The Community Earth System Model version 2 (CESM2)


Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO, USA

Atmospheric Chemistry Observations and Modeling Laboratory, National Center for Atmospheric Research, Boulder, CO, USA

Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder, Boulder, CO, USA

Institute for Marine and Atmospheric Research Utrecht, Utrecht University, The Netherlands

Computational and Information Systems Laboratory, National Center for Atmospheric Research, Boulder, CO, USA

Department of Earth, Environmental, and Planetary Sciences, Brown University, Providence, RI, USA

Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA

Department of Physics, University of Toronto, Toronto, ON, Canada

Department of Earth System Science, University of California Irvine, Irvine, CA, USA

Applied Mathematics / Earth and Environmental Sciences, Columbia University, New York, NY, USA

Atmospheric Science and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA

Corresponding author: Gokhan Danabasoglu (gokhan@ucar.edu)
Key Points:

- Community Earth System Model version 2 includes many substantial science and infrastructure improvements since its previous version.

- Pre-industrial control and historical simulations were performed with low-top and high-top comprehensive chemistry atmospheric models.

- Comparisons to observations are improved relative to previous versions, including major reductions in radiation and precipitation biases.

Keywords: Community Earth System Model (CESM); Global coupled Earth system modeling; Pre-industrial and historical simulations; Coupled model evaluation
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 pre-industrial 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” (140 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 short-wave 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 observations. 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. 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 (CMIP6).
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, to the addition of wind-driven waves on the model's ocean surface. The solutions from a large set of simulations that include integrations for the pre-industrial 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, starting with the Climate System Model [CSM; Boville and Gent, 1998], the first coupled climate model to achieve a stable present-day control simulation without any so-called flux corrections which were required to stabilize coupled simulations even in the absence of changed radiative forcing. This was followed by the Community Climate System Model version 2 [CCSM2; Kiehl and 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 it as among the
best climate models in the world [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 CESM 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 pre-industrial (PI) control simulation; an abrupt quadrupling of CO₂ concentration simulation; a 1% per year CO₂ 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 more than 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 were made available in December 2018 (CESM2.1.0) and in June 2019 (CESM2.1.1) from which many of these simulations can be run as out-of-the-box configurations.

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 encountered during the development process and the strategies adopted to address them. Initial conditions and forcing datasets 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 the historical
simulations that include variables from many component models are presented in Sections 7-11. These CESM2 highlights or improvements 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 (ECS). Finally, Section 13 provides a summary and discussion. 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 (Fig. 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.

Figure 1. Schematic of the CESM2 component models.
2.1 Atmosphere

The Community Atmosphere Model version 6 (CAM6) uses the same Finite Volume (FV) dynamical core [Lin and Rood, 1997] as in CCSM4 and CESM1, but it includes many changes to the existing representation of physical processes [Neale et al., personal communication]. 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]. CLUBB is a direct replacement for the separate shallow-convection, boundary layer, and grid-scale condensation schemes present in CAM5 [Bogenschutz et al., 2013]. It is a continuous, high-order closure representation of moist turbulence with a novel use of simple, assumed binormal probability density functions (PDFs) describing subgrid scale variations in key temperature, humidity, and vertical velocity fields. The Morrison-Gettelman cloud microphysics scheme [MG1; Morrison and Gettelman, 2008] has been replaced by an updated version [MG2; Gettelman and 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 following Hoose et al. [2010] rather than just temperature. 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 scheme (TMS), 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 [Bacmeister et al., personal communication] 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 with a model top at 3.6 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 (WACCM6) in CESM2 is configured identically to CAM6, except that it uses 70 vertical levels with a model top at 6x10^-6 hPa (about 140 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 (SO2). As described by Gettelman et al. [2019a], 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 short-
wave 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 sub-diurnal
coupling of the ocean model [Robert, 1966; Asselin, 1972; Williams, 2011]; and subsequently,
use of a one-hour 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 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 5500 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 3500 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 co-limitation (N, P, Si, and Fe), 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 CaCO3 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 and Martiny, 2015], burial of material
at the seafloor that matches riverine inputs in the pre-industrial climate, both semi-labile 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 dataset. 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 through 2005. Outside this period, the fluxes are held constant using the closest temporal value [Long et al., personal communication].

Version 3.14 of the NOAA WaveWatch-III ocean surface wave prediction model [Tolman, 2009] is incorporated in CESM2 as a new component. 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]. The wave model is solved on a coarse resolution, near-global grid (78.4°S – 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. 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 zone [Roach et al., 2018], wave effects on the mean flow that become important at mesoscale-permitting ocean model resolutions [e.g., McWilliams and Fox-Kemper, 2013; Suzuki and 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
[McWilliams and Sullivan, 2000; Li et al., 2016], 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 and Fox-Kemper, 2017; Reichl and 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 and 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 [Löfverström et al., personal
communication; Muntjewerf et al., personal communication]. 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; van Kampenhout et al., 2019a]. 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) 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, co-limitation, and temperature acclimation); soil hydrology improvements (frozen soil hydrology updates, spatially-explicit soil depth, dry surface layer control on soil evaporation, and updated groundwater scheme); snow model updates (new snow cover fraction parameterization, new fresh snow density parameterization based on temperature and wind resulting in more realistic denser snow packs, 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 snow pack allowing for firm formation); 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; a completely updated lake model; vertically-resolved soil biogeochemistry scheme that allows for simulation of permafrost carbon; 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; inclusion of a new fire model that includes representation of natural and anthropogenic fire triggers and suppression as well as agricultural, deforestation, and peat fires; multiple urban classes and updated urban energy model; 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, 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; 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. 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. Further descriptions, assessments, and characterization of CLM5 compared to prior CLM versions in land-only configurations are provided in van Kampenhout et al. [2017], Lawrence et al. [2019], Wieder et al. [2019], Fisher et al. [2019], Kennedy et al. [2019], Bonan et al. [2019], and Oleson and Feddema [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; Li et al., 2013]. In standard CESM2 configurations, the MOSART grid resolution is 0.5° x 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$^{-1}$), MOSART additionally
simulates time-varying main channel flow velocity (m s$^{-1}$) 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 dataset [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 re-written coupling infrastructure
that includes modularity, multi-instance capabilities for data assimilation, and new support for
land-ice coupling. It also includes re-written 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 datasets 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 minutes. 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 minutes. 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 4320 processors on the
Cheyenne Supercomputer, corresponding to about 3450 cpu hours per simulation year. The
CESM2(WACCM6) version has a throughput of about 4 simulated years per day on 3780
processors on the same machine, corresponding to about 22700 cpu hours per simulation year,
roughly 7 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: 1152 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 which 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 in with WACCM6, so fewer processors are needed for the ocean to keep pace.

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 (LS) 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 LS 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 LS ice cover. The more extensive LS 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 LS 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 fifty 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%; LS SSTs that remain around freezing point with salinities < 34 psu; curtailed LS 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 LS ice cover after a sufficiently long simulation (see Section 4) to provide ocean and sea-ice initial conditions going forward.

Figure 2. 30-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].
The second unacceptable feature was an unrealistic cooling of global-mean surface temperatures in the historical simulations (see Fig. 5, 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 pre-industrial 1850s. This behavior first appeared when the new CMIP6-based emissions datasets were applied in our simulations. The same model versions with CMIP5-based emissions did not cool in this way. The observational time series (see Fig. 5, 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 datasets (which were resolved in the final release) and their application to CAM6, particularly for anthropogenic sulfur. Simulations with the corrected datasets did show some reductions of the unrealistic cooling, but the cooling behavior remained, particularly during the mid-1940s-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. The modifications are described in detail by Gettelman et al. [2019b]. 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. First, a 3-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 LS 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 LS 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 CESM2(CAM6) and CESM2(WACCM6) PI control simulations. Thus, in total, these ocean
potential temperature and salinity initial conditions represent states integrated for about 1070 years after initialization starting from the PHC2 datasets.

The land and 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 repeatedly cycled through multiple years of surface forcing that were extracted from simulation 297. Twenty-one years of forcing from this coupled model run were used, in order to capture some aspects of interannual variability. The land model was spunup 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-yr 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; Lindsay et al., personal communication].
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 (Fig. 3). The 3 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.

Figure 3. Schematic of the CESM2 PI control and historical simulations. CESM2(WACCM6) PI simulation was started first (black) to obtain the forcing datasets 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.
5  Forcing Datasets for Pre-Industrial and Historical Simulations

CESM2(CAM6) PI control and historical simulations used several forcing datasets 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 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 et al. [2019a] and Emmons 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) H2O production rates due to CH4 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. 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 datasets to perform the CESM2(CAM6) simulations is provided in Fig. 3.

The tropospheric oxidant species (O₃, OH, NO₃, and HO₂) 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 mid-month 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 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 VolcanEASM database [Neely and Schmidt, 2016] are used. In addition, observed carbonyl sulfide concentrations are specified as time-varying boundary conditions. As shown in Fig. 4, this leads to a realistic simulation of the stratospheric aerosol optical depth (SAOD) compared to the CMIP6 data which 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 that persists during volcanically quiescent periods. Historical variability in this background is largely due to variability in the source gas carbonyl sulfide (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.

Figure 4. Time evolution of the globally-averaged stratospheric aerosol optical depth (SAOD) as
simulated by CESM2(WACCM6) (3-member average, black) and CMIP6 database (red). There
is uncertainty in the time, location, and intensity of volcanic emissions, especially further back in
time.

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 (CH₄) 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 CH₄ molecule results in the production of two molecules of H₂O [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 CFCs, in CESM2(CAM6), they are specified as a global average mixing ratio derived from the lower boundary conditions provided by CMIP6. CESM2(WACCM6) uses CFC-11*, which is CFC-11eq minus the species that are advected by WACCM6.

Updated land cover and land use data have been created for CLM5. The new dataset 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 dataset is provided in
Lawrence et al. [2019] and references therein.

6 Pre-Industrial Control Spin-up and Tuning

The CESM2(CAM6) and CESM2(WACCM6) PI control simulations were integrated for 1200
and 500 years, respectively. In such long control simulations, it is necessary to aim for top-of-
the-atmosphere (TOA) global- and time-mean heat fluxes of $<|0.1|$ W m$^{-2}$ to avoid
unrealistically warm or cold climate states towards the end of these simulations that will be
subsequently used to initialize historical integrations. To achieve this desired level of TOA heat
balance, the coupled model was tuned using two sets of adjustments. The first involved very
minor changes in the so-called “gamma” parameter in CLU BB, which controls the width of the
assumed Gaussian PDF for subgrid vertical velocities. Wider PDFs for these vertical velocities
with a large gamma value imply 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 result in 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 snow melt was lowered from $-1.0^\circ$C to $-1.5^\circ$C and the maximum
melting snow grain radius size was increased from 1000 microns to 1500 microns, both
adjustments again with respect to those used in CESM1 simulations. These adjustments are
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 in the control simulations and remained fixed at these values in all subsequent experiments.

The impacts of these tunings were assessed in decadal to multi-decadal length simulations prior to the start of the main PI control simulations. The CESM2(CAM6) and CESM2(WACCM6) integrations used identical tuning parameters, and no changes or corrections were made throughout the respective control integrations. The global-mean TOA heat flux imbalances are +0.05 and +0.06 W m$^{-2}$ in 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 simulations, this heat gain is solely reflected in the ocean model, as the land and sea-ice components show minor heat losses. The global-mean heat 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 temperature increases from its initial value of 3.92°C to about 4.23°C by year 1200 in CESM2(CAM6). Similarly, it increases from 3.92°C 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 x 10$^{-5}$ PSU century$^{-1}$ in both CESM2(CAM6) and CESM(WACCM6). This salinity trend is very similar to those of CCSM4 and CESM1(LENS) PI control simulations.
7 Surface Temperatures

The model global-mean surface temperature anomaly time series for the historical period are presented in Fig. 5 in comparison with observations. A reference period of 1850-1870 is chosen to easily identify warming since the pre-industrial 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°C and 0.4°C at the end of the integration period in CESM2(CAM6) and CESM2(WACCM6) simulations, respectively. Although the warming trends between 1975-2000 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., Solomon et al., 2010; Kaufmann et al., 2011; Kosaka and Xie, 2013; Meehl et al., 2014]. 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.

Figure 5. Time series of the global-mean surface 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-member 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 datasets. The latter two time series are referenced to the mean of HADCRU45 for the 1870-1890 period.

The 1985-2014 average SST distributions from CESM2(CAM6), CESM2(WACCM6), and CESM1(LENS) ensemble-means are presented in Fig. 6 in comparison to the Extended Reconstructed Sea Surface Temperature version 5 dataset [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) solutions are almost identical in their mean SSTs. 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 essentially reflect the SST differences in the corresponding PI control simulations from the ERSSTv5 dataset for the pre-industrial 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, i.e., 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].
Figure 6. 1985-2014 average sea surface temperature (in °C) from (top) ERSST 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.
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 7 shows the 1985-2014 average 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 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\(^{-1}\) in CESM1(LENS) down to 0.25 and 0.23 mm day\(^{-1}\) in CESM2(CAM6) and CESM2(WACCM6), respectively. The rms error also decreases substantially by about 20\%, from 0.87 mm day\(^{-1}\) in CESM1(LENS) to 0.71 mm day\(^{-1}\) 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 US 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.
Figure 7. 1985-2014 average total precipitation (in mm day\(^{-1}\)) 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.
8.2 Short-wave cloud forcing

A dramatic improvement is seen in the radiative properties and, in particular, the time-averaged short-wave cloud forcing (Fig. 8). In comparison to the Clouds and the Earth’s Radiant Energy System – Energy Balanced and Filled [CERES-EBAF; Loeb et al., 2009] dataset, the reflective biases of the clouds in CESM1(LENS) throughout the tropics over land and ocean, often in excess of −30 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 re-adjustment 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 rms 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.
Figure 8. 2000-2013 average short-wave cloud forcing (in W m$^{-2}$) (a) from the CERES-EBAF dataset and (b-d) for model minus observation differences. Only one ensemble member is used for model simulations.

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 (Fig. 9) \cite{Neale et al., personal communication}. The observation-based Tropical Rainfall Measuring Mission – ECMWF Re-Analysis-Interim (TRMM/ERAI) 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 9. Lag correlation relationship between total precipitation averaged in the Indian Ocean and total precipitation (color) and 850-mb zonal wind (contour) at all Indo-Pacific longitudes from (a) TRMM/ERA datasets and (b-d) model simulations. Based on daily data filtered with a 20-100 day Lanczos filter. Only one ensemble member is used for model simulations.
8.4 El Niño – Southern Oscillation

ENSO is arguably the most important mode of tropospheric climate variability on interannual timescales. 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 [Richter and Rasch, 2008; Neale et al., 2008; Deser et al., 2012]. These improvements are retained in CESM2. Figure 10 compares spatial composites of December through February (DJF) anomalies of precipitation, surface air temperature over land and SST over ocean, and sea level pressure (SLP) from observations, CESM2(CAM6), CESM2(WACCM6), and CESM1(LENS). Again, CESM2(CAM6) and CESM2(WACCM6) solutions are very similar to each other with the exception of a stronger ENSO amplitude in the former, and the spatial patterns are generally well simulated by all model versions. However, there are two notable improvements in CESM2(CAM6) and CESM2(WACCM6) with respect to CESM1(LENS). First, the orientation of SLP 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).

---

**Figure 10.** Spatial composites of December through February (DJF) ENSO anomalies of precipitation (top row; in mm day\(^{-1}\)), 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), CESM2(WACCM6), and CESM1(LENS). The top panels also show the 6 mm day\(^{-1}\) DJF-mean contour. The normalized December Niño 3.4 timeseries subject to a 3-point binomial filter are used to composite all years greater than 1 standard deviation (El Niño) and all years less than -1 standard deviation (La Niña). Observations are from GPCP for precipitation (1979-2017), Extended Reconstructed SST version 5 (ERSSTv5) for SST (1920-2017), Berkeley Earth Surface Temperatures (BEST) for 2 m air temperature (1920-2017), and ECMWF Re-Analysis for 20\(^{th}\) century (ERA20C) combined with ERAI for PSL (1920-2017).
846 8.5 Stratospheric ozone

847 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 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 11 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 through 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 interannual variability of meteorology in the Southern Hemisphere polar vortex.
Figure 11. October 90°-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 simulations while the ensemble means are shown by the solid lines.

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 12 presents the 1995-2014 average AMOC distributions from
CESM2(CAM6) and CESM2(WACCM6) in comparison to CESM1(LENS). These are for the
Eulerian-mean, i.e., 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 = 10⁶ m³ s⁻¹)
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 trans-basin
transport estimates have become available. Among these, the RAPID array at 26.5°N has the
longest record. For the April 2004 – February 2017 period, the mean and standard deviation of
the estimated NADW maximum transport at 26.5°N are 17.0 ± 3.3 Sv [Smeed et al., 2018;
Frajka-Williams et al., 2019]. 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 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, 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 is about 1-5 Sv at 24.5°N based on sparse datasets 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).

Figure 12. 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 counter-clockwise circulations, respectively.

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 Fig. 13, 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 $28.9 \times 10^3$ 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 $21.4 \times 10^3$ and $15.4 \times 10^3$ 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 $20.2 \times 10^3$ km$^3$. The thicker sea-ice in CESM2(WACCM6) is related to springtime Arctic clouds, which are thicker as a result of increased aerosol concentrations predicted from the interactive chemistry module [Gettelman et al., 2019a; DuVivier et al., personal communication]. The differences in sea-ice thickness 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 (Fig. 13e). 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 $1.6 \times 10^6$ km$^2$, and the ice melt-back is delayed by about one 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 [DuVivier et al., personal communication].

Figure 13. The annual-mean sea-ice thickness 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 for the
(g) Northern Hemisphere and (h) Southern Hemisphere. All values are averages for 1979-2014. The observations are estimates from the PIOMAS dataset. The analysis uses 6, 3, and 40 ensemble members for CESM2(CAM6), CESM2(WACCM6), and CESM1(LENS), respectively.

10.2 Land-ice surface mass balance

Figure 14 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) [Vizcaino 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 observed SMB reasonably well for both ice sheets. Improvements relative to CESM1(CAM4) can be attributed in part to more realistic snow/firm 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 [Lenaerts et al., personal communication]. Remaining biases for Greenland include excessive snowfall in the south, likely associated with unresolved topography [van Kampenhout et al., 2019b], excessive rainfall [van Kampenhout et al., 2019a], and a generally too positive SMB bias in the north.
Figure 14. Surface mass balance (SMB in mm water equivalent yr$^{-1}$) 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.
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 datasets and metrics for each variable and variable-to-variable comparisons, are shown in Fig. 15 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 grey in the left panel of Fig. 15) 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 suggest 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) – e.g., 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 datasets 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 spinup 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 spinup process was supposed to set soil carbon stocks to near zero, but due to very small but non-zero 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.
Figure 15. 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 at http://www.cesm.ucar.edu/experiments/cesm2.0/land/diagnostics/clm_diag_ILAMB.html.

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 CO₂ concentrations. Consequently, in CESM1 land-use and land-cover change
carbon loss fluxes dominate over the CO₂ 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 PgC (Fig. 16). In CESM2(CAM6) and
CESM2(WACCM6), the CO₂ 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 Fig. 15),
which translate to improved annual cycle amplitudes of atmospheric CO₂ concentrations in
emissions-driven simulations.
Figure 16. Land carbon accumulation for the period 1960-2014 for 3 historical ensemble members of CESM2(CAM6), CESM2(WACCM6), and CESM1(LENS) along with observational estimates. Observed estimates are from the Global Carbon Project [Le Quéré et al., 2016]. Grey shading shows uncertainty range for the observational estimates, using an uncertainty of 0.8 Pg C yr\(^{-1}\) and calculated assuming that the errors are additive in time as in Koven et al. [2013].

12 Equilibrium Climate Sensitivity

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 is usually estimated using a slab ocean model (SOM) configuration, because such a configuration equilibrates much faster compared to one with a full-depth ocean model (about 50 vs. > 1000 years). Furthermore, ECS values obtained with a SOM versus a full-depth ocean component are comparable to each other.
[Danabasoglu and Gent, 2009]. Therefore, SOM configurations were used to estimate the ECS of both model versions. The control SOM experiments use a pre-industrial value of CO₂ set at 280 ppmv. These respective control simulations are compared with a second set of SOM experiments in which the CO₂ concentration is instantaneously doubled to 560 ppmv (2xCO₂). All experiments are integrated for > 70 years to ensure that global-mean surface temperature has equilibrated and TOA radiative imbalances are near zero. Differenting the equilibrated surface temperatures (last 30-year averages) between the respective control and 2xCO₂ experiments, the ECS values are computed as 5.3°C for CESM2(CAM6) and 5.1°C for CESM2(WACCM6). These values represent significant increases from an ECS of around 4.0°C in CESM1 [Gettelman et al., 2012]. Values of climate sensitivity derived from fully-coupled 4xCO₂ runs using the Gregory method [Gregory et al., 2004] are consistent with these SOM-based values [Bacmeister et al., personal communication]. The evolution of CESM2's climate sensitivity through several development versions of the model is examined in detail in Gettelman et al. [2019b] as well as Bacmeister et al. [personal communication]. Both of these studies suggest 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.

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 datasets 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 1200-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 US 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 SLP anomaly teleconnection patterns associated with ENSO over the North Pacific; a global land carbon accumulation trend that matches the observed 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 thicker
springtime Arctic clouds that result from increased aerosol concentrations predicted from the
interactive chemistry module.

Figure 17 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 datasets. Indeed, CESM2 simulations have notable
improvements in many of the metrics considered here compared to those of CESM1(LENS).
Figure 17. 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).

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. The latter class of biases can have significant climate impacts, but no robust remedy exists to alleviate them.

The path to finalizing a CESM2 version that produced acceptable solutions 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 datasets, 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 and 0.4°C in CESM2(CAM6) and CESM2(WACCM6), respectively, 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 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). These values represent a
significant increase from an ECS of 4.0°C calculated for CESM1. 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 et al., 2019; also Bacmeister et al.,
personal communication].

The CESM community has been participating in over 20 CMIP6-endorsed MIPs. The solutions
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 [Fasullo, personal communication].
Model and Data Availability

Previous and current CESM versions are freely available at www.cesm.ucar.edu/models/cesm2/.

The CESM solutions / datasets used in this study are also freely available from the Earth System Grid Federation (ESGF) at esgf-node.llnl.gov/search/cmip6 or from the NCAR Digital Asset Services Hub (DASH) at data.ucar.edu or from the links provided from the CESM web site at www.cesm.ucar.edu.

Acknowledgments

The CESM project is supported primarily by the National Science Foundation (NSF). This material is based upon work supported by the National Center for Atmospheric Research (NCAR), which is a major facility sponsored by the NSF under Cooperative Agreement No. 1852977. Computing and data storage resources, including the Cheyenne supercomputer (doi:10.5065/D6RX99HX), were provided by the Computational and Information Systems Laboratory (CISL) at NCAR. We thank all the scientists, software engineers, and administrators who contributed to the development of CESM2. GPCP Precipitation data are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at https://www.esrl.noaa.gov/psd/. The implementation of the WaveWatch-III has benefitted from the input and collaborations with the wave model developers at the NOAA National Centers for Environmental Prediction (NCEP).

Appendix A: Construction of forcing datasets 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 mid-month date of 16 January 1860 are derived from the average of January averages for the 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.

*Table A1. Forcing datasets date information*

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


Frajka-Williams, E., I. J. Ansorge, J. Baehr, H. L. Bryden, M. Paz Chidichimo, S. A.

Cunningham, G. Danabasoglu, S. Dong, K. A. Donohue, S. Elipot, P. Heimbach, N. P.

Holliday, R. Hummels, L. C. Jackson, J. Karstensen, M. Lankhorst, I. A. Le Bras, M. S.

Lozier, E. L. McDonagh, C. S. Meinen, H. Mercier, B. I. Moat, R. C. Perez, C. G.

Piecuch, M. Rhein, M. A. Srokosz, K. E. Trenberth, S. Bacon, G. Forget, G. Goni, D.

Kieke, J. Koelling, T. Lamont, G. D. McCarthy, C. Mertens, U. Send, D. A. Smeed, S.


69


Hemispheric and large-scale land surface air temperature variations: An extensive
revision and an update to 2010, *J. Geophys. Res.*, 117, D05127, doi:


Danabasoglu, J. Edwards, M. Holland, P. Kushner, J.-F. Lamarque, D. Lawrence, K.
Lindsay, A. Middleton, E. Munoz, R. Neale, K. Oleson, L. Polvani, and M. Vertenstein
(2015), The Community Earth System Model (CESM) Large Ensemble project: A
community resource for studying climate change in the presence of internal climate

Gentine (2019), Implementing plant hydraulics in the Community Land Model, version 5,

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

Knutti, R., D. Masson, and A. Gettelman (2013), Climate model genealogy: Generation CMIP5

Kosaka, Y., and S.-P. Xie (2013), Recent global-warming hiatus tied to equatorial Pacific surface


Landschützer, N. Lefèvre, A. Lenten, S. Lienert, D. Lombardozzi, J. R. Melton, N.

Metzl, F. Millero, P. M. S. Monteiro, D. R. Munro, J. E. M. S. Nabel, S. Nakaoka, K.


T. Takahashi, H. Tian, B. Tilbrook, I. T. van der Laan-Luijkhx, G. R. van der Werf, N.


Lenssen, N., G. Schmidt, J. Hansen, M. Menne, A. Persin, R. Ruedy, and D. Zyss (2019),

Improvements in the uncertainty model in the Goddard Institute for Space Studies

Surface Temperature (GISTEMP) analysis, *J. Geophys. Res. Atmos.*, doi:

10.1029/2018JD029522 (in press).


do dissolved organic matter cycling in the Community Earth System Model,


NESDIS 18, 346 pp.

Li, H. Y., M. S. Wigmosta, H. Wu, M. Y. Huang, Y. H. Ke, A. M. Coleman, and L. R. Leung

(2013), A physically based runoff routing model for land surface and Earth system


Moore, J. K., S. C. Doney, and K. Lindsay (2004), Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model, Global Biogeochemical Cycles, 18,


Moore, J. K., K. Lindsay, S. C. Doney, M. C. Long, and K. Misumi (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, J. Climate, 26, 9291-9312, doi: 10.1175/JCLI-D-12-00566.1.


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


Williams, P. D. (2011), The RAW filter: An improvement to the Robert – Asselin filter in semi-

Thickness distribution of Antarctic sea ice, *J. Geophys. Res., 113*, C05S92, doi:

World Meteorological Organization (2010), *Scientific Assessment of Ozone Depletion*, WMO

Karspeck, K. Lindsay, M. C. Long, H. Teng, and N. S. Lovenduski (2018), Predicting
near-term changes in the Earth System: A large ensemble of initialized decadal prediction
1867-1886, doi: 10.1175/BAMS-D-17-0098.1.

Zhang, R., R. Sutton, G. Danabasoglu, Y.-O. Kwon, R. Marsh, S. G. Yeager, D. E. Amrhein, and
C. M. Little (2019), A review of the role of the Atlantic meridional overturning