Key Points:
- Community Earth System Model Version 2 includes many substantial science and infrastructure improvements since its previous version
- Preindustrial control and historical simulations were performed with low-top and high-top with comprehensive chemistry atmospheric models
- Comparisons to observations are improved relative to previous versions, including major reductions in radiation and precipitation biases

Correspondence to:
G. Danabasoglu,
gokhan@ucar.edu

The Community Earth System Model Version 2 (CESM2)

G. Danabasoglu[1], J.-F. Lamarque[1], J. Bacmeister[1], D. A. Bailey[1], A. K. DuVivier[1], J. Edwards[1], L. K. Emmons[2], J. Fasullo[1], R. Garcia[2], A. Gettelman[1,2], C. Hannay[1], M. M. Holland[1], W. G. Large[1], P. H. Lauritzen[1], D. M. Lawrence[1], J. T. M. Lenaerts[3], K. Lindsay[1], W. H. Lipscomb[1], M. J. Mills[2], R. Neale[1], K. W. Oleson[1], B. Otto-Bliesner[1], A. S. Phillips[1], W. Sacks[1], S. Tilmes[2], L. van Kampenhout[4], M. Vertenstein[1], A. Bertini[1], J. Dennis[5], C. Deser[1], C. Fischer[1], B. Fox-Kemper[6], J. E. Kay[7], D. Kinnison[2], P. J. Kushner[8], V. E. Larson[9], M. C. Long[1], S. Mickelson[5], J. K. Moore[10], E. Nienhouse[5], L. Polvani[11], P. J. Rasch[12], and W. G. Strand[1]

[1]Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO, USA, [2]Atmospheric Chemistry Observations and Modeling Laboratory, National Center for Atmospheric Research, Boulder, CO, USA, [3]Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder, Boulder, CO, USA, [4]Institute for Marine and Atmospheric Research Utrecht, Utrecht University, Utrecht, Netherlands, [5]Computational and Information Systems Laboratory, National Center for Atmospheric Research, Boulder, CO, USA, [6]Department of Earth, Environmental, and Planetary Sciences, Brown University, Providence, RI, USA, [7]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA, [8]Department of Physics, University of Toronto, Toronto, ON, Canada, [9]Department of Mathematical Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI, USA, [10]Department of Earth System Science, University of California Irvine, Irvine, CA, USA, [11]Applied Mathematics/Earth and Environmental Sciences, Columbia University, New York, NY, USA, [12]Atmospheric Science and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA

Abstract An overview of the Community Earth System Model Version 2 (CESM2) is provided, including a discussion of the challenges encountered during its development and how they were addressed. In addition, an evaluation of a pair of CESM2 long preindustrial control and historical ensemble simulations is presented. These simulations were performed using the nominal 1° horizontal resolution configuration of the coupled model with both the "low-top" (40 km, with limited chemistry) and "high-top" (130 km, with comprehensive chemistry) versions of the atmospheric component. CESM2 contains many substantial science and infrastructure improvements and new capabilities since its previous major release, CESM1, resulting in improved historical simulations in comparison to CESM1 and available observations. These include major reductions in low-latitude precipitation and shortwave cloud forcing biases; better representation of the Madden-Julian Oscillation; better El Niño-Southern Oscillation-related teleconnections; and a global land carbon accumulation trend that agrees well with observationally based estimates. Most tropospheric and surface features of the low- and high-top simulations are very similar to each other, so these improvements are present in both configurations. CESM2 has an equilibrium climate sensitivity of 5.1–5.3 °C, larger than in CESM1, primarily due to a combination of relatively small changes to cloud microphysics and boundary layer parameters. In contrast, CESM2′s transient climate response of 1.9–2.0 °C is comparable to that of CESM1. The model outputs from these and many other simulations are available to the research community, and they represent CESM2′s contributions to the Coupled Model Intercomparison Project Phase 6.

Plain Language Summary The Community Earth System Model (CESM) is an open-source, comprehensive model used in simulations of the Earth′s past, present, and future climates. The newest version, CESM2, has many new technical and scientific capabilities ranging from a more realistic representation of Greenland′s evolving ice sheet, to the ability to model in detail how crops interact with the larger Earth system, to improved representation of clouds and rain, and to the addition of wind-driven waves on the model′s ocean surface. The data sets from a large set of simulations that include integrations for the preindustrial conditions (1850s) and for the 1850-2014 historical period are available to the community, representing CESM2′s contributions to the Coupled Model Intercomparison Project Phase 6 (CMIP6).

1. Introduction

The Community Earth System Model Version 2 (CESM2) is the latest generation of the coupled climate/Earth system models developed as a collaborative effort between scientists, software engineers, and students from the National Center for Atmospheric Research (NCAR), universities, and other research institutions. The CESM2 builds on many successes of its predecessors. The first global coupled climate model that did not use any flux corrections, which were previously required to stabilize coupled simulations even in the absence of changed radiative forcing, was developed at NCAR. Washington and Meehl (1989) performed several short (multidecadal) simulations with this model to study climate sensitivity. The first coupled climate model to achieve a stable present-day control simulation without any flux corrections after long multi-century integrations was the Climate System Model (CSM; Boville & Gent, 1998). This latter effort was followed by the Community Climate System Model Version 2 (CCSM2; Kiehl & Gent, 2004), the CCSM3 (Collins et al., 2006), and the CCSM4 (Gent et al., 2011). Additional capabilities such as interactive carbon-nitrogen cycling, global dynamic vegetation and land use change due to anthropogenic activities, a marine biogeochemistry module, a dynamic ice sheet model, and new chemical and physical processes for direct and indirect aerosol effects were added to develop CCSM into an Earth system model. With these new capabilities, the model was renamed the Community Earth System Model (CESM1; Hurrell et al., 2013).

These CCSM and CESM versions have been at the forefront of national and international efforts to understand and predict the behavior of Earth's climate. Output from numerous simulations using CCSM and CESM has been routinely used in studies to better understand the processes and mechanisms responsible for climate variability and change. Most of these studies make use of CCSM's and CESM's contributions to the various phases of the Coupled Model Intercomparison Project (CMIP). Evaluation of those contributions has identified CCSM and CESM as among the most realistic climate models in the world based on a few metrics that compare the model outputs against present-day observationally based data sets (Knutti et al., 2013). In addition, CESM1 has also been used in key activities to advance our understanding of the climate system and its variability and predictability, by supplementing CESM1's contributions to the CMIPs with other community driven science efforts. These include the CESM1 Large Ensemble (CESM1(LENS); Kay et al., 2015), the CESM Decadal Prediction Large Ensemble (CESM-DPLE; Yeager et al., 2018), the CESM Last Millennium Ensemble (CESM-LME; Otto-Bliesner et al., 2016), and the CESM Stratospheric Aerosol Geoengineering Large Ensemble (GLENS; Tilmes et al., 2018).

The CESM2 was released to the community in June 2018 (available at www.cesm.ucar.edu:/models/cesm2/). Simulations with the model have been done as contributions to the CMIP Phase 6 (CMIP6; Eyring et al., 2016) using the nominal 1° horizontal resolution configuration with both the so-called "low-top" (limited chemistry) and "high-top" with comprehensive chemistry versions of the atmospheric model (see section 2). Following the notation introduced in Hurrell et al. (2013), these configurations will be referred to as CESM2(CAM6) and CESM2(WACCM6), respectively. CMIP6-required core simulations are the so-called Diagnostic, Evaluation, and Characterization of Klima (DECK) experiments that consist of a long preindustrial (PI) control simulation; an abrupt quadrupling of CO_2 concentration simulation; a 1% per year CO_2 concentration increase simulation; and an AMIP (Atmospheric Model Intercomparison Project) simulation forced with prescribed observed sea surface temperatures (SSTs) and sea-ice concentrations. These are complemented by multiple simulations of the historical (1850–2014) period. In addition, CESM2 is participating in about 20 Model Intercomparison Projects (MIPs) within CMIP6, including the ScenarioMIP, which requests future-projection simulations driven by alternative plausible futures of emissions and land use (O'Neill et al., 2016). The output fields from these CESM2 simulations have been posted on the Earth System Grid Federation (ESGF; https://esgf-node.llnl.gov/search/cmip6). To expedite the use of CESM2 by the community primarily for CMIP6-related science and simulations, two *incremental* releases of CESM2 with the same base code as in the June 2018 release were made available in December 2018 (CESM2.1.0) and in June 2019 (CESM2.1.1). These two releases contain additional configurations from which many of the CMIP6 experiments can be run as out-of-the-box simulations.

This manuscript provides an overview of the CESM2 as well as an evaluation of the solutions from CESM2 (CAM6) and CESM2(WACCM6) PI control simulations and from the subsequent ensemble sets of historical simulations with 11 and 3 members, respectively. Section 2 summarizes the major new features and capabilities introduced in each component model since CESM1. Section 3 briefly documents challenges
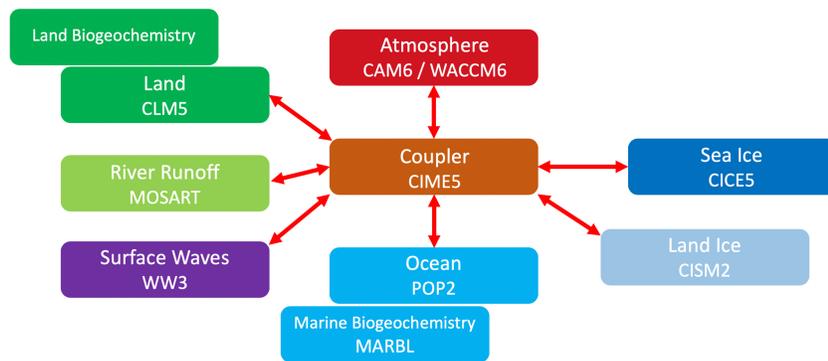
Figure 1. Schematic of the CESM2 component models.

encountered during the development process and the strategies adopted to address them. Initial conditions and forcing data sets used in the simulations are detailed in sections 4 and 5, respectively. Information on model tuning, spin-up, and drifts in the PI control simulations is presented in section 6. Highlights from primarily the historical simulations that include variables from many component models are presented in sections 7–11. CESM2 results are shown with respect to CESM1 PI control and CESM1(LENS) historical simulations as well as available observations. Section 12 briefly discusses the new model's equilibrium climate sensitivity and transient climate response. Finally, section 13 provides a summary and discussion. While much of our analysis makes use of the full set of ensemble members available, in a few instances fewer (or even one) ensemble members are used for clarity and simplicity, as many of the simulation characteristics or improvements are robust across the ensemble members. Solutions from the PI control and historical simulations as well as many other CESM2 DECK and MIP simulations are analyzed and documented in more detail in the manuscripts collected as part of the *AGU CESM2 Virtual Special Issue*.

2. CESM2 and Its Components

The CESM2 is an open-source community coupled model consisting of ocean, atmosphere (both low-top and high-top with comprehensive chemistry), land, sea-ice, land-ice, river, and wave models that exchange states and fluxes via a coupler (Figure 1). It contains many substantial science and infrastructure improvements and capabilities since its previous version, CESM1. In this section, each component model is introduced, and new developments are summarized. As noted below, the individual version numbers across component models differ, reflecting their independent development paths.

2.1. Atmosphere

The Community Atmosphere Model Version 6 (CAM6) uses the same Finite Volume (FV) dynamical core (Lin & Rood, 1997) as in CCSM4 and CESM1, but it includes many changes to the existing representation of physical processes. With the exception of radiation (Rapid Radiative Transfer Model for General circulation models, RRTMG; Iacono et al., 2008), all parameterizations have undergone some level of change. The primary change is the inclusion of the unified turbulence scheme, Cloud Layers Unified By Binormals (CLUBB; Golaz et al., 2002; Larson, 2017). CLUBB unifies the description of cloudy turbulent layers by directly replacing the separate shallow-convection, boundary layer, and grid-scale condensation schemes present in CAM5 (Bogenschutz et al., 2013). It is a high-order closure representation of moist turbulence that closes many terms by novel use of a simple, multivariate binormal probability density function (PDF) describing subgrid scale variations in temperature, humidity, and vertical velocity. The Morrison-Gettelman cloud microphysics scheme (MG1; Morrison & Gettelman, 2008) has been replaced by an updated version (MG2; Gettelman & Morrison, 2015), which is now able to forecast, instead of diagnose, the mass and number concentration of falling condensed species (rain and snow). Notably, mixed phase ice nucleation depends on aerosols, rather than just temperature, following Hoose et al. (2010), as implemented in CESM2 by Wang et al. (2014) with modifications of Shi et al. (2015). Aerosols are treated using the Modal Aerosol Model version 4 (MAM4; Liu et al., 2016). A new subgrid orographic form drag parameterization (Beljaars et al., 2004) replaces the Turbulent Mountain Stress (TMS) scheme and now operates with a diagnosed profile that may extend above the surface level. An anisotropic orographic gravity wave

drag scheme following Scinocca and Mcfarlane (2000) represents vertically propagating gravity waves and near-surface drag as a function of mountain ridge orientation and height to more accurately represent the dependence on near-surface flow direction and mountain orientation. The workhorse version of CAM6 uses a nominal 1° (1.25° in longitude and 0.9° in latitude) horizontal resolution with 32 vertical levels and a model top at 2.26 hPa (about 40 km). This version of the model has relatively coarse stratospheric representation (and no prognostic chemistry module for ozone and other stratospheric constituents) and is thus referred to as a "low-top" model.

The Whole Atmosphere Community Climate Model Version 6 (WACCM) in CESM2 is configured identically to CAM6, except that it uses 70 vertical levels and its model top is at 4.5×10^{-6} hPa (about 130 km). CAM6 and WACCM6 share the same vertical level structure between the surface and 87 hPa. Compared to CAM6, this version of the model has superior stratospheric representation and is thus referred to as a "high-top" model. WACCM6 features a comprehensive chemistry mechanism with a description of tropospheric, stratospheric, and upper-atmospheric chemistry. The default WACCM6 chemical mechanism contains 228 prognostic chemical species with reactions relevant for the whole atmosphere: Troposphere, Stratosphere, Mesosphere, and Lower Thermosphere (TSMLT). In particular, it includes an extensive representation of secondary organic aerosols (Tilmes et al., 2019). MAM4 has been modified to incorporate a new prognostic stratospheric aerosol capability (Mills et al., 2016). The modifications include mode width changes, growth of sulfate aerosol into the coarse mode, and the evolution of stratospheric sulfate aerosol from natural and anthropogenic emissions of source gases, including carbonyl sulfide (OCS) and volcanic sulfur dioxide (SO_2). As described by Gettelman, Mills, et al. (2019) and as also shown below, WACCM6 simulates nearly the same climate as CAM6 in CESM2, with slight differences due to aerosol responses to chemistry, and upper atmospheric processes above the CAM6 model top. WACCM6 also simulates internal variability in the stratosphere, including Stratospheric Sudden Warming (SSW) events on the intraseasonal timescales and the explicitly-resolved Quasi-Biennial Oscillation.

2.2. Ocean

The CESM2 ocean component is the same as in CCSM4 and CESM1 but features several physics and numerical improvements; it is a level-coordinate model based on the Parallel Ocean Program Version 2 (POP2; Smith et al., 2010), as described in Danabasoglu et al. (2012). The physics advances include a new parameterization for mixing effects in estuaries, which improves the representation of the exchange of freshwater between the terrestrial and marine branches of the hydrologic cycle (Sun et al., 2019); increased mesoscale eddy (isopycnal) diffusivities at depth to improve the representation of passive tracers; use of prognostic chlorophyll for shortwave absorption; use of salinity-dependent freezing-point together with the sea-ice model (Assur, 1958); and a new Langmuir mixing parameterization in conjunction with the new wave model component (Li et al., 2016; also see below). The numerical improvements include a new iterative solver for the barotropic mode to reduce communication costs, particularly advantageous for high-resolution simulations on large processor counts (Huang et al., 2016); a tracer-conserving time filtering scheme based on an adaption of the Robert-Asselin filter to enable subdiurnal coupling of the ocean model (Asselin, 1972; Robert, 1966; Williams, 2011); and subsequently, use of a 1-hr coupling frequency to explicitly resolve (sub) diurnal and inertial periods. In addition, the K-Profile vertical mixing Parameterization (KPP; Large et al., 1994) is incorporated via the Community ocean Vertical Mixing (CVMix) framework, and the Caspian Sea is treated as a lake in the land model, and thus, it is no longer included in the ocean model as a marginal sea. The model uses spherical coordinates in the Southern Hemisphere. In the Northern Hemisphere, the grid North Pole is displaced into Greenland. The horizontal resolution is nominal 1° with a uniform resolution of 1.125° in the zonal direction. The resolution varies significantly in the meridional direction with the finest resolution of 0.27° at the equator. In the Southern Hemisphere, the resolution monotonically increases to 0.53° at 32°S, remaining constant further south. In the Northern Hemisphere high latitudes, the finest resolution is about 0.38°, occurring in the northwestern Atlantic Ocean, and the coarsest resolution is about 0.64°, located in the northwestern Pacific Ocean. There are 60 levels in the vertical, with a maximum depth of 5,500 m. The vertical resolution is uniform at 10 m in the upper 160 m and increases monotonically to a maximum of 250 m by a depth of 3,500 m.

Ocean biogeochemistry in CESM2 is represented by the Marine Biogeochemistry Library (MARBL), configured to implement an updated version of what has previously been known as the Biogeochemistry

Elemental Cycle (BEC; Moore et al., 2002, 2004, 2013). MARBL represents multiple nutrient colimitation (N, P, Si, and Fe) and includes three explicit phytoplankton functional groups (diatoms, diazotrophs, and pico/nano phytoplankton) and one implicit group (calcifiers), and there is one zooplankton group (Moore et al., 2004). MARBL includes fully prognostic carbonate chemistry and simulates sinking particulate organic matter implicitly, subject to ballasting by mineral dust, biogenic $CaCO_3$, and Si following Armstrong et al. (2002). Major updates relative to CESM1 include a representation of subgrid-scale variations in light (Long et al., 2015), variable C:P stoichiometry (Galbraith & Martiny, 2015), burial of material at the seafloor that matches riverine inputs in the preindustrial climate, both semilabile and refractory dissolved organic material pools (Letscher et al., 2015), prognostic oceanic emission of ammonia (Paulot et al., 2015), and an explicit ligand tracer that complexes iron. Atmospheric deposition of iron is computed prognostically as a function of dust and black carbon deposition simulated by CAM6. Riverine nutrient, carbon, and alkalinity fluxes are supplied to the ocean from a data set. These fluxes are held constant using data from GlobalNEWS (Mayorga et al., 2010) with the exception of inorganic nitrogen and phosphorus fluxes, which are taken from the Integrated Model to Assess the Global Environment-Global Nutrient Model (IMAGE-GNM) (Beusen et al., 2015, 2016), and vary from 1900 to 2005. Outside this period, the fluxes are held constant using the closest temporal value.

Version 3.14 of the NOAA WaveWatch-III ocean surface wave prediction model (Tolman, 2009) is incorporated in CESM2 as a new component. The wave model is solved on a coarse resolution, near-global grid (78.4°S to 78.4°N) with a resolution of 4° in longitude and 3.25° in latitude with 25 frequency and 24 direction bins, optimized for climate applications of the Stokes drift, called the ww3a grid. The waves are driven by winds, air-sea temperature differences, mixed layer depth, and ocean currents provided through the coupler, and they deliver three new benefits. First, the waves provide the forcing necessary for including Langmuir, or wave-driven, turbulence in the KPP vertical mixing scheme in POP2 (Li et al., 2016). Second, wave statistics from CMIP6 experiments will be contributed to the Coordinated Ocean Wave Climate Project (COWCLIP) experiment (Hemer et al., 2012), which studies changes in wave climate under climate change. The CESM2 contribution is expected to be unique in COWCLIP comparison as the waves will be run in a fully coupled system. Third, wave statistics are now available for development of other wave-related process studies and parameterizations, such as drag and gas transfer coefficients that depend on sea state (Reichl et al., 2014), waves in the marginal ice zones (Roach et al., 2018), wave effects on the mean flow that become important at mesoscale-permitting ocean model resolutions (e.g., McWilliams & Fox-Kemper, 2013; Suzuki & Fox-Kemper, 2016), and other climate effects of waves (Cavaleri et al., 2012). The CESM2 default version of Langmuir mixing includes only enhanced mixing within the mixed layer (Li et al., 2016; McWilliams & Sullivan, 2000), but a scheme with more accurate entrainment at the mixed layer base is also available (Li et al., 2017). When the prognostic wave model is not used, a statistical theory prediction of the needed variables of the wave state based on empirical relationships observed in surface waves can be used instead. This "theory wave" feature in CVMix provides similar Langmuir mixing effects at a much lower cost (Li & Fox-Kemper, 2017; Reichl & Li, 2019), with a minimal loss of accuracy for the Langmuir mixing application.

2.3. Sea Ice

CESM2 uses CICE Version 5.1.2 (CICE5; Hunke et al., 2015) as its sea-ice component. Compared to CESM1, the sea-ice model now incorporates mushy-layer thermodynamics (Turner & Hunke, 2015), with a prognostic vertical profile of ice salinity. To accomplish this, the model parameterizes the gravity drainage of brine, melt water flushing within the ice, and salinity effects of snow-ice formation. The sea-ice vertical resolution has been enhanced from four layers in CESM1 to eight layers in CESM2 to better resolve the salinity and temperature profiles. Also, the snow model now resolves three layers to represent the vertical temperature profile. The melt pond parameterization has been updated (Hunke et al., 2013) and now accounts for the fact that ponds preferentially form on undeformed sea ice. In conjunction with the ocean model, a salinity-dependent freezing temperature is incorporated (Assur, 1958). CICE5 shares the same horizontal grid with the ocean component.

2.4. Land Ice

The land-ice modeling capabilities introduced in CESM1 (Hurrell et al., 2013; Lipscomb et al., 2013) have been extended in CESM2. CESM1 used the Glimmer Community Ice Sheet Model (a serial code with

simplified shallow-ice dynamics) and was limited to one-way coupling; the ice sheet could respond dynamically to surface mass balance (SMB) forcing from the land model, but land surface topography and surface types could not evolve. The ice sheet component of CESM2 is the Community Ice Sheet Model Version 2.1 (CISM2.1; Lipscomb et al., 2019), a state-of-the-art parallel model with higher-order velocity solvers (valid for fast-flowing ice streams and ice shelves) and improved treatments of basal sliding, iceberg calving, and other physical processes. The simulations in this manuscript assume fixed ice sheets, but CESM2 also supports two-way coupling between the Greenland ice sheet and the land and atmosphere models. For two-way coupling, land surface types and surface elevation evolve as the ice sheet advances and retreats, with total water mass conserved. The ice sheet model typically is run on the Greenland domain at 4-km grid resolution with a depth-integrated higher-order solver (Goldberg, 2011), a pseudo-plastic basal sliding law (Aschwanden et al., 2016), and with floating ice immediately calved. CESM2 simulations with dynamic Antarctic and paleo ice sheets are not yet supported but are under development. With either fixed or dynamic ice sheets, the SMB and surface temperature over ice sheets are computed in CLM5 in multiple elevation classes for each glaciated grid cell (Lipscomb et al., 2013). By default, these calculations are done for any CESM2 simulation with an active land model, unlike CESM1 in which ice sheet SMB was computed only in a few custom simulations (e.g., Vizcaíno et al., 2013). For Greenland, the SMB and surface temperature are passed to the coupler and downscaled to the finer CISM grid, with linear interpolation between elevation classes in the vertical direction and bilinear interpolation in the horizontal. For Antarctica, the SMB is computed in multiple elevation classes in CLM5 but is not downscaled by the coupler.

2.5. Land

The Community Land Model Version 5 (CLM5; Lawrence et al., 2019) includes an extensive suite of new and updated processes and parameterizations. Collectively, these developments improve the model's hydrological and ecological realism and enhance the representation of anthropogenic land use activities on climate and the carbon cycle. Specifically, the main updates are as follows: photosynthesis updates (canopy scaling, colimitation (Bonan et al., 2011), and temperature acclimation (Lombardozzi et al., 2015)); soil hydrology improvements (frozen soil hydrology updates (Swenson et al., 2012), spatially explicit soil depth (Brunke et al., 2016), dry surface layer control on soil evaporation (Swenson & Lawrence, 2014), and updated groundwater scheme (Swenson & Lawrence, 2015)); snow model updates (new snow cover fraction parameterization (Swenson & Lawrence, 2012), new fresh snow density parameterization based on temperature and wind resulting in more realistic denser snowpacks, new fresh grain size parameterization based on temperature, revised canopy snow interception, canopy snow radiation, and snow unloading parameterizations, and 10-m snow water equivalent maximum snowpack allowing for firn formation (van Kampenhout et al., 2017)); introduction of a mechanistic plant hydraulics and hydraulic redistribution scheme, which explicitly models water transport through roots, stems, and leaves according to a simple hydraulic framework and which replaces the empirical soil moisture stress function (Kennedy et al., 2019); a completely updated lake model (Subin et al., 2012); vertically resolved soil biogeochemistry scheme that allows for simulation of permafrost carbon (Koven et al., 2013); addition of a global crop model that represents planting, harvest, grain fill, and grain yields for six crop types and includes time-evolving, spatially explicit fertilization rates and irrigated areas (Levis et al., 2018); inclusion of a new fire model that has representation of natural and anthropogenic fire triggers and suppression as well as agricultural, deforestation, and peat fires (F. Li et al., 2013; Li & Lawrence, 2017); multiple urban classes and updated urban energy model (Oleson & Feddema, 2019); and improved representation of plant N dynamics to address known deficiencies in CLM4. This latter improvement is accomplished by (1) introducing flexible plant carbon : nitrogen (C:N) stoichiometry (Ghimire et al., 2016), which circumvents the problematic CLM4 separation of potential and actual productivity fluxes that effectively decoupled leaf-level C exchange from associated water and energy fluxes; (2) explicitly simulating how photosynthetic capacity responds to environmental conditions through the Leaf Utilization of Nitrogen for Assimilation (LUNA) module (Ali et al., 2016); and (3) accounting for how changes in N availability have consequences for plant productivity through the Fixation and Uptake of Nitrogen (FUN) module, which calculates the carbon costs of N acquisition and allocates C expenditure on N uptake among biological fixation, active uptake, and retranslocation (Shi et al., 2016). CLM5 receives instantaneous N fluxes directly through the coupler from WACCM6 when it is used. Otherwise, CLM5 uses monthly N-deposition derived from prior CESM2(WACCM6) simulations. Also, irrigation is applied either with water taken from rivers, if available, or diffusely from the ocean if there is not enough river water. Detailed description, assessment, and

characterization of CLM5 performance compared to prior CLM versions in land-only configurations are provided in Lawrence et al. (2019), Wieder et al. (2019), Fisher et al. (2019), and Bonan et al. (2019).

2.6. River Transport

The River Transport Model (RTM) used in CESM1 has been replaced with the Model for Scale Adaptive River Transport (MOSART; H. Y. Li et al., 2013). In standard CESM2 configurations, the MOSART grid resolution is 0.5° × 0.5°. In each MOSART grid cell, surface runoff received from CLM5 is routed across hillslopes and then discharged along with subsurface runoff, also received from CLM5, into a tributary subnetwork. This tributary subnetwork then feeds into the main river channel of that grid cell that links river water from upstream grid cells to downstream grid cells. A primary difference between RTM and MOSART is how river flow is calculated. RTM uses a simple linear reservoir method wherein the flux of water transferred from an RTM grid cell to its downstream neighbor is dependent only on the river water storage in that grid cell, the mean distance between grid cells, and a globally constant effective water velocity. In MOSART, river flow rates are calculated explicitly via the physically based kinematic wave method, a commonly used method in hydrology that is based on mass and momentum equations. As a consequence, whereas RTM only simulates streamflow (m^3/s), MOSART additionally simulates time-varying main channel flow velocity (m/s) and water depth, as well as subgrid surface water flow in hillslopes and tributaries. MOSART requires detailed information on river hydrography (i.e., parameters describing river and tributary width, depth, average slope, roughness coefficient, and dominant river length), which is derived from the global 1-km HydroSheds data set (Lehner et al., 2008).

2.7. Coupling

CESM2 is accompanied by a new collaborative software infrastructure for building and running the model system as well as for controlling state and flux exchanges among the components. This framework is known as the Common Infrastructure for Modeling the Earth (CIME; http://github.com/ESMCI/cime). CIME contains a significantly rewritten coupling infrastructure that includes modularity, multi-instance capabilities for data assimilation, and new support for land-ice coupling. It also includes rewritten data components that have new Python capabilities and are more easily extensible. As in CESM1, the data components can replace prognostic components, thereby enabling component feedbacks to be easily controlled. Examples of such configurations include ocean - sea-ice coupled integrations in which the atmospheric data sets are provided by a data atmosphere model, and a land-only setup where the land model is the only prognostic component. CIME also provides a new Case Control System (CCS) for configuring, compiling, and executing complex Earth system model experiments. CCS is an object-oriented set of Python utilities that provide many new capabilities for easier portability, case generation, and user customization. It records details of the model version used as well as user customization of the experimental setup. Additionally, CCS has system and unit testing functionality that leads to a more robust and flexible code base. CIME can be ported and tested in a stand-alone mode, without any of the CESM prognostic components, as a first step in porting CESM to other machines. Finally, CIME contains additional tools, such as a new statistical consistency test, that allow users to quickly verify if their port of CESM2 is successful.

The atmosphere, land, wave, and sea-ice components communicate state information and fluxes through the coupler every atmospheric time step of 30 min. The only fluxes calculated in the coupler are those between the atmosphere and ocean, and the coupler communicates them to the ocean component every hour and to the atmosphere every 30 min. The land-ice component communicates with the coupler once a day. The nominal 1° horizontal version of the CESM2(CAM6) has a throughput of about 30 simulated years per day on 4,320 processors on the Cheyenne Supercomputer, corresponding to about 3,450 cpu hours per simulation year. The CESM2(WACCM6) version has a throughput of about 4 simulated years per day on 3,780 processors on the same machine, corresponding to about 22,700 cpu hours per simulation year, roughly seven times more expensive than CESM2(CAM6). In both CESM2(CAM6) and CESM2 (WACCM6) configurations, the atmosphere is scaled to near the maximum allowed by the FV dycore: 1,152 tasks by 3 threads. Although CAM6 and WACCM6 can run on more processors, it is not cost-effective to do so. Thus, the lower processor count in CESM2(WACCM6) simulations is due to the ocean component that uses 256 tasks by 3 threads in CESM2(CAM6) but only 24 tasks by 3 threads in CESM2(WACCM6). This is because the atmospheric model is much slower with WACCM6, so fewer processors are needed for the ocean to keep pace.
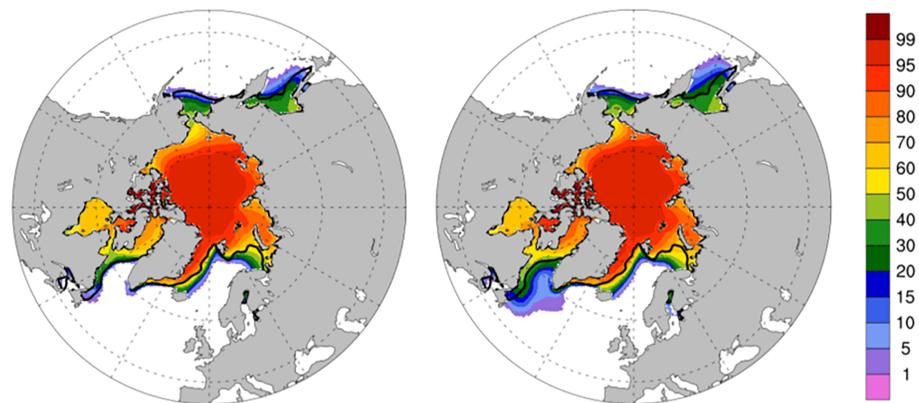
Figure 2. Thirty-year mean Arctic sea-ice concentration distributions (in %) from CESM2(CAM6) PI simulations (left) without and (right) with excessive sea-ice formation in the Labrador Sea region. The black lines are the 15% concentration values from SSM/I observations for the 1981–2005 period (Cavalieri et al., 1996).

3. Development Strategy and Challenges

Evaluations of the component model developments summarized above were done first by the respective developers using both component-only and fully-coupled configurations with various CCSM4 and CESM1 versions. This is in contrast to the process used during the CCSM4 development where initial assessments were done primarily in component-only simulations. Preliminary versions of the component models were then brought together to start the first CESM2 coupled simulations in October 2015. During this development and evaluation process, particular attention was paid to conservation of heat and freshwater across the component models, climate drifts, *reasonableness* of mean climate and of variability in various fields such as El Niño-Southern Oscillation (ENSO) and Atlantic meridional overturning circulation (AMOC) as well as making sure that the model simulations were able to reproduce observed key features of the 20th century climate such as evolution of global-mean surface temperatures, present-day SSTs, and sea-ice distributions and trends. Unfortunately, the path to a fully-coupled CESM2 version that produced acceptable simulations for CMIP6 turned out to be much longer than anticipated and very challenging. Along the way, errors and bugs were discovered and addressed. Progress was particularly delayed due to two major unacceptable features that were not obviously related to bugs or implementation errors. The first concerned the formation of unrealistically extensive sea-ice cover over the Labrador Sea region in PI control simulations. The second was that the global-mean surface temperature time series in historical simulations showed a significant and quite unrealistic cooling particularly during the second half of the 20th century, with end-of-the-century temperatures being actually colder than in the 1850s. Substantial human and computational resources were dedicated to address these major challenges, as well as less serious issues, as described below. After executing over 300 simulations, the vast majority being fully coupled, ranging in duration from several decades to centennial, a near-final CESM2 version was created in April 2018.

The formation of unrealistically extensive sea-ice cover over the Labrador Sea region, extending into the northern North Atlantic, in PI control simulations proved to be very challenging to address robustly. In fact, this feature—not supported by available observations—emerged twice during the development process. Figure 2 shows 30-year average sea-ice concentration distributions from two CESM2(CAM6) PI simulations with and without such an extensive Labrador Sea ice cover. The more extensive Labrador Sea ice extent is also accompanied by enhanced (and more unrealistic) sea ice in the Sea of Okhotsk region. In simulations with extensive ice cover, this feature emerges as early as around year 40 and as late as around year 105 relative to initialization of the coupled simulation. In no simulation did the Labrador Sea ice cover become unrealistically extensive after year 105 if it had not already emerged earlier in the simulation, indicating that the model likely has multiple equilibria and the trajectory through phase space following initialization passes near a bifurcation point. It was found that extensive ice cover persisted when the simulations were extended by 50 or more years further. It was not feasible, given time constraints, to undertake longer integrations to investigate long-term behavior of this feature. Simulation characteristics when extensive ice cover is

present include annual-mean sea-ice concentrations that are usually around 60% in the Labrador Sea; Labrador Sea SSTs that remain around freezing point with salinities <34 psu; curtailed Labrador Sea deep water formation/convection; compensating enhanced deep-water formation to the east of Greenland and in the Irminger Sea; and a slightly weaker AMOC maintained by the switch in the location of deep-water formation. Detailed analyses of many simulations with extensive ice cover in comparison to simulations without this feature, unfortunately, did not identify any robust mechanisms involving a particular process or phenomenon such as surface fluxes, liquid and solid runoff, lateral advection, vertical mixing, and ice drag as instigators or drivers. Indeed, when ensemble simulations were performed using the same model configuration but with only differences in their initial atmospheric potential temperatures introduced by round-off level perturbations, some members showed extensive ice cover, and others did not. Given this experience, a practical—though not fully satisfactory—approach was adopted to use an ensemble member without an extensive Labrador Sea ice cover after a sufficiently long simulation (see section 4) to provide ocean and sea-ice initial conditions going forward.

The second unacceptable feature was an unrealistic cooling of global-mean surface temperatures in the historical simulations (see Figure 7, green line). The cooling started around the 1920s, mildly at first and then getting worse during roughly the 1950–1980 period, with surface temperature anomalies exceeding −0.4 °C with respect to the 1850s. Although the model warmed after the 1980s, the end-of-the-century temperatures remained unrealistically colder than in the preindustrial 1850s. This behavior first appeared when the new CMIP6-based emissions data sets were applied in our simulations. The same model versions with CMIP5-based emissions did not cool in this way. The observational time series (see Figure 7, black and gray lines) show decadal fluctuations that have significant contributions from internal variability. The cooling described above clearly exceeds this baseline internal variability range, and this is what is addressed here. The goal is not to reproduce exactly the 20th century global-mean surface temperature time series in each realization but to eliminate the unrealistic cooling behavior. Initial investigations uncovered some inconsistencies and errors in the CMIP6 emissions data sets (which were resolved in their final release) and their application to CAM6, particularly for anthropogenic sulfur. Simulations with the corrected data sets did show some reductions of the unrealistic cooling, but the cooling behavior remained, particularly during the mid-1940s to 1980 period, with end-of-the-century temperatures only slightly warmer than in the 1850s (not shown). Analysis of CAM6 against recent volcanic sulfur emission events indicated a potentially excessive response (Malavelle et al., 2017), which could be traced to the impact of aerosols on the liquid water path. Several modifications were, therefore, made to aerosol—cloud interaction processes, focusing on the warm rain formation process (autoconversion and accretion) to reduce the magnitude of these effects. As a result of these modifications, the global historical (2000 minus 1850) aerosol effective radiative forcing changed from −1.82 W/m^2 in CAM5 with CMIP5-based emissions to −1.67 W/m^2 in CAM6 with CMIP6-based emissions (see Gettelman, Hannay, et al., 2019, for further details). These changes were instrumental in enabling the model to simulate a realistic 20th century surface temperature time series as discussed below (see section 7), but they did not target a particular equilibrium climate sensitivity value (see section 12).

4. Initial Conditions

The ocean and sea-ice initial conditions used in both CESM2(CAM6) and CESM2(WACCM6) PI control simulations were obtained through the following series of integrations (Figure 3). First, a three-member ensemble set of simulations was started from the CESM1(LENS) PI control integration (Kay et al., 2015) at Year 402 (all simulations start from 01 January of a given year), which itself was initialized from the January-mean potential temperature and salinity from the Polar Science Center Hydrographic Climatology (PHC2; Steele et al., 2001; Levitus et al., 1998). Two ensemble members developed excessive Labrador Sea ice cover, starting at about Year 80. The third member (Simulation 262c) did not show such a behavior. Therefore, it was decided to use this simulation to provide ocean and sea-ice states for initializing subsequent simulations. Indeed, the simulation was later extended to >350 years to confirm that the Labrador Sea ice cover remained realistic. A series of simulations with minor adjustments and bug fixes in the coupled model were then performed, starting with a simulation (293) from Year 161 of Simulation 262c. This simulation was followed by another (Simulation 297) that started from Year 130 of Simulation 293. Finally, a third simulation (299) was initialized from Year 249 of Simulation 297. The ocean and sea-ice states from year 134 of the final simulation (299) were used as the initial conditions for the main
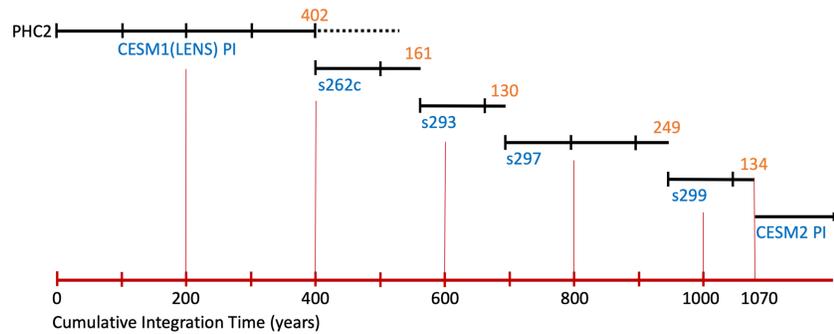
Figure 3. Schematic of the sequence of simulations used to obtain ocean and sea-ice initial conditions for the CESM2 PI control simulations. The simulation names are given in blue. The orange numbers represent the year of a simulation from which a subsequent simulation started. The black horizontal lines represent the length of a given simulation in years. The red line shows the cumulative integration time in years. The black and red tick marks indicate 100-year intervals. CESM1(LENS) PI control simulation started from the PHC2 observations.

CESM2(CAM6) and CESM2(WACCM6) PI control simulations. Thus, in total, these ocean potential temperature and salinity initial conditions represent states integrated for about 1,070 years after initialization starting from the PHC2 data sets.

The runoff initial conditions also come from Year 134 of Simulation 299 for both CESM2(CAM6) and CESM2(WACCM6) PI controls. Although their origins can be traced back to Year 161 of Simulation 262c, their in-between paths differ from that of ocean and sea-ice initial conditions. The atmospheric initial conditions come from Year 134 of Simulation 299 for CESM2(CAM6) PI control and Year 11 of a preliminary CESM2(WACCM6) PI simulation (Simulation 298) for the main CESM2(WACCM6) PI control.

Initial conditions for the land states and ocean biogeochemical tracers, including abiotic radiocarbon, were obtained with long spin-up runs of the land and ocean models, respectively. In these spin-up runs, the active atmospheric component was replaced with a data-atmosphere component that simply read in and repeatedly cycled through multiple years of surface forcing data sets that were needed to construct surface fluxes. Twenty-one years of forcing data sets extracted from Simulation 297 were used, in order to capture some aspects of interannual variability. The land model was spun up from bare ground, utilizing the accelerated decomposition mode to equilibrate the slow carbon pools, followed by several hundred years of normal mode spin-up (Lawrence et al., 2019). The ocean tracer spin-up was applied only to the biogeochemical tracers of the ocean model. To avoid drift of the ocean physical state, it was reset to its initial condition at the beginning of each 21-year cycle, keeping it synchronized with the surface forcing. A subset of the ocean biogeochemical tracers was spun up using a Newton-Krylov based solver (Lindsay, 2017).

The initial conditions for the 11 members of the CESM2(CAM6) historical simulations come from Years 501, 601, 631, 661, 691, 721, 751, 781, 811, 841, and 871 of the corresponding PI control simulation (Figure 4). The three members of the CESM2(WACCM6) historical simulations are initialized from Years 56, 61, and 71 of the corresponding PI control integration. The late and early start dates for the CESM2(CAM6) and CESM2 (WACCM6) historical simulations simply reflect their respective PI control integration lengths and are not intended to avoid or sample any particular variability characteristics.

5. Forcing Data Sets for Preindustrial and Historical Simulations

CESM2(CAM6) PI control and historical simulations used several forcing data sets that were obtained from the corresponding CESM2(WACCM6) simulations. This is a new approach in CESM2 adopted to ensure that a physically consistent pair of low- and high-top atmospheric models that share the same code base and contain the same physics in their common range of altitude is used. It represents a degree of compatibility and consistency not available to most modeling groups. CESM2(WACCM6) uses CMIP6-provided forcings wherever appropriate but calculates itself chemical and aerosol constituents from emissions that are based on the anthropogenic and biomass burning inventories specified by CMIP6. Anthropogenic emissions for 1850 to 2014 are provided from the Community Emissions Data System (CEDS; Hoesly et al., 2018). Biomass
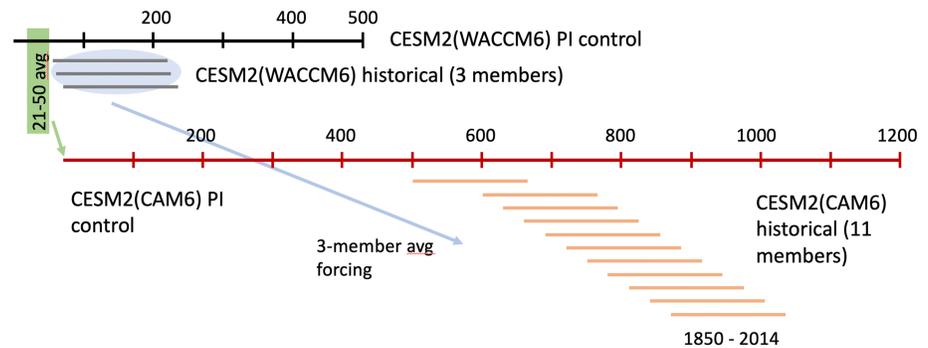
Figure 4. Schematic of the CESM2 PI control and historical simulations. CESM2(WACCM6) PI simulation was started first (black) to obtain the forcing data sets used for the CESM2(CAM6) PI integration (red). These forcings were based on the average of Years 21–50 (green shading). Three historical CESM2(WACCM6) simulations were started from Years 56, 61, and 71 of the corresponding PI simulation (gray lines). As described in section 5, three-member ensemble averages of these WACCM6 simulations (blue shading) were then used to obtain the forcings needed for the CESM2 (CAM6) historical simulations (orange lines). The CESM2(CAM6) historical simulations started from Years 501, 601, 631, 661, 691, 721, 751, 781, 811, 841, and 871 of the corresponding PI simulation. The black and red tick marks indicate 100-year intervals.

burning emissions are described by van Marle et al. (2017) and are all specified at the surface (without any plume-rise or specified vertical distribution of the emissions). Further details of the emissions and chemistry used in CESM2(WACCM6), including the specification of ozone-depleting substances, are described in Gettelman, Mills, et al. (2019). CESM2(CAM6) simulations use the CMIP6-provided forcings except for (i) tropospheric oxidants for chemistry; (ii) stratospheric ozone for radiative effects; (iii) stratospheric aerosols for radiative effects; (iv) H_2O production rates due to CH_4 oxidation in the stratosphere; and (v) N deposition to the land and ocean components. These WACCM6-based forcings are described below.

First, CESM2(CAM6) PI simulation 299 was run with prescribed forcings from a CESM2(WACCM6) simulation (295). Then the CESM2(WACCM6) PI control was conducted, using initial states for the ocean, sea-ice, and land components from Simulation 299 (see section 4). Distributions of tropospheric oxidants, stratospheric ozone, stratospheric aerosol, and methane oxidation rates were derived from the average of Years 21–50 of this CESM2(WACCM6) PI simulation and used in the subsequent CESM2(CAM6) PI control. For historical simulations, an ensemble of three CESM2(WACCM6) simulations was run first for the 1850–2015 period. Corresponding forcings were derived from the CESM2(WACCM6) ensemble average, for use in the CESM2(CAM6) historical simulations. The averaging of 30 years in the PI control and of three ensemble members in the historical simulations is intended to reduce the influence of natural variability on these forcings. A schematic summary of our integration strategy to obtain the necessary forcing data sets to perform the CESM2(CAM6) simulations is provided in Figure 4.

The tropospheric oxidant species (O_3, OH, NO_3, and HO_2) are prescribed in CESM2(CAM6), which uses them in gas-phase and aerosol chemistry calculations to allow tropospheric aerosol formation. Monthly averaged values are used in both the PI control and historical simulations. For the CESM2(CAM6) PI control, a repeating 12-month annual cycle is derived from averaged values over Years 21–50 of the CESM2 (WACCM6) PI control simulation. CESM2(CAM6) interpolates cyclically between midmonth date values. For CESM2(CAM6) historical simulations, 12-month annual cycles are provided at quasi-decadal frequency (see Appendix A). CESM2(CAM6) interpolates annual cycles for years between those provided.

Stratospheric ozone and stratospheric aerosol are radiatively active species important to the radiative budget, temperature, general circulation, chemistry, and cloud physics of CESM2. In CESM2(CAM6), stratospheric distributions of these species are prescribed from 5-day zonal means of CESM2(WACCM6) output. Ozone volume mixing ratio is prescribed for the entire vertical extent of the model for use in radiative transfer calculations.

A major advance in CESM2(WACCM6) is the development of prognostic stratospheric aerosols derived from gas-phase sulfur emissions. As described in Mills et al. (2016), this new feature provides CESM2 with stratospheric aerosol forcing that is much more consistent with observations than previous climatologies
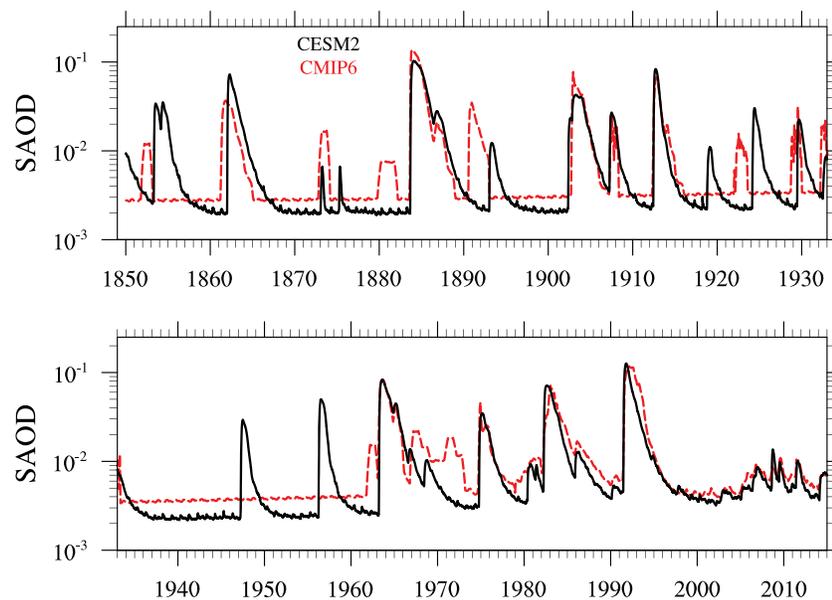
Figure 5. Time evolution of the globally averaged stratospheric aerosol optical depth (SAOD) at 550 nm as simulated by CESM2(WACCM6) (three-member average) and prescribed in CESM2(CAM6) (black) in comparison to that provided by CMIP6 (red). There is uncertainty in the time, location, and intensity of volcanic emissions, especially further back in time.

developed for climate models. This self-consistent treatment of volcanic aerosols and chemistry has been shown to produce realistic radiative responses to eruptions over the 1979–2015 period (Mills et al., 2017; Schmidt et al., 2018). For CMIP6, volcanic SO_2 emissions for 1850–2015 from VolcanEESM database (Neely & Schmidt, 2016) are used. In addition, observed OCS concentrations are specified as time-varying boundary conditions. As shown in Figure 5, this leads to a realistic simulation of the stratospheric aerosol optical depth (SAOD) compared to the CMIP6 data that are constrained by satellite observations starting in 1979. Before this date, when the information on volcanic emissions is sparser, CESM2(WACCM6) and CMIP6 SAOD can differ; for example, the eruptions of Hekla (1947) and Bezymianny (1956) are missing from the CMIP6 database. Prior to the satellite era, there is also a very small (~0.001) difference between CESM2(WACCM6) and CMIP6 data in the background component of SAOD at 550 nm that persists during volcanically quiescent periods. Historical variability in this background is largely due to variability in the source gas OCS, which WACCM6 accounts for as a time-varying lower boundary condition for OCS (Montzka et al., 2004). CMIP6 prescribes a time-varying background based on the 1999–2005 average, scaled following the calculations of Sheng et al. (2015).

Although CESM2(CAM6) does not include the interactive chemistry needed to realistically simulate the development of stratospheric aerosol following a major volcanic eruption, model consistency is maintained by prescribing aerosol properties above the tropopause in CESM2(CAM6) using 5-day zonally averaged output from CESM2(WACCM6). Prescribed stratospheric aerosol properties are the sulfate mass mixing ratio and diameter of each of the three sulfate-bearing modes of MAM4, aerosol surface area density, and ice condensation nuclei number concentration. Each of these properties is prescribed consistently with its use in CESM2(WACCM6), which includes prognostic stratospheric aerosol in MAM4. As a result, the radiative impacts of stratospheric aerosol from large volcanic eruptions are consistent between CESM2(CAM6) and CESM2(WACCM6) simulations. For the CESM2(CAM6) PI control, a 5-day zonal-mean annual cycle was created from the average of Years 21–50 of the CESM2(WACCM6) PI control. For the CESM2(CAM6) historical simulations, 5-day zonal-mean values were calculated for each year over the 1850–2014 historical period from the average of the three CESM2(WACCM6) historical ensemble members. The 5-day frequency allows resolution of the impacts of major volcanic eruptions, as well as the annual development of polar ozone loss (Neely et al., 2014). The averaging of three CESM2(WACCM6) historical ensemble members is especially important to the evolution of stratospheric aerosol resulting from volcanic eruptions, which can be particularly sensitive to the circulation present at the time of the eruption.

The oxidation of methane (CH_4) is an important source of water vapor to the middle atmosphere. Because water vapor is a greenhouse gas, production rates of water vapor in CESM2(CAM6) are prescribed using methane oxidation rates from CESM2(WACCM6). It is assumed that the oxidation of each CH_4 molecule results in the production of two molecules of H_2O (Nicolet, 1970). For the CESM2(CAM6) PI control, a repeating 12-month annual cycle is derived from averaged values over Years 21–50 of the CESM2 (WACCM6) PI control simulation. For CESM2(CAM6) historical simulations, 12-month annual cycles are provided at the same quasi-decadal frequency as the tropospheric oxidants (see above and Appendix A).

Regarding Chlorofluorocarbons (CFCs), in CESM2(CAM6), they are specified as a global average mixing ratio derived from the lower boundary conditions provided by CMIP6. The radiative transfer module in both CESM2(CAM6) and CESM2(WACCM6) considers only CFC-12 and CFC-11eq (equivalent). The CFC-11eq time series is supplied by CMIP6 (Meinshausen et al., 2017). Because CESM2(WACCM6) includes most of the chemical species contained in CFC-11eq, the global distribution of CFC-11eq is first derived interactively in WACCM6. The derived global average distribution at the surface is then scaled to match a given CMIP6 scenario that contains additional species not included in the WACCM6 chemistry, that is, CFC-11eq = CFC-11eq [CMIP6]/CFC-11eq [WACCM6]).

Updated land cover and land use data have been created for CLM5. The new data set combines updated versions of present-day satellite land cover descriptions with the past and future transient land use time series from the Land Use Harmonization Version 2 product (LUH2, http://luh.umd.edu/). CLM5 (and ocean biogeochemistry) uses monthly N-deposition derived from CESM2(WACCM6) simulations. Additional information on this data set is provided in Lawrence et al. (2019) and references therein.

## 6. Preindustrial Control Spin-Up and Tuning

The CESM2(CAM6) and CESM2(WACCM6) PI control simulations were integrated for 1,200 and 500 years, respectively. In such long control simulations, it is necessary to aim for a top-of-the-atmosphere (TOA) global- and time-mean energy balance of $<|0.1|W/m^2$ to avoid unrealistically warm or cold climate states toward the end of these simulations that will be subsequently used to initialize historical integrations. Such a small TOA energy balance is indeed desired in CESM2, because there are no additional external sources or sinks of energy, such as geothermal heat fluxes in the ocean bottom. To achieve this level of TOA energy balance, the coupled model was tuned using two sets of adjustments. The first involved very minor changes in the so-called "gamma_coef" parameter in CLUBB, which controls how bimodal the PDF of subgrid vertical velocity is. A larger value of gamma_coef implies a less bimodal, longer-tailed PDF for vertical velocity. The longer tails of velocity, in turn, lead to stronger entrainment at the top of the planetary boundary layer, which in turn reduces the amount and thickness of low-clouds, particularly in marine stratocumulus layers. The second set of adjustments were done in CICE5. Because the sea ice was deemed to be too thick in the PI simulations—in comparison to the present-day observations—to produce a realistic sea-ice distribution at present day, the bulk dry snow grain radius standard deviation was reduced to 1.25 from its CESM1 value of 1.75. In addition, the temperature for the onset of snowmelt was lowered from −1.0 to −1.5 °C, and the maximum melting snow grain radius size was increased from 1,000 to 1,500 microns, both adjustments again with respect to those used in CESM1 simulations. All these adjustments were only applied to snow on sea ice, and they are certainly within the observationally based acceptable ranges of these tunable sea-ice parameters. The changes collectively result in a lower snow and sea-ice albedo and hence thinner sea-ice overall. These parameters were adjusted identically in the two PI control simulations, and no additional changes were made throughout the control integrations. Furthermore, the tuning parameters remained fixed at these values in all subsequent experiments.

The impacts of these tunings were assessed in decadal to multidecadal length simulations prior to the start of the main PI control simulations. Figure 6 (left panel) shows the time series of the global-mean TOA energy imbalances (fluxes) for the two control simulations where the imbalances are +0.05 and +0.06 W/m^2 for CESM2(CAM6) and CESM2(WACCM6), respectively, averaged over their integration lengths. These imbalances are +0.03 W/m^2 during the last 500 years of CESM2(CAM6) and +0.06 W/m^2 during the last 200 years of CESM2(WACCM6). While these TOA imbalances are much smaller than the imbalance of about −0.15 W/m^2 in CCSM4 (Gent et al., 2011), they are larger than the exceptionally small TOA imbalance of near-zero in the CESM1(LENS) control simulation. In both CESM2(CAM6) and CESM2(WACCM6) control
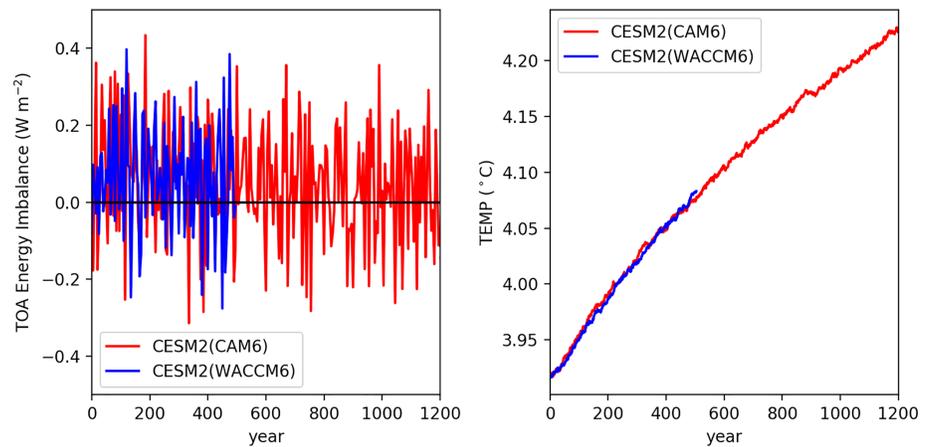
Figure 6. Time series of (left) 5-year average, global-mean TOA energy imbalance (flux) and (right) annual- and volume-mean ocean potential temperature from the CESM2(CAM6) and CESM2(WACCM6) PI control simulations.

simulations, this energy gain is solely reflected in the ocean model, as the land and sea-ice components show minor energy losses. The global-mean energy gain by the ocean model is about 0.07 W/m^2 during the last 500 years of CESM2(CAM6) and about 0.10 W/m^2 during the last 200 years of CESM2(WACCM6). As a result, the global-mean ocean potential temperature (Figure 6, right panel) increases from its initial value of 3.92 to about 4.23 °C by Year 1,200 in CESM2(CAM6). Similarly, it increases from 3.92 to about 4.08 °C in CESM2(WACCM6). Although the details of the depth ranges and magnitudes of this warming differ across the ocean basins, depths below about 2 km show warming in all major basins (not shown). The global-mean ocean salinity decreases slightly by about 5×10^{-5} psu per century in both CESM2(CAM6) and CESM (WACCM6). This salinity trend is very similar to those of CCSM4 and CESM1(LENS) PI control simulations.
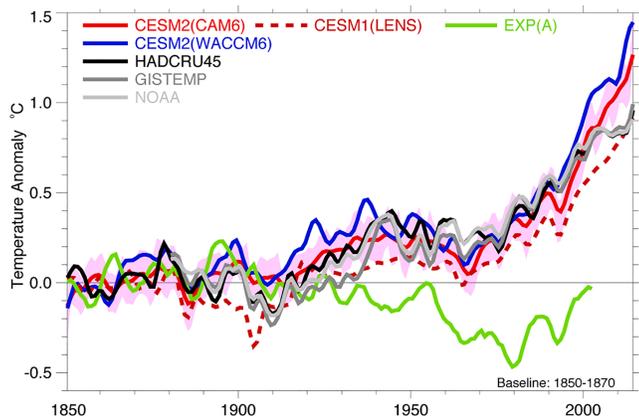
Figure 7. Time series of the global-mean surface (2-m air) temperature anomaly relative to the respective 1850–1870 averages during the historical period from observations, CESM2(CAM6), CESM2(WACCM6), and CESM1(LENS) ensemble simulations. The solid red line and the pink shading show the 11-member ensemble-mean and its spread for CESM2 (CAM6) simulations. The solid blue and dashed red lines show the 3- and 40-member ensemble means for CESM2(WACCM6) and CESM1(LENS) simulations, respectively. The green line, EXP(A), shows a sample time series from an earlier simulation with unrealistic cooling particularly during the second half of the 20th century. The CESM1 simulations use the Representative Concentration Pathway 8.5 (RCP8.5) forcing for the 2006–2014 period. The observations are from the HADCRU45, GISTEMP, and NOAAGlobalTemp data sets. The latter two time series are referenced to the mean of HADCRU45 for the 1870–1890 period.
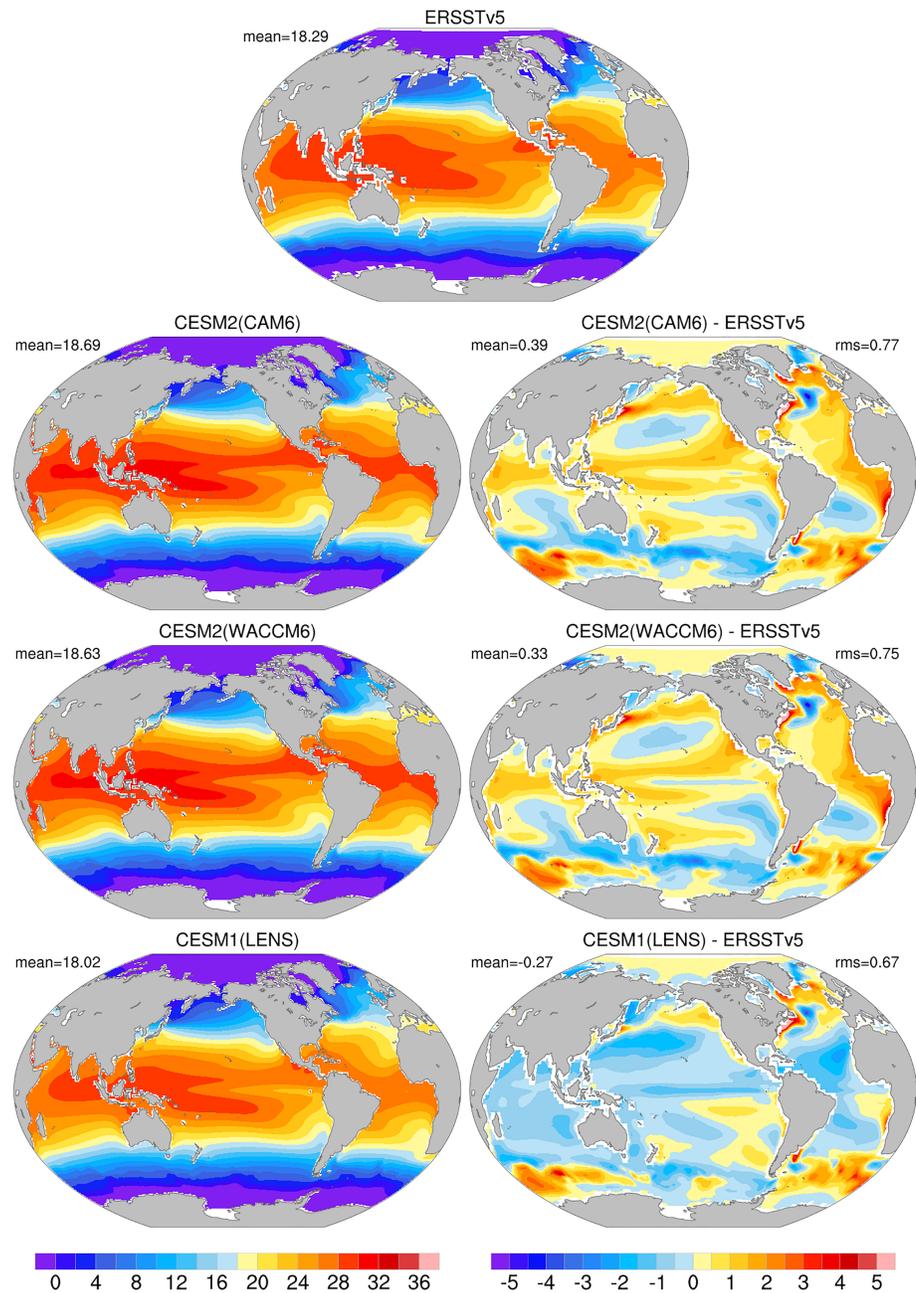
7. Surface Temperatures

The model global-mean surface (2-m air) temperature anomaly time series for the historical period are presented in Figure 7 in comparison with observations. A reference period of 1850–1870 is chosen to easily identify warming since the preindustrial conditions. For CESM2 (CAM6), both the ensemble-mean and its spread are shown. Only the ensemble-mean time series are included from CESM2(WACCM6) and CESM1(LENS) simulations for clarity. The observations are from the Hadley Centre-Climate Research Unit Temperature Anomalies (HADCRU4.5; Jones et al., 2012); the Goddard Institute for Space Studies Surface Temperature Analysis (GISTEMP; Lenssen et al., 2019); and the National Oceanic and Atmospheric Administration merged land-ocean global surface temperature analysis (NOAAGlobalTemp; Vose et al., 2012). In general, the CESM2 (WACCM6) time series tend to be slightly warmer than those of CESM2(CAM6). Both sets of simulations capture the observed low-frequency variability, and the observations are usually within the ensemble spread of CESM2(CAM6) simulations with a few notable exceptions. One such example is the dip around 1910 evident in observations, but absent in model simulations. This discrepancy may be related to incorrect and/or missing representations of impacts of volcanic eruptions in both models and observations during the 1900–1910 period. Another major discrepancy is seen during the 2000–2014 period with model ensemble-mean temperature anomalies being warmer than observed by about 0.2 to 0.4 °C at the end of the integration period in

CESM2 simulations. Although the warming trends between 1975 and 2000 are comparable between these two model simulations and observations, no CESM2 ensemble member is able to capture the observed slowdown in the warming trend during roughly 2000–2012. This is not surprising as internal variability as well as anthropogenic and natural forcings are known to play a role in the creation of this decadal-scale slowdown (e.g., Deser et al., 2017; Kaufmann et al., 2011; Kosaka & Xie, 2013; Meehl et al., 2014; Solomon et al., 2010). Indeed, as discussed in Meehl et al. (2014), a much larger ensemble than presently used is needed as such a slowdown is present only in a very small fraction of the available CMIP5 simulations (10 out of 262). In comparison to CESM2 simulations and observations, CESM1 (LENS) ensemble-mean time series show generally colder temperature anomalies, usually remaining below or near the lower extent of the CESM2(CAM6) spread. CESM1(LENS) does not capture the observed decadal-scale slowdown, either.

The 1985–2014 average SST (0- to 10-m average in the ocean) distributions from CESM2(CAM6), CESM2 (WACCM6), and CESM1(LENS) ensemble means are presented in Figure 8 in comparison to the Extended Reconstructed Sea Surface Temperature Version 5 data set (ERSSTv5; Huang et al., 2017). There are two salient features in the figure, particularly evident in the model-observations difference distributions. First, CESM2(CAM6) and CESM2(WACCM6) mean SSTs are almost identical to each other. Second, both model SSTs are similarly warmer than observations and those of CESM1(LENS). Indeed, CESM2(CAM6) and CESM2(WACCM6) global-mean SSTs are warmer than observations by 0.39 and 0.33 °C, respectively, while CESM1(LENS) is colder than observations by 0.27 °C. These global-mean differences largely reflect the SST differences in the corresponding PI control simulations from the ERSSTv5 data set for the preindustrial period, which are computed as 0.33 °C for CESM2(CAM6), 0.18 °C for CESM2(WACCM6), and −0.18 ° C for CESM1(LENS) over 30-year periods close to the mean initial start dates of the respective historical ensemble members. When root-mean-square (rms) differences from observations are considered, both CESM2(CAM6) and CESM2(WACCM6) have slight degradations of the mean SST simulations with rms differences of 0.77 and 0.75 °C, respectively, compared to 0.67 °C for CESM1(LENS). The SST difference distributions depict several persistent biases. The warm-cold SST difference dipole in the North Atlantic associated with errors in the separation and path of the Gulf Stream-North Atlantic Current system is present in all model simulations with comparable magnitudes. The warm biases in the Eastern boundary upwelling regions, that is, off the coasts of California, South America, and South Africa, are more pronounced in CESM2(CAM6) and CESM2(WACCM6) than in CESM1(LENS). These CESM2 biases are much larger in their magnitude (and spatial extent) than the global- and time-mean SST differences between CESM2 and CESM1(LENS), so the larger CESM2 biases cannot be solely attributed to the warmer mean state of these simulations. Reducing these upwelling-related biases has been rather challenging in 1° horizontal resolution models. In CESM1, the upwelling biases were reduced due to more realistic wind stress curl in a model configuration with much higher horizontal resolutions in the ocean and atmospheric models than used in the present simulations (Small et al., 2015).

Seasonal-average surface (2-m) air temperature model minus observations differences over land for December-January-February (DJF) and June-July-August (JJA) for the 1985–2014 period from CESM2 (CAM6), CESM2(WACCM6), and CESM1(LENS) are shown in Figure 9. The observations are from the Berkeley Earth Surface Temperatures data set (BEST; Rohde et al., 2013). As in the SST distributions, CESM2(CAM6) and CESM2(WACCM6) distributions are very similar to each other, and, on average, CESM2 is warmer than CESM1. Both the global mean bias and rms error are appreciably lower in CESM2 than in CESM1 in all seasons. For example, the DJF mean bias (rms error) has been reduced from −2.24 ° C (1.90 °C) in CESM1(LENS) to 0.89 °C (1.51 °C) and 0.67 °C (1.38 °C) in CESM2(CAM6) and CESM2 (WACCM6), respectively. In CESM2, there is a notably damped annual cycle—warm biases in the winter season and cold biases in the summer season—in the northern high latitudes. There is also a large summertime warm bias over the central United States, which partially coincides with a dry bias (see Figure 10), and may be related to the absence of propagating mesoscale convective systems at the 1° model resolution. Another interesting bias is the warm bias over Northeast India, which exists in all seasons except for March-April-May (MAM) (not shown). This region is among the most heavily irrigated regions in the world and previous studies have shown that the irrigation-induced cooling is likely mitigating against climate warming in this area (Thiery et al., 2020). Though CESM2 includes active irrigation by default, irrigation in India is applied mainly in the MAM season due to a limitation in the crop planting date scheme that
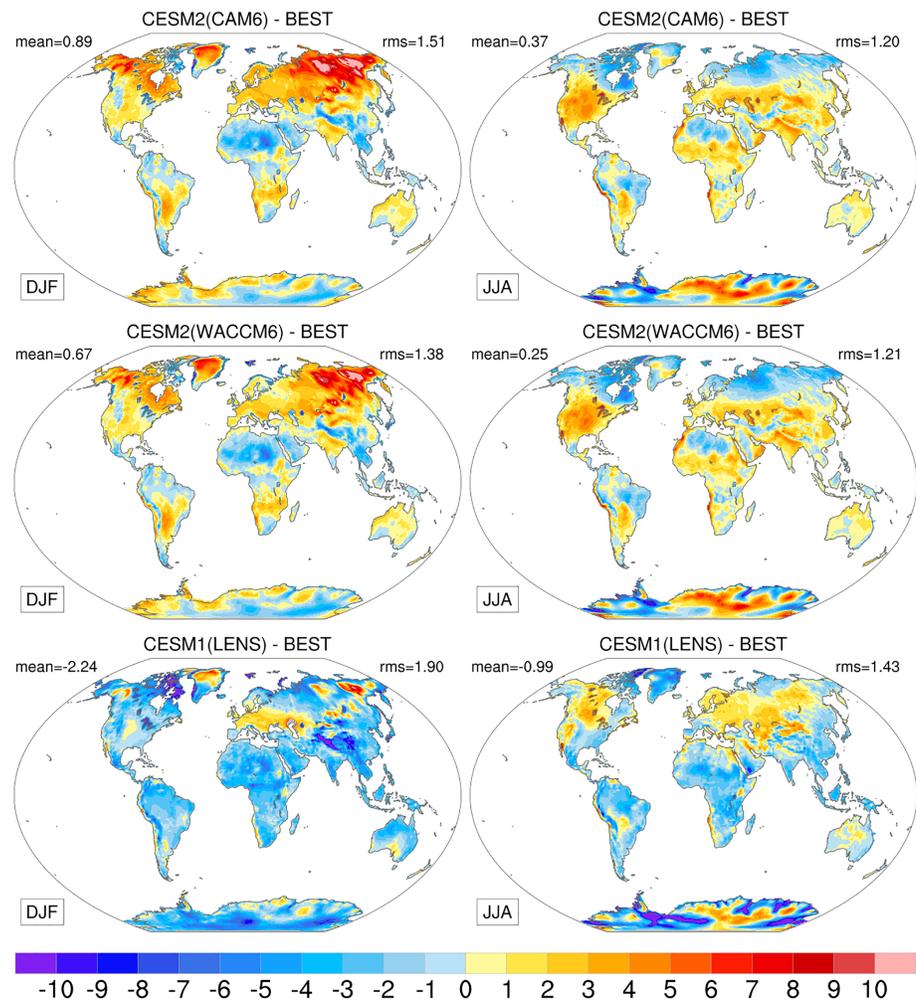
Figure 8. The 1985–2014 average sea surface temperature (in °C) from (top) ERSST observations and (left) model simula-tions. The right panels show the respective model minus observations difference distributions. The model distributions represent full ensemble means. The global means for each panel and the root-mean-square error for the difference dis-tributions are also provided.

incorrectly results in crops being planted in the spring over India. In reality, over much of Northern India, the two main cropping and irrigation seasons are in summer and in winter. Improved simulation of crop planting dates, therefore, may result in reduced surface air temperature biases in this region.

8. Atmospheric Fields

8.1. Precipitation

The numerous improvements incorporated into the atmospheric model components have a positive impact on many aspects of the climate of the fully-coupled CESM2 system. Figure 10 shows the 1985–2014 average

Figure 9. Seasonal-average surface (2-m) air temperature model minus observations differences (in °C) over land for DJF and JJA for the 1985–2014 period from CESM2(CAM6), CESM2(WACCM6), and CESM1(LENS). The observations are from the BEST data set. The model distributions represent full ensemble means. The global-mean biases and the root-mean-square errors are also provided.

precipitation from CESM2(CAM6), CESM2(WACCM6), and CESM1(LENS) ensemble means in comparison to observational estimates from the Global Precipitation Climatology Project (GPCP, Huffman et al., 2009). As in the SST and surface temperature over land distributions, CESM2(CAM6) and CESM2(WACCM6) simulations produce nearly identical mean precipitation fields. Both sets of simulations clearly depict a systematic reduction of most biases from CESM1(LENS) to CESM2, from a mean bias of 0.33 mm/day in CESM1(LENS) down to 0.25 and 0.23 mm/day in CESM2(CAM6) and CESM2(WACCM6), respectively. The rms error also decreases substantially by about 20%, from 0.87 mm/day in CESM1(LENS) to 0.71 mm/day in CESM2. These improvements arise primarily from the reduction of the major tropical wet biases of the Indian Ocean, the South Pacific Convergence Zone (SPCZ), and the tropical Atlantic Inter-Tropical Convergence Zone (ITCZ). Over land, the dipole of Andean wet and Amazon dry biases is decreased, and excessive precipitation over Australia and Western United States is reduced. Nevertheless, there are a few degraded regions, including a slightly larger negative bias off Brazil in CESM2 that extends into the South Atlantic Ocean.
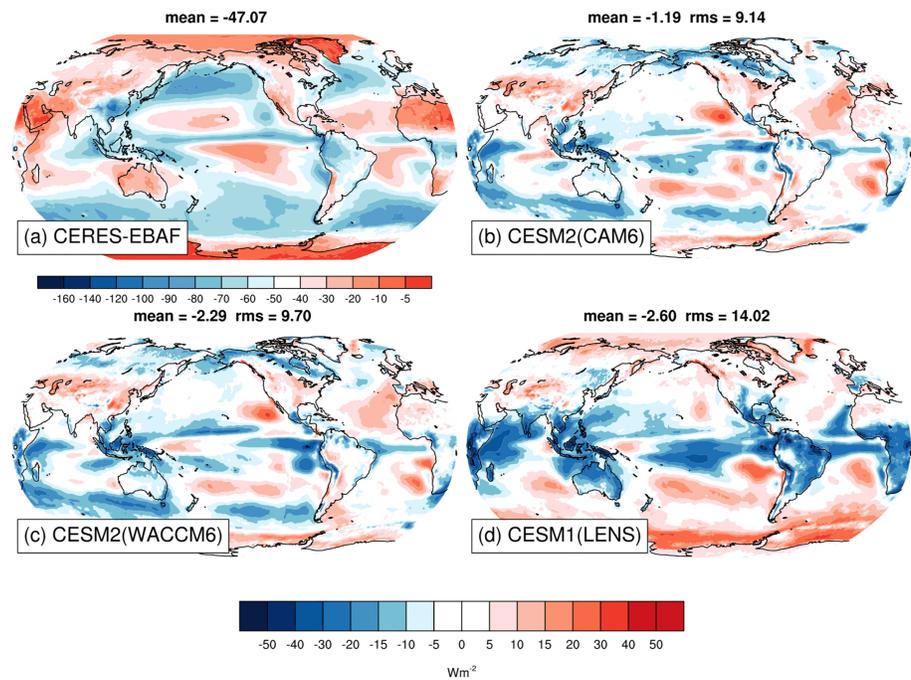
8.2. Shortwave and Longwave Cloud Forcing

A dramatic improvement is seen in the radiative properties and, in particular, the time-averaged shortwave cloud forcing (Figure 11). In comparison to the Clouds and the Earth's Radiant Energy System-Energy
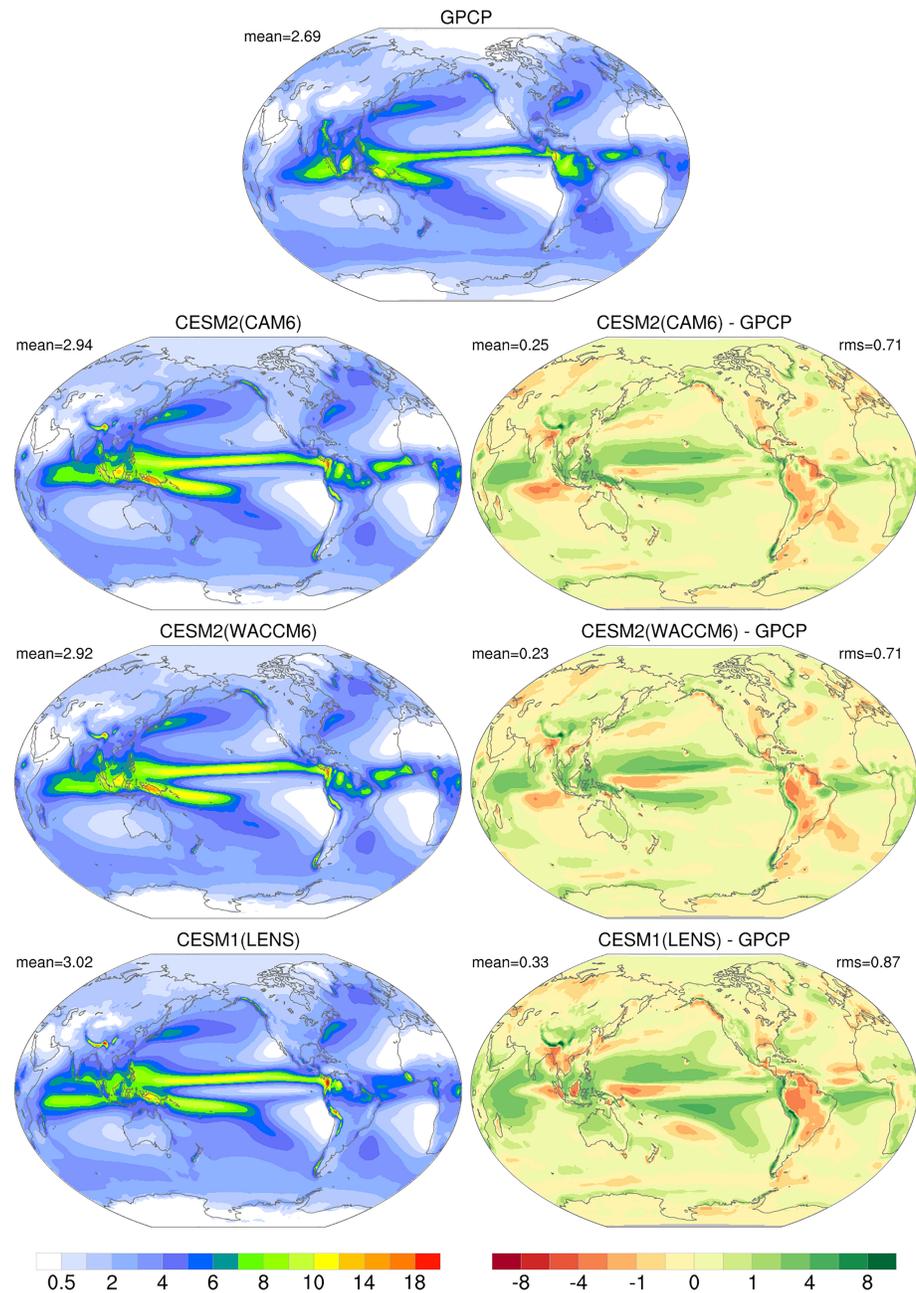
Figure 10. The 1985–2014 average total precipitation (in mm/day) from (top) GPCP observations and (left) model simulations. The right panels show the respective model minus observations difference distributions. The model distributions represent full ensemble means. The global means for each panel and the root-mean-square error for the difference distributions are also provided.

Balanced and Filled (CERES-EBAF; Loeb et al., 2009) data set, the reflective biases of the clouds in CESM1 (LENS) throughout the tropics over land and ocean, often in excess of $-30 \, \text{W/m}^2$, are substantially alleviated in both CESM2(CAM6) and CESM2(WACCM6). This improvement is due to several changes in different cloud regimes. The interaction of CLUBB with the CAM6 deep convective parameterization enabled a readjustment of tropical deep convection over land to reduce biases. Net rms for shallow clouds is reduced with CLUBB. At higher latitudes, low clouds were insufficiently bright in CESM1(LENS). Changes in mixed-phase cloud properties in the MG2 microphysics scheme produce more supercooled liquid water in CESM2 and result in more low clouds and more reflective clouds. Quantitatively, the global-mean rms

Figure 11. The 2000–2013 average shortwave cloud forcing (in W/m^2) (a) from the CERES-EBAF data set and (b–d) for model minus observation differences. Only one ensemble member is used for model simulations.

error is substantially reduced from about 14 W/m^2 in CESM1(LENS) to about 9.1 W/m^2 in CESM2(CAM6) and 9.7 W/m^2 in CESM2(WACCM6). The subtropical stratocumulus regions show a few large-scale improvements. For the most part however, deficiencies in cloud forcing seen in CESM1(LENS) are still seen in the CESM2(CAM6) and CESM2(WACCM6) stratocumulus regions.

In contrast, the improvements in the time-averaged long-wave cloud forcing (Figure 12) are much more modest compared to those of the shortwave cloud forcing. In particular, weak cloud forcing biases in CESM2(CAM6) and CESM2(WACCM6) remain similar to those of CESM1(LENS) outside of the tropics, illustrating the continued challenge of representing ice microphysics and their optical characteristics. Within the tropics, the Indo-Pacific improvements in cloud forcing largely reflect the patterns of tropical precipitation changes and associated upper tropospheric cloud and cloud water. These are most apparent in the reduction of the double ITCZ bias in the Pacific Ocean and the east-west precipitation distribution in the Indian Ocean. Quantitatively, the global-mean rms errors are about 8.3 W/m^2 in CESM2(CAM6) and 7.6 W/m^2 in CESM2(WACCM6), both representing modest reductions from the rms error of about 9.6 W/m^2 in CESM1(LENS).

8.3. Madden-Julian Oscillation

An aspect of variability that was poorly represented in CESM1 is the Madden-Julian Oscillation (MJO). In CESM2, minor changes to the Zhang-McFarlane deep convection scheme and the addition of CLUBB have combined to generate a much improved MJO (Figure 13). The observation-based Tropical Rainfall Measuring Mission-European Centre for Medium-Range Weather Forecast (ECMWF) Reanalysis-Interim (TRMM/ERAI; Kummerow et al., 2000; Dee et al., 2011) product shows convectively coupled coherent wind and precipitation features that propagate in quadrature from the Indian Ocean into the West Pacific. In CESM1(LENS), these features had incorrect westward propagation in the Indian Ocean and a lack of West Pacific variability. In both CESM2(CAM6) and CESM2(WACCM6), these adverse characteristics are largely absent, such that coherent eastward propagation now occurs from the Western Indian Ocean through the Maritime Continent and into the Central Pacific. Correlations are slightly weaker in CESM2 (WACCM6) than in CESM2(CAM6). Remaining bias characteristics are slightly weak coupling and phase speed.
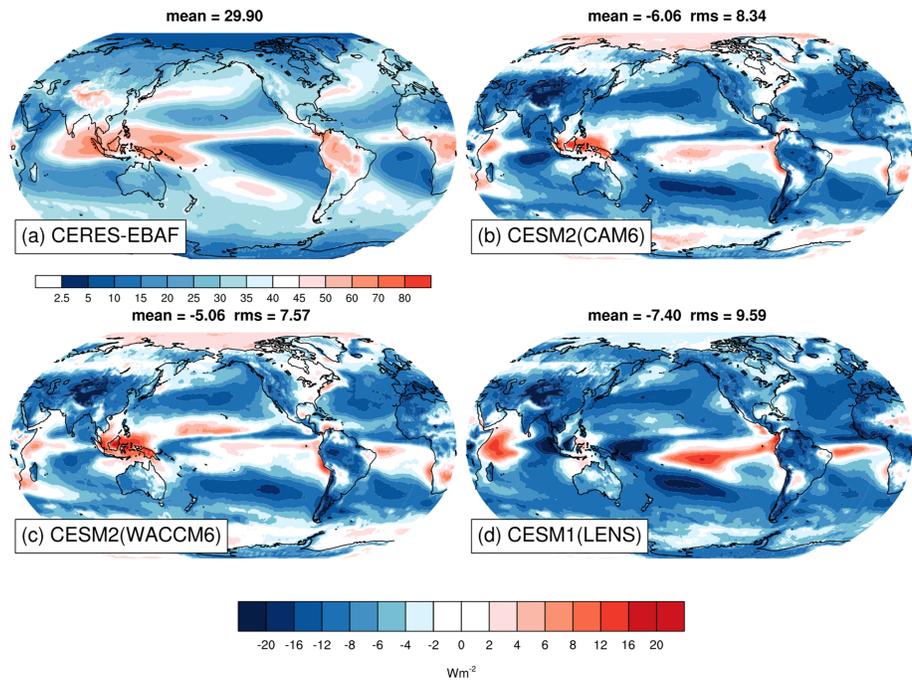
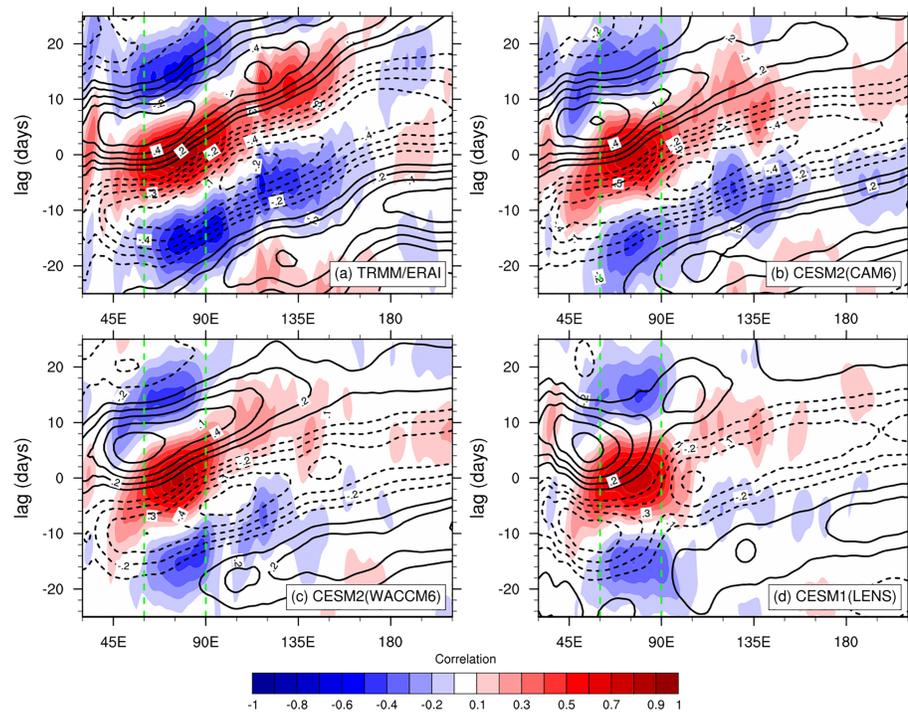**Figure 12.** Same as in Figure 11 except for longwave cloud forcing.

Figure 13. Lag correlation relationship between total precipitation averaged in the Indian Ocean (between 60º–90ºE as shown by the green dash lines) and total precipitation (color) and 850-mb zonal wind (contour) at all Indo-Pacific longitudes from (a) TRMM/ERAI data sets and (b–d) model simulations. The latitude range for averaging is 10ºS to 10ºN. Based on daily data filtered with a 20- to 100-day Lanczos filter. Only one ensemble member is used for model simulations.

ENSO Spatial Composites (DJF^{+1})
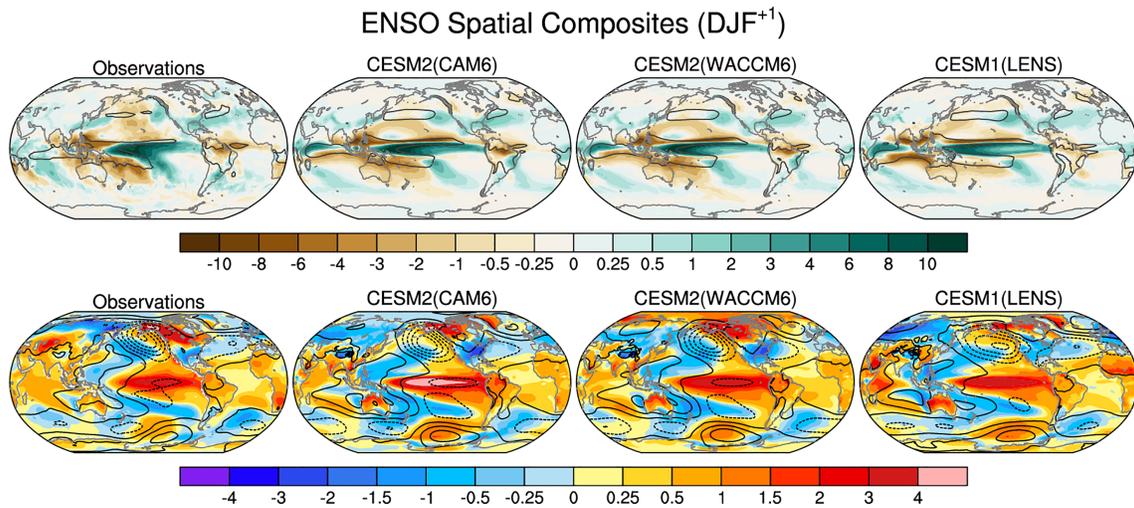
Figure 14. Spatial composites of DJF ENSO anomalies of precipitation (top row; in mm/day), surface air temperature over land and SST over ocean (bottom row; color shading; in °C), and sea level pressure (bottom row; contours; −16 to 16 hPa by 2 hPa with negative contours dashed) from observations, CESM2(CAM6) PI control for Years 100–1,199, CESM2(WACCM6) PI control for Years 50–499, and CESM1(LENS) PI control for Years 500–1,599. The 6 mm/day DJF climatological mean precipitation contour is shown in black for each data set in the top row. The normalized December Niño 3.4 time series are used to composite all years greater than 1 standard deviation (El Niño) and all years less than −1 standard deviation (La Niña), and the ENSO composite is formed by subtracting the La Niña composite from the El Niño composite. Observations are from GPCP for precipitation (1979–2017), ERSSTv5 for SST (1920–2017), BEST for 2-m air temperature (1920–2017), and ECMWF Reanalysis for the 20th century (ERA20C; Poli et al., 2016) combined with ERA-Interim (Berrisford et al., 2009) for sea level pressure (1920–2017).

8.4. El Niño-Southern Oscillation

ENSO is arguably the most important mode of tropospheric climate variability on interannual time scales. Because it was one of the worst aspects in CCSM3 simulations with a dominant period of 2 years, improving ENSO was the highest priority during CCSM4 development. As a result of changes in the atmospheric deep convection parameterization, ENSO characteristics were improved in CCSM4 and CESM1 (Deser et al., 2012; Neale et al., 2008; Richter & Rasch, 2008). These improvements are retained in CESM2. Figure 14 compares ENSO (El Niño minus La Niña) spatial composites of DJF anomalies of precipitation, surface air temperature over land and SST over ocean, and sea level pressure from observations and CESM2(CAM6), CESM2(WACCM6), and CESM1(LENS) PI control simulations. Again, CESM2(CAM6) and CESM2 (WACCM6) are very similar to each other with the exception of a slightly stronger ENSO amplitude in the former, and the spatial patterns are generally well simulated by all model versions. (The ENSO composite value of DJF SST in the Niño 3.4 region is significantly different between CESM2(CAM6) and CESM2 (WACCM6) at the 95% confidence level based on a 2-sided Student's t test [1.38 °C vs. 1.26 °C], for reasons that remain to be understood.) There are two notable improvements in CESM2(CAM6) and CESM2 (WACCM6) with respect to CESM1(LENS). First, the orientation of the sea level pressure teleconnection over the North Pacific, in particular its northwest-southeast tilt, is more realistic with implications for the circulation-induced air temperature anomalies over western North America. Second, the representation of the tropical precipitation anomaly pattern is improved, with more prominent drying over the maritime continent and SPCZ, and a broader equatorial maximum over the western half of the Pacific. These improvements are associated with a better representation of the climatological precipitation distribution as evidenced by a more prominent northwest-southeast oriented SPCZ that is confined to the western Pacific and less of a double ITCZ structure as indicated by the contours in the top row. The precipitation anomaly pattern in CESM2(CAM6) and CESM2(WACCM6) also shows a stronger east-west Walker Circulation component, more in line with observations, compared to the dominant narrow Hadley Circulation response in CESM1(LENS).

8.5. Stratospheric Ozone

WACCM6 is able to accurately simulate the evolution of stratospheric ozone over the late 20th century. In particular, WACCM6 can produce the halogen-induced ozone loss in the Southern Hemisphere polar
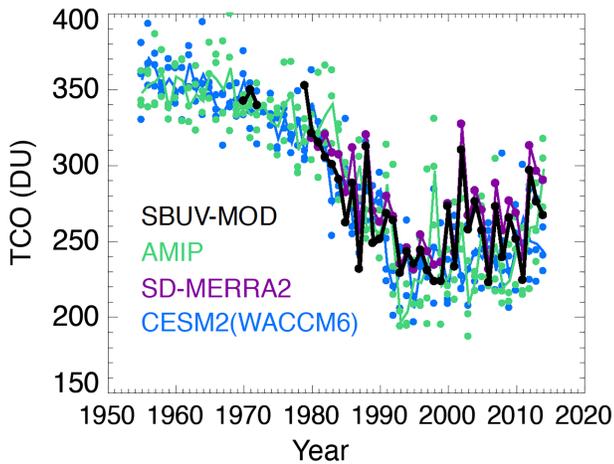
Figure 15. October 90° to 60°S Total Column Ozone (TCO, in Dobson Units: DU) from CESM2(WACCM6) simulations (blue), fixed SST (AMIP-type) simulations (green), Specified Dynamics simulation (SD-MERRA2; purple), and observations (SBUV-MOD; black). For CESM2(WACCM6) and AMIP, dots represent individual ensemble members while the ensemble means are shown by the solid lines.

stratosphere in spring. This occurs due to the activation of chlorine and bromine via heterogeneous chemistry on the surface of polar stratospheric cloud particles (WMO, 2010). Figure 15 illustrates the annual evolution of the Southern Hemisphere polar cap Total Column Ozone (TCO) in October. Coupled historical and AMIP-type simulations with WACCM6 are able to broadly represent the ozone loss from the 1970s to 1990s. "Specified Dynamics" (SD) simulations, wherein the model is constrained with MERRA-2 (Rienecker et al., 2011) meteorology (SD-MERRA2), indicate that much of the variability in observations can be explained by inter-annual variability of meteorology in the Southern Hemisphere polar vortex.

9. Atlantic Meridional Overturning Circulation

One of the climatically most important ocean circulations is the AMOC, representing a zonally integrated view of the Atlantic Ocean circulation. Through its associated heat and salt transports, AMOC significantly impacts spatial and temporal variability in the North Atlantic and surrounding areas, even influencing the global climate. Many studies link the low-frequency variability of the North Atlantic SSTs to fluctuations in AMOC (Zhang et al., 2019 and references therein). Figure 16 presents the 1995–2014 average AMOC distributions from CESM2(CAM6) and CESM2(WACCM6) in comparison to CESM1(LENS). These are for the Eulerian-mean, that is, resolved flow, component only as the parameterized contributions are relatively small. The figure shows that AMOC spatial structure and transports are indistinguishable between CESM2(CAM6) and CESM2(WACCM6). The primary pattern is the clockwise circulation cell (positive contours) associated with the southward flowing North Atlantic Deep Water (NADW). The maximum NADW transport is about 24 Sv (1 Sv = 10^6 m³/s) in CESM2(CAM6) and CESM2(WACCM6), while it is about 26 Sv in CESM1(LENS). There are no basin-scale observational estimates for AMOC, and only recently a few transbasin transport estimates have become available. Among these, the RAPID array at 26.5°N has the longest record. For the April 2004 to February 2017 period, the mean and standard deviation of the estimated NADW maximum transport at 26.5°N are 17.0 ± 3.3 Sv (Frajka-Williams et al., 2019; Smeed et al., 2018). Acknowledging the different averaging periods, the model maximum transports at this latitude are larger than the RAPID estimate by about 2–3 Sv with CESM1(LENS) showing the largest magnitude.

The penetration depth of the NADW cell as measured by the depth of the zero-contour line is similarly shallower in CESM2(CAM6) and CESM2(WACCM6) than in CESM1(LENS). Given that the ocean component in all model versions uses the same overflow parameterization (Danabasoglu et al., 2010) to represent dense-water overflows through the Denmark Strait and Faroe Bank Channel, the differences in penetration depth
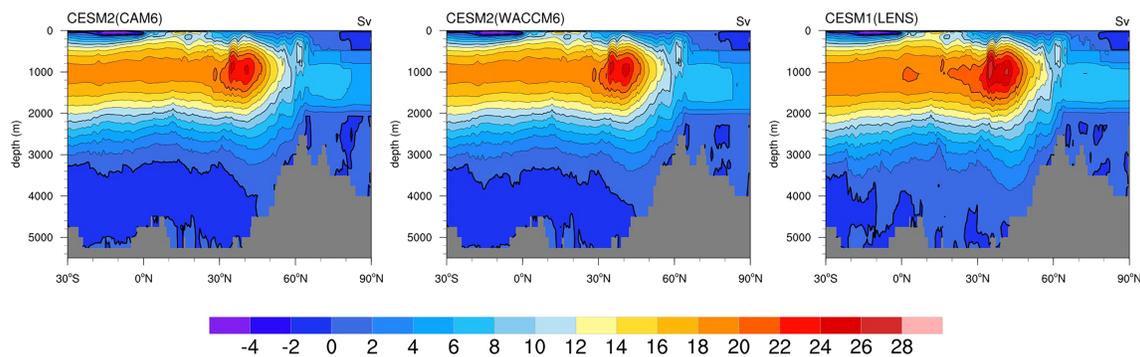
Figure 16. The 1995–2014 average Eulerian-mean Atlantic meridional overturning circulation (in Sv) from (left) CESM2(CAM6), (middle) CESM2(WACCM6), and (right) CESM1(LENS). The model distributions represent full ensemble means. The positive and negative contours indicate clockwise and counterclockwise circulations, respectively.

between the CESM1 and CESM2 simulations reflect differences in water mass properties, partially arising from surface flux differences.

The counterclockwise (negative contours) circulation region in the abyssal ocean below the NADW cell represents the northward flowing Antarctic Bottom Water (AABW). In comparison to very limited observational estimates, all the simulations suffer from a chronic bias of very anemic AABW transport due to too weak AABW formation, with CESM2(CAM6) and CESM2(WACCM6) showing only a marginally broader and stronger AABW cell than in CESM1(LENS). The observational range for the AABW transport into the Atlantic basin is about 1–5 Sv at 24.5°N based on sparse data sets dating back to 1957 (Frajka-Williams et al., 2011). Both CESM2 simulations have a maximum transport of about 0.5 Sv, compared to near zero in CESM1(LENS). Such weak deep overturning circulations lead to weak ventilation at depth evident in tracer distributions particularly in the North Pacific, compared to available observations (not shown).

10. Sea-Ice and Land-Ice Fields

10.1. Sea-Ice Thickness and Extent

The annual-mean sea-ice thickness and extent are shown in Figure 17, averaged for 1979–2014. For the Arctic, sea-ice in both CESM2 simulations is thinner than in CESM1(LENS), which has a total annual-mean Arctic ice volume of 28.9×10^3 km^3. The sea ice is considerably thicker in the simulations with CESM2 (WACCM6) than with CESM2(CAM6), with the respective total annual-mean Arctic ice volumes of 21.4×10^3 and 15.4×10^3 km^3. The CESM2(WACCM6) simulations agree well with observationally based estimates from the Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS; Schweiger et al., 2011), which simulates a 1979–2014 average Arctic ice volume of 20.2×10^3 km^3. The thicker sea ice in CESM2(WACCM6) is related to liquid Arctic clouds, which are thicker as a result of increased aerosol concentrations predicted from the interactive chemistry module (Gettelman, Mills, et al., 2019). Specifically, the interactive oxidants with full chemistry in CESM2(WACCM6) mean that aerosols will evolve differently. Because the aerosols are not produced locally in remote regions such as the Arctic, the increase in aerosol concentrations over the Arctic is due to a lengthening of the aerosol lifetime. The anticorrelation between aerosols and oxidants is meteorologically dependent and so may not be seen in monthly average oxidants fields used in CESM2(CAM6), leading to a different aerosol evolution, particularly in such remote areas. The differences in sea-ice thickness between the two model configurations have consequences for the annual cycle of ice extent. In particular, with the thin ice bias in the CESM2(CAM6) simulations, more melting occurs during the summer months, with a resulting low summer ice extent as compared to satellite passive-microwave observations (Figure 17g). In contrast, CESM2(WACCM6) and CESM1(LENS) have an excellent simulation of summer ice extent compared to observations. In both CESM2(CAM6) and CESM2 (WACCM6) simulations, the ice extent is low in winter compared to observations. This is in contrast to CESM1(LENS), which simulates a winter ice extent that is in very good agreement with observations.

In the Southern Hemisphere, the sea ice is nearly identical in CESM2(CAM6) and CESM2(WACCM6), both somewhat thinner than the sea ice simulated in CESM1(LENS). As in observations, a primarily seasonal ice pack is present, and the ice remains thinner than 1 m in the annual mean over most of the sea-ice domain for all model versions. Sea-ice thicker than 1 m forms in the Amundsen/Bellingshausen and Weddell Seas. This is in reasonable agreement with the limited in situ ice thickness observations (Worby et al., 2008). The ice extent annual cycle is also reasonably well simulated in both CESM2 simulations. The maximum simulated ice cover is biased low by about 1.6×10^6 km^2, and the ice melt-back is delayed by about 1 month relative to observations but reaches a similar minimum ice cover in February. This is in contrast to CESM1(LENS), which has a good simulation of the winter maximum ice extent but retains too much ice at the summer minimum.
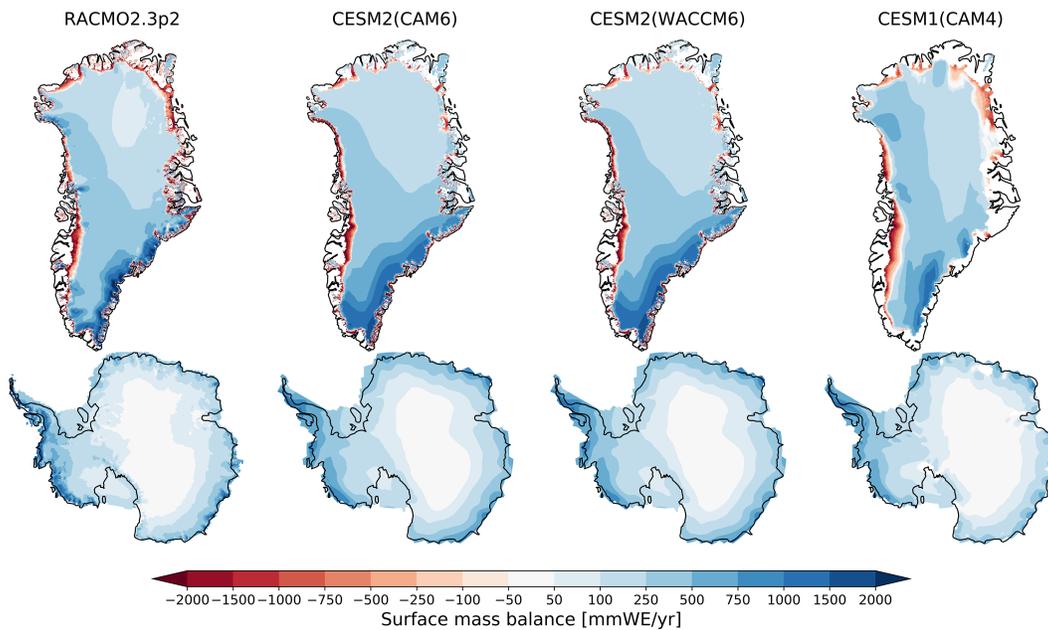
10.2. Land-Ice Surface Mass Balance

Figure 18 shows the simulated Greenland and Antarctic SMB for the 1961–1990 mean of CESM2(CAM6) and CESM2(WACCM6) historical simulations in comparison to the regional atmospheric climate model RACMO2 (Noël et al., 2018; van Wessem et al., 2018) and to CESM1(CAM4) (Vizcaíno et al., 2013). (We show CESM1(CAM4) results because ice sheet SMB was not computed for CESM1(LENS) simulations.) CESM2(CAM6) and CESM2(WACCM6) solutions are very similar to each other, both reproducing the
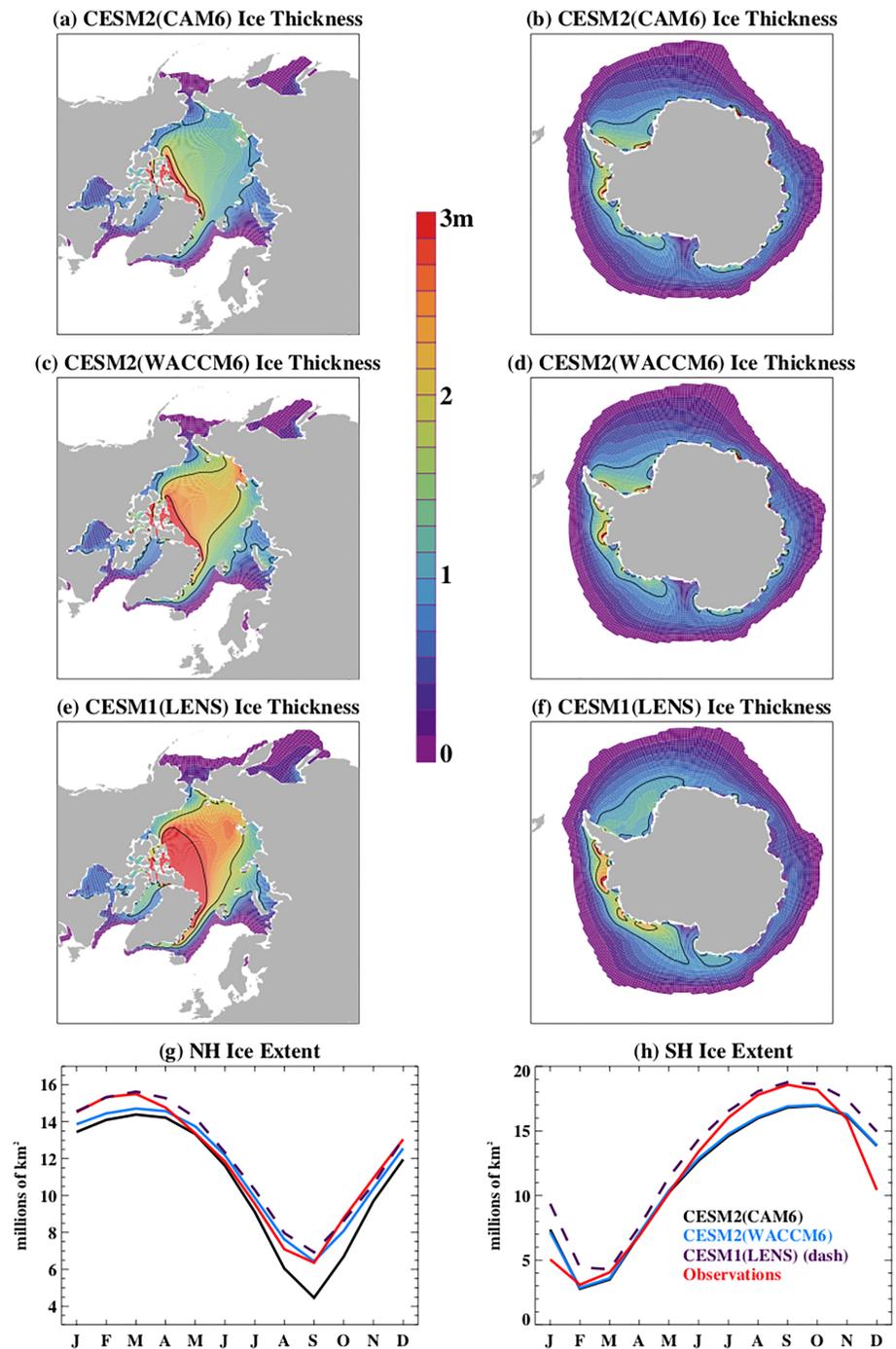
Figure 17. The annual-mean sea-ice thickness (in m) in CESM2(CAM6) (a and b), CESM2(WACCM6) (c and d), and CESM1(LENS) (e and f) simulations. The annual cycle of sea-ice extent (in km^2) for the (g) Northern Hemisphere and (h) Southern Hemisphere. All values are averages for 1979–2014. The observations are estimates from the PIOMAS data set. The analysis uses 6, 3, and 40 ensemble members for CESM2(CAM6), CESM2(WACCM6), and CESM1(LENS), respectively. In (h), CESM2(CAM6) and CESM2(WACCM6) lines overlap.

RACMO2 SMB reasonably well for both ice sheets. Improvements relative to CESM1(CAM4) can be attributed in part to more realistic snow/firn physics in CLM5 (van Kampenhout et al., 2017), and to the Beljaars et al. (2004) orographic form drag scheme and new cloud microphysics in CAM6. Remaining biases for Greenland include excessive snowfall in the south, likely associated with unresolved

Figure 18. Surface mass balance (SMB in mm water equivalent per year) of the Greenland ice sheet (top row) and Antarctic ice sheet (bottom row) from statistically downscaled output of the regional atmospheric climate model RACMO2.3p2, and from CESM2(CAM6), CESM2(WACCM6), and CESM1(CAM4). The CESM results are from historical simulations averaged over 1961–1990, with eleven ensemble members for CESM2(CAM6), three members for CESM2(WACCM6), and one member for CESM1(CAM4). For Greenland, CESM2 SMB has been downscaled using multiple elevation classes to a 5-km grid for CESM1(CAM4) and to a 4-km CISM grid for both versions of CESM2. For Antarctica, both CESM1 and CESM2 SMB are shown on the land model grid without downscaling. RACMO2.3p2 data are averaged over 1961–1990 for Greenland (Noël et al., 2018) and over 1979–2010 for Antarctica (van Wessem et al., 2018) at a resolution of 11 and 27 km, respectively.

topography, excessive rainfall (van Kampenhout et al., 2019), and a generally too positive SMB bias in the north.

11. Land Fields

The land model variables are assessed with the International Land Model Benchmarking (ILAMBv2.1; Collier et al., 2018) package. The ILAMBv2 used here integrates analysis for 30 variables evaluated against more than 60 data products. ILAMB package results include statistics, maps, time series, and metrics for relative bias, rms error, seasonal cycle phase, spatial distribution, and interannual variability, as well as variable-to-variable assessments that evaluate functional relationships between variables. ILAMB summary diagrams, which integrate scores across several data sets and metrics for each variable and variable-to-variable comparisons, are shown in Figure 19 for three ensemble members each of CESM1(LENS) and CESM2(CAM6)—CESM2(WACCM6) results are very similar to those of CESM2(CAM6). For most variables assessed, CESM2(CAM6) better matches the observations than CESM1(LENS). Except for precipitation, all the land climate variables (shown in gray in the left panel of Figure 19) show modest improvement, indicating an improved surface climate simulation in CESM2(CAM6) due to a combination of improvements in CAM6 and CLM5. The land carbon, water, and energy variable improvements are also seen in land-only simulations forced with historical climate reconstructions (not shown; see Lawrence et al., 2019, for detailed assessment), which suggests that improvements in these variables are predominantly a result of CLM5 physics and biogeochemistry model development rather than an outcome of a better-simulated surface climate in CESM2(CAM6). The consistently improved variable-to-variable scores suggests improved process representation in CLM5. These functional relationship scores are improved even for comparisons involving a mean quantity that is slightly degraded in CESM2(CAM6)—for example, land precipitation is slightly degraded, but precipitation versus gross primary productivity and precipitation versus burned area are improved. This result supports the interpretation that the improvements derive mainly from CLM5 developments. Insight into the sources of improvement is provided in Lawrence et al. (2019) and Wieder et al. (2019).

A caveat here is that ILAMB results should be interpreted carefully and that an improvement or degradation in a particular ILAMB variable should not be overly interpreted as indicative of whether or not results that are dependent on that variable are reliable or not (see Lawrence et al., 2019, and Bonan et al., 2019). For example, soil carbon is degraded according to the ILAMB metric, but other aspects of soil carbon such as the soil carbon turnover time appear to be improved (Lawrence et al., 2019), potentially reflecting a lack of consistency across observational data sets or a limitation in the applied ILAMB metric. A note of caution on the soil carbon field in CESM2 is that due to an error in the model spin-up process, unrealistically large soil carbon stocks are seen for some high-latitude grid cells where vegetation does not survive the cold climate (e.g., Canadian Archipelago). For these grid cells, the spin-up process was supposed to set soil carbon stocks to near zero, but due to very small but nonzero vegetation stocks, the soil carbon stocks were not set to zero as intended. Model users can resolve this problem (which was fixed in CESM2.1.1) by screening out large soil carbon stocks when vegetation carbon is zero.

One of the clearest improvements in CESM2 and CLM5 is the simulated global land carbon accumulation trends. CESM1/CLM4 produced an unrealistically strong nutrient limitation on photosynthesis, which limited the capacity in CESM1 for carbon uptake in response to increases in atmospheric CO_2 concentrations. Consequently, in CESM1 land use and land-cover change carbon loss fluxes dominate over the CO_2 fertilization response, and CESM1 simulates an accumulated carbon loss over the period 1960 to 2004, whereas the observationally-based estimates indicate a carbon gain of ~40 Pg C (Figure 20). In CESM2(CAM6) and CESM2(WACCM6), the CO_2 fertilization response appears to be more reasonable and, when this is combined with updates to the land use and land-cover change carbon fluxes, results in a late 20th century accumulated carbon trend that agrees well with observationally based estimates. Further discussion on the source of improvement for this feature of CESM2, which is also seen in land-only simulations, is provided in Lawrence et al. (2019), Wieder et al. (2019), and Bonan et al. (2019). In addition to the improved accumulated carbon, CESM2 simulations also show evidence of considerable improvements in the amplitude of the annual cycle of net ecosystem exchange, especially at northern high latitudes (see Carbon Dioxide variable metrics in Figure 19), which translate to improved annual cycle amplitudes of atmospheric CO_2 concentrations in emissions-driven simulations.

12. Equilibrium Climate Sensitivity and Transient Climate Response

An important emergent property of the coupled model is its Equilibrium Climate Sensitivity (ECS; e.g., Stocker et al., 2013), defined as the equilibrium change in global-mean surface temperature after a doubling of CO_2 concentration. ECS can be estimated using a slab ocean model (SOM) configuration, because (1) such a configuration equilibrates much faster compared to one with a full-depth ocean model (about 50 vs. >1,000 years) and (2) ECS values obtained with a SOM versus a full-depth ocean component are comparable to each other (Danabasoglu & Gent, 2009). Therefore, SOM configurations are used here to estimate the ECS of both model configurations. The control SOM experiments use a preindustrial value of CO_2 set at 280 ppmv. These respective control simulations are compared with a second set of SOM experiments in which the CO_2 concentration is instantaneously doubled to 560 ppmv (2xCO_2). All SOM experiments are integrated long enough to ensure that global-mean surface temperature has equilibrated and TOA radiative imbalances are near zero. Differencing the equilibrated surface temperatures (SST over the ice-free oceans and a radiative temperature over land and sea-ice) between the respective control and 2xCO_2 experiments, the ECS values are computed as 5.3 °C for CESM2(CAM6) (Gettelman, Hannay, et al., 2019) and 5.1 °C for CESM2 (WACCM6). These values represent significant increases from an ECS of around 4.0 °C in the SOM version of CESM1(CAM5) (Gettelman et al., 2012).

Values of climate sensitivity derived from the first 150 years of fully-coupled 4xCO2 runs using the Gregory method (so-called Effective Climate Sensitivity; Gregory et al., 2004) are 5.3 °C for CESM2(CAM6) and 4.8 °C for CESM2(WACCM6) compared to 4.3 °C in CESM1(CAM5) obtained with the same method (Meehl et al., 2013). However, this level of relatively good agreement with the above SOM-based estimates of ECS is deceptive, because the Gregory method estimate of climate sensitivity for CESM2 is highly sensitive to the number of simulation years used in the calculation, due to nonlinearities in the relationship between TOA energy imbalance and temperature, with larger estimates of climate sensitivity as more years are used.
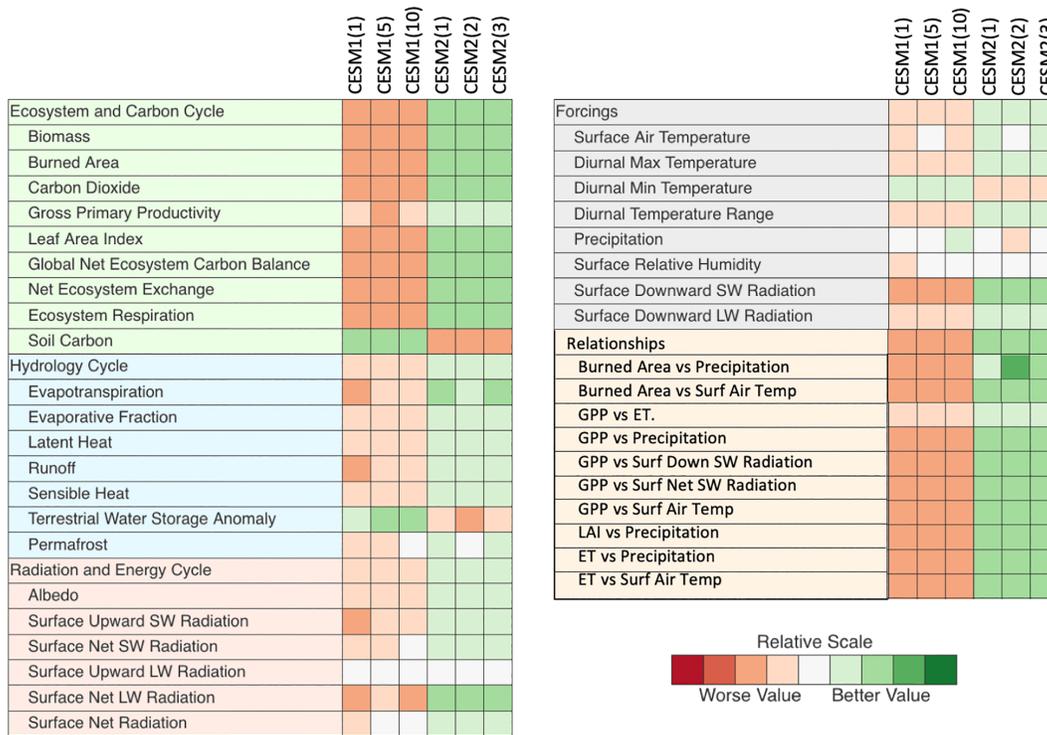
Figure 19. ILAMB summary diagrams for three ensemble members each of CESM1(LENS) and CESM2(CAM6) historical simulations. The left panel shows the single-variable assessment summary for land variables. The upper part of the right panel shows assessment of the land *forcing* variables as simulated by the atmosphere model. The lower part of the right panel shows a variable-to-variable comparison summary. See Collier et al. (2018) for details on all metrics. Full ILAMB package results for both coupled and land-only simulations are available online (at http://www.cesm.ucar.edu/experiments/cesm2.0/land/diagnostics/clm_diag_ILAMB.html).

The evolution of CESM2′s climate sensitivity through several development versions of the model is examined in detail in Gettelman, Hannay, et al. (2019). This study suggests that the increased climate sensitivity in CESM2 has arisen from a combination of relatively small changes to cloud microphysics and boundary layer parameters as well as from land parameter changes that were introduced late in the development process, along with the introduction of more realistic clouds with CLUBB and MG2.
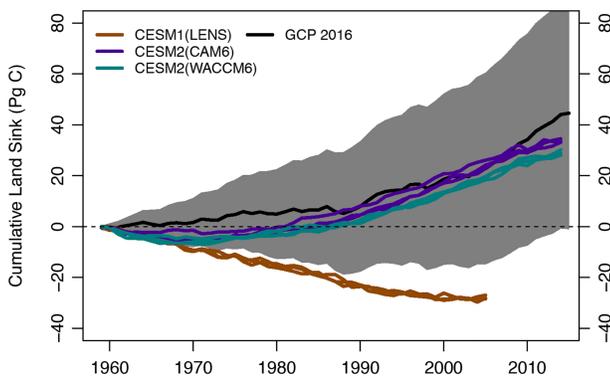
Figure 20. Land carbon accumulation for the period 1960–2014 for three historical ensemble members of CESM2(CAM6), CESM2(WACCM6), and CESM1(LENS) along with observationally based estimates. The latter are from the Global Carbon Project (Le Quéré et al., 2016). Gray shading shows uncertainty range for the observational estimates, using an uncertainty of 0.8 Pg C/year and calculated assuming that the errors are additive in time as in Koven et al. (2013).

Another important emergent property of the coupled model is its Transient Climate Response (TCR). It is calculated from a pair of fully coupled simulations: a control simulation, again usually for PI conditions, and a 1% per year CO_2 concentration increase simulation. Following the ESMValTool (Eyring et al., 2020) procedure, TCR in CESM2 is calculated as the difference of the 20-year average (Years 60–79) global mean surface temperature, centered around the time of CO_2 doubling at about Year 70, from the 140-year linear trend (fit) of the corresponding PI control, overlapping the 1% per year CO_2 increase run. The TCR values are obtained as 2.0 °C for CESM2(CAM6) and 1.9 °C for CESM2(WACCM6). These values are very comparable to a TCR value of 2.3 °C for CESM1(CAM5) reported by Meehl et al. (2013). Processes leading to CESM2′s TCR value are being investigated. Preliminary analysis suggests that, in CESM2, while tropics and the Southern Ocean warm faster, the Northern Hemisphere at middle and high latitudes warms at a slower rate, in comparison to CESM1.

It is important to stress that all estimates of ECS and TCR are subject to substantial uncertainties arising from the details of the calculation procedures. These uncertainties include the selection of reference periods for
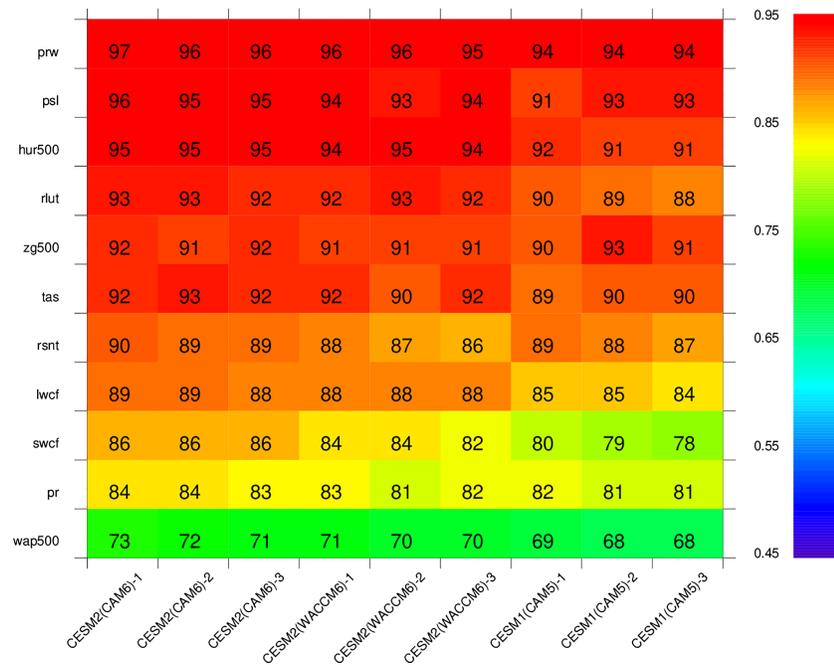
Figure 21. Mean pattern correlations across climatological mean, seasonal contrasts (JJA-DJF), and ENSO teleconnections for three ensemble members of CESM2(CAM6), CESM2(WACCM6), and CESM1(LENS) for various fields: precipitable water (prw), sea level pressure (psl), 500-hPa relative humidity (hur500), outgoing longwave radiation (rlut), 500-hPa geopotential height (zg500), 2-m air temperature (tas), top of atmosphere net shortwave flux (rsnt), longwave and shortwave cloud forcing (lwcf, swcf), rainfall (pr), and 500-hPa vertical velocity (wap500).

the PI simulations, the length of the averaging periods, and whether detrending is applied or not. For the above CESM values, such details may result in 0.1–0.2 °C differences.

13. Summary and Discussion

The new version of the Community Earth System Model, CESM2, contains many substantial science and infrastructure improvements and capabilities. The CESM project is contributing simulations to the CMIP6 effort with two CESM2 model configurations differing only in their atmospheric components. While one configuration uses the low-top CAM6, CESM2(CAM6), the other employs WACCM6 with its high-top and comprehensive chemistry, CESM2(WACCM6). Both configurations use the same nominal 1° horizontal resolution. A noteworthy new approach in CESM2 is that CESM2(WACCM6) PI and historical simulations were performed first to obtain some of the necessary forcing data sets to run the corresponding CESM2 (CAM6) simulations. This was critical to ensure that, for the first time for CESM, a physically consistent pair of low- and high-top atmospheric models are used. This manuscript documents the CESM2 development process and evaluates solutions from the 1,200-year CESM2(CAM6) and 500-year CESM2(WACCM6) PI control integrations and from the subsequent historical simulations. The focus has been primarily on the historical simulations, as they can be more easily and directly compared with modern observations.

The new developments and capabilities in CESM2 result in many improvements in solutions in comparison to available observations and those of CESM1. These improvements include major reductions in the tropical wet biases of the Indian Ocean, the SPCZ, and the Atlantic ITCZ, generally with less of a double ITCZ structure in both the Pacific and Atlantic Basins; reductions in the dipole of Andean wet and Amazon dry biases as well as in Australian and Western U.S. excessive precipitation biases; substantial reductions in the reflective biases of the clouds throughout the tropics; a much better representation of MJO characteristics with coherent eastward propagation of wind and precipitation anomalies; more realistic representation of the northwest-southeast tilt of the sea level pressure anomaly teleconnection patterns associated with ENSO over the North Pacific; a global land carbon accumulation trend that matches the observationally based

trend remarkably well; and better representation of the SMB for both the Greenland and Antarctic ice sheets. An important finding is that CESM2(CAM6) and CESM2(WACCM6) historical simulations produce solutions that are very similar to each other in most variables. So these improvements are present in both sets of simulations. An exception is the Arctic sea-ice thickness with CESM2(WACCM6) showing thicker sea ice than in CESM2(CAM6) in better agreement with observations. The thicker sea ice in CESM2 (WACCM6) is related to liquid Arctic clouds, which are thicker as a result of increased aerosol concentrations predicted from the interactive chemistry module.

Figure 21 presents a summary analysis comparing CESM2 solutions to those of CESM1(LENS). The figure shows mean pattern correlations across climatological mean, seasonal contrasts (JJA-DJF), and ENSO teleconnections for the first three ensemble members of CESM2(CAM6), CESM2(WACCM6), and CESM1 (LENS) historical simulations for 11 atmospheric fields with respect to many observational-based data sets. Indeed, CESM2 simulations have notable improvements in many of the metrics considered here compared to those of CESM1(LENS).

Undoubtedly, many biases still remain in the model solutions. These biases range from relatively minor ones, such as local precipitation errors without discernable global impacts to biases that can have significant climate impacts such as the thin Arctic sea ice in CESM2(CAM6) simulations. The sea-ice bias can perhaps be tuned away at the expense of possible appearance of other biases. There are also much more persistent biases such as the incorrect separation and path of the Gulf Stream-North Atlantic Current system, with accompanying large SST and surface salinity biases. This latter class of biases can have significant climate impacts as well, but no robust remedy exists to alleviate them in noneddy-permitting/noneddy-resolving ocean models usually used in climate simulations.

The path to finalizing a CESM2 version that produced acceptable simulations took much longer than anticipated. The process was hindered particularly by two major unacceptable features in simulations. The first concerned formation of unrealistically extensive sea-ice cover over the Labrador Sea region in PI control simulations. Unfortunately, extensive analysis of simulations did not reveal a particular process (or processes) as the culprit. To avoid further delays, a practical approach was adopted that involved performing several member ensemble simulations, which differed at round-off levels in their atmospheric potential temperature initial conditions, and then using one of the members without an extensive Labrador Sea ice cover to provide the ocean and sea-ice initial conditions. The second issue was that the global-mean surface temperature time series in historical simulations showed a substantial and unrealistic cooling, particularly during the second half of the 20th century, with end-of-the-century temperatures being colder than in the 1850s. Adoption of updated CMIP6 emissions data sets, along with several modifications in the aerosol-cloud interaction processes based on observational data from recent volcanic sulfur emission events, resulted in surface temperature time series that evolve realistically during the historical period. Nevertheless, there are several discrepancies in the details of the modeled and observed time series. For example, the model surface temperatures are warmer in the early part of the 21st century than in observations by 0.2 to 0.4 °C due at least partly to the missing decadal-scale slowdown in warming during roughly 2000–2012 in the simulations. However, such a difference is not unexpected because the observed slowdown is associated with internal climate variability in addition to anthropogenic and natural forcings. This warm bias is also reflected in the SST and over-land surface temperature distributions.

An emergent property of climate simulations is their ECS, which in CESM2 is calculated as 5.3 °C for CESM2 (CAM6) and 5.1 °C for CESM2(WACCM6) based on CO_2 doubling experiments with CESM2 SOM configurations. These values represent a significant increase from an ECS of 4.0 °C calculated for CESM1, also using a SOM configuration. Climate sensitivities based on the Gregory method from fully-coupled 4xCO$_2$ runs are calculated as 5.3 °C for CESM2(CAM6) and 4.8 °C for CESM2(WACCM6), again larger than the value of 4.3 °C in CESM1(CAM5) obtained with the same method (Meehl et al., 2013). However, these latter estimates are highly sensitive to the number of simulation years used in the calculation due to nonlinearities in the relationship between TOA energy imbalance and temperature. Our analysis suggests that the increased climate sensitivity in CESM2 has arisen from a combination of seemingly small changes to cloud microphysics and boundary layer parameters as well as from land parameter changes that were introduced late in the development process, along with the introduction of more realistic clouds with CLUBB and MG2 (see Gettelman, Hannay, et al., 2019). In contrast, the CESM2 TCR values, calculated as 2.0 °C for CESM2

Table A1
Forcing Data Sets Date Information

Seasonal cycle date	WACCM6 experiment	Years averaged
1849	PI control	21–50
1860	Historical	1856–1864
1870	Historical	1866–1874
1880	Historical	1876–1884
1890	Historical	1886–1894
1900	Historical	1896–1904
1910	Historical	1906–1914
1920	Historical	1916–1924
1930	Historical	1926–1934
1940	Historical	1936–1944
1950	Historical	1946–1954
1960	Historical	1956–1965
1970	Historical	1966–1974
1980	Historical	1976–1984
1990	Historical	1986–1994
2000	Historical	1996–2004
2010	Historical	2006–2014
2013	Historical	2012–2014
2015	Historical	2012–2014

(CAM6) and 1.9 °C for CESM2(WACCM6), are comparable to a TCR value of 2.3 °C for CESM1(CAM5) reported in Meehl et al. (2013).

The CESM community has been participating in about 20 CMIP6-endorsed MIPs. The data sets from these MIPs and all the DECK simulations are being made available on the ESGF site for use of the broader research community. These simulations, including the PI control and historical integrations presented here, are analyzed in more detail in over 70 manuscripts collected in the *AGU CESM2 Virtual Special Issue*. Our preliminary analyses suggest that the CESM2 simulations exhibit agreement with satellite era observations of the climate mean state, seasonal cycle, and interannual variability that is among the closest of coupled climate models in the present CMIP6 archive.

Appendix A.: Construction of Forcing Data Sets From CESM2(WACCM6) Historical Simulations for Use in Historical Simulations With CESM2(CAM6)

For CESM2(CAM6) historical simulations, 12-month annual cycles are provided at quasi-decadal frequency (Table A1). The first year in the forcing file is 1849, which is derived from Years 21–50 of the CESM2 (WACCM6) PI control simulation. Subsequent annual cycles are provided using the ensemble average of the three CESM2(WACCM6) historical simulations. One annual cycle per decade is provided from 1860 to 2010, each averaged over 9 years centered on the year. For example, oxidants provided for the midmonth date of 16 January 1860 are derived from the average of January averages for Years 1856–1864 of three CESM2(WACCM6) historical ensemble members. An annual cycle is also provided for 2015, averaged from Years 2012–2014, as the end year needed for time interpolation. CESM2(CAM6) interpolates annual cycles for years between those provided.

References

Ali, A. A., Xu, C., Rogers, A., Fisher, R. A., Wullschleger, S. D., Massoud, E. C., et al. (2016). A global scale mechanistic model of photosynthetic capacity (LUNA V1.0). *Geoscientific Model Development*, 9, 587–606. https://doi.org/10.5194/gmd-9-587-2016

Armstrong, R. A., Lee, C., Hedges, J. I., Honjo, S., & Wakeham, S. G. (2002). A new, mechanistic model for organic carbon fluxes in the ocean based on the quantitative association of POC with ballast minerals. *Deep-Sea Research*, 49, 219–236. https://doi.org/10.1016/S0967-0645(01)00101-1

Aschwanden, A., Fahnestock, M. A., & Truffer, M. (2016). Complex Greenland outlet glacier flow captured. *Nature Communications*, 7, 10524. https://doi.org/10.1038/ncomms10524

Asselin, R. (1972). Frequency filter for time integrations. *Monthly Weather Review*, 100, 487–490.

Assur, A. (1958). Composition of sea ice and its tensile strength. In *Arctic sea ice; Conference held at Easton, Maryland, February 24–27, 1958*, v. 598 of Publs. Natl. Res. Coun. Wash., pages 106-138, Washington, DC, US.

Beljaars, A. C. M., Brown, A. R., & Wood, N. (2004). A new parameterization of turbulent orographic form drag. *Quarterly Journal of the Royal Meteorological Society*, 130, 1327–1347. https://doi.org/10.1256/qj.03.73

Berrisford, P., Dee, D. P. K. F., Fielding, K., Fuentes, M., Kallberg, P., Kobayashi, S., & Uppala, S. (2009). The ERA-interim archive. ERA report series, (1), 1-16, Reading, U. K.

Beusen, A. H. W., Bouwman, A. F., Van Beek, L. P. H., Mogollón, J. M., & Middelburg, J. J. (2016). Global riverine N and P transport to ocean increased during the 20th century despite increased retention along the aquatic continuum. *Biogeosciences*, 13, 2441–2451. https://doi.org/10.5194/bg-13-2441-2016

Beusen, A. H. W., Van Beek, L. P. H., Bouwman, A. F., Mogollón, J. M., & Middelburg, J. J. (2015). Coupling global models for hydrology and nutrient loading to simulate nitrogen and phosphorus retention in surface water—Description of IMAGE-GNM and analysis of performance. *Geoscientific Model Development*, 8, 4045–4067. https://doi.org/10.5194/gmd-8-4045-2015

Bogenschutz, P. A., Gettelman, A., Morrison, H., Larson, V. E., Craig, C., & Schanen, D. P. (2013). Higher-order turbulence closure and its impact on climate simulation in the Community Atmosphere Model. *Journal of Climate*, 26, 9655–9676. https://doi.org/10.1175/JCLI-D-13-00075.1

Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., et al. (2011). Improving canopy processes in the Community Land Model Version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *Journal of Geophysical Research*, 116, G02014. https://doi.org/10.1029/2010JG001593

Bonan, G. B., Lombardozzi, D. L., Wieder, W. R., Oleson, K. W., Lawrence, D. M., Hoffman, F. M., & Collier, N. (2019). Model structure and climate data uncertainty in historical simulations of the terrestrial carbon cycle (1850-2014). *Global Biogeochemical Cycles*, 33(10), 1310–1326. https://doi.org/10.1029/2019GB006175

Boville, B. A., & Gent, P. R. (1998). The NCAR Climate System Model, version one. *Journal of Climate*, 11, 1115–1130.

Brunke, M. A., Broxton, P., Pelletier, J., Gochis, D., Hazenberg, P., Lawrence, D. M., et al. (2016). Implementing and evaluating variable soil thickness in the Community Land Model, version 4.5 (CLM4.5). *Journal of Climate*, 29, 3441–3461. https://doi.org/10.1175/JCLI-D-15-0307.1

Cavaleri, L., Fox-Kemper, B., & Hemer, M. (2012). Wind waves in the coupled climate system. *Bulletin of the American Meteorological Society*, *93*, 1651–1661.

Cavalieri, D. J., Parkinson, C. L., Gloersen, P., & Zwally, H. J. (1996). updated yearly. Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS passive microwave data, version 1. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: https://doi.org/10.5067/8GQ8LZQVL0VL.

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., et al. (2018). The International Land 1 Model Benchmarking (ILAMB) system: Design, theory, and implementation. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2731–2754. https://doi.org/10.1029/2018MS001354

Collins, W. D., Bitz, C. M., Blackmon, M. L., Bonan, G. B., Bretherton, C. S., Carton, J. A., et al. (2006). The Community Climate System Model Version 3 (CCSM3). *Journal of Climate*, *19*, 2122–2143.

Danabasoglu, G., Bates, S. C., Briegleb, B. P., Jayne, S. R., Jochum, M., Large, W. G., et al. (2012). The CCSM4 ocean component. *Journal of Climate*, *25*, 1361–1389. https://doi.org/10.1175/JCLI-D-11-00091.1

Danabasoglu, G., & Gent, P. R. (2009). Equilibrium climate sensitivity: Is it accurate to use a slab ocean model? *Journal of Climate*, *22*, 2494–2499. https://doi.org/10.1175/2008JCLI2596.1

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

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*, 553–597.

Deser, C., Guo, R., & Lehner, F. (2017). The relative contributions of tropical Pacific sea surface temperatures and atmospheric internal variability to the recent global warming hiatus. *Geophysical Research Letters*, *44*, 7945–7954. https://doi.org/10.1002/2017GL074273

Deser, C., Phillips, A. S., Tomas, R. A., Okumura, Y., Alexander, M. A., Capotondi, A., et al. (2012). ENSO and Pacific decadal variability in Community Climate System Model version 4. *Journal of Climate*, *25*, 2622–2651. https://doi.org/10.1175/JCLI-D-11-00301.1

Eyring, V., Bock, L., Lauer, A., Righi, M., Schlund, M., Andela, B., et al. (2020). ESMValTool v2.0 – Extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP. *Geoscientific Model Development Discussion*. https://doi.org/10.5194/gmd-2019-291 (in review)

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

Fisher, R. A., Wieder, W. R., Sanderson, B. M., Koven, C. D., Oleson, K. W., Xu, C., et al. (2019). Parametric controls on vegetation responses to biogeochemical forcing in the CLM5. *Journal of Advances in Modeling Earth Systems*, *11*(9), 2879–2895. https://doi.org/10.1029/2019MS001609

Frajka-Williams, E., Ansorge, I. J., Baehr, J., Bryden, H. L., Paz Chidichimo, M., Cunningham, S. A., et al. (2019). Atlantic meridional overturning circulation: Observed transport and variability. *Frontiers in Marine Science*, *6*, 260. https://doi.org/10.3389/fmars.2019.00260

Frajka-Williams, E., Cunningham, S. A., Bryden, H., & King, B. A. (2011). Variability of Antarctic bottom water at 24.5°N in the Atlantic. *Journal of Geophysical Research*, *116*, C11026. https://doi.org/10.1029/2011JC007168

Galbraith, E. D., & Martiny, A. C. (2015). A simple nutrient-dependence mechanism for predicting the stoichiometry of marine ecosystems. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 8199–8204. https://doi.org/10.1073/pnas.1423917112

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., et al. (2011). The Community Climate System Model Version 4. *Journal of Climate*, *24*, 4973–4991. https://doi.org/10.1175/2011JCLI4083.1

Gettelman, A., Hannay, C., Bacmesiter, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., et al. (2019). High climate sensitivity in the Community Earth System Model Version 2 (CESM2). *Geophysical Research Letters*, *46*, 8329–8337. https://doi.org/10.1029/2019GL083978

Gettelman, A., Kay, J. E., & Shell, K. M. (2012). The evolution of climate sensitivity and climate feedbacks in the Community Atmosphere Model. *Journal of Climate*, *25*, 1453–1469. https://doi.org/10.1175/JCLI-D-11-00197.1

Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R., et al. (2019). The Whole Atmosphere Community Climate Model version 6 (WACCM6). *Journal of Geophysical Research: Atmospheres*, *124*(23), 12,380–12,403. https://doi.org/10.1029/2019JD030943

Gettelman, A., & Morrison, H. (2015). Advanced two-moment bulk microphysics for global models. Part I: Off-line tests and comparison with other schemes. *Journal of Climate*, *28*, 1268–1287. https://doi.org/10.1175/JCLI-D-14-00102.1

Ghimire, B., Riley, W. J., Koven, C. D., Mu, M. Q., & Randerson, J. T. (2016). Representing leaf and root physiological traits in CLM improves global carbon and nitrogen cycling predictions. *Journal of Advances in Modeling Earth Systems*, *8*, 598–613. https://doi.org/10.1002/2015MS000538

Golaz, J.-C., Larson, V. E., & Cotton, W. R. (2002). A PDF-based model for boundary layer clouds. Part I: Method and model description. *Journal of the Atmospheric Sciences*, *59*, 3540–3551.

Goldberg, D. N. (2011). A variationally derived, depth-integrated approximation to a higher-order glaciological flow model. *Journal of Glaciology*, *57*, 157–170. https://doi.org/10.3189/002214311795306763

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

Hemer, M. A., Wang, X. L., Weisse, R., & Swail, V. R. (2012). Advancing wind-waves climate science: The COWCLIP project. *Bulletin of the American Meteorological Society*, *93*, 791–796.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., et al. (2018). Historical (1750-2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geoscientific Model Development*, *11*, 369–408. https://doi.org/10.5194/gmd-11-369-2018

Hoose, C., Kristjánsson, J. E., Chen, J.-P., & Hazra, A. (2010). A classical-theory-based parameterization of heterogeneous ice nucleation by mineral dust, soot, and biological particles in a global climate model. *Journal of the Atmospheric Sciences*, *67*, 2483–2503. https://doi.org/10.1175/2010JAS3425.1

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

Huang, X., Tang, Q., Tseng, Y., Hu, Y., Baker, A. H., Bryan, F. O., et al. (2016). P-CSI v1.0, An accelerated barotropic solver for the high-resolution ocean model component in the Community Earth System Model v2.0. *Geoscientific Model Development*, *9*, 4209–4225. https://doi.org/10.5194/gmd-9-4209-2016

Huffman, G. J., Adler, R. F., Bolvin, D. T., & Gu, G. (2009). Improving the global precipitation record: GPCP Version 2.1. *Geophysical Research Letters*, *36*, L17808. https://doi.org/10.1029/2009GL040000

Hunke, E. C., Hebert, D. A., & Lecomte, O. (2013). Level-ice melt ponds in the Los Alamos sea ice model, CICE. *Ocean Modelling*, *71*, 26–42. https://doi.org/10.1016/j.ocemod.2012.11.008

Hunke, E. C, Lipscomb, W. H., Turner, A. K., Jeffery, N., & Elliott, S. (2015). CICE: The Los Alamos Sea Ice Model. Documentation and Software User's Manual. Version 5.1. *T-3 Fluid Dynamics Group*, Los Alamos National Laboratory, Tech. Rep. LA-CC-06-012.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., et al. (2013). The Community Earth System Model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, *94*, 1339–1360. https://doi.org/10.1175/BAMS-D-12-00121.1

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, *113*, D13103. https://doi.org/10.1029/2008JD009944

Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., & Morice, C. P. (2012). Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research*, *117*, D05127. https://doi.org/10.1029/2011JD017139

Kaufmann, R. K., Kauppi, H., Mann, M. L., & Stock, J. H. (2011). Reconciling anthropogenic climate change with observed temperature 1998–2008. *Proceedings of the National Academy of Sciences*, *108*, 11790–11793. https://doi.org/10.1073/pnas.1102467108

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The Community Earth System Model (CESM) Large Ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*, 1333–1349. https://doi.org/10.1175/BAMS-D-13-00255.1

Kennedy, D., Swenson, S., Oleson, K. W., Lawrence, D. M., Fisher, R., da Costa, A. C. L., & Gentine, P. (2019). Implementing plant hydraulics in the Community Land Model, Version 5. *Journal of Advances in Modeling Earth Systems*, *11*, 485–513. https://doi.org/10.1029/2018ms001500

Kiehl, J. T., & Gent, P. R. (2004). The Community Climate System Model, version 2. *Journal of Climate*, *17*, 3666–3682.

Knutti, R., Masson, D., & Gettelman, A. (2013). Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters*, *40*, 1194–1199. https://doi.org/10.1002/grl.50256

Kosaka, Y., & Xie, S.-P. (2013). Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature*, *501*, 403–407. https://doi.org/10.1038/nature12534

Koven, C. D., Riley, W. J., Subin, Z. M., Tang, J. Y., Torn, M. S., Collins, W. D., et al. (2013). The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4. *Biogeosciences*, *10*, 7109–7131. https://doi.org/10.5194/bg-10-7109-2013

Kummerow, C., Simpson, J., Thiele, O., Barnes, W., Chang, A. T., Stocker, E., et al. (2000). The status of the Tropical Rainfall Measuring Mission (TRMM) after two years in orbit. *Journal of Applied Meteorology*, *39*, 1965–1982. https://doi.org/10.1175/1520-0450(2001)040<1965:TSOTTR>2.0.CO;2

Large, W. G., McWilliams, J. C., & Doney, S. C. (1994). Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Reviews of Geophysics*, *32*, 363–403.

Larson, V. E., (2017). CLUBB-SILHS: A parameterization of subgrid variability in the atmosphere. arXiv:1711.03675v2 [physics.ao-ph].

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., et al. (2019). The Community Land Model Version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, *11*, 4245–4287. https://doi.org/10.1029/2018MS001583

Le Quéré, C., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., Peters, G. P., et al. (2016). Global Carbon Budget 2016. *Earth System Science Data*, *8*, 605–649. https://doi.org/10.5194/essd-8-605-2016

Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos Transactions American Geophysical Union*, *89*, 93–94.

Lenssen, N., Schmidt, G., Hansen, J., Menne, M., Persin, A., Ruedy, R., & Zyss, D. (2019). Improvements in GISTEMP uncertainty model. *Journal of Geophysical Research: Atmospheres*, *124*, 6307–6326. https://doi.org/10.1029/2018JD029522

Letscher, R. T., Moore, J. K., Teng, Y. C., & Primeau, F. (2015). Variable C:N:P stoichiometry of dissolved organic matter cycling in the Community Earth System Model. *Biogeosciences*, *12*, 209–221. https://doi.org/10.5194/bg-12-209-2015

Levis, S., Badger, A., Drewniak, B., Nevison, C., & Ren, X. L. (2018). CLMcrop yields and water requirements: Avoided impacts by choosing RCP 4.5 over 8.5. *Climatic Change*, *146*, 501–515. https://doi.org/10.1007/s10584-016-1654-9

Levitus, S., T. Boyer, M. Concright, D. Johnson, T. O'Brien, J. Antonov, et al. (1998). Introduction, Vol. I, World Ocean Database 1998, NOAA Atlas NESDIS 18, 346 pp.

Li, F., & Lawrence, D. M. (2017). Role of fire in the global land water budget during the twentieth century due to changing ecosystems. *Journal of Climate*, *30*, 1893–1908. https://doi.org/10.1175/JCLI-D-16-0460.1

Li, F., Levis, S., & Ward, D. S. (2013). Quantifying the role of fire in the Earth system—Part 1: Improved global fire modeling in the Community Earth System Model (CESM1). *Biogeosciences*, *10*, 2293–2314. https://doi.org/10.5194/bg-10-2293-2013

Li, H. Y., Wigmosta, M. S., Wu, H., Huang, M. Y., Ke, Y. H., Coleman, A. M., & Leung, L. R. (2013). A physically based runoff routing model for land surface and Earth system models. *Journal of Hydrometeorology*, *14*, 808–828. https://doi.org/10.1175/Jhm-D-12-015.1

Li, Q., & Fox-Kemper, B. (2017). Assessing the effects of Langmuir turbulence on the entrainment buoyancy flux in the ocean surface boundary layer. *Journal of Physical Oceanography*, *47*, 2863–2886.

Li, Q., Fox-Kemper, B., Breivik, O., & Webb, A. (2017). Statistical models of global Langmuir mixing. *Ocean Modelling*, *113*, 95–114.

Li, Q., Webb, A., Fox-Kemper, B., Craig, A., Danabasoglu, G., Large, W. G., & Vertenstein, M. (2016). Langmuir mixing effects on global climate: WAVEWATCH III in CESM. *Ocean Modelling*, *103*, 145–160. https://doi.org/10.1016/j.ocemod.2015.07.020

Lin, S. J., & Rood, R. B. (1997). An explicit Flux-Form Semi-Lagrangian shallow water model on the sphere. *Quarterly Journal of the Royal Meteorological Society*, *123*, 2477–2498.

Lindsay, K. (2017). A Newton-Krylov solver for fast spin-up of online ocean tracers. *Ocean Modelling*, *109*, 33–43. https://doi.org/10.1016/j.ocemod.2016.12.001

Lipscomb, W. H., Fyke, J. G., Vizcaíno, M., Sacks, W. J., Wolfe, J., Vertenstein, M., et al. (2013). Implementation and initial evaluation of the Glimmer Community Ice Sheet Model in the Community Earth System Model. *Journal of Climate*, *26*, 7352–7371. https://doi.org/10.1175/JCLI-D-12-00557.1

Lipscomb, W. H., Price, S. F., Hoffman, M. J., Leguy, G. R., Bennett, A. R., Bradley, S. L., et al. (2019). Description and evaluation of the Community Ice Sheet Model (CISM) v. 2.1. *Geoscientific Model Development*, *12*, 387–424. https://doi.org/10.5194/gmd-12-387-2019

Liu, X., Ma, P. L., Wang, H., Tilmes, S., Singh, B., Easter, R. C., et al. (2016). Description and evaluation of a new four-mode version of the Modal Aerosol Module (MAM4) within Version 5.3 of the Community Atmosphere Model. *Geoscientific Model Development*, *9*, 505–522. https://doi.org/10.5194/gmd-9-505-2016

Loeb, N. A., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., et al. (2009). Toward optimal closure of the Earth's top-of-atmosphere radiation budget. *Journal of Climate*, *22*, 748–766.

Lombardozzi, D. L., Bonan, G. B., Smith, N. G., Dukes, J. S., & Fisher, R. A. (2015). Temperature acclimation of photosynthesis and respiration: A key uncertainty in the carbon cycle-climate feedback. *Geophysical Research Letters*, *42*, 8624–8631. https://doi.org/10.1002/2015GL065934

Long, M. C., Lindsay, K., & Holland, M. M. (2015). Modeling photosynthesis in sea ice covered waters. *Journal of Advances in Modeling Earth Systems*, *7*, 1189–1206. https://doi.org/10.1002/2015MS000436

Malavelle, F. F., Haywood, J. M., Jones, A., Gettelman, A., Clarisse, L., Bauduin, S., et al. (2017). Strong constraints on aerosol-cloud interactions from volcanic eruptions. *Nature*, *546*, 485–491. https://doi.org/10.1038/nature22974

Mayorga, E., Seitzinger, S. P., Harrison, J. A., Dumont, E., Beusen, A. H. W., Bouwman, A. F., et al. (2010). Global Nutrient Export from WaterSheds 2 (NEWS 2): Model development and implementation. *Environmental Modelling and Software[R]*, *25*, 837–853. https://doi.org/10.1016/j.envsoft.2010.01.007

McWilliams, J. C., & Fox-Kemper, B. (2013). Oceanic wave-balanced surface fronts and filaments. *Journal of Fluid Mechanics*, *730*, 464–490. https://doi.org/10.1017/jfm.2013.348

McWilliams, J. C., & Sullivan, P. (2000). Vertical mixing by Langmuir circulations. *Spill Science and Technology Bulletin*, *6*, 225–237.

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*, 898–902. https://doi.org/10.1038/NCLIMATE2357

Meehl, G. A., Washington, W. M., Arblaster, J. M., Hu, A., Teng, H., Kay, J. E., et al. (2013). Climate change projections in CESM1(CAM5) compared to CESM4. *Journal of Climate*, *26*, 6287–6308. https://doi.org/10.1175/JCLI-D-12-00572.1

Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., et al. (2017). Historical greenhouse concentrations for climate modelling (CMIP6). *Geoscientific Model Development*, *10*, 2057–2116. https://doi.org/10.5194/gmd-10-2057-2017

Mills, M. J., Richter, J. H., Tilmes, S., Kravitz, B., MacMartin, D. G., Glanville, A. A., et al. (2017). Radiative and chemical response to interactive stratospheric sulfate aerosols in fully coupled CESM1(WACCM). *Journal of Geophysical Research: Atmospheres*, *122*, 13,061–13,078. https://doi.org/10.1002/2017JD027006

Mills, M. J., Schmidt, A., Easter, R., Solomon, S., Kinnison, D. E., Ghan, S. J., et al. (2016). Global volcanic aerosol properties derived from emissions, 1990–2014, using CESM1(WACCM). *Journal of Geophysical Research: Atmospheres*, *121*, 2332–2348. https://doi.org/10.1002/2015JD024290

Montzka, S. A., Aydin, M., Battle, M., Butler, J. H., Saltzman, E. S., Hall, B. D., et al. (2004). A 350-year atmospheric history for carbonyl sulfide inferred from Antarctic firn air and air trapped in ice. *Journal of Geophysical Research*, *109*, D22302. https://doi.org/10.1029/2004JD004686

Moore, J. K., Doney, S. C., Kleypas, J. A., Glover, D. M., & Fung, I. Y. (2002). An intermediate complexity marine ecosystem model for the global domain. *Deep Sea Research*, *49*, 403–462. https://doi.org/10.1016/S0967-0645(01)00108-4

Moore, J. K., Doney, S. C., & Lindsay, K. (2004). Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model. *Global Biogeochemical Cycles*, *18*, GB4028. https://doi.org/10.1029/2004GB002220

Moore, J. K., Lindsay, K., Doney, S. C., Long, M. C., & Misumi, K. (2013). Marine Ecosystem Dynamics and Biogeochemical Cycling in the Community Earth System Model [CESM1(BGC)]: Comparison of the 1990s with the 2090s under the RCP4.5 and RCP8.5 scenarios. *Journal of Climate*, *26*, 9291–9312. https://doi.org/10.1175/JCLI-D-12-00566.1

Morrison, H., & Gettelman, A. (2008). A new two-moment bulk stratiform cloud microphysics scheme in the NCAR community atmosphere model (CAM3), Part I: Description and numerical tests. *Journal of Climate*, *21*, 3642–3659.

Neale, R. B., Richter, J. H., & Jochum, M. (2008). The impact of convection on ENSO: From a delayed oscillator to a series of events. *Journal of Climate*, *21*, 5904–5924.

Neely, R. R. III, Marsh, D. R., Smith, K. L., Davis, S. M., & Polvani, L. M. (2014). Biases in Southern Hemisphere climate trends induced by coarsely specifying the temporal resolution of stratospheric ozone. *Geophysical Research Letters*, *41*, 8602–8610. https://doi.org/10.1002/2014GL061627

Neely, R. R. III, & Schmidt, A. (2016). VolcanEESM: Global volcanic sulphur dioxide (SO2) emissions database from 1850 to present—Version 1.0. Centre for Environmental Data Analysis, 04 February 2016, doi: https://doi.org/10.5285/76ebdc0b-0eed-4f70-b89e-55e606bcd568.

Nicolet, M. (1970). Ozone and hydrogen reactions. *Annales de Géophysique*, *26*, 531–546.

Noël, B., van de Berg, W. J., van Wessem, J. M., van Meijgaard, E., van As, D., Lenaerts, J. T. M., et al. (2018). Modelling the climate and surface mass balance of polar ice sheets using RACMO2—Part 1: Greenland (1958–2016). *The Cryosphere*, *12*, 811–831. https://doi.org/10.5194/tc-12-811-2018

Oleson, K. W., & Feddema, J. (2019). Parameterization and surface data improvements and new capabilities for the Community Land Model Urban (CLMU). *Journal of Advances in Modeling Earth Systems.* https://doi.org/10.1029/2018MS001586

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*, 3461–3482.

Otto-Bliesner, B. L., Brady, E. C., & Fasullo, J. (2016). Climate variability and change since 850 CE: An ensemble approach with the Community Earth System Model. *Bulletin of the American Meteorological Society*, *97*, 735–754. https://doi.org/10.1175/BAMS-D-14-00233.1

Paulot, F., Jacob, D. J., Johnson, M. T., Bell, T. G., Baker, A. R., Keene, W. C., et al. (2015). Global oceanic emission of ammonia: Constraints from seawater and atmospheric observations: Ocean Ammonia Emissions. *Global Biogeochemical Cycles*, *29*, 1165–1178. https://doi.org/10.1002/2015GB005106

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., et al. (2016). ERA-20C: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, *29*, 4083–4097. https://doi.org/10.1175/JCLI-D-15-0556.1

Reichl, B. G., Hara, T., & Ginis, I. (2014). Sea state dependence of the wind stress over the ocean under hurricane winds. *Journal of Geophysical Research: Oceans*, *119*(1), 30–51. https://doi.org/10.1002/2013JC009289

Reichl, B. G., & Li, Q. (2019). A parameterization with a constrained potential energy conversion rate of vertical mixing due to Langmuir turbulence. *Journal of Physical Oceanography*. https://doi.org/10.1175/JPO-D-18-0258.1

Richter, J. H., & Rasch, P. J. (2008). Effects of convective momentum transport on the atmospheric circulation in the Community Atmosphere Model, Version 3. *Journal of Climate*, *21*, 1487–1499.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., et al. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of Climate*, *24*(14), 3624–3648. https://doi.org/10.1175/JCLI-D-11-00015.1

Roach, L. A., Horvat, C., Dean, S. M., & Bitz, C. M. (2018). An emergent sea ice floe size distribution in a global coupled ocean-sea ice model. *Journal of Geophysical Research: Oceans*, *123*(6), 4322–4337. https://doi.org/10.1029/2017JC013692

Robert, A. J. (1966). The integration of a low order spectral form of the primitive meteorological equations. *Journal of the Meteorological Society of Japan*, *44*, 237–245.

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., et al. (2013). Berkeley earth temperature averaging process. *Geoinformatics & Geostatistics: An Overview*, *1*. https://doi.org/10.4172/gigs.1000103

Schmidt, A., Mills, M. J., Ghan, S., Gregory, J. M., Allan, R. P., Andrews, T., et al. (2018). Volcanic radiative forcing from 1979 to 2015. *Journal of Geophysical Research: Atmospheres*, *123*, 12491–12508. https://doi.org/10.1029/2018JD028776

Schweiger, A., Lindsay, R., Zhang, J., Steele, M., & Stern, H. (2011). Uncertainty in modeled Arctic sea ice volume. *Journal of Geophysical Research*, *116*, C00D06. https://doi.org/10.1029/2011JC007084

Scinocca, J., & Mcfarlane, N. (2000). The parametrization of drag induced by stratified flow over anisotropic orography. *Quarterly Journal of the Royal Meteorological Society*, *126*, 2353–2394. https://doi.org/10.1256/smsqj.56801

Sheng, J.-X., Weisenstein, D. K., Luo, B.-P., Rozanov, E., Stenke, A., Anet, J., et al. (2015). Global atmospheric sulfur budget under volcanically quiescent conditions: Aerosol-chemistry-climate model predictions and validation. *Journal of Geophysical Research: Atmospheres*, *120*, 256–276. https://doi.org/10.1002/2014JD021985

Shi, M., Fisher, J. B., Brzostek, E. R., & Phillips, R. P. (2016). Carbon cost of plant nitrogen acquisition: Global carbon cycle impact from an improved plant nitrogen cycle in the Community Land Model. *Global Change Biology*, *22*, 1299–1314. https://doi.org/10.1111/gcb.13131

Shi, X., Liu, X., & Zhang, K. (2015). Effects of preexisting ice crystals on cirrus clouds and comparison between different ice nucleation parameterizations with the Community Atmosphere Model (CAM5). *Atmospheric Chemistry and Physics*, *15*, 1503–1520. https://doi.org/10.5194/acp-15-1503-2015

Small, R. J., Curchitser, E., Hedstrom, K., Kauffman, B., & Large, W. G. (2015). The Benguela upwelling system: Quantifying the sensitivity to resolution and coastal wind representation in a global climate model. *Journal of Climate*, *28*, 9409–9432. https://doi.org/10.1175/JCLI-D-15-0192.1

Smeed, D. A., Josey, S. A., Beaulieu, C., Johns, W. E., Moat, B. I., Frajka-Williams, E., et al. (2018). The North Atlantic Ocean is in a state of reduced overturning. *Geophysical Research Letters*, *45*, 1527–1533. https://doi.org/10.1002/2017GL076350

Smith, R., P. Jones, B. Briegleb, F. Bryan, G. Danabasoglu, J. Dennis, et al. (2010). The Parallel Ocean Program (POP) reference manual, Ocean component of the Community Climate System Model (CCSM), LANL Tech. Report, LAUR-10-01853, 141 pp.

Solomon, S., Rosenlof, K. H., Portmann, R. W., Daniel, J. S., Davis, S. M., Sanford, T. J., & Plattner, G.-K. (2010). Contributions of stratospheric water vapor to decadal changes in the rate of global warming. *Science*, *327*, 1219–1223.

Steele, M., Morley, R., & Ermold, W. (2001). PHC: A global ocean hydrography with a high quality Arctic Ocean. *Journal of Climate*, *14*, 2079–2087.

Stocker, T. F., Qin, D., Plattner, G. K., Tignor, M., Allen, S. K., Boschung, J., et al. (2013). *IPCC, 2013: Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change* (1535 pp.). Cambridge, UK and New York, NY, USA: Cambridge University Press.

Subin, Z. M., Riley, W. J., & Mironov, D. (2012). An improved lake model for climate simulations: Model structure, evaluation, and sensitivity analyses in CESM1. *Journal of Advances in Modeling Earth Systems*, *4*, M02001. https://doi.org/10.1029/2011MS000072

Sun, Q., Whitney, M. M., Bryan, F. O., & Tseng, Y.-H. (2019). Assessing the skill of the improved treatment of riverine freshwater in the Community Earth System Model (CESM) relative to a new salinity climatology. *Journal of Advances in Modeling Earth Systems*, *11*(5), 1189–1206. https://doi.org/10.1029/2018MS001349

Suzuki, N., & Fox-Kemper, B. (2016). Understanding Stokes forces in the wave-averaged equations. *Journal of Geophysical Research: Oceans*, *121*, 1–18. https://doi.org/10.1002/2015JC011566

Swenson, S., & Lawrence, D. (2014). Assessing a dry surface layer-based soil resistance parameterization for the Community Land Model using GRACE and FLUXNET-MTE data. *Journal of Geophysical Research: Atmospheres*, *119*, 10299–10312. https://doi.org/10.1002/2014JD022314

Swenson, S., & Lawrence, D. (2015). A GRACE-based assessment of interannual groundwater dynamics in the Community Land Model. *Water Resources Research*, *51*(11), 8817–8833. https://doi.org/10.1002/2015WR017582

Swenson, S. C., & Lawrence, D. (2012). A new fractional snow-covered area parameterization for the Community Land Model and its effect on the surface energy balance. *Journal of Geophysical Research*, *117*, D21107. https://doi.org/10.1029/2012JD018178

Swenson, S. C., Lawrence, D. M., & Lee, H. (2012). Improved simulation of the terrestrial hydrological cycle in permafrost regions by the Community Land Model. *Journal of Advances in Modeling Earth Systems*, *4*, M08002. https://doi.org/10.1029/2012MS000165

Thiery, W., Visser, A. J., Fischer, E. M., Hauser, M., Hirsch, A. L., Lawrence, D. M., et al. (2020). Warming of hot extremes alleviated by expanding irrigation. *Nature Communications*, *11*, 290. https://doi.org/10.1038/s41467-019-14075-4

Tilmes, S., Hodzic, A., Emmons, L. K., Mills, M. J., Gettelman, A., Kinnison, D. E., et al. (2019). Climate forcing and trends of organic aerosols in the Community Earth System Model (CESM2). *Journal of Advances in Modeling Earth Systems*, *11*. https://doi.org/10.1029/MS001827

Tilmes, S., Richter, J. H., Kravitz, B., MacMartin, D. G., Mills, M. J., Simpson, I. R., et al. (2018). CESM1(WACCM) Stratospheric Aerosol Geoengineering Large Ensemble (GLENS) project. *Bulletin of the American Meteorological Society*, *99*(11), 2361–2371. https://doi.org/10.1175/BAMS-D-17-0267.1

Tolman, H. L. (2009). User manual and system documentation of WAVEWATCH III TM version 3.14. Technical note, MMAB Contribution, 276, p.220.

Turner, A. K., & Hunke, E. C. (2015). Impacts of a mushy-layer thermodynamic approach in global sea-ice simulations using the CICE sea ice model. *Journal of Geophysical Research: Oceans*, *120*, 1253–1275. https://doi.org/10.1002/2014JC010358

van Kampenhout, L., Lenaerts, J. T. M., Lipscomb, W. H., Sacks, W. J., Lawrence, D. M., Slater, A. G., & van den Broeke, M. R. (2017). Improving the representation of polar snow and firn in the Community Earth System Model. *Journal of Advances in Modeling Earth Systems*, *9*, 2583–2600. https://doi.org/10.1002/2017/MS000988

van Kampenhout, L., Rhoades, A. M., Herrington, A. R., Zarzycki, C. M., Lenaerts, J. T. M., Sacks, W. J., & van den Broeke, M. R. (2019). Regional grid refinement in an Earth system model: Impacts on the simulated Greenland surface mass balance. *The Cryosphere*. https://doi.org/10.5194/tc-13-1547-2019

van Marle, M. J. E., Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A.-L., Field, R. D., et al. (2017). Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fire models (1750–2015). *Geoscientific Model Development*, *10*, 3329–3357. https://doi.org/10.5194/gmd-10-3329-2017

van Wessem, J. M., van de Berg, W. J., Noël, B. P. Y., van Meijgaard, E., Amory, C., Birnbaum, G., et al. (2018). Modelling the climate and surface mass balance of polar ice sheets using RACMO2—Part 2: Antarctica (1979–2016). *The Cryosphere*, *12*, 1479–1498. https://doi.org/10.5194/tc-12-1479-2018

Vizcaíno, M., Lipscomb, W. H., Sacks, W. J., van Angelen, J. H., Wouters, B., & van den Broeke, M. R. (2013). Greenland surface mass balance as simulated by the Community Earth System Model. Part I: Model evaluation and 1850-2005 results. *Journal of Climate*, *26*, 7793–7812. https://doi.org/10.1175/JCLI-D-12-00615.1

Vose, R. S., Arndt, D., Banzon, V. F., Easterling, D. R., Gleason, B., Huang, B., et al. (2012). NOAA's merged land-ocean surface temperature analysis. *Bulletin of the American Meteorological Society*, *1677-1685*. https://doi.org/10.1175/BAMS-D-11-00241.1

Wang, Y., Liu, X., Hoose, C., & Wang, B. (2014). Different contact angle distributions for heterogeneous ice nucleation in the Community Atmospheric Model version 5. *Atmospheric Chemistry and Physics*, *14*, 10411–10430. https://doi.org/10.5194/acpd-14-10411-2014

Washington, W. M., & Meehl, G. A. (1989). Climate sensitivity due to increased CO2: Experiments with a coupled atmosphere and ocean general circulation model. *Climate Dynamics*, *4*, 1–38.

Wieder, W. R., Lawrence, D. M., Fisher, R. A., Bonan, G. B., Cheng, S. J., Goodale, C. L., et al. (2019). Beyond static benchmarking: Using experimental manipulations to evaluate land model assumptions. *Global Biogeochemical Cycles*, *33*(10), 1289–1309. https://doi.org/10.1029/2018GB006141

Williams, P. D. (2011). The RAW filter: An improvement to the Robert-Asselin filter in semi-implicit integrations. *Monthly Weather Review*, *139*, 1996–2007.

Worby, A. P., Geiger, C., Paget, M. J., van Woert, M., Ackley, S. F., & DeLiberty, T. (2008). Thickness distribution of Antarctic sea ice. *Journal of Geophysical Research*, *113*, C05S92. https://doi.org/10.1029/2007JC004254

World Meteorological Organization (2010). *Scientific assessment of ozone depletion, WMO Report*. Geneva: World Meteorological Organization.

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

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