The DOE E3SM Model Version 2: Overview of the physical model

Jean-Christophe Golaz[1], Luke P. Van Roekel[2], Xue Zheng[1], Andrew Roberts[2], Jonathan D Wolfe[2], Wuyin Lin[3], Andrew Bradley[4], Qi Tang[1], Mathew E Maltrud[2], Ryan M Forsyth[1], Chengzhu Zhang[1], Tian Zhou[5], Kai Zhang[5], Charles Sutton Zender[6], Mingxuan Wu[5], Hailong Wang[5], Adrian K Turner[2], Balwinder Singh[5], Jadwiga H. Richter[7], Yi Qin[1], Mark R. Petersen[2], Azamat Mametjanov[8], Po-Lun Ma[5], Vincent E Larson[9], Jayesh Krishna[8], Noel D. Keen[10], Nicole Jeffery[2], Elizabeth C Hunke[2], Walter M. Hannah[1], Oksana Guba[4], Brian M Griffin[9], Yan Feng[8], Darren Engwirda[2], Alan V. Di Vittorio[10], Cheng Dang[11], LeAnn Conlon[2], Chih-Chieh Chen[7], Michael Brunke[12], Gautam Bisht[5], James J Benedict[2], Xylar S Asay-Davis[2], Yuying Zhang[1], Xubin Zeng[12], Shaocheng Xie[1], Phillip Justin Wolfram Jr.[2], Tom Vo[1], Milena Veneziani[2], Teklu Kidane Tesfa[5], Sarat Sreepathi[13], Andrew G. Salinger[4], Michael J. Prather[6], Salil Mahajan[13], Qing Li[14], Philip W Jones[2], Robert L Jacob[8], J E Jack Reeves Eyre[15], Gunther W Huebler[9], Xianglei Huang[16], Benjamin R Hillman[4], Bryce E Harrop[5], James G Foucar[4], Yilin Fang[5], Darin Comeau[2], Peter Martin Caldwell[1], Tony Bartoletti[1], Karthik Balaguru[5], Mark A Taylor[4], Renata McCoy[1], L. Ruby Leung[5], and David Craig Bader[1]

[1]Lawrence Livermore National Laboratory (DOE)
[2]Los Alamos National Laboratory (DOE)
[3]Brookhaven National Laboratory (DOE)
[4]Sandia National Laboratories (DOE)
[5]Pacific Northwest National Laboratory (DOE)
[6]University of California, Irvine
[7]National Center for Atmospheric Research (UCAR)
[8]Argonne National Laboratory (DOE)
[9]University of Wisconsin-Milwaukee
[10]Lawrence Berkeley National Laboratory (DOE)
[11]Joint Center for Satellite Data Assimilation
[12]The University of Arizona
[13]Oak Ridge National Laboratory (DOE)
[14]The Hong Kong University of Science and Technology (Guangzhou)
[15]NOAA NCEP/CPC
[16]University of Michigan-Ann Arbor

November 30, 2022

Abstract

This work documents version two of the Department of Energy's Energy Exascale Earth System Model (E3SM). E3SM version 2 (E3SMv2) is a significant evolution from its predecessor E3SMv1, resulting in a model that is nearly twice as fast and with a simulated climate that is improved in many metrics. We describe the physical climate model in its lower horizontal resolution configuration consisting of 110 km atmosphere, 165 km land, 0.5° river routing model, and an ocean and sea ice with mesh spacing varying between 60 km in the mid-latitudes and 30 km at the equator and poles. The model performance is evaluated by means of a standard set of Coupled Model Intercomparison Project Phase 6 (CMIP6) Diagnosis, Evaluation, and Characterization of Klima (DECK) simulations augmented with historical simulations as well as simulations to evaluate impact of different forcing agents.

The simulated climate is generally realistic, with notable improvements in clouds and precipitation compared to E3SMv1. E3SMv1 suffered from an excessively high equilibrium climate sensitivity (ECS) of 5.3 K. In E3SMv2, ECS is reduced to 4.0 K which is now within the plausible range based on a recent World Climate Research Programme (WCRP) assessment. However, E3SMv2 significantly underestimates the global mean temperature in the second half of the historical record. An analysis of single-forcing simulations indicates that correcting the historical temperature bias would require a substantial reduction in the magnitude of the aerosol-related forcing.

The DOE E3SM Model Version 2: Overview of the physical model

Jean-Christophe Golaz[1], Luke P. Van Roekel[2], Xue Zheng[1], Andrew F. Roberts[2], Jonathan D. Wolfe[2], Wuyin Lin[3], Andrew M. Bradley[4], Qi Tang[1], Mathew E. Maltrud[2], Ryan M. Forsyth[1], Chengzhu Zhang[1], Tian Zhou[5], Kai Zhang[5], Charles S. Zender[6], Mingxuan Wu[5], Hailong Wang[5], Adrian K. Turner[2], Balwinder Singh[5], Jadwiga H. Richter[7], Yi Qin[1], Mark R. Petersen[2], Azamat Mametjanov[8], Po-Lun Ma[5], Vincent E. Larson[9,5], Jayesh Krishna[8], Noel D. Keen[10], Nicole Jeffery[2], Elizabeth C. Hunke[2], Walter M. Hannah[1], Oksana Guba[4], Brian M. Griffin[9], Yan Feng[8], Darren Engwirda[2], Alan V. Di Vittorio[10], Cheng Dang[11,12*], LeAnn M. Conlon[2], Chih-Chieh-Jack Chen[7], Michael A. Brunke[13], Gautam Bisht[5], James J. Benedict[2], Xylar S. Asay-Davis[2], Yuying Zhang[1], Xubin Zeng[13], Shaocheng Xie[1], Phillip J. Wolfram[2], Tom Vo[1], Milena Veneziani[2], Teklu K. Tesfa[5], Sarat Sreepathi[14], Andrew G. Salinger[4], Michael J. Prather[11], Salil Mahajan[14], Qing Li[2,15*], Philip W. Jones[2], Robert L. Jacob[8], J. E. Jack Reeves Eyre[13,16*], Gunther W. Huebler[9], Xianglei Huang[17], Benjamin R. Hillman[4], Bryce E. Harrop[5], James G. Foucar[4], Yilin Fang[5], Darin S. Comeau[2], Peter M. Caldwell[1], Tony Bartoletti[1], Karthik Balaguru[5], Mark A. Taylor[4], Renata B. McCoy[1], L. Ruby Leung[5], David C. Bader[1]

[1]Lawrence Livermore National Laboratory, Livermore, CA, USA
[2]Los Alamos National Laboratory, Los Alamos, NM, USA
[3]Brookhaven National Laboratory, Upton, NY, USA
[4]Sandia National Laboratories, Albuquerque, NM, USA
[5]Pacific Northwest National Laboratory, Richland, WA, USA
[6]Departments of Earth System Science and Computer Science, University of California, Irvine, CA, USA
[7]Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO, USA
[8]Argonne National Laboratory, Lemont, IL, USA
[9]Department of Mathematical Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI, USA
[10]Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[11]Department of Earth System Science, University of California, Irvine, CA, USA
[12]Joint Center for Satellite Data Assimilation, Boulder, CO, USA
[13]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA
[14]Oak Ridge National Laboratory, Oak Ridge, TN, USA
[15]The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, Guangdong, China
[16]NOAA NCEP/CPC, College Park, MD, USA
[17]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA
[*]Current affiliation

Key Points:

- E3SMv2 is nearly twice as fast as E3SMv1 with a simulated climate that is improved in many metrics (e.g. precipitation and clouds).
- Climate sensitivity is substantially lower with a now plausible ECS of 4.0 K (compared to an unlikely value of 5.3 K in E3SMv1).
- E3SMv2 underestimates the warming in the late historical period due to excessive aerosol-related forcing.

Corresponding author: Chris Golaz, golaz1@llnl.gov

Abstract

This work documents version two of the Department of Energy's Energy Exascale Earth System Model (E3SM). E3SM version 2 (E3SMv2) is a significant evolution from its predecessor E3SMv1, resulting in a model that is nearly twice as fast and with a simulated climate that is improved in many metrics. We describe the physical climate model in its lower horizontal resolution configuration consisting of 110 km atmosphere, 165 km land, 0.5° river routing model, and an ocean and sea ice with mesh spacing varying between 60 km in the mid-latitudes and 30 km at the equator and poles. The model performance is evaluated by means of a standard set of Coupled Model Intercomparison Project Phase 6 (CMIP6) Diagnosis, Evaluation, and Characterization of Klima (DECK) simulations augmented with historical simulations as well as simulations to evaluate impact of different forcing agents.

The simulated climate is generally realistic, with notable improvements in clouds and precipitation compared to E3SMv1. E3SMv1 suffered from an excessively high equilibrium climate sensitivity (ECS) of 5.3 K. In E3SMv2, ECS is reduced to 4.0 K which is now within the plausible range based on a recent World Climate Research Programme (WCRP) assessment. However, E3SMv2 significantly underestimates the global mean temperature in the second half of the historical record. An analysis of single-forcing simulations indicates that correcting the historical temperature bias would require a substantial reduction in the magnitude of the aerosol-related forcing.

Plain Language Summary

The U.S. Department of Energy recently released version two of its Energy Exascale Earth System Model (E3SM). E3SMv2 experienced a significant evolution in many of its model components (most notably the atmosphere and sea ice models), and its supporting software infrastructure. In this work, we document the computational performance of E3SMv2 and analyze its ability to reproduce the observed climate. To accomplish this, we utilize the standard Diagnosis and Evaluation and Characterization of Klima (DECK) experiments augmented with historical simulations for the period (1850-2015). We find that E3SMv2 is nearly twice as fast as its predecessor and more accurately reproduces the observed climate in a number of metrics, most notably clouds and precipitation. We also find that the model's simulated response to increasing carbon dioxide (the Equilibrium Climate Sensitivity) is much more realistic. Unfortunately, E3SMv2 underestimates the global mean surface temperature compared to observations during the second half of historical period. Using sensitivity experiments, where forcing agents (carbon dioxide, aerosols) are selectively disabled in the model, we determine that correcting this problem would require a strong reduction in the impact of aerosols.

1 Introduction

The U.S. Department of Energy (DOE) Energy Exascale Earth System Model (E3SM) project (https://e3sm.org) was conceived from the confluence of energy mission needs and disruptive changes in scientific computing technology. E3SM aims to optimize the use of DOE resources to meet the science needs of DOE. The long-term goal of the E3SM project is to address the challenge of actionable predictions of Earth system variability and change, with an emphasis on the most critical scientific questions facing the nation and DOE (Leung et al., 2020).

Version one of E3SM (E3SMv1) was first released in 2018 as a physical climate model with a lower horizontal resolution configuration (110-km atmosphere, 60-to-30 km ocean; Golaz et al., 2019) followed by a higher resolution configuration (25-km atmosphere, 18-to-6 km ocean; Caldwell et al., 2019). The lower resolution configuration served as the starting point for a biogeochemistry configuration (E3SMv1.1; Burrows et al., 2020) and a cryosphere configuration (E3SMv1.2; Comeau et al., 2022).

Version two E3SM is a significant evolution from version one. Herein we describe the changes made in E3SM version 2 (E3SMv2) in each model component and the supporting infrastructure. We further diagnose its performance relative to E3SMv1. E3SMv2 includes significant improvements to component model structure and physical parameterizations. The result is a model that is nearly twice as fast as version one with a simulated climate that is improved in many metrics. Also new to E3SMv2 is the introduction of fully coupled regionally refined mesh (RRM) configurations. Although simulations with the RRM will be the subject of forthcoming manuscripts, the validation herein will provide a benchmark for RRM configurations.

As with E3SMv1, we focus on the physical climate model at lower resolution with a 110 km atmosphere, 165 km land, 0.5° river routing model, and an ocean and sea ice with mesh spacing varying between 60 km in the mid-latitudes and 30 km at the equator and poles. The vertical grids remain the same as in E3SMv1 with 72 layers and a top at approximately 60 km in the atmosphere and 60 layers in the ocean. We focus our analysis on the CMIP6 Diagnosis, Evaluation, and Characterization of Klima (DECK) and historical simulations (Eyring et al., 2016). E3SMv2 DECK simulations reveal a number of improvements in the simulated mean climate and variability: equilibrium climate sensitivity, precipitation, shortwave cloud radiative effects, ozone hole, aerosol absorption and sea ice. Yet despite numerous improvements, a number of important biases remain including a weak Atlantic Meridional Overturning Circulation and an inability to appropriately simulate the historical temperature record. To diagnose the latter bias we conduct an ensemble of simulations following the Detection and Attribution Model Intercomparison Project (DAMIP) protocol (Gillett et al., 2016). Using a decomposition analysis, we find that an overly strong aerosol effect is responsible for this bias and further that if this effect can be reduced, other reductions in regional radiation, temperature, and other biases can be expected. These results also show that even though E3SMv2 has shortcomings, it can still serve as a useful tool for numerous future studies.

E3SM was originally branched from an early developmental version of CESM2 (CESM2; Danabasoglu et al., 2020). The river routing, ocean and sea ice components as well as the atmosphere dynamical core and stratospheric chemistry are now different, while the atmosphere physics, the land model and the coupler retain similarities to CESM2. E3SMv2 is the second release of a CMIP6-class model for E3SM. E3SMv2 also serves as a foundation for additional upcoming configurations targeting DOE applications: (i) an RRM configuration with a high resolution region (25-km atmosphere, 14-km ocean) centered over North America, (ii) a biogeochemistry configuration with interactive carbon, nitrogen and phosphorous cycles and (iii) a cryosphere configuration with RRM over the Southern Ocean and ice-shelf cavities.

We begin in Section 2 with a description of the changes in E3SMv2 for each model component. In Sub-section 2.6 we describe important improvements to energy conservation in the coupled system and our coupled tuning strategy for E3SMv2. Section 3 details computational performance and factors leading to the nearly doubling of throughput. Section 4 details the simulation campaign and analysis of the simulated climate in each portion of the campaign. Section 5 presents an examination of the historical temperature record bias and the potential impact of altering the contribution of aerosols and greenhouse gases on the simulated climate. We end with summary and conclusions in Section 6.

2 Model description

2.1 Atmosphere

2.1.1 Dynamical core

The dynamical core in EAMv2 is greatly upgraded from v1 with a new nonhydrostatic option (not employed here). The dynamical core solves the equations of motion in a rotating reference frame with the shallow atmosphere approximation, hyperviscosity based turbulence closure and the option to apply the hydrostatic approximation. The code is implemented in the High Order Method Modeling Environment (HOMME) (J. Dennis et al., 2005; J. M. Dennis et al., 2011; Evans et al., 2013). The equations are formulated following Taylor et al. (2020) using a terrain following mass based vertical coordinate (Kasahara, 1974; Laprise, 1992). The simulations presented here use the hydrostatic approximation. The nonhydrostatic configuration adds additional prognostic equations for vertical velocity and geopotential height and is used for E3SM's cloud resolving simulations (Caldwell et al., 2021). The prognostic equations consist of the time-reversible adiabatic terms (Taylor et al., 2020), a ∇^4 hyperviscosity (J. M. Dennis et al., 2011; Guba et al., 2014), and a sponge layer at the model top (described below). For the adiabatic terms, we use a structure preserving formulation in order to preserve the discrete Hamiltonian and produce an energetically consistent model.

The horizontal discretization uses the collocated mimetic spectral finite element method from Taylor & Fournier (2010). Within each element the prognostic variables are represented by degree p polynomials with $p = 3$ and order of accuracy $n_p = 4$. EAMv2 uses new separate parameterized physics and dynamics grids. Hannah et al. (2021) describe these grids, the remap algorithms to transfer data between the grids, and the new topography file format to support these grids. The grids are the same as introduced in Herrington et al. (2019), but in EAMv2, the high-order remap method is local to each element except for some halo data for extremal mixing ratio values. Thus, EAMv2's grid remap algorithms work without modification in RRM configurations. While some of the initialization infrastructure is part of the physics infrastructure, the dynamical core provides the remap algorithms. In EAMv2, the pg2 configuration is used, meaning each element has a 2×2 subgrid for a total of four physics columns. Thus, the total number of physics columns in a simulation is 4/9 the number used in EAMv1 for a given element grid. The dynamics grid has an average grid spacing of 110 km, while the physics grid and, as a result, the land grid have an average grid spacing of 165 km.

The vertical discretization uses a Lorenz staggered extension of the mimetic centered difference from Simmons & Burridge (1981). The vertical grid remains the same as in EAMv1 with 72 layers and a top at approximately 60 km. With the vertical staggering, prognostic variables are located at level midpoints, with the exception of the vertical velocity and the geopotential, which are located at level interfaces. For the vertical transport terms, we use a vertically Lagrangian approach adapted from Lin (2004). The timestepping algorithm, unchanged from EAMv1, is the high-CFL 5 stage 3rd order accurate Runge-Kutta method from Guerra & Ullrich (2016).

There are several sources of dissipation in the dynamical core. The ∇^4 hyperviscosity is the largest. It is applied to all prognostic variables and on every model layer. For the model-top sponge layer, we apply a ∇^2 Laplacian operator in the top 6 model layers to all prognostic variables. The strength is proportional to the model layer reference pressure, following Lauritzen et al. (2011). In addition, vertical dissipation is introduced by the monotone vertical remap operator. A smaller amount of dissipation is also generated by the Runge-Kutta timestepping. In EAMv1, we used additional divergence damping in order to control noise when running with realistic topography. This was implemented by separating the hyperviscosity into compressible and rotational components and using a larger hyperviscosity coefficient for the compressible component. EAMv2 has a more accurate pressure gradient formulation which improves the treatment of topography and no longer needs nor uses additional divergence damping.

The dynamical core's passive tracer transport method is a new interpolation semi-Lagrangian (ISL) scheme called Islet (Bradley et al., 2021). A high-order ISL method using the natural Gauss-Lobatto-Legendre (GLL) element-local interpolant is unstable; thus, Islet provides modified element-local interpolation basis functions that obey a necessary condition for stability. EAMv2 uses the lowest-order Islet basis set, the one for $n_p = 4$. Because the model code was frozen before the Islet bases were finalized, the formulation of the $n_p = 4$ stable basis set is slightly different than reported in Bradley et al. (2021), but this difference has essentially no impact. To achieve global mass conservation, shape preservation, and mass-tracer consistency, Islet uses element-local and global versions of the communication-efficient density reconstructor (CEDR) described in Algorithm 3.1 of Bradley et al. (2019). The ISL scheme's time step can be, and in EAMv2 is, longer than the vertical remap time step of the dynamics. In integrating from time t_1 to time t_2, Lagrangian levels at time t_2 are reconstructed from data on the reference grid at times t_1 and t_2. Then horizontal velocity at time t_2 is remapped to the Lagrangian levels. Finally, departure points within each Lagrangian level are computed at time t_1. Then 2D advection within each level can proceed as usual. In this time step configuration, the CEDR must be applied to the 3D data rather than separately to each level because the reconstructed levels do not conserve mass within each level; thus, corrections must be applied among levels as well as within each level. In EAMv2's lower resolution configuration, the vertical remap time step is two times larger than the dynamics time step, and the passive tracer transport is six times larger. Like the rest of the dynamical core, Islet works without modification in RRM configurations.

2.1.2 Updated atmosphere physics

As in EAMv1 (Rasch et al., 2019; Xie et al., 2018), EAMv2 represents subgrid turbulent transport and cloud macrophysics by use of the Cloud Layers Unified By Binormals (CLUBB) parameterization (Golaz et al., 2002; V. E. Larson, 2017). In EAMv2, CLUBB represents all stratiform and shallow cumulus clouds, but not deep convective clouds. CLUBB prognoses various subgrid moments of turbulence, heat content, and moisture, and the moments are used to estimate a multivariate subgrid probability density function (PDF). The PDF is then used to diagnose liquid cloud fraction and cloud liquid water via a saturation adjustment. CLUBB is called immediately before the microphysics.

The main update of CLUBB for EAMv2 is that CLUBB's internal call order has been changed so that CLUBB's subgrid moments are prognosed first, and the PDF is estimated immediately afterward. This leaves a saturation-adjusted state for the microphysics. This call order eliminates the unrealistic pockets of supersaturation that were left for the microphysics to handle in EAMv1. Another update of CLUBB is that its code has been refactored in order to improve computational performance. For instance, arrays were restructured to permit contiguous memory access. Loops were rearranged in order to allow calculations with no data dependencies to be done in parallel. Asymptotic values of functions were approximated analytically in order to avoid the unnecessary calculation of expensive special functions.

The deep convection scheme (G. J. Zhang & McFarlane, 1995, ZM hereafter) in EAMv2 is the same as that in EAMv1, except that ZM adopts two updates described in Xie et al. (2019) to improve its simulated precipitation, in particular the diurnal cycle. The new ZM feature combines the dynamic Convective Available Potential Energy (dCAPE) trigger proposed in Xie & Zhang (2000) with an unrestricted air parcel launch level (ULL) approach used in Y.-C. Wang et al. (2015) (hereafter the dCAPE-ULL trigger). The dCAPE trigger provides a dynamic constraint for preconditioning of convection-favoring environments and prevents CAPE from being released spontaneously. The ULL trigger removes the constraint that convection is always rooted within the boundary layer, as is often assumed in deep convection schemes. Thus, it captures mid-level convection by detecting atmospheric instability above the boundary layer. As shown in Xie et al. (2019), the use of the dCAPE-ULL trigger helps address the "too frequent, too weak" precipitation issue — a long-standing climate model bias — as well as capture the nocturnal elevated convection systems which are often seen downstream of major mountains associated with the propagation of Mesoscale Convective Systems (MCSs) but missed in most climate models including E3SM. It also significantly improves the phase of the diurnal cycle of precipitation over both land and ocean.

After releasing EAMv1, Ma et al. (2022) proposed a set of recalibrated atmospheric parameters in the deep convection scheme, the microphysics scheme, and the CLUBB turbulence and macrophysics scheme (hereafter EAMv1p). Many of these parameter changes have been carried over to EAMv2. A new feature in EAMv1p is the inclusion of surface wind speed enhancements from the gustiness associated with turbulence, shallow and deep convection in the surface flux calculations over land and ocean (Ma et al., 2022; Harrop et al., 2018; Redelsperger et al., 2000).

In the deep convection scheme, the parcel buoyancy considers the subgrid temperature perturbation from the CLUBB scheme in addition to a constant value of 0.8 K used in EAMv1. A new tunable parameter with a default value of 2.0, $zmconv\_tp\_fac$ (see Table A1), is introduced to scale the square root of the CLUBB subgrid temperature variance to be the subgrid temperature perturbation. Additionally, the parameters related to the autoconversion rate, detrained ice cloud effective radius, and cloud fraction in deep convective clouds are reduced, while the parameters related to the downdraft mass flux fraction and the impact of the surface temperature change are enhanced compared to EAMv1.

A number of tunable parameters in the CLUBB scheme have been updated in EAMv1p to improve both stratocumulus and shallow cumulus clouds. Briefly, EAMv1p separated the setting of several damping coefficients at low skewness ($X*a$) and high skewness ($X*b$), recalibrated transition factors between the two regimes ($X*c$), and adjusted parameters controlling the low cloudiness (e.g., mu, $C8$, $C1$, C_k10) to increase stratocumulus clouds and reduce shallow cumulus clouds. To better represent clouds and precipitation in subtropical low cloud regimes, the liquid cloud accretion enhancement factor and the exponent coefficient for liquid cloud autoconversion rate in the microphysics scheme have been updated as well. For ice and mixed-phase clouds, the overly suppressed scaling factor (0.1) for the Wegener–Bergeron–Findeisen (WBF) process in EAMv1 has been updated to be 0.7. The Aitken mode sulfate aerosol size threshold for homogeneous ice nucleation is increased. The minimum subgrid vertical velocity for liquid droplet nucleation is reduced from 0.2 to 0.1 m/s in EAMv2.

Based on atmosphere-only and coupled simulations performed during the tuning process, EAMv2 keeps tunable parameters related to liquid droplet sedimentation, ice particle fall speed, and the lateral entrainment of deep convection the same as EAMv1 instead of EAMv1p (see Table A1 for details).

The effective aerosol radiative forcing ($\mathrm{ERF_{aer}}$) estimated in E3SMv1 is about -1.6 $\mathrm{Wm^{-2}}$ (Golaz et al., 2019), which is relatively large compared to other CMIP6 models (Smith et al., 2020). After applying the EAMv1p parameter tuning proposed by Ma et al. (2022), the simulated magnitude of $\mathrm{ERF_{aer}}$ shortwave and longwave components is reduced significantly, but the change in net $\mathrm{ERF_{aer}}$ is small due to the compensation between longwave and shortwave. Clouds are more susceptible to aerosol perturbations under relatively clean conditions. Based on analysis of developmental configurations (to be documented in a separate work), unrealistically-small cloud droplet number concentrations (e.g., $< 10\,\mathrm{cm^{-3}}$) frequently appeared, especially in mid- and high-latitude regions. As a temporary remedy, a lower bound ($10\,\mathrm{cm}^{-3}$) is applied to the simulated cloud droplet number concentration in EAMv2. Results show that it reduces the net $\mathrm{ERF_{aer}}$ magnitude by 0.3-0.4 $\mathrm{Wm^{-2}}$, which agrees with findings from previous studies (e.g. Hoose et al., 2009). The lower bound value is also consistent with other CMIP6 models (e.g. Mignot et al., 2021). We note however that this is not a cure for the problem. Additional efforts are planned to improve the simulated aerosol and cloud properties in pristine regions and reduce $\mathrm{ERF_{aer}}$ in a more physical manner for future versions of E3SM.

EAMv2 employs the same orographic and non-orographic gravity wave (GW) parameterization as EAMv1, following Richter et al. (2010), which includes separate representation of orographic GWs (McFarlane, 1987), convective GWs (Beres et al., 2004), and GWs generated by frontal systems (Charron & Manzini, 2002). Tunable parameters in the orographic and frontal GW parameterizations remain the same as in EAMv1. In EAMv1, the period of the quasi-biennial oscillation (QBO) in the tropical stratospheric zonal mean wind was only 18 months as compared to 28 months in observations (Richter et al., 2019). In order to arrive at a more realistic representation of the QBO in EAMv2, several combinations of tunable parameters in the Beres et al. (2004) parameterization were explored, focusing on the convective fraction (CF) and efficiency with which convection generates GWs, *effgw_beres*, starting with the setting that improved the QBO in EAMv1 described in Richter et al. (2019) (*effgw_beres*=0.35 from 0.4, CF=8% from 5%). Based on sensitivity simulations performed in parallel with the pre-industrial spinup simulation, CF was changed from 8% to 10% (*gw_convect_hcf* = 1/CF = 10), and *effgw_beres* remained 0.35 (Table A1), resulting in a QBO period of $\sim$ 21 months in the pre-industrial control. Due to changes in tropical variability (Kelvin and mixed-Rossby gravity waves) related to the convective parameterization changes described above, the amplitude of the QBO in E3SMv2 is weaker than in observations.

2.1.3 Atmospheric chemistry

The atmospheric chemistry in EAMv1 was the O3v1 model with prognostic stratospheric ozone by the linearized chemistry (Linoz v2) (Hsu & Prather, 2009) and the prescribed tropospheric ozone with the v1.0 input4MIPS ozone data set (Hegglin et al., 2016). The prescribed tropospheric ozone data only contained decadal monthly zonal climatology of latitude-pressure values. Due to the sharp cross-tropopause ozone gradient, unphysical ozone distributions was simulated in the vicinity of the tropopause when the modelled tropopause was higher than that of the prescribed data, assigning stratospheric ozone abundances to the tropospheric model grid boxes. Since ozone interacts with the radiation transfer code in E3SM, such ozone deficiencies impacted the solar heating and radiative forcing.

In EAMv2, we implemented the O3v2 model (Tang et al., 2021) to overcome the limitations in the O3v1 model by replacing the prescribed ozone data with a passive ozone tracer in the troposphere. Ozone is transported from the stratosphere into the troposphere and decays within the lowest four model layers (below 1 km) with a 48-hour e-folding to 30 ppb (parts per billion by mole fraction). The choice of 30 ppb is based on observations (Ziemke et al., 2019) and gives a tropospheric ozone mass similar to full chemistry models. O3v2 is capable of interacting with the tropopause changes and hence captures the naturally sharp ozone cross-tropopause gradient. Moreover, the ozone sink at the lower boundary in O3v2 allows us to diagnose the stratosphere-troposphere exchange flux of ozone, an important tropospheric ozone budget term, which was not possible with O3v1. The ozone hole is simulated following Cariolle et al. (1990) to represent the rapid chlorine-induced ozone depletion at cold temperatures, but the polar stratospheric cloud (PSC) temperature threshold is increased to 197.5 K in the EAMv2 from 193 K in the EAMv1 due to a warmer Antarctic winter pole. More details about O3v2 in E3SM are documented by Tang et al. (2021).

2.1.4 Aerosol

The aerosol model in EAMv2 is based on EAMv1 (H. Wang et al., 2020) which itself evolved from the four mode version of Modal Aerosol Module (MAM4) in CAM5.3 (Liu et al., 2016) that represents the major aerosol species within four internally mixed size modes, and incorporated the new treatments of aerosol processes related to new particle formation, secondary organic aerosol formation, aerosol convective transport and wet removal, resuspension, and deposition and mixing with snow grain. These new treatments in EAMv1 led to significant improvements in characterizing global distributions of aerosols and interactions with clouds and radiation. The development and evaluation of aerosol representation in the E3SMv1 coupled model simulations with both standard resolution (Golaz et al., 2019) and high resolution (Caldwell et al., 2019) configurations have mainly focused on the global budgets and annual mean constraints of aerosol optical depth (AOD) with observational estimates in the present-day conditions. While the total/speciated AOD and direct radiative effects are constrained to a large extent, further analysis of E3SMv1 simulations suggested that the shortwave absorption of aerosols is too strong in the model compared with observations especially over the dusty regions (Feng et al., 2022). The heating effect in the atmosphere due to the overestimated dust absorption could lead to changes of the lower tropospheric stability and affect the model-simulated clouds and precipitation.

In E3SMv2, we updated dust refractive indices in the shortwave bands with the observationally derived values from the AERONET measurements (Dubovik et al., 2000), which replace the strongly absorbing dust properties used in E3SMv1 (Hess et al., 1998). Additionally, we implemented a different dust particle size distribution (Kok, 2011) in E3SMv2 for calculating fractional dust emission fluxes into the accumulation and coarse modes. Kok et al. (2017) suggests that dust size distributions at emission in current global climate models under-represent the coarse-mode (>1 μm) dust particles in the atmosphere. For the same dust optical depth, coarse-mode dust particles would result in larger longwave warming and less shortwave cooling than the fine particles, resulting in a less cooling net effect of dust aerosols. Compared to E3SMv1, the new size distribution implemented to E3SMv2 (Kok, 2011) predicts more particles in larger dust sizes: about 1.1% in the accumulation mode and 98.9% in the coarse mode, which is consistent with the recent measurements (Kok et al., 2017) but can substantially change the dust transport to remote regions (Wu et al., 2020). With these updates in E3SMv2, dust emissions are re-tuned for the globally constrained dust optical depth of 0.03 ± 0.005 (Ridley et al., 2016).

2.2 Ocean

Due to development priorities, the ocean component (the Model for Prediction Across Scales-Ocean: MPAS-Ocean) in E3SMv2 is mostly unchanged from E3SMv1 (Petersen et al., 2018, 2019). The underlying spatial discretization (Thuburn et al., 2009) is applied to the primitive equations with a free surface (Ringler et al., 2013), with 60 layers using a z-star vertical coordinate (Petersen et al., 2015; Reckinger et al., 2015). The parameterizations of unresolved physics, such as the Gent & Mcwilliams (1990) parameterization for mesoscale eddy transport, and K-Profile Parameterization (KPP, Large et al., 1994; Van Roekel et al., 2018) for vertical mixing, remain largely the same with minor update.

In E3SMv2, Redi isopycnal mixing is introduced following the triad formulation from Griffies et al. (1998). The Gent-McWilliams mesoscale eddy mixing parameterization continues to utilize a globally constant value for the bolus coefficient. However, in development of E3SMv2, a series of sensitivity simulations were conducted to find a more optimal value of the Gent-McWilliams bolus kappa parameter. Based on these simulations, a value of 900 $\mathrm{m^2\,s^{-1}}$ was chosen. This value is half of that used in E3SMv1. The reduction improved the surface salinity bias and increased Antarctic Circumpolar Circulation (ACC) transport (not shown). A smaller globally constant value (400 $\mathrm{m^2\,s^{-1}}$) is utilized for Redi isopycnal mixing. In the Redi parameterization, slope tapering is a slightly modified version of Danabasoglu & Williams (1995) with a critical slope parameter of 0.01. We also implemented the stratification-based tapering from Danabasoglu & Marshall (2007).

In addition to the improvements in model physics, a sign error in the high order reconstruction of tracer values on cell edges, was discovered in the flux corrected tracer transport advection scheme. A set of simulations was conducted to determine the impact of this bug. The percentage change in ocean heat content (OHC) due to the bug fix at various levels is shown in Fig. S1. In a broad sense the effect of the bug fix was to increase ocean heat content, although there are broad swaths of decrease in OHC in the North Pacific and ACC in the upper ocean (Figs. S1a-c). In the deep ocean, the bug fix resulted in weaker OHC anomalies (Fig. S1d).

2.3 Sea ice

Improvements have been made to the column physics, coupling, and analysis of E3SM's sea ice component (MPAS-Seaice) since the E3SMv1 configuration described by Turner et al. (2021). Here we expand on innovations new to E3SMv2. The core Delta-Eddington radiative transfer of Briegleb & Light (2007) has been updated to the Dang et al. (2019) SNICAR-AD model, ensuring radiative consistency across all snow surfaces, including on land, ice sheets and sea ice. The SNICAR-AD radiative transfer code includes five-band snow single-scattering properties, two-stream Delta-Eddington approximation with the adding–doubling technique, and parameterization for correcting the near-infrared (NIR) snow albedo biases when solar zenith angle exceeds 75° (Dang et al., 2019). However, radiative coupling with the atmosphere still integrates across just two bands (visible and NIR) separated at 700nm, which does not fully exploit the five-band capability; an expansion of the coupling bands is planned for E3SMv3.

A new snow-on-sea ice-morphology has been added to E3SMv2 that includes the effects of wind redistribution: losses to leads and meltponds, and the piling of snow against ridges. Snow grain radius, now a prognosed tracer field on sea ice, evolves according to temperature gradient and wet snow metamorphisms and feeds back to the SNICAR-AD radiative model up to a dry maximum of 2800 μm. Fresh snow falls at a grain radius of 54.5 μm, and five vertical snow layers replace the previous single snow layer atop each of the five sea ice thickness categories retained from E3SMv1. The combined default configurations of the new radiative and snow schemes were found to minimally impact the climatic state of sea ice reported in this manuscript, but greater parametric sensitivities are explored in a sister paper in preparation.

The most significant improvement to the sea ice climate since E3SMv1 was achieved with coupling changes associated with mushy-layer thermodynamics. Whereas the basal temperature of the ice was held fixed at -1.8 °C in E3SMv1, the new version of the model assumes the mushy liquidus basal temperature from the sea ice as described by Turner & Hunke (2015). Conversion of frazil ice from MPAS-Ocean with a fixed reference salinity of 4 PSU to the mushy layer now conserves to computational accuracy over a 500-year control integration. This was achieved by exchanging additional mass between the upper ocean and sea ice model to accommodate an assumed 25% mushy liquid content assumed from heat and mass transferred adiabatically from the MPAS-Ocean frazil scheme active from a depth of 100 m. In addition to achieving perfect heat and mass conservation between sea ice and ocean models, this improvement greatly reduces a negative sea ice thickness bias in the summer Arctic reported by Golaz et al. (2019) for E3SMv1; it only minimally impacts Southern Ocean sea ice mass that was better simulated as compared to northern hemisphere sea ice in E3SMv1. Note that E3SM does not use virtual ice-ocean fluxes, but instead full volume and heat flux exchange consistent with a Boussinesq ocean model as described by Campin et al. (2008).

In addition to these core physics improvements, E3SMv2 includes a number of structural additions to the sea ice model. E3SMv2 has significantly increased output to better diagnose behavior and compare against seasonal extremes and data. For example, daily Ice Numerals for Arctic shipping (Aksenov et al., 2017) are easily derived from this output, commensurate with the new E3SMv2 marine mesh that resolves major Arctic shipping channels (Section 2.4).

E3SMv2 now also includes a prescribed-extent ice mode for MPAS-Seaice based on that found in the Community Ice CodE (CICE) in E3SMv1 and CESM (Bailey et al., 2011). This mode is needed for AMIP (Atmospheric Model Intercomparison Project) style simulations where a full prognostic sea ice model is not desired but sea ice surface fluxes, albedos, snow depth, and surface temperature are needed by the atmosphere model and are calculated by the vertical thermodynamics module of the sea ice component. The mode is intended for atmosphere sensitivity experiments and does not conserve energy or mass. In this mode, sea ice thermodynamics is active but sea ice dynamics are disabled and at each time step ice area and thickness are reset to specified values. Ice area is interpolated in time and space from an input data set, while ice thickness in grid cells containing sea ice is set to 2 m in the Northern hemisphere and 1 m in the Southern hemisphere. During each adjustment snow volume is adjusted to preserve the snow thickness prognosed in the previous time step. Snow temperatures are reset to the surface temperature, as prognosed in the previous time step, while ice temperatures are set so that the ice temperature gradient is linear, with the ice temperature at the top equal to the prognosed surface temperature, and equal to the sea freezing temperature at the base of the ice. The vertical ice salinity profile is reset to the profile from Bitz & Lipscomb (1999).

2.4 Unstructured marine mesh generation

Generation of the unstructured Centroidal Voronoi-type meshes (e.g. Ringler et al., 2008) used in the ocean and sea ice components of E3SMv2 is handled using the JIGSAW library (Engwirda, 2017), enabling the creation of complex, variable-resolution meshes to resolve regional ocean (Hoch et al., 2020), sea ice (Turner et al., 2021) and land-ice (Hoffman et al., 2018) dynamics. Compared to E3SMv1, improvements to the robustness, efficiency, and flexibility of our meshing workflows has been targeted — employing a multi-paradigm mesh generation strategy that combines ‘off-centre’ Delaunay-refinement and ‘hill-climbing’ optimization approaches (Engwirda & Ivers, 2016; Engwirda, 2018) to build the Spherical Centroidal Voronoi Tessellations (SCVTs) used in the MPAS-Ocean and MPAS-Seaice dynamical cores. Key to improved robustness in E3SMv2 is the elimination of invalid grid configurations centered around obtuse triangles, in which a lack of geometrical consistency between adjacent computational cells would lead to breakdowns in the numerical discretization used by the ocean dynamical core. Difficulties associated with the generation of valid meshes limited the application of variable mesh resolution in E3SMv1, restricting model configurations to quasi-uniform resolution cases. These effects are remedied in E3SMv2, with our enhanced optimization strategies leading to the generation of valid, well-conditioned meshes in complex, regionally-refined configurations. Equally important are improvements to E3SM’s COMPASS (Configuration Of MPAS Setups) package — a Python-based scripting environment that allows modelers to readily customize mesh and model configurations based on proximity to geographic features, climatological state, and user-defined inputs, with geometric tuning parameters that are easy to adjust on the fly. COMPASS tracks mesh provenance data associated with the creation of each new E3SM configuration to support model regression testing and ensure long-term reproducibility. Overall, improvements to the unstructured meshing workflows in E3SMv2 has led to significantly improved turnaround in the mesh design, simulation, and analysis process, reducing the time required to complete various MPAS mesh-related tasks from days-to-weeks in E3SMv1 to minutes-to-hours in E3SMv2. As a consequence of these improvements, the E3SMv2 coastline is more realistic across the globe. As one example, E3SMv2 includes key shipping routes in the Canadian Archipelago that were missing from E3SMv1 (Figure 1), eliciting improved archipelagic through-flow.

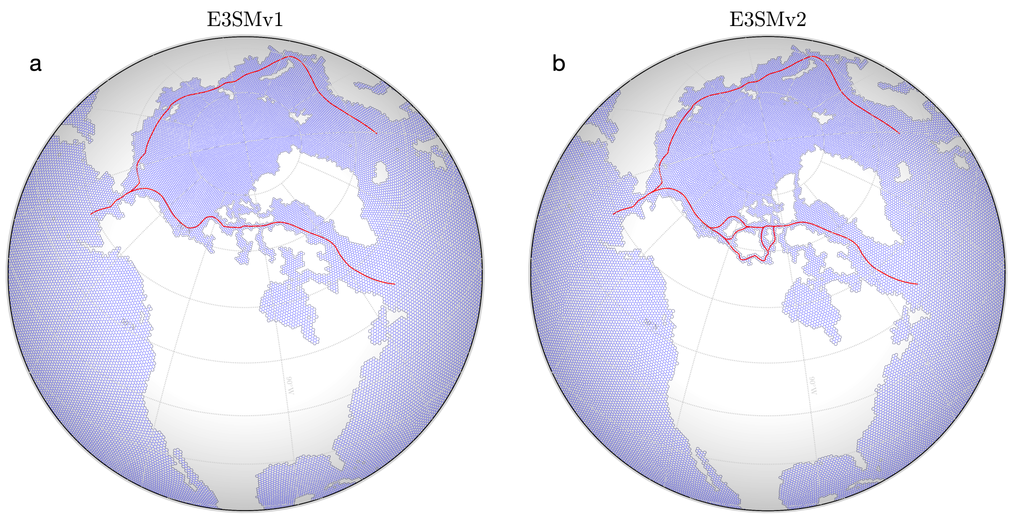


Figure 1. Comparison of the (a) old and (b) new standard resolution E3SM unstructured marine mesh, highlighting improved geographic acuity in E3SMv2 including Arctic coastal shipping channels fitting standard routes published by the Arctic Council (2009) (red).

2.5 Land and river

The physics configuration of E3SM Land Model version 2 (ELMv2) used in E3SMv2 is similar to E3SMv1 (Golaz et al., 2019). ELMv2 simulates hydrologic and thermal processes in vegetation, snow, and soil for different land cover types, which include bare soils, vegetated surfaces, lakes, glaciers, and urban areas. Present-day leaf area index (LAI) is prescribed using satellite data and photosynthesis and is not limited by leaf nutrients. The prescribed vegetation distribution has been updated for E3SMv2 to resolve inconsistencies across platforms in translating land use to changes in plant functional types. ELMv2 includes the new shortwave radiation model SNICAR-AD for snow also used for sea ice as described in section 2.3.

The river routing component in E3SMv2 (Model for Scale Adaptive River Transport, MOSARTv2) takes the runoff produced by ELM and routes it to the river mouth as freshwater input to the ocean component. The physics scheme and configuration is the same as used in E3SMv1 standard resolution (Golaz et al., 2019). Specifically, MOSARTv2 uses the kinematic wave approach to route streamflow across hillslopes, tributaries, and main river stems on an eight-direction-based river network (Li et al., 2013) at 0.5° latitude-longitude spatial resolution.

There are a number of new features developed in ELMv2 and MOSARTv2 since the release of v1 that were not activated in E3SMv2. The soil erosion model of Tan et al. (2018) has been implemented and simulations showed that 5% of the newly fixed land organic carbon in the continental United States (CONUS) is displaced annually by soil erosion (Tan et al., 2020). OpenACC directives were added in ELM to include support for GPUs and a 1km ELM simulation over the CONUS was successfully performed on the Oak Ridge National Laboratory's Summit supercomputer (D. Wang et al., 2020). The plant hydraulics model of Kennedy et al. (2019) has been implemented to more mechanistically account for water stress on vegetation. MOSARTv2 now includes a two-way irrigation scheme which allows the irrigation in ELMv2 to be constrained by the surface water availability calculated by MOSARTv2 (Zhou et al., 2020). The surface water made available for the irrigation includes the water storage in river channels and the reservoirs estimated by a water management scheme introduced in MOSARTv2 (Voisin et al., 2013). MOSARTv2 also includes a newly-developed flood inundation scheme which adds a floodplain storage defined by the local Digital Elevation Model (DEM) along the main river channel (Luo et al., 2017). Once activated, this scheme allows the water exchanges between main river channel and floodplain and thus outputs the inundated fraction for each gridcell. Although inactive in E3SMv2, the impact of these new features have been evaluated in separate studies described above and will be evaluated as a whole in fully-coupled simulation campaigns planned in the future.

2.6 Coupled system

As in E3SMv1, the coupler/driver for E3SMv2 is cpl7 (Craig et al., 2012). The driver of cpl7 performs the integration of the coupled model and provides the "main" for the single executable. cpl7 relies on the Model Coupling Toolkit (MCT; J. Larson et al., 2005) for inter-component communication and remapping operations.

2.6.1 Mapping weights

The remapping operations are performed using mapping weights precomputed by external tools for each grid pair using two different algorithms. Nearly all maps in both directions use the TempestRemap conservative, monotone map (Ullrich & Taylor, 2015; Ullrich et al., 2016). In the case of the atmosphere's pg2 grid and the ocean's Voronoi grid, TempestRemap implements an L^2 projection between the finite-volume grids. The requirement of monotonicity implies the projection must use the constant-function basis rather than a high-order reconstruction. This map type is used for all fluxes and most states in the coupled model. The second map type is bilinear interpolation from ESMF (Hill et al., 2004). This map type is used to transfer state from the atmosphere to the ocean and sea ice.

2.6.2 Energy conservation

EAM and its predecessor CAM (Neale et al., 2012) are designed around the principle that each parameterization conserves energy. Therefore, the change of energy in the atmosphere should be equal to the difference in net fluxes at the top of the model and the surface. A long-term average of the energy change should be close to zero if the model conserves energy, since storage in the atmosphere is minimal.

EAMv1 contained a few energy leaks. For example, one source of leaks is the presence of a limiter for water forms (K. Zhang et al., 2018), but this source is small. In EAMv1, we recognized the gravity wave drag (GWD) parameterization as the source of the largest energy leak. In the orographic gravity waves parameterization, the change in kinetic energy was not properly accounted for. After a fix, the energy imbalance in the atmosphere is reduced from 0.07 $\mathrm{Wm^{-2}}$ to 0.01 $\mathrm{Wm^{-2}}$. Figure S2 depicts energy imbalance for atmosphere simulations with and without the GWD energy fix.

MPAS-Ocean utilizes a fixed two band exponential formulation for penetrating shortwave radiation. For grid cells with shallow bottom depths, a portion of the penetrating shortwave radiation reaches the bottom of the ocean. In E3SMv1, this portion of the shortwave radiation was not accounted for, resulting in a globally averaged energy leak of approximately 0.25 $\mathrm{Wm^{-2}}$. In E3SMv2, the shortwave radiation that reaches the bottom of the ocean is added to the bottom layer. In development of E3SMv2, we found that this change had minimal impact on the large scale ocean climate.

After these energy conservation errors in the atmosphere and the ocean were addressed, we realized that the coupled system was no longer in energy balance compared to E3SMv1. Further investigation led to the energy correction term incorporated in E3SMv1 to account for the inconsistent definition of energy in the ocean and atmosphere (see Golaz et al., 2019, Appendix A). While conceptually correct, the computation of that correction term was based on all the precipitation, when instead it should have included only precipitation over ocean and ocean runoff. Precipitation over land should not have been included because the land model ELM does not take into account heat carried by precipitation. The energy imbalance was corrected by calculating the needed energy to bring fluxes of water to a common temperature with the ocean, and then pass the globally averaged value as a correction term to be applied in the atmosphere every coupling time step.

2.6.3 Coupled tuning

The coupled tuning objectives for the pre-industrial control simulation were similar to Golaz et al. (2019):

1. Near-zero long-term average net top-of-atmosphere (TOA) energy flux and total ocean heat content (OHC) in equilibrium.
2. Minimum long-term drift in global mean surface air temperature.
3. Reasonable absolute global mean surface air temperature.

Furthermore, spatial root meam square error (RMSE) against observations for key climate variables (e.g., annual mean SST, annual and seasonal precipitation, TOA radiation, cloud radiative effect, sea surface wind stress, etc.) from the E3SM Diagnostic package (C. Zhang et al., 2022) are also considered. Tuning was performed iteratively at component levels and with the coupled system under perpetual pre-industrial (1850) forcings.

In the atmosphere, we conducted short atmosphere sensitivity tests with repeating SST and sea ice annual cycle ("F2010") to estimate the impact of individual parameters on the modeled precipitation, cloud radiative forcing and other climate state variables. Promising atmospheric configurations were then evaluated with longer Atmospheric Model Intercomparison Project (AMIP) simulations (prescribed SST for year 1980-2015) before being tested in pre-industrial coupled mode. Results from the coupled simulation then fed back into another round of atmospheric tuning.

Periodically, we also performed atmospheric simulations to evaluate cloud feedback and aerosol ERF to inform the atmospheric tuning. Specifically, we estimated the cloud feedback using Cess-like simulations (Cess et al., 1989) by comparing the differences between an 11-year AMIP standard simulation (year 1980-1990) and the same simulation except with globally +4K SST (Ringer et al., 2014). The aerosol ERF was estimated with time slice simulations (e.g. Hansen, 2005) consisting of a 9-year 2010 simulation vs a 2010 simulation except with 1850 aerosol emissions. To estimate the aerosol ERF more efficiently, we also used short (1 year after 3-month spin-up) nudged simulations with 2010 and 1850 aerosol emissions (all other external forcings kept as year 2010 conditions), where the horizontal winds were nudged towards model output from a baseline simulation. Previous studies (K. Zhang et al., 2022; S. Zhang et al., 2022) showed good agreement in the global and regional annual mean aerosol ERF estimates between the free-running and nudged simulations in E3SMv1.

Component-level development and tuning for the ocean also relied on simulations forced with atmospheric reanalyses (Tsujino et al., 2018) to guide the tuning of the Gent-McWilliams bolus kappa parameter and the newly implemented Redi isopynal mixing scheme.

As in E3SMv1, the last step was a final tuning of the CLUBB parameter $clubb\_c14$ in the coupled system to minimize long-term drift by adjusting shortwave cloud radiative effects (SWCRE) in the low-cloud regimes.

Pre-industrial simulations were the only coupled simulations performed before the model was frozen. In particular, no idealized CO_2 or test historical simulations were performed before finalizing E3SMv2.

3 Computational performance

3.1 Performance comparison of v1 and v2 simulations

This section examines computational performance using a set of atmosphere-only and fully coupled simulations. Relative to EAMv1, EAMv2 is approximately twice as efficient for primarily two reasons: faster passive tracer transport and fewer physics columns. E3SMv2 is also approximately twice as efficient because the ocean dynamics time step is three times larger than in E3SMv1. The sea ice component is slower in v2 than in v1 because of additional snow layers.

Performance benchmark simulations were performed on the ANL Chrysalis cluster. Chrysalis has 512 compute nodes. Each node of the cluster has two AMD Epyc 7532 "Rome" 2.4 GHz processors, and each processor has 32 cores, for a total of 64 cores per node. Each node has 256GB 16 channel DDR4 3200MHz memory. The interconnect hardware is Mellanox HDR200 InfiniBand and uses the fat tree topology. The model code was compiled with Intel release 20200925 with GCC version 8.3.1 compatibility and run with OpenMPI 4.1.1 provided in the Mellanox HPC-X Software Toolkit.

All throughput values reported in this section are derived using the maximum time (minimum throughput) over all MPI processes. Only the total throughput value is fully accurate, as it is computed using the top-level wallclock time of the simulation, excluding initialization; component and subcomponent throughput values are approximations because these lower-level timers are not associated with global synchronization points. The simulations are run with one MPI process per core and no OpenMP threading. A throughput data point corresponds to one simulation run for three months with the default input/output (I/O) configuration and one restart file at the simulation end. For these tests, both v1 and v2 simulations use the new SCORPIO (Software for Caching Output and Reads for Parallel I/O) I/O library; thus, performance differences in these simulations are due to components' computational and I/O volume differences rather than I/O library differences. Performance improvements from SCORPIO are documented separately in Section 3.2.

Figure 2 summarizes the performance of E3SMv2 relative to E3SMv1 on the lower resolution E3SMv1 and E3SMv2 pre-industrial control simulations. Figure 2a plots total throughput versus the number of computer nodes. The models provide a small number of optimized layouts, available using the names XS (v2 only), S, M, L. In addition, the figure shows small-node-count simulations using a simple stacked layout ("st") in which each component runs serially with respect to the others, and all components share the same processors. Each simulation's data point is annotated with its throughput in simulated years per day (SYPD) and layout. Comparing S, M, and L layouts between models, v2 is at least 1.97 times more efficient than v1. Figure 2b illustrates this efficiency difference by plotting the throughput-resource product for each component as a rectangle for the L layouts. The atmosphere (ATM), sea ice (ICE), coupler (CPL), land (LND),

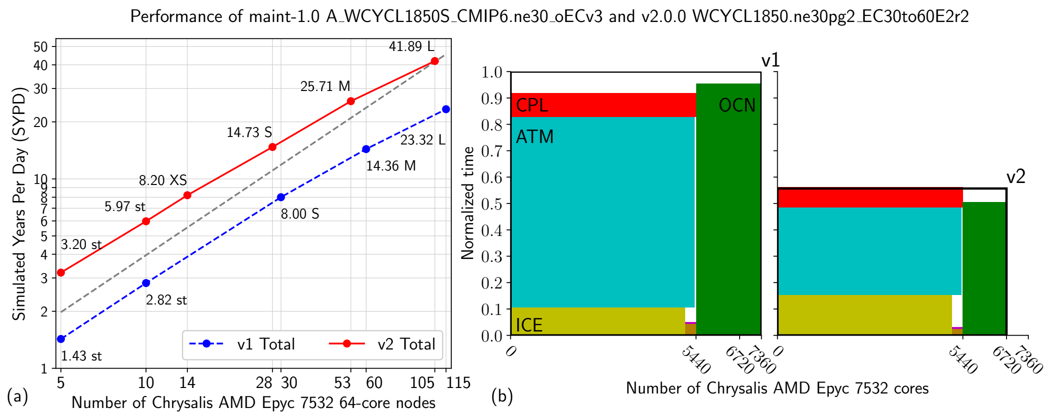


Figure 2. Performance of the lower resolution E3SMv1 and E3SMv2 pre-industrial control simulations. (a) Throughput vs. number of nodes. PE layouts XS, S, M, L are provided as part of the models. Points annotated with "st" use a simple stacked layout in which each component runs serially with respect to the others, and all components share the same processors. (b) Throughput-resource product plots. Each component has one rectangle. A rectangle has the area given by the product of throughput and number of nodes. In v2, the atmosphere and ocean components have substantially smaller throughput-resource products.

and river runoff (ROF; LND and ROF are too small to label) components run on one set of nodes, while the ocean (OCN) component runs on another set. An unfilled rectangle having "v1" or "v2" at the top-right corner shows the total product; because the throughput value of each component is approximate, the filled rectangles do not sum to the total throughput value.

Figure 3 focuses on just the atmosphere component using prescriped SST and sea ice simulations. In E3SMv2, by default MPAS-Seaice now replaces CICE in such configurations (see Section 2.3). However, we use CICE for this study for three reasons. First, MPAS-Seaice requires a partition file for each process decomposition, and one goal of this study is to run simulations with a large number of decompositions. With CICE, we do not need to generate a decomposition file for each one. Second, MPAS-Seaice is slower than CICE, and it must run on an MPAS grid; the combined slowdown from each of these would reduce the precision of our analysis of just the atmosphere component's performance in this study. Finally, v1 must use CICE, so a comparison of just the changes to the atmosphere component is best done by using CICE in the v2 simulations as well.

Figure 3a shows total throughput of the simulation and approximate throughputs of the dynamical core ("dycore") and passive tracer transport. A subset of data points are annotated with throughput values. Passive tracer transport is at least six to at least eight times faster in v2 than in v1. Two details are apparent in this plot. First, the dynamical core is sensitive to the element decomposition, while the rest of the model is sensitive to the finer physics column decomposition. Thus, between 43 and 68 nodes, performance of the dynamical core subcomponents plateaus or slightly degrades, since in this range an increase in node count provides no improvement to the most-burdened MPI processes. Nonetheless, total throughput is roughly monotonically increasing even in this node count range. Second, representative node counts are chosen to favor, generally separately, v1 and v2 in roughly equal numbers. Thus, there are closely spaced pairs of points in this same range to show the best available throughputs of both model versions.

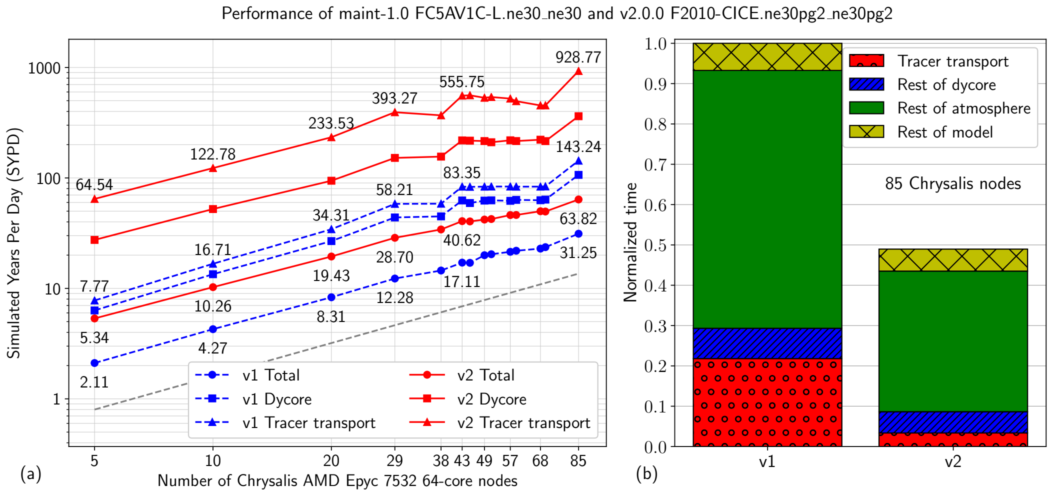


Figure 3. Performance of the lower resolution EAMv1 and EAMv2 atmosphere simulations. (a) Throughput vs. number of nodes. PE layouts are simple stacked layouts. (b) Proportion of time spent in each subcomponent, with the total time for v1 normalized to 1.

Figure 3b decomposes performance of the 85-node simulations into the same subcomponents. Only each full-height bar is fully accurate; subcomponent proportions are approximate. Again, tracer transport in v2 is over six times faster than in v1, speeding up the dynamical core by over three times in this case. The total model speedup is a little over two times in this case, with the speedup outside of the dynamical core coming from the reduction in number of physics columns.

3.2 File Input/Output

The EAM and E3SM simulations discussed above used the SCORPIO library for reading input data and writing simulation output to the file system. To improve the I/O write performance, the library caches and rearranges output data among MPI processes before using low-level I/O libraries, such as NetCDF, Parallel NetCDF (PnetCDF), and the Adaptable IO System (ADIOS), to write the data to the file system. In all the simulation campaigns we used PnetCDF as the low-level I/O library in SCORPIO, and I/O accounted for less than 4% of the total runtime of the simulation.

To measure the I/O improvements in the model, we compared the old version of the I/O library, SCORPIO CLASSIC (based on PIO, J. M. Dennis et al., 2012), used by E3SMv1 with the new version of the library, SCORPIO, used by E3SMv2 by running E3SMv1 benchmark simulation on Chrysalis with the S, M and L configurations. The simulation was run for 90 simulated days and generated $\sim$ 30GB of history and restart model output for each configuration. We found that SCORPIO provides a higher write and read performance than SCORPIO CLASSIC for all the model configurations. SCORPIO provides a consistent write throughput of 3-3.5 GB/s for all the configurations while the write throughput of SCORPIO CLASSIC drops from 1.9 GB/s for the S configuration to 356 MB/s for the L configuration. The time to read the model input data stays relatively constant for SCORPIO with the different model configurations while it increases exponentially with the number of MPI processes for SCORPIO CLASSIC. The time to read the model input data is $\sim$ 40% higher for SCORPIO CLASSIC compared to SCORPIO for the S and M model configurations, and for the L model configuration the time to read data with SCORPIO CLASSIC is 3.3 times the time taken with SCORPIO. The total time, including reads and writes, spent in I/O by both the libraries was less than

8% of the total runtime for all the model configurations except the L configuration with the SCORPIO CLASSIC library, where I/O accounted for 25% of the total runtime.

4 Simulation Campaign

Table 1 summarizes the E3SMv2 simulation campaign. All simulations were configured to adhere to the CMIP6 specifications as closely as possible and rely on the same boundary files as E3SMv1 (Golaz et al., 2019). The CMIP6 DECK plus historical simulations (Eyring et al., 2016) include the pre-industrial control (*piControl*) spanning a total of 500 years, idealized CO_2 simulations (*1pctCO2*, *abrupt-4xCO2*; 150 years each) and a five-member ensemble of historical simulations (*historical_N*; 1850-2014). These simulations were initialized from *piControl* on Jan 1 of various years as indicated in Table 1. AMIP simulations (prescribed SST and sea ice extent) were also performed to cover the entire period for which CMIP6 provides surface boundary conditions (1870-2014). Atmosphere, land and river initial conditions for *amip_N* were taken from year 1870 of the corresponding *historical_N* coupled simulations.

To understand the relative importance of different forcing agents, a set of DAMIP (Detection and Attribution Model Intercomparison Project; Gillett et al., 2016) historical simulations was performed. They consist of five-member ensembles with well-mixed greenhouse-gas-only (*hist-GHG*) and anthropogenic aerosol related (*hist-aer*). Instead of natural-only historical simulations as in Gillett et al. (2016), we opted for a third set with all agents active except well-mixed GHG and aerosols (*hist-all-xGHG-xaer*). This non-standard choice was motivated by a desire to include all forcing agents in our decomposition (including land-use and ozone).

Finally, we performed a set of simulations following RFMIP (Radiative Forcing Model Intercomparison Project; Pincus et al., 2016) with slight updates to the protocol (https://rfmip.leeds.ac.uk/rfmip-erf). These simulations are designed to estimate time-varying total effective radiative forcing (ERF) and aerosol-related ERF. Three sets of prescribed SST and sea ice simulations are performed with SST and sea ice derived from a 500-year average of *piControl*. *piClim-control* is the control simulation with all forcing agents held at their 1850 values. *piClim-histall* activates all time varying forcing agents, whereas *piClim-histaer* only activates time varying agents related to anthropogenic aerosols and their precursors.

The entire simulation campaign was performed on the DOE-E3SM Chrysalis cluster located at Argonne National Laboratory. E3SMv2 experienced only a single model crash during the nearly 3000 simulated years. The failure occurred during year 121 of *abrupt-4xCO2* ensemble member 301. The failure was overcome by rerunning and toggling a flag in the coupler ("BFBFLAG") that changes order of arithmetic operations. This introduces a "butterfly effect" sufficient to alter the weather and avoid the original failure point.

4.1 Pre-industrial control

The pre-industrial control simulation (*piControl*) was initialized after a 1000-year long spin-up simulation, itself initialized from ocean and sea ice states derived from a one-year forced ocean-sea ice simulation. During the spin-up, the model configuration was final, except for a small retuning of the gravity wave drag parameterization that was introduced at year 800 to improve the period of the QBO as described in Section 2.1.2.

The climate simulated by E3SMv2 is very stable throughout the 500-year *piControl* as demonstrated in Figure 4. The net TOA radiation (Fig. 4a) averages to -0.05 Wm^{-2} with no trend. This value is sufficiently close (compared to anthropogenic forcing) to the ideal value of 0 Wm^{-2} for a fully equilibrated and perfectly energy conserving model.

Table 1. Summary of E3SMv2 simulations.

Label	Description	Period	Ens.	Initialization
	Fully coupled (atmosphere, ocean, sea ice, land and river)			
piControl	Pre-industrial control	500 years	-	Pre-industrial spinup
1pctCO2	Prescribed 1% yr^{-1} CO_2 increase	150 years	1	*piControl* (101)
abrupt-4xCO2	Abrupt CO_2 quadrupling	150 years	2	*piControl* (101, 301)
historical_N	Historical	1850-2014	5	*piControl* (101, 151, 201, 251, 301)
hist-GHG	DAMIP well-mixed greenhouse-gas-only historical	1850-2014	5	*piControl* (101, 151, 201, 251, 301)
hist-aer	DAMIP anthropogenic-aerosol-only historical	1850-2014	5	*piControl* (101, 151, 201, 251, 301)
hist-all-xGHG-xaer	Other forcing historical (all forcing except GHG and aer)	1850-2014	5	*piControl* (101, 151, 201, 251, 301)
	Prescribed SST and sea ice extent (atmosphere, thermodynamic sea ice, land and river)			
amip_N	Atmosphere with prescribed SSTs and sea ice concentration	1870-2014	3	*historical_N* (1870)
piClim-control	RFMIP baseline control	50 years	-	Pre-industrial spinup
piClim-histall	RFMIP time-varying ERF all agents	1850-2014	3	*piClim-Control* (21, 31, 41)
piClim-histaer	RFMIP time-varying ERF aerosols	1850-2014	3	*piClim-Control* (21, 31, 41)

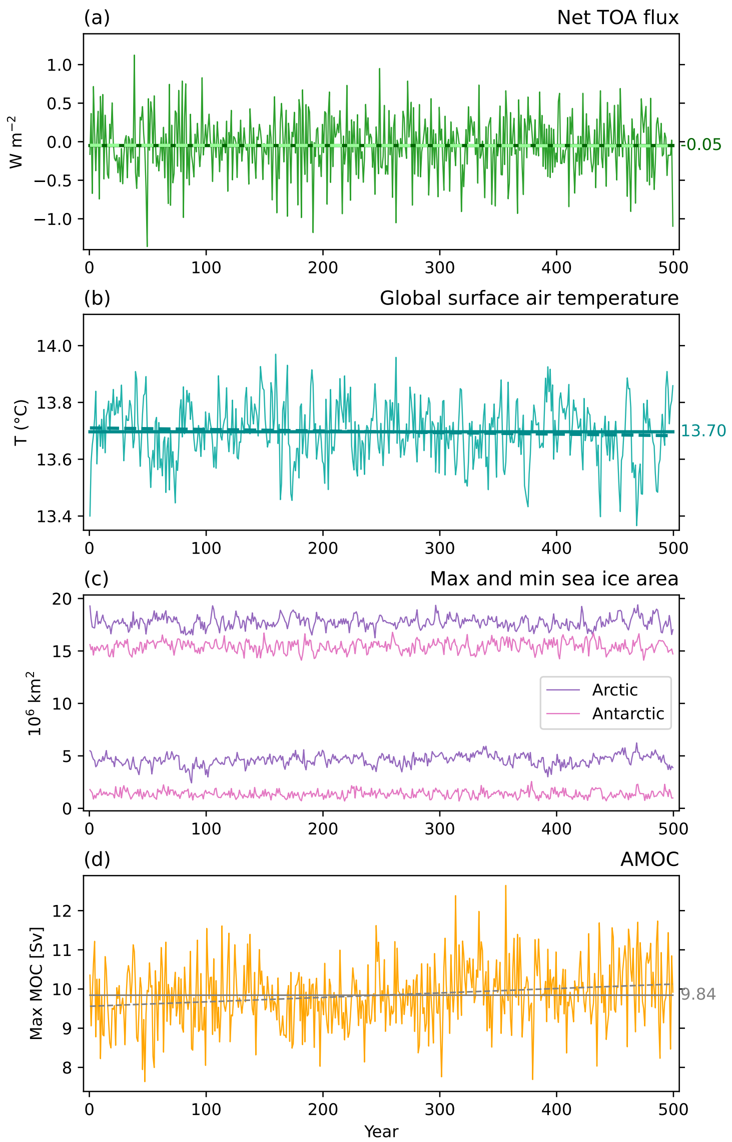


Figure 4. Time evolution of annual (a) global mean net top-of-atmosphere (TOA) radiation (positive down), (b) global mean surface air temperature, (c) maximum and minimum of total sea ice area for the Arctic and Antarctic, and (d) maximum Atlantic Meridional Overturning Circulation (AMOC) at 26.5°N below 500-m depth in the piControl simulation. Dashed lines in (a), (b), and (d) represent linear trends. The solid straight line in (a) is the mean TOA energy imbalance of -0.05 Wm^{-2}, while the solid straight line in (d) is the mean annual maximum AMOC of 9.84 Sv.

The global mean surface temperature averages to 13.70 °C with a very small downward trend (dashed line in Fig. 4b). The average temperature is very similar to E3SMv1 and consistent with estimated warming and the present-day global temperature of 14.0±0.5°C by Jones et al. (1999) for the period 1961-1990 and with leading reanalyses datasets (14.3 to 14.6 °C) for the period 1979-2008 (Hawkins & Sutton, 2016). Along with the global mean temperature, maximum and minimum seasonal sea ice areas for the Arctic and Antarctic are stable as well (Fig. 4c).

Finally, the maximum AMOC in E3SMv2 is quite weak, similarly to that in E3SMv1. The AMOC in Fig. 4d is weaker than the value in Golaz et al. (2019) (∼11 Sv). However, during the E3SMv2 development it was discovered that the published AMOC did not include the contribution of the parameterized mesoscale eddies. In the North Atlantic the influence of the Gent-McWilliams parameterization opposes the resolved AMOC. When the eddy bolus velocity is included in the v1 calculation, the AMOC is very similar (∼9.5 Sv) to that in E3SMv2.

4.2 Climate sensitivity and effective radiative forcing

Included in the DECK simulations are two idealized CO_2 simulations designed to estimate the model response (sensitivity) to CO_2-forcing at different time horizons. The equilibrium climate sensitivity (ECS) is defined as the equilibrium surface temperature change resulting from a doubling in CO_2 concentrations. Because it is not practical to run a model to equilibrium, ECS is approximated by linear regression of TOA radiation vs surface temperature in a 150-year *"abrupt-4xCO2"* simulation (Gregory et al., 2004), often referred to as "effective climate sensitivity". Response on shorter time scales is measured by the transient climate response (TCR). TCR is defined as the change in surface temperature averaged for a 20-year period around the time of CO_2 doubling from a *1pctCO2* simulation. TCR depends on both climate sensitivity and ocean heat uptake rate.

Figure 5 illustrates the time evolution of annual-average surface air temperature from the E3SMv1 and E3SMv2 idealized CO_2 simulations, as well as their linear regression. ECS is reduced from 5.3 K in E3SMv1 to 4.0 K in E3SMv2, a substantial reduction. TCR is reduced as well, but by a smaller relative fraction from 2.93 K to 2.41 K. The effective CO_2 radiative forcing is also reduced by approximately 10% (3.34 to 2.98 Wm^{-2}).

For comparison, Meehl et al. (2020) evaluated ECS and TCR for 37 CMIP6 models. ECS ranged between 1.8 and 5.6 K, with 6 models above 5 K including E3SMv1. The multimodel mean ECS was 3.7 K with a standard deviation of 1.1 K. TCR ranged from 1.3 to 3.0 K, with E3SMv1 having the largest value. The multimodel mean TCR was 2.0 K with a standard deviation of 0.4 K. E3SMv2 is now within one standard deviation of multimodel mean for both ECS and TCR, but still on the high side.

World Climate Research Programme (WCRP) researchers conducted a recent assessment of the equilibrium climate sensitivity following multiple lines of evidence (Sherwood et al., 2020). They arrived at a 66% confidence range of 2.6–3.9 K for their baseline calculation and 2.3–4.5 K under their robustness tests. The broader 5–95% confidence ranges were 2.3–4.7 K, respectively 2.0–5.7 K. E3SMv1 with an ECS of 5.3 K is rather unrealistic as it lies outside of most of those ranges. On the other hand, E3SMv2 has a high, but plausible ECS of 4.0 K.

Although a part of the reduction in ECS stems from the reduced effective radiative forcing in E3SMv2 (from 3.34 to 2.98 Wm^{-2}), it is mainly due to the reduced total climate feedback. Applying the radiative kernel method (Soden et al., 2008) implemented in the E3SM cloud feedback diagnostic package (Qin, 2022) to decompose the climate feedback into different components, we find the reduced cloud feedback (E3SMv1: 0.93 $Wm^{-2}K^{-1}$; E3SMv2: 0.72 $Wm^{-2}K^{-1}$), especially over the marine low cloud regions, contributes the most to the reduction in total climate feedback, whereas the changes in other non-cloud feedbacks are negligible. Sensitivity tests on model changes in E3SMv2 atmosphere physics indicate that the dCAPE-ULL convective trigger in the ZM scheme and the updated CLUBB tuning parameters play leading roles in reducing the marine low cloud feedbacks in E3SMv2.

We also evaluate the evolution of the effective radiative forcing (ERF) from pre-industrial to present-day conditions using RFMIP simulations (Table 1). ERF_{total} is the difference in net TOA radiation between *piClim-histall* and *piClim-control* and ERF_{aer} the difference between *piClim-histaer* and *piClim-control*. Their time evolutions are shown in Figure 6 along with their counterparts from E3SMv1 (computed with a comparable but slightly different methodology, see Golaz et al., 2019). The time evolutions of ERF_{total} and ERF_{aer} are nearly identical between E3SMv2 and E3SMv1. ERF_{total} remains close to zero until the late 1900's, except for dips during explosive volcanic eruptions. Averaging over the last 20 years reveals small differences between the two models. The aerosol forcing is slightly reduced in magnitude (-1.52 vs -1.65 Wm^{-2}), but the the total forcing does not increase as a result. In fact it is reduced (+1.00 vs +1.10 Wm^{-2}), likely as a consequence of the smaller CO_2 ERF (Fig. 5).

Another assessment was conducted under the auspices of the WCRP with the goal of bounding the aerosol radiative forcing (Bellouin et al., 2020). Following multiple lines of evidence, the assessment arrived at a 68% confidence interval for the total aerosol effective radiative forcing of -1.6 to -0.6 Wm^{-2}, or -2.0 to -0.4 Wm^{-2} with a 90% likelihood. With a forcing of -1.52 Wm^{-2}, E3SMv2 is close to the lower bound but within the narrower confidence interval.

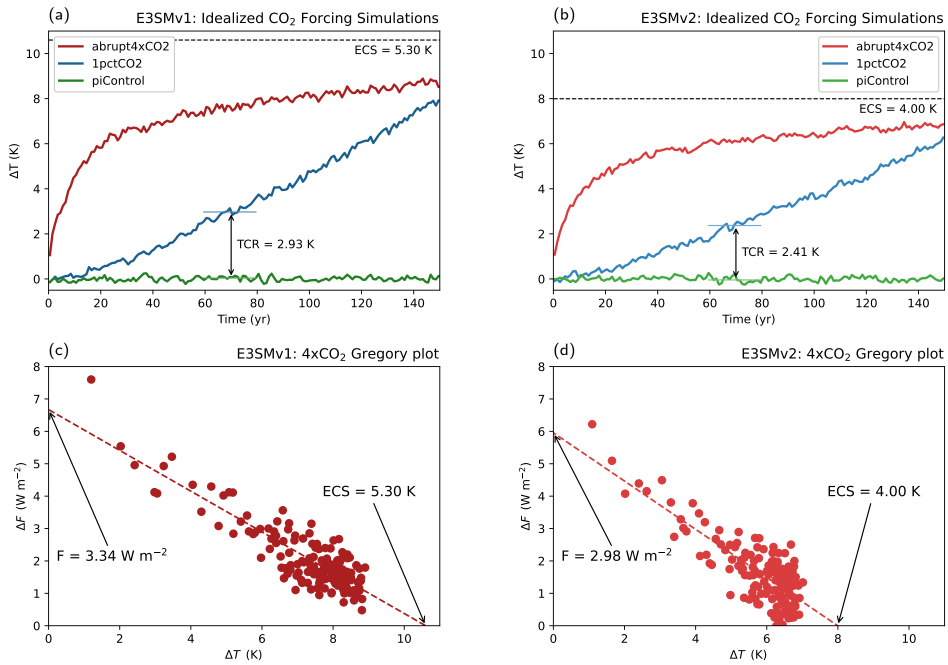


Figure 5. (a-b): time evolution of annual global mean air surface temperature anomalies for the idealized CO_2 forcing simulations *abrupt-4xCO2* (red), *1pctCO2* (blue) and the control simulation (*piControl*; green) for E3SMv1 and E3SMv2. The transient climate response (TCR) is computed as an 20-year average around time of doubling (year 70). (c-d) Gregory regression to estimate effective climate sensitivity (ECS) and effective $2xCO_2$ radiative forcing (F).

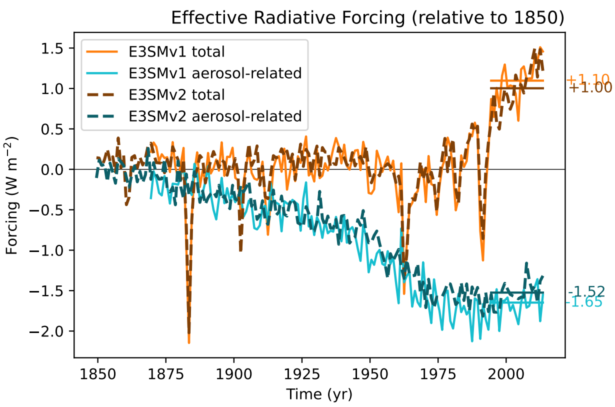


Figure 6. Time evolution of annual global mean total ERF (brown) and aerosol-related ERF (blue) for E3SMv1 and E3SMv2. Horizontal lines and adjacent values denote averages from 1995 to 2014.

4.3 Historical ensemble

To facilitate comparisons between model and observations, the bulk of the analysis focuses on the historical simulations. For climatologies, we select the last 30 years (1985-2014) of the ensemble members.

Figure 7 provides a broad overview of the model performance. Spatial RMSE against observations or ranalysis products are computed for annual and sesonal averages with the E3SM Diags package (C. Zhang et al., 2022). The first historical ensemble members of E3SM are depicted with triangles, blue for E3SMv1 and red for E3SMv2. They are compared against 52 CMIP6 models shown with box-and-whisker plots (minimum, 25th, 75th percentile, maximum). Underlying E3SM Diags comparison figures are available on-line (`https://portal.nersc.gov/project/e3sm/CMIP6_comparison_1985-2014_E3SMv2_golaz_etal_2022/`). For most fields, E3SMv2 outperforms E3SMv1. Notable improvements include precipitation and sea-level pressure. The simulated precipitation in E3SMv2 is now competitive with the upper quartile of the CMIP6 ensemble. While sea-level pressure is also much improved, it is still only about average compared to CMIP6. Consistent with sea-level pressure, zonal wind at 850 hPa also improves. E3SMv2, similarly to E3SMv1, has a good representation of TOA radiation fields, moderately improving upon v1 for most fields and seasons. Unfortunately, two fields suffer from a degradation in E3SMv2 as compared to E3SMv1. For the zonal wind at 200 hPa, the degradation is partly associated with the change in stratospheric ozone chemistry (i.e., O3v2) (Tang et al., 2021, their Figure 10), but the differences between E3SMv2 and E3SMv1 in Figure 7 are larger than those between E3SMv1+O3v2 and E3SMv1, suggesting that other factors contribute as well. The degradation in surface air temperature over land is largely attributable to poor simulation of the historical temperature record (see Sections 4.3.6 and 5 below).

4.3.1 Radiation and Clouds

Annual net top-of-atmosphere (TOA) radiative flux in E3SMv1 and v2 is depicted in Figure 8 in comparison with observations from CERES-EBAF Ed4.1 (Loeb et al., 2018). The simulated global mean value is nearly identical between the two versions at +0.3 Wm^{-2}, lower than the observational estimate (but consistent with the smaller warming; Figure 6 and Section 4.3.6). Many regional biases are reduced in E3SMv2, including positive biases over stratocumulus regions, as well as negative biases over tropical

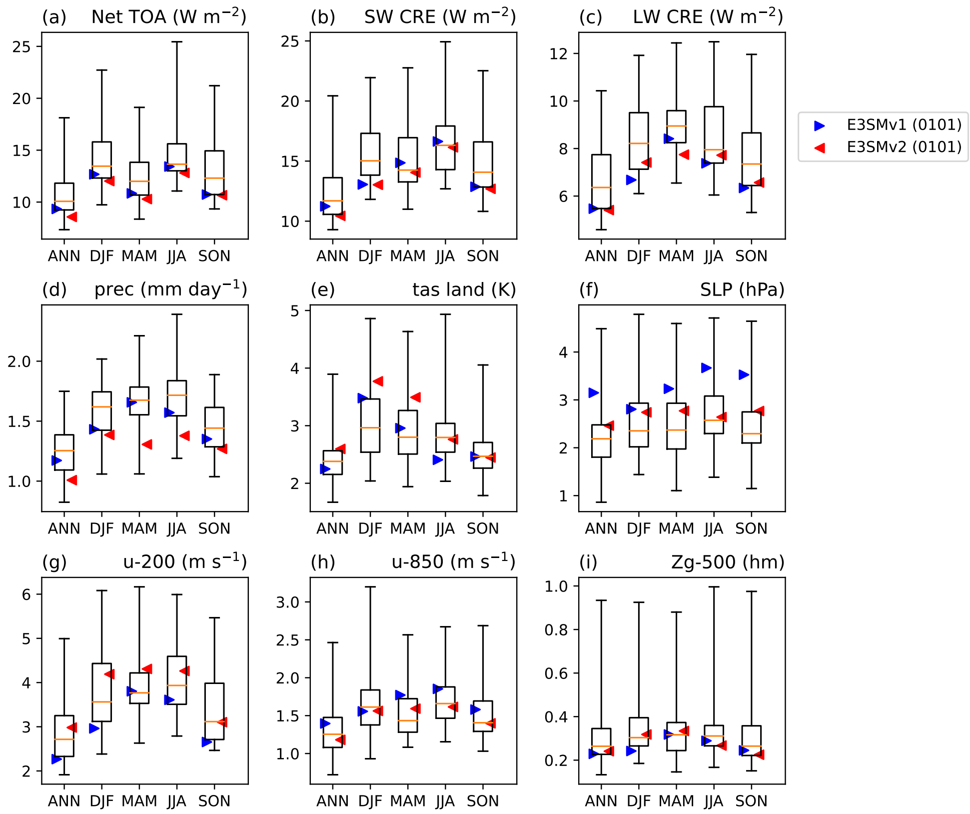


Figure 7. Comparison of RMSE (1985–2014) of an ensemble of 52 CMIP6 models (first historical members r1i1p1f1) with the first historical members of E3SMv1 (blue triangles) and E3SMv2 (red triangles). Box and whiskers show 25th, 75th percentile, minimum and maximum RMSE for the CMIP6 ensemble. Spatial RMSE against observations are computed for annual and sesonal averages with the E3SM Diags package (C. Zhang et al., 2022). Fields shown include TOA net radiation (a), TOA SW and LW cloud radiative effects (b, c), precipitation (d), surface air temperature over land (e), sea-level pressure (f), 200- and 850-hPa zonal wind (g, h), and 500-hPa geopotential height (i). TOA = top-of-atmosphere; SW = shortwave; CRE = cloud radiative effects; LW = longwave; DJF = December–February; MAM = March–April; JJA = June–August; SON = September–November; RMSE = root-mean-square error. The mean climatology of the reference observational and reanalysis datasets are derived from: CERES-EBAF Ed4.1 (Loeb et al., 2018) (2001-2018) for (a, b and c), GPCP2.3 (Adler et al., 2018) (1979-2017) for (d) and ERA5 (Hersbach et al., 2020) (1979-2019) for (e, f, g and h). Due to data availability, not all models are included for every variable. Complete data is available in Table S1.

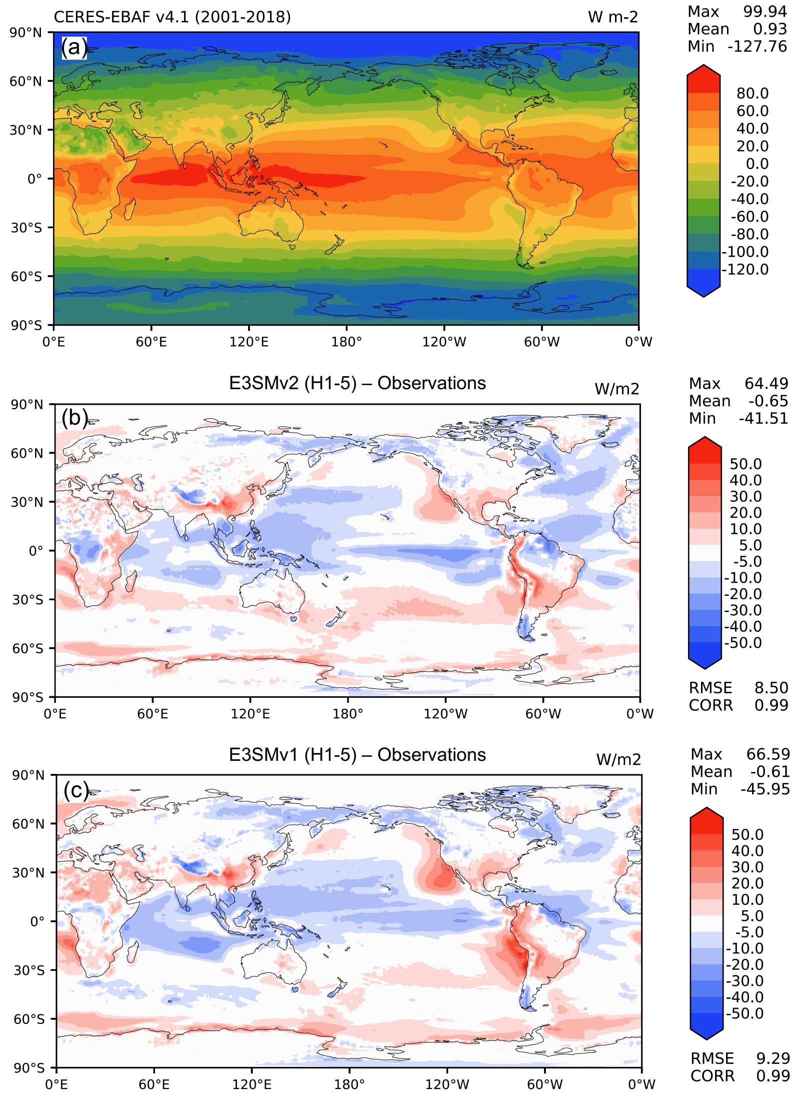


Figure 8. Annual net top-of-atmosphere (TOA) radiative flux (W/m^2): (a) CERES-EBAF Ed4.1 observational estimate (2001-2018), (b) model bias from the 5-member ensemble of E3SMv2 historical coupled simulations (1985–2014), and (c) model bias from the 5-member ensemble of E3SMv1 historical coupled simulations (1985–2014). RMSE = root-mean-square error. CORR = correlation coefficient between observation and model.

and subtropical Pacific, Indian, and Atlantic oceans, resulting in an overall smaller RMSE (8.5 vs 9.3 Wm^{-2}).

Figure 9, 10 demonstrate that both the TOA shortwave and longwave cloud radiative effects are improved in the E3SMv2 historical ensemble compared with the E3SMv1 historical ensemble in terms of RMSE and the pattern correlation. Overall, the global mean SWCRE in E3SMv2 is weaker than in E3SMv1 by $\sim 1.5\mathrm{Wm}^{-2}$, and the LWCRE is weaker by $\sim 0.6\mathrm{Wm}^{-2}$. The positive TOA SWCRE bias associated with the stratocumulus decks over eastern ocean basins, especially right off the coasts of California, Peru and Chile, and the southern West Africa, is clearly reduced, while the negative SWCRE bias associated with the cumulus regimes over central/western tropical oceans is slightly reduced as well. The improvement in the marine boundary layer cloud regimes is mainly from the updated CLUBB tuning paramters (Ma et al., 2022). The TOA LWCRE bias is reduced over the equatorial Pacific and the intertropical convergence zone (ITCZ), which is associated with the improved precipitation over these areas (described in Section 4.3.2 below). The positive TOA LWCRE bias is also slightly reduced over the Southern Ocean.

The enhanced Wegener-Bergeron-Findeisen (WBF) efficiency and the update to the ZM scheme significantly increase ice water in mixed-phase clouds, which also weakens SWCRE in the Southern Hemisphere (e.g. $\sim 30\,°\mathrm{S}$ in Figure 9). The liquid condensate mass fraction as a function of temperature at all latitudes between $30\,°\mathrm{S}$–$80\,°\mathrm{S}$ (Figure 11) from both E3SMv1 and E3SMv2 historical coupled simulations demonstrate that the updated atmosphere features and tuning parameters in E3SMv2 significantly increase ice cloud mass fraction in the temperature range between $-10\,°\mathrm{C}$ and $-50\,°\mathrm{C}$, which is closer to the observational estimate (Y. Zhang et al., 2019).

We further quantify the improvements in the subtropical stratocumulus decks compared to E3SMv1 following Brunke et al. (2019). We define the decks as the areas within 30° latitude by 35° longitude boxes in the Northeast Pacific (NEP), Northeast Atlantic (NEA), Southeast Pacific (SEP), Southeast Atlantic (SEA), and the Southern Indian Ocean (SIO) where low cloud cover > 45, the LCC45+ decks. E3SMv2 LCC is generally improved falling more within the observational spread represented by three satellite and in-situ based climatologies [the Cloud-Aerosol Lidar and Infrared Pathfinder (CALIPSO) satellite GCM-Oriented CALIPSO Cloud Product (GOCCP), the International Satellite Cloud Climatology Project (ISCCP) D2 product, and the Extended Edited Cloud Reports Archive (EECRA)] (Figure 12).

The cloud changes that lead to the SWCRE improvements can be explained by the spatial errors in the simulated LCC45+ cloud decks with respect to GOCCP which are defined as in Brunke et al. (2019). An example of these for the seasons of maximum LCC for each region in Figure 12 is given in Figure 13. For "apples-to-apples" comparisons, the model output from the Cloud Feedback Model Intercomparison Project Observation Simulator Package (COSP) CALIPSO satellite simulator is used. Centroid distances (Figure 13a) measure the distance between the centroid of the seasonal mean cloud deck in GOCCP and the model. Smaller centroid distances are better than large ones. Area ratios (Figure 13b) are the ratio of the area of the model's deck to that of the satellite to measure cloud deck size errors. Finally, overlap ratios (Figure 13c) are the fraction of the union of the model and satellite cloud decks in which there is overlap. This synthesizes the effects of location, size, and shape errors in the simulated cloud decks. Both of these ratios should be close to 1 for minimal errors.

Figure 13 shows that E3SMv2 improves most the representation of the widely studied subtropical stratocumulus cloud decks in the NEP, NEA, and SEP. In these regions, centroid distances are decreased and overlap ratios are similar to or increased to values closer to 1. Area ratios are improved in all regions with values closer to 1 except NEA. Similar results are found in all other seasons.

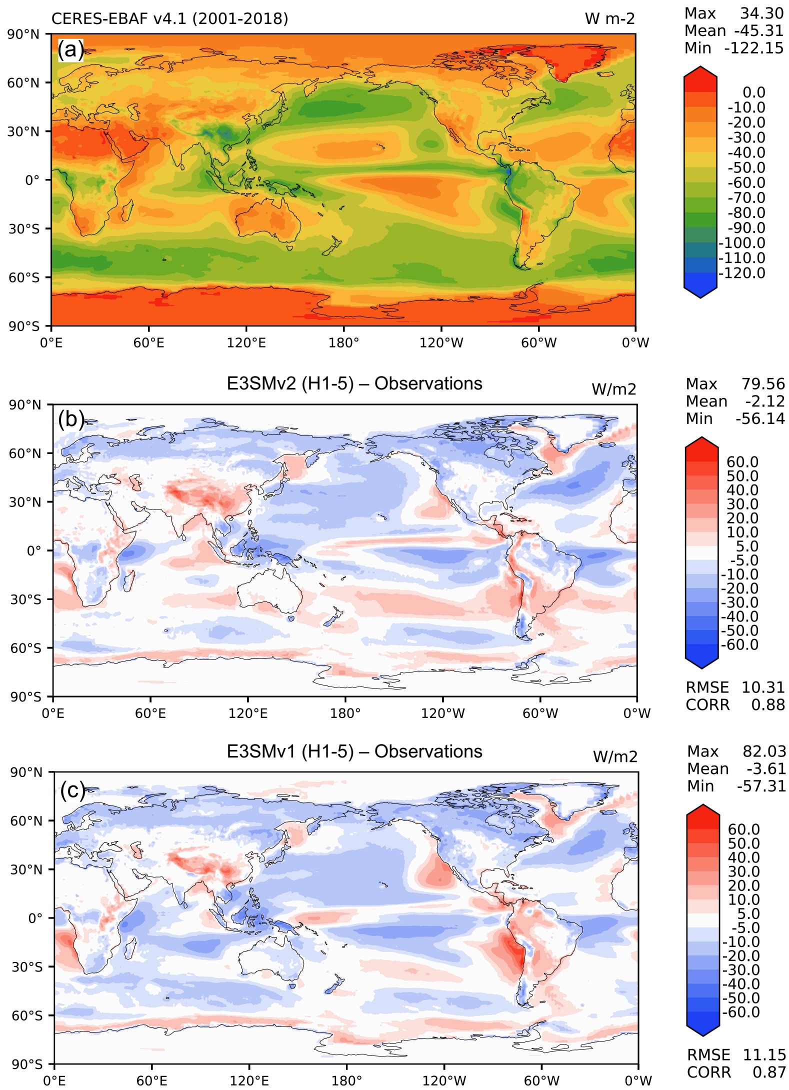


Figure 9. Annual top-of-atmosphere shortwave cloud radiative effect (W/m^2): (a) CERES-EBAF Ed4.1 observational estimate (2001-2018), (b) model bias from the 5-member ensemble of E3SMv2 historical coupled simulations (1985–2014), and (c) model bias from the 5-member ensemble of E3SMv1 historical coupled simulations (1985–2014). RMSE = root-mean-square error. CORR = correlation coefficient between observation and model.

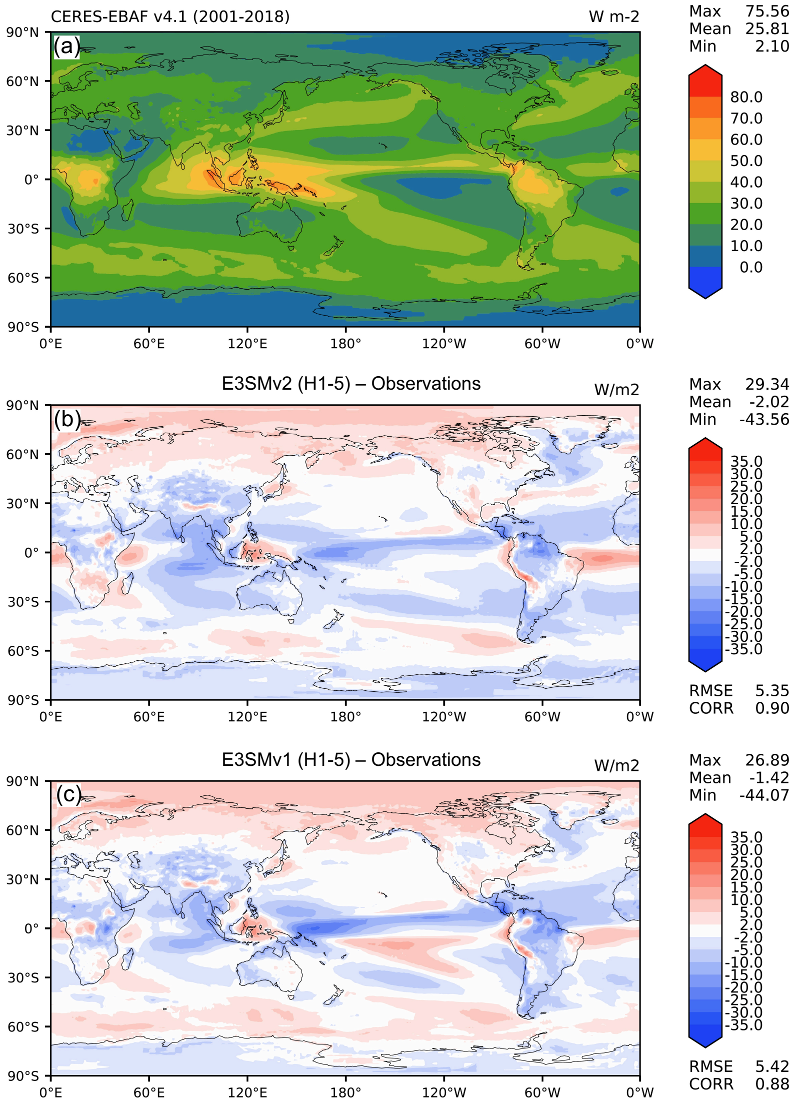


Figure 10. Annual top-of-atmosphere longwave cloud radiative effect (Wm^{-2}): (a) CERES-EBAF Ed4.1 observational estimate (2001-2018), (b) model bias from the 5-member ensemble of E3SMv2 historical coupled simulations (1985–2014), and (c) model bias from the 5-member ensemble of E3SMv1 historical coupled simulations (1985–2014). RMSE = root-mean-square error. CORR = correlation coefficient between observation and model.

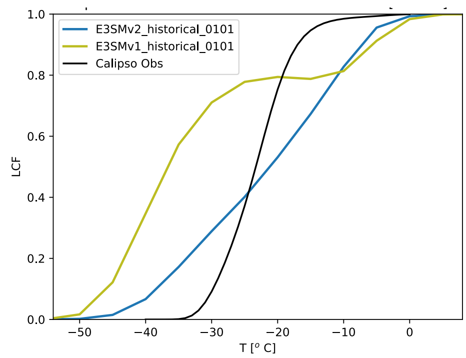


Figure 11. Diagnosed mixed-phase partitioning based on the monthly model output in the 30-80°S latitude band from (blue line) the E3SMv2 historical coupled simulation (1985–2014), (olive line) the E3SMv1 historical coupled simulations (1985–2014), and (black line) observations from Hu et al. (2010)

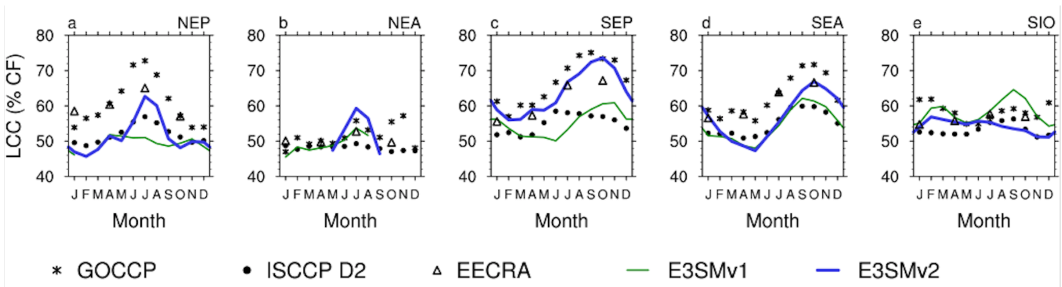


Figure 12. The mean low cloud cover (LCC) for each of the LCC45+ cloud decks (see text for definitions) for the 30° latitude by 35° longitude boxes over the Northeast Pacific (NEP), Northeast Atlantic (NEA), Southeast Pacific (SEP), Southeast Atlantic (SEA), and the Southern Indian Ocean (SIO).

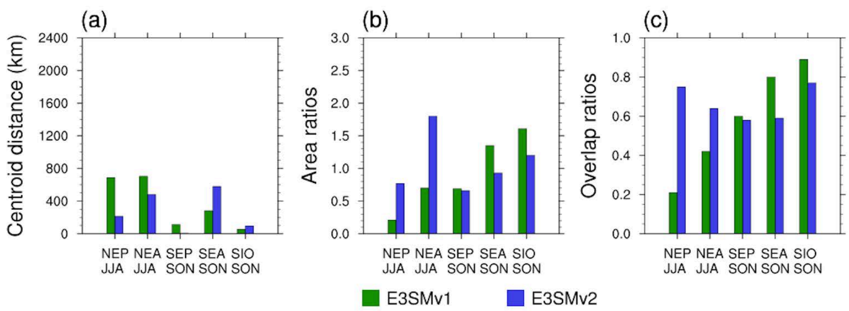


Figure 13. Centroid distances, area ratios, and overlap ratios of the LCC45+ decks in June-August (JJA) for the two Northern Hemisphere regions and in September-November (SON) for the Southern Hemisphere regions.

4.3.2 Precipitation

The model bias in annual precipitation from E3SMv2 shows notable improvement compared with that in E3SMv1 (Figure 14). The biases are clearly reduced in the Tropical Pacific ocean, Maritime continent, Central America and the Amazon. The updated ZM tuning parameters, the dCAPE-ULL convective trigger, and the inclusion of the gustiness effects and the subgrid temperature variance are found to reduce the regional biases of annual mean precipitation (Xie et al., 2019; Ma et al., 2022).

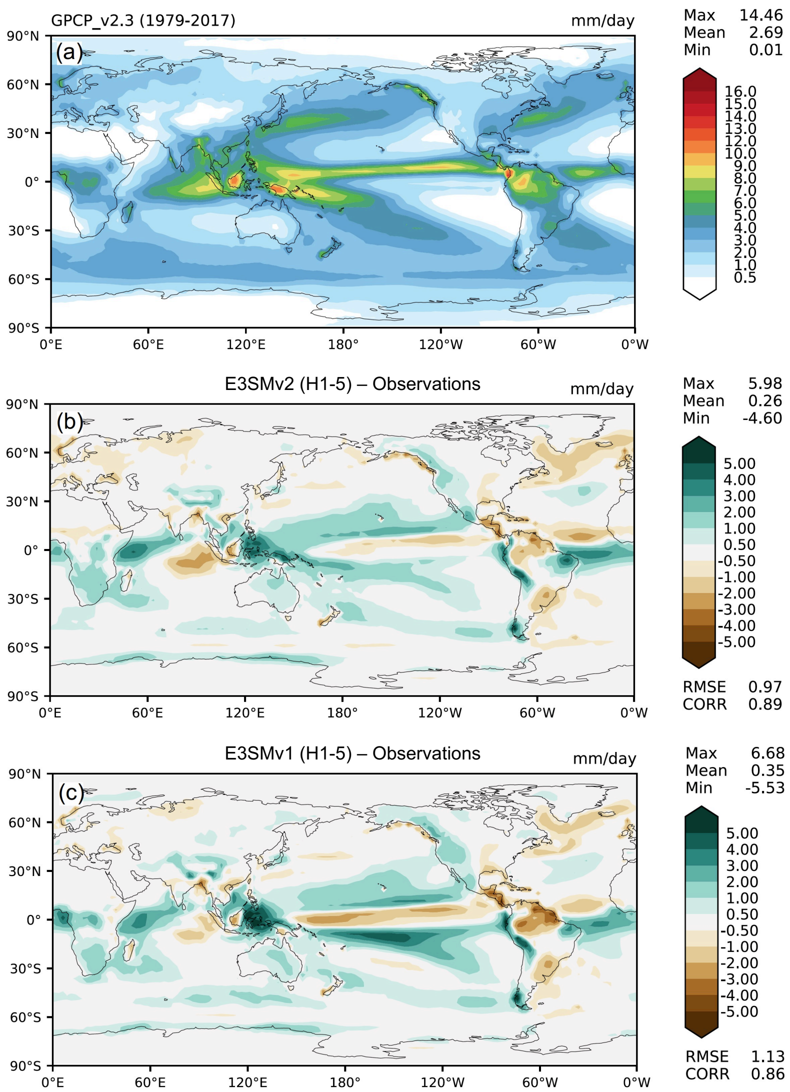


Figure 14. Annual precipitation rate (mm/day): (a) Global Precipitation Climatology Project v2.3 observational estimate (1979-2017), (b) model bias from the 5-member ensemble of E3SMv2 historical coupled simulations (1985–2014), and (c) model bias from the 5-member ensemble of E3SMv1 historical coupled simulations (1985–2014). RMSE = root-mean-square error. CORR = correlation coefficient between observation and model.

As described in section 2.1.2, the dCAPE-ULL convective trigger is expected to broadly improve the simulation of diurnal precipitation. This can be clearly seen in Figure 15, which shows the comparison of the time phase (color) and amplitude (color density) of diurnal precipitation between TRMM, and E3SMv2 and E3SMv1 historical simulations over the tropics. The improvements are most evident in the diurnal peak phase. Over the oceans, E3SMv2 captures the observed widespread morning peaks, particularly along the primary precipitation bands, where on average the peak precipitation occurs 3 hours too early in E3SMv1. Over the Maritime continent region, E3SMv2 closely reproduces the observed early evening peaks over land and the transition to morning peaks towards the coasts and open oceans, while E3SMv1 has too-early diurnal precipitation peaks from noon to early afternoon over land and similarly much earlier peaks around midnight in the coastal regions. Over the tropical continents, including Africa, South America, and South Asia, the observed diurnal peaks occur from late afternoon to early evening. While the diurnal precipitation peaks in E3SMv1 are nearly phase-locked to insolation over these land masses, the phase-locking behaviors are totally avoided in E3SMv2, which actually sees the peak phases delayed by several hours. However, the improvement in simulating diurnal timing phases does not translate to diurnal amplitude. This is presumably due to lack of skill in simulating meso-scale convective systems in coarse resolution models. Furthermore, while the diurnal amplitudes are weaker in both models compared to observations, they are somewhat degraded from E3SMv1 to E3SMv2 particularly over weakly precipitating subtropical oceans.

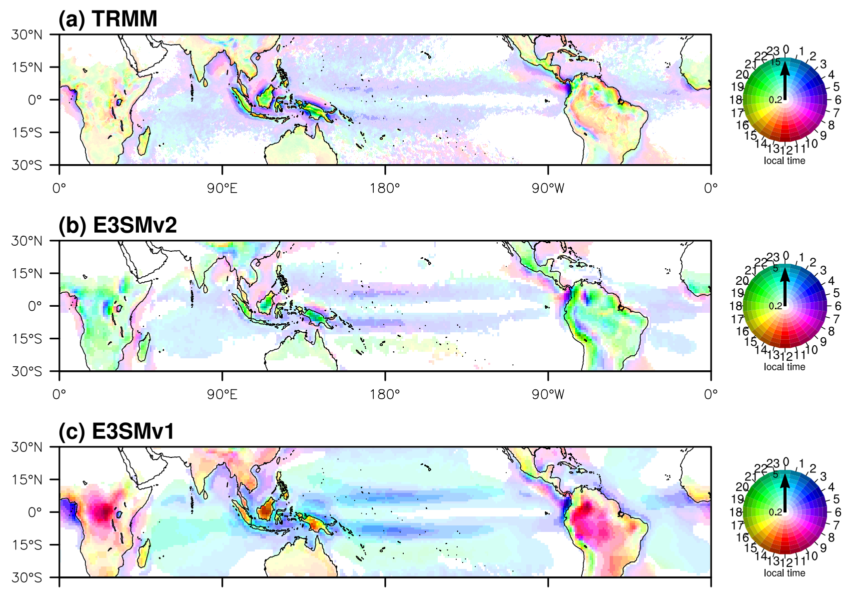


Figure 15. Annual mean time phase (color) and amplitude (color density) of the first diurnal harmonic of 3-hourly total precipitation (mm/day) from (a) TRMM (1998-2013), and historical simulations (1985-2014) of (b) E3SMv2 and (c) E3SMv1. Note the difference in the upper bound of the plotted amplitude ranges for TRMM (15 mm/day) and models (5 mm/day). Areas with diurnal amplitude less than 0.2 mm/day are left blank.

After the model was finalized, it was observed that the dCAPE trigger, independent of the ULL trigger and other model settings, induces a checkerboard grid-level noise pattern in a number of output fields, including total grid-box cloud water liquid and ice paths, when these fields are temporally instantaneous or averaged over not more than several days. Figure S3 illustrates this issue by comparing a daily average output of the total grid-box cloud liquid water path in two lower resolution atmosphere simulations with the dCAPE trigger on and off.

4.3.3 Tropical variability

As in Golaz et al. (2019) we examine the E3SMv2 variability of El Niño Southern Oscillation (ENSO) via wavelet analysis (Torrence & Compo, 1998) of the Niño 3.4 SST for the *piControl* and historical simulations in Fig. 16. In this figure the *piControl* has again been divided into five 100-year intervals. The 90% confidence interval is shown as the dashed black line. ENSO variability in E3SMv2 shows a number of similarities to E3SMv1 (compare to Golaz et al., 2019, their Fig. 20). Again E3SMv2 shows a very robust peak of variability at short periods (∼2.5 years), which is similar to E3SMv1 and shorter than ERSSTv4 (thick black line). While a longer period (6-9 years) remains in the *piControl*, the mean for the five 100-year intervals has reduced relative to E3SMv1. This longer term variability is weaker than simulated in other CMIP5 and CMIP6 models (see Orbe et al., 2020, their Fig. 10a) and observations (black line in Fig. 16). The intermediate periods (3-6 years) seen in ERSSTv4 are not well captured in E3SMv2. The spatial SST response to ENSO is shown in Fig. S4. The magnitude of SST response (approximately 2.5°C) in the *piControl* and historical ensemble mean (panels b and c) is consistent with E3SMv1, other CMIP models, and observations (Golaz et al., 2019; Brown et al., 2020). However, the center of response is shifted too far westward, which is consistent with other models.

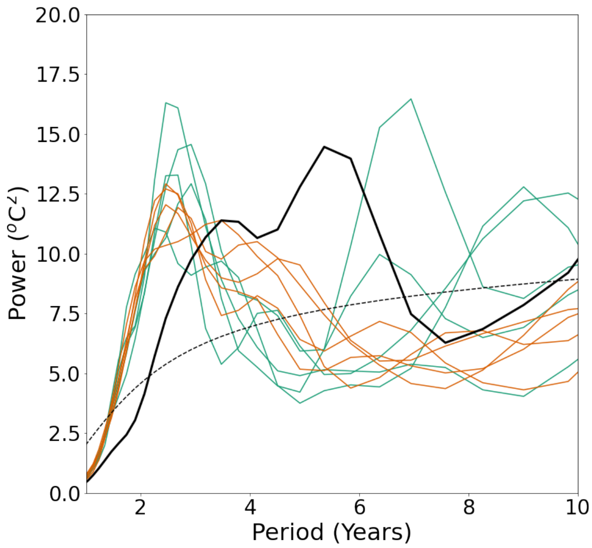


Figure 16. El Niño–Southern Oscillation (ENSO; Nino3.4) variability of the pre-industrial (PI) control simulation and historical ensemble. The Morlet wavelet of degree 6 is used (e.g., Torrence & Compo, 1998). The PI control (green lines) has been divided into five 100-year sections, each Historical ensemble member is shown as an orange line. ERSSTv4 data (W. Liu et al., 2015) is shown as the thick black line. The 90% confidence interval is shown as the dashed black line.

The Madden-Julian oscillation (MJO; Madden & Julian, 1971), the dominant mode of tropical variability on subseasonal (10-100 day) scales, is a key contributor to ENSO events (C. Zhang & Gottschalck, 2002), monsoon activity (Wheeler & McBride, 2012), extratropical atmospheric blocking episodes (Henderson et al., 2016), tropical cyclone formation (Maloney & Hartmann, 2000), and weather extremes (Higgins et al., 2000; Matsueda & Takaya, 2015; Mundhenk et al., 2016). Its accurate representation in numerical models is essential for weather and climate prediction (Vitart & Robertson, 2018), yet a satisfactory depiction of the MJO remains elusive (Jiang et al., 2015; Ahn et al., 2020). Figure 17 shows the distribution of tropical precipitation spectral power, normalized by a smoothed background spectrum, in zonal wavenumber-frequency space (Wheeler & Kiladis, 1999). Results from an E3SMv2 historical simulation (Fig. 17b) indicate slightly lower power values for equatorial Rossby waves and the MJO and a MJO peak that is at a higher frequency compared to observations (Fig. 17a). Relative to E3SMv1 (see Golaz et al. (2019) and Orbe et al. (2020) for details), precipitation normalized power in the broad MJO spectral region has increased and shifted to higher frequencies (Fig. 17c). Both E3SMv2 and E3SMv1 dramatically underestimate precipitation variability associated with atmospheric Kelvin waves and other synoptic-scale disturbances. Lag correlations of equatorial precipitation and 850 hPa zonal wind with Indian Ocean precipitation (Figure 18) suggest some improvement in MJO propagation across the Maritime Continent in E3SMv2 compared to E3SMv1, as evidenced by more consistent red shading eastward to 125°E. In both E3SMv2 and E3SMv1, the quadrature phasing of precipitation and zonal wind resembles that in observations, but the MJO phase speed begins to exceed the observed 5.5 m s^{-1} reference value (dashed green line) east of 120°E and especially in E3SMv2. A more detailed evaluation of tropical subseasonal variability in E3SMv2 will be presented in a forthcoming manuscript.

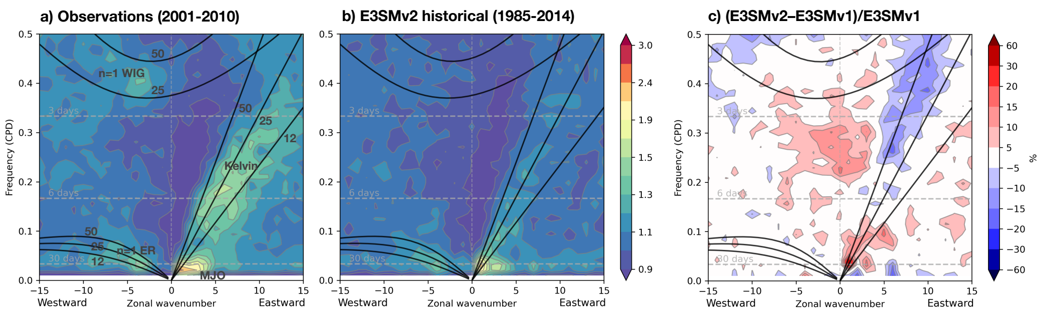


Figure 17. Tropical zonal wavenumber-frequency power spectra of the component of precipitation that is symmetric about the Equator for (a) observations (Tropical Rainfall Measuring Mission product 3B42v7) from 2001-2010 and (b) the 1985–2014 period from an E3SMv2 historical simulation. Plotted values represent the summed power from 15°S–15°N divided by the smoothed background power (the "normalized" power). Solid black lines indicate shallow water dispersion curves for equivalent depths of 12, 25, and 50 m. Prominent wave types are labeled: westward inertia-gravity (n=1 WIG), Kelvin, equatorial Rossby (n=1 ER), and the Madden-Julian oscillation (MJO). (c) The change, expressed as a percent difference, in the normalized spectral power between E3SMv2 and E3SMv1 historical simulations for the period 1985–2014.

4.3.4 Ozone

The stratospheric column ozone (SCO) of the historical ensemble mean of E3SMv2 is compared with the satellite observations from the Ozone Monitoring Instrument (OMI) and the Microwave Limb Sounder (MLS) at 60°S to 60°N, where the satellite observations have good quality all year round. Figure 19 shows the climatology of SCO zonal mean annual cycle from years 1995–2014 of E3SMv2 historical simulations and years 2005–2017 of the OMI+MLS observations. We chose different years of simulations from that of the observations to facilitate the comparison with the O3v2 model results in the E3SMv1 reported by Tang et al. (2021). The E3SMv2 historical simulations match the observed

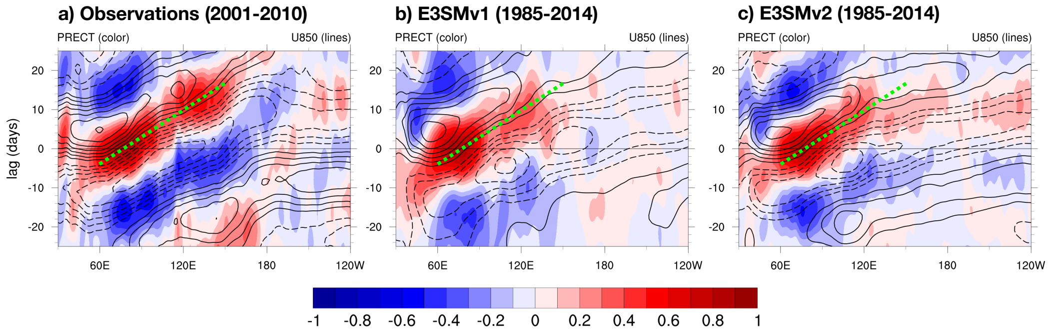


Figure 18. Latitudinally averaged (10°S–10°N) precipitation (colors) and 850 hPa zonal wind (lines) anomalies lag correlated with precipitation in the Indian Ocean region (60°–90°E, 10°S–10°N) for (a) observations from 2001–2010 (Tropical Rainfall Measuring Mission [TRMM] precipitation and Modern-Era Retrospective Analysis for Research and Applications [MERRA] wind), (b) the 1985–2014 period from an E3SMv1 historical simulation, and (c) the 1985–2014 period from an E3SMv2 historical simulation. The dashed green line in (a) represents the observed Madden-Julian oscillation phase speed ($5.5\,\mathrm{m\,s^{-1}}$) in precipitation and is copied to panels (b) and (c) for reference. The line contour interval is 0.1, solid lines indicate positive correlations, dashed lines indicate negative correlations, and the zero correlation line is omitted. Anomalies, defined as departures from the smoothed seasonal cycle, are bandpass filtered to retain 20-100 day signals prior to correlation.

SCO seasonal phase and pattern, but generally overestimate the SCO magnitude except over the Northern Hemisphere (NH) mid-latitudes and near 30°S from March to September. Comparing to the E3SMv1 SCO in Fig. 1d of Tang et al. (2021), the E3SMv2 SCO better matches observations in the SH mid-latitudes, but is worse in the NH mid-latitudes. This E3SMv1-E3SMv2 difference in the SCO is likely associated with the QBO and GW retuning for the E3SMv2.

The evolution of the Antarctic ozone hole during the historical time period reflects the combined effect of dynamics, physics, and chemistry. The NASA Ozone Watch website (https://ozonewatch.gsfc.nasa.gov, last access: October 11, 2021) archives the daily records of the Antarctic ozone hole area (where the total column ozone (TCO) is less than 220 DU) and minimum TCO in the SH based on daily TCO observational data. Figures 20a and b compare the yearly E3SMv2 historical ensemble mean time series with the yearly Ozone Watch observations for the SH minimum TCO and the ozone hole area, respectively. Both the yearly model and observational results are based on the daily data from July 1 to December 31 of each year.

The Antarctic ozone hole emerges about 1980 after the build up of anthropogenic chlorouorocarbons (CFCs) reach a threshold that initiates rapid, catalytic destruction of ozone within the Antarctic stratospheric polar vortex (Molina & Rowland, 1974; Farman et al., 1985). The ozone hole simulation in E3SMv2 is weaker than observed in terms of minimum TCO (Figure 20a,c) and areal extent of the ozone hole (Figure 20b,d). Given the 50 DU high bias for ozone-hole minimum TCO (Figure 20c), the temporal history of the ozone hole, from onset to partial recovery, is well matched in E3SMv2 (Figure 20a). In terms of seasonality, the E3SMv2 ozone hole begins almost a month later and recovers almost a month earlier. The cause of this is not the ozone chemical model, as it works well in other atmospheric models, but is likely to be related to the formation and persistence of the wintertime vortex. The ozone hole is created chemically, but its size and duration depend on the vortex remaining isolated from the mid-latitude stratosphere through-

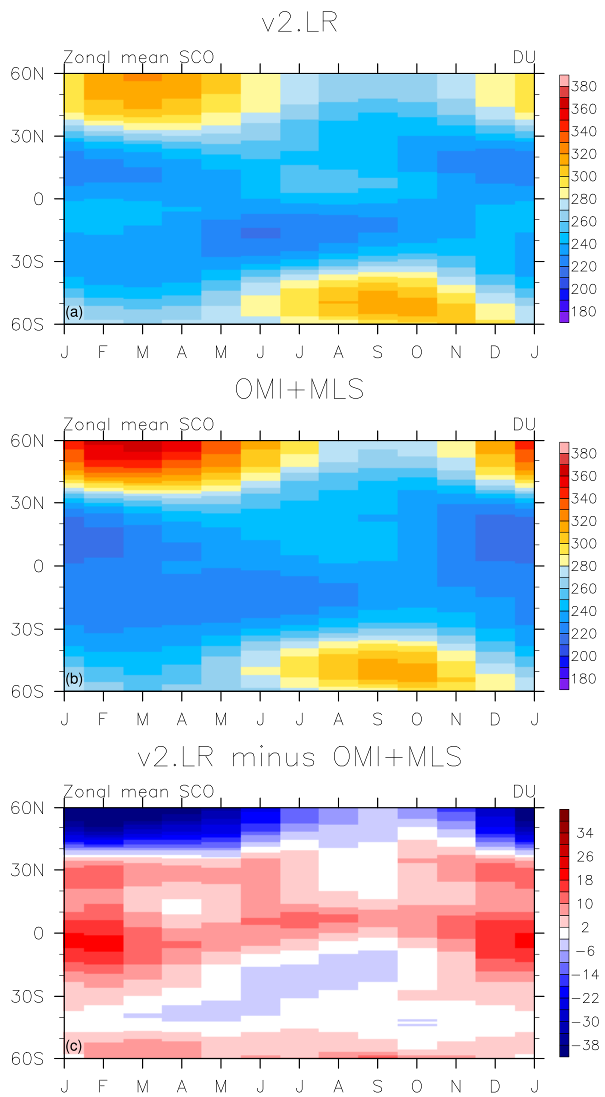


Figure 19. Climatology of zonal mean annual cycle of stratospheric column ozone (SCO, in Dobson units (DU)). The panels are (a) E3SMv2 ensemble mean of historical simulations from years 1995–2014; (b) OMI+MLS observations from years 2005–2017; (c) The differences in SCO of E3SMv2 minus OMI+MLS.

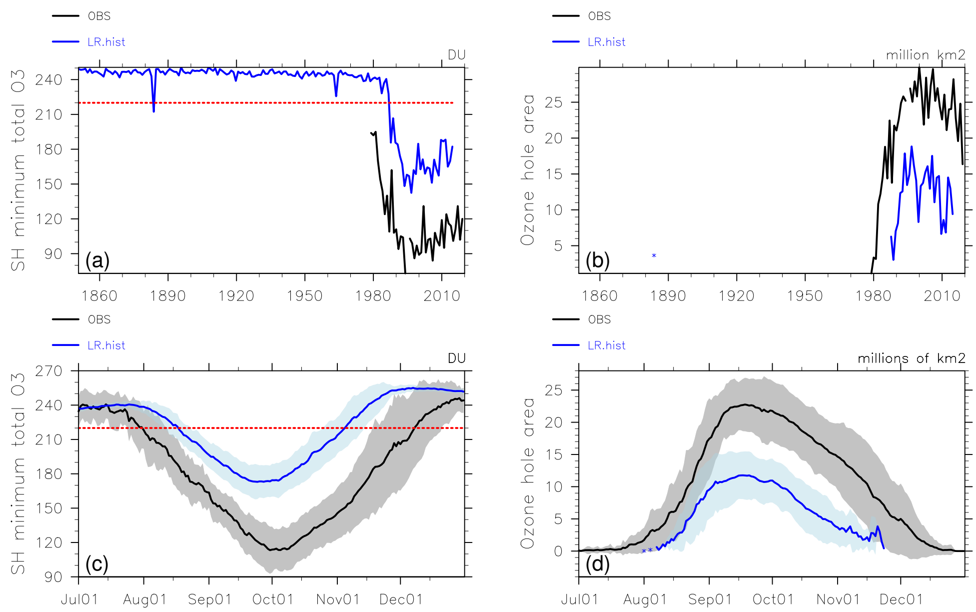


Figure 20. Ozone hole results as shown in the historical time series (top) and daily mean climatology and variance (bottom) of the SH minimum total column ozone (left, unit: DU) and the SH maximum ozone hole area (right, area with total ozone < 220 DU, unit: million km^2) based on the daily data from July 1 to December 31. In the bottom panels, the lines indicate the multi-year average (observations in black from years 1990–2019 and models in blue from years 1990–2014), and shading covers ±1 standard deviation.

out most of the lower stratosphere. The E3SMv2 ozone hole interannual variability (IAV, shaded areas in Figure 20c,d), scaled to the size of the ozone hole, matches the observations, indicating that the vortex IAV is similar to observations. It is possible that the weaker ozone hole in E3SMv2 could be improved with a colder stratosphere, or parametrically, by increasing the PSC temperature threshold.

4.3.5 Aerosols

The global distribution of annual mean AOD at 550 nm from E3SMv2 and E3SMv1 historical simulations (2000-2014) is compared with observational composite (Kinne et al., 2013) in Figure 21. Model results are not included for this comparison over regions where the observations are not available, e.g., in the high latitudes. E3SMv1 and v2 realistically capture the broad regional distribution in AOD, but E3SMv2 has a stronger positive bias than E3SMv1 in the global mean (0.034 vs. 0.013) compared to the observational composite, although the low bias over mid-latitude source regions is improved in E3SMv2. Larger positive biases in E3SMv2 than E3SMv1 are found over tropical and subtropical oceans. Decomposition of the total AOD into major aerosol species is provided in Table 2. The positive biases are mostly due to an increase in anthropogenic aerosol species, particularly sulfate and secondary organic aerosol (SOA). The global annual mean burdens of sulfate and SOA have an increase of 1.03 and 0.95 Tg, respectively, in the E3SMv2 historical simulations (2000-2014) compared to E3SMv1 (Fig. S5). The global annual mean burdens of other anthropogenic aerosol species are also larger in E3SMv2 than those in E3SMv1, although both model simulations use the same set of CMIP6 emissions, in-

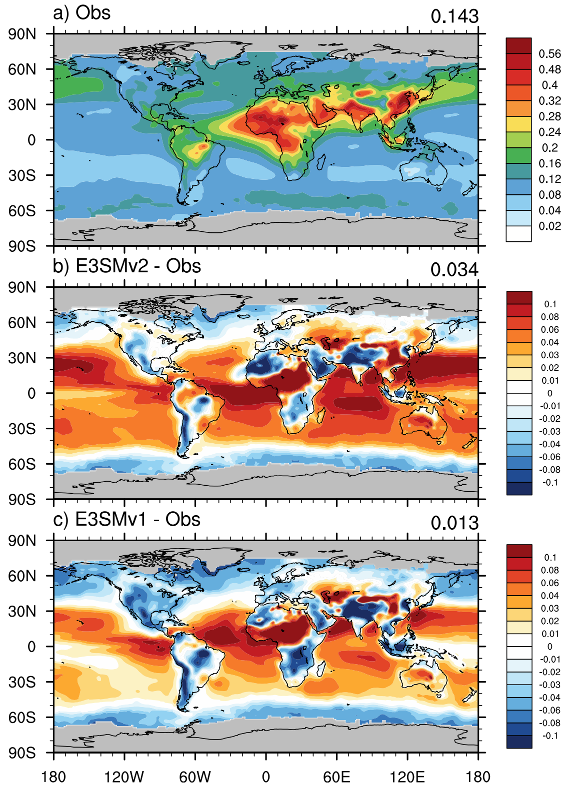


Figure 21. Spatial distributions of global annual mean (a) aerosol optical depth (AOD) from observational composite (Kinne et al., 2013) and the AOD difference between (b) E3SMv2 and (c) E3SMv1, respectively, from the historical simulations (2000-2014) and the observational composite. Areas with gray shading in polar regions indicate missing value. The number at the top-right of each panel represents the global mean.

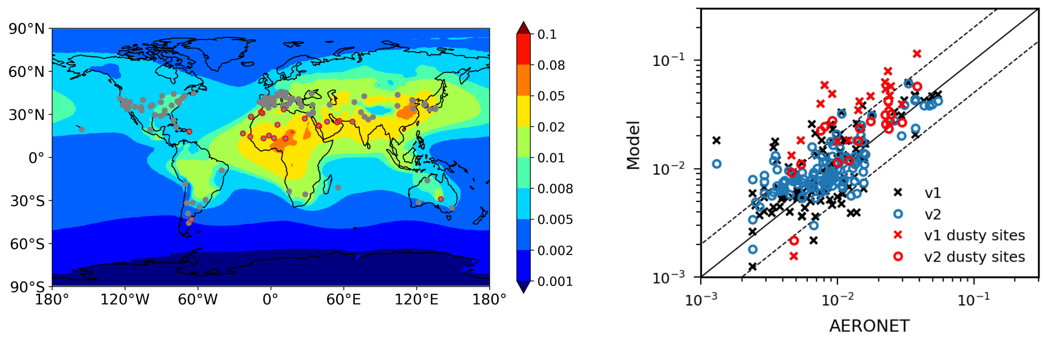
Table 2. Global and annual mean AOD at 550 nm for total aerosol and major aerosol types

AOD (2000-2014)	Total	Dust	Sea salt	Sulfate	POM[a]	BC[a]	SOA[a]
E3SMv1 (DECK)	0.146	0.032	0.049	0.024	0.007	0.0049	0.029
E3SMv2 (historical)	0.166	0.028	0.049	0.033	0.009	0.0063	0.040

[a]POM (particulate organic matter), BC (black carbon), and SOA (secondary organic aerosol)

dicating that the aerosol removal in E3SMv2 is weaker than in E3SMv1. This might be an unintended consequence of intensive cloud and precipitation parameter tuning for EAMv2. Natural aerosols (e.g., dust and sea salt) in E3SMv2 have small changes in their global burdens, as their emissions are scaled to match the global constraints of dust or sea salt optical depth.

In addition to AOD, aerosol absorption of sunlight is also an important parameter in determining the aerosol radiative impacts. As discussed in Section 2.1.4, dust refractive indices in the shortwave were updated in E3SMv2. This leads to better agreement in the simulated aerosol absorption optical depth (AAOD) at 550 nm, as shown in Fig. 22, compared with the long-term average AAOD (2006-2015) derived from the ground-based AERONET measurements (Holben et al., 1998). The compiled AERONET data for AAOD are available at a total of 139 stations globally, and 19 of them with aerosol Ångström exponent <0.8 are denoted as the dusty sites, which are located near the ma-



(a) E3SMv2 AAOD (b) Comparison with AERONET

Figure 22. (a) Global and annual mean aerosol absorption optical depth (AAOD) at 550nm with E3SMv2 for the time period of 2000-2014. The gray dots overlaid on top denote the locations of 139 AERONET stations, of which those circled in red denote the 19 dusty sites, and (b) comparison with the AAOD observations derived from AERONET between 2006-2015 (Holben et al., 1998)

jor dust source regions. Compared to E3SMv1, E3SMv2 simulates smaller AAODs over all the dusty sites, and the calculated multi-site mean is 0.024, reducing the overestimation of E3SMv1 (0.044) by nearly a factor of two against the observations (0.017). Over the other AERONET sites, AAODs in E3SMv2 are generally larger than those in E3SMv1 mainly due to the increased BC burden. Overall, E3SMv2 improves from E3SMv1 (0.017) by predicting a smaller AAOD (0.014) averaged over all the AERONET sites, similar to the observed mean (0.012). The spatial correlation between the modeled and observed AAOD is noticeably improved in E3SMv2, for a larger correlation coefficient (0.83) with the AERONET data than that of E3SMv1 (0.72). Stronger correlation with the observed AOD is also found over the AERONET sites, implying a better representation of aerosol spatial distributions in E3SMv2.

The improvement in the modeled aerosol absorption leads to less aerosol heating in the atmosphere and more aerosol cooling at the top of the atmosphere over the dust-influenced regions in E3SMv2, while the opposite effects occur over the BC-dominated regions. Additionally, we also updated the representation of dust size distribution in emission by accounting for more coarse particles in E3SMv2, which would decrease the net cooling effect of dust but the impact is less than the enhanced cooling due to the lowered dust absorption (Feng et al., 2022).

4.3.6 Historical temperature record

We now compare the time evolution of the global mean surface air temperature in E3SM with the observed historical record. We select the HadCRUT5-Analysis product (Morice et al., 2021); other products are available but the differences are minor compared to the differences with E3SM. Figure 23 shows the temperature anomalies normalized with respect to 1850-1899. As discussed previously (Golaz et al., 2019), E3SMv1 failed to accurately simulate the record by underestimating the warming starting around 1930 but eventually caught up to with the observed record near 2010 because it overestimated the pace of warming from 1990 onward. This was attributed to excessively strong aerosol-related forcing and high climate sensitivity. While both have improved in E3SMv2 – slightly for the aerosol-related forcing and significantly for the sensitivity – E3SMv2 further underestimates the global mean surface temperature during the second half of the record.

E3SMv2 diverges from E3SMv1 around 1930 and remains colder for the remainder of the record. A more in-depth analysis of this shortcoming is provided in Section 5.

As mentioned above, no historical test simulations were performed prior to finalizing E3SMv2. Once the model development was concluded and the first historical simulation complete, the E3SM project made a pragmatic decision to be transparent and release the model version and accompanying simulations, rather than delay in an attempt to correct the problem with the simulation of the global mean temperature in the historical record.

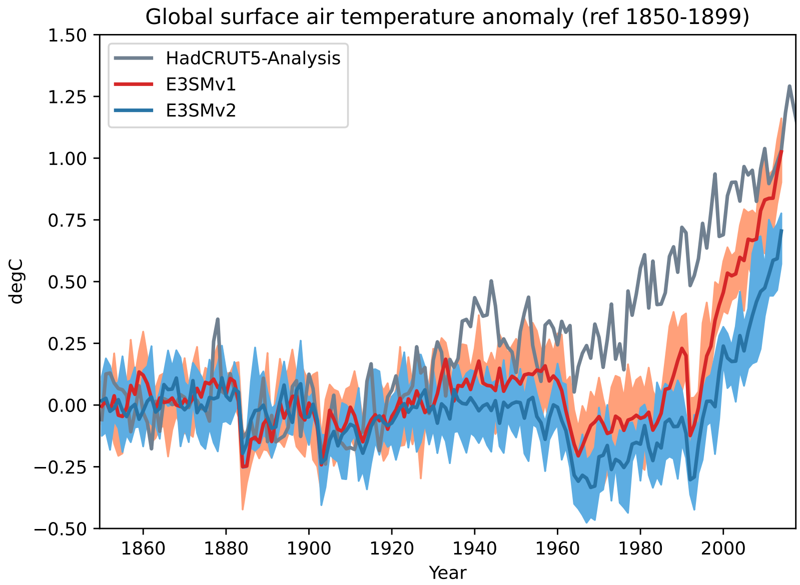


Figure 23. Time evolution of annual global mean surface air temperature anomalies (with respect to 1850-1899). Comparison between observations from HadCRUT5-Analysis (grey), E3SMv1 ensemble mean (red) and range (orange) and E3SMv2 ensemble mean (dark blue) and range (light blue).

5 Historical record: role of GHG vs aerosols

To understand why E3SMv2 fails to accurately simulate the second half of the historical temperature record, we analyze an ensemble of coupled simulations spanning 1850-2014, but selectively activating only certain time varying forcing agents:

- well-mixed greenhouse gases only ("GHG"),
- aerosol and aerosol precursors only, including interactions with clouds ("aer"),
- everything-else, all forcing agents except well-mixed GHG and aerosol ("other").

This decomposition is similar to the DAMIP protocol (Gillett et al., 2016), except for the everything-else configuration, which is similar to natural forcing but includes additional forcing terms (in particular land-use and ozone). We chose this particular decomposition so that all the forcing agents are accounted for within the set. Five ensemble members were run for each decomposition, initialized identically to the five-member ensemble of historical simulations.

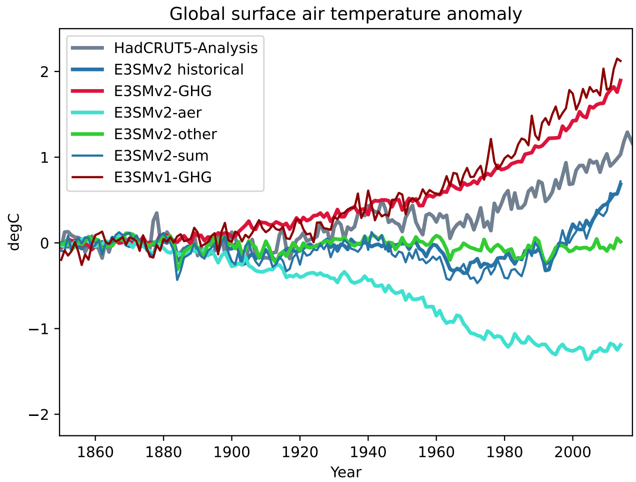


Figure 24. Global annual surface air temperature anomalies for model and observations (gray). For E3SMv2, the decomposition includes contributions from only GHG (red), only aer (turquoise), and other (green). The E3SMv2 historical is in blue, with the sum of individual terms in thin blue. Also shown is E3SMv1 with GHG only forcing (dark red). Observations from HadCRUT5-Analysis are normalized with respect to 1850-1899. Model results are normalized with respect to the 500-year *piControl* simulation.

The time evolution of annual global mean surface air temperature is depicted in Fig. 24. As expected, the dominant forcings are GHG (red) and aerosol-related (turquoise). The remaining forcings (green) show inter-annual variations (mostly from volcanic eruptions and the solar cycle) with little long term trend. A summation over the decomposition (thin blue) recovers the original historical ensemble (thick blue) very well, indicating that the decomposition is linear. The GHG and aerosol contributions almost perfectly mirror each other until approximately 1960, thus explaining the lack of net warming until then (Fig. 23). It is only after the aerosol-related forcing stabilizes around 1990 due to pollution control in North America and Europe that the GHG starts to dominate and E3SMv2 warms as a whole. As discussed previously, E3SMv2 has a lower TCR and ECS compared to E3SMv1. As a result, the warming from GHG alone is weaker than in v1 (dark red; Zheng et al., 2021). The two models diverge mostly after 1960 which helps explain why E3SMv2 remains colder longer.

Equipped with this decomposition and under the assumption of linearity, we can investigate hypothetical configurations with different relative strengths of GHG and aerosol. We can write any variable ψ as:

$$\psi_{\text{all}} = \psi_{\text{piControl}} + \alpha_{\text{GHG}} \left(\psi_{\text{GHG}} - \psi_{\text{piControl}}\right) + \alpha_{\text{aer}} \left(\psi_{\text{aer}} - \psi_{\text{piControl}}\right) + \left(\psi_{\text{other}} - \psi_{\text{piControl}}\right) \quad (1)$$

This reconstruction is conceptually similar to Neelin et al. (2010), but applied to different forcing terms rather than physics parameter perturbations. Setting $\alpha_{\text{GHG}} = \alpha_{\text{aer}} = 1$ recovers the all-forcing configuration as long as the decomposition is linear. We call this configuration “composite base”. Linearity is a very good approximation for annual global averages (Fig 24). It also holds well for two-dimensional and three-dimensional climatological fields as demonstrated in Fig. S6: RMSE for the composite base configuration (red stars) and E3SMv2 (red triangles) are very similar for most fields and seasons.

We note however that individual terms in Eq. 1 are derived from five-member averages, and therefore the reconstruction is not expected to realistically capture natural multidecadal variability. While multidecadal variability plays an important role (e.g. Zeng & Geil, 2016) it is clearly not sufficient to explain the mismatch between E3SMv2 and observations (Fig. 23).

We can vary α_{GHG} and α_{aer} in Eq. 1 (with ψ set to surface air temperature) to construct hypothetical composite model configurations. Varying α_{GHG} modulates the model response to GHG (akin to modulating TCR and the shorter time periods in ECS), while α_{aer} modulates the model response to aerosols (akin to modulating the magnitude of the aerosol-related forcing and feedback).

We construct a loss function that quantifies the mismatch between modeled and observed surface air temperature separately in the northern (NH) and southern hemispheres (SH):

$$F = \sum_{SH,NH} \left(\sum_{yr=1950}^{2014} (\bar{T}_{\mathrm{model}} - \bar{T}_{\mathrm{obs}})^2 \right)^{1/2} \tag{2}$$

We opt to separately account for SH and NH due to the strong asymmetry in aerosol forcing. We also select the latter part of the historical record (1950-2014) when observational uncertainties are smaller. Changing those assumptions (global average, entire historical record) does not fundamentally change the results.

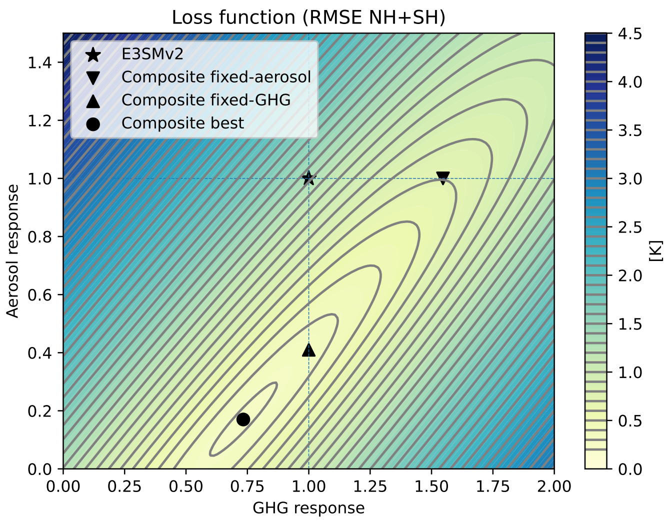


Figure 25. Loss function from Eq. 2. Star represents E3SMv2, circle global minimum, and triangles local minima by minimizing along a single dimension.

The loss function F is shown in Fig. 25 as a function of α_{GHG} and α_{aer}. The surface depicts a broad valley oriented diagonally. The global minimum (composite best) is situated at $\alpha_{\mathrm{GHG}} = 0.73$ and $\alpha_{\mathrm{aer}} = 0.17$, indicating that improving the historical temperature record simulated by E3SMv2 would require a modest reduction in response from GHG, but a very substantial one from the aerosols. Also shown in Fig. 25 are two local minima. One holding GHG constant (composite fixed-GHG; $\alpha_{\mathrm{GHG}} = 1$ and $\alpha_{\mathrm{aer}} = 0.41$) and one holding aerosol constant (composite fixed-aerosol; $\alpha_{\mathrm{GHG}} = 1.55$ and $\alpha_{\mathrm{aer}} = 1$). The first local minimum is much closer to the global one compared to the second one, confirming that aerosols are the dominant source of the mismatch.

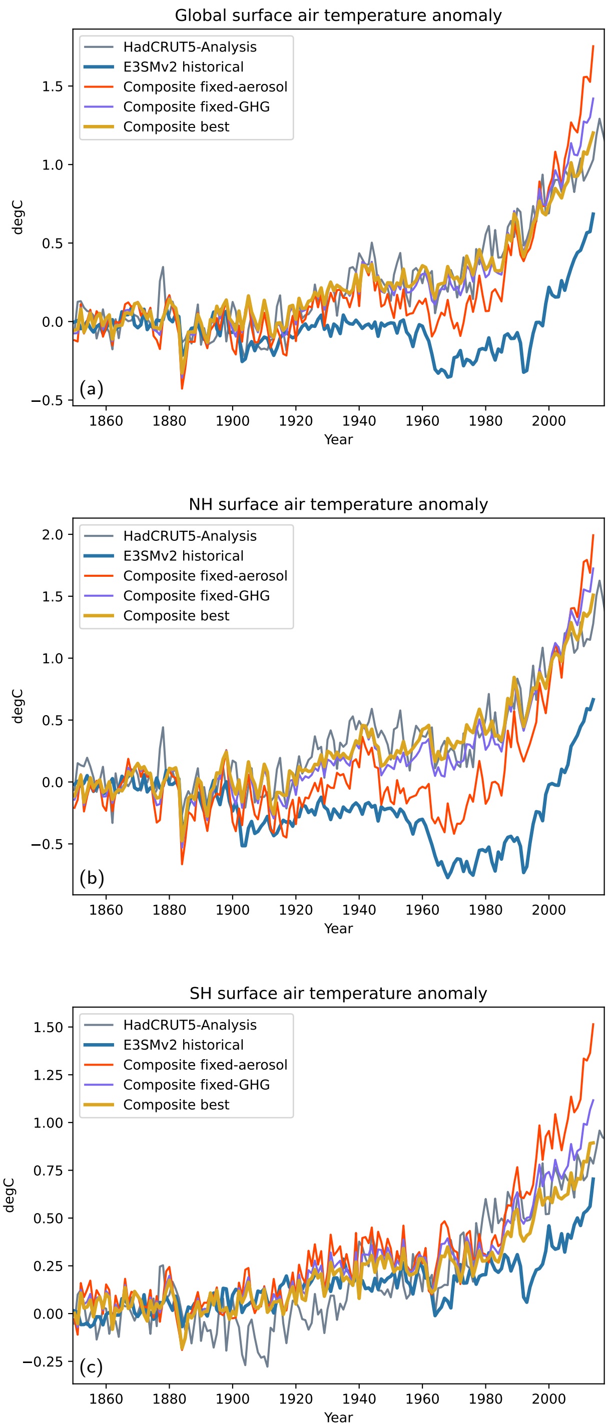


Figure 26. Surface air temperature anomalies (with respect to 1850-1899) for (a) global, (b) northern hemisphere and (c) southern hemisphere. Lines shown include observations (HadCRUT5-Analysis, grey), E3SMv2 (blue), and composite configurations from Fig. 25 (red, purple, gold).

This can be further illustrated by constructing global and hemispheric temperature time series corresponding to these composite configurations (Fig. 26). The composite best solution (gold) corresponding to the global minimum improves considerably upon E3SMv2 and matches the historical record best for each region (global, NH, SH). Composite fixed-GHG (purple) also does an adequate job, but with some indication of excessive warming in the 2000s due to its higher response to GHG. Composite fixed-aerosol (orange), which increases the response of GHG to balance the strong aerosol cooling fails to match the historical record well. This confirms the argument that higher sensitivity cannot adequately compensate for excessive aerosol forcing owing to the presence of a plateau in the aerosol forcing and hemispheric asymmetry (e.g. Zhao et al., 2018; Albright et al., 2021).

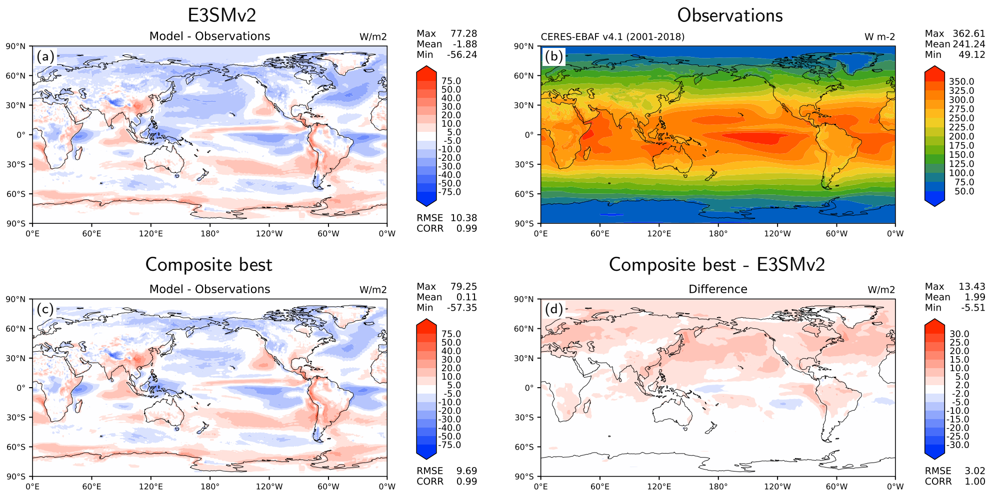


Figure 27. Net TOA SW radiation: observations (CERES-EBAF 4,1; b), model error for E3SMv2 (a), composite best configuration from Fig. 25 (c) and difference between E3SMv2 and composite best configuration (d). Model averages are computed over 1985-2014.

Finally, we also reconstruct climatological fields for the period 1985-2014 using Eq. 1. Figure 27 shows the top-of-atmosphere SW net radiation. Remarkably, the NH negative bias in E3SMv2 (blue shading in Fig 27a) is greatly reduced in composite best (Fig. 27c) which becomes much closer to observations regionally, especially over the N Atlantic and N Pacific oceans. Global metrics also improve with a reduced mean bias (0.11 vs -1.88 W/m^2) and RMSE (9.69 vs 10.38 W/m^2). A similar picture emerges for the sea-surface temperature (Fig. 28) with substantial reductions in regional cold biases in the NH. SH SST biases are essentially unchanged, pointing to a different cause.

Taken together, our results indicate that a substantial reduction in the aerosol forcing would not only improve the match with the historical temperature record, but also improve aspects of the present-day climatology. Other fields, for example precipitation exhibit much smaller impact as seen in Figure S6 by comparing the gold (composite best) and red stars (composite base). This is reassuring in the sense that E3SMv2, despite its shortcomings, can still serve as a useful model for many studies.

5.1 Impacts on Polar Climate

In the historical ensemble (Fig. 29), Northern Hemisphere sea ice extent and volume both increase over the time period 1850-1978, and decrease after the mid-1980s, as

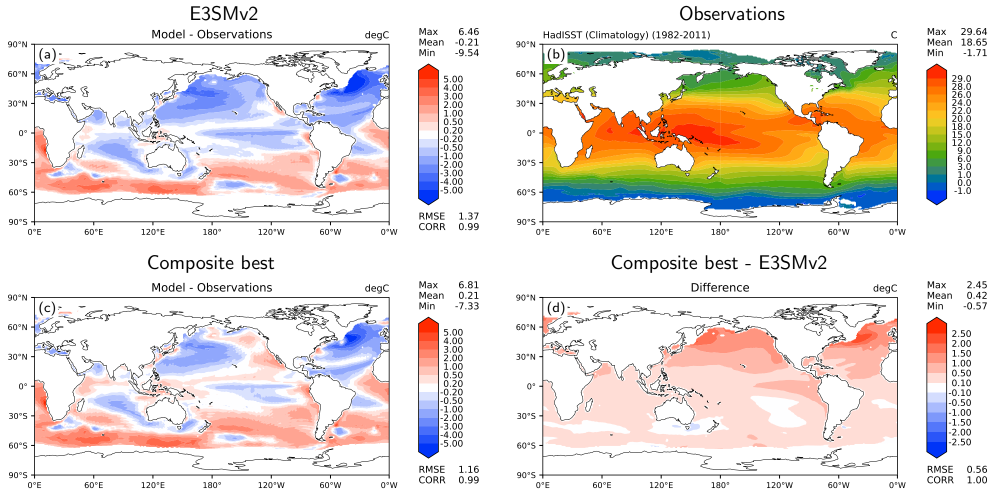


Figure 28. Same as Fig. 27 but for SST. Sea ice covered regions are excluded from the averaging.

observed. This behavior is consistent with changes in the ocean heat content (Fig. 29e) and surface air temperature anomalies for the historical simulations (Fig. 26b).

The maximum ice extent in the Arctic is larger in v2 than in v1, while it is smaller in the Southern Hemisphere. The minimum ice extent is similar for v1 and v2 in both hemispheres. E3SMv1's large, cold SST bias in the North Atlantic and associated anomalous sea ice in the Labrador Sea remains in v2, although it improves in the aerosol sensitivity "composite" simulations (Fig. 28c). Unlike v1, which featured both warm and cold SST biases in the Northern Hemisphere, the Northern Hemisphere in v2 is too cold over its entirety (only regions outside of the sea ice pack are shown in Fig. 28a), and so greater sea ice extent in v2 is not surprising. The Southern Hemisphere is still biased warm, but not as badly as in v1, and sea ice in the Southern Ocean is not extensive enough compared with the climate data record, year-round, in v2.

Trends during the satellite era (Fig. 29, right panels) indicate that the model extent is decreasing faster than observed in the Arctic, consistent with the faster increase in surface air temperatures than observed (Fig. 26b). The ice extent trend has the opposite sign compared with observations in the Antarctic, as in many other models, and the change in volume extremes (Δ) between 1850 and 2015 is decreasing.

A counter-intuitive result is that extremes in the ice extent and volume in the historical simulations (left column of Fig. 29) are generally larger than in the pre-industrial control, with a greater range of variability. However, this behavior is consistent with the aerosol forcing biases discussed in Section 5. Cloud radiative effects is still biased positive in the polar regions, although v2 has improved over v1.

The net effect of improvements to the radiative and snow schemes in v2 only minimally impacts the climatic state of sea ice, indicating that biases in prior v1 simulations were not fundamentally due to critical faults in these parameterizations. Lack of conservation in the ice-ocean mass coupling scheme played a much more important role; the correction of mass exchanges between the upper ocean and sea ice models to account for brine content in the sea ice thickens the Arctic ice pack in summer, reducing a bias from v1 (Fig. 29c, left column), while minimally impacting ice in the Southern Ocean (Fig. 29d).

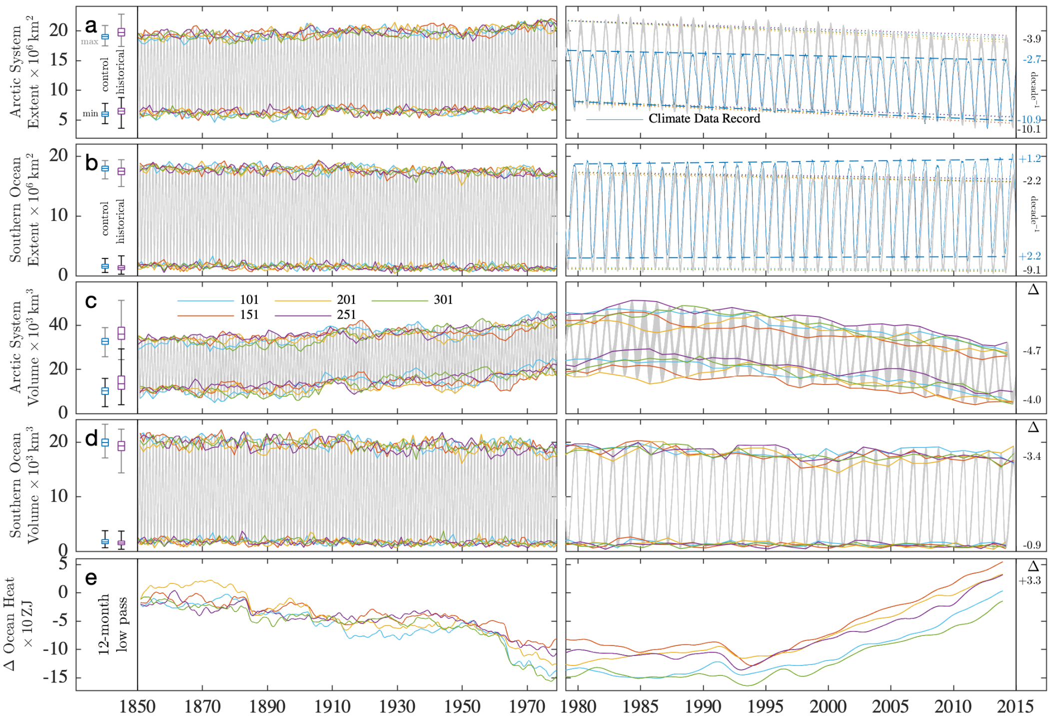


Figure 29. Daily sea ice extent (a, b) and volume (c, d) evolution across five ensemble members for the Northern and Southern Hemispheres, respectively, divided at the beginning of the core passive-microwave observation period in 1979 and compared to the change in 12-month filtered total ocean heat content from the start of the historical period in (e). Box plots in the left column compare annual extremes from daily values of the 500-year pre-industrial control (blue) with the industrial-era 5-member ensemble (purple). Trace colors for the year of the control simulation from which the ensemble members were spawned are indicated in (c; 101, 151, 201, 251, 300). Linear decadal trend in annual maximum and minimum daily extent is indicated in the right column for the ensemble mean of each ensemble trend line from 1979 to 2015, as compared to the Meier et al. (2017) NOAA Climate Data Record for (a) and (b). The right column in (c), (d) and (e) indicates the change (Δ) in the ensemble mean of volume extremes and non-filtered ocean content between 1850 and 2015.

With this mass-conserving scheme, the maximum and minimum sea ice areas are now stable in both hemispheres for the 500-year pre-industrial simulations, as shown in Fig. 4c.

6 Summary and conclusion

By design, E3SMv2 represents an evolution from E3SMv1 and as such resembles E3SMv1 in many aspects. There are nevertheless notable differences that justified a new model release and associated simulation campaign.

- E3SMv2 is approximately twice as fast (or efficient if measured in terms of power) compared to v1 (Fig. 2). The efficiency gains are achieved in the atmosphere and ocean components. In the atmosphere, they arise from a new semi-Lagrangian tracer transport method and a new grid for physics calculations (Fig. 3). The gain in the ocean is due to a longer timestep.
- The atmospheric physics, while based on the same basic set of parameterizations as v1, underwent significant retuning in v2. Many improvements from the intermediate EAMv1p configuration (Ma et al., 2022) are incorporated with additional changes to further improve clouds and precipitation (e.g. Figs. 9,11,12,13,14).
- A new convective trigger function for the deep convection (Xie et al., 2019) significantly improves the phase of the diurnal cycle of precipitation, but the amplitude remains weaker than observed (Fig. 15).
- E3SMv2 captures important modes of variability such as ENSO (Fig. 16) and MJO (Fig. 18). However, the ENSO spectrum has excessive energy at short periods ($\sim$2.5 years) and is too weak for longer periods (6-9 years). MJO phase speed is realistic west of 125°E, but then exceeds observations east of it. Tropical variability is significantly too weak (Fig. 17).
- A more realistic treatment of ozone is implemented (Tang et al., 2021). It captures the seasonal cycle of stratospheric column ozone (Fig. 19) and the ozone hole in the historical period, although the size is underestimated (Fig. 20).
- Dust aerosol optical properties and particle size distributions are revised, resulting in a better prediction of mean AAOD over dusty AERONET sites (Fig. 22). Burdens of sulfate and SOA aerosols increase as an unintended consequence of cloud tuning efforts, giving rise to a slightly overestimated global mean AOD despite regional improvements (Fig. 21).
- E3SMv2 is less sensitive to GHG forcing (Fig. 5). ECS is reduced significantly compared to v1 (4.0 K vs 5.3 K) which is mostly attributable to a smaller cloud feedback. The ECS value of 4.0 K is plausible as assessed by WCRP (Sherwood et al., 2020). This is a substantial achievement compared to the unrealistically high sensitivity of E3SMv1. On shorter time scales, TCR is also reduced to 2.4 from 2.9 K.
- The effective aerosol forcing ($\mathrm{ERF_{aer}}$ = -1.5 $\mathrm{Wm^{-2}}$) remains essentially unchanged in E3SMv2 (Fig. 6). This value is within the likely range assessed by WCRP (Bellouin et al., 2020). Some changes were made in v2 that reduced the magnitude of $\mathrm{ERF_{aer}}$, but their impact was negated by changes elsewhere in the cloud physics (convection).
- E3SMv2 significantly underestimates the global mean temperature in the second half of the historical temperature record (Fig. 23). An analysis of single-forcing simulations indicate that correcting the historical record would require a substantial reduction in the magnitude of $\mathrm{ERF_{aer}}$ (60 to 80%), and possibly a more modest reduction in the model's response to GHG (Figs. 25, 26). A reduction in $\mathrm{ERF_{aer}}$ would furthermore reduce regional biases in TOA radiative fluxes and SST (Fig. 27, 28). Other fields are less impacted (e.g. precipitation; Fig. S6), indicating that E3SMv2 can still serve as a useful tool despite its shortcomings.

- Proper conservation of mass in ocean/sea-ice exchanges increased Arctic sea ice volume, improving a low-thickness bias from v1, while impacting the Southern Ocean ice pack very little. Changes to the radiation and snow physics parameterizations had little net effect, highlighting the importance of coupled interactions over internal sea ice processes in the climate system (Hunke, 2010). The sea ice simulations shown here are largely consistent with the overall climatic environment, including excessively cool surface air and ocean temperatures.

This release of E3SMv2 serves as a starting point for additional configurations. They include regionally refined configurations with higher resolution over North America and, separately, the Southern Ocean. A configuration with interactive biogeochemistry is also under development. While E3SMv2 improves upon its predecessor in many aspects, significant work remains. The highest priorities for future releases of E3SM are addressing the weak AMOC and the poor historical temperature record.

Appendix A Atmosphere configuration

Table A1: List of the atmospheric tuning parameters. Note: the value of *microp_aero_wsubmin* was set to 0.001 for v1p and v2 based on Ma et al. (2021). However, an additional lower bound is present in the code that effectively resets it to 0.1 consistent with Ma et al. (2022).

Scheme	**Parameter**	**v2**	**v1**	**v1p**	**Short Description**
CLUBB	*clubb_c14*	2.5	1.06	2.0	Dissipation of $\overline{u'^2}$ and $\overline{v'^2}$
	clubb_c1	2.4	1.335	2.4	Low-skewness value of dissipation of $\overline{w'^2}$
	clubb_c1b	2.8	1.335	2.8	High-Skw value of dissipation of $\overline{v'^2}$
	clubb_c1c	0.75	1.0	0.75	Smoothness of transition between high-Skw and low-Skw for the dissipation of $\overline{v'^2}$
	clubb_c6rtb	7.5	6.0	7.5	High-Skw value of pressure damping of water flux
	clubb_c6rtc	0.5	1.0	0.5	Smoothness of transition between high-Skw and low-Skw for the pressure damping of water flux
	clubb_c6thlb	7.5	6.0	7.5	High-Skw value of pressure damping of heat flux
	clubb_c6thlc	0.5	1.0	0.5	Smoothness of transition between high-Skw and low-Skw for the pressure damping of heat flux
	clubb_c8	5.2	4.3	5.2	Pressure damping of $\overline{w'^3}$
	clubb_c11	0.7	0.8	0.7	Buoyancy damping of $\overline{2'^3}$ at low Skw
	clubb_c11b	0.2	0.35	0.2	Buoyancy damping of $\overline{2'^3}$ at high Skw
	clubb_c11c	0.85	0.5	0.85	Smoothness of transition between high Skw and low Skw for the buoyancy damping of $\overline{2'^3}$
	clubb_c_k10	0.35	0.3	0.35	Coefficient of momentum diffusivity, Kh_zm
	clubb_c_k10h	0.35	0.3	0.35	Coefficient of thermodynamic diffusivity, Kmh_zm
	clubb_gamma_coef	0.12	0.32	0.12	Constant of the width of PDF in w-coordinate
	clubb_gamma_coefb	0.28	0.32	0.28	High-skw value of gamma coefficient
	clubb_gamma_coefc	1.2	5.0	1.2	Smoothness of transition between values of gamma coefficient

	clubb_mu	$5e^{-4}$	$1e^{-3}$	$5e^{-4}$	Fractional parcel entrainment rate per unit height [1/m]
	clubb_wpxp_l_thresh	100.0	60	100	Threshold in length scale below which extra damping is applied to C6 and C7 functions [m]
	clubb_ice_deep	$14e^{-6}$	$16e^{-6}$	$14e^{-6}$	Radius of ice particles detrained from deep convection [m]
	$cldfrc\_dp1$	0.018	0.045	0.018	parameter for deep convective cloud fraction
	clubb_use_sgv	True	False	True	Enables subgrid features gustiness, tpert, and thv fix
	clubb_ipdf_call_placement	1	2	1	Select the placement of the call to CLUBB's PDF: 1 - before advancing CLUBB's predictive fields, 2 - after, 3 - both before and after
ZM	*zmconv_alfa*	0.14	0.1	0.14	Maximum downdraft mass flux fraction
	zmconv_c0_lnd	0.002	0.007	0.002	Autoconversion coefficient over land for deep convection
	zmconv_c0_ocn	0.002	0.007	0.002	Autoconversion coefficient over ocean for deep convection
	zmconv_dmpdz	-0.7e^{-3}	-0.7e^{-3}	-1.2e^{-3}	Parcel fractional mass entrainment rate
	zmconv_mx_bot_lyr_adj	1	2	1	Bottom layer adjustment for setting "launching" level of maximum moist static energy
	zmconv_tp_fac	2	0	2	Tpert scale factor in ZM deep convection scheme
MG2	*cld_sed*	1.0	1.0	1.8	Scale factor for cloud droplet sedimentation
	ice_sed_ai	500	500	1200	Cloud ice fall speed parameter
	micro_mg_berg_eff_factor	0.7	0.1	0.7	Efficiency factor for WBF processes
	micro_mg_accre_enhan_fac	1.75	1.5	1.75	Accretion enhancement factor
	prc_exp1	-1.4	-1.2	-1.4	Tunable exponent coefficient for autoconversion
	micro_mincdnc	10.D6	0.0	0.0	Minimum cloud droplet number concentration imposed when $micro\_mincdnc > 0$ $[m^{-3}]$

nucleate	*so4_sz_thresh_icenuc*	$0.08e^{-6}$	$0.05e^{-6}$	$0.08e^{-6}$	Aitken mode SO2 size threshold for ice nucleation
microp aero	*microp_aero_wsubmin*	0.1	0.2	0.1 See note in caption	Minimum subgrid vertical velocity
aerosol	*seasalt_emis_scale*	0.6	0.85	0.6	Tuning factor for sea salt aerosol emission
dust	*dus_emis_fact*	1.5	2.05	2.8	Tuning parameter for dust emissions
Linoz	*linoz_psc_t*	197.5	193.0	193.0	Tunable Linoz PSC ozone loss temperature threshold (K)
Gravity wave drag	*gw_convect_hcf*	10.0	20.0	20.0	Heating rate conversion factor associated with convective gravity waves
	effgw_beres	0.35	0.40	0.40	Efficiency associated with convective gravity waves from the Beres scheme
	effgw_oro	0.375	0.25	0.25	Efficiency associated with orographic gravity waves

Acknowledgments

This research was supported as part of the Energy Exascale Earth System Model (E3SM) project, funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research. E3SM production simulations were performed on a high-performance computing cluster provided by the BER Earth System Modeling program and operated by the Laboratory Computing Resource Center at Argonne National Laboratory. Developmental simulations were also performed using BER Earth System Modeling program's Compy computing cluster located at Pacific Northwest National Laboratory. Additional developmental simulations, as well as post-processing and data archiving of production simulations used resources of the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This work was partially supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement 1852977. Portions of this study were supported by the Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program of the U.S. Department of Energy's Office of Biological and Environmental Research (BER) via NSF Interagency Agreement 1844590.

All model codes may be accessed on the GitHub repository at `https://github.com/E3SM-Project/E3SM`. A maintenance branch (`maint-2.0`; `https://github.com/E3SM-Project/E3SM/tree/maint-2.0`) has been specifically created to reproduce these simulations. Bit-for-bit results with the original simulations on identical machines will be maintained on that branch for as long as the computing environment supports it. Complete native model output is accessible directly on NERSC at `https://portal.nersc.gov/archive/home/projects/e3sm/www/WaterCycle/E3SMv2/LR`. A subset of the data reformatted following CMIP conventions is available through the DOE Earth System Grid Federation (ESGF) at `https://esgf-node.llnl.gov/projects/e3sm` (NOTE TO REVIEWERS: CONVERSION AND PUBLICATION IN CMIP FORMAT IS ON-GOING). Performance data and scripts for Figures 2 and 3 are available at `https://github.com/`

`E3SM-Project/perf-data/tree/main/v2-overview/chrysalis-perf-study`; see the readme.txt file there for further details.

Lawrence Livermore National Laboratory is operated by Lawrence Livermore National Security, LLC, for the U.S. Department of Energy, National Nuclear Security Administration under Contract DE-AC52-07NA27344. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. Pacific Northwest National Laboratory is operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RL01830. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References

Adler, R., Sapiano, M., Huffman, G., Wang, J.-J., Gu, G., Bolvin, D., … Shin, D.-B. (2018). The global precipitation climatology project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, *9*(4), 138. doi: 10.3390/atmos9040138

Ahn, M.-S., Kim, D., Kang, D., Lee, J., Sperber, K. R., Gleckler, P. J., … Kim, H. (2020, June). MJO propagation across the maritime continent: Are CMIP6 models better than CMIP5 models? *Geophys. Res. Lett.*, *47*(11). doi: 10.1029/2020gl087250

Aksenov, Y., Popova, E. E., Yool, A., Nurser, A. J., Williams, T. D., Bertino, L., & Bergh, J. (2017). On the future navigability of Arctic sea routes: High-resolution projections of the Arctic Ocean and sea ice. *Mar. Policy*, *75*. doi: 10.1016/j.marpol.2015.12.027

Albright, A. L., Proistosescu, C., & Huybers, P. (2021). Origins of a relatively tight lower bound on anthropogenic aerosol radiative forcing from bayesian analysis of historical observations. *Journal of Climate*, *34*(21), 8777–8792. doi: 10.1175/jcli-d-21-0167.1

Arctic Council. (2009). *Arctic Marine Shipping Assessment 2009 Report* (Summary Report). Arctic Council's Protection of the Arctic Marine Environment (PAME). Retrieved from `http://hdl.handle.net/11374/54`

Bailey, D., Holland, M., Hunke, E., Lipscomb, B., Briegleb, B., Bitz, C., & Schramm, J. (2011). *Community Ice CodE (CICE) User's Guide – Version 4.0* (User Guide). National Center of Atmospheric Research. Retrieved from `https://www.cesm.ucar.edu/models/cesm1.0/cice/ice_usrdoc.pdf`

Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., … Stevens, B. (2020). Bounding global aerosol radiative forcing of climate change. *Reviews of Geophysics*, *58*(1). doi: 10.1029/2019rg000660

Beres, J. H., Alexander, M. J., & Holton, J. R. (2004). A method of specifying the gravity wave spectrum above convection based on latent heating properties and background wind. *Journal of the Atmospheric Sciences*, *61*(3), 324 - 337. doi: 10.1175/1520-0469(2004)061⟨0324:AMOSTG⟩2.0.CO;2

Bitz, C. M., & Lipscomb, W. H. (1999). An energy-conserving thermodynamic model of sea ice. *Journal of Geophysical Research: Oceans*, *104*(C7), 15669-15677. doi: 10.1029/1999JC900100

Bradley, A. M., Bosler, P. A., & Guba, O. (2021). Islet: Interpolation semi-lagrangian element-based transport. *Geosci. Model Dev. Discuss. [preprint]*, *2021*, 1–48. doi: 10.5194/gmd-2021-296

Bradley, A. M., Bosler, P. A., Guba, O., Taylor, M. A., & Barnett, G. A. (2019). Communication-efficient property preservation in tracer transport. *SIAM Journal on Scientific Computing*, *41*(3), C161–C193. doi: 10.1137/18m1165414

Briegleb, B., & Light, B. (2007). *A Delta-Eddington Multiple Scattering Parameterization for Solar Radiation in the Sea Ice Component of the Community Climate System Model* (NCAR Technical Note Nos. NCAR/TN-472+STR). National Center of Atmospheric Research. doi: 10.5065/D6B27S71

Brown, J. R., Brierley, C. M., An, S.-I., Guarino, M.-V., Stevenson, S., Williams, C. J. R., … Zheng, W. (2020). Comparison of past and future simulations of ENSO in CMIP5/PMIP3 and CMIP6/PMIP4 models. *Climate of the Past*, *16*(5), 1777–1805. doi: 10.5194/cp-16-1777-2020

Brunke, M. A., Ma, P.-L., Eyre, J. E. J. R., Rasch, P. J., Sorooshian, A., & Zeng, X. (2019). Subtropical marine low stratiform cloud deck spatial errors in the E3SMv1 Atmosphere Model. *Geophysical Research Letters*, *46*(21), 12598–12607. doi: 10.1029/2019gl084747

Burrows, S. M., Maltrud, M., Yang, X., Zhu, Q., Jeffery, N., Shi, X., … Leung, L. R. (2020). The DOE E3SM v1.1 biogeochemistry configuration: Description and simulated ecosystem-climate responses to historical changes in forcing. *Journal of Advances in Modeling Earth Systems*, *12*(9). doi: 10.1029/2019ms001766

Caldwell, P. M., Mametjanov, A., Tang, Q., Van Roekel, L. P., Golaz, J. C., Lin, W., … Zhou, T. (2019). The DOE E3SM coupled model version 1: Description and results at high resolution. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4095–4146. doi: 10.1029/2019MS001870

Caldwell, P. M., Terai, C. R., Hillman, B. R., Keen, N. D., Bogenschutz, P. A., Lin, W., … et al. (2021). Convection-permitting simulations with the E3SM global atmosphere model. *J. Adv. Modeling Earth Systems*, *13*. doi: 10.1029/2021MS002544

Campin, J.-M., Marshall, J., & Ferreira, D. (2008). Sea ice-ocean coupling using a rescaled vertical coordinate z*. *Ocean Model.*, *24*(1-2), 1–14. doi: 10.1016/J.Ocemod.2008.05.005

Cariolle, D., Lasserre-Bigorry, A., Royer, J.-F., & Geleyn, J.-F. (1990). A general circulation model simulation of the springtime antarctic ozone decrease and its impact on mid-latitudes. *Journal of Geophysical Research: Atmospheres*, *95*(D2), 1883-1898. doi: 10.1029/JD095iD02p01883

Cess, R. D., Potter, G. L., Blanchet, J. P., Boer, G. J., Ghan, S. J., Kiehl, J. T., … Yagai, I. (1989). Interpretation of cloud-climate feedback as produced by 14 atmospheric general circulation models. *Science*, *245*(4917), 513-516. doi: 10.1126/science.245.4917.513

Charron, M., & Manzini, E. (2002). Gravity waves from fronts: Parameterization and middle atmosphere response in a general circulation model. *Journal of the Atmospheric Sciences*, *59*(5), 923 - 941. doi: 10.1175/1520-0469(2002)059⟨0923:GWFFPA⟩2.0.CO;2

Comeau, D., Asay-Davis, X. S., Begeman, C. B., Hoffman, M. J., Lin, W., Petersen, M. R., … Turner, A. K. (2022). The DOE E3SM v1.2 cryosphere configuration: Description and simulated antarctic ice-shelf basal melting. *Journal of Advances in Modeling Earth Systems*, *14*(2). doi: 10.1029/2021ms002468

Craig, A. P., Vertenstein, M., & Jacob, R. (2012). A new flexible coupler for earth system modeling developed for CCSM4 and CESM1. *The International Journal of High Performance Computing Applications*, *26*(1), 31–42. doi: 10.1177/1094342011428141

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., … Strand, W. G. (2020, February). The Community Earth System Model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*(2). doi: 10.1029/2019ms001916

Danabasoglu, G., & Marshall, J. (2007). Effects of vertical variations of thickness diffusivity in an ocean general circulation model. *Ocean Modelling*, *18*(2), 122–141. doi: 10.1016/j.ocemod.2007.03.006

Danabasoglu, G., & Williams, J. C. M. (1995). Sensitivity of the global ocean circulation to parameterizations of mesoscale tracer transports. *Journal of Climate*, *8*(12), 2967–2987. doi: 10.1175/1520-0442(1995)008⟨2967:sotgoc⟩2.0.co;2

Dang, C., Zender, C. S., & Flanner, M. G. (2019). Intercomparison and improvement of two-stream shortwave radiative transfer schemes in earth system models for a unified treatment of cryospheric surfaces. *The Cryosphere*, *13*(9), 2325–2343. doi: 10.5194/tc-13-2325-2019

Dennis, J., Fournier, A., Spotz, W. F., St-Cyr, A., Taylor, M. A., Thomas, S. J., & Tufo, H. (2005). High-resolution mesh convergence properties and parallel efficiency of a spectral element atmospheric dynamical core. *The International Journal of High Performance Computing Applications*, *19*(3), 225–235. doi: 10.1177/1094342005056108

Dennis, J. M., Edwards, J., Evans, K. J., Guba, O., Lauritzen, P. H., Mirin, A. A., … Worley, P. H. (2011). CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model. *The International Journal of High Performance Computing Applications*, *26*(1), 74–89. doi: 10.1177/1094342011428142

Dennis, J. M., Edwards, J., Loy, R., Jacob, R., Mirin, A. A., Craig, A. P., & Vertenstein, M. (2012). An application-level parallel I/O library for Earth system models. *The International Journal of High Performance Computing Applications*, *26*(1), 43-53. doi: 10.1177/1094342011428143

Dubovik, O., Smirnov, A., Holben, B. N., King, M. D., Kaufman, Y. J., Eck, T. F., & Slutsker, I. (2000). Accuracy assessments of aerosol optical properties retrieved from aerosol robotic network (AERONET) sun and sky radiance measurements. *Journal of Geophysical Research: Atmospheres*, *105*(D8), 9791–9806. doi: 10.1029/2000jd900040

Engwirda, D. (2017). JIGSAW-GEO (1.0): locally orthogonal staggered unstructured grid generation for general circulation modelling on the sphere. *Geoscientific Model Development*, *10*(6), 2117–2140. doi: 10.5194/gmd-10-2117-2017

Engwirda, D. (2018). Generalised primal-dual grids for unstructured co-volume schemes. *Journal of Computational Physics*, *375*, 155–176. doi: 10.1016/j.jcp.2018.07.025

Engwirda, D., & Ivers, D. (2016). Off-centre Steiner points for Delaunay-refinement on curved surfaces. *Computer-Aided Design*, *72*, 157–171. doi: 10.1016/j.cad.2015.10.007

Evans, K., Lauritzen, P., Mishra, S., Neale, R., Taylor, M., & Tribbia, J. (2013). AMIP simulation with the CAM4 spectral element dynamical core. *J. Climate*, *26*(3), 689–709. doi: 10.1175/jcli-d-11-00448.1

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. doi: 10.5194/gmd-9-1937-2016

Farman, J. C., Gardiner, B. G., & Shanklin, J. D. (1985). Large losses of total ozone in antarctica reveal seasonal ClOx/NOx interaction. *Nature*, *315*(6016), 207–210. doi: 10.1038/315207a0

Feng, Y., Wang, H., Rasch, P. J., Zhang, K., Lin, W., Tang, Q., … Yu, H. (2022). Global dust cycle and direct radiative effect in E3SM version 1: Impact of increasing model resolution. *Earth and Space Science Open Archive*, 50. doi: 10.1002/essoar.10510950.1

Gent, P. R., & Mcwilliams, J. C. (1990). Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, *20*(1), 150–155. doi: 10.1175/1520-0485(1990)020⟨0150:IMIOCM⟩2.0.CO;2

Gillett, N. P., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., … Tebaldi, C. (2016). The Detection and Attribution Model Intercomparison Project (DAMIP v1.0) contribution to CMIP6. *Geoscientific Model Development*, *9*(10), 3685–3697. doi: 10.5194/gmd-9-3685-2016

Golaz, J.-C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., … Zhu, Q. (2019). The DOE E3SM Coupled Model Version 1: Overview and Evaluation at Standard Resolution. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2089-2129. doi: 10.1029/2018MS001603

Golaz, J.-C., Larson, V. E., & Cotton, W. R. (2002). A PDF-based model for boundary layer clouds. Part I: Method and model description. *J. Atmos. Sci.*, *59*, 3540–3551. doi: 10.1175/1520-0469(2002)059⟨3540:apbmfb⟩2.0.co;2

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., … Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, *31*(3). doi: 10.1029/2003gl018747

Griffies, S. M., Gnanadesikan, A., Pacanowski, R. C., Larichev, V. D., Dukowicz, J. K., & Smith, R. D. (1998). Isoneutral diffusion in a z-coordinate ocean model. *Journal of Physical Oceanography*, *28*(5), 805–830. doi: 10.1175/1520-0485(1998)028⟨0805:idiazc⟩2.0.co;2

Guba, O., Taylor, M., Ullrich, P., Overfelt, J., & Levy, M. (2014). The spectral element method on variable resolution grids: Evaluating grid sensitivity and resolution-aware numerical viscosity. *Geosci. Model Dev.*, *7*, 4081-4117. doi: 10.5194/gmdd-7-4081-2014

Guerra, J. E., & Ullrich, P. A. (2016). A high-order staggered finite-element vertical discretization for non-hydrostatic atmospheric models. *Geoscientific Model Development*, *9*(5), 2007–2029. doi: 10.5194/gmd-9-2007-2016

Hannah, W. M., Bradley, A. M., Guba, O., Tang, Q., Golaz, J.-C., & Wolfe, W. (2021). Separating physics and dynamics grids for improved computational efficiency in spectral element Earth System Models. *J. Adv. Model Earth Sy.*, *13*(7), e2020MS002419. doi: 10.1029/2020ms002419

Hansen, J. (2005). Efficacy of climate forcings. *Journal of Geophysical Research*, *110*(D18). doi: 10.1029/2005jd005776

Harrop, B. E., Ma, P.-L., Rasch, P. J., Neale, R. B., & Hannay, C. (2018). The role of convective gustiness in reducing seasonal precipitation biases in the tropical west pacific. *Journal of Advances in Modeling Earth Systems*, *10*(4), 961-970. doi: 10.1002/2017MS001157

Hawkins, E., & Sutton, R. (2016). Connecting climate model projections of global temperature change with the real world. *Bulletin of the American Meteorological Society*, *97*(6), 963–980. doi: 10.1175/bams-d-14-00154.1

Hegglin, M., Kinnison, D., Lamarque, J.-F., & Plummer, D. (2016). *CCMI ozone in support of CMIP6 - version 1.0.* Earth System Grid Federation. doi: 10.22033/ESGF/input4MIPs.1115

Henderson, S. A., Maloney, E. D., & Barnes, E. A. (2016). The influence of the madden–julian oscillation on northern hemisphere winter blocking. *J. Climate*, *29*(12), 4597–4616. doi: 10.1175/JCLI-D-15-0502.1

Herrington, A. R., Lauritzen, P. H., Taylor, M. A., Goldhaber, S., Eaton, B. E., Bacmeister, J. T., ... Ullrich, P. A. (2019). Physics–dynamics coupling with element-based high-order Galerkin methods: Quasi-equal-area physics grid. *Mon. Weath. Rev.*, *147*(1), 69–84. doi: 10.1175/mwr-d-18-0136.1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. doi: 10.1002/qj.3803

Hess, M., Koepke, P., & Schult, I. (1998). Optical properties of aerosols and clouds: The software package OPAC. *Bulletin of the American Meteorological Society*, 831-844. doi: 10.1175/1520-0477(1998)079⟨0831:opoaac⟩2.0.co;2

Higgins, R., Schemm, J. E., Shi, W., & Leetmaa, A. (2000). Extreme precipitation events in the western united states related to tropical forcing. *J. Climate*, *13*(4), 793–820. doi: 10.1175/1520-0442(2000)013⟨0793:EPEITW⟩2.0.CO;2

Hill, C., DeLuca, C., Suarez, M., & Da Silva, A. (2004). The architecture of the Earth System Modeling Framework. *Comput. Sci. Eng.*, *6*(1), 18–28. doi: 10.1109/mcise.2004.1255817

Hoch, K. E., Petersen, M. R., Brus, S. R., Engwirda, D., Roberts, A. F., Rosa, K. L., & Wolfram, P. J. (2020). MPAS-Ocean simulation quality for variable-resolution North American coastal meshes. *Journal of Advances in Modeling Earth Systems*, *12*(3). doi: 10.1029/2019ms001848

Hoffman, M. J., Perego, M., Price, S. F., Lipscomb, W. H., Zhang, T., Jacobsen, D., ... Bertagna, L. (2018). MPAS-Albany Land Ice (MALI): a variable-resolution ice sheet model for earth system modeling using voronoi grids. *Geoscientific Model Development*, *11*(9), 3747–3780. doi: 10.5194/gmd-11-3747-2018

Holben, B. N., Eck, T. F., Slutsker, I., Tanre, D., Buis, J. P., Setzer, A., ... Smirnov, A. (1998). AERONET - a federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment*, *66*(1), 1-16. doi: 10.1016/s0034-4257(98)00031-5

Hoose, C., Kristjansson, J. E., Iversen, T., Kirkevag, A., Seland, O., & Gettelman, A. (2009). Constraining cloud droplet number concentration in GCMs suppresses the aerosol indirect effect. *Geophysical Research Letters*, *36*(12). doi: 10.1029/2009GL038568

Hsu, J., & Prather, M. J. (2009). Stratospheric variability and tropospheric ozone. *Journal of Geophysical Research: Atmospheres*, *114*(D6). doi: 10.1029/2008JD010942

Hu, Y., Rodier, S., Xu, K.-m., Sun, W., Huang, J., Lin, B., ... Josset, D. (2010). Occurrence, liquid water content, and fraction of supercooled water clouds from combined caliop/iir/modis measurements. *Journal of Geophysical Research: Atmospheres*, *115*(D4). doi: 10.1029/2009JD012384

Hunke, E. C. (2010). Thickness sensitivities in the CICE sea ice model. *Ocean Mod.*, *34*, 137–149. doi: 10.1016/j.ocemod.2010.05.004

Jiang, X., Waliser, D. E., Xavier, P. K., Petch, J., Klingaman, N. P., Woolnough, S. J., ... Zhu, H. (2015). Vertical structure and physical processes of the madden-julian oscillation: Exploring key model physics in climate simulations. *J. Geophys. Res.: Atmospheres*, *120*(10), 4718–4748. doi: 10.1002/2014jd022375

Jones, P. D., New, M., Parker, D. E., Martin, S., & Rigor, I. G. (1999). Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics*, *37*(2), 173-199. doi: 10.1029/1999rg900002

Kasahara, A. (1974). Various vertical coordinate systems used for numerical weather prediction. *Mon. Weath. Rev.*, *102*, 509–522. doi: 10.1175/1520-0493(1974)102⟨0509:vvcsuf⟩2.0.co;2

Kennedy, D., Swenson, S., Oleson, K. W., Lawrence, D. M., Fisher, R., Lola da Costa, A. C., & Gentine, P. (2019). Implementing plant hydraulics in the community land model, version 5. *Journal of Advances in Modeling Earth Systems*, *11*(2), 485–513. doi: 10.1029/2018ms001500

Kinne, S., O'Donnel, D., Stier, P., Kloster, S., Zhang, K., Schmidt, H., ... Stevens, B. (2013). MAC-v1: A new global aerosol climatology for climate studies. *Journal of Advances in Modeling Earth Systems*, *5*, 704-740. doi: 10.1002/jame.20035

Kok, J. F. (2011). Does the size distribution of mineral dust aerosols depend on the wind speed at emission? *Atmospheric Chemistry and Physics*, *11*(19), 10149–10156. doi: 10.5194/acp-11-10149-2011

Kok, J. F., Ridley, D. A., Zhou, Q., Miller, R. L., Zhao, C., Heald, C. L., ... Haustein, K. (2017). Smaller desert dust cooling effect estimated from analysis of dust size and abundance. *Nature Geoscience*, *10*(4), 274–278. doi: 10.1038/ngeo2912

Laprise, R. (1992). The Euler equations of motion with hydrostatic pressure as an independent variable. *Mon. Weath. Rev.*, *120*(1), 197–207. doi: 10.1175/1520-0493(1992)120⟨0197:teeomw⟩2.0.co;2

Large, W. G., Mcwilliams, J. C., & Doney, S. C. (1994). Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Reviews of Geophysics*, *32*(4), 363–403. doi: 10.1029/94RG01872

Larson, J., Jacob, R., & Ong, E. (2005). The Model Coupling Toolkit: A new Fortran90 toolkit for building multiphysics parallel coupled models. *The International Journal of High Performance Computing Applications*, *19*(3), 277-292. doi: 10.1177/1094342005056115

Larson, V. E. (2017). *CLUBB-SILHS: A parameterization of subgrid variability in the atmosphere.* arXiv. doi: 10.48550/ARXIV.1711.03675

Lauritzen, P. H., Mirin, A. A., Truesdale, J., Raeder, K., Anderson, J. L., Bacmeister, J., & Neale, R. B. (2011). Implementation of new diffusion/filtering operators in the CAM-FV dynamical core. *The International Journal of High Performance*

Computing Applications, *26*(1), 63–73. doi: 10.1177/1094342011410088
Leung, L. R., Bader, D. C., Taylor, M. A., & McCoy, R. B. (2020). An Introduction to the E3SM Special Collection: Goals, Science Drivers, Development, and Analysis. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2019MS001821. doi: 10.1029/2019MS001821
Li, H. Y., Wigmosta, M. S., Wu, H., Huang, M., Ke, Y., Coleman, A. M., & Leung, L. R. (2013). A physically based runoff routing model for land surface and earth system models. *Journal of Hydrometeorology*, *14*, 808-828. doi: 10.1175/JHM-D-12-015.1
Lin, S.-J. (2004). A vertically Lagrangian finite-volume dynamical core for global models. *Mon. Weath. Rev.*, *132*, 2293–2397. doi: 10.1175/1520-0493(2004)132⟨2293:avlfdc⟩2.0.co;2
Liu, Ma, P. L., Wang, H., Tilmes, S., Singh, B., Easter, R. C., . . . Rasch, P. J. (2016). Description and evaluation of a new four-mode version of the Modal Aerosol Module (MAM4) within version 5.3 of the Community Atmosphere Model. *Geoscientific Model Development*, *9*(2), 505-522. doi: 10.5194/gmd-9-505-2016
Liu, W., Huang, B., Thorne, P. W., Banzon, V. F., Zhang, H.-M., Freeman, E., . . . Woodruff, S. D. (2015). Extended reconstructed sea surface temperature version 4 (ERSST.v4): Part II. Parametric and structural uncertainty estimations. *Journal of Climate*, *28*(3), 931–951. doi: 10.1175/JCLI-D-14-00007.1
Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., . . . Kato, S. (2018). Clouds and the earth's radiant energy system (CERES) energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition-4.0 data product. *Journal of Climate*, *31*(2), 895–918. doi: 10.1175/jcli-d-17-0208.1
Luo, X., Li, H. Y., Ruby Leung, L., Tesfa, T. K., Getirana, A., Papa, F., & Hess, L. L. (2017). Modeling surface water dynamics in the Amazon basin using MOSART-Inundation v1.0: Impacts of geomorphological parameters and river flow representation [Journal Article]. *Geoscientific Model Development*, *10*, 1233-1259. doi: 10.5194/gmd-10-1233-2017
Ma, P.-L., Harrop, B. E., Larson, V. E., Neale, R., Gettelman, A., Morrison, H., . . . Leung, L. R. (2021). Better calibration of cloud parameterizations and subgrid effects increases the fidelity of E3SM Atmosphere Model version 1. *Geoscientific Model Development Discussions*, *2021*, 1–57. doi: 10.5194/gmd-2021-298
Ma, P.-L., Harrop, B. E., Larson, V. E., Neale, R. B., Gettelman, A., Morrison, H., . . . Leung, L. R. (2022). Better calibration of cloud parameterizations and subgrid effects increases the fidelity of the E3SM Atmosphere Model version 1. *Geoscientific Model Development*, *15*(7), 2881–2916. doi: 10.5194/gmd-15-2881-2022
Madden, R. A., & Julian, P. R. (1971). Detection of a 40–50 day oscillation in the zonal wind in the tropical pacific. *J. Atmos. Sci.*, *28*(5), 702–708. doi: 10.1175/1520-0469(1971)028⟨0702:doadoi⟩2.0.co;2
Maloney, E. D., & Hartmann, D. L. (2000). Modulation of hurricane activity in the Gulf of Mexico by the Madden-Julian Oscillation. *Nature*, *287*(5460), 2002–2004. doi: 10.1126/science.287.5460.2002
Matsueda, S., & Takaya, Y. (2015). The global influence of the Madden–Julian Oscillation on extreme temperature events. *J. Climate*, *28*(10), 4141–4151. doi: 10.1175/jcli-d-14-00625.1
McFarlane, N. A. (1987). The effect of orographically excited gravity wave drag on the general circulation of the lower stratosphere and troposphere. *Journal of Atmospheric Sciences*, *44*(14), 1775 - 1800. doi: 10.1175/1520-0469(1987)044⟨1775:TEOOEG⟩2.0.CO;2
Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J.-F., Stouffer, R. J., . . . Schlund, M. (2020). Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 earth system models. *Science*

Advances, *6*(26). doi: 10.1126/sciadv.aba1981

Meier, W., Fetterer, F., Savoie, M., Mallory, S., Duerr, R., & Stroeve, J. (2017). NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 3. *Electron. Media*. doi: 10.7265/N59P2ZTG

Mignot, J., Hourdin, F., Deshayes, J., Boucher, O., Gastineau, G., Musat, I., ... Silvy, Y. (2021). The tuning strategy of IPSL-CM6a-LR. *Journal of Advances in Modeling Earth Systems*, *13*(5). doi: 10.1029/2020ms002340

Molina, M., & Rowland, F. (1974). Stratospheric sink for chlorofluoromethanes: chlorine atom-catalysed destruction of ozone. *Nature*, *249*, 810–812. doi: 10.1038/249810a0

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., ... Simpson, I. R. (2021). An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set. *Journal of Geophysical Research: Atmospheres*, *126*(3). doi: 10.1029/2019jd032361

Mundhenk, B. D., Barnes, E. A., & Maloney, E. D. (2016). All-season climatology and variability of atmospheric river frequencies over the North Pacific. *J. Climate*, *29*(13), 4885–4903. doi: 10.1175/JCLI-D-15-0655.1

Neale, R. B., Chen, C.-C., Gettelman, A., Lauritzen, P. H., Park, S., Williamson, D. L., ... Taylor, M. A. (2012). *Description of the NCAR Community Atmosphere Model (CAM 5.0)* (NCAR Technical Note Nos. NCAR/TN-486+STR). National Center of Atmospheric Research. doi: 10.5065/wgtk-4g06

Neelin, J. D., Bracco, A., Luo, H., McWilliams, J. C., & Meyerson, J. E. (2010). Considerations for parameter optimization and sensitivity in climate models. *Proceedings of the National Academy of Sciences*, *107*(50), 21349–21354. doi: 10.1073/pnas.1015473107

Orbe, C., Roekel, L. V., Adames, Á. F., Dezfuli, A., Fasullo, J., Gleckler, P. J., ... Zhao, M. (2020). Representation of modes of variability in six U.S. climate models. *J. Climate*, *33*(17), 7591–7617. doi: 10.1175/jcli-d-19-0956.1

Petersen, M., Asay-Davis, X., Jacobsen, D., Maltrud, M., Ringler, T., Van Roekel, L., & Wolfram, P. (2018). MPAS Ocean User's Guide V6. *Zenodo*. doi: 10.5281/zenodo.1246893

Petersen, M., Asay-Davis, X. S., Berres, A. S., Chen, Q., Feige, N., Hoffman, M. J., ... Woodring, J. L. (2019). An Evaluation of the Ocean and Sea Ice Climate of E3SM Using MPAS and Interannual CORE-II Forcing. *Journal of Advances in Modeling Earth Systems*, *11*(5), 1438–1458. doi: 10.1029/2018MS001373

Petersen, M., Jacobsen, D., Ringler, T. D., Hecht, M. W., & Maltrud, M. E. (2015). Evaluation of the arbitrary Lagrangian–Eulerian vertical coordinate method in the MPAS-Ocean model. *Ocean Modelling*, *86*, 93–113. doi: 10.1016/j.ocemod.2014.12.004

Pincus, R., Forster, P. M., & Stevens, B. (2016). The Radiative Forcing Model Intercomparison Project (RFMIP): experimental protocol for CMIP6. *Geoscientific Model Development*, *9*(9), 3447–3460. doi: 10.5194/gmd-9-3447-2016

Qin, Y. (2022). *qinyia/diag_feedback_E3SM: First release.* Zenodo. doi: 10.5281/zenodo.6354226

Rasch, P. J., Xie, S., Ma, P.-L., Lin, W., Wang, H., Tang, Q., ... Yang, Y. (2019). An overview of the atmospheric component of the Energy Exascale Earth System Model. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2377–2411. doi: 10.1029/2019ms001629

Reckinger, S. M., Petersen, M. R., & Reckinger, S. J. (2015). A study of overflow simulations using MPAS-Ocean: Vertical grids, resolution, and viscosity. *Ocean Modelling*, *96*, 291–313. doi: 10.1016/j.ocemod.2015.09.006

Redelsperger, J.-L., Guichard, F., & Mondon, S. (2000). A parameterization of mesoscale enhancement of surface fluxes for large-scale models. *Journal of Climate*, *13*(2), 402 - 421. doi: 10.1175/1520-0442(2000)013⟨0402:

APOMEO⟩2.0.CO;2

Richter, J. H., Chen, C.-C., Tang, Q., Xie, S., & Rasch, P. J. (2019). Improved simulation of the QBO in E3SMv1. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3403-3418. doi: 10.1029/2019MS001763

Richter, J. H., Sassi, F., & Garcia, R. R. (2010). Toward a physically based gravity wave source parameterization in a general circulation model. *Journal of the Atmospheric Sciences*, *67*(1), 136 - 156. doi: 10.1175/2009JAS3112.1

Ridley, D. A., Heald, C. L., Kok, J. F., & Zhao, C. (2016). An observationally constrained estimate of global dust aerosol optical depth. *Atmospheric Chemistry and Physics*, 15097-15117. doi: 10.5194/acp-16-15097-2016

Ringer, M. A., Andrews, T., & Webb, M. J. (2014). Global-mean radiative feedbacks and forcing in atmosphere-only and coupled atmosphere-ocean climate change experiments. *Geophysical Research Letters*, *41*(11), 4035-4042. doi: 10.1002/2014GL060347

Ringler, T., Ju, L., & Gunzburger, M. (2008). A multiresolution method for climate system modeling: Application of spherical centroidal voronoi tessellations. *Ocean Dynamics*, *58*(5-6), 475–498. doi: 10.1007/s10236-008-0157-2

Ringler, T., Petersen, M., Higdon, R. L., Jacobsen, D., Jones, P. W., & Maltrud, M. (2013). A multi-resolution approach to global ocean modeling. *Ocean Modelling*, *69*, 211–232. doi: 10.1016/j.ocemod.2013.04.010

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., ... Zelinka, M. D. (2020). An assessment of earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, *58*. doi: 10.1029/2019rg000678

Simmons, A. J., & Burridge, D. M. (1981). An energy and angular momentum conserving vertical finite-difference scheme and hybrid vertical coordinates. *Mon. Weath. Rev.*, *109*, 758–766. doi: 10.1175/1520-0493(1981)109⟨0758:aeaamc⟩2.0.co;2

Smith, C. J., Kramer, R. J., Myhre, G., Alterskjær, K., Collins, W., Sima, A., ... Forster, P. M. (2020). Effective radiative forcing and adjustments in cmip6 models. *Atmospheric Chemistry and Physics*, *20*(16), 9591–9618. doi: 10.5194/acp-20-9591-2020

Soden, B. J., Held, I. M., Colman, R., Shell, K. M., Kiehl, J. T., & Shields, C. A. (2008). Quantifying climate feedbacks using radiative kernels. *Journal of Climate*, *21*(14), 3504–3520. doi: 10.1175/2007JCLI2110.1

Tan, Z., Leung, L. R., Li, H.-Y., & Tesfa, T. (2018). Modeling sediment yield in land surface and earth system models: Model comparison, development, and evaluation. *Journal of Advances in Modeling Earth Systems*, *10*(9), 2192–2213. doi: 10.1029/2017ms001270

Tan, Z., Leung, L. R., Li, H.-Y., Tesfa, T., Zhu, Q., & Huang, M. (2020). A substantial role of soil erosion in the land carbon sink and its future changes. *Global Change Biology*, *26*(4), 2642–2655. doi: 10.1111/gcb.14982

Tang, Q., Prather, M. J., Hsu, J., Ruiz, D. J., Cameron-Smith, P. J., Xie, S., & Golaz, J.-C. (2021). Evaluation of the interactive stratospheric ozone (O3v2) module in the E3SM version 1 Earth system model. *Geoscientific Model Development*, *14*(3), 1219–1236. doi: 10.5194/gmd-14-1219-2021

Taylor, M. A., & Fournier, A. (2010). A compatible and conservative spectral element method on unstructured grids. *J. Comput. Phys.*, *229*, 5879– 5895. doi: 10.1016/j.jcp.2010.04.008

Taylor, M. A., Guba, O., Steyer, A., Ullrich, P. A., Hall, D. M., & Eldrid, C. (2020). An energy consistent discretization of the nonhydrostatic equations in primitive variables. *Journal of Advances in Modeling Earth Systems*, *12*(1). doi: 10.1029/2019MS001783

Thuburn, J., Ringler, T. D., Skamarock, W. C., & Klemp, J. B. (2009). Numerical representation of geostrophic modes on arbitrarily structured C-grids. *Journal of Computational Physics*, *228*(22), 8321–8335. doi: 10.1016/j.jcp.2009.08.006

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, *79*(1), 61–78. doi: 10.1175/1520-0477(1998)079⟨0061:apgtwa⟩2.0.co;2

Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G., ... Yamazaki, D. (2018). JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do). *Ocean Modelling*, *130*, 79–139. doi: 10.1016/j.ocemod.2018.07.002

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

Turner, A. K., Lipscomb, W. H., Hunke, E. C., Jacobsen, D. W., Jeffery, N., Engwirda, D., ... Wolfe, J. D. (2021). MPAS-seaice (v1.0.0): Sea-ice dynamics on unstructured Voronoi meshes. *Geoscientific Model Development Discussions*, 1–46. doi: 10.5194/gmd-2021-355

Ullrich, P. A., Devendran, D., & Johansen, H. (2016). Arbitrary-order conservative and consistent remapping and a theory of linear maps: Part II. *Mon. Weather Rev.*, *144*(4), 1529–1549. doi: 10.1175/mwr-d-15-0301.1

Ullrich, P. A., & Taylor, M. A. (2015). Arbitrary-order conservative and consistent remapping and a theory of linear maps: Part I. *Mon. Weather Rev.*, *143*(6), 2419–2440. doi: 10.1175/mwr-d-14-00343.1

Van Roekel, L., Adcroft, A., Danabasoglu, G., Griffies, S. M., Kauffman, B., Large, W., ... Schmidt, M. (2018). The KPP boundary layer scheme for the ocean: Revisiting its formulation and benchmarking one-dimensional simulations relative to LES. *Journal of Advances in Modeling Earth Systems*, *10*, 2647–2685. doi: 10.1029/2018MS001336

Vitart, F., & Robertson, A. W. (2018). The sub-seasonal to seasonal prediction project (s2s) and the prediction of extreme events. *npj Climate and Atmos. Sci.*, *1*(1). doi: 10.1038/s41612-018-0013-0

Voisin, N., Li, H. Y., Ward, D., Huang, M., Wigmosta, M., & Leung, L. R. (2013). On an improved sub-regional water resources management representation for integration into earth system models [Journal Article]. *Hydrology and Earth System Sciences*, *17*, 3605-3622. doi: 10.5194/hess-17-3605-2013

Wang, D., Thorton, P. E., Schwartz, P. D., & Yuan, F. (2020). Early experience in ultra-scale E3SM land model development on SUMMIT. In *Proceedings of 10th international congress on environmental modelling and software.* Retrieved from `https://www.osti.gov/servlets/purl/1822096`

Wang, H., Easter, R. C., Zhang, R., Ma, P.-L., Singh, B., Zhang, K., ... Yoon, J.-H. (2020). Aerosols in the E3SM version 1: New developments and their impacts on radiative forcing. *Journal of Advances in Modeling Earth Systems*, *12*(1). doi: 10.1029/2019MS001851

Wang, Y.-C., Pan, H.-L., & Hsu, H.-H. (2015). Impacts of the triggering function of cumulus parameterization on warm-season diurnal rainfall cycles at the atmospheric radiation measurement southern great plains site. *Journal of Geophysical Research - Atmosphere*, *120*(20), 10681-10702. doi: 10.1002/2015JD023337

Wheeler, M., & Kiladis, G. N. (1999). Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber–frequency domain. *J. Atmos. Sci.*, *56*(3), 374–399. doi: 10.1175/1520-0469(1999)056⟨0374:CCEWAO⟩2.0.CO;2

Wheeler, M., & McBride, J. (2012). Australasian monsoon. In K. M. W. Lau & E. D. Waliser (Eds.), *Intraseasonal variability in the atmosphere-ocean climate system* (pp. 147–197). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-13914-7_14

Wu, M., Liu, X., Yu, H., Wang, H., Shi, Y., Yang, K., ... Ke, Z. (2020). Understanding processes that control dust spatial distributions with global climate models and satellite observations. *Atmospheric Chemistry and Physics*, *20*, 13835-13855. doi: 10.5194/acp-20-13835-2020

Xie, S., Lin, W., Rasch, P. J., Ma, P.-L., Neale, R., Larson, V. E., … Zhang, Y. (2018). Understanding cloud and convective characteristics in version 1 of the E3SM Atmosphere Model. *Journal of Advances in Modeling Earth Systems*, *10*(10), 2618–2644. doi: 10.1029/2018ms001350

Xie, S., Wang, Y.-C., Ma, H.-Y., Tang, Q., Tang, S., Zheng, X., … Zhang, M. H. (2019). Improved diurnal cycle of precipitation in E3SM with a revised convective triggering function. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2290-2310. doi: 10.1029/2019MS001702

Xie, S., & Zhang, M. (2000). Impact of the convection triggering function on single-column model simulations. *Journal of Geophysical Research - Atmosphere*, *105*(D11), 14983-14996. doi: 10.1029/2000JD900170

Zeng, X., & Geil, K. (2016). Global warming projection in the 21st century based on an observational data-driven model. *Geophysical Research Letters*, *43*(20), 10,947-10,954. doi: https://doi.org/10.1002/2016GL071035

Zhang, C., Golaz, J.-C., Forsyth, R., Vo, T., Xie, S., Shaheen, Z., … others (2022). The E3SM Diagnostics package (E3SM Diags v2.6): A python-based diagnostics package for earth system models evaluation. *Geoscientific Model Development Discussions*, 1–35. doi: 10.5194/gmd-2022-38

Zhang, C., & Gottschalck, J. (2002). SST anomalies of ENSO and the madden–julian oscillation in the equatorial pacific. *J. Climate*, *15*(17), 2429–2445. doi: 10.1175/1520-0442(2002)015⟨2429:saoeat⟩2.0.co;2

Zhang, G. J., & McFarlane, N. A. (1995). Sensitivity of climate simulations to the parameterization of cumulus convection in the canadian climate centre general circulation model. *Atmosphere-Ocean*, *33*(3), 407-446. doi: 10.1080/07055900.1995.9649539

Zhang, K., Rasch, P. J., Taylor, M. A., Wan, H., Leung, R., Ma, P.-L., … Xie, S. (2018). Impact of numerical choices on water conservation in the E3SM Atmosphere Model version 1 (EAMv1). *Geoscientific Model Development*, *11*(5), 1971–1988. doi: 10.5194/gmd-11-1971-2018

Zhang, K., Zhang, W., Wan, H., Rasch, P. J., Ghan, S. J., Easter, R. C., … Leung, R. L. (2022). Effective radiative forcing of anthropogenic aerosols in E3SMv1: historical changes, causality, decomposition, and parameterization sensitivities. *Atmospheric Chemistry and Physics Discussions*, *2022*, 1–49. doi: 10.5194/acp-2021-1087

Zhang, S., Zhang, K., Wan, H., & Sun, J. (2022). Further improvement and evaluation of nudging in the e3sm atmosphere model version 1 (eamv1). *Geoscientific Model Development Discussions*, *2022*, 1–37. doi: 10.5194/gmd-2022-10

Zhang, Y., Xie, S., Lin, W., Klein, S. A., Zelinka, M., Ma, P.-L., … Ma, H.-Y. (2019). Evaluation of Clouds in Version 1 of the E3SM Atmosphere Model With Satellite Simulators. *Journal of Advances in Modeling Earth Systems*, *11*(5), 1253-1268. doi: 10.1029/2018MS001562

Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., … Xiang, B. (2018). The GFDL global atmosphere and land model AM4.0/LM4.0: 2. model description, sensitivity studies, and tuning strategies. *Journal of Advances in Modeling Earth Systems*, *10*(3), 735–769. doi: 10.1002/2017ms001209

Zheng, X., Li, Q., Zhou, T., Tang, Q., Van Roekel, L. P., & Golaz, J.-C. (2021). Description of historical and future projection simulations by the global coupled E3SMv1.0 model as used in CMIP6. *Geoscientific Model Development Discussions*, *2021*, 1–40. doi: 10.5194/gmd-2021-312

Zhou, T., Leung, L. R., Leng, G., Voisin, N., Li, H., Craig, A. P., … Mao, Y. (2020). Global irrigation characteristics and effects simulated by fully coupled land surface, river, and water management models in e3sm [Journal Article]. *Journal of Advances in Modeling Earth Systems*, *12*(10). doi: 10.1029/2020ms002069

Ziemke, J. R., Oman, L. D., Strode, S. A., Douglass, A. R., Olsen, M. A., McPeters, R. D., ... Taylor, S. L. (2019). Trends in global tropospheric ozone inferred from a composite record of toms/omi/mls/omps satellite measurements and the merra-2 gmi simulation. *Atmospheric Chemistry and Physics*, *19*(5), 3257–3269. doi: 10.5194/acp-19-3257-2019

Supporting Information for "The DOE E3SM Model Version 2: Overview of the physical model"

Jean-Christophe Golaz[1], Luke P. Van Roekel[2], Xue Zheng[1], Andrew F. Roberts[2], Jonathan D. Wolfe[2], Wuyin Lin[3], Andrew M. Bradley[4], Qi Tang[1], Mathew E. Maltrud[2], Ryan M. Forsyth[1], Chengzhu Zhang[1], Tian Zhou[5], Kai Zhang[5], Charles S. Zender[6], Mingxuan Wu[5], Hailong Wang[5], Adrian K. Turner[2], Balwinder Singh[5], Jadwiga H. Richter[7], Yi Qin[1], Mark R. Petersen[2], Azamat Mametjanov[8], Po-Lun Ma[5], Vincent E. Larson[9,5], Jayesh Krishna[8], Noel D. Keen[10], Nicole Jeffery[2], Elizabeth C. Hunke[2], Walter M. Hannah[1], Oksana Guba[4], Brian M. Griffin[9], Yan Feng[8], Darren Engwirda[2], Alan V. Di Vittorio[10], Cheng Dang[11,12*], LeAnn M. Conlon[2], Chih-Chieh-Jack Chen[7], Michael A. Brunke[13], Gautam Bisht[5], James J. Benedict[2], Xylar S. Asay-Davis[2], Yuying Zhang[1], Xubin Zeng[13], Shaocheng Xie[1], Phillip J. Wolfram[2], Tom Vo[1], Milena Veneziani[2], Teklu K. Tesfa[5], Sarat Sreepathi[14], Andrew G. Salinger[4], Michael J. Prather[11], Salil Mahajan[14], Qing Li[2,15*], Philip W. Jones[2], Robert L. Jacob[8], J. E. Jack Reeves Eyre[13,16*], Gunther W. Huebler[9], Xianglei Huang[17], Benjamin R. Hillman[4], Bryce E. Harrop[5], James G. Foucar[4], Yilin Fang[5], Darin S. Comeau[2], Peter M. Caldwell[1], Tony Bartoletti[1], Karthik Balaguru[5], Mark A. Taylor[4], Renata B. McCoy[1], L. Ruby Leung[5], David C. Bader[1]

[1]Lawrence Livermore National Laboratory, Livermore, CA, USA

[2]Los Alamos National Laboratory, Los Alamos, NM, USA

[3]Brookhaven National Laboratory, Upton, NY, USA

April 20, 2022, 4:52pm

[4]Sandia National Laboratories, Albuquerque, NM, USA

[5]Pacific Northwest National Laboratory, Richland, WA, USA

[6]Departments of Earth System Science and Computer Science, University of California, Irvine, CA, USA

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

[8]Argonne National Laboratory, Lemont, IL, USA

[9]Department of Mathematical Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

[10]Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[11]Department of Earth System Science, University of California, Irvine, CA, USA

[12]Joint Center for Satellite Data Assimilation, Boulder, CO, USA

[13]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA

[14]Oak Ridge National Laboratory, Oak Ridge, TN, USA

[15]The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, Guangdong, China

[16]NOAA NCEP/CPC, College Park, MD, USA

[17]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA

[*]Current affiliation

Contents of this file

1. Figures S3 to S5

Additional Supporting Information (Files uploaded separately)

1. Captions for large Tables S1 to Sx (if larger than 1 page, upload as separate excel file)

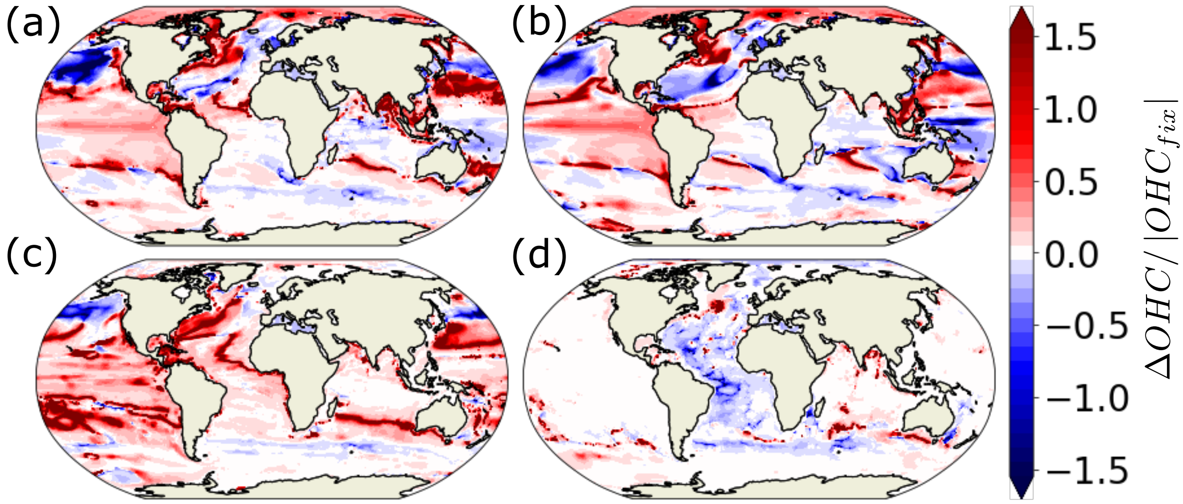


Figure S1. Percent change in ocean heat content anomalies between the simulation with the advection bug and with the bug fixed, i.e., $(OHC_{fix} - OHC_{bug})/|OHC_{fix}|$, (a) Full depth, (b) 0-700m, (c) 700-2000m, and (d) 2000m - Bottom.

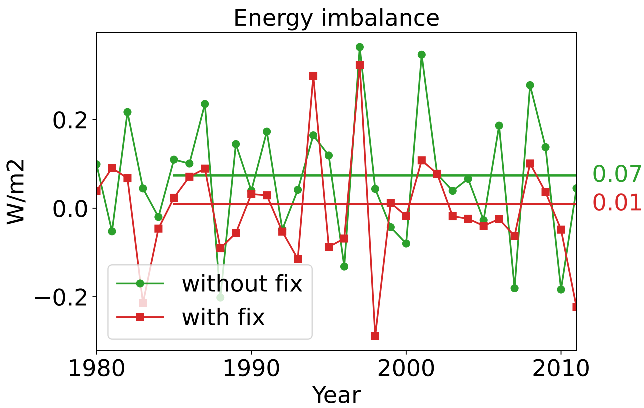


Figure S2. Energy imbalance (diagnosed as the difference between the net fluxes at the top and the surface) for atmosphere simulations with and without energy fix in the gravity wave drag parametrization. Horizontal lines and corresponding values to the right of the plot indicate average values of the imbalance.

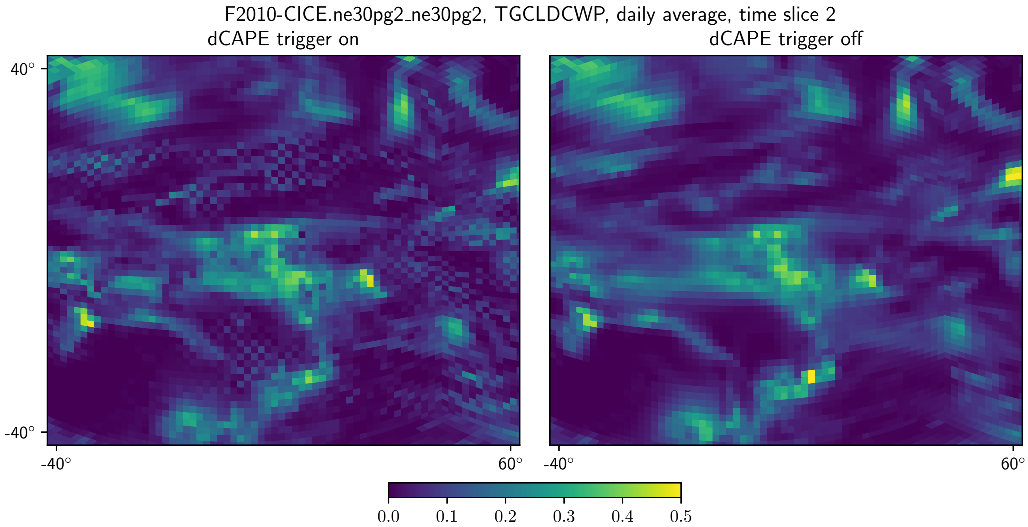


Figure S3. Daily average output of the total grid-box cloud liquid water path (TGCLDLWP) field in the second time slice of two low-resolution atmosphere simulations (F2010-CICE.ne30pg2_ne30pg2) with the dCAPE trigger on and off.

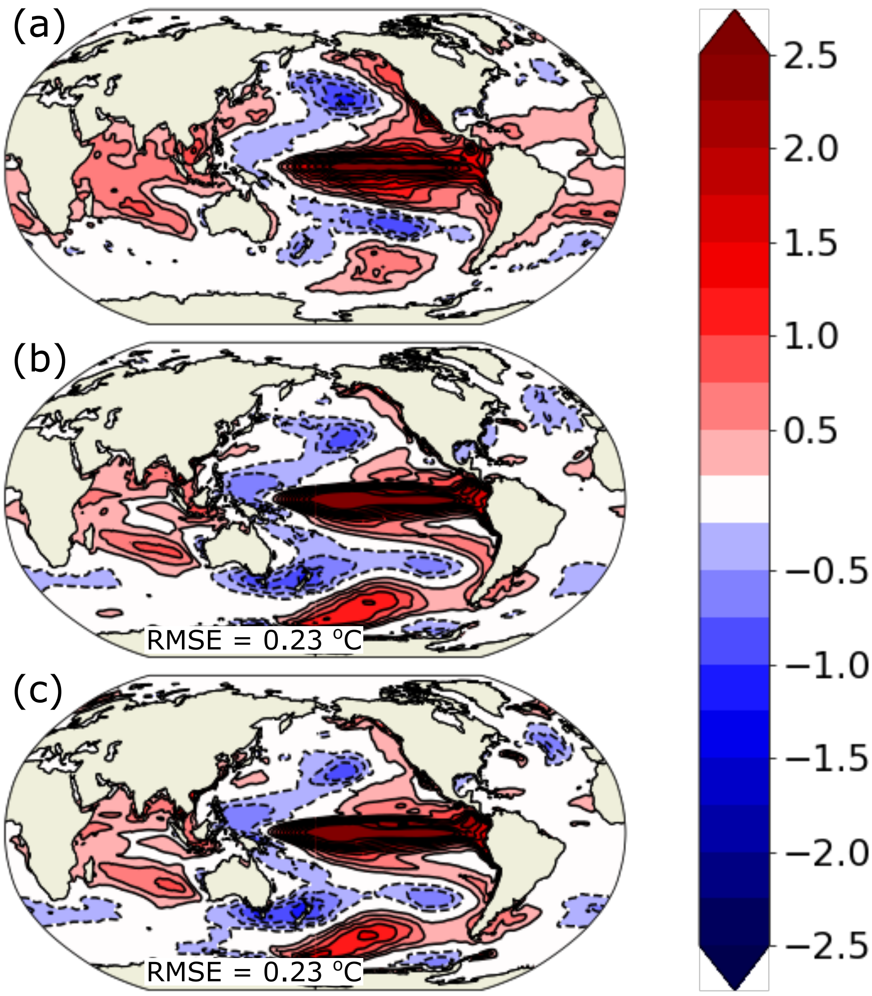


Figure S4. (a–c) Difference of composite El Niño events and composite La Niña events for the HadleyISST data set, the E3SMv2 historical ensemble (1850–2015), and the pre-industrial control, respectively. El Niño events are defined as periods when the Niño 3.4 SST anomaly exceeds 0.8 °C for more than six consecutive months. The La Niña criterion is Niño 3.4 SST anomaly less then -0.8 °C for more than 6 months (these definitions are consistent with Menary et al., 2018). When an El Niño–Southern Oscillation event is identified, the SST is averaged from November to March. For model output, every ensemble member contributes to the mean composite.

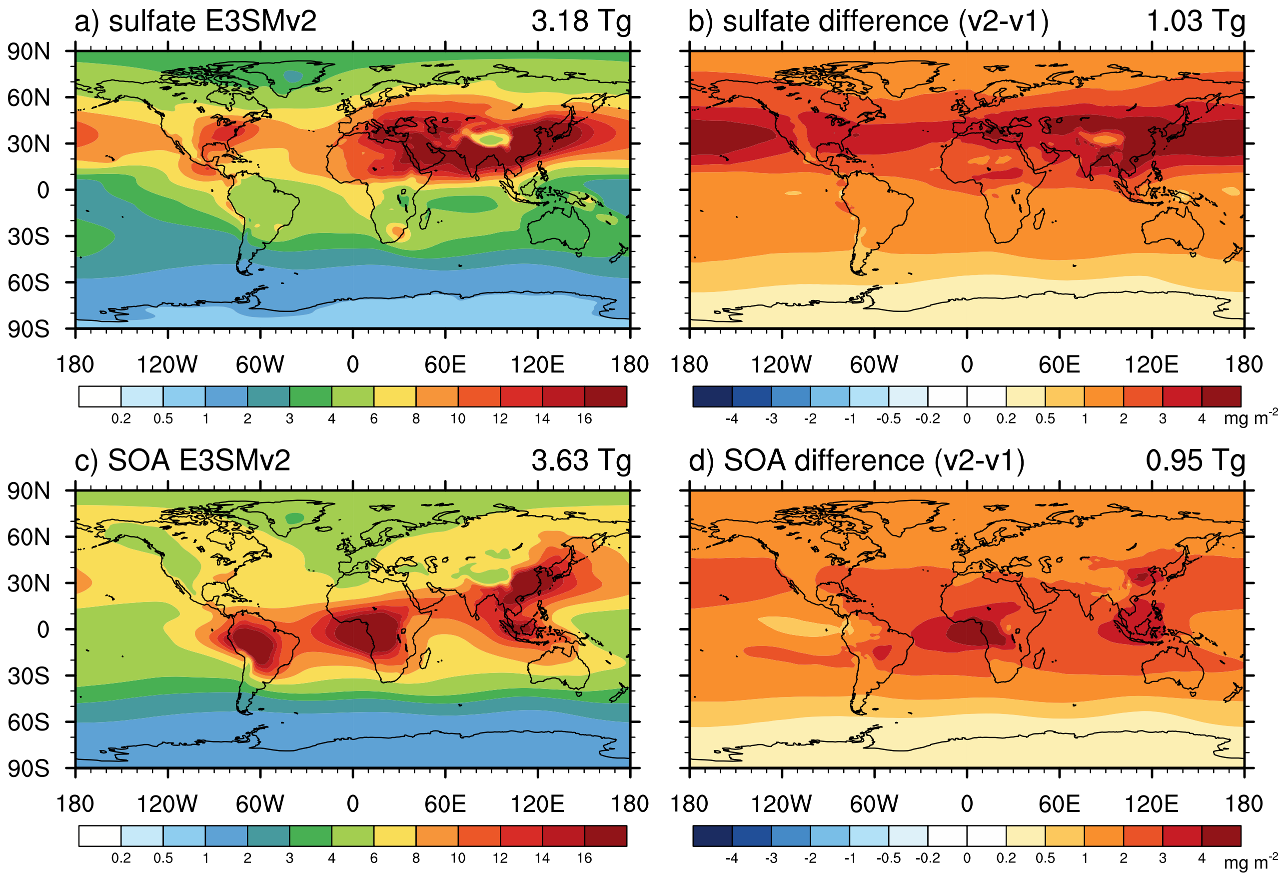


Figure S5. Spatial distributions of global annual mean (2000-2014) (a) sulfate burden from E3SMv2 historical simulations, (b) sulfate burden differences between E3SMv2 and E3SMv1 historical simulations, (c) SOA burden from E3SMv2, and (d) SOA burden differences between E3SMv2 and E3SMv1.

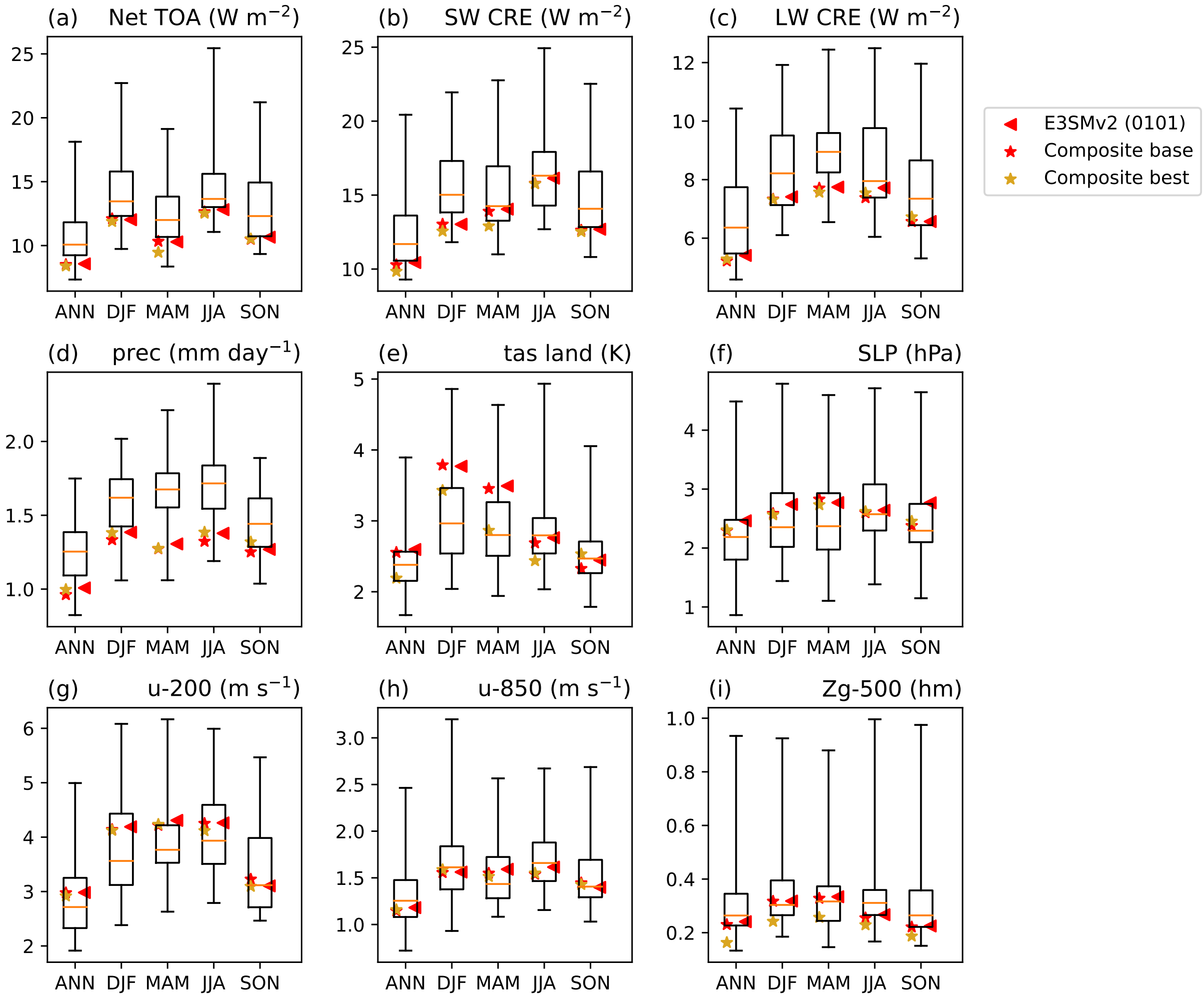


Figure S6. Same as Figure 7 but showing first historical member of E3SMv2 (red triangles) and composite configurations. Red stars (composite base) and gold stars (composite best) refer to hypothetical composite configurations generated by linear combination of single-forcing simulations described in Section 5. Complete data is available in Table S1.

Table S1. Data from Figures 7 and S6 is available in external file '`cmip6.csv`'. Rows correspond to CMIP6 models (first member of historical simulations) or E3SMv2 configurations and column correspond to different fields and seasons. Values are RMSE against relevant observations. Missing values (models for which a specific variable is not available) are indicated by '`--`'. Underlying E3SM Diags comparison figures are available on-line (`https://portal.nersc.gov/project/e3sm/CMIP6_comparison_1985-2014_E3SMv2_golaz_etal_2022/`) . See main text for additional information.