIP5 CO₂-emission-forced simulations from ESM2M (red) and ESM2G (green), the CMIP5 CO₂-emission-forced simulations from ESM4.1 (blue), and the ESM4.1 “esm” emission control simulation (no emissions; cyan) and biases (b; ppm) in CMIP5 CO₂-emission-forced simulations from ESM2M (red) and ESM2G (green), the CMIP5 CO₂-emission-forced simulations from ESM4.1 (blue) relative to the standard CO₂-concentration-forced simulation. Also shown are the integrated uptake (d; PgC) over time for combined land and ocean biosphere in the emissions control (black), and historical emissions (red) simulations, as well as a comparison of the net land and ocean uptake referenced to the control under emissions (green and dark blue, respectively) and concentrations (magenta and light blue, respectively). The latitudinal structure of total uptake (d; PgC) over the historical simulations is shown integrated from the South to North Poles in the emissions control for the ocean (black) and land and the net land and ocean uptake referenced to the control under emissions (green and dark blue, respectively) and concentrations (magenta and light blue, respectively).

The seasonal amplitude of atmospheric CO₂ variability in historical simulations forced by CO₂ emissions with atmospheric CO₂ allowed to freely evolve are compared with results from the previous generation ESM2G, ESM4.1 and observations from NOAA Global Monitoring Division (*Dr. Pieter Tans, NOAA/ESRL (www.esrl.noaa.gov/gmd/ccgg/trends/) and Dr. Ralph Keeling, Scripps Institution of Oceanography (scrippsco2.ucsd.edu/)* in Figure 34 (ESM2G chosen over ESM2M as reference for its superior representation of land-atmosphere CO₂ fluxes; [Dunne *et al.*, 2013]). ESM4.1 exhibits marked improvement in comparison with marine sites in both seasonal CO₂ amplitude (Figure 34; left panels) with RMSE decreasing by roughly half - from 3.3 ppm to 1.6 ppm. This improvement in fidelity comes through both reduction of local air-sea CO₂ flux amplitude in Southern Hemisphere oceans directly as well as reduced expression of the high tropical land-atmosphere fluxes expressed over the ocean. In the northern latitudes, ESM4.1 exhibits higher seasonal CO₂ amplitudes than ESM2G more consistent with observations. Similarly, dramatic improvement is shown in the representation of interannual variability (IAV) from ESM2G to ESM4 with the RMSE reduced from 0.6 to 0.2.

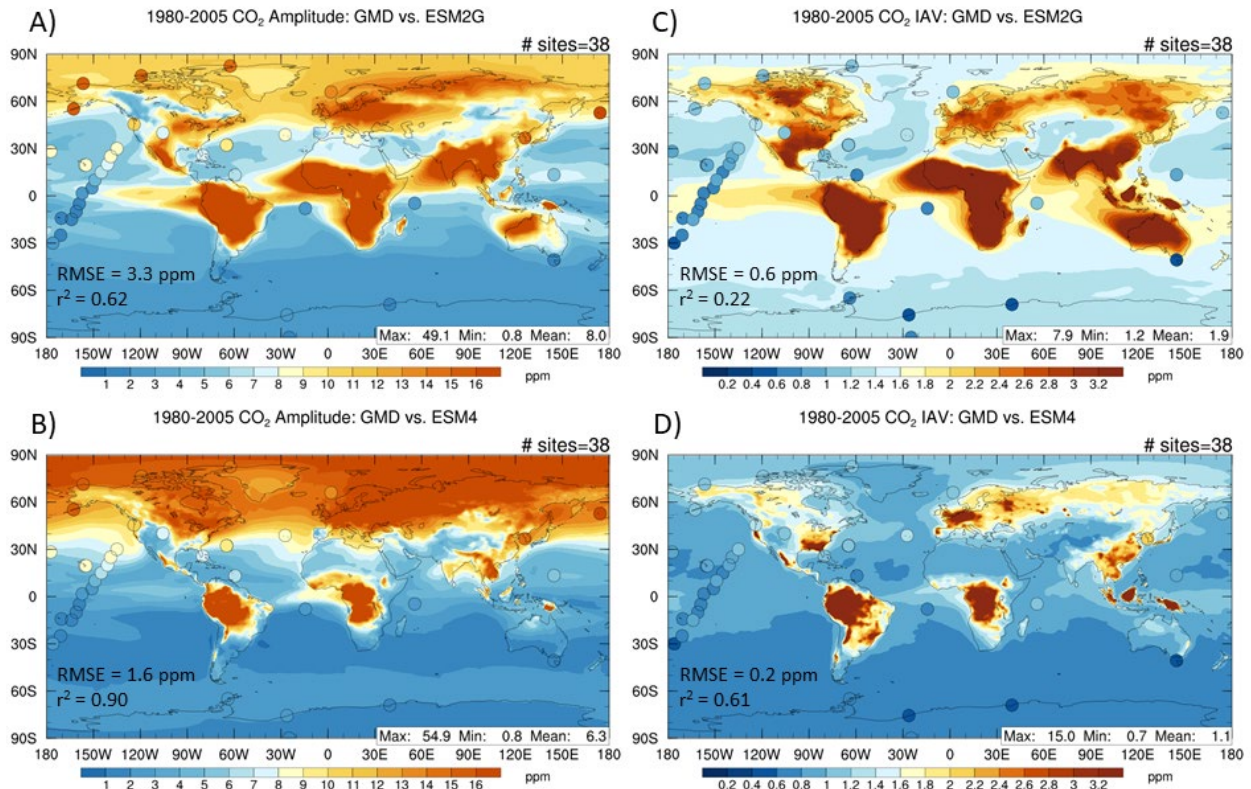


Figure 34: Spatial distribution of atmospheric seasonal CO₂ (ppm) amplitude (left) and CO₂ interannual variability (IAV, right) overlaid with NOAA-GMD observations for ESM2G (top; A and B, respectively) and ESM4 (bottom; C and D, respectively). GMD sites with at least 15-year

observational record are selected for evaluation. Root mean square error (ϵ ; ppm) is shown on each plot.

Comparison of the seasonal cycle over a selection of long time series at pristine sites in the NOAA GMD dataset show marked improvement in ESM4.1 over ESM2G across latitudes from North to South as shown in Figure 35. In all cases shown in Figure 35 except Ascension, the correlation coefficient of the seasonal cycle has increased and magnitude of interannual variability (IAV) has decreased, both improving the correspondence of ESM4.1 to the observations relative to ESM2G. At Barrow, the seasonal amplitude has increased in ESM4.1 with deeper and longer duration summer drawdown relative to ESM2G illustrating the substantial improvements to the representation of the seasonal cycle of net community production. Similarly, the seasonal cycle at Mauna Loa also exhibits much improved amplitude and phasing consistent with observations, similarly highlighting land carbon cycle improvements from ESM2G to ESM4.1. South of the equator from Ascension Island through Tutuila (Samoa) ESM4.1 avoids the strong drawdown in CO₂ associated with synchrony of peak Amazon and African rainforest net community production in the boreal spring in ESM2G [Dunne *et al.*, 2013]. While this lease the amplitude much improved, the phasing of the seasonal cycle still leaves considerable improvement. Moving to the South Pole, ESM4.1 avoids the strong seasonal cycle in air-sea gas flux that was seen in ESM2G associated with the complex interplay of ocean physical and biological controls that has been the subject of recent analysis [e.g. Mongwe *et al.*, 2016]. One such improvement is the dramatic reduction in Southern Ocean warm bias (Figure 14) relative to ESM2M and ESM2G. Improvements relating to ESM4.1 COBALT phytoplankton bloom phenology are probably also at play [Stock *et al.*, in review] and worthy of further analysis in the context of CMIP6 intercomparison.

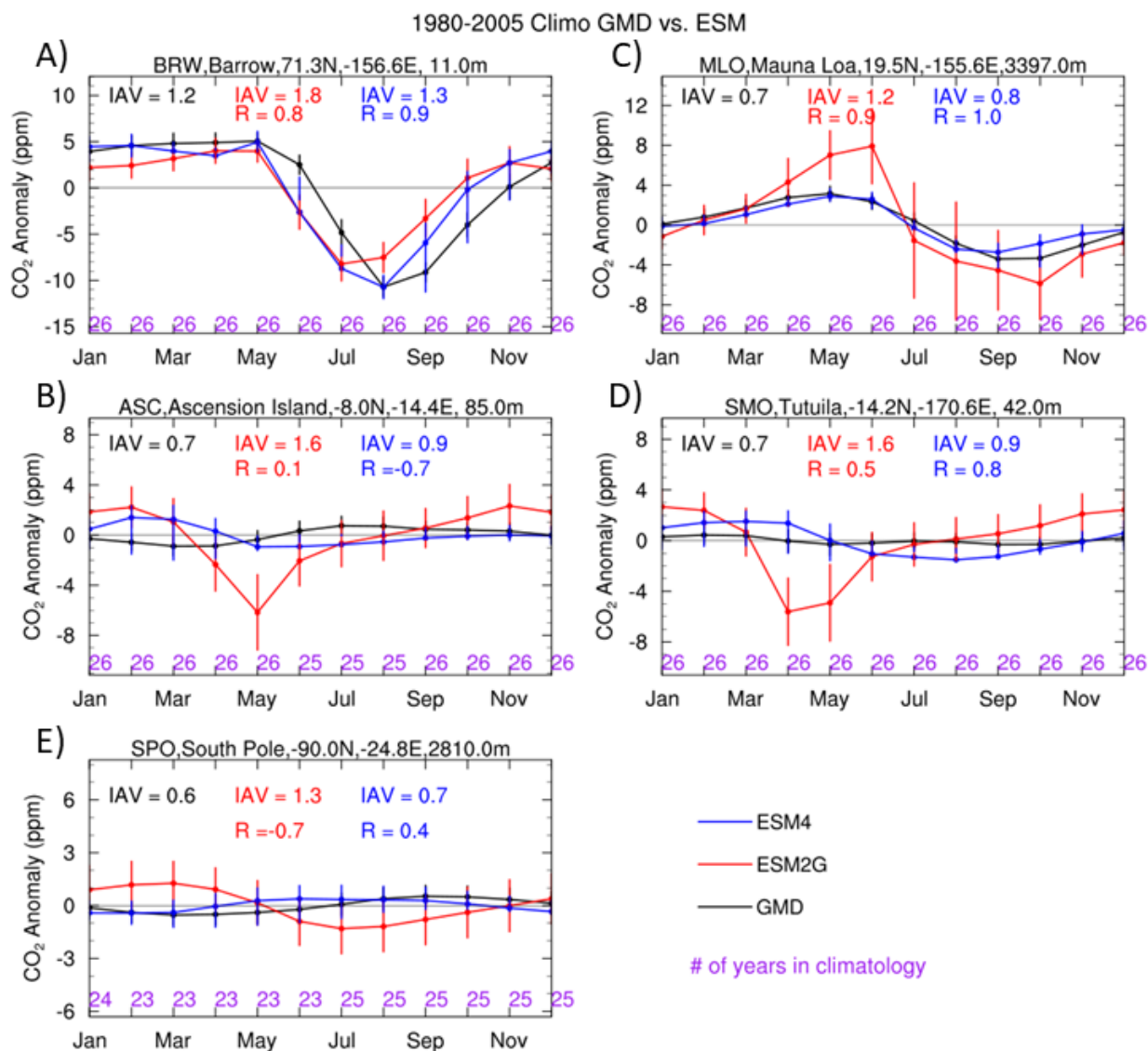


Figure 35: Climatological seasonal cycle of atmospheric CO₂ in NOAA/GMD observations (black), ESM2G (red) and ESM4.1 (blue). CO₂ interannual variability (IAV; ppm) and correlation coefficient (R) at each site is also shown.

4.17 Tropospheric and total column Ozone

As a final consideration of model fidelity, we consider the key role of interactive chemistry in ESM4.1 through evaluation of tropospheric and total column ozone against observations. We find substantial improvements in the representation of tropospheric ozone across GFDL model generations as we compare global tropospheric ozone column (60°N to 60°S) with observational

estimates from the Aura Ozone Monitoring Instrument/Microwave Limb Sounder (OMI/MLS) satellite instruments of the Tropospheric Ozone Column [Ziemke *et al.*, 2019] for both CM3 and ESM4.1 (Figure 36). While CM3 exhibited high biases over nearly all regions except the subpolar Antarctic, ESM4.1 exhibits a general interhemispheric pattern of high values in the northern midlatitudes and over continents and low values over the Southern Ocean. Over Oceania, the greater ozone in ESM4.1 versus CM3 is related to the differences in biomass burning emissions. While CM3 (Figure 36, upper panels) exhibited an average high bias = 19.6%, the high bias in ESM4 has been reduced to 8.5% possibly because of cancellation of the positive and negative biases in the two hemispheres. Similarly, the RMSE has reduced considerably from 6.9 Dobson Units (DU) in CM3 to 4.8 DU in ESM4.1.

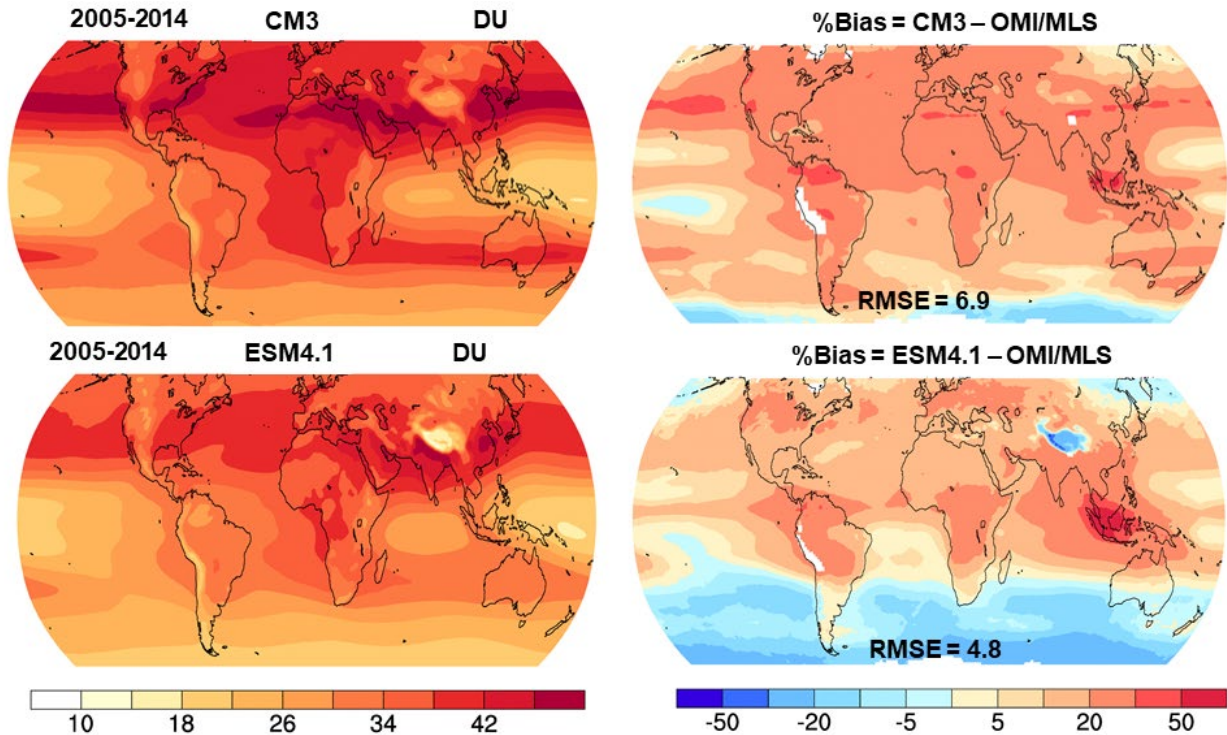


Figure 36: Tropospheric ozone column in CM3 (upper left; Dobson Units, DU), ESM4.1 (lower left; DU), and the % bias compared to the OMI/MLS satellite estimate of the Tropospheric Ozone Column (Ziemke *et al.*, 2019) for CM3 (upper right; %) and ESM4.1 (lower right; %). RMSE is provided in DU.

Comparison of the CM3 and ESM4.1 to total column ozone similarly observations demonstrates substantial improvements across model generations. Figure 37 compares CM3 and ESM4.1 simulated time series of total column ozone from 1980 to 2015 against two datasets: Multi-Satellite Merged Total Column merged NASA and NOAA product from Frith *et al.* [2013] (SBUV; open triangles) and version 3.4 of the National Institute of Water and Atmospheric Research - Bodeker Scientific (NIWA; closed circles) total column ozone (TCO) database from Bodeker *et al.* [2005]. Comparisons are shown for the annual average globally, in the tropics and southern and northern midlatitudes, for March in the Arctic, and October in the Antarctic.

Globally (Figure 37a). Overall, ESM4.1 compares better against observations of TCO than CM3. Absolute values of TCO for CM3 and ESM4.1 deviate from the observations by about 10 DU in opposite directions with CM3 biased high and ESM4.1 biased low. Both the models generally capture the trend in global TCO. ESM4.1 is able to better capture the observed evolution of global mean TCO as indicated by the greater correlation coefficients for ESM4.1 (0.84, 0.90) compared with CM3 (0.79, 0.73).

Regional means and temporal patterns in tropical TCO are improved in ESM4.1 compared to CM3 while comparisons in the extra-tropics are mixed. In the tropics (Figure 37b), absolute TCO in ESM4.1 is within about 5 DU of observations as opposed to ~10 DU in CM3. *Eyring et al.* [2013] attributed CM3's high bias to an overestimation of the ozone maximum at around 10 hPa which has been substantially improved in ESM4.1 (not shown). Consistent with observations, both the models simulate negligible TCO trends in the tropics, however ESM4.1 exhibits greater skill. In the northern mid-latitudes (Figure 37c), ESM4.1 is farther from observations than CM3 although with similar skill in simulating the observed TCO evolution. The comparison is opposite for the southern mid-latitudes (Figure 37d), i.e., ESM4.1 is much closer to observed absolute values than CM3 though with similar correlations. *Eyring et al.* [2013] attributed CM3's excess TCO at southern midlatitudes to the excessive strength of the Brewer-Dobson circulation from the tropics (where TCO is already too high) to midlatitudes. ESM4.1 notably improves this bias in CM3. In the Arctic in March (Figure 37e), ESM4.1 reproduces the observed absolute TCO slightly better than CM3, however both have low skill in reproducing the observed evolution. In the Antarctic in October (Figure 37f), ESM4.1 and CM3 are similar in their comparison against observations both in terms of trends and absolute values.

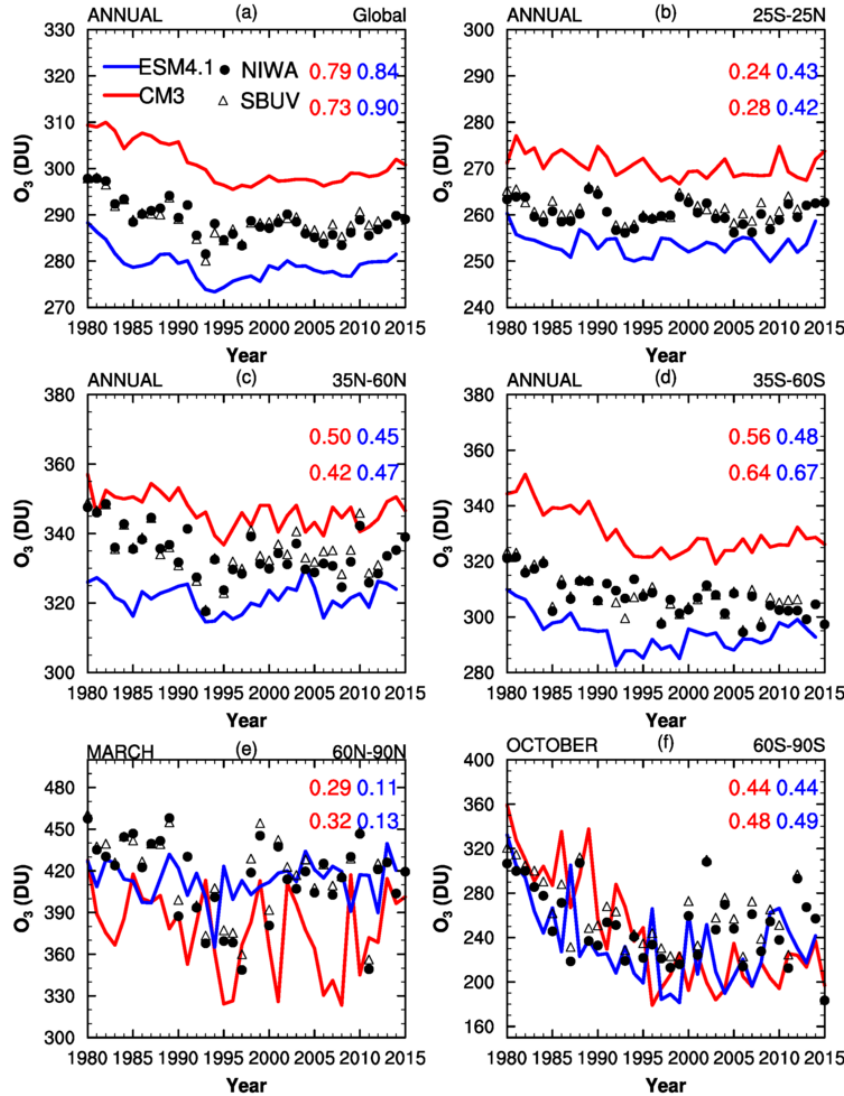


Figure 37: Comparison of time series of total ozone column (TCO; DU) for the annual mean (a) global mean (90°S-90°N), (b) tropics (25°S-25°N), (c) northern mid-latitudes (35°N-60°N), (d) southern mid-latitudes (35°S-35°N), and for the March and October mean in the (e) Arctic (60°N-90°N), and (f) Antarctic (60°S-90°S), respectively, from CM3 (red) and ESM4.1 (blue) against observations from the multi-satellite merged NASA and NOAA TCO product (Frith et al., 2013; SBUV; open triangles) and version 3.4 of the NIWA-BS TCO database (Bodeker et al., 2005; NIWA; closed circles). The numbers in each panel indicate linear correlation coefficient (R) for model against each of the measurement datasets (top for NIWA and bottom for SBUV).

5 Discussion

In the previous section we described a broad suite of coupled model simulation characteristics to complement the component-level descriptions for ESM4.1 atmospheric chemistry [Horowitz et al., in review], land [Shevliakova et al., in review], and ocean biogeochemistry [Stock et al., in

review]. In contrast with GFDL's CM4.0 effort which focused on ocean resolution for physical climate interactions, the ESM4.1 effort focused on revision of CM4.0 towards comprehensiveness of Earth system interactions, including fully interactive Earth system representation of the carbon, dust and iron cycles, and substantially improved interactions of reduced nitrogen and biogenic volatile organic carbon. Relative to CM4.0, we also find that the reduction of ocean model resolution from $\frac{1}{4}$ to $\frac{1}{2}$ and freeing up of the many chemistry, aerosol and ecosystem interactions with climate leads to slight degradation in many of the baseline simulation characteristics, though the ability to tune the role of mesoscale eddies and a more sophisticated atmospheric aerosol and land model in ESM4.1 also leads to several improvements relative to CM4.0. These include low upper 700 m and global thermal drift and carbon drift (Figure 2), zonal wind stress in the Pacific (Figure 9), penetration of CFC's into the deep (700-2000 m) Southern Ocean (Figure 22), weaker 2-m atmospheric cold bias (Figure 11) and resulting representation of sea ice (Figure 17, Figure 18), improved SO_4 and sea salt distributions (Figure 6) and improved tropospheric ozone (Figure 36). While reduced ocean resolution, more interactive chemistry, and revision of aerosol parameterizations leads to some differences in simulation characteristics with CM4.0, many of which are not fully understood, the most important of these differences is likely those related to climate sensitivity, aerosol forcing, and the quality of the 20th century temperature evolution. Considering ESM4.1 in comparison to previous coupled chemistry (CM3; [Donner *et al.*, 2011]) and even more so coupled carbon-climate [Dunne *et al.*, 2012; 2013] illustrates the marked improvements that have resulted from AM4.0 [Zhao *et al.*, 2018a, 2018b] and AM4.1 [Horowitz *et al.*, in review] in radiation (Figures 3-5) and precipitation (Figure 12-13), OM4 [Adcroft *et al.*, 2019] in surface temperature (Figure 14), salinity (Figure 15) and mixed layer depths and sea ice (Figure 16-17), LM4.1 [Shevliakova *et al.*, in review] in representation of Köppen climate types (Figure 24) and seasonal and interannual CO_2 variability (Figure 31-32), COBALT [Stock *et al.*, in review] in low drift in carbon (Figure 2) and good representation of natural and anthropogenic fluxes (Figure 33) including improved seasonality in the Southern Hemisphere (Figure 31, Figure 33) through fully coupled CM4.0 [Held *et al.*, 2019] and ESM4.1 development. Southern Ocean winds (Figure 9) and sea ice (Figure 16 and 17) are markedly improved in ESM4.1 relative to CM4.0 and previous generation models. Misrepresentation of Köppen climate types is reduced by approximately half from the CM2.1 climate used in ESM2M to ESM4.1 (Figure 24). Representation of atmospheric chemistry and aerosol cycling are strongly improved even as they are made more mechanistically consistent (Figure 6). CO_2 interactions are similarly improved in ESM4.1 relative to previous generation models in the seasonal and interannual scales (Figure 31-32). Representation of many modes of climate variability are markedly improved in this generation of models as described for CM4.0 in Held *et al.* [2019] including ENSO teleconnections to North America and the Indian Ocean (Figure 29) and the pattern and strength of AMO (Figure 30). Climate sensitivity in ESM4.1 is lower than CM4.0 and compares well with historical observations of 20th century warming (Figure 32).

As with all development efforts, the end product involves many trade-offs and imperfect solutions that would undoubtedly be addressed/improved in future model versions. In atmosphere development, the implications of differences in gravity wave drag between CM4.0's 33 layers and ESM4.1's 49 layers were only briefly considered and should be re-evaluated in future development efforts to improve the representation of upper atmosphere equatorial winds (Figure 8). Similarly, the many changes associated with transition to interactive chemistry and aerosol cycling incorporated through both coupled interactions with chemistry and separately to improve the mechanistic consistency and other parameterization updates were only briefly considered. In ocean development, parameterization of submesoscale baroclinic mixed layer eddies is likely too strong leading to shallow mixed layers relative to CM4.0. Mixing by internal tides was also parameterized too strongly relative to the intended values used in CM4.0. On land, the last minute increase in near infrared albedo on snow over glacier to the upper limit of the observational constraint was an unfortunately necessary medicine to stimulate quasi-permanent surface heat extraction from the polar Southern Ocean and prevent unacceptable subsurface accumulation in Southern Ocean heating in earlier versions. A more comprehensive assessment of the role of land and sea ice albedo, sea ice dynamics, cloud interaction and ocean parameterizations is left to future development efforts.

Other developmental challenges having incrementally improved in ESM4.1 remain without a clear path forward. The long standing double ITCZ bias, low stratus and associated warm SST biases in Eastern Boundary Current regions, cold subtropical gyres and warm Southern Ocean remain lingering issues with important consequences for accurate representation of Köppen climates, ocean biogeography, and the carbon cycle and inviting future development. In the atmosphere, representation of interactive aerosol dust (Figure 6) and ozone (Figure 36, Figure 37) continues to be a challenge. On land, the reason for strong land-use induced CO₂ emissions over the historical period contributing to elevated CO₂ when the model is driven by emissions (Figure 33) remain a topic of ongoing research. In terms of the tropical coupled atmosphere-ocean representation, challenges include the shallow tropical mixed layers, excessive tropical rainfall and associated cyclonic wind stress curl, and better simulation of tropical instability waves (TIWs) and their effects on upper-ocean vertical mixing near the equator. Based on our experience with much higher resolution models such as CM2.6 [Griffies *et al.*, 2015], we expect that some of these challenges will be met by energizing the explicit ocean mesoscale. Other aspects of high resolution atmosphere-ocean coupling should also be explored. On land, promising avenues of exploration include incorporation of the nitrogen cycle (Sulman *et al.*, 2019) and hillslope tiling [Subin *et al.*, 2014] versions of LM3 and sub-grid parameterization of topography for precipitation partitioning [Leung and Ghan, 1998]. In atmospheric chemistry, more detailed comparisons with observations and incorporation of revised chemical mechanisms are being explored.

Some of these challenges such as centennial scale Southern Ocean variability driving surface climate variability in this class of GFDL models is a new characteristic of variability. Further work will be necessary to understand and better represent the associated mechanisms of Southern Ocean heat capacitance and its potential to provide sudden climate warming both in models and the real climate system.

6 Conclusions

We have described the baseline model configuration and simulation characteristics of GFDL's Earth System Model version 4.1 (ESM4.1) which builds on component and coupled model developments at GFDL over the 2013-2018 period for contribution of coupled carbon, chemistry, and climate simulations as part of the fifth phase of the Coupled Model Intercomparison Project. Based on GFDL's fourth generation framework, we find dramatic improvements in both carbon cycle and chemistry-climate interactions in ESM4.1 relative to the previous generation GFDL ESM2M/ESM2G and CM3 models. We find ESM4.1 to capture much of the baseline simulations characteristics of CM4.0. These improvements in physical climate features across generations of GFDL coupled carbon and chemistry-climate models include atmospheric radiation, sea surface temperature and salinity, sea ice extent, minimum and maximum mixed layer depth, interior ocean properties and atmospheric carbon cycle dynamics. Notably, ESM4.1 is able to represent key regional features of historical climate that were problematic in the previous generation coupled chemistry-climate CM3 [Donner *et al.*, 2010] and the focus of considerable subsequent work to understand and improve [Golaz *et al.*, 2013; Zhao *et al.*, 2016].

The ability of ESM4.1 to compete in terms of fidelity with a state of the art physical model like CM4.0 [Held *et al.*, 2019] while also representing dynamic interactions in carbon and chemistry is a major achievement in synchronization of model development over previous efforts that built either chemistry or carbon cycle representations on the previous generation physical model framework such as ESM2M/ESM2G [Dunne *et al.*, 2012] having rested entirely on the physics of CM2.1 [Delworth *et al.*, 2006] and CM3 [Donner *et al.*, 2011] including ostensibly similar ocean and land components. Major differences include ESM4.1 improvements relative to CM4.0 in: 1) Southern Ocean mode and intermediate water ventilation, 2) aerosol representation in the Southern Ocean, 3) reduction in ocean heat uptake, and 4) reduced transient and equilibrium climate sensitivity relative to CM3 and CM4.0. Fidelity concerns in ESM4.1 relative to CM4.0 include: 1) moderate degradation in sea surface temperature biases including partial return of previous generation cold biases in the North Atlantic and warm biases in the Southern Ocean, 2) moderate degradation in aerosol representation in some regions, 3) moderate degradation in upper atmosphere equatorial zonal winds, and 4) the presence of strong centennial scale climate modulation by Southern Ocean convection. Overall, ESM4.1 represents an important advance in coupled carbon-chemistry-climate modeling and should prove applicable to a broad range of studies in both physical climate variability and forced response as well as

impact studies across climate, land and ocean ecosystems, air quality and carbon feedback applications.

Acknowledgements

We thank GFDL's Director, V. Ramaswamy and Deputy Director, Whit Anderson, for supporting the vision behind this work for consolidating chemistry-carbon-climate interactions in GFDL's Fourth generation effort, and the tireless support of GFDL's Modeling Systems and Operations for keeping ESM4.1 development and implementation going through many technical challenges. Cathy Raphael provided graphics support including the model schematic. The work benefited from internal GFDL reviews by Leo Donner and Liping Zhang. Data is provided at <https://doi.org/10.22033/ESGF/CMIP6.1407>. Model code is provided at <http://doi.org/10.5281/zenodo.3836405>. The input data is provided at ftp://data1.gfdl.noaa.gov/users/ESM4/ESM4Documentation/GFDL-ESM4/inputData/ESM4_rundir.tar.gz

References

- Adcroft, A., & Hallberg, R. (2006). On methods for solving the oceanic equations of motion in generalized vertical coordinates. *Ocean Modelling*, **11**(1-2), 224-233.
- Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C. O., ... & Held, I. The GFDL Global Ocean and Sea Ice Model OM4.0 (2019). Model Description and Simulation Features. *Journal of Advances in Modeling Earth Systems*, **11**(10), DOI:10.1029/2019MS001726.
- Adler, R.F., G.J. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin (2003) The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979 - Present). *J. Hydrometeor.*, **4**(6), 1147-1167.
- Adler, R. F., M. R. P. Sapiiano, G. J. Huffman, J.-J. Wang, G. Gu, D. Bolvin, L. Chiu, U. Schneider, A. Becker, E. Nelkin, P. Xie, R. Ferraro, and D.-B. Shin (2018) The Global Precipitation Climatology Project (GPCP) Monthly Analysis (New Version 2.3) and a Review of 2017 Global Precipitation. *Atmosphere*, **9**, 138. doi:10.3390/atmos9040138.
- Arora, V. K., Boer, G. J., Friedlingstein, P., Eby, M., Jones, C. D., Christian, J. R., ... & Hajima, T. (2013). Carbon-concentration and carbon-climate feedbacks in CMIP5 Earth system models. *Journal of Climate*, **26**(15), 5289-5314.
- Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., ... & Chamberlain, M. A. (2019). Carbon-concentration and carbon-climate feedbacks in CMIP6 models, and their comparison to CMIP5 models. *Biogeosciences Discussions*, 1-124.

- Atwood, A. R., D. S. Battisti, A. T. Wittenberg, W. G. H. Roberts, and D. J. Vimont (2017). Characterizing unforced multi-decadal variability of ENSO: A case study with the GFDL CM2.1 coupled GCM. *Climate Dyn.*, **49** (7-8), 2845-2862. doi: 10.1007/s00382-016-3477-9
- Austin, J., Horowitz, L. W., Schwarzkopf, M. D., Wilson, R. J., & Levy, H. (2013). Stratospheric ozone and temperature simulated from the preindustrial era to the present day. *Journal of Climate*, **26**(11), 3528-3543.
- Balmaseda, M. A., K. Mogensen, and A. T. Weaver (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quart. J. Roy. Meteor. Soc.*, **139**, 1132-1161. doi:10.1002/qj.2063
- Bates, K. H., & Jacob, D. J. (2019). A new model mechanism for atmospheric oxidation of isoprene: global effects on oxidants, nitrogen oxides, organic products, and secondary organic aerosol. *Atmospheric Chemistry & Physics*, 19(14).
- Beadling, R. L., Russell, J. L., Stouffer, R. J., Goodman, P. J., & Mazloff, M. (2019). Assessing the quality of Southern Ocean circulation in CMIP5 AOGCM and Earth System Model simulations. *Journal of Climate*, <https://doi.org/10.1175/JCLI-D-19-0263.1>.
- Blanchard, J. L., Jennings, S., Holmes, R., Harle, J., Merino, G., Allen, J. I., ... & Barange, M. (2012). Potential consequences of climate change for primary production and fish production in large marine ecosystems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**(1605), 2979-2989.
- Bodeker, G. E., H. Shiona, and H. Eskes (2005). Indicators of Antarctic ozone depletion, *Atmos. Chem. Phys.*, **5**, 2603–2615, doi:10.5194/acp-5-2603-2005.
- Boucher, O., Halloran, P. R., Burke, E. J., Doutriaux-Boucher, M., Jones, C. D., Lowe, J., ... & Wu, P. (2012). Reversibility in an Earth System model in response to CO2 concentration changes. *Environmental Research Letters*, **7**(2), 024013.
- Briegleb, B.P., and B. Light (2007). A delta-Eddington multiple scattering parameterization for solar radiation in the sea ice component of the Community Climate System Model, NCAR Technical Note TN-472+STR.
- Bronselaer, B., Winton, M., Russell, J., Sabine, C. L., & Khatiwala, S. (2017). Agreement of CMIP5 simulated and observed ocean anthropogenic CO2 uptake. *Geophysical Research Letters*, **44**(24), 12-298.
- Brovkin, V., Boysen, L., Arora, V. K., Boisier, J. P., Cadule, P., Chini, L., ... & Hurtt, G. C. (2013). Effect of anthropogenic land-use and land-cover changes on climate and land carbon storage in CMIP5 projections for the twenty-first century. *Journal of Climate*, **i**(18), 6859-6881.
- Capotondi, A., A. Wittenberg, and S. Masina, 2006: Spatial and temporal structure of tropical Pacific interannual variability in 20th century coupled simulations. *Ocean Modelling*, **15**, 274-298. doi: 10.1016/j.ocemod.2006.02.004
- Cavalieri, D. J., Parkinson, C. L., Gloersen, P., and Zwally, H. J. (1996). Sea ice concentrations from Nimbus-7SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 1. NASA DAAC at the Natl. Snow and Ice Data Cent. <https://doi.org/10.5067/8GQ8LZQVL0VL>

- Chen, C., M. A. Cane, A. T. Wittenberg, and D. Chen, (2017). ENSO in the CMIP5 simulations: Life cycles, diversity, and responses to climate change. *Journal of Climate*, **30** (2), 775-801. doi: 10.1175/JCLI-D-15-0901.1
- Collins, W. J., Lamarque, J. F., Schulz, M., Boucher, O., Eyring, V., Hegglin, M. I., ... & Smith, S. J. (2017). AerChemMIP: quantifying the effects of chemistry and aerosols in CMIP6. *Geoscientific Model Development*, **10**(2), 585-607.
- Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., ... & Brönnimann, S. (2011). The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, **137**(654), 1-28.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... & Bechtold, P. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137**(656), 553-597.
- Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A., ... & Durachta, J. W. (2006). GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *Journal of Climate*, **19**(5), 643-674.
- Delworth, T. L., A. Rosati, W. Anderson, A. J. Adcroft, V. Balaji, R. Benson, K. Dixon, S. M. Griffies, H.-C. Lee, R. C. Pacanowski, G. A. Vecchi, A. T. Wittenberg, F. Zeng, and R. Zhang (2012). Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *Journal of Climate*, **25**, 2755-2781. doi: 10.1175/JCLI-D-11-00316.1
- Delworth, T. L., Zeng, F., Zhang, L., Zhang, R., Vecchi, G. A., & Yang, X. (2017). The central role of ocean dynamics in connecting the North Atlantic Oscillation to the extratropical component of the Atlantic Multidecadal Oscillation. *Journal of Climate*, **30**(10), 3789-3805.
- Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J. H., Dunne, K. A., ... & Harrison, M. J. (2020). SPEAR: The Next Generation GFDL Modeling System for Seasonal to Multidecadal Prediction and Projection. *Journal of Advances in Modeling Earth Systems*, **12**(3), e2019MS001895.
- Dentener, F., Kinne, S., Bond, T., Boucher, O., Cofala, J., Generoso, S., ... & Marelli, L. (2006). Emissions of primary aerosol and precursor gases in the years 2000 and 1750 prescribed data-sets for AeroCom. *Atmos. Chem. Phys*, **6**, 4321-4344.
- Deser, C., A. S. Phillips, R. A. Tomas, Y. Okumura, M. A. Alexander, A. Capotondi, J. D. Scott, Y. -O. Kwon, and M. Ohba (2012). ENSO and Pacific Decadal Variability in Community Climate System Model Version 4. *J. Climate*, **25**, 2622-2651, doi: 10.1175/JCLI-D-11-00301.1
- Dickson, A. G. and Riley, J. P. (1979) The estimation of acid dissociation constants in seawater media from potentiometric titrations with strong base. I. The ionic product of water – KW, *Mar. Chem.*, **7**, 89–99.

- Donner, L. J., Soden, B. J., & Seman, C. J. (1996). Use of ISCCP Data to Evaluate a GCM Parameterization for Ice Clouds. *WORLD METEOROLOGICAL ORGANIZATION-PUBLICATIONS-WMO TD*, II-39.
- Donner, L. J., Wyman, B. L., Hemler, R. S., Horowitz, L. W., Ming, Y., Zhao, M., ... & Austin, J. (2011). The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3. *Journal of Climate*, **24**(13), 3484-3519.
- Donohue, K. A., Tracey, K. L., Watts, D. R., Chidichimo, M. P., & Chereskin, T. K. (2016). Mean Antarctic circumpolar current transport measured in Drake Passage. *Geophysical Research Letters*, **43**, 11,760–11,767. <https://doi.org/10.1002/2016GL070319>
- Dunne, J. P., J. G. John, A. J. Adcroft, S. M. Griffies, R. W. Hallberg, E. Shevliakova, R. J. Stouffer, W. Cooke, K. A. Dunne, M. J. Harrison, J. P. Krasting, S. L. Malyshev, P. C. D. Milly, P. J. Phillips, L. T. Sentman, B. L. Samuels, M. J. Spelman, M. Winton, A. T. Wittenberg, and N. Zadeh, (2012). GFDL's ESM2 global coupled climate-carbon Earth System Models, Part I: Physical formulation and baseline simulation characteristics. *Journal of Climate*, **25**, 6646-6665. doi: 10.1175/JCLI-D-11-00560.1
- Dunne, J. P., John, J. G., Shevliakova, E., Stouffer, R. J., Krasting, J. P., Malyshev, S. L., ... & Dunne, K. A. (2013). GFDL's ESM2 global coupled climate-carbon earth system models. Part II: carbon system formulation and baseline simulation characteristics. *Journal of Climate*, **26**(7), 2247-2267.
- Eyring, V., Arblaster, J. M., Cionni, I., Sedláček, J., Perlwitz, J., Young, P. J., ... & Faluvegi, G. (2013). Long-term ozone changes and associated climate impacts in CMIP5 simulations. *Journal of Geophysical Research: Atmospheres*, **118**(10), 5029-5060.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development (Online)*, **9**(LLNL-JRNL-736881).
- Evans, S, P Ginoux, S Malyshev, and E Shevliakova (2016). Climate-vegetation interaction and amplification of Australian dust variability. *Geophysical Research Letters*, **43**(22), DOI:[10.1002/2016GL071016](https://doi.org/10.1002/2016GL071016).
- Flanner, M. G., Zender, C. S., Randerson, J. T., & Rasch, P. J. (2007). Present-day climate forcing and response from black carbon in snow. *Journal of Geophysical Research: Atmospheres*, **112**(D11).
- Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., ... & Forest, C. (2014). Evaluation of climate models. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 741-866). Cambridge University Press.
- Fox-Kemper, B., Danabasoglu, G., Ferrari, R., Griffies, S. M., Hallberg, R. W., Holland, M.M., et al. (2011). Parameterization of mixed layer eddies. III: Implementation and impact in

- global ocean climate simulations. *Ocean Modelling*, 39(1), 61–78.
<https://doi.org/10.1016/j.ocemod.2010.09.002>
- Frith, S. M. et al. (2013). Multi-Satellite Merged Ozone (O3) Profile and Total Column 1 Month Zonal Mean L3 Global 5.0 degree Latitude Zones V1, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC),
 10.5067/MEASURES/OZONE/DATA310
- Frölicher, T. L., Sarmiento, J. L., Paynter, D. J., Dunne, J. P., Krasting, J. P., & Winton, M. (2015). Dominance of the Southern Ocean in anthropogenic carbon and heat uptake in CMIP5 models. *Journal of Climate*, 28(2), 862-886.
- Frölicher, T. L., Winton, M., & Sarmiento, J. L. (2014). Continued global warming after CO₂ emissions stoppage. *Nature Climate Change*, 4(1), 40.
- Ganachaud, A. & C. Wunsch (2003). Large-scale ocean heat and freshwater transports during the World Ocean Circulation Experiment. *Journal of Climate*, 16, 696–705.
- Guo, H., Golaz, J. C., Donner, L. J., Wyman, B., Zhao, M., & Ginoux, P. (2015). CLUBB as a unified cloud parameterization: Opportunities and challenges. *Geophysical Research Letters*, 42(11), 4540-4547.
- Gehlen, M., Séférian, R., Jones, D. O., Roy, T., Roth, R., Barry, J. P., ... & Joos, F. (2014). Projected pH reductions by 2100 might put deep North Atlantic biodiversity at risk.
- Gerber, S., Hedin, L. O., Oppenheimer, M., Pacala, S. W., & Shevliakova, E. (2010). Nitrogen cycling and feedbacks in a global dynamic land model. *Global Biogeochemical Cycles*, 24(1).
- Gillett, N. P., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., ... & Tebaldi, C. (2016). Detection and attribution model intercomparison project (damip). *Geoscientific Model Development*, 9(10), 3685-3697.
- Gnanadesikan, A., & Stouffer, R. J. (2006). Diagnosing atmosphere-ocean general circulation model errors relevant to the terrestrial biosphere using the Köppen climate classification. *Geophysical Research Letters*, 33(22).
- Golaz, J. C., Salzmann, M., Donner, L. J., Horowitz, L. W., Ming, Y., & Zhao, M. (2011). Sensitivity of the aerosol indirect effect to subgrid variability in the cloud parameterization of the GFDL atmosphere general circulation model AM3. *Journal of Climate*, 24(13), 3145-3160.
- Golaz, J. C., Horowitz, L. W., & Levy, H. (2013). Cloud tuning in a coupled climate model: Impact on 20th century warming. *Geophysical Research Letters*, 40(10), 2246-2251.
- Guo, H., J. G. John, C. Blanton, C. McHugh, S. Nikonov, A. Radhakrishnan, N Zadeh, V. Balaji, J. Durachta, C. Dupuis, R. Menzel, T. Robinson, S. Underwood, H. Vahlenkamp, K. Dunne, Krista, P. P. G. Gauthier, P. Ginoux, S. M. Griffies, R. W. Hallberg, M. Harrison, W. Hurlin, S. Malyshev, V. Naik, F. Paulot, D. J. Paynter, J. Ploshay, D. M. Schwarzkopf, C. J. Seman, A. Shao, L. Silvers, B. Wyman, Y. Zeng, A. Adcroft, J. P. Dunne, I. M. Held, J. P. Krasting, John P.; Horowitz, Larry W.; Milly, E. Shevliakova, M. Winton, and M. Zhao (2018).

- NOAA-GFDL GFDL-CM4 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation. <https://doi.org/10.22033/ESGF/CMIP6.8594>
- Graham, F. S., A. T. Wittenberg, J. N. Brown, S. J. Marsland, and N. J. Holbrook (2017). Understanding the double peaked El Niño in coupled GCMs. *Climate Dynamics*, **48** (5), 2045-2063. doi: 10.1007/s00382-016-3189-1
- Griffies, S. M., Winton, M., Anderson, W. G., Benson, R., Delworth, T. L., Dufour, C. O., ... & Wittenberg, A. T. (2015). Impacts on ocean heat from transient mesoscale eddies in a hierarchy of climate models. *Journal of Climate*, **28**(3), 952-977.
- Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., & Geron, C. (2006). Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). *Atmospheric Chemistry and Physics*, **6**(11), 3181-3210.
- Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., & Wang, X. (2012). The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2. 1): an extended and updated framework for modeling biogenic emissions. *Geoscientific Model Development*, **5**(6), 1471-1492.
- Guilyardi, E., H. Bellenger, M. Collins, S. Ferrett, W. Cai, and A. Wittenberg (2012). A first look at ENSO in CMIP5. *CLIVAR Exchanges*, **17**, 29-32. ISSN: 1026-0471.
- Held, I. M., H. Guo, A. Adcroft, J. P. Dunne, L. W. Horowitz, J. P. Krasting, E. Shevliakova, M. Winton, M. Zhao, M. Bushuk, A. T. Wittenberg, B. Wyman, B. Xiang, R. Zhang, W. G. Anderson, V. Balaji, L. J. Donner, K. A. Dunne, J. W. Durachta, P. P. G. Gauthier, P. Ginoux, J.-C. Golaz, S. M. Griffies, R. Hallberg, L. Harris, M. J. Harrison, W. J. Hurlin, J. G. John, P. Lin, S.-J. Lin, S. Malyshev, R. Menzel, P. C. D. Milly, Y. Ming, V. Naik, D. J. Paynter, F. Paulot, V. Ramaswamy, B. G. Reichl, T. E. Robinson, A. Rosati, C. J. Seman, L. G. Silvers, S. D. Underwood, and N. Zadeh, 2019: Structure and Performance of GFDL's CM4.0 Climate Model. *Journal of Advances in Modeling Earth Systems*, **11**. <https://doi.org/10.1029/2019MS001829>.
- Heymsfield, A. J., & Donner, L. J. (1990). A scheme for parameterizing cirrus cloud ice water content in general circulation models. *FIRE Science Results* 1989, 3079, 411.
- Hoffman, F. M., Randerson, J. T., Arora, V. K., Bao, Q., Cadule, P., Ji, D., ... & Obata, A. (2014). Causes and implications of persistent atmospheric carbon dioxide biases in Earth System Models. *Journal of Geophysical Research: Biogeosciences*, **119**(2), 141-162.
- Horowitz, L. W., Walters, S., Mauzerall, D. L., Emmons, L. K., Rasch, P. J., Granier, C., ... & Tyndall, G. S. (2003). A global simulation of tropospheric ozone and related tracers: Description and evaluation of MOZART, version 2 (DOI 10.1029/2002JD002853). *JOURNAL OF GEOPHYSICAL RESEARCH-ALL SERIES-*, **108**(24; SECT 4), ACH-16.
- Horowitz, L. W., V. Naik, F. Paulot, P. A. Ginoux, J. P. Dunne, J. Mao, J. Schnell, Xi Chen³, M. Lin, P. Lin, S. Malyshev, D. Paynter, E. Shevliakova, M. Zhao (in review). The GFDL Global Atmospheric Chemistry-Climate Model AM4.1: Model Description and Simulation Characteristics. *Journal of Advances in Modeling Earth Systems*.

- Huang, B., V.F. Banzon, E. Freeman, J. Lawrimore, W. Liu, T.C. Peterson, T.M. Smith, P.W. Thorne, S.D. Woodruff, and H. Zhang (2015). Extended Reconstructed Sea Surface Temperature Version 4 (ERSST.v4). Part I: Upgrades and Intercomparisons. *Journal of Climate*, **28**, 911–930. doi:10.1175/JCLI-D-14-00006.1
- Huang, B., P.W. Thorne, V.F. Banzon, T. Boyer, G. Chepurin, J.H. Lawrimore, M.J. Menne, T.M. Smith, R.S. Vose, and H. Zhang (2017). Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons. *Journal of Climate*, **30**, 8179–8205, doi:10.1175/JCLI-D-16-0836.1
- Jaeglé, L., Quinn, P. K., Bates, T. S., Alexander, B., & Lin, J. T. (2011). Global distribution of sea salt aerosols: new constraints from in situ and remote sensing observations. *Atmospheric Chemistry and Physics*, **11**(7), 3137–3157.
- Jansen, M. F., Adcroft, A. J., Hallberg, R., & Held, I. M. (2015). Parameterization of eddy fluxes based on a mesoscale energy budget. *Ocean Modelling*, **92**, 28– 41.
- Jansen, M. F., Adcroft, A., Khani, S., & Kong, H. (2019). Towards an energetically consistent, resolution aware parameterization of ocean mesoscale eddies. *Journal of Advances in Modeling Earth Systems*.
- John, J. G., Stock, C. A., & Dunne, J. P. (2015). A more productive, but different, ocean after mitigation. *Geophysical Research Letters*, **42**(22), 9836–9845, doi:10.1002/2015GL066160.
- Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., ... & Jung, M. (2016). C4MIP–The Coupled Climate–Carbon Cycle Model Intercomparison Project: experimental protocol for CMIP6. *Geoscientific Model Development*, **9**(8), 2853–2880.
- Keller, D. P., Lenton, A., Scott, V., Vaughan, N. E., Bauer, N., Ji, D., ... & Zickfeld, K. (2018). The carbon dioxide removal model intercomparison project (CDRMIP): rationale and experimental protocol for CMIP6. *Geoscientific Model Development*, **11**(3), 1133–1160.
- Kerr, R. A. (2000). A North Atlantic climate pacemaker for the centuries. *Science*, **288**, 1984–1985, doi:10.1126/science.288.5473.1984.
- Kessler, W. S. (2002). Is ENSO a cycle or a series of events? *Geophys. Res. Lett.*, **29**(23), 2125, doi:10.1029/2002GL015924.
- Key, R. M., Kozyr, A., Sabine, C. L., Lee, K., Wanninkhof, R., Bullister, J. L., ... & Peng, T. H. (2004). A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP). *Global biogeochemical cycles*, **18**(4).
- Kirtman, B.P. (1997). Oceanic Rossby Wave Dynamics and the ENSO Period in a Coupled Model. *Journal of Climate*, **10**, 1690–1704.
- Köppen, W (1923) Dle Klimate der Erde. Walter de Gruyter, Berlin
- Krasting, J. P., Dunne, J. P., Shevliakova, E., & Stouffer, R. J. (2014). Trajectory sensitivity of the transient climate response to cumulative carbon emissions. *Geophysical Research Letters*, **41**(7), 2520–2527.
- Krasting, J. P.; John, J. G; Blanton, C.; McHugh, C.; Nikonov, S.; Radhakrishnan, A., et al. (2018). NOAA-GFDL GFDL-ESM4 model output prepared for CMIP6 CMIP. Earth System Grid Federation. <https://doi.org/10.22033/ESGF/CMIP6.1407>

- Laufkötter, C., John, J. G., Stock, C. A., & Dunne, J. P. (2017). Temperature and oxygen dependence of the remineralization of organic matter. *Global Biogeochemical Cycles*, **31**(7), 1038-1050.
- Larkin, N.K. and D.E. Harrison (2002). ENSO Warm (El Niño) and Cold (La Niña) Event Life Cycles: Ocean Surface Anomaly Patterns, Their Symmetries, Asymmetries, and Implications. *Journal of Climate*, **15**, 1118–1140, [https://doi.org/10.1175/1520-0442\(2002\)015<1118:EWENOA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1118:EWENOA>2.0.CO;2)
- Lawrence, D. M., Hurtt, G. C., Arneth, A., Brovkin, V., Calvin, K. V., Jones, A. D., ... & Seneviratne, S. I. (2016). The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: rationale and experimental design. *Geoscientific Model Development*, **9**(9), 2973-2998.
- Lee, K., Kim, T.-W., Byrne, R. H., Millero, F. J., Feely, R. A., and Liu, Y.-M. (2010). The universal ratio of boron to chlorinity for the North Pacific and North Atlantic oceans, *Geochim. Cosmochim. Ac.*, **74**, 1801–1811, doi:10.1016/j.gca.2009.12.027.
- Lee, T., Waliser, D. E., Li, J. L. F., Landerer, F. W., & Gierach, M. M. (2013). Evaluation of CMIP3 and CMIP5 wind stress climatology using satellite measurements and atmospheric reanalysis products. *Journal of Climate*, **26**(16), 5810-5826.
- Leung, L.R., and S.J. Ghan (1998). Parameterizing Subgrid Orographic Precipitation and Surface Cover in Climate Models. *Mon. Wea. Rev.*, **126**(12), 3271-3291.
- Leuning, R. (1995). A critical appraisal of a combined stomatal-photosynthesis model for C3 plants. *Plant, Cell & Environment*, **18**(4), 339-355.
- Li, J., Mao, J., Min, K.-E., Washenfelder, R. A., Brown, S. S., Kaiser, J., Keutsch, F. N., Volkamer, R., Wolfe, G. M., Hanisco, T. F., Pollack, I. B., Ryerson, T. B., Graus, M., Gilman, J. B., Lerner, B.M., Warneke, C., deGouw, J.A., Middlebrook, A.M., Liao, J., Welti, A., Henderson, B. H., McNeill, V. Faye, Hall, S. R., Ullmann, K., Donner, L. J., Paulot, F., and Horowitz, L. W. (2016) Observational constraints on glyoxal production from isoprene oxidation and its contribution to organic aerosol over the Southeast United States, *J. Geophys. Res.-Atmos.*, **121**, 9849–9861, <https://doi.org/10.1002/2016JD025331>.
- Lin, M., Fiore, A. M., Horowitz, L. W., Cooper, O. R., Naik, V., Holloway, J., ... & Ryerson, T. B. (2012). Transport of Asian ozone pollution into surface air over the western United States in spring. *Journal of Geophysical Research: Atmospheres*, **117**(D21).
- Lin, S. J. (2004). A “vertically Lagrangian” finite-volume dynamical core for global models. *Monthly Weather Review*, **132**(10), 2293-2307.
- Locarnini, R. A., A. V. Mishonov, J. I. Antonov, T. P. Boyer, H. E. Garcia, O. K. Baranova, M. M. Zweng, C. R. Paver, J. R. Reagan, D. R. Johnson, M. Hamilton, and D. Seidov, 2013. World Ocean Atlas 2013, Volume 1: Temperature. S. Levitus, Ed., A. Mishonov Technical Ed.; NOAA Atlas NESDIS 73, 40 pp.
- Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., ... & Kato, S. (2018). Clouds and the earth’s radiant energy system (CERES) energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition-4.0 data product. *Journal of Climate*, **31**(2), 895-918.

- Lohmann, U., Sausen, R., Bengtsson, L., Cubasch, U., Perlwitz, J., & Roeckner, E. (1993). The Köppen climate classification as a diagnostic tool for general circulation models. *Climate Research*, **3**, 177-193.
- Lotze, H. K., Tittensor, D. P., Bryndum-Buchholz, A., Eddy, T. D., Cheung, W. W., Galbraith, E. D., ... & Bopp, L. (2019). Global ensemble projections reveal trophic amplification of ocean biomass declines with climate change. *Proceedings of the National Academy of Sciences*, **116**(26), 12907-12912.
- Mao, J., Paulot, F., Jacob, D. J., Cohen, R. C., Crounse, J. D., Wennberg, P. O., ... & Horowitz, L. W. (2013). Ozone and organic nitrates over the eastern United States: Sensitivity to isoprene chemistry. *Journal of Geophysical Research: Atmospheres*, **118**(19), 11-256.
- Mårtensson, E. M., Nilsson, E. D., de Leeuw, G., Cohen, L. H., & Hansson, H. C. (2003). Laboratory simulations and parameterization of the primary marine aerosol production. *Journal of Geophysical Research: Atmospheres*, **108**(D9).
- Meehl, G. A., Teng, H., & Arblaster, J. M. (2014). Climate model simulations of the observed early-2000s hiatus of global warming. *Nature Climate Change*, **4**(10), 898.
- Millero, F. J.: Carbonate constants for estuarine waters, *Mar. Freshwater Res.*, **61**, 139–142, doi:10.1071/MF09254, 2010.
- Mongwe, N. P., Chang, N., & Monteiro, P. M. (2016). The seasonal cycle as a mode to diagnose biases in modelled CO₂ fluxes in the Southern Ocean. *Ocean Modelling*, **106**, 90-103.
- Morice, C.P., Kennedy, J.J., Rayner, N.A. and Jones, P.D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 dataset. *Journal of Geophysical Research*, **117**, D08101, doi:10.1029/2011JD017187
- Nachtergaele, F. O., van Velthuisen, H. , Wiberg, D., Batjes, N. H., Dijkshoorn, J. A., van Engelen, V. W. P., Fischer, G., Jones, A., Montanarella, L., Petri, M., Prieler, S., Teixeira, E., Shi, X. (2014). Harmonized World Soil Database (HWSD). ISRIC - World Soil Information.
- Naik, V., Horowitz, L. W., Fiore, A. M., Ginoux, P., Mao, J., Aghedo, A. M., & Levy, H. (2013). Impact of preindustrial to present-day changes in short-lived pollutant emissions on atmospheric composition and climate forcing. *Journal of Geophysical Research: Atmospheres*, **118**(14), 8086-8110.
- New, M. G., M. Hulme, and P. D. Jones (1999). Representing twentieth century space-time climate variability. part I: Development of a 1961 – 90 mean monthly terrestrial climatology, *J. Clim.*, **12**, 829 – 856.
- O'Neill, B. C., C. Tebaldi, D. P. van Vuuren, V. Eyring, P. Friedlingstein, G. Hurtt, R. Knutti et al (2016). "The scenario model intercomparison project (ScenarioMIP) for CMIP6." *Geoscientific Model Development* **9**(9), 3461-3482.
- Orr, J. C. (2002) Global Ocean Storage of Anthropogenic Carbon (GOSAC), final report, EC Environment and Climate Programme, Institute Pierre Simon Laplace, Paris
- Orr, J. C., & Epitalon, J. M. (2015). Improved routines to model the ocean carbonate system: mocsy 2.0. *Geoscientific Model Development*, **8**(3), 485-499.

- Orr, J. C., Najjar, R. G., Aumont, O., Bopp, L., Bullister, J. L., Danabasoglu, G., ... & Griffies, S. M. (2017). Biogeochemical protocols and diagnostics for the CMIP6 Ocean Model Intercomparison Project (OMIP). *Geoscientific Model Development*, **10**(6), 2169-2199.
- Pai, S. J., Heald, C. L., Pierce, J. R., Farina, S. C., Marais, E. A., Jimenez, J. L., ... & Shilling, J. E. (2020). An evaluation of global organic aerosol schemes using airborne observations.
- Paulot, F., Jacob, D. J., Johnson, M. T., Bell, T. G., Baker, A. R., Keene, W. C., ... & Stock, C. A. (2015). Global oceanic emission of ammonia: Constraints from seawater and atmospheric observations. *Global Biogeochemical Cycles*, **29**(8), 1165-1178.
- Paulot, F., Ginoux, P., Cooke, W. F., Donner, L. J., Fan, S., Lin, M. Y., ... & Horowitz, L. W. (2016). Sensitivity of nitrate aerosols to ammonia emissions and to nitrate chemistry: implications for present and future nitrate optical depth. *Atmospheric Chemistry and Physics*, **16**(3), 1459-1477.
- Paulot, F., Fan, S., & Horowitz, L. W. (2017). Contrasting seasonal responses of sulfate aerosols to declining SO₂ emissions in the Eastern US: Implications for the efficacy of SO₂ emission controls. *Geophysical Research Letters*, **44**(1), 455-464.
- Paulot, F., Paynter, D., Winton, M., Ginoux, P., Zhao, M., & Horowitz, L. W. (2020). Revisiting the impact of sea salt on climate sensitivity. *Geophysical Research Letters*, **47**(3), e2019GL085601.
- Paulot, F., C. Stock, J. G. John, N. Zadeh, and L. W. Horowitz (in review) Ocean ammonia outgassing: modulation by CO₂ and anthropogenic nitrogen deposition. *Journal of Advances in Modeling Earth Systems*.
- Phillips, A. S., C. Deser, and J. Fasullo (2014). A New Tool for Evaluating Modes of Variability in Climate Models. *EOS*, **95**, 453-455, doi: 10.1002/2014EO490002.
- Pincus, R., Forster, P. M., & Stevens, B. (2016). The radiative forcing model intercomparison project (RFMIP): Experimental protocol for CMIP6. *Geoscientific Model Development (Online)*, **9**(9).
- Prospero, J. M., and Lamb, P. J. (2003). African droughts and dust transport to the Caribbean: Climate change implications. *Science*, **302**(5647), 1024-1027.
- Purves, D. W., Lichstein, J. W., Strigul, N., & Pacala, S. W. (2008). Predicting and understanding forest dynamics using a simple tractable model. *Proceedings of the National Academy of Sciences*, **105**(44), 17018-17022.
- Rabin, S. S., Magi, B. I., Shevliakova, E., & Pacala, S. W. (2015). Quantifying regional, time-varying effects of cropland and pasture on vegetation fire. *Biogeosciences*, **12**(22), 6591-6604.
- Randerson, J. T., Lindsay, K., Munoz, E., Fu, W., Moore, J. K., Hoffman, F. M., ... & Doney, S. C. (2015). Multicentury changes in ocean and land contributions to the climate-carbon feedback. *Global Biogeochemical Cycles*, **29**(6), 744-759.
- Rayner, N. A. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., ... & Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine

- air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, 108(D14).
- Reichl, B. G., & Hallberg, R. (2018). A simplified energetics based planetary boundary layer (ePBL) approach for ocean climate simulations. *Ocean Modelling*, 132, 112-129.
- Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., et al. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of Climate*, 24, 3624–3648.
- Resplandy, L., Bopp, L., Orr, J. C., & Dunne, J. P. (2013). Role of mode and intermediate waters in future ocean acidification: Analysis of CMIP5 models. *Geophysical Research Letters*, 40(12), 3091-3095.
- Resplandy, L., Keeling, R. F., Rödenbeck, C., Stephens, B. B., Khatiwala, S., Rodgers, K. B., ... & Tans, P. P. (2018). Revision of global carbon fluxes based on a reassessment of oceanic and riverine carbon transport. *Nature Geoscience*, 11(7), 504.
- Richter I, Xie S-P (2008) On the origin of equatorial Atlantic biases in coupled general circulation models. *Climate Dynamics* 31:587–598
- Russell, J. L., Stouffer, R. J., & Dixon, K. W. (2006). Intercomparison of the Southern Ocean circulations in the IPCC coupled model control simulations. *Journal of Climate*, 19, 4560–4575.
- Sander, S. P., J. Abbatt, J. R. Barker, J. B. Burkholder, R. R. Friedl, D. M. Golden, R. E. Huie, C. E. Kolb, M. J. Kurylo, G. K. Moortgat, V. L. Orkin and P. H. Wine (2011). Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies, Evaluation No. 17, JPL Publication 10-6, Jet Propulsion Laboratory, Pasadena. <http://jpldataeval.jpl.nasa.gov>.
- Santanello, J. A., Jr., and Coauthors, 2018: Land–atmosphere interactions: The LoCo perspective. *Bull. Amer. Meteor. Soc.*, <https://doi.org/10.1175/BAMS-D-17-0001.1>,
- Sarmiento, J. L., Gruber, N., Brzezinski, M. A., & Dunne, J. P. (2004). High-latitude controls of thermocline nutrients and low latitude biological productivity. *Nature*, 427(6969), 56-60.
- Savoie, D. L., and J. M. Prospero (1989). Comparison of oceanic and continental sources of non-sea-salt sulphate over the Pacific Ocean, *Nature*, 339, 685–687.
- Schnell, J. L., Naik, V., Horowitz, L. W., Paulot, F., Mao, J., Ginoux, P., ... & Ram, K. (2018). Exploring the relationship between surface PM 2.5 and meteorology in Northern India. *Atmospheric Chemistry and Physics*, 18(14), 10157-10175.
- Séférian, R., Gehlen, M., Bopp, L., Resplandy, L., Orr, J. C., Marti, O., ... & Lindsay, K. (2016). Inconsistent strategies to spin up models in CMIP5: implications for ocean biogeochemical model performance assessment. *Geoscientific Model Development*, 9(5).
- Shevliakova, E. et al. (in review) GFDL land model LM4.1. *Journal of Advances in Modeling Earth Systems*.
- Shindell, D. T., Lamarque, J. F., Schulz, M., Flanner, M., Jiao, C., Chin, M., ... & Milly, G. (2013). Radiative forcing in the ACCMIP historical and future climate simulations. *Atmospheric Chemistry and Physics*, 13(6), 2939-2974.

- Stock, C. A., Dunne, J. P., & John, J. G. (2014). Global-scale carbon and energy flows through the marine planktonic food web: An analysis with a coupled physical–biological model. *Progress in Oceanography*, **120**, 1-28.
- Stock, C. A., J. P. Dunne, S. Fan, P. Ginoux, J. John, J. Krasting, C. Laufkötter, F. Paulot, and N. Zadeh (in review). Ocean Biogeochemistry in GFDL’s Earth System Model 4.1 and its Response to Increasing Atmospheric CO₂. *Journal of Advances in Modeling Earth Systems*.
- Stockdale, T. N., Busalacchi, A. J., Harrison, D. E., & Seager, R. (1998). Ocean modeling for ENSO. *Journal of Geophysical Research: Oceans*, **103**(C7), 14325-14355.
- Subin, Z. M., Milly, P. C., Sulman, B. N., Malyshev, S., & Shevliakova, E. (2014). Resolving terrestrial ecosystem processes along a subgrid topographic gradient for an earth-system model. *Hydrology & Earth System Sciences Discussions*, 11(7).
- Sulman, B. N., Phillips, R. P., Oishi, A. C., Shevliakova, E., & Pacala, S. W. (2014). Microbe-driven turnover offsets mineral-mediated storage of soil carbon under elevated CO₂. *Nature Climate Change*, **4**(12), 1099.
- Sulman, B. N., E Shevliakova, E R Brzostek, S N Kivlin, and S Malyshev, et al. (2019) Diverse mycorrhizal associations enhance terrestrial C storage in a global model. *Global Biogeochemical Cycles*, **33**(4), DOI:[10.1029/2018GB005973](https://doi.org/10.1029/2018GB005973).
- Talley, L. D., J. L. Reid, and P. E. Robbins (2003). Data-based meridional overturning streamfunctions for the global ocean. *Journal of Climate*, **16**, 3213–3226.
- Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, **93**(4), 485-498.
- Timmermann, R., Beckmann, A., & Hellmer, H. H. (2001). The role of sea ice in the fresh-water budget of the Weddell Sea, Antarctica. *Annals of Glaciology*, **33**, 419-424.
- Todd-Brown, K. E., Randerson, J. T., Post, W. M., Hoffman, F. M., Tarnocai, C., Schuur, E. A., & Allison, S. D. (2013). Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences*, **10**(3), 1717-1736.
- Trenberth, K. E., & Caron, J. M. (2001). Estimates of meridional atmosphere and ocean heat transports. *Journal of Climate*, **14**(16), 3433-3443.
- Trenberth, K. E., & D. J. Shea (2006). Atlantic hurricanes and natural variability in 2005, *Geophys. Res. Lett.*, **33**, L12704, doi:10.1029/2006GL026894
- Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., ... & Li, X. (2005). The ERA-40 re-analysis. *Quarterly Journal of the royal meteorological society*, **131**(612), 2961-3012.
- Vecchi, G. A., T. Delworth, R. Gudgel, S. Kapnick, A. Rosati, A. T. Wittenberg, F. Zeng, et al., 2014: On the seasonal forecasting of regional tropical cyclone activity. *Journal of Climate*, **27**, 7994-8016. doi: 10.1175/JCLI-D-14-00158.1
- Vose, R. S., Arndt, D., Banzon, V. F., Easterling, D. R., Gleason, B., Huang, B., ... & Reynolds, R. W. (2012). NOAA's merged land–ocean surface temperature analysis. *Bulletin of the American Meteorological Society*, **93**(11), 1677-1685.

- Ward, D S., E Shevliakova, S Malyshev, and S Rabin (2018). Trends and variability of global fire emissions due to historical anthropogenic activities. *Global Biogeochemical Cycles*, **32**(1), DOI:[10.1002/2017GB005787](https://doi.org/10.1002/2017GB005787).
- Weng, E. S., Malyshev, S., Lichstein, J. W., Farrior, C. E., Dybzinski, R., Zhang, T., ... & Pacala, S. W. (2015). Scaling from individual trees to forests in an Earth system modeling framework using a mathematically tractable model of height-structured competition. *Biogeosciences*, **12**(9), 2655-2694.
- Westervelt, D. M., Horowitz, L. W., Naik, V., Tai, A. P. K., Fiore, A. M., & Mauzerall, D. L. (2016). Quantifying PM_{2.5}-meteorology sensitivities in a global climate model. *Atmospheric environment*, **142**, 43-56.
- Winton, M., Adcroft, A., Dunne, J. P., Held, I. M., Shevliakova, E., Zhao, M., ... & Paynter, D. (2020). Climate Sensitivity of GFDL's CM4. 0. *Journal of Advances in Modeling Earth Systems*, **12**(1), e2019MS001838.
- Wittenberg, A. T., A. Rosati, N.-C. Lau, and J. J. Ploshay (2006). GFDL's CM2 global coupled climate models, Part III: Tropical Pacific climate and ENSO. *Journal of Climate*, **19**, 698-722. doi: 10.1175/JCLI3631.1
- Wittenberg, A. T., (2009). Are historical records sufficient to constrain ENSO simulations? *Geophys. Res. Lett.*, **36**, L12702. doi: 10.1029/2009GL038710
- Wittenberg, A. T., A. Rosati, T. L. Delworth, G. A. Vecchi, and F. Zeng (2014). ENSO modulation: Is it decadal predictability? *Journal of Climate*, **27**, 2667-2681. doi: 10.1175/JCLI-D-13-00577.1
- Wittenberg, A. T., G. A. Vecchi, T. L. Delworth, A. Rosati, W. Anderson, W. F. Cooke, S. Underwood, F. Zeng, S. Griffies, and S. Ray (2018). Improved simulations of tropical Pacific annual-mean climate in the GFDL FLOR and HiFLOR coupled GCMs. *Journal of Advances in Modeling Earth Systems*, **10** (12), 3176-3220. doi:10.1029/2018MS001372
- Wolf, A., Anderegg, W. R., & Pacala, S. W. (2016). Optimal stomatal behavior with competition for water and risk of hydraulic impairment. *Proceedings of the National Academy of Sciences*, **113**(46), E7222-E7230.
- Yin, L., Fu, R., Shevliakova, E., & Dickinson, R. E. (2013). How well can CMIP5 simulate precipitation and its controlling processes over tropical South America?. *Climate Dynamics*, **41**(11-12), 3127-3143.
- Vavrus, S., & Waliser, D. (2008). An improved parameterization for simulating Arctic cloud amount in the CCSM3 climate model. *Journal of Climate*, **21**(21), 5673-5687.
- Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., ... & Taylor, K. E. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, **47**(1), e2019GL085782.
- Zhang, L., Delworth, T. L., Cooke, W., & Yang, X. (2019). Natural variability of Southern Ocean convection as a driver of observed climate trends. *Nature Climate Change*, **9**(1), 59.
- Zhang, J. L., & M. Steele (2007). Effect of vertical mixing on the Atlantic Water layer circulation in the Arctic Ocean, *J. Geophys. Res.*, **112**, C04S04, doi:10.1029/2006JC003732.

- Zhao, M. (2014). An investigation of the connections among convection, clouds, and climate sensitivity in a global climate model. *Journal of Climate*, 27(5), 1845-1862.
- Zhao, M., Golaz, J. C., Held, I. M., Ramaswamy, V., Lin, S. J., Ming, Y., ... & Guo, H. (2016). Uncertainty in model climate sensitivity traced to representations of cumulus precipitation microphysics. *Journal of Climate*, 29(2), 543-560.
- Zhao, M., Golaz, J. C., Held, I. M., Guo, H., Balaji, V., Benson, R., ... & Dunne, K. (2018a). The GFDL global atmosphere and land model AM4. 0/LM4. 0: 1. Simulation characteristics with prescribed SSTs. *Journal of Advances in Modeling Earth Systems*, 10(3), 691-734.
- Zhao, M., Golaz, J. C., Held, I. M., Guo, H., Balaji, V., Benson, R., ... & Dunne, K. (2018b). The GFDL global atmosphere and land model AM4. 0/LM4. 0: 2. Model description, sensitivity studies, and tuning strategies. *Journal of Advances in Modeling Earth Systems*, 10(3), 735-769.
- Ziemke, J. R., Oman, L. D., Strode, S. A., Douglass, A. R., Olsen, M. A., McPeters, R. D., ... & Thompson, A. M. (2019). Trends in global tropospheric ozone inferred from a composite record of TOMS/OMI/MLS/OMPS satellite measurements and the MERRA-2 GMI simulation. *Atmospheric Chemistry and Physics*, 19(5), 3257-3269.
- Zuidema, P., C. Alvarez, S. J. Kramer, L. Custals, M. Izaguirre, P. Sealy, J. M. Prospero, and E. Blades (2019). Is Summer African Dust Arriving Earlier to Barbados? The Updated Long-Term In Situ Dust Mass Concentration Time Series from Ragged Point, Barbados, and Miami, Florida. *Bulletin of the American Meteorological Society* 100, 1981-1986.
- Zweng, M.M, J.R. Reagan, J.I. Antonov, R.A. Locarnini, A.V. Mishonov, T.P. Boyer, H.E. Garcia, O.K. Baranova, D.R. Johnson, D.Seidov, M.M. Biddle (2013). World Ocean Atlas 2013, Volume 2: Salinity. S. Levitus, Ed., A. Mishonov Technical Ed.; NOAA Atlas NESDIS 74, 39 p.p