GISS-E2.1: Configurations and Climatology

Maxwell Kelley[1,2], Gavin A. Schmidt[1], Larissa Nazarenko[3,1], Ron L. Miller[1], Susanne E. Bauer[1], Reto Ruedy[2,1], Gary L. Russell[1], Igor Aleinov[3,1], Mike Bauer[3,1], Rainer Bleck[4,5], Vittorio Canuto[1], Grégory Cesana[3,1], Ye Cheng[3,1], Thomas L. Clune[6], Ben Cook[1], Anthony D. Del Genio[1], Gregory S. Elsaesser[7,1], Greg Faluvegi[3,1], Nancy Y. Kiang[1], Daehyun Kim[8], Andrew A. Lacis[1], Anthony Leboissetier[2,1], Allegra N. LeGrande[1], Ken K. Lo[2,1], John C. Marshall[9], Sonali McDermid[10], Elaine E. Matthews[1], Keren Mezuman[3,1], Lee T. Murray[11], Valdar Oinas[2,1], Clara Orbe[1], Carlos Pérez García-Pando[12], Jan P. Perlwitz[13,1], Michael J. Puma[3,1], David Rind[1], Anastasia Romanou[1], Drew T. Shindell[13], Shan Sun[5], Nick Tausnev[2,1], Kostas Tsigaridis[3,1], George Tselioudis[1], Ensheng Weng[3,1], Jingbo Wu[7,1], Mao-Sung Yao[2,1]

[1]NASA Goddard Institute for Space Studies, New York NY
[2]SciSpace LLC, New York NY
[3]Center for Climate Systems Research, Earth Institute, Columbia University, New York NY
[4]CIRES, University of Colorado
[5]NOAA/ESRL/Global Systems Division, Boulder CO
[6]Goddard Space Flight Center, Greenbelt MD
[7]Department of Applied Physics and Applied Mathematics, Columbia University, New York NY
[8]University of Washington, Seattle WA
[9]Massachusetts Institute of Technology, Cambridge MA
[10]Department of Environmental Studies, New York University, New York, NY
[11]Department of Earth and Environmental Sciences, University of Rochester, Rochester NY
[12]Barcelona Supercomputing Center, Barcelona Spain
[13]Climate, Aerosol, and Pollution Research, LLC, Bronx NY
[14]Nicholas School of the Environment, Duke University, Durham NC

Key Points:

- GISS E2.1 is an updated climate model version for use within the CMIP6 project.
- Atmospheric composition is calculated consistently in all model versions.
- Results demonstrate a very significant improvement in skill in a climate model without changes to atmospheric resolution.

Corresponding author: Gavin A. Schmidt, gavin.a.schmidt@nasa.gov

Abstract

This paper describes the GISS-E2.1 contribution to the Coupled Model Intercomparison Project, Phase 6 (CMIP6). This model version differs from the predecessor model (GISS-E2) chiefly due to parameterization improvements to the atmospheric and ocean model components, rather than increases in resolution. Model skill when compared to modern era climatologies is significantly higher than in previous versions. Additionally, updates in how forcings and forcing uncertainty are tackled have a material impact on the results. In particular, there have been specific improvements in representations of modes of variability (such as the Madden-Julian Oscillation and other modes in the Pacific) and significant improvements in the simulation of the climate of the Southern Oceans, including sea ice. The effective climate sensitivity to 2×CO_2 is slightly higher than previously at 2.7–3.1°C (depending on version), and is a result of lower radiative forcing and stronger positive feedbacks.

Plain Language Summary

This paper describes and documents the latest iteration of the NASA GISS climate model which will be used for understanding historical climate change and to make projections for the future. We compare the model output to a wide range of observations over the recent era (1979–2014) and show that there has been a significant increase in how well the model performs compared to the previous version from 2014, though some persistent biases remain. The model has a temperature response to the increase of carbon dioxide that is slightly higher than it used to be, but is well within the range expected from observational and past climate constraints.

1 Introduction

The evaluation and assessment of climate models that are being used for attribution of past change and projections of future change has, for the last two decades, been dominated by the Coupled Model Intercomparison Project (CMIP). This is an internationally organised project run by the community and with almost universal participation from climate modeling groups across the world. The latest iteration (Phase 6) started accepting data in 2018 [*Eyring et al.*, 2016] in anticipation of the upcoming Intergovernmental Panel on Climate Change (IPCC) 6th Assessment Report (AR6) due in April 2021.

Climate modeling at the Goddard Institute for Space Studies (GISS) has a long pedigree dating back to the late 1970s [*Hansen et al.*, 1983, 1997, 2002] and has participated in almost all phases of the CMIP project, notably in CMIP3 and CMIP5 [*Schmidt et al.*, 2006, 2014]. Community experience over the last decade has demonstrated that constrained structural diversity in climate modeling is essential for elucidating important connections between processes and outcomes, and GISS models, with their distinct pedigree, have an important and continuing role to play in providing part of that diversity [*Knutti et al.*, 2013]. However, for that role to be successful, GISS needs to maintain and improve model realism (better process inclusion and higher skill) and continue participation in international and national climate model assessment projects. These projects allow model developers to benefit from the very broad scrutiny of results in these public archives from interested researchers and users across the world.

This paper is a description and an initial assessment of the GISS-E2.1 climate model, the first GISS contribution to CMIP6. This model version was developed as part of a long term strategy to improve model performance as much as possible without a significant jump in computational resources, building from the GISS-E2 models used in CMIP5. This exercise could be seen as the result of a much longer tuning process than is generally undertaken with a new model [*Schmidt et al.*, 2017]. Other GISS contributions to CMIP6 will include more significant leaps (higher vertical and horizontal resolutions, new grids and advection schemes, higher model top, new moist process schemes etc.) and will be reported in *Rind et al.* [2020] and elsewhere. Descriptions of specific results for the composition modeling, historical runs and future projections will also be described elsewhere [*Bauer and Tsigaridis*, 2020; *Miller et al.*, 2020].

1.1 Nomenclature

The series of GISS ModelE versions used in this and previous CMIP iterations, have been GISS-E-R, GISS-E-H (in CMIP3, with the R and H denoting different ocean models [*Schmidt et al.*, 2006; *Hansen et al.*, 2007; *Sun and Bleck*, 2006]) followed by GISS-E2-R and GISS-E2-H in CMIP5 [*Schmidt et al.*, 2014], and GISS-E2.1-G and GISS-E2.1-H (in CMIP6). Other CMIP6 versions will be GISS-E2.2-G/H and GISS-E3-G. Some versions (denoted by -CC) also include an interactive carbon cycle [*Romanou et al.*, 2014]. In CMIP5, there were three formal versions of the models that varied according to the degree of interactivity in atmospheric composition (`physics-version=1,2, or 3`). In CMIP6, `physics-version=2` has been dropped, `physics-version=1` denoted as NINT (for non-interactive) uses offline ozone and aerosol fields from `physics-version=3` the OMA model as described in [*Bauer and Tsigaridis*, 2020] and two new aerosol schemes have been added: TOMAS (denoted by `physics-version=4`) [*Lee and Adams*, 2012] and MATRIX (`physics-version=5`) [*Bauer et al.*, 2008], which will be described elsewhere. Additionally, there is an additional parameter in the CMIP6 database related to forcings. The `f#` number is used to denote different sets of forcing data (either concentrations, emissions or similar). In the initial E2.1 submission two versions have been made available for the historical runs; `f1` and `f2` which have different ozone forcings (see section 2.1.3). Documentation of these conventions in all GISS CMIP6 submissions will be maintained and updated at `https://data.giss.nasa.gov/modelE/cmip6/`.

2 Model code changes

Code changes since GISS-E2-R/H [*Schmidt et al.*, 2014] consist of replacement or structural variation of some parameterizations, updating of input files, bug fixes, and retuning of specific parameters. These changes have been driven by internal and external identification of unsatisfactory performance, desired improvements in physical realism in parameterizations, and updates of observational data sets used either as input or evaluation. This section lays out the drivers of specific changes and the specific changes made. Notably, with the exception of additional layers in the ocean models (8 in E2.1-G, 6 in E2.1-H), no other changes were made to the horizontal or vertical resolution in any component.

The main focus of the developments were to address unrealistic aspects in the CMIP5 simulations, notably poor Southern Ocean SST and sea ice (a common problem in across CMIP5 [*Hyder et al.*, 2018]), excessive ocean mixing, and precipitation pattern biases which were evident in *Schmidt et al.* [2014]. Additionally, through the intense analysis by the wider community of the CMIP5 simulations, additional issues were identified that led to subsequent bug fixes or re-calibrations of the code (for instance the assessment in *Prather et al.* [2017] led to a rexamination of the ozone chemistry, and the authors of *Hezel et al.* [2012] alerted us to an issue with snow cover over sea ice). Lastly, new functionality was required to accommodate more complex emission input data and irrigation effects. The specifics of the changes are outlined in the following sections.

2.1 Atmospheric processes

As stated above, atmospheric resolution is the same as in the CMIP5 model, including the number of layers. However, a change was made to the manner in which terrain-following (sigma) layers in the troposphere transition to constant-pressure layers in the stratosphere. In E2, the transition is abrupt, occurring at 150 hPa. For E2.1, the option was activated to use a smooth transition, centered at 100 hPa with a half-width of approximately 30 hPa. This change removes some artifacts previously seen in the diagnostics but negatively impacts the stratosphere circulation.

2.1.1 Radiative Transfer

The total solar irradiance has been updated based on new satellite calibrations [*Kopp and Lean*, 2011] to have a base value of 1361 W $m^{-}2$ (compared to 1366 W $m^{-}2$ in GISS-E2) though this is not expected to have any impact on the climatology or sensitivity once the models have been retuned for radiative balance [*Rind et al.*, 2014]. Spectral irradiance values have also been updated to the latest estimates [*Coddington et al.*, 2016].

Further calibration of the GISS-E2 radiation framework against line-by-line results led to a few improvements for E2.1. Most notably, non-continuum absorption of shortwave radiation by water vapor was significantly increased, thereby rectifying a problem subsequently highlighted in analyses of the CMIP5 ensemble [*DeAngelis et al.*, 2015]. In the longwave region, a systematic increase of OLR of a few W $m^{-}2$ was the main outcome of optimizations of lookup tables for finer model layering and larger training sets of atmospheric profiles. HITRAN 2012 spectroscopy [*Rothman et al.*, 2013] was also incorporated, with negligible impact. The improvements to clear-sky SW and LW skill relative to E2 and other schemes can be seen in the intercomparison of *Pincus et al.* [2015].

A small but consequential error in the snow masking of vegetation (where a constant snow density was used instead of the computed predicted snow density) was fixed, thereby reducing the area fraction of old, compacted snow and hastening springtime snowmelt.

2.1.2 Clouds, convection and boundary layer

As described in *Kim et al.* [2012], *Del Genio et al.* [2012] and *Del Genio et al.* [2015], modifications to the cumulus parameterization in GISS-E2 led to much greater variability associated with the Madden-Julian Oscillation (MJO) in GISS-E2.1. GISS-E2.1 retains the basic entraining double plume updraft-downdraft framework used in GISS-E2, but with the following changes: (1) The entrainment rate coefficient of the more weakly entraining plume is increased from 0.3 to 0.4, thus increasing the sensitivity of convection to environmental humidity; (2) The partitioning between convective precipitation that descends and has the potential to evaporate in the environment rather than in the downdraft is increased from 0 percent to 50 percent, thus increasing the sensitivity of humidity to convection; (3) downdraft buoyancy, which was determined solely by temperature in GISS-E2, is now based on virtual temperature including condensate loading; (4) A previous limit on the cumulus mass flux that inadvertently resulted in zero entrainment rates at high altitudes in strongly convecting environments was eliminated.

The most impactful E2.1 update to the stratiform cloud parameterization concerns the treatment of glaciation in the mixed-phase temperature range. In E2, glaciation in a given gridcell is a probabilistically timed event after which no supercooled liquid can exist or form until all ice has disappeared and the phase decision can "reset" for a new cloud. Within the single-phase cloud condensate framework inherited from E2, E2.1 attempts to model glaciation in a more continuous manner via a temperature-dependent autoconversion rate of supercooled liquid to ice precipitation. This rate is rapid at the homogeneous freezing temperature of -35°C and decreases linearly toward the warm-cloud autoconversion rate at -5°C. Relative to the new-cloud reset mechanism in E2, this "virtual" mixed-phase representation significantly increases the amount of supercooled water cloud in the Southern Ocean and the Arctic in E2.1. Radiation metrics, rather than liquid water path, were the primary means of evaluating this change.

The regime-specific threshold relative humidity for stratiform cloud formation in E2 was dependent upon moist convective activity, resolved vertical motion, and altitude (near the surface). Convective area also restricted the maximum coverage of stratiform cloud. The

E2.1 code was modified as follows: (1) the coverage restriction is no longer applied above convective cloud top, (2) the dependence on vertical motion was dropped, since its application criterion did not distinguish fronts from other structures, and (3) altitude is taken to be relative to local planetary boundary layer (PBL) height rather than a fixed 850 hPa, better demarcating cloud-topped boundary layers from the free troposphere.

The modifications of the turbulence parameterization within and above the PBL [*Yao and Cheng*, 2012] from GISS-E2 include 1) the non-local vertical transport scheme for virtual potential temperature, specific humidity, and other scalars is updated from the [*Holtslag and Moeng*, 1991] scheme to the more robust *Holtslag and Boville* [1993] scheme; 2) employing the turbulence length scale formulation obtained from the large eddy simulation data by *Nakanishi* [2001]; 3) using the more realistic "Richardson number criterion" rather than the "TKE criterion" to calculate the PBL height, following *Troen and Mahrt* [1986] and *Holtslag and Boville* [1993] and 4) modifying the similarity law near the surface in extreme stability conditions [*Zeng et al.*, 1998]. With the above modifications, the relative humidity and low cloud cover have better vertical structures due to greater transport of water vapor in the PBL. The differences in the diagnosed PBL height between the E2.1 and E2 versions correlate well with the differences in the total cloud distribution over oceans. This newer parameterization leads to improvement in cloud and radiation fields in the extratropics. Tropical low clouds were not specifically targeted, as they require finer layering at low levels and a cloud-enabled PBL scheme.

2.1.3 Composition and chemistry

The basic NINT simulations that are the focus of this paper do not have interactive composition, but the background fields of ozone and aerosol concentrations are derived from simulations of the interactive OMA version of the model, run under AMIP conditions [*Bauer and Tsigaridis*, 2020]. Thus the numerous, relatively minor updates and improvements to the composition modules affected these runs and so are described here for completeness.

All anthropogenic and biomass burning emissions of short-lived species were updated to CMIP6 specifications [*Hoesly et al.*, 2018; *van Marle et al.*, 2017], and are now prescribed annually, rather than by decadal interpolation as in CMIP5. Coding changes include: (1) calculating solar input to photolysis code using higher wavelength resolution; (2) updating the photolysis calculations to use up to 3 sets of temperature-dependent cross sections rather than 2; (3) harmonizing the heterogeneous chemistry reaction rate calculations in the stratosphere to use the identical surface areas as those in the radiation code (typically satellite-derived extinction values); (4) correcting the passing of ozone from chemistry to radiation to use ozone only rather than all odd oxygen (which had led to warm biases in the uppermost stratosphere); (5) updating reaction rate coefficients from the JPL 2000 to the 2011 compendium [*Sander et al.*, 2011]; (6) removing an imposed minimum tracer value which had led to large mixing ratios in high latitude grid boxes at high altitudes where total air masses are small; (7) expanding the representation of reactions including atomic hydrogen (no longer limited to specific pressure ranges); (8) expanding aircraft emissions to include more species; (9) correcting the amount of ozone input in photolysis calculations to use the gridbox top rather than the mid-gridbox value, which led to ozone chemistry biases [*Prather et al.*, 2017]. The harmonization of aerosol surface areas in (3) identified a coding error that led to large underestimates in volcanic aerosol surface areas for chemistry in the stratosphere. The two sets of runs denoted by `f1` and `f2` forcings reflect the impacts of that change.

Overall performance of the composition diagnostics is fairly similar to E2, based on comparison with the trace gas observations made in *Shindell et al.* [2013a]. A detailed analysis suggests that over the US and China, the model is slightly high biased in terms of the simulated tropospheric ozone column relative to Tropospheric Emission Spectrometer (TES) measurements (Fig. 1) and substantially low biased in terms of aerosol optical depth relative to MISR observations [*Seltzer et al.*, 2017].

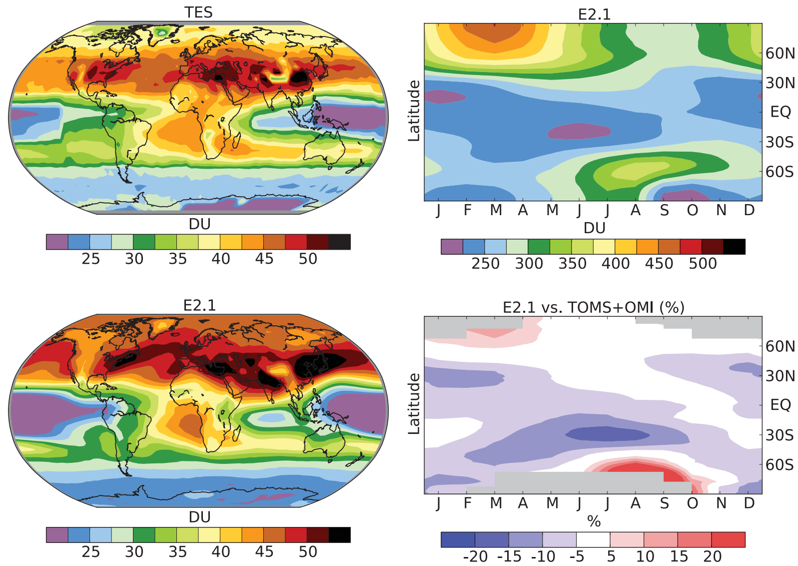


Figure 1. Left column: Annual average 2005–2009 tropospheric column ozone (DU) in TES observations (top) and in E2.1 (bottom). The tropopause is defined using the NCEP 2005–2009 monthly values for TES and the model's internally calculated values for E2.1. Right column: 2000–2010 average of zonal mean, seasonal total column ozone (DU) in E2.1 (top) and the percent difference of this relative to TOMS + OMI observations for the same years (100×(model-obs)/obs) (bottom).

Several updates were made to lightning NO_x production in the chemistry module. The default flash rate parameterization remains a function of convective cloud depth, separately determined over land and sea [*Price and Rind*, 1994]. However, the calculation is now done using altitude above ground level rather than sea level, eliminating spurious lightning over high-altitude regions such as Antarctica. The land and marine flash rate equations are separately tuned to reproduce the respective present-day mean values from the Lightning Imaging Sensor (LIS) and Optical Transient Detector (OTD) satellite climatology [*Cecil et al.*, 2014]. Flash rates are converted to column NO_x production rates using a fixed NO_x-yield per flash assumption. These are then distributed vertically from the surface to the local cloud-top height using the unimodal probability distribution functions of *Ott et al.* [2010] instead of the earlier bimodal distribution of *Pickering et al.* [1998]. The NO_x-yield per flash is determined such the model reproduces the present-day methane chemical lifetime of 9.7 yr [*Prather et al.*, 2012]. This results in 290 mol N per flash, yielding a global mean of 6.4 Tg N yr^{-1}. This is slightly lower than in E2 (7.3 Tg N yr^{-1}) [*Shindell et al.*, 2013b] and falls within the relatively large range of estimates for the present-day lightning NO_x source (2–8 Tg N yr^{-1}) [*Murray*, 2016].

The aerosol module OMA [*Bauer and Tsigaridis*, 2020], simulating sulfate, nitrate, ammonium, carbonaceous aerosols (black carbon and organic carbon, including the NOx-dependent formation of SOA and methanesulfonic acid formation), dust and sea-salt has remained mainly unchanged since CMIP5 (where it was called TCADI), with the exception of (1) increasing in-cloud ammonia dissolution to account for dissociation, thereby remedying the overabundance of nitrate aerosol in E2 [*Nazarenko et al.*, 2017; *Mezuman et al.*, 2016] (2) the parameterized e-folding time of hydrophobic to hydrophilic BC in OMA, a proxy for its aging lifetime, has been tuned in the CMIP6 OMA simulation to match that of MATRIX [*Bauer et al.*, 2008], which does include physically-based aging calculations as part of

Pressure level (hPa)	Avg. diff. AMIP	Avg. diff. coupled	Avg. bias AMIP	Avg. bias coupled	Std. dev. of observations
125	43.5	65.8	9.9	45.1	92.9
200	21.7	27.3	1.2	7.7	52.2
300	13.4	15.2	7.0	8.3	25.6
500	9.6	10.9	6.2	7.5	11.7
900	8.0	8.8	3.6	4.5	8.9

Table 1. Ozone differences and biases (ppbv) between model E2.1-G versions and sonde climatologies. Sonde data primarily from the 1990s and early 2000s; model from early 2000s.

the aerosol microphysics. The new aging timescale for OMA has been evaluated using ice cores and HIPPO flight campaign data in [*Bauer et al.*, 2013]. And (3) updates to the dust representation as discussed below. An evaluation of the aerosol scheme using CMIP6 forcings against satellite, surface network and ice core data is discussed in [*Bauer and Tsigaridis*, 2020].

We updated the heterogeneous chemistry calculations for the formation of nitrate and sulfate coatings on the surface of soil dust particles by uptake of nitric acid and sulfur dioxide, respectively, which were originally described by *Bauer et al.* [2004] and *Bauer and Koch* [2005]. Dust properties are now retrieved from the dust module, instead of being defined separately in the heterogeneous chemistry module, to make those properties consistent with the rest of the model. This concerns the boundaries of the six dust bins (0.1–0.2, 0.2–0.5, 0.5–1, 1–2, 2–4, and 4–8 μm particle diameter), for which coatings on dust particles are calculated in the model, the dust particle densities, and the weights that are used to partition the dust mass of the total clay size range (0.1–2 μm), which is advected as a bulk in the model, among the four clay bins. The weights reflect the size distribution of dust, compared to the previous version where inadvertently only the largest clay bin was considered. An erroneous calculation of the dust number concentration, which led to an overestimate of the number was also corrected. The net effect of the changes are reduced masses of sulfate and nitrate coating on dust by one order of magnitude due to lower uptake of the precursor gases sulfur dioxide and nitric acid, respectively. The global masses of latter in the atmosphere are larger by about 6% and 9%, respectively, with significantly larger increases over North Africa, Middle East, and Central Asia, where dust concentration is elevated. In turn, particulate nitrate aerosols is up to five times higher over equatorial Africa and India and sulfate aerosol is up to 50% higher in the northern hemisphere.

The default dust aerosol tracers in the OMA-version follow the approach of [*Cakmur et al.*, 2006], with the difference that the emitted relative silt and clay fractions of total dust and the emitted total dust mass are optimized in two successive steps, instead of simultaneously. The two-step approach reduces the emitted relative fraction of clay-dust mass (now about 8% of all dust mass over the size range 0.1–32 μm for OMA), which makes the model better conforming with recently published research on the global size distribution of dust in the atmosphere [*Kok et al.*, 2017].

Ozone distributions used in the NINT models are generally similar to those in prior versions. Changes to chemistry have resulted in modest improvements to comparisons with observational data in the troposphere (Table 1). For example, the average bias near the surface (900 hPa) has been reduced from 6.6 (22%) in E2 [*Shindell et al.*, 2013a] to 3.6 (12%) in E2.1. This will help make for more realistic impact studies, which in the past required substantial bias adjustments [*Shindell et al.*, 2018]. The modeled ozone hole in this configuration is biased as the Brewer-Dobson circulation to high-latitudes is too strong in winter, leading to ozone and temperature overestimates during that season. This creates large positive biases in the lowermost stratosphere and upper troposphere from June through August over 60-90S and smaller, but again positive, biases from December through April over 60-90N.

Comparison of the tropospheric column ozone with observations from the Tropospheric Emission Spectrometer (TES) show that the model captures many features of ozone's distribution (Fig. 1). The effect of the wintertime positive biases in the lower stratosphere are clearly visible in model overestimates of tropospheric columns poleward of 50°N and 70°S. Such comparisons are highly sensitive to the tropopause definition [*Shindell et al.*, 2013a], which is in turn sensitive to stratospheric temperature biases and so typically any widespread ozone biases seen here reflect only small differences in the altitude of the tropopause relative to observations. The model captures the maximum over the Atlantic off the west coast of Africa and the minima over the equatorial Pacific and Indian Oceans. As in E2, the minima over the eastern tropical Pacific is too low, however, and this is likely to again dominate biases in long wave radiative fluxes due to ozone [*Bowman et al.*, 2013]. The distribution of column ozone is well represented over most of the NH mid-latitudes, though the magnitude is roughly 2-4 DU too large. The global area-weighted average column in the model is 35.4 DU, very similar to the 35.9 DU from the TES observations. Spatial correlations are broadly similar to those in E2, with an R^2 correlation against TES of 0.86 (compared to 0.85 in E2) and a value of 0.68 against the tropospheric column estimate obtained from OMI minus MLS observations (compared to 0.71 in E2).

2.1.4 Gravity wave drag

E2.1 includes orographic and frontal sources of parameterized gravity waves as in E2. Systematic re-optimization of the scheme was not performed, but two updates required re-calibration of tuning factors: (1) saturation momentum flux was reduced by a factor of approximately 2 as a result of correcting its definition (2) the metric for the presence of fronts (deformation at 700 hPa) was corrected, increasing its magnitude. The orographic wave coefficient was thus reduced and the threshold deformation magnitude for generation of frontal waves was increased.

2.2 Ocean processes

We used two ocean model versions with E2.1 which are denoted E2.1-G (coupling to the GISS Ocean v1 (GO1)), and E2.1-H (coupling to HYCOM). This section describes the updates in each since CMIP5.

2.2.1 GISS Ocean v1

For gross ocean structure and transport metrics, the most impactful updates to E2.1-G are in the parameterizations of mesoscale eddies and vertical mixing. In addition, a high-order advection scheme [*Prather*, 1986] and finer upper-ocean layering (an increase from 32 total layers to 40) sharpened the representation of frontal and thermocline structures in regions of weak parameterized mixing. The updates outlined here will be described more completely elsewhere, as part of parameter sensitivity studies.

A fundamental update to mesoscale eddy transport was the correction of an error in the definition of neutral surfaces in E2-R which drastically reduced the restratification effect. Through the lens of ocean-only simulations and inter-model comparisons of temperature/salinity drifts and circulation metrics such as AMOC and ACC strength, subsequent work explored the consequences of controlled variations in the magnitude and structure of the mesoscale eddy diffusivity [*Marshall et al.*, 2017; *Romanou et al.*, 2017]. Those efforts informed the creation of a moderate-complexity 3D mesoscale diffusivity for E2.1-G whose primary differences from the E2-R scheme are: (1) surface-intensified eddies, in the form of an exponential decay of diffusivity with depth, where the location-dependent decay scale is equal to $[|\rho_h z|]/[|\rho_h|]$, [] denotes vertical averaging, and ρ_h is the horizontal gradient of potential density; (2) replacement of Rossby radius by a geographically constant nominal length scale $L = 39$ km in the $L^2 T_{eady}^{-1}$ baroclinicity scaling of diffusivity retained from E2-R (J. Marshall, pers. comm.); (3) qualitative representation of the Coriolis element in the discarded Rossby radius by a factor $1/max(.05, sin(|latitude|))$ multiplying the diffusivity. The location dependence in (1) permits eddies to restratify the Southern Ocean over a large depth range, consistent with observed density structure there, while not overacting in other regions of the World Ocean (such as the North Atlantic, where the aforementioned ocean-only experiments indicated that deep mesoscale effects can suppress the AMOC). Simplifications (2) and (3) preserve the large-scale structure of the diffusivity distribution and its interactivity while eliminating unconstrained small-scale structure. E2.1-G also adopts a new representation of mesoscale transport expressed in local quasi-isopycnal layering, circumventing some of the difficulties associated with the skew-flux representation that was employed in E2-R.

The E2.1-G vertical diffusivity now includes a contribution from tidal dissipation. AMOC sensitivity to this effect is exploited as a (model-specific) constraint on the considerable uncertainties surrounding this process. Exploratory coupled simulations, lacking the stabilizing effects of relaxation toward climatological surface salinity and a prescribed atmospheric state, systematically developed a runaway haline stratification at high northern latitudes that was the proximate cause of a weak AMOC and excessive northern hemisphere sea ice. The sole parameterization change in any atmosphere or ocean component found able to sustain a strong AMOC was tidally driven mixing, which occurs in the shallow waters bordering the North Atlantic using the dissipation distribution generated by *Jayne* [2009].

Ventilation of marginal seas through their connecting straits has been increased via two mechanisms in E2.1-G, reducing salinity biases there. For straits deep enough that density contrasts can drive strong opposing flows at the surface and depth, the finer upper-ocean layering in E2.1-G resolves this structure, in conjunction with a slight tuning of strait depths. Secondly, horizontal diffusivity was increased in straits that are shallow or have weaker density contrasts. The first mechanism impacted the Red and Black seas, and the second the Baltic and Hudson. The first is the sole ventilation mechanism for straits narrower than the nominal resolution, which are parameterized using the *Russell et al.* [1995] 1-dimensional channel scheme that lacks horizontal mixing.

2.2.2 HYCOM

HYCOM is a hybrid-isopycnal ocean model that was used with previous coupled ModelE versions [*Sun and Bleck*, 2006; *Romanou et al.*, 2013]. Compared to the version used in E2, for E2.1-H, the number of vertical layers was increased to 32 from 26, and the refined equatorial mesh was removed (since it no longer provided a demonstrable increase in skill). HYCOM has traditionally used σ_2 as its vertical coordinate: potential density referenced to a pressure nominally corresponding to 2km depth. At pressures far from this reference, stable in-situ stratification may be misdiagnosed as unstable according to potential density, impacting the layering scheme and vertical mixing. To ensure a monotonic potential density profile in the upper ocean under conditions of stable in-situ stratification there, E2.1-H employs σ_1 (potential density referenced to 1km). This change eliminated spurious deep convection in the Southern Ocean which inhibited formation of the summer halocline and limited sea ice extent. The resulting degradation of the abyssal diagnosis of stratification was found to be benign.

The virtual salt flux formulation of surface freshwater fluxes, employed by HYCOM for consistency with its barotropic/baroclinic mode-splitting scheme, was corrected to conserve global salt, thereby eliminating a net source that resulted in significant positive biases in E2-H salinity. Other fixes to ocean-atmosphere-ice flux coupling include (1) interpolation between grids, (2) elimination of slight inaccuracies in the sea ice mass and heat fluxes, and (3) a modification to the land topography along the coastline to reduce flux biases in atmospheric grid-boxes with average land heights significantly above sea level.

2.3 The cryosphere

Common to both ocean models as in E2, the sea ice component of E2.1 retains the overall framework of E2, excepting the treatment of salt as a material constituent. Algorithmic changes within the framework made the most direct contributions to differences with E2 climatology, and include (1) correction of an inadvertent snow-to-ice transformation during vertical regridding, thereby increasing snow thickness and surface albedo, (2) removal of a 10% floor on lead fraction for conditions typical of the Antarctic winter, (3) closure of leads for thick-ice conditions typical in the Arctic, thereby reducing wintertime heat flux and ice growth there, (4) independent horizontal advection of snow mass. Thermodynamics now follows the "Brine Pocket" (BP) parameterization [*Bitz and Lipscomb*, 1999; *Schmidt et al.*, 2004], and thus salt plays a more active role in E2.1 sea ice, affecting its specific heat and melt rates. Processes relevant to the salt budget (e.g. gravity drainage and flushing of meltwater) are consistently processed with the BP physics. The switch from the previous 'Saline Ice' thermodynamics in E2 to the BP one in E2.1 lead to a slight increase in multiyear sea ice thickness and of sea ice area in the Arctic, a alight reduction of the Antarctic sea ice area as well as a more physically realistic vertical profile of the salinity in the ice.

2.4 Land surface processes

2.4.1 Irrigation and Groundwater

While transient historical changes to irrigation was implemented as a forcing in E2 [*Cook et al.*, 2014; *Krakauer et al.*, 2016], it was not included in the standard CMIP5 submissions. In E2.1, irrigation is now a standard component. Seasonal irrigation amounts from *Wisser et al.* [2010] are applied to the grid box land surface. The water is drawn first from the local surface water system (including rivers and lakes), and if that is insufficient, it is assumed to be drawn from an external groundwater source (which is tracked diagnostically). Groundwater is assumed to have the same temperature as the soil, and has default tracer values. Groundwater recharge is not accounted for, and so there is a small increase in total water mass (and eventually, sea level) associated with the net global groundwater draw in these simulations. These affects have a complex impact on freshwater delivery to the oceans (and hence sea level). Irrigation from local surface water sources, leads to increased soil moisture and reduced river outflow, while net removals of groundwater add freshwater to the climate system, to the tune of about 0.2 mm yr^{-1} of global sea level equivalent in 2010.

2.4.2 Vegetation

As in E2, all vegetation properties affecting physical climate, with the exception of canopy conductance, are prescribed in the simulations described here, whose primary update was the incorporation of satellite-derived distributions of vegetation characteristics, as described below. Like E2, E2.1 sees vegetation properties via the Ent Terrestrial Biosphere Model (Ent TBM), a demographic dynamic global vegetation model (DGVM) whose functionalities are gradually being coupled to ModelE [*Kiang*, 2012; *Kim et al.*, 2015; *Kiang et al.*, in prep.], including carbon cycle interactivity (Ito et al., in prep.). Prescribed interannual variation of vegetation is limited to land use and land cover change (LULCC), by which historical crop and pasture cover is used to rescale the natural vegetation cover fractions in a grid cell. CMIP6 protocol implementations are described more fully in *Miller et al.* [2020] and Ito et al. (in prep.).

The Ent Global Vegetation Structure Data set (Ent GVSD) [*Kiang et al.*, in prep.] has been created as part of ongoing Ent TBM development and is now used for E2.1, replacing E2 prescriptions of vegetation cover, type, height, and LAI based on *Matthews* [1983]. Ent GVSD satellite data sources include land cover types and monthly varying LAI from the Moderate Resolution Imaging Spectroradiometer (MODIS) [*Gao et al.*, 2008; *Myneni et al.*, 2002; *Tian et al.*, 2002a,b; *Yang et al.*, 2006], and tree heights from *Simard et al.* [2011], who utilized 2005 data from the Geoscience Laser Altimeter System (GLAS) aboard the ICESat (Ice, Cloud, and land Elevation Satellite). Specific leaf area (mass carbon per leaf area) data from the TRY database of leaf traits [*Kattge et al.*, 2011] was classified for the Ent TBM 13 plant functional types (PFTs). These observed spatial distributions and leaf trait parameters together allow an estimate of vegetation demography (density of plants per area), and achieve equilibrium behavior in plant-atmosphere carbon exchange and internal plant carbon balances for late 20th C. to early 21st C. climate. The water stress algorithm, which controls the availability of soil water for transpiration, was replaced in E2.1 with a more commonly-used soil water deficit-based one [*Porporato et al.*, 2001; *Rodriguez-Iturbe*, 2000], with the goal of improving transpiration, by distinguishing soil moisture levels at which onset of water stress happens for different plant functional types.

The overall effect of these updates upon surface albedo was significant in some regions, though the overall impact upon physical climate modest compared to other components. Ent PFTs are mapped to the E2 vegetation types for radiative purposes in E2.1; reclassification of cover types directly increased the surface albedo of Australia and eastern South America by several percent. High northern latitudes became brighter via increased snow masking, though this effect was compensated by the correction to snow masking described in the radiation section. Canopy conductances generally decreased using the new LAIs.

3 Simulation design and configurations

The GISS models are designed so that any experiment can be run with an appropriate level of interactivity and complexity - some experiments require the aerosol and chemistry fields to respond to and influence the surface climate, while other simulations focus on one-way impacts. In earlier iterations, NINT historical simulations relied on calculated concentrations of aerosols and tropospheric ozone from a prior generation of models. For instance, the NINT simulations in CMIP5 (using GISS-E2-R or GISS-E2-H) used fields from *Koch et al.* [2011] which were calculated using the CMIP3 model (GISS-E). In CMIP3, the aerosol and ozone fields were from the SI2000 version of the model [*Koch*, 2001; *Koch et al.*, 1999] and thus were not strictly consistent with the composition changes generated in the same-generation interactive models (OMA or MATRIX aerosol microphysical versions) or the specified emission paths. Additionally, many key interactions present in the (computationally expensive) interactive runs (such as ozone responses to volcanoes or solar activity changes) were not represented in the CMIP5 NINT runs.

For CMIP6 we have striven for an increased coherence between forcings and model physics. Namely, we have generated all the historical composition fields using an ensemble of AMIP runs (1860–2014) with the updated interactive OMA version and annually-resolved CMIP6 emissions [*Bauer and Tsigaridis*, 2020]. The time needed to generate new composition fields slows down production, but the resulting NINT simulations have more fidelity to the real world and reflect more processes, while being 3–4 times faster to run when compared to interactive composition versions.

3.1 Pre-industrial boundary conditions

There are a few notable changes from CMIP5 for "pre-industrial" (PI) conditions, which is a slight misnomer, since conditions around 1850 cannot be considered to be unaffected by industrialization, agriculture and fossil fuel use (through the background greenhouse gas levels) and explicit background levels of land use and land cover change, including irrigation [*Hawkins et al.*, 2017]. We now include a background level of irrigation along with background levels of land use/land cover (LULC) alterations and anthropogenic aerosols. The emissions from biomass burning are taken from the standard CMIP6 specifications, but include an (uncertain) anthropogenic component. The spin-up under PI conditions is always greater than 500 years, but the procedure does not include pre-1850 transient changes that might be expected to still have been responsible for ocean heat content anomalies at that time

[*Stenchikov et al.*, 2009; *Gregory*, 2010]. Nonetheless, the difference in sub-surface ocean conditions from reality in 1850 are significantly larger than the impact of prior transient volcanic effects (compared to a suitable averaged background level). Experience from simulations of the last millennium in CMIP5 suggests that the differences in 20th Century transient climate resulting from this choice are minimal.

3.2 Historical Transients

As mentioned above, radiatively active atmospheric composition (ozone and aerosols) is taken from AMIP experiments using E2.1 (OMA) using CMIP6-prescribed annual emissions of aerosols, their precursors and other short-lived reactive chemical species. Well-mixed greenhouse gases, solar activity changes (affecting TSI and the spectral irradiance), LULC (including irrigation), were specified using a mix of approaches [*Miller et al.*, 2020]. Volcanic aerosols were prescribed using pre-computed aerosol depth and effective particle radius [*Thomason et al.*, 2018], though we will also be using interactive emission-driven volcanic effects in some future CMIP6 simulations [*LeGrande et al.*, 2016].

It is important to note that there is substantial uncertainty in some of these drivers over time, especially in the aerosols, solar activity, and early big volcanic eruptions. We therefore plan to incorporate this uncertainty in the CMIP6 historical simulations. Different realizations of the forcings suite are denoted by the `f` number in the `ripf` designation of each individual run in the CMIP6 archive. The first submissions with E2.1-G/E2.1-H have two variations (`f1` and `f2`) which differ due to a correction of the stratospheric ozone chemistry code in the OMA simulations (see section 2.1.3). The `f2` simulations have an improved stratospheric ozone response to volcanoes and a better representation of the Antarctic ozone hole, but are not detectably different in tropospheric climate or response.

4 Coupled Model Tuning

Model tuning for E2.1 followed the procedure described in *Schmidt et al.* [2017]. In the atmosphere-only simulation using the `f1` 1850 pre-industrial conditions, the parameters in the cloud schemes that control the threshold relative humidity and the critical ice mass for condensate conversion are used to achieve global radiative balance and a reasonable global mean albedo. Tuning of convective entrainment was also used to enhance MJO variability [*Del Genio et al.*, 2015].

Upon coupling the ocean and atmosphere models, there is an initial drift to a quasi-stable equilibrium which is judged on overall terms for realism, including the overall skill in the climatological metrics for zonal mean temperature, surface temperatures, sea level pressure, short and long wave radiation fluxes, precipitation, lower stratospheric water vapor, and seasonal sea ice extent. For the configuration to be acceptable, drifts have to be relatively small and quasi-stable behavior of the North Atlantic meridional circulation and other ocean metrics, including the Antarctic Circumpolar Current, are required. ENSO-related metrics are also monitored, but they were not specifically tuned for, since the underlying tropical Pacific SST climatology was not considered to be a feasible tuning target using E2.1 vertical resolution, cloud, and boundary layer schemes. In practice, longer spin-up integrations help reduce drift, and the model state once stabilized can be assessed for suitability. Large drifts at the start of an integration have often been reduced by different tuning choices that either affect surface atmospheric fluxes or (more usually) ocean mixing.

Composition tuning (as described below) is done with pre-industrial and present-day fully interactive simulations (including chemistry and aerosols and indirect effects) and the non-interactive versions use the composition derived from those simulations and the same tuning. Specifically, in the NINT transient simulations, the AIE was preset to have a value of -1 Wm^{-2} in 2000 as it was in the CMIP5 simulations [*Miller et al.*, 2014]. Configurations with interactive aerosols have free latitude to produce whatever forcing is calculated. Additionally, in previous model versions we used a temperature threshold for the formation of polar stratospheric clouds (and hence the heterogeneous chemistry associated with them) [*Shindell et al.*, 2013a] which was tuned so that the polar ozone hole timing is correct despite potential biases in polar vortex temperatures. However, this was not used in E2.1. This model does, however, maintain prior practice of tuning photolysis rates at short wavelengths (<200 nm) for N_2O and O_2 that corrects for problems in stratospheric circulation that otherwise lead to biases in high latitude concentrations of NO_x and O_3. Also, with respect to dust aerosols, emissions are tuned so that the model can match retrieved aerosol optical depths for the present-day [*Miller et al.*, 2006], similarly tuning of the lightning parameterization (and associated source for NO_x) is done against modern observations of flash rate, and tropospheric ozone amounts.

Note that the atmospheric component was tuned using the pre-industrial `f1` background ozone and aerosols. Upon switching to the `f2` background, there was a slight drift in the coupled model. Prior to any historical runs with the `f2` forcings, the coupled model was run a further 100 years to reach a new quasi-equilibrium.

We do not fine tune for an exact global mean surface temperature, since that is effectively precluded by the long spin-up times and limited resources available. Similarly, no tuning was done for climate sensitivity or for performance in a simulation with transient forcing or hindcasts.

5 Climatology 1979–2014

As was seen in the results shown in *Schmidt et al.* [2014], the impact of interactivity in the aerosol or chemistry parts of the model have limited impacts on the climatologies. In addition, while in E2, there was a substantive difference in the composition fields between NINT and TCADI simulations, that is no longer the case in E2.1, though composition-related interactivity may have an greater impact on the variability. We therefore only show the ensemble mean climatology from the standard NINT simulations (10 members for E2.1-G, 5 members for E2.1-H), in both spatial patterns, zonal and global means compared to updated observed climatologies for the satellite period (or as close as possible). All diagnostics are from the `f2` historical simulations unless otherwise stated. We include the zonal mean diagnostics from the E2.1-G `f1` forcings ensemble for completeness, but the differences are small. Note that the map projection uses Equal Earth [*Šavrič et al.*, 2018] and that we now plot zonal means with an area weighted x-axis to minimise visual distortion.

5.1 Global mean diagnostics

Table 2 summarizes a standard set of global mean diagnostics for the NINT versions of the GISS-E2.1 models (with `f2` forcings) and a comparison with updated observations and previous model versions [*Schmidt et al.*, 2014]. Notable improvements are in the global mean temperature, precipitation, and sensible heat fluxes. The net radiative imbalance over this period is also in better comparison with updated estimates from NODC . There are notable biases in total column water vapour (7% too high), and LW cloud forcing (some 20 to 25% too low, though still better than previously). Lower stratospheric water vapour is deficient, consistent with a too cold tropopause. The TOA radiative fluxes are tuned for in pre-industrial atmosphere-only simulations and are therefore not truly predictive. Differences between the coupled models with different ocean modules are small compared to differences with the observations at the global mean level.

Field	E2.1-G	E2.1-H	E2-R	E2-H	Observations
Surface air temp. (°C)	14.1	14.5	14.9	15.6	14.3±0.5[J]
Planetary Albedo	30.4	30.2	29.9	29.7	30[E]/29.4[SEA]
Cloud cover (%)	59.9	59.8	62	62	68[SRK]
Precip. (mm day^{-1})	2.97	2.98	3.17	3.21	2.9[G]
Snowfall (mm day^{-1})	0.24	0.23	0.19	0.17	0.18[L08]/0.12[SEA]
Atmos. water (mm)	26.7	26.8	23.8	24.0	24.9[O]
Energy fluxes (W m^{-2}):					
TOA Absorbed SW	236.9	237.5	239.5	240.3	240.2[SEA]/239.4[T]
TOA Outgoing LW	236.5	237.1	238.8	239.5	239.7[SEA]/238.5[T]
Surf. Abs. SW	161.5	161.9	169.5	170.1	165[SEA]/169[T]
Surf. Down. LW	345.8	347.4	341	344	345.6[SEA]/343[T]
Surf. Net LW (up)	50.5	50.7	56.9	56.9	52.4[SEA]/57[T]
Sensible heat flux	23.9	23.9	19.3	19.0	24[SEA]/17[T]
Latent heat flux	85.8	86.2	91.9	92.8	88[SEA]/82[T]
TOA SW cld. forcing	-48.8	-48.1	-48.9	-48.5	-45.4[C]
TOA LW cld. forcing	21.1	21.1	18.8	19.0	25.9[C]
TOA Net. Rad. Imb.	0.42	0.39	0.66	0.62	0.41±0.03[NN]
Trop. lower strat. water vapor minima (ppmv)	3.0	2.8	4.5	4.4	3.8±0.3[D]
Zonal mean tropopause temp. (min., DJF) (°C)	-81	-82	-80	-80	-80
Hadley Circ. (10^9 kg s^{-1}) (DJF)	205	207	206	208	170–238[S]

Table 2. Global annual ensemble mean model features over the period 1979-2014 (1980-2004 for the E2 models) and key diagnostics compared to observations or best estimates. Cloud cover is estimated based on clouds with optical thickness >0.1 . [J] *Jones et al.* [1999] with updates, [C] CERES EBAF Ed4.1 *Loeb et al.* [2019], [T] *Trenberth et al.* [2009] and updates, [G] GPCP V2.3/TRMM TMPA V7 *Huffman et al.* [2007, 2009], [O] Obs4MIPs, [NN] Dervied from NOAA NCEI ocean heat content data, [D] *Dessler* [1998], [S] *Stachnik and Schumacher* [2011], [SEA] *Stephens et al.* [2012], [SRK] *Stubenrauch et al.* [2013]

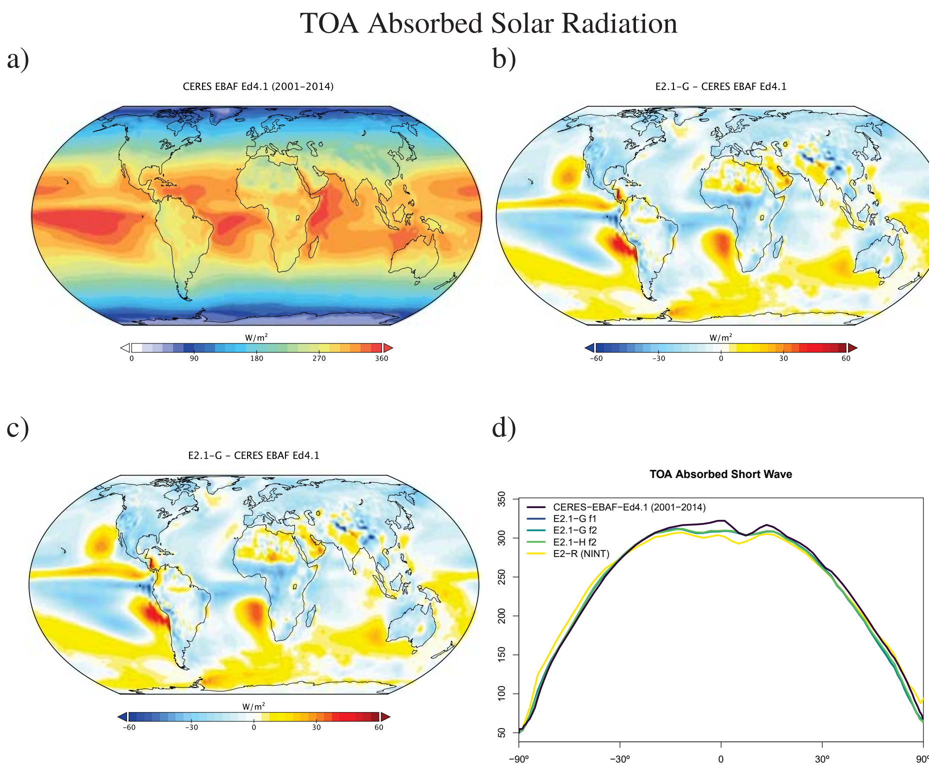


Figure 2. a) Annual climatology of TOA Absorbed Short Wave (W m^{-}2) in CERES EBAF Ed4.1 [*Loeb et al.*, 2019]. b) Difference of E2.1-G from CERES. c) Difference from E2.1-H. d) Absolute Zonal means, including E2.1-G (f1 and f2), E2.1-H and the earlier model version, E2-R.

5.2 Radiation and Clouds

Radiation diagnostics are now compared to the latest balanced CERES product (EBAF Ed4.1) *Loeb et al.* [2019]. Improvements since E2 are clearest in the Southern Ocean, where excessive SW absorption has been greatly ameliorated, but also in the tropics, although obvious biases associated with the marine stratus regions in the eastern ocean basins still exist (figs. 2 and 3). Notably the sign of the biases in the Arctic have changed in SW absorption. There is a lack of asymmetry across the tropics (which is clear in the observations), with the southern tropics characterised by excessive water vapor and cloud forcing. Note that Southern Ocean estimates of TOA absorbed SW (fig. 2) are somewhat better constrained than SW cloud radiative forcing (fig. 7).

Cloud fraction diagnostics have been upgraded to the ISCCP-H product for 1984–2014 [*Young et al.*, 2018]. The overall patterns are slightly improved in the tropics and mid-latitudes, but the persistent biases (in the marine stratus regions) remain clear (figs. 4 and 5). The bias in low cloud over sea ice regions may be an artifact. The improvements are clearer in the SW CRF diagnostic (fig. 7), and in the high latitudes at least for the LW cloud radiative forcing which remains overall too low (except in the erroneously cloudy tropical mid-Pacific (fig. 8). The cloud top pressure/cloud optical depth histograms (fig. 6) show that the model has improved its “too few - too bright" low cloud problem, as low cloud cover has increased and optical thickness has decreased in relation to the E2 version.

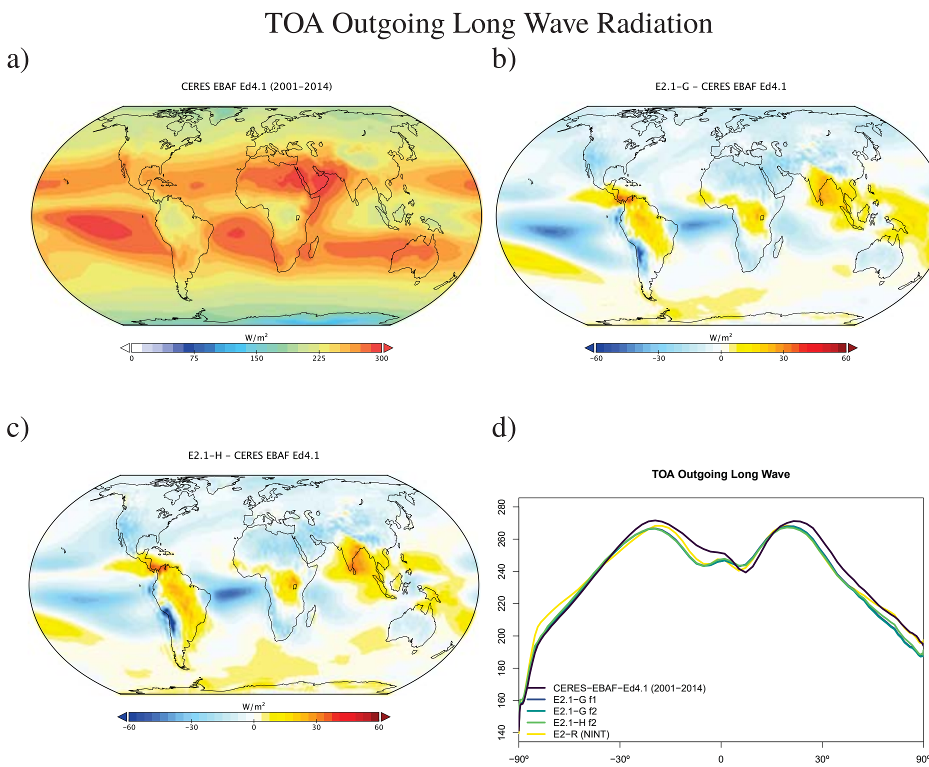


Figure 3. Annual climatology of TOA outgoing long wave data in data and models, as in fig. 2.

Additional information from comparisons of an earlier E2.1 version with active-sensor satellite observations (not shown) confirms an improvement of the low cloud cover in the high latitudes and over the trade wind regions while large biases remain over the stratocumulus regions in the tropics and subtropics. This low cloud bias might alter the strength of the low cloud feedbacks in response to global warming [*Cesana et al.*, 2019]. The large high-cloud positive bias found in E2 [*Cesana and Waliser*, 2016] has been mostly removed except in the southern tropics, where the overestimate of total cloud cover (fig. 4) comes from an excess of very high clouds (above 16 km), which are not present in satellite observations. Finally, the amount of supercooled water cloud with respect to ice cloud is overestimated on average [*Cesana et al.*, 2016]. As a result, the negative cloud-phase feedback [*Ceppi et al.*, 2016; *Tan et al.*, 2016] might be underestimated in E2.1 while it was overestimated in E2.

Atmospheric hydrological observations come from two blended data products via the Obs4MIPS archive [*Gleckler et al.*, 2011; *Teixeira et al.*, 2014; *Ferraro et al.*, 2015]. The precipitable water vapor is a blend of the RSS product over ocean [*Wentz and Schabel*, 2000; *Wentz et al.*, 2007] and MERRA-2 (over land) from the CREATE-MRE project [*Potter et al.*, 2018] while the precipitation product is a blend of TRMM satellite estimates over ocean [*Huffman et al.*, 2007; *Adler et al.*, 2009] and GPCP [*Huffman et al.*, 2009] Version 2.3 satellite-gauge calibrated precipitation over land. Precipitable water vapor discrepancies (fig. 9) are larger than in the previous model in the tropics, where the lack of asymmetry is readily apparent. The largest biases in water vapor coincide with the excessive LW CRF. This is also consistent with overall precipitation biases (fig. 10) which show a classic double-ITCZ problem in the Pacific, although one that is diminished in magnitude compared to E2. Excessive land precipitation in the Western Pacific Warm Pool has also been greatly ame-

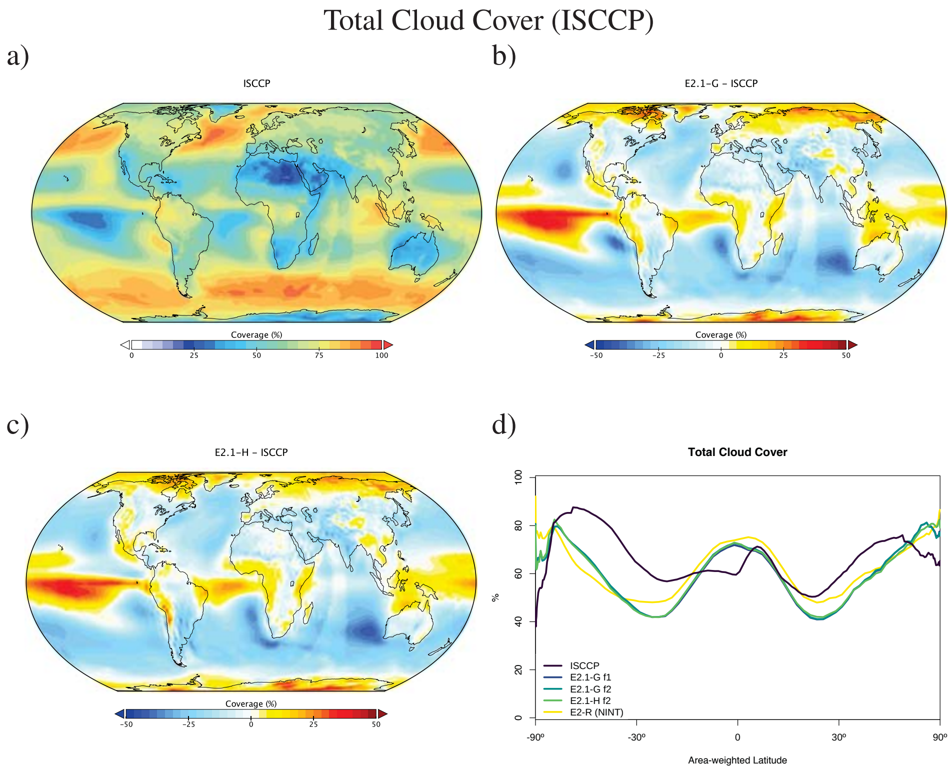
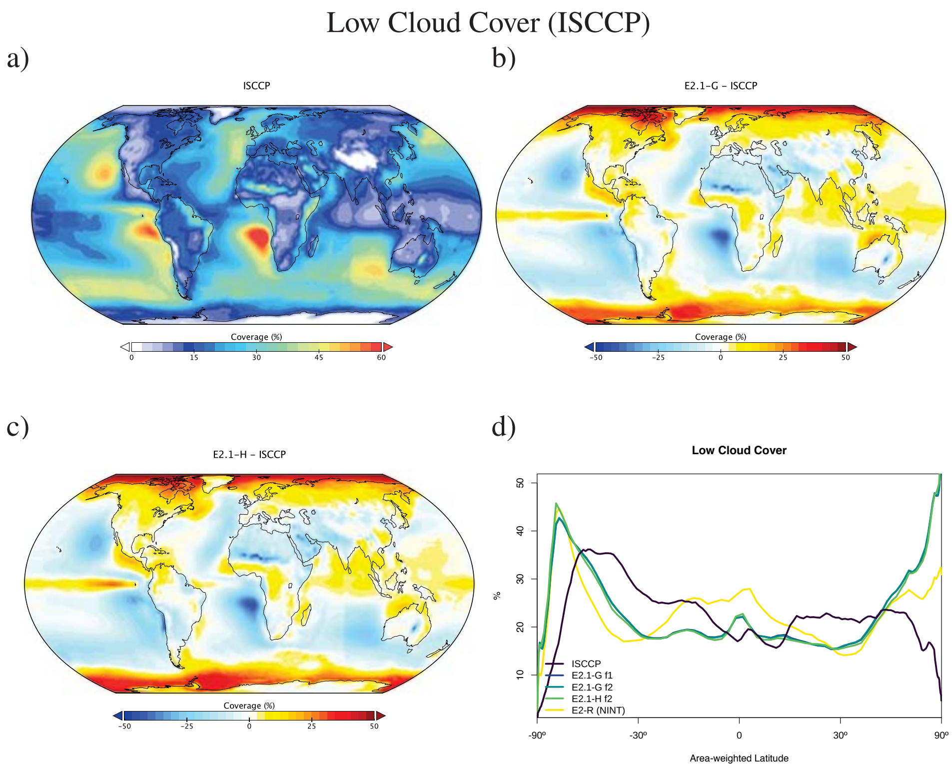


Figure 4. Annual climatology of Total Cloud Cover as seen by ISCCP-H, figure description as in fig. 2.

liorated. Note too, that part of the reduced bias in rainfall is due to upgrades in the observational product.

Snowfall biases are noticeable in the zonal mean (fig. 11), particularly in the Arctic, where excessive snowfall is related to wintertime cold biases in both models.



647 **Figure 5.** Annual climatology of Low Cloud Cover as seen by ISCCP-H, figure description as in fig. 2.

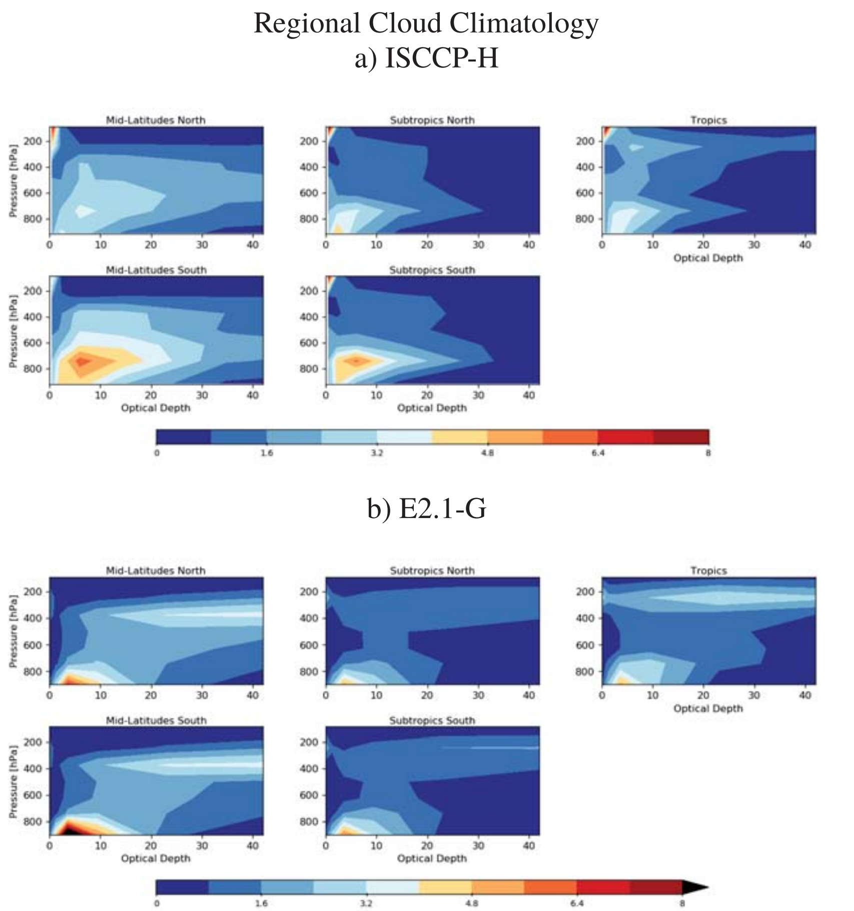
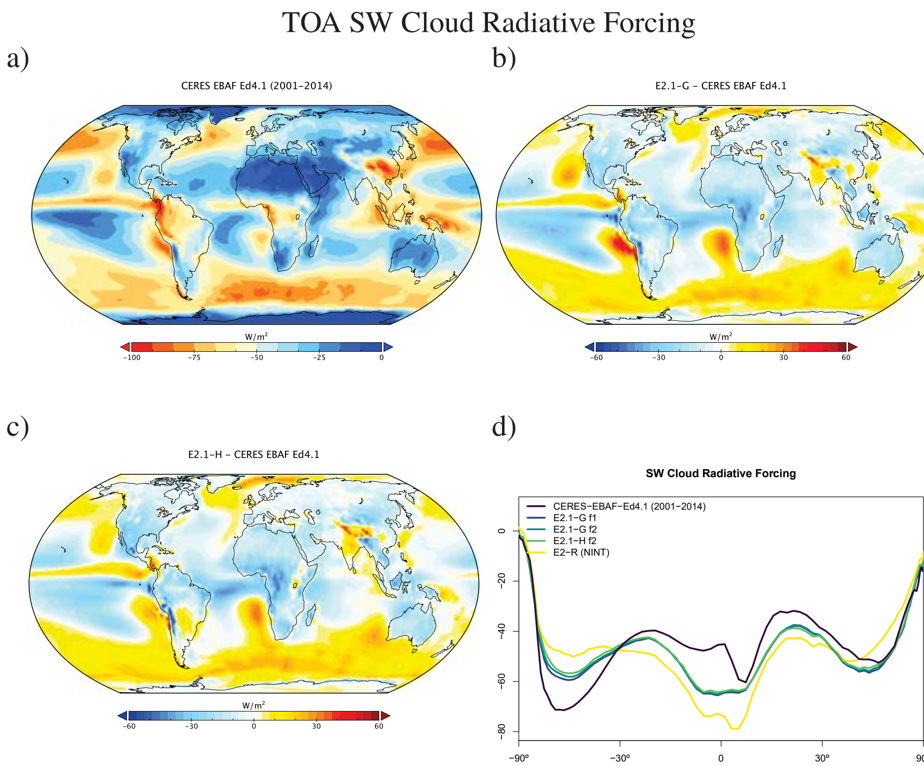
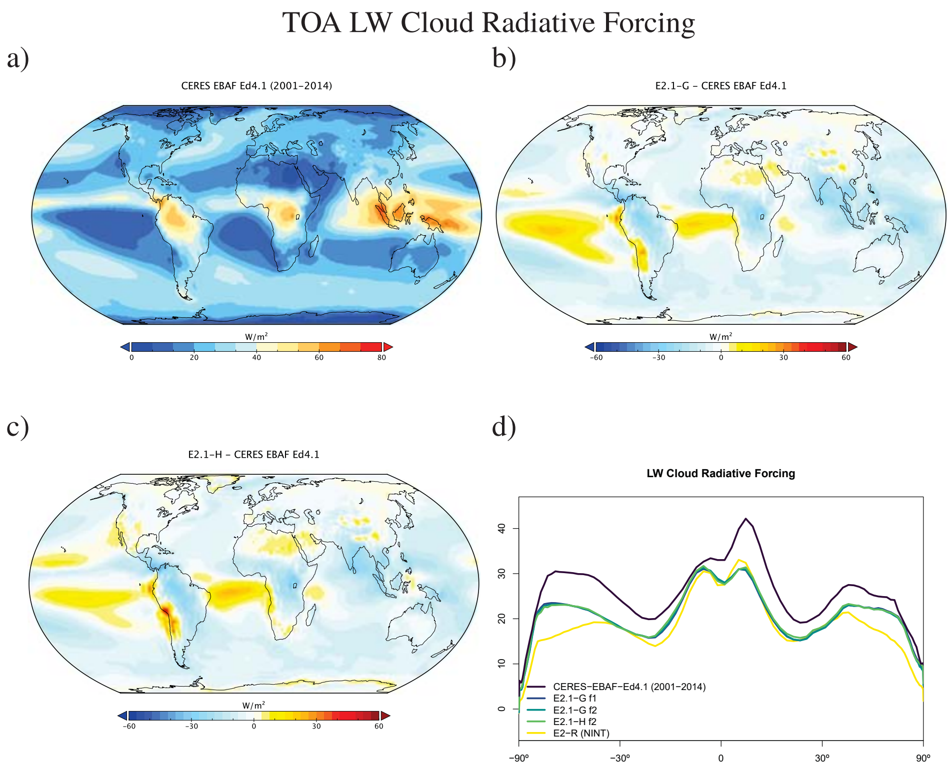


Figure 6. Climatology of cloud occurrence as a function of optical depth and pressure for five latitudinal bands as seen by ISCCP (60°N–30°N, 30°N–15°N, 15°N–15°S, , 15°S–30°S and 30°S–60°S). a) Data from ISCCP-H [*Young et al.*, 2018]. b) Data from the ensemble mean E2.1-G results. (Results from E2.1-H are indistinguishable).



652 **Figure 7.** Annual climatology of short wave cloud radiative forcing, figure description as in fig. 2



653 **Figure 8.** Annual climatology of long wave cloud radiative forcing, figure description as in fig. 2

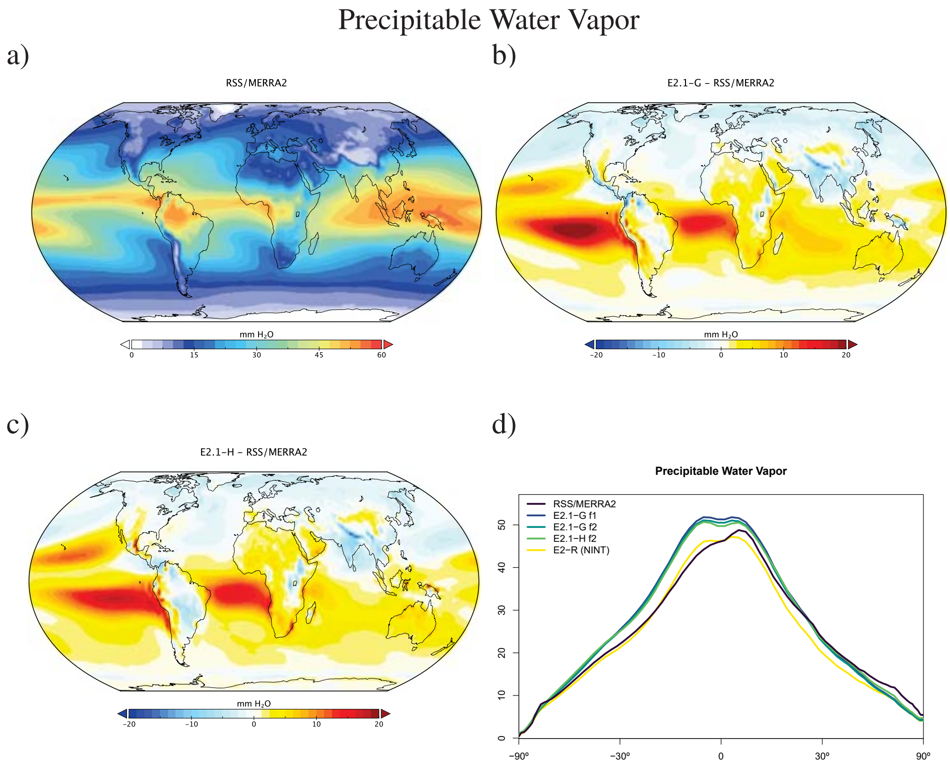
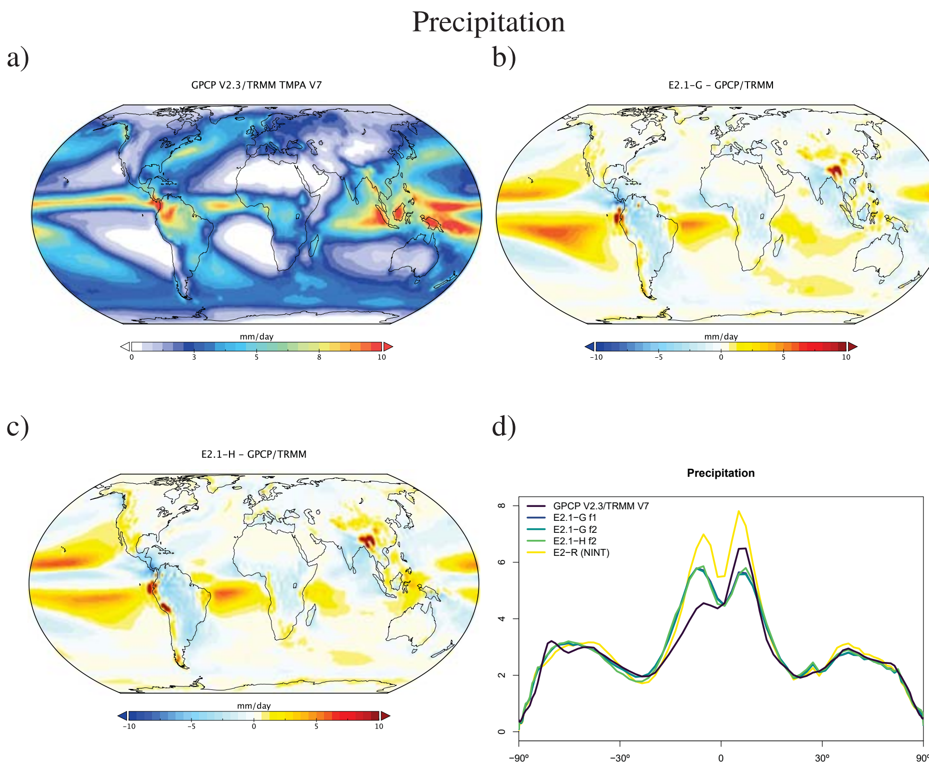
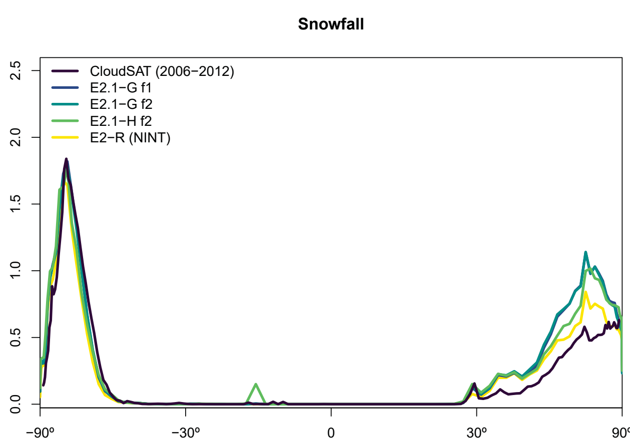


Figure 9. Annual climatology of precipitable water vapour, figure description as in fig. 2. Data derived from a blend of RSS and MERRA2 products over ocean and land respectively.



656 **Figure 10.** Annual climatology of precipitation. Figure description is as fig. 2.



657 **Figure 11.** Annual climatology of snowfall compared to CloudSAT data *Liu* [2008].

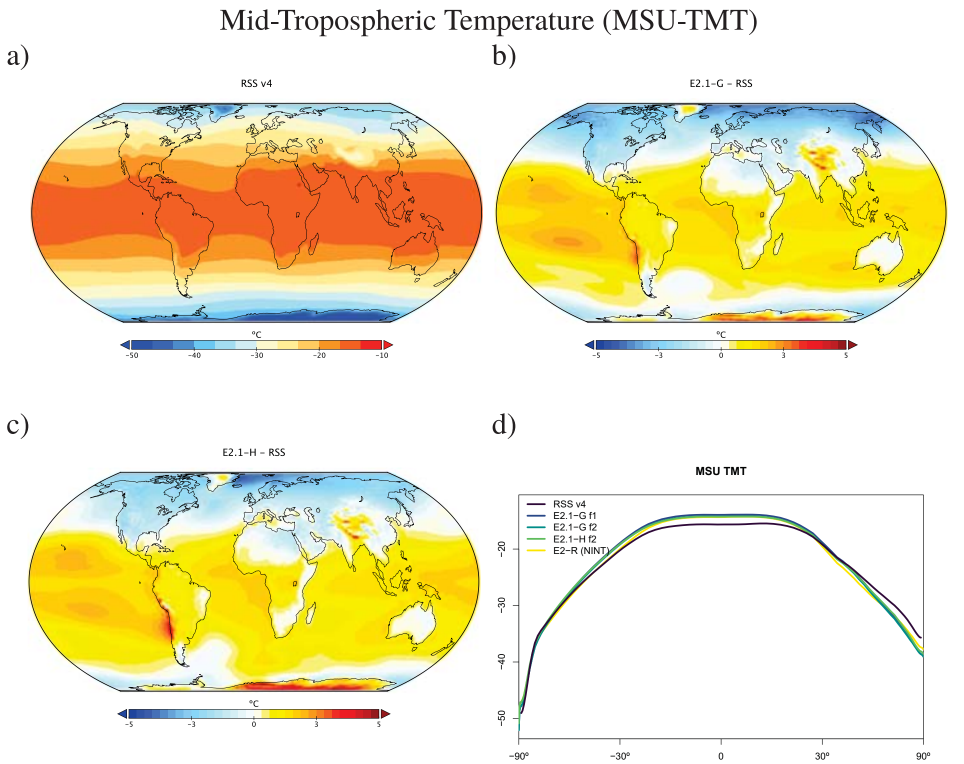


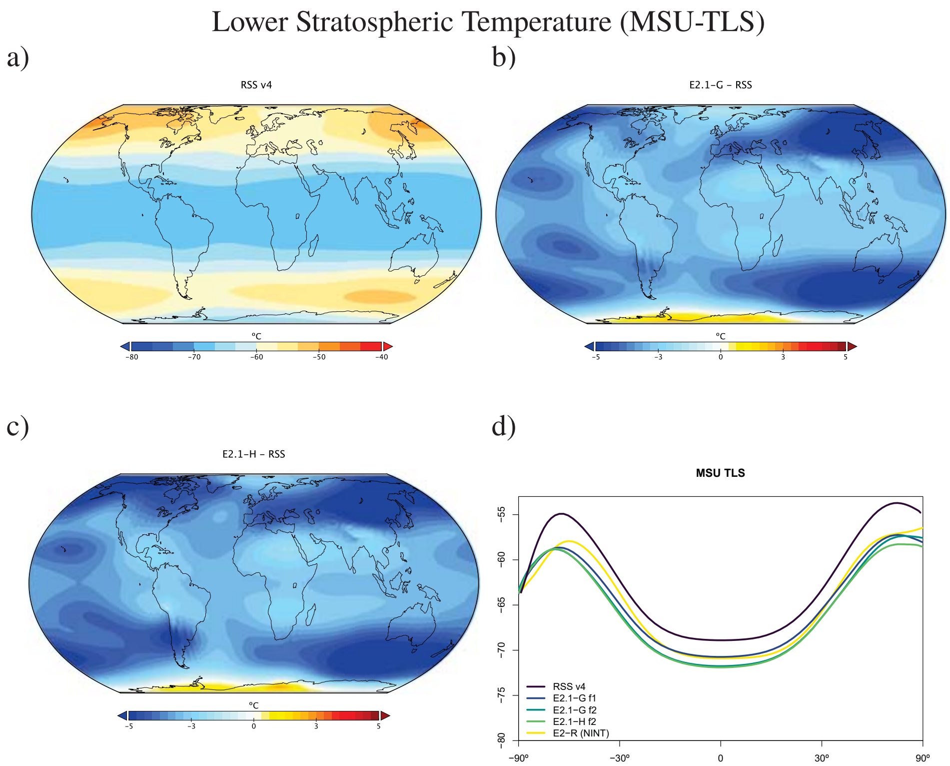
Figure 12. Annual climatology of MSU TMT. Observational data comes from RSS (1979–2014) (version 4.0) [*Mears and Wentz*, 2016]. Figure description is as fig. 2.

5.3 Satellite-derived Atmospheric Temperatures

The structure of temperature through the atmosphere plays a large role in defining fingerprints of climate change forcings, and so we compare the models to the Microwave Sounding Unit (MSU) and Stratospheric Sounding Unit (SSU) 1979–2014 brightness temperature climatologies (figs. 12, 13, 14). We highlight results from the mid-troposphere (TMT), the lower stratosphere (TLS) and middle stratosphere (SSU Channel 2) which have global weightings centered on 600, 70 and 4 hPa, respectively (though with substantial tails) [*Mears and Wentz*, 2016; *Zou and Qian*, 2016]. We use a static weighting function to estimate the channels, which though slightly less accurate than a radiative transfer calculation that takes into account surface emissivity, atmospheric water vapor, and temperature profiles [*Shah and Rind*, 1995], does not produce significantly different results.

Starting with MSU-TMT (fig. 12), the models show a notable warm bias in the tropics and sub-tropics, indicating a slightly steeper lapse rate in the troposphere than observed, and a cold bias in the northern high latitudes. Warm biases over high topography may be an artifact of the diagnostic comparison.

In the lower stratosphere (fig. 13), the models are anomalously cold, though partially the worse comparison to observations since E2 is related to an warmer climatology in the latest RSS version 4.0 [*Mears and Wentz*, 2016]. The mid and upper stratosphere (as illustrated by the SSU-2 channel, fig. 14) is too warm - particularly in the polar regions. This overall pattern of stratospheric temperature error is consistent with, but not completely explained by, a too weak Brewer-Dobson circulation in this relatively low-top model.



681 **Figure 13.** Annual climatology of MSU TLS. Observational data comes from RSS (1979–2014) (version
682 4.0) [*Mears and Wentz*, 2016]. Figure description is as fig. 2.

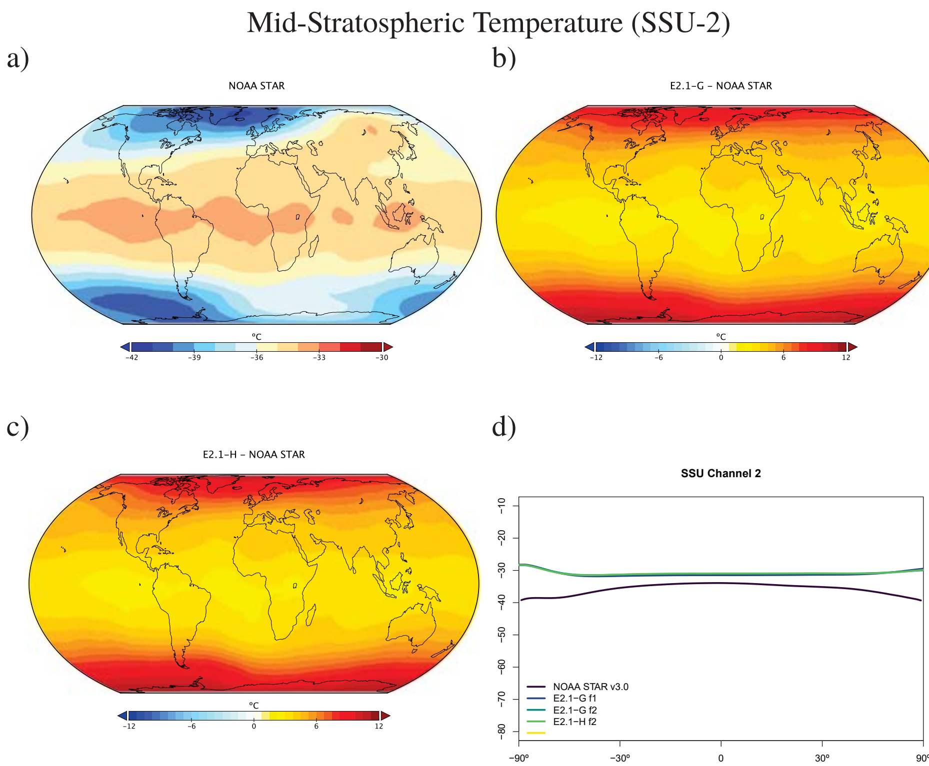


Figure 14. Annual climatology of SSU Channel 2. Observational data comes from NOAA STAR (1979–2014) (version 3.0) [*Zou and Qian*, 2016]. Figure description is as fig. 2.

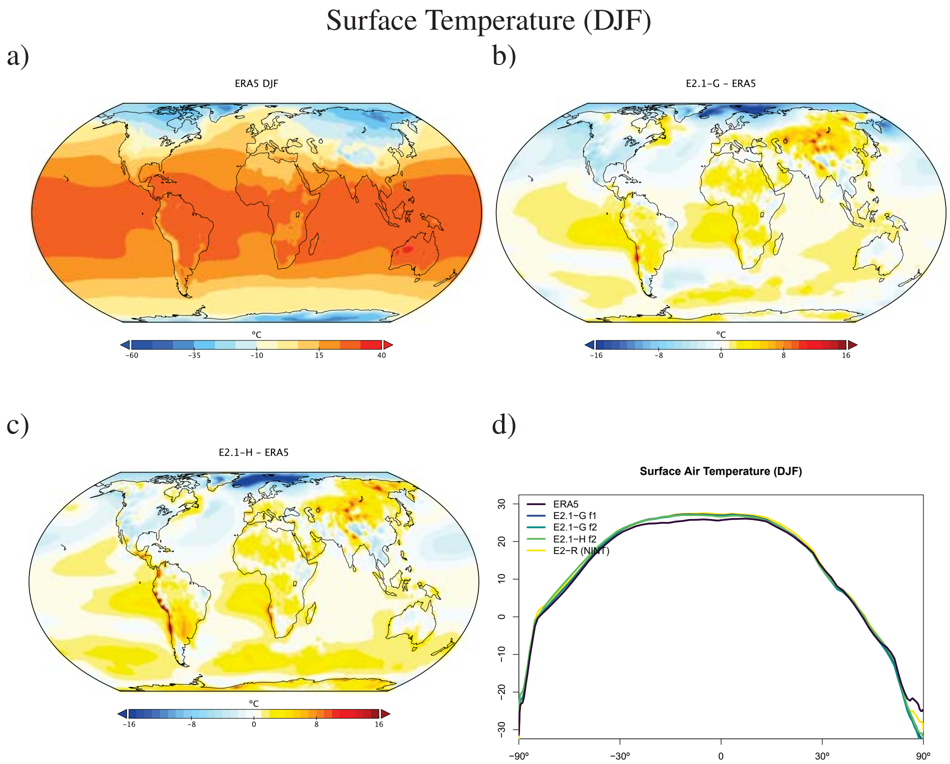


Figure 15. DJF climatology of surface air temperature. Figure description is as fig. 2.

5.4 Surface Fields

Surface field observations are taken from ERA5 [*Copernicus Climate Change Service (C3S)*, 2017]. Overall biases in the surface temperature fields (figs. 16 and 15) are similar to the situation seen in CMIP5, though the magnitude of errors in the Southern Oceans are notably reduced (consistent with the improvements of cloud and radiation diagnostics discussed above). Land errors are reduced, though the winter cool bias in the Arctic is slightly increased.

Sea level pressure biases are quite different between the two ocean model versions (figs. 18 and 17), with E2.1-G having more excessive high pressure in the tropics than in E2.1-H. This is partially explained by the higher than observed water vapor in the models, and the topographic change made in the HYCOM land-ocean grid which increased surface pressure over land (and through conservation, decreased it over the oceans). In the northern summer, both models fail to generate as large a jet stream gradient as observed. However, the overall pattern of surface wind stress is improved from E2 (fig. 19), with a notably more poleward and stronger maxima in the mid-to-high latitudes. There remains a westward bias in the eastern tropical Pacific.

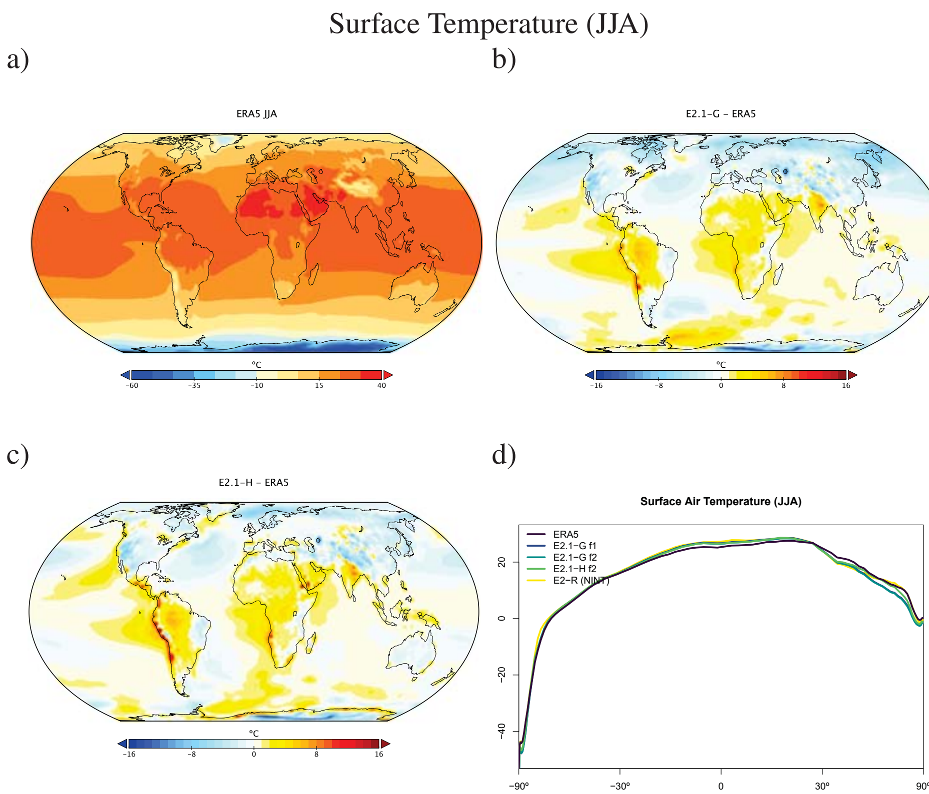


Figure 16. JJA climatology of surface air temperature. Figure description is as fig. 2.

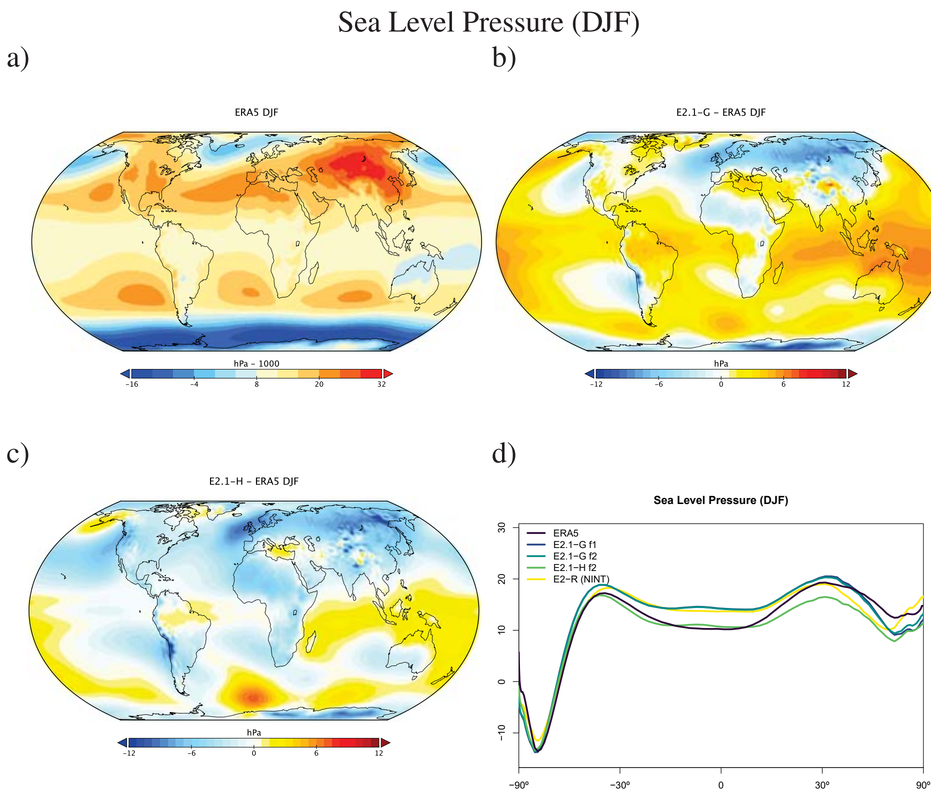


Figure 17. DJF climatology of sea level pressure (including water vapor mass in the diagnostic, even though it is not seen by the dynamics). Figure description is as fig. 2.

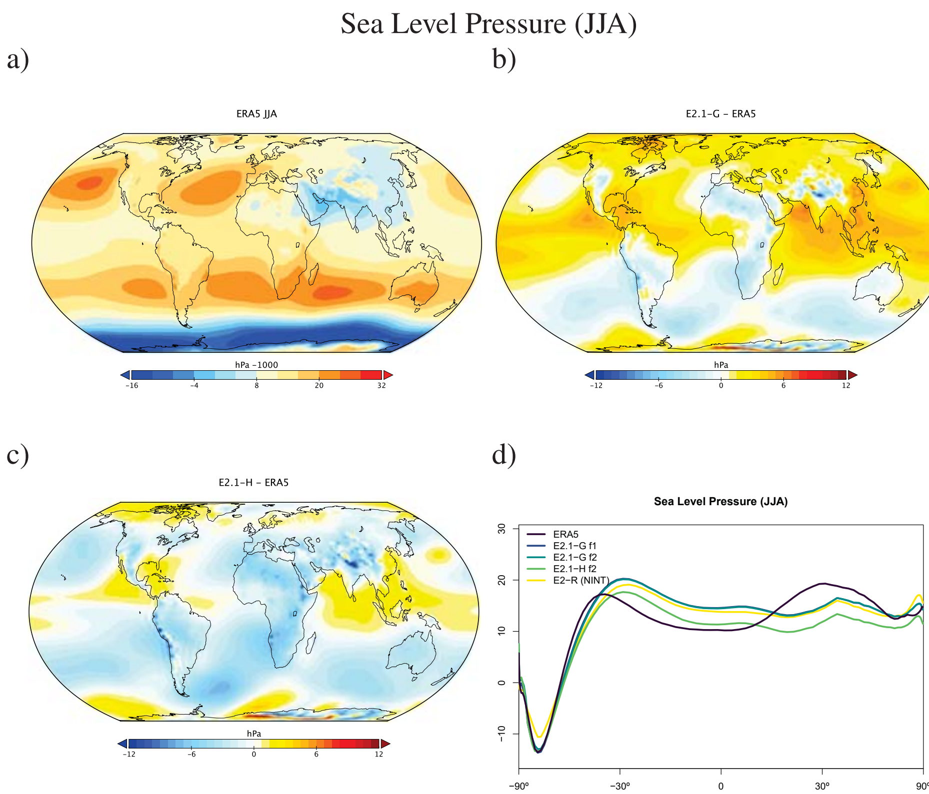
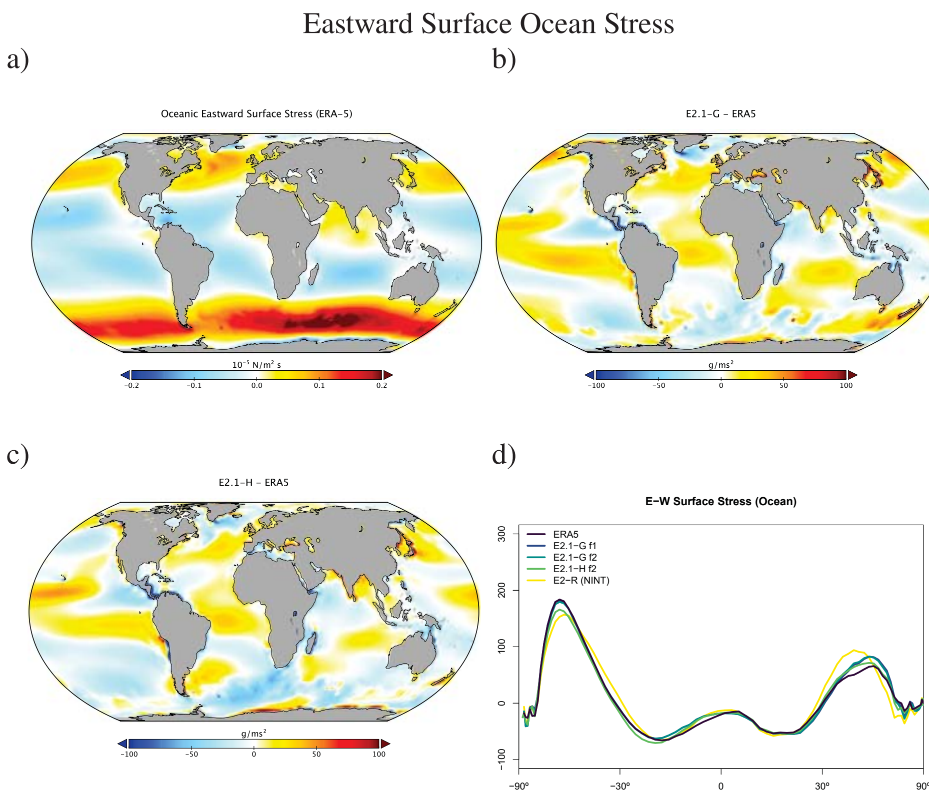


Figure 18. JJA climatology of sea level pressure. Figure description is as fig. 2.



706 **Figure 19.** Annual climatology of Eastward surface ocean stress. Figure description is as fig. 2.

River	E2.1-G	E2.1-H	E2-R	E2-H	Observations
Amazon	251±8	280	198–236	229–300	545
Congo	22±1	36	35–69	41–82	106
Brahmaputra-Ganges	125±4	81	68–86	110–140	105
Yangtze	108±2	111	85–100	191-210	78
Lena	44±1	41	32–34	29–31	40
Ob	52±1	38	47–52	80–89	33
St. Lawrence	56±1	35	53–55	27–28	29
Mackenzie	23±1	29	28–29	31	24

Table 5. Annual mean discharge from selected rivers (km^3 $month^{-1}$). Ensemble mean and standard deviation for E2.1-G, range of model values from E2-R/H. Observations from *Fekete et al.* [2001].

Runoff from the major rivers can be compared to observational data [*Fekete et al.*, 2001] (Table 5). In the tropics, runoff is severely deficient in the Amazon basin and African rain forests (due to insufficient rainfall) where skill has not increased compared to earlier model versions. High latitude rivers are, however, more consistently modeled.

5.5 Ocean

We focus here on the diagnostics that most impact the coupled simulation and are straightforwardly comparable to observations. More detailed description and analysis of E2.1 ocean circulation and structure will be presented elsewhere.

Sea surface temperature biases (fig. 20) are still dominated by the errors in the marine stratus regions, though are (again) improved in the Southern Ocean region. Arctic biases are colder than before. Overall, tropical temperatures are slightly warm, particularly in the southern tropics, which is consistent with the errors in precipitable water vapour, clouds and radiation seen above. Remarkably, the difference seen across the ocean models is quite small.

Salinity biases in E2.1-G are far smaller than in E2-R, particularly in marginal seas, but also in the open ocean (fig. 21). Clear positive biases are obvious near major river mouths (consistent with insufficient river outflow seen in Table 5).

For HYCOM, the biases in surface salinity (fig. 21c) have been totally reversed, in part due to the correction to virtual salt fluxes, from a large excess salinity in E2-H, to an overall underestimated salinity in E2.1-H, though with a reduced overall error. Arctic biases are noticeably reduced, possibly associated with the implementation of the BP ice thermodynamics.

Ocean transports are also greatly improved, notably the Drakes Passage where offsets to the observed transport are much less than previously in both models (Table 6). Fluxes through the Gulf Stream and Kuroshio Current are reasonable, but slightly higher than inferred from observations. The mass and heat transports at 26°N from the N. Atlantic overturning circulation in E2.1-H are in good agreement with direct observations [*McCarthy et al.*, 2015; *Smeed et al.*, 2019; *Johns et al.*, 2011], but larger in E2.1-G.. Poleward heat transports peak above 1 PW at ~20°N, this is significantly higher than the estimates derived from a ocean state estimation approach [*Forget and Ferreira*, 2019] (fig. 22), but in reasonable agreement with direct heat flux estimates [*Ganachaud and Wunsch*, 2003]. Poleward transports in the southern oceans in E2.1-G are much more consistent with both direct measurements and ocean state estimates.

Sensitivity experiments with a reduced tidal mixing efficiency in E2.1-G suggested that tuning of this parameter could match the target Atlantic overturning transport metric at 26°N and the *Forget and Ferreira* [2019] heat transport there, but with the penalty of unacceptably

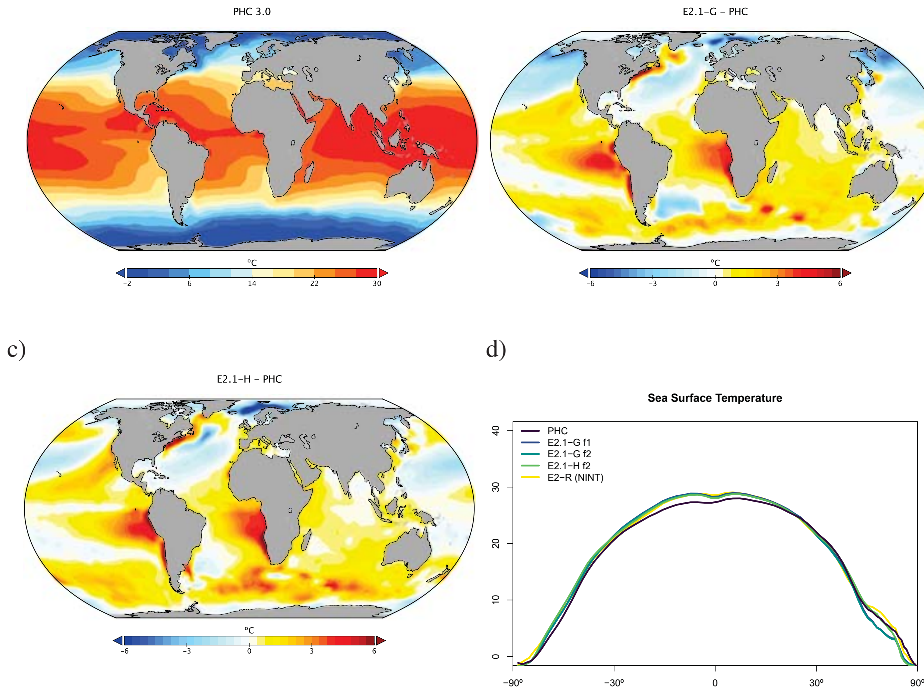


Figure 20. Annual climatology of sea surface temperature compared to the PHC 3.0 product. Figure description is as fig. 2.

increasing cold biases in northern midlatitudes and the Arctic. Such compromises will be revisited in future model versions having improved cloud radiative forcing and atmospheric transports. Ocean-only experiments with an E2.1-G prototype [*Romanou et al.*, 2017] indicate that its CFC uptake is best matched in configurations having weaker AMOC magnitudes than those realized here, with implications for heat and carbon uptake.

5.6 Cryosphere

Figures 23 show that modeled seasonal cycles of sea ice in the Arctic and Antarctic have improved, although the absolute value of Arctic sea ice is still too large. Improvements in snow depth, meltpond extent, and albedo (as compared to SHEBA data [*Curry et al.*, 2001]) are clear (fig. 24) and due to model fixes and the inclusion of the Brine-Pocket thermodynamics scheme. Southern Ocean improvements are largely due to a more stratified ocean and an associated reduction of vertical mixing in both E2.1-G and E2.1-H, as opposed to changes in sea ice physics or properties (as has been the case previously [*Liu et al.*, 2003]). Brighter clouds in the Southern Ocean in E2.1 also cool temperatures and aid ice formation there.

Diagnostic	E2.1-G	E2.1-H	E2-R	E2-H	Observations
N. Atl. MOC (Max)	27.2	20.4±0.3	27.2±0.7	24.5± 0.8	-
N. Atl. MOC (26°N)	24.8±0.4	17.8±0.3	18.4±0.3	22.4±0.6	≈18[R19]
Atl. Heat (26°N)	1.21±0.01	1.06±0.01	0.97±0.01	0.99±0.02	1.3±0.4[J11]/0.88±0.01[F19]
ACC (Drake Pass.)	150±1	178±1	254± 1	192±2	130[P88]/173[D16]
Gulf Stream	55±1	48.2±0.3	49± 1	39.8±0.8	≈35[R11]
Kuroshio	49±1	67±2	64±1	71.7±0.5	≈57[I01]
Bering Str.	0.16±0.003	-0.55±0.01	0.16±0.01	0.45±0.01	0.8±0.2[W05]
Indonesian throughflow	18.9±0.2	18.4±0.2	11.5±0.2	17.6±0.3	15[S09]

Table 6. Selected ocean mass (Sv) and heat (PW) fluxes. Range is standard deviation of the 1979–2014 average from 5 ensemble members for each configuration. Observations: [R19] *McCarthy et al.* [2015]; *Smeed et al.* [2019] (estimate at 26°N); [P88] *Petersen* [1988]; [D16] *Donohue et al.* [2016]; [J11] *Johns et al.* [2011]; [I01] *Imawaki et al.* [2001]; [W05] *Woodgate et al.* [2005]; [S09] *Sprintall et al.* [2009]; [F19] *Forget and Ferreira* [2019]

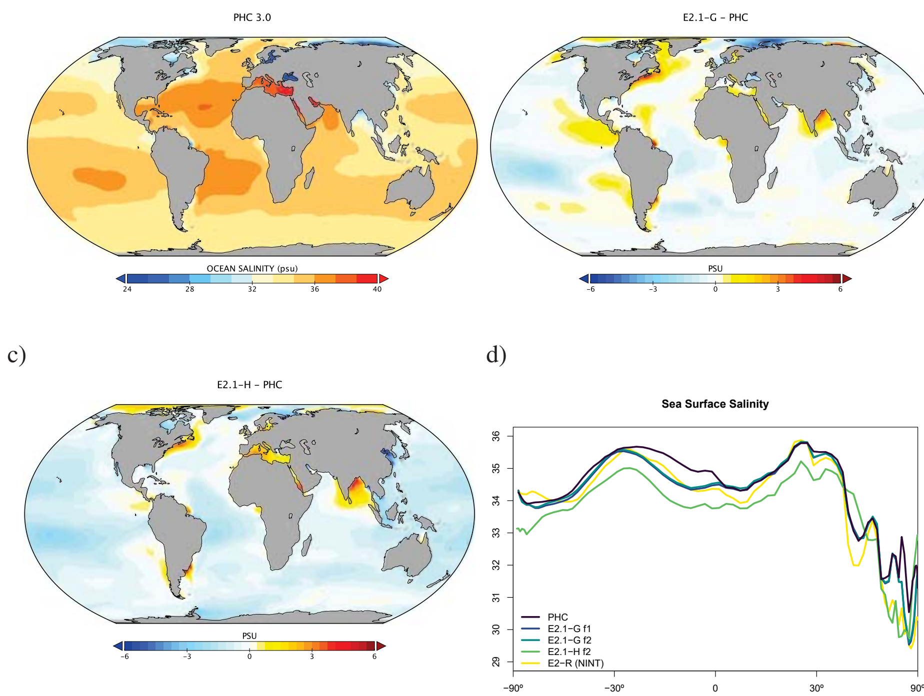


Figure 21. Annual climatology of sea surface salinity compared to the PHC 3.0 product. Figure description is as fig. 2.

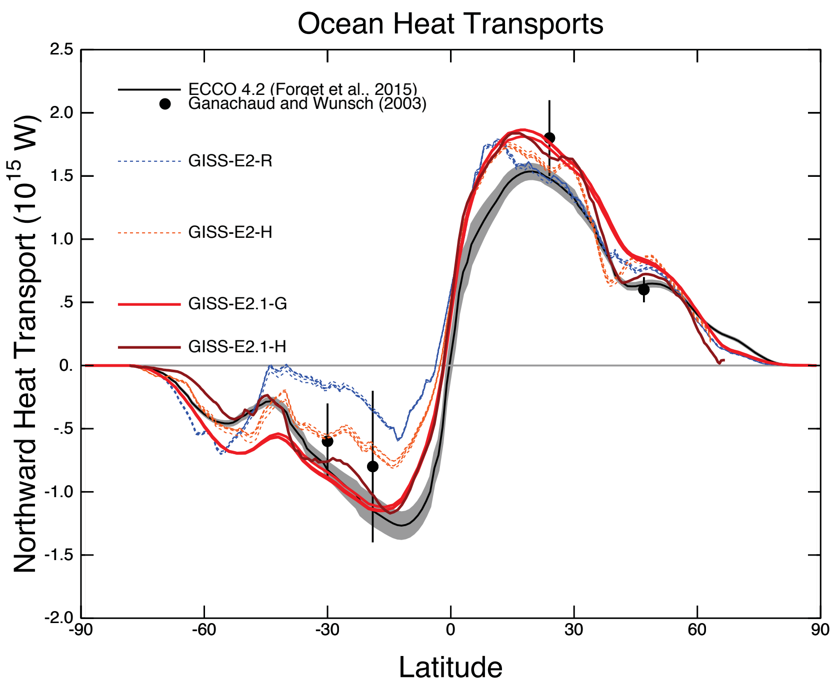


Figure 22. Annual mean global northward ocean heat transports. Comparisons of the models with mean estimates from 1992–2011 from the ECCO ocean state estimate (v4 release 2) with 95% confidence intervals on the mean derived from the interannual variability [*Forget et al.*, 2015; *Forget and Ferreira*, 2019] and oceanographic estimates [*Ganachaud and Wunsch*, 2003].

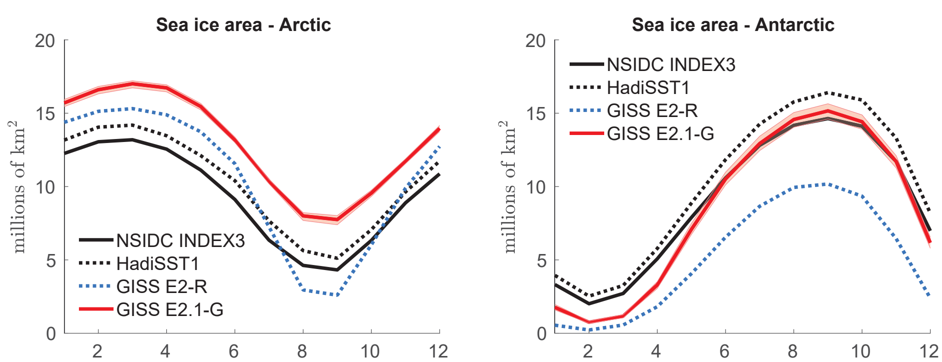


Figure 23. Annual climatology of sea ice area in both hemispheres in E2-R (blue dashed) and E2.1-G (red). Observational data comes from NSIDC (1979–2014), after correction for the Arctic polar 'hole' [*Fetterer et al.*, 2011] and HadISST1 (1979–2014) [*Rayner et al.*, 2011]. The ensemble mean climatology is plotted for E2-R (1979–2012) and E2.1-G (1979–2014, with spread across E2.1-G ensemble members in pink).

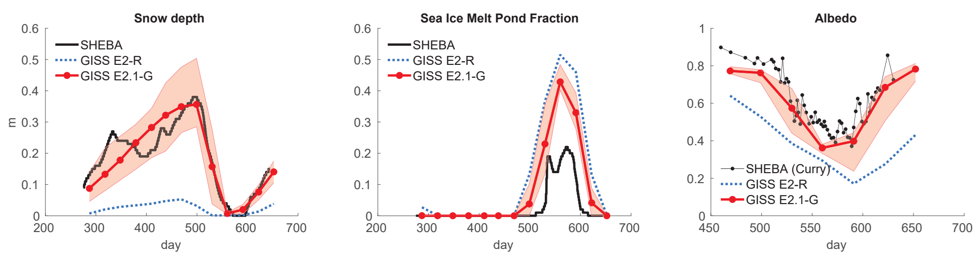


Figure 24. Spot comparisons of the E2-R (blue dashed) and E2.1-G (red) simulations against the SHEBA measurements for snow depth, meltpond fraction and albedo [*Curry et al.*, 2001]. Ensemble spread for E2.1-G is in pink.

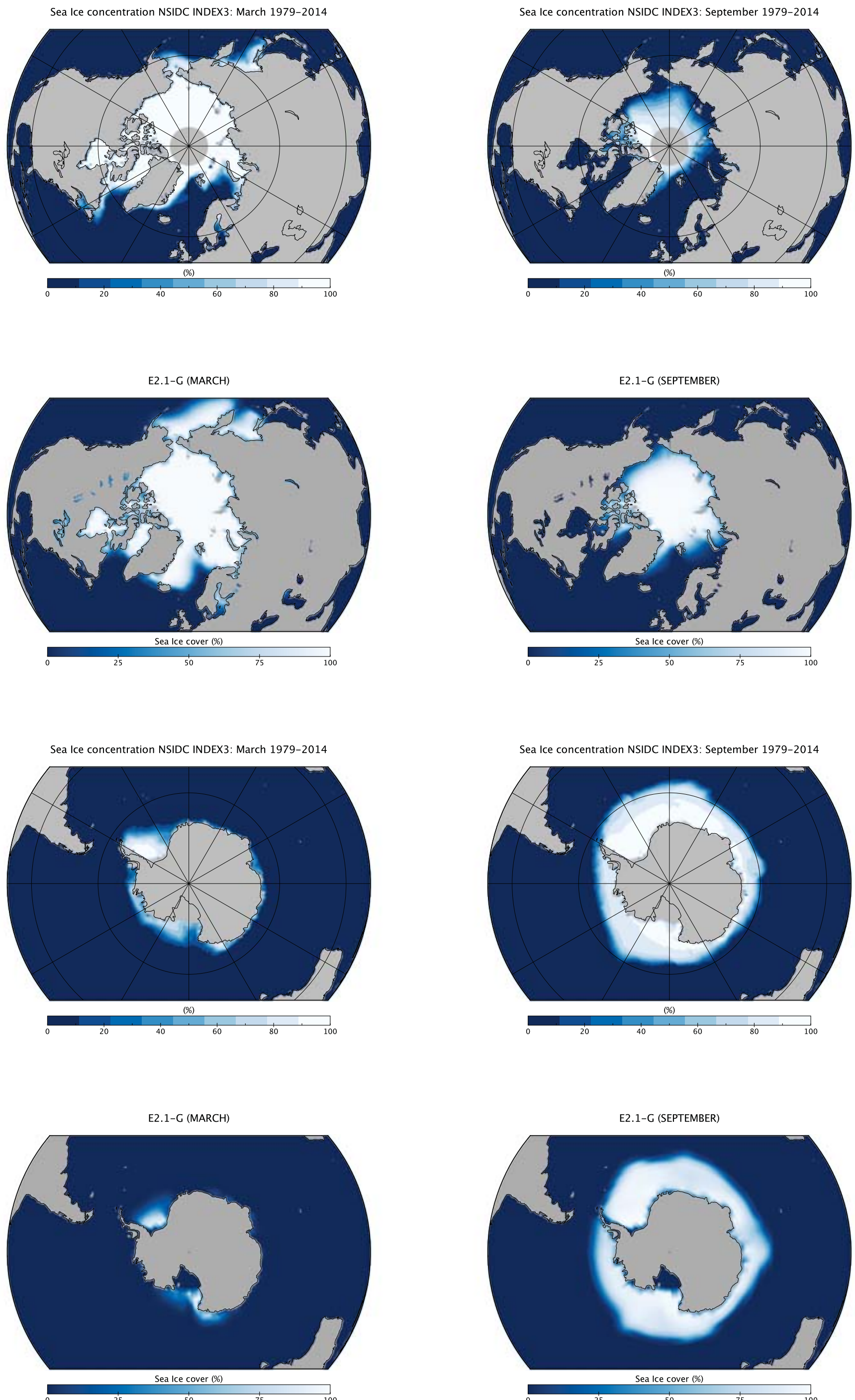


Figure 25. Sea ice concentration (%) for March (left column) and September (right column) in the NSIDC observations and E2.1-G simulations. Rows a) and b) Arctic, and rows c) and d) Antarctic. E2.1-H results are similar.

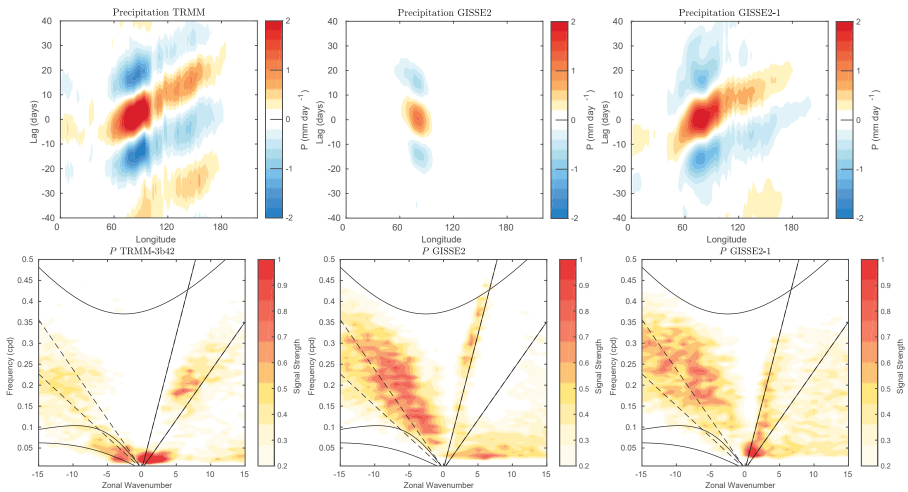


Figure 26. Comparison of MJO signals and propagation in the TRMM data (release 3b24), [*Iguchi et al.*, 2000] and in E2-R and in E2.1-G simulations. (Top) Hovmöller plots of MJO propagation. (Bottom) Wheeler-Kiladis diagrams for tropical wave motion [*Wheeler and Kiladis*, 1999]. Figures courtesy of Ángel Adames.

5.7 Model internal variability

As model processes have become more sophisticated and the base climatology has become more realistic, the realism of the patterns of internal variability has also improved. We focus here on ENSO, the PDO and the MJO because the improvements over previous models have been most dramatic. Notably, while the MJO was a specific target for model improvement through the model development process, the changes in ENSO and PDO patterns emerged as part of the overall improvement in skill.

The MJO improvement is highlighted in figure 26, where the lack of MJO-related activity and lack of propagating features in the Pacific in E2-R was very clear in comparison with an analysis of the TRMM data. However, in E2.1-G, the improvement in propagation and in the wavenumber/frequency plot [*Wheeler and Kiladis*, 1999] is evident.

For the longer term tropical modes, ENSO and the PDO, there have been large improvements in the patterns of associated temperature variability (figs. 27a and 28a) across CMIP generations, and particularly since CMIP5. However, that improvement must be tempered by a recognition that the spectral signature of ENSO has not improved. In all versions of E2, there was insufficient overall variance, and in particularly a deficit in variability at 3–7 years (overall standard deviations were 0.60°C for E2-R and 0.67°C for E2-H, compared to ∼1°C in the ERSST5 observations). However, in E2.1-G and E2.1-H the 2 to 4-year variability is now too strong (figs. 27b). The overall Nino3.4 standard deviation is modestly too high (1.2°C) in E2.1-G and too low in E2.1-H (0.75°C), compared to ∼1°C in the ERSST5 observations [*Huang et al.*, 2017]. This nonetheless impacts the interannual variability in these simulations, even for the global mean, leading us to increase the number of ensemble members in the E2.1-G historical simulations in order to better assess the forced responses.

The larger overall ENSO variability and unrealistically peaked spectral signature in E2.1-G relative to E2.1-H suggest that ocean interior structure and damping mechanisms exert as much influence as atmospheric processes. Some of the latter have been quantified in feedback form for E2.1-G following Figure 7 in *Bellenger et al.* [2014]. Specifically, the wind-stress (positive) feedback is $9.8{\times}10^{-3}$ N m^{-2} °C^{-1}, 20% weaker than in ERA40, and

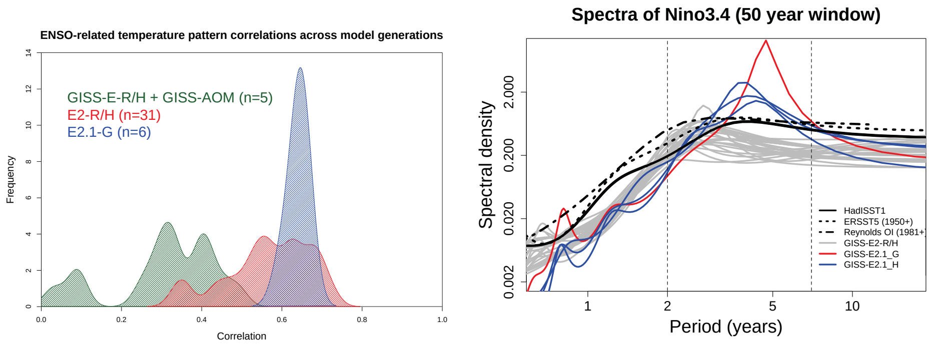


Figure 27. a) Improvement of pattern correlations of ENSO to the observations over GISS model generation (from CMIP3 to CMIP6). Calculations via the Climate Variability Data Portal (CVDP [*Phillips et al.*, 2014]). b) Spectra of Nino3.4 variability in 50 year segments from the PI-controls compared to various observational products.

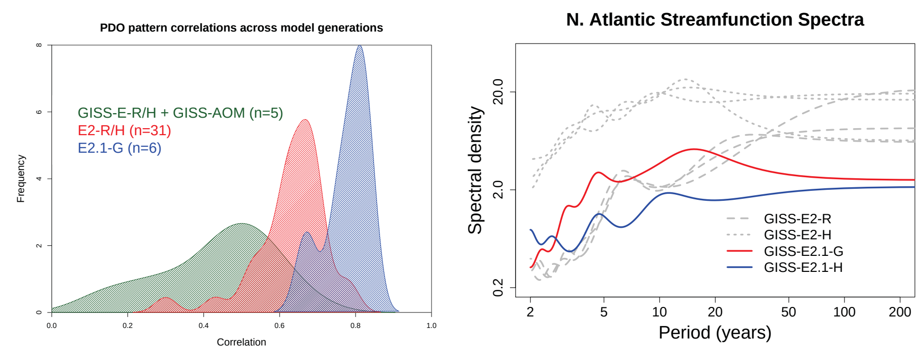


Figure 28. a) Improvement of pattern correlations of the PDO to the observations over GISS model generation (from CMIP3 to CMIP6). Calculations via the Climate Variability Data Portal (CVDP [*Phillips et al.*, 2014]). b) Spectra of variability in the N. Atlantic annual mean maximum streamfunction (derived from a detrended 1000 years of PI-control simulation).

the surface-flux (negative) feedback is -12.5 W m^{-2} °C^{-1}, 30% weaker than observed. In a sensitivity test (similar to one reported in *Rind et al.* [2020]), we applied a change to the atmospheric convection scheme that led to reduced ENSO amplitude and a shift of the peak to shorter periods. Both of the feedback coefficients are significantly smaller in that simulation, suggesting that its ENSO improvement occurred for the wrong reasons, and overall model skill was not enhanced. This remains an active area of model testing, although we anticipate that it will require a substantial improvement of marine stratus biases (as a function of increased vertical resolution and better moist physics) before tuning for the correct ENSO feedbacks will become worthwhile.

In the North Atlantic, where decadal and greater variability is associated with the overturning streamfunction there are mixed changes. There is greater variability at 8–15 yrs for E2.1-G compared to E2-R, but significantly less variability in E2.1-H compared to E2-H. The standard deviation of the detrended annual streamfunction maximum is 1.4 Sv for E2.1-G, and 1.0 for E2.1-H (compared to 1.3 and 3.2 Sv in E2-R and E2-H respectively). This can be compared to the interannual variability in the meridional overturning circulation at 26°N [*McCarthy et al.*, 2015; *Smeed et al.*, 2019] of ~1.3 Sv.

Field	E2.1-G (`f2`)	E2.1-G (`f1`)	E2.1-H (`f2`)	E2-R	E2-H
OTR	**0.68**	**0.68**	0.67	0.66	0.63
ASR	0.84	0.84	**0.85**	0.79	0.78
MSU-TMT	0.89	0.88	**0.90**	**0.90**	**0.90**
MSU-TLS	0.64	0.69	0.62	**0.73**	0.71
TOTAL CLOUD	**0.32**	**0.32**	0.31	0.19	0.17
LOW CLOUD	**0.35**	**0.35**	0.34	0.16	0.12
SLP (DJF)	0.76	0.75	**0.81**	0.78	0.71
SLP (JJA)	0.82	0.82	**0.83**	0.79	0.75
SAT (DJF)	**0.90**	**0.90**	0.89	**0.90**	0.88
SAT (JJA)	**0.90**	**0.90**	**0.90**	**0.90**	0.87
PRECIP	**0.52**	0.51	0.51	0.50	0.45
EWSS	**0.81**	**0.81**	0.77	0.78	0.71
SST	0.90	0.90	0.90	**0.91**	0.86
SSS	**0.73**	**0.73**	0.57	0.63	0.54

Table 7. Arcsin-Mielke scores across model configurations for selected fields as referenced above (see fig. 29 for the field definitions, though we also add sea surface temperature (SST) and salinity (SSS) in this table). The highest score across the coupled models for each field is highlighted. Note that for the E2 models, the output data is from 1979–2004, while the target climatologies are as described above.

5.8 Summary Statistics

We are interested both in how model evolution affects skill scores, but also in how the GISS model compares to similarly functional models in the CMIP5 and CMIP6 ensembles. Improvements across the board are seen in standard large scale climatological tests as seen in the Taylor diagram comparing E2-R (NINT) with E2.1-G (NINT) (Fig. 29). The improvements are largest in fields that were the worst performing in CMIP5 (clouds, precipitation), though still positive for even well-simulated fields. As in previous papers, we can calculate an Arcsin-Mielke score (between 0 and 1) [*Watterson*, 1996] for a suite a standard variables (Table 7). These reflect the same general tendencies. Differences between the `f1` and `f2` ensembles are barely perceptible (except for MSU-TLS which is a little better in the `f1` ensemble).

Any overall ranking of performance is by necessity ad hoc, and not determinative of every metric, but across a range of measures, the E2.1-G (`f2`) is the best performing configuration considered here. There are small degradations of skill for the MSU diagnostics (though not for the trends [*Miller et al.*, 2020]). E2.1-H has slightly better SLP patterns, but the differences in atmospheric variables is minor, especially compared to the improvements of all E2.1 configurations with respect to E2.

6 Climate Sensitivities

As part of the DECK simulations requested by CMIP6, we performed a number of idealized simulations (1pct4xCO2, abrupt4xCO2) as well as some related simulations (abrupt2xCO2 with the coupled and q-flux ocean versions) (all performed with the `f1` background composition). The summary of various metrics of climate sensitivity (along with the comparison to the previous models) is seen in Table 8. We note that the effective CS as calculated by the Gregory method [*Gregory et al.*, 2004] almost always underestimates the true long term ECS by 10 to 20% [*Dunne et al.*, 2020]. The perhaps more relevant TCR is slightly larger in the E2.1 models than previously, consistent with a smaller rate of mixing of heat into the ocean (and slightly smaller overall radiative imbalance (Table 2)

Model version & configuration	ECS_{qflux}	CS_{Eff}	ECS from $4\times CO_2$	ECS from $2\times CO_2$	TCR
E2.1-G (NINT)	3.0	2.7	3.2	3.6	1.8
E2.1-H (NINT)	"	3.1	3.5	3.4	1.9
E2.1-G (OMA)	2.9	2.6			1.6
E2.1-H (OMA)	"	3.1			2.0
E2.1-G (MATRIX)					1.8
E2.1-G (MATRIX)					2.0
E2.1-G (TOMAS)	3.1				
E2-R (NINT)	2.7	2.1	2.3	2.6	1.4
E2-H (NINT)	"	2.3	2.5		1.7
E2-R (TCADI/OMA)	3.0	2.4			1.6
E2-H (TCADI/OMA)	"	2.5			1.8

Table 8. Climate Sensitivities to $2\times CO_2$ (°C) estimated multiple ways (note that not all calculations have been completed with all versions). Equilibrium Climate Sensitivity (ECS) is defined from multi-millennial coupled simulations, or from a q-flux (slab ocean) model (ECS_{qflux}). CS_{Eff} is from a linear extrapolation of yr 1–150 results in the abrupt4xCO2 simulations [*Gregory et al.*, 2004]. Transient Climate Responses (TCRs) are taken from year 70 in the 1pct4xCO2 simulation.

The relative stability of the climate sensitivity from E2 to E2.1 is however due to two counteracting influences. First, as discussed in *Miller et al.* [2020], the effective radiative forcing associated with a doubling of CO_2 is 15% smaller (3.59 compared to 4.19 W m^{-2}) in the E2.1 model than it was in E2 and closer to the canonical 3.7 W m^{-2} [*Myhre et al.*, 2013]. This is mostly explained by higher water vapor content and greater LW cloud forcing which reduce the baseline contribution of CO_2 to longwave opacity, and hence reduce the sensitivity to opacity changes. Secondly, the changes to cloud feedbacks associated with the increase in supercooled cloud water make the overall cloud feedbacks more positive (by reducing the negative cloud phase feedback). Thus the impact to $2\times CO_2$ is only slightly changed, though the normalised sensitivity has increased substantially from 0.62 to 1.00 °C W^{-1} m^2 (using the ECS from $2\times CO_2$), or similarly from 0.58 to 0.87 W^{-1} m^2 (using the long-term response to $4\times CO_2$).

7 Conclusions

As computational resources increase, the temptation at many climate modeling centers is to increase resolution (and therefore compute time) such that the overall throughput of the model stays roughly constant. In contrast to that strategy though, the increment from GISS-E2 to GISS-E2.1 versions focused instead on fixes, better calibrations and in a few cases, improved parameterizations. This was embarked on in parallel with a far more extensive upgrade to the E3 code (including, new topologies, new dynamical cores, higher horizontal and vertical resolution, and new moist physics) which will be reported elsewhere. The question then arises, as to whether the first strategy can provide a worthwhile increase in skill with negligible costs of additional runtime, more efficiently than the second. The answer to that is a definitive yes.

Skill scores in E2.1 are consistently (though not universally) higher in fields that were specifically tuned for as well as in emergent properties (such as the PDO patterns) that were not. Improvements are physically coherent across fields, particularly in the Southern Ocean where the most change has been seen in the ocean, atmosphere and cryosphere.

Nonetheless, we also note the limitations of this approach and the stubborn persistence of long-term biases. Notably, while cloud properties improved, the lack of sufficient marine stratus is still apparent. Similarly, the persistence of a double ITCZ, and excessive symmetry in the zonal mean tropical diagnostics has not been ameliorated to any significant extent. These features have however been almost eliminated in the E3 simulations which have had the benefit of higher resolution, greatly improved moist physics and more comprehensive calibration [*Cesana et al.*, 2019]. It's also apparent that minor retunings are not able to compensate for a model top that is too low for a realistic stratospheric circulation or to generate a quasi-biennial oscillation [*Orbe et al.*, 2020; *Rind et al.*, 2014].

Within the broader constellation of the multi-model ensembles used in CMIP, true structural diversity continues to be a necessary component for any multi-model projection to have a hope of spanning the 'truth' [*Knutti et al.*, 2013]. Better-calibrated lower resolution models and more sophisticated higher resolution models here can play a significant role in expanding that diversity and avoiding the potential danger of similar, and perhaps problematic, new assumptions being adopted by all model groups as they jointly improve such features as cloud and aerosol microphysics [*Gettelman et al.*, 2019; *Andrews et al.*, 2019; *Golaz et al.*, 2019]. The apparent increase in climate sensitivity to doubled CO_2 in some of the next-generation models [*Forster et al.*, 2019; *Dunne et al.*, 2020] whether realistic or not, is very concerning. If this is a reflection of the real world, climate impacts are likely to be greater than we have up to now anticipated, and if it is not, then it raises serious questions about model independence and underlines the importance of true structural diversity. We simply note that the model sensitivity seen in the E2.1 models ($\sim 3°$C) is near the center of the traditionally accepted range of 1.5 to 4°C. While the justification for that range has improved enormously since the Charney report [*Charney et al.*, 1979], the latest assessments do not challenge it [*Sherwood et al.*, 2020].

8 Data and code availability

All data from the piControl, historical, abrupt4xCO2, and 1pctCO2 simulations discussed here are publicly available in the CMIP6 archive through multiple nodes of the Earth System Grid Federation. The code used corresponds to the E2.1 tag in the ModelE git repository available from the NCCS CDS system.

Acknowledgments

We'd like to thank Ángel Adames and John Fasullo for providing figures and data from their multi-model comparisons. Climate modeling at GISS is supported by the NASA Modeling, Analysis and Prediction program, and resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center. The water vapor and precipitation datasets used in this work were obtained from the obs4MIPs `https://esgf-node.llnl.gov/projects/obs4mips` project hosted on the Earth System Grid Federation `https://esgf.llnl.gov`. MSU data are produced by Remote Sensing Systems and sponsored by the NOAA Climate and Global Change Program and are available at `www.remss.com`. SSU data are from NOAA / NESDIS Center for Satellite Applications and Research. Ocean Heat Content data were taken from NOAA NCEI `https://www.nodc.noaa.gov/OC5/3M_HEAT_CONTENT/`.

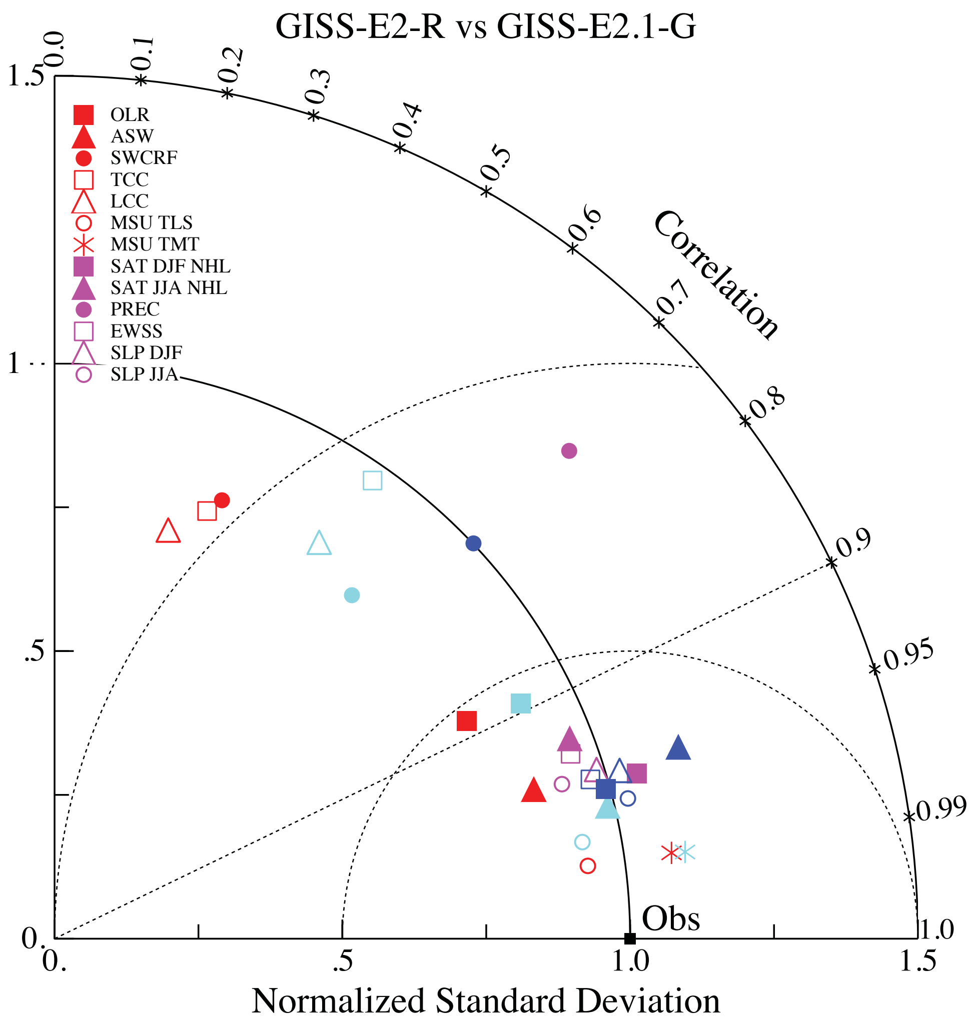


Figure 29. Summary Taylor diagram for selected quantities showing the difference in performance for E2.1-G (light and dark blue symbols) compared to E2-R (red and purple) for different fields. The change in each field can be tracked by going from the red (purple) symbol to the corresponding light blue (dark blue) one. Data sources: CERES EBAF 4d1b: Outgoing Longwave Radiation (OLR) and Absorbed Solar Radiation (ASR) (60°S–60°N); RSS v4 MSU-TMT and MSU-TLS; ISCCP-H Total Cloud Cover, Low Cloud Cover (60°S–60°N), ERA-5 oceanic Sea Level Pressure (SLP) (DJF and JJA), SAT (DJF and JJA) and Eastward Surface Stress (EWSS); TRMM/GPCP Precipitation.

References

Adler, R. F., J.-J. Wang, G. Gu, and G. J. Huffman (2009), A ten-year tropical rainfall climatology based on a composite of TRMM products, *Journal of the Meteorological Society of Japan*, *87A*, 281–293, doi:10.2151/jmsj.87a.281.

Andrews, T., M. B. Andrews, A. Bodas-Salcedo, G. S. Jones, T. Kuhlbrodt, J. Manners, M. B. Menary, J. Ridley, M. A. Ringer, A. A. Sellar, C. A. Senior, and Y. Tang (2019), Forcings, feedbacks, and climate sensitivity in HadGEM3-GC3.1 and UKESM1, *Journal of Advances in Modeling Earth Systems*, doi:10.1029/2019ms001866.

Bauer, S. E., and D. Koch (2005), Impact of heterogeneous sulfate formation at mineral dust surfaces on aerosol loads and radiative forcing in the Goddard Institute for Space Studies general circulation model, *J. Geophys. Res.*, *110*, D17202, doi:10.1029/2005JD005870.

Bauer, S. E., and K. Tsigaridis (2020), The end of the anthropogenic aerosol era?, *Journal of Advances in Modeling Earth Systems*, submitted.

Bauer, S. E., Y. Balkanski, M. Schulz, D. A. Hauglustaine, and F. Dentener (2004), Global modeling of heterogeneous chemistry on mineral aerosol surfaces: Influence on tropospheric ozone chemistry and comparison to observations, *J. Geophys. Res.*, *109*(D2), doi:10.1029/2003JD003868.

Bauer, S. E., D. L. Wright, D. Koch, E. R. Lewis, R. McGraw, L.-S. Chang, S. E. Schwartz, and R. Ruedy (2008), MATRIX (Multiconfiguration Aerosol TRacker of mIXing state): an aerosol microphysical module for global atmospheric models, *Atmos. Chem. Phys.*, *8*, 6003–6035.

Bauer, S. E., A. Bausch, L. Nazarenko, K. Tsigaridis, B. Xu, R. Edwards, M. Bisiaux, and J. McConnell (2013), Historic and future black carbon deposition on the three ice caps: Ice-core measurements and model simulations from 1850 to 2100, *J. Geophys. Res.*, *118*, doi:10.1002/jgrd.50612.

Bellenger, H., E. Guilyardi, J. Leloup, M. Lengaigne, and J. Vialard (2014), ENSO representation in climate models: from CMIP3 to CMIP5, *Climate Dynamics*, *42*(7–8), 1999–2018, doi:10.1007/s00382-013-1783-z.

Bitz, C., and W. Lipscomb (1999), An energy-conserving thermodynamic model of sea ice, *Journal of Geophysical Research*, *104*, 15,669–15,677.

Bowman, K. W., D. T. Shindell, H. M. Worden, J. Lamarque, P. J. Young, D. S. Stevenson, Z. Qu, M. de la Torre, D. Bergmann, P. J. Cameron-Smith, W. J. Collins, R. Doherty, S. B. Dalsøren, G. Faluvegi, G. Folberth, L. W. Horowitz, B. M. Josse1, Y. H. Lee, I. A. MacKenzie, G. Myhre, T. Nagashima, V. Naik, D. A. Plummer, S. T. Rumbold, R. B. Skeie, S. A. Strode1, K. Sudo, S. Szopa, A. Voulgarakis, G. Zeng, S. S. Kulawik, A. M. Aghedo, , and J. R. Worden (2013), Evaluation of ACCMIP outgoing longwave radiation from tropospheric ozone using TES satellite observations, *Atmospheric Chemistry Physics*, *13*, 4057–4072.

Cakmur, R. V., R. L. Miller, J. Perlwitz, I. V. Geogdzhayev, P. Ginoux, D. Koch, K. E. Kohfeld, I. Tegen, and C. S. Zender (2006), Constraining the magnitude of the global dust cycle by minimizing the difference between a model and observations, *J. Geophys. Res.*, *111*, D06207, doi:10.1029/2005JD005791.

Cecil, D. J., D. E. Buechler, and R. J. Blakeslee (2014), Gridded lightning climatology from TRMM-LIS and OTD: Dataset description, *Atmos. Res.*, *135-136*, 404–414, doi:10.1016/j.atmosres.2012.06.028.

Ceppi, P., D. L. Hartmann, and M. J. Webb (2016), Mechanisms of the negative shortwave cloud feedback in middle to high latitudes, *Journal of Climate*, *29*(1), 139–157, doi:10.1175/jcli-d-15-0327.1.

Cesana, G., and D. E. Waliser (2016), Characterizing and understanding systematic biases in the vertical structure of clouds in CMIP5/CFMIP2 models, *Geophysical Research Letters*, *43*(19), 10,538–10,546, doi:10.1002/2016gl070515.

Cesana, G., H. Chepfer, D. Winker, B. Getzewich, X. Cai, O. Jourdan, G. Mioche, H. Okamoto, Y. Hagihara, V. Noel, and M. Reverdy (2016), Using in situ airborne measurements to evaluate three cloud phase products derived from CALIPSO, *Journal of Geo-*

physical Research: Atmospheres, *121*(10), 5788–5808, doi:10.1002/2015jd024334.

Cesana, G., A. D. Del Genio, A. S. Ackerman, M. Kelley, G. Elsaesser, A. M. Fridlind, Y. Cheng, and M.-S. Yao (2019), Evaluating models' response of tropical low clouds to SST forcings using CALIPSO observations, *Atmospheric Chemistry and Physics*, *19*(5), 2813–2832, doi:10.5194/acp-19-2813-2019.

Charney, J., A. Arakawa, D. J. Baker, B. Bolin, R. E. Dickenson, R. M. Goody, C. E. Leith, H. M. Stommel, and C. I. Wunsch (1979), Carbon dioxide and climate:A scientific assessment, *Tech. rep.*, National Academy of Sciences, Woods Hole, MA.

Coddington, O., J. L. Lean, P. Pilewskie, M. Snow, and D. Lindholm (2016), A solar irradiance climate data record, *Bulletin of the American Meteorological Society*, *97*(7), 1265–1282, doi:10.1175/bams-d-14-00265.1.

Cook, B. I., S. P. Shukla, M. J. Puma, and L. S. Nazarenko (2014), Irrigation as an historical climate forcing, *Climate Dynamics*, *44*(5-6), 1715–1730, doi:10.1007/s00382-014-2204-7.

Copernicus Climate Change Service (C3S) (2017), ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate.

Curry, J. A., J. L. Schramm, D. K. Perovich, and J. O. Pinto (2001), Applications of SHEBA/FIRE data to evaluation of snow/ice albedo parameterizations, *J. Geophys. Res.*, *106*, 15,345–15,355, doi:10.1029/2000JD900311.

DeAngelis, A. M., X. Qu, M. D. Zelinka, and A. Hall (2015), An observational radiative constraint on hydrologic cycle intensification, *Nature*, *528*(7581), 249–253, doi:10.1038/nature15770.

Del Genio, A. D., Y. Chen, D. Kim, and M.-S. Yao (2012), The MJO transition from shallow to deep convection in CloudSat/CALIPSO data and GISS GCM simulations, *Journal of Climate*, *25*(11), 3755–3770, doi:10.1175/jcli-d-11-00384.1.

Del Genio, A. D., J. Wu, A. B. Wolf, Y. Chen, M.-S. Yao, and D. Kim (2015), Constraints on cumulus parameterization from simulations of observed MJO events, *Journal of Climate*, *28*(16), 6419–6442, doi:10.1175/jcli-d-14-00832.1.

Dessler, A. E. (1998), A reexamination of the "stratospheric fountain" hypothesis, *Geophys. Res. Lett.*, *25*, 4165–4168.

Donohue, K. A., K. L. Tracey, D. R. Watts, M. P. Chidichimo, and T. K. Chereskin (2016), Mean Antarctic Circumpolar Current transport measured in Drake Passage, *Geophysical Research Letters*, *43*(22), 11,760–11,767, doi:10.1002/2016GL070319.

Dunne, J. P., M. Winton, J. Bacmeister, G. Danabasoglu, A. Gettelman, J.-C. Golaz, C. Hannay, G. A. Schmidt, J. P. Krasting, L. R. Leung, L. Nazarenko, L. A. Sentman, R. J. Stouffer, and J. D. Wolfe (2020), Estimation of model long term equilibrium climate sensitivity from 150-year simulations, *Geophys. Res. Lett.*, submitted.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor (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.

Fekete, B. M., C. J. Vörösmarty, and R. B. Lammers (2001), Scaling gridded river networks for macroscale hydrology: Development, analysis, and control of error, *Water Resources Research*, *37*(7), 1955–1967, doi:10.1029/2001wr900024.

Ferraro, R., D. E. Waliser, P. Gleckler, K. E. Taylor, and V. Eyring (2015), Evolving Obs4MIPs to Support Phase 6 of the Coupled Model Intercomparison Project (CMIP6), *Bulletin of the American Meteorological Society*, *96*(8), ES131–ES133, doi:10.1175/bams-d-14-00216.1.

Fetterer, F., K. Knowles, W. Meier, and M. Savoie (2011), Sea ice index, Boulder, Colorado USA: National Snow and Ice Data Center. Digital media.

Forget, G., and D. Ferreira (2019), Global ocean heat transport dominated by heat export from the tropical pacific, *Nature Geoscience*, *12*(5), 351–354, doi:10.1038/s41561-019-0333-7.

Forget, G., J.-M. Campin, P. Heimbach, C. N. Hill, R. M. Ponte, and C. Wunsch (2015), ECCO version 4: An integrated framework for non-linear inverse modeling and global ocean state estimation, *Geoscientific Model Development*, *8*(10), 3071–3104, doi:10.5194/gmd-8-3071-2015.

Forster, P. M., A. C. Maycock, C. M. McKenna, and C. J. Smith (2019), Latest climate models confirm need for urgent mitigation, *Nature Climate Change*, doi:10.1038/s41558-019-0660-0.

Ganachaud, A., and C. Wunsch (2003), Large-scale ocean heat and freshwater transports during the World Ocean Circulation Experiment, *J. Climate*, *16*, 696–705.

Gao, F., J. T. Morisette, R. E. Wolfe, G. Ederer, J. Pedelty, E. Masuoka, R. Myneni, B. Tan, and J. Nightingale (2008), An algorithm to produce temporally and spatially continuous MODIS-LAI time series, *IEEE Geoscience and Remote Sensing Letters*, *5*(1), 60–64, doi:10.1109/lgrs.2007.907971.

Gettelman, A., C. Hannay, J. T. Bacmeister, R. B. Neale, A. G. Pendergrass, G. Danabasoglu, J.-F. Lamarque, J. T. Fasullo, D. A. Bailey, D. M. Lawrence, and M. J. Mills (2019), High climate sensitivity in the Community Earth System Model version 2 (CESM2), *Geophysical Research Letters*, *46*(14), 8329–8337, doi:10.1029/2019gl083978.

Gleckler, P., R. Ferraro, and D. Waliser (2011), Improving use of satellite data in evaluating climate models, *Eos, Transactions American Geophysical Union*, *92*(20), 172–172, doi:10.1029/2011eo200005.

Golaz, J.-C., P. M. Caldwell, L. P. Van Roekel, M. R. Petersen, Q. Tang, J. D. Wolfe, G. Abeshu, V. Anantharaj, X. S. Asay-Davis, D. C. Bader, et al. (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.

Gregory, J. M. (2010), Long-term effect of volcanic forcing on ocean heat content, *Geophysical Research Letters*, *37*(22), L22701, doi:10.1029/2010gl045507.

Gregory, J. M., W. J. Ingram, M. A. Palmer, G. S. Jones, P. A. Stott, R. B. Thorpe, J. A. Lowe, T. C. Johns, and K. D. Williams (2004), A new method for diagnosing radiative forcing and climate sensitivity, *Geophys. Res. Lett.*, *31*, L03205, doi:10.1029/2003GL018747.

Hansen, J., M. Sato, R. Ruedy, A. Lacis, K. Asamoah, K. Beckford, S. Borenstein, E. Brown, B. Cairns, B. Carlson, B. Curran, S. de Castro, L. Druyan, P. Etwarrow, T. Ferede, M. Fox, D. Gaffen, J. Glascoe, H. Gordon, S. Hollandsworth, X. Jiang, C. Johnson, N. Lawrence, J. Lean, J. Lerner, K. Lo, J. Logan, A. Luckett, M. P. McCormick, R. McPeters, R. Miller, P. Minnis, I. Ramberran, G. Russell, P. Russell, P. Stone, I. Tegen, S. Thomas, L. Thomason, A. Thompson, J. Wilder, R. Willson, and J. Zawodny (1997), Forcings and chaos in interannual to decadal climate change, *Journal of Geophysical Research: Atmospheres*, *102*(D22), 25,679–25,720, doi:10.1029/97jd01495.

Hansen, J., M. Sato, L. Nazarenko, R. Ruedy, A. Lacis, D. Koch, I. Tegen, T. Hall, D. Shindell, B. Santer, P. Stone, T. Novakov, L. Thomason, R. Wang, Y. Wang, D. Jacob, S. Hollandsworth, L. Bishop, J. Logan, A. Thompson, R. Stolarski, J. Lean, R. Willson, S. Levitus, J. Antonov, N. Rayner, D. Parker, and J. Christy (2002), Climate forcings in Goddard Institute for Space Studies SI2000 simulations, *Journal of Geophysical Research*, *107*(D18), doi:10.1029/2001jd001143.

Hansen, J., M. Sato, R. Ruedy, P. Kharecha, A. Lacis, R. L. Miller, L. Nazarenko, K. Lo, G. A. Schmidt, G. Russell, I. Aleinov, S. Bauer, E. Baum, B. Cairns, V. Canuto, M. Chandler, Y. Cheng, A. Cohen, A. D. Genio, G. Faluvegi, E. Fleming, A. Friend, T. Hall, C. Jackman, J. Jonas, M. Kelley, N. Y. Kiang, D. Koch, G. Labow, J. Lerner, S. Menon, T. Novakov, V. Oinas, J. Perlwitz, J. Perlwitz, D. Rind, A. Romanou, R. Schmunk, D. Shindell, P. Stone, S. Sun, D. Streets, N. Tausnev, D. Thresher, N. Unger, M. Yao, and S. Zhang (2007), Climate simulations for 1880-2003 with GISS modelE, *Clim. Dynam.*, *29*, 661–696, doi:10.1007/s00382-007-0255-8.

Hansen, J. E., G. L. Russell, D. Rind, P. Stone, A. Lacis, R. Ruedy, and L. Travis (1983), Efficient three-dimensional models for climatic studies, *Mon. Wea. Rev.*, *111*, 609–662, doi:10.1175/1520-0493(1983)111<0609:ETDGMF>2.0.CO;2.

Hawkins, E., P. Ortega, E. Suckling, A. Schurer, G. Hegerl, P. Jones, M. Joshi, T. J. Osborn, V. Masson-Delmotte, J. Mignot, P. Thorne, and G. J. van Oldenborgh (2017), Estimating changes in global temperature since the pre-industrial period, *Bulletin of the American Meteorological Society*, doi:10.1175/bams-d-16-0007.1.

Hezel, P. J., X. Zhang, C. M. Bitz, B. P. Kelly, and F. Massonnet (2012), Projected decline in spring snow depth on Arctic sea ice caused by progressively later autumn open ocean freeze-up this century, *Geophysical Research Letters*, *39*(17), L17505, doi:10.1029/2012gl052794.

Hoesly, R. M., S. J. Smith, L. Feng, Z. Klimont, G. Janssens-Maenhout, T. Pitkanen, J. J. Seibert, L. Vu, R. J. Andres, R. M. Bolt, T. C. Bond, L. Dawidowski, N. Kholod, J. ichi Kurokawa, M. Li, L. Liu, Z. Lu, M. C. P. Moura, P. R. O'Rourke, and Q. Zhang (2018), Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), *Geoscientific Model Development*, *11*(1), 369–408, doi:10.5194/gmd-11-369-2018.

Holtslag, A. A. M., and B. A. Boville (1993), Local versus nonlocal boundary-layer diffusion in a global climate model, *J. Climate*, *6*, 1825–1842.

Holtslag, A. A. M., and C.-H. Moeng (1991), Eddy diffusivity and countergradient trasport in the convective atmospheric boundary layer, *J. Atmos. Sci.*, *48*, 1690–1698.

Huang, B., P. W. Thorne, V. F. Banzon, T. Boyer, G. Chepurin, J. H. Lawrimore, M. J. Menne, T. M. Smith, R. S. Vose, and H.-M. Zhang (2017), Extended reconstructed sea surface temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons, *Journal of Climate*, *30*(20), 8179–8205, doi:10.1175/jcli-d-16-0836.1.

Huffman, G. J., D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, and E. F. Stocker (2007), The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *Journal of Hydrometeorology*, *8*(1), 38–55, doi:10.1175/jhm560.1.

Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu (2009), Improving the global precipitation record: GPCP version 2.1, *Geophysical Research Letters*, *36*(17), doi:10.1029/2009gl040000.

Hyder, P., J. M. Edwards, R. P. Allan, H. T. Hewitt, T. J. Bracegirdle, J. M. Gregory, R. A. Wood, A. J. S. Meijers, J. Mulcahy, P. Field, K. Furtado, A. Bodas-Salcedo, K. D. Williams, D. Copsey, S. A. Josey, C. Liu, C. D. Roberts, C. Sanchez, J. Ridley, L. Thorpe, S. C. Hardiman, M. Mayer, D. I. Berry, and S. E. Belcher (2018), Critical southern ocean climate model biases traced to atmospheric model cloud errors, *Nature Communications*, *9*(1), doi:10.1038/s41467-018-05634-2.

Iguchi, T., T. Kozu, R. Meneghini, J. Awaka, and K. Okamoto (2000), Rain-profiling algorithm for the TRMM precipitation radar, *Journal of Applied Meteorology*, *39*(12), 2038–2052.

Imawaki, S., H. Uchida, H. Ichikawa, M. Fukasawa, and S. Umatani (2001), Satellite altimeter monitoring the Kuroshio Transport south of Japan, *Geophys. Res. Lett.*, *28*, 17–20, doi:10.1029/2000GL011796.

Jayne, S. R. (2009), The impact of abyssal mixing parameterizations in an ocean general circulation model, *Journal of Physical Oceanography*, *39*(7), 1756–1775, doi:10.1175/2009jpo4085.1.

Johns, W. E., M. O. Baringer, L. M. Beal, S. A. Cunningham, T. Kanzow, H. L. Bryden, J. J. M. Hirschi, J. Marotzke, C. S. Meinen, B. Shaw, and R. Curry (2011), Continuous, array-based estimates of Atlantic Ocean heat transport at 26.5^{o}N, *J. Climate*, *24*, 2429–2449, doi:10.1175/2010JCLI3997.1.

Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor (1999), Surface air temperature and its variations over the last 150 years, *Revs. Goephys.*, *37*, 173–199.

Kattge, J., S. Diaz, S. Lavorel, C. Prentice, P. Leadley, G. Bonisch, E. Garnier, M. Westoby, P. B. Reich, I. J. Wright, J. H. C. Cornelissen, C. Violle, S. P. Harrison, P. M. van Bodegom, M. Reichstein, B. J. Enquist, N. A. Soudzilovskaia, D. D. Ackerly, M. Anand, O. Atkin, M. Bahn, T. R. Baker, D. Baldocchi, R. Bekker, C. C. Blanco, B. Blonder, W. J.

Bond, R. Bradstock, D. E. Bunker, F. Casanoves, J. Cavender-Bares, J. Q. Chambers, F. S. Chapin, J. Chave, D. Coomes, W. K. Cornwell, J. M. Craine, B. H. Dobrin, L. Duarte, W. Durka, J. Elser, G. Esser, M. Estiarte, W. F. Fagan, J. Fang, F. Fernandez-Mendez, A. Fidelis, B. Finegan, O. Flores, H. Ford, D. Frank, G. T. Freschet, N. M. Fyllas, R. V. Gallagher, W. A. Green, A. G. Gutierrez, T. Hickler, S. I. Higgins, J. G. Hodgson, A. Jalili, S. Jansen, C. A. Joly, A. J. Kerkhoff, D. Kirkup, K. Kitajima, M. Kleyer, S. Klotz, J. M. H. Knops, K. Kramer, I. Kuhn, H. Kurokawa, D. Laughlin, T. D. Lee, M. Leishman, F. Lens, T. Lenz, S. L. Lewis, J. Lloyd, J. Llusia, F. Louault, S. Ma, M. D. Mahecha, P. Manning, T. Massad, B. E. Medlyn, J. Messier, A. T. Moles, S. C. Muller, K. Nadrowski, S. Naeem, U. Niinemets, S. Nollert, A. Nuske, R. Ogaya, J. Oleksyn, V. G. Onipchenko, Y. Onoda, J. Ordonez, G. Overbeck, W. A. Ozinga, et al. (2011), TRY - a global database of plant traits, *Global Change Biology*, *17*(9), 2905–2935, doi:10.1111/j.1365-2486.2011.02451.x.

Kiang, N. Y. (2012), Description of the NASA GISS vegetation dynamics model, *Tech. rep.*, NASA.

Kiang, N. Y., C. Montes, I. Aleinov, C. Schaaf, E. Fisher, Q. Sun, D. Carrer, T. Yao, F. Zhao, and Z. Wang (in prep.), The Ent global vegetation structure dataset: Versions 1.x, *Tech. report*, NASA.

Kim, D., A. H. Sobel, A. D. Del Genio, Y.-H. Chen, S. J. Camargo, M. S. Yao, M. Kelley, and L. Nazarenko (2012), The tropical subseasonal variability simulated in the NASA GISS general circulation model, *J. Clim.*, *25*, 4641–4659, doi:10.1175/JCLI-D-11-00447.1.

Kim, Y., P. R. Moorcroft, I. Aleinov, M. Puma, and N. Y. Kiang (2015), Variability of phenology and fluxes of water and carbon with observed and simulated soil moisture in the Ent Terrestrial Biosphere Model (Ent TBM version 1.0.1.0.0), *Geoscientific Model Development*, *8*, 3837–3865, doi:doi:10.5194/gmdd-8-5809-2015.

Knutti, R., D. Masson, and A. Gettelman (2013), Climate model genealogy: Generation CMIP5 and how we got there, *Geophys. Res. Lett.*, *40*, 1194âĂŞ–1199, doi:10.1002/grl.50256.

Koch, D. (2001), Transport and direct radiative forcing of carbonaceous and sulfate aerosols in the GISS GCM, *J. Geophys. Res.*, *106*, 20,311–20,332.

Koch, D., D. Jacob, I. Tegen, D. Rind, and M. Chin (1999), Tropospheric sulfur simulation and sulfate direct radiative forcing in the Goddard Institute for Space Studies general circulation model, *Journal of Geophysical Research: Atmospheres*, *104*(D19), 23,799–23,822, doi:10.1029/1999jd900248.

Koch, D., S. Bauer, A. Del Genio, G. Faluvegi, J. R. McConnell, S. Menon, R. L. Miller, D. Rind, R. Ruedy, G. A. Schmidt, and D. Shindell (2011), Coupled aerosol-chemistry-climate twentieth century transient model investigation: Trends in short-lived species and climate responses, *J. Climate*, *24*, 2693–2714, doi:10.1175/2011JCLI3582.1.

Kok, J. F., D. A. Ridley, Q. Zhou, R. L. Miller, C. Zhao, C. L. Heald, D. S. Ward, S. Albani, and K. Haustein (2017), Smaller desert dust cooling effect estimated from analysis of dust size and abundance, *Nature Geoscience*, *10*, 274–278, doi:doi:10.1038/NGEO2912.

Kopp, G., and J. L. Lean (2011), A new, lower value of total solar irradiance: Evidence and climate significance, *Geophys. Res. Lett.*, *38*, L01706, doi:10.1029/2010GL045777.

Krakauer, N. Y., M. J. Puma, B. I. Cook, P. Gentine, and L. Nazarenko (2016), Ocean–atmosphere interactions modulate irrigation's climate impacts, *Earth System Dynamics*, *7*(4), 863–876, doi:10.5194/esd-7-863-2016.

Lee, Y.-H., and P. J. Adams (2012), A fast and efficient version of the TwO-Moment Aerosol Sectional (TOMAS) global aerosol microphysics model, *Aerosol. Sci. Technol.*, *46*, 678–689, doi:10.1080/02786826.2011.643259.

LeGrande, A. N., K. Tsigaridis, and S. E. Bauer (2016), Role of atmospheric chemistry in the climate impacts of stratospheric volcanic injections, *Nature Geoscience*, *9*(9), 652–655, doi:10.1038/ngeo2771.

Liu, G. (2008), Deriving snow cloud characteristics from CloudSat observations, *J. Geophys. Res.*, *113*, D00A09, doi:10.1029/2007JD009766.

Liu, J., G. A. Schmidt, D. G. Martinson, D. Rind, G. L. Russell, and X. Yuan (2003), Sensitivity of sea ice to physical parameterisations in the GISS global climate model, *J. Geophys. Res.*, *108*, 3053, doi:10.1029/2001JC001167.

Loeb, N. G., F. G. Rose, S. Kato, D. A. Rutan, W. Su, H. Wang, D. R. Doelling, W. L. Smith, and A. Gettelman (2019), Towards a consistent definition between satellite and model clear-sky radiative fluxes, *Journal of Climate*, doi:10.1175/jcli-d-19-0381.1.

Marshall, J., J. R. Scott, A. Romanou, M. Kelley, and A. Leboissetier (2017), The dependence of the ocean's MOC on mesoscale eddy diffusivities: A model study, *Ocean Modelling*, *111*, 1–8, doi:10.1016/j.ocemod.2017.01.001.

Matthews, E. (1983), Global vegetation and land use: New high-resolution data bases for climate studies, *Journal of Climate and Applied Meteorology*, *22*, 474–487.

McCarthy, G., D. Smeed, W. Johns, E. Frajka-Williams, B. Moat, D. Rayner, M. Baringer, C. Meinen, J. Collins, and H. Bryden (2015), Measuring the atlantic meridional overturning circulation at 26°n, *Progress in Oceanography*, *130*, 91–111, doi:10.1016/j.pocean.2014.10.006.

Mears, C. A., and F. J. Wentz (2016), Sensitivity of satellite-derived tropospheric temperature trends to the diurnal cycle adjustment, *Journal of Climate*, *29*(10), 3629–3646, doi:10.1175/jcli-d-15-0744.1.

Mezuman, K., S. E. Bauer, and K. Tsigaridis (2016), Evaluating secondary inorganic aerosols in three dimensions, *Atmospheric Chemistry and Physics*, *16*(16), 10,651–10,669, doi:10.5194/acp-16-10651-2016.

Miller, R. L., R. V. Cakmur, J. Perlwitz, I. V. Geogdzhayev, P. Ginoux, K. E. Kohfeld, D. Koch, C. Prigent, R. Ruedy, G. A. Schmidt, and I. Tegen (2006), Mineral dust aerosols in the NASA Goddard Institute for Space Sciences ModelE atmospheric general circulation model, *J. Geophys. Res.*, *111*, D06208, doi:10.1029/2005JD005796.

Miller, R. L., G. A. Schmidt, L. S. Nazarenko, N. Tausnev, R. Ruedy, M. Kelley, K. K. Lo, I. Aleinov, M. Bauer, S. Bauer, R. Bleck, V. Canuto, Y. Cheng, T. L. Clune, A. Del Genio, G. Faluvegi, J. E. Hansen, R. J. Healy, N. Y. Kiang, D. Koch, A. A. Lacis, A. N. LeGrande, J. Lerner, S. Menon, V. Oinas, J. Perlwitz, M. J. Puma, D. Rind, A. Romanou, G. L. Russell, M. Sato, D. T. Shindell, S. Sun, K. Tsigaridis, N. Unger, A. Voulgarakis, M.-S. Yao, and J. Zhang (2014), CMIP5 historical simulations (1850-2012) with GISS ModelE2, *J. Adv. Model. Earth Syst.*, *6*, 441–477, doi:10.1002/2013MS000266.

Miller, R. L., G. A. Schmidt, L. Nazarenko, S. E. Bauer, M. Kelley, R. Ruedy, G. L. Russell, I. Aleinov, M. Bauer, R. Bleck, V. Canuto, Y. Cheng, T. L. Clune, A. D. Del Genio, G. S. Elsaesser, G. Faluvegi, N. Y. Kiang, D. Kim, A. A. Lacis, A. Leboissetier, A. N. LeGrande, K. K. Lo, J. C. Marshall, E. E. Matthews, K. Mezuman, L. T. Murray, V. Oinas, C. Orbe, C. Pérez García-Pando, J. P. Perlwitz, M. J. Puma, D. Rind, A. Romanou, D. T. Shindell, S. Sun, N. Tausnev, K. Tsigaridis, G. Tselioudis, E. Weng, J. Wu, and M.-S. Yao (2020), CMIP6 historical simulations (1850-2014) with GISS ModelE2.1, *JAMES*, (submitted).

Murray, L. T. (2016), Lightning NO_x and Impacts on Air Quality, *Curr. Pollution. Rep.*, *2*(2), 115–133, doi:10.1007/s40726-016-0031-7.

Myhre, G., D. Shindell, F.-M. Bréon, W. Collins, J. Fuglestvedt, J. Huang, D. Koch, J.-F. Lamarque, D. Lee, B. Mendoza, T. Nakajima, A. Robock, G. Stephens, T. Takemura, and H. Zhang (2013), Anthropogenic and natural radiative forcing, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Myneni, R. B., S. Hoffman, Y. Knyazikhin, J. L. Privette, J. Glassy, Y. Tian, Y. Wang, X. Song, Y. Zhang, G. R. Smith, A. Lotsch, M. Friedl, J. T. Morisette, P. Votava, R. R. Nemani, and S. W. Running (2002), Global products of vegetation leaf area and fraction absorbed par from year one of MODIS data, *Remote Sensing of Environment*, *83*(1-2), 214–231.

Nakanishi, M. (2001), Improvement of the Mellor–Yamada turbulence closure model based on Large-Eddy Simulation data, *Boundary-Layer Meteorology*, *99*(3), 349–378, doi:10.1023/a:1018915827400.

Nazarenko, L., D. Rind, K. Tsigaridis, A. D. D. Genio, M. Kelley, and N. Tausnev (2017), Interactive nature of climate change and aerosol forcing, *Journal of Geophysical Research: Atmospheres*, *122*(6), 3457–3480, doi:10.1002/2016jd025809.

Orbe, C., L. V. Roekel, Á. F. Adames, G. Danabasoglu, A. Dezfuli, J. Fasullo, P. J. Gleckler, J. Lee, W. Li, L. Nazarenko, G. A. Schmidt, K. R. Sperber, and M. Zhao (2020), Representation of modes of variability in 6 U.S. climate models, *J. Climate*, submitted.

Ott, L. E., K. E. Pickering, G. L. Stenchikov, D. J. Allen, A. J. DeCaria, B. A. Ridley, R.-F. Lin, S. Lang, and W.-K. Tao (2010), Production of lightning NO_x and its vertical distribution calculated from three-dimensional cloud-scale chemical transport model simulations, *J. Geophys. Res. Atmos.*, *115*, doi:10.1029/2009JD011880.

Petersen, R. G. (1988), On the transport of the Antarctic Circumpolar Current through Drake Passage and its relation to wind, *J. Geophys. Res.*, *93*, 13,993–14,004, doi:10.1029/JC093iC11p13993.

Phillips, A. S., C. Deser, and J. T. Fasullo (2014), Evaluating modes of variability in climate models, *Eos, Transactions American Geophysical Union*, *95*, 453–455, doi:10.1002/2014EO490002.

Pickering, K. E., Y. Wang, W.-K. Tao, C. G. Price, and J.-F. Müller (1998), Vertical distributions of lightning NO_x for use in regional and global chemical transport models, *J. Geophys. Res.*, *103*(D23), 31203, doi:10.1029/98JD02651.

Pincus, R., E. J. Mlawer, L. Oreopoulos, A. S. Ackerman, S. Baek, M. Brath, S. A. Buehler, K. E. Cady-Pereira, J. N. S. Cole, J.-L. Dufresne, M. Kelley, J. Li, J. Manners, D. J. Paynter, R. Roehrig, M. Sekiguchi, and D. M. Schwarzkopf (2015), Radiative flux and forcing parameterization error in aerosol-free clear skies, *Geophysical Research Letters*, *42*(13), 5485–5492, doi:10.1002/2015gl064291.

Porporato, A., F. Laio, L. Ridolfi, and I. Rodriguez-Iturbe (2001), Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress: III. Vegetation water stress, *Advances in Water Resources*, *24*(7), 725–744, doi:https://doi.org/10.1016/S0309-1708(01)00006-9.

Potter, G. L., L. Carriere, J. Hertz, M. Bosilovich, D. Duffy, T. Lee, and D. N. Williams (2018), Enabling reanalysis research using the collaborative reanalysis technical environment (CREATE), *Bulletin of the American Meteorological Society*, *99*(4), 677–687, doi:10.1175/bams-d-17-0174.1.

Prather, M. J. (1986), Numerical advection by conservation of second order moments, *J. Geophys. Res.*, *91*, 6671–6680.

Prather, M. J., C. D. Holmes, and J. Hsu (2012), Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry, *Geophys. Res. Lett.*, *39*, L09803, doi:10.1029/2012GL051440.

Prather, M. J., X. Zhu, C. M. Flynn, S. A. Strode, J. M. Rodriguez, S. D. Steenrod, J. Liu, J.-F. Lamarque, A. M. Fiore, L. W. Horowitz, J. Mao, L. T. Murray, D. T. Shindell, and S. C. Wofsy (2017), Global atmospheric chemistry – which air matters, *Atmospheric Chemistry and Physics*, *17*(14), 9081–9102, doi:10.5194/acp-17-9081-2017.

Price, C. G., and D. H. Rind (1994), Possible implications of global climate change on global lightning distributions and frequencies, *J. Geophys. Res.*, *99*(D5), 10,823–10,831, doi:10.1029/94JD00019.

Rayner, D. J., J.-M. Hirschi, T. Kanzow, W. E. Johns, S. A. Cunningham, P. G. Wright, E. F.-W. and. H. L. Bryden, C. S. Meinen, M. O. Baringer, J. Marotzke, and L. M. Beal (2011), Monitoring the Atlantic Meridional Overturning Circulation, *Deep Sea Res. I*, *48*, 1744–1753, doi:10.1016/j.dsr2.2010.10.056.

Rind, D., C. Orbe, J. Jonas, L. Nazarenko, T. Zhou, M. Kelley, A. Lacis, D. Shindell, G. Faluvegi, G. Russell, M. Bauer, G. A. Schmidt, A. Romanou, and N. Tausnev (2020), GISS Model E2.2: A climate model optimized for the middle atmosphere. Part 1: Model structure, climatology, variability and climate sensitivity, *J. Geophys. Res. Atmos.*, submitted.

Rind, D. H., J. L. Lean, and J. Jonas (2014), The impact of different absolute solar irradiance values on current climate model simulations, *J. Climate*, *27*(3), 1100–1120, doi:10.1175/jcli-d-13-00136.1.

Rodriguez-Iturbe, I. (2000), Ecohydrology: A hydrologic perspective of climate-soil-vegetation dynamies, *Water Resources Research*, *36*(1), 3–9, doi:10.1029/1999WR900210.

Romanou, A., W. Gregg, J. Romanski, M. Kelley, R. Bleck, R. Healy, L. Nazarenko, G. Russell, G. Schmidt, S. Sun, and N. Tausnev (2013), Natural air–sea flux of CO_2 in simulations of the NASA-GISS climate model: Sensitivity to the physical ocean model formulation, *Ocean Modelling*, *66*, 26–44, doi:10.1016/j.ocemod.2013.01.008.

Romanou, A., J. Romanski, and W. W. Gregg (2014), Natural ocean carbon cycle sensitivity to parameterizations of the recycling in a climate model, *Biogeosciences*, *11*(4), 1137–1154, doi:10.5194/bg-11-1137-2014.

Romanou, A., J. Marshall, M. Kelley, and J. Scott (2017), Role of the ocean's AMOC in setting the uptake efficiency of transient tracers, *Geophysical Research Letters*, *44*(11), 5590–5598, doi:10.1002/2017gl072972.

Rothman, L., I. Gordon, Y. Babikov, A. Barbe, D. C. Benner, P. Bernath, M. Birk, L. Bizzocchi, V. Boudon, L. Brown, A. Campargue, K. Chance, E. Cohen, L. Coudert, V. Devi, B. Drouin, A. Fayt, J.-M. Flaud, R. Gamache, J. Harrison, J.-M. Hartmann, C. Hill, J. Hodges, D. Jacquemart, A. Jolly, J. Lamouroux, R. L. Roy, G. Li, D. Long, O. Lyulin, C. Mackie, S. Massie, S. Mikhailenko, H. Müller, O. Naumenko, A. Nikitin, J. Orphal, V. Perevalov, A. Perrin, E. Polovtseva, C. Richard, M. Smith, E. Starikova, K. Sung, S. Tashkun, J. Tennyson, G. Toon, V. Tyuterev, and G. Wagner (2013), The HITRAN2012 molecular spectroscopic database, *Journal of Quantitative Spectroscopy and Radiative Transfer*, *130*, 4–50, doi:10.1016/j.jqsrt.2013.07.002.

Russell, G. L., J. R. Miller, and D. Rind (1995), A coupled atmosphere-ocean model for transient climate change, *Atmosphere-Ocean*, *33*(4), 683–730.

Sander, S. P., J. Abbatt, J. R. Barker, J. B. Burkholder, R. R. Friedl, D. M. Golden, R. E. Huie, C. E. Kolb, M. J. Kurylo, G. K. Moortgat, V. L. Orkin, and P. H. Wine (2011), Chemical kinetics and photochemical data for use in atmospheric studies, evaluation no. 17.

Šavrič, B., T. Patterson, and B. Jenny (2018), The Equal Earth map projection, *International Journal of Geographical Information Science*, *33*(3), 454–465, doi:10.1080/13658816.2018.1504949.

Schmidt, G. A., C. M. Bitz, U. Mikolajewicz, and L. B. Tremblay (2004), Ice-ocean boundary conditions for coupled models, *Ocean Modelling*, *7*, 59–74.

Schmidt, G. A., R. Ruedy, J. E. Hansen, I. Aleinov, N. Bell, M. Bauer, S. Bauer, B. Cairns, V. Canuto, Y. Cheng, A. Del Genio, G. Faluvegi, A. D. Friend, T. M. Hall, Y. Hu, M. Kelley, N. Y. Kiang, D. Koch, A. A. Lacis, J. Lerner, K. K. Lo, R. L. Miller, L. Nazarenko, V. Oinas, J. Perlwitz, J. Perlwitz, D. Rind, A. Romanou, G. L. Russell, M. Sato, D. T. Shindell, P. H. Stone, S. Sun, N. Tausnev, D. Thresher, and M.-S. Yao (2006), Present day atmospheric simulations using GISS ModelE: Comparison to in-situ, satellite and reanalysis data, *J. Clim.*, *19*, 153–192, doi:10.1175/JCLI3612.1.

Schmidt, G. A., M. Kelley, L. Nazarenko, R. Ruedy, G. L. Russell, I. Aleinov, M. Bauer, S. Bauer, M. K. Bhat, R. Bleck, V. Canuto, Y. Chen, Y. Cheng, T. L. Clune, A. Del Genio, R. de Fainchtein, G. Faluvegi, J. E. Hansen, R. J. Healy, N. Y. Kiang, D. Koch, A. A. Lacis, A. N. LeGrande, J. Lerner, K. K. Lo, E. E. Matthews, S. Menon, R. L. Miller, V. Oinas, A. O. Oloso, J. Perlwitz, M. J. Puma, W. M. Putman, D. Rind, A. Romanou, M. Sato, D. T. Shindell, S. Sun, R. A. Syed, N. Tausnev, K. Tsigaridis, N. Unger, A. Voulgarakis, M.-S. Yao, and J. Zhang (2014), Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, *J. Adv. Model. Earth Syst.*, *6*, 141–184, doi:

10.1002/2013MS000265.

Schmidt, G. A., D. Bader, L. J. Donner, G. S. Elsaesser, J.-C. Golaz, C. Hannay, A. Molod, R. B. Neale, and S. Saha (2017), Practice and philosophy of climate model tuning across six US modeling centers, *Geoscientific Model Development*, *10*(9), 3207–3223, doi:10.5194/gmd-10-3207-2017.

Seltzer, K., D. Shindell, G. Faluvegi, and L. Murray (2017), Evaluating modeled impact metrics for human health, agriculture growth, and near-term climate, *Journal of Geophysical Research: Atmospheres*, *122*(24), 13,506–13,524, doi:10.1002/2017JD026780.

Shah, K. P., and D. Rind (1995), Use of microwave brightness temperatures with a general circulation model, *J. Geophys. Res.*, *100*, 13,841–13,874.

Sherwood, S., M. J. Webb, J. D. Annan, K. C. Armour, P. M. Forster, J. Hargreaves, G. Hegerl, S. A. Klein, K. D. Marvel, E. J. Rohling, M. Watanabe, T. Andrews, P. Braconnot, C. Bretherton, G. L. Foster, Z. Hausfather, A. S. von der Heydt, R. Knutti, T. Mauritsen, J. R. Norris, C. Proistosescu, M. Rugenstein, G. A. Schmidt, K. B. Tokarska, and M. D. Zelinka (2020), A combined assessment of Earth's climate sensitivity, *Revs. Geophys.*, in review.

Shindell, D., G. Faluvegi, K. Seltzer, and C. Shindell (2018), Quantified, localized health benefits of accelerated carbon dioxide emissions reductions, *Nature Climate Change*, *8*(4), 291–295, doi:10.1038/s41558-018-0108-y.

Shindell, D. T., O. Pechony, A. Voulgarakis, G. Faluvegi, L. S. Nazarenko, J.-F. Lamarque, K. Bowman, G. Milly, B. Kovari, R. Ruedy, and G. A. Schmidt (2013a), Interactive ozone and methane chemistry in GISS-E2 historical and future simulations, *Atmos. Chem. Phys.*, pp. 2653–2689, doi:10.5194/acp-13-2653-2013.

Shindell, D. T., O. Pechony, A. Voulgarakis, G. Faluvegi, L. S. Nazarenko, J.-F. Lamarque, K. Bowman, G. Milly, B. Kovari, R. Ruedy, and G. A. Schmidt (2013b), Interactive ozone and methane chemistry in GISS-E2 historical and future simulations, *Atmos. Chem. Phys.*, pp. 2653–2689, doi:10.5194/acp-13-2653-2013.

Simard, M., N. Pinto, J. B. Fisher, and A. Baccini (2011), Mapping forest canopy height globally with spaceborne lidar, *Journal of Geophysical Research-Biogeosciences*, *116*, G04021, doi:10.1029/2011jg001708.

Smeed, D., B. I. Moat, D. Rayner, W. E. Johns, M. O. Baringer, D. L. Volkov, and E. Frajka-Williams (2019), Atlantic meridional overturning circulation observed by the RAPID-MOCHA-WBTS (RAPID-Meridional Overturning Circulation and Heatflux Array-Western Boundary Time Series) array at 26N from 2004 to 2018., doi:10.5285/8CD7E7BB-9A20-05D8-E053-6C86ABC012C2.

Sprintall, J., S. E. Wijffels, R. Molcard, and I. Jaya (2009), Direct estimates of the Indonesian Throughflow entering the Indian Ocean: 2004–2006, *J. Geophys. Res.*, *114*, C07001, doi:10.1029/2008JC005257.

Stachnik, J. P., and C. Schumacher (2011), A comparison of the Hadley circulation in modern reanalyses, *J. Geophys. Res.*, *116*, D22102, doi:10.1029/2011JD016677.

Stenchikov, G., T. L. Delworth, V. Ramaswamy, R. J. Stouffer, A. Wittenberg, and F. Zeng (2009), Volcanic signals in oceans, *Journal of Geophysical Research*, *114*(D16), doi:10.1029/2008jd011673.

Stephens, G. L., J. Li, M. Wild, C. A. Clayson, N. Loeb, S. Kato, T. L'Ecuyer, P. W. Stackhouse, M. Lebsock, and T. Andrews (2012), An update on Earth's energy balance in light of the latest global observations, *Nat. Geosci.*, *5*, 691–696, doi:10.1038/ngeo1580.

Stubenrauch, C. J., W. B. Rossow, S. Kinne, S. Ackerman, G. Cesana, H. Chepfer, L. D. Girolamo, B. Getzewich, A. Guignard, A. Heidinger, B. C. Maddux, W. P. Menzel, P. Minnis, C. Pearl, S. Platnick, C. Poulsen, J. Riedi, S. Sun-Mack, A. Walther, D. Winker, S. Zeng, and G. Zhao (2013), Assessment of global cloud datasets from satellites: Project and database initiated by the GEWEX radiation panel, *Bull. Amer. Meteor. Soc.*, *94*, 1031—1049, doi:10.1175/BAMS-D-12-00117.1.

Sun, S., and R. Bleck (2006), Multi-century simulations with the coupled GISS-HYCOM climate model: Control experiments, *Clim. Dynam.*, *26*, 407–428, doi:10.1007/

s00382-005-0091-7.

Tan, I., T. Storelvmo, and M. D. Zelinka (2016), Observational constraints on mixed-phase clouds imply higher climate sensitivity, *Science*, *352*(6282), 224–227, doi:10.1126/science.aad5300.

Teixeira, J., D. Waliser, R. Ferraro, P. Gleckler, T. Lee, and G. Potter (2014), Satellite observations for CMIP5: The genesis of Obs4MIPs, *Bulletin of the American Meteorological Society*, *95*(9), 1329–1334, doi:10.1175/bams-d-12-00204.1.

Thomason, L. W., N. Ernest, L. Millán, L. Rieger, A. Bourassa, J.-P. Vernier, G. Manney, B. Luo, F. Arfeuille, and T. Peter (2018), A global space-based stratospheric aerosol climatology: 1979–2016, *Earth System Science Data*, *10*(1), 469–492, doi:10.5194/essd-10-469-2018.

Tian, Y., C. Woodcock, Y. Wang, J. Privette, N. Shabanov, L. Zhou, Y. Zhang, W. Buermann, J. Dong, B. Veikkanen, T. Hame, K. Andersson, M. Ozdogan, Y. Knyazikhin, and R. Myneni (2002a), Multiscale analysis and validation of the MODIS LAI product - I. Uncertainty assessment, *Remote Sensing of Environment*, *83*(3), 414–431.

Tian, Y., C. Woodcock, Y. Wang, J. Privette, N. Shabanov, L. Zhou, Y. Zhang, W. Buermann, J. Dong, B. Veikkanen, T. Hame, K. Andersson, M. Ozdogan, Y. Knyazikhin, and R. Myneni (2002b), Multiscale analysis and validation of the MODIS LAI product. II. Sampling strategy. remote sens. environ., 83:431-441., *Remote Sensing of Environment*, *83*(3), 431–441.

Trenberth, K. E., J. T. Fasullo, and J. Kiehl (2009), Earth's global energy budget, *Bull. Amer. Meteor. Soc.*, *90*, 311–324, doi:10.1175/2008BAMS2634.1.

Troen, I. B., and L. Mahrt (1986), A simple model of the atmospheric boundary layer: Sensitivity to surface evaporation, *Boundary-Layer Meteorology*, *37*(1-2), 129–148, doi:10.1007/bf00122760.

van Marle, M. J. E., S. Kloster, B. I. Magi, J. R. Marlon, A.-L. Daniau, R. D. Field, A. Arneth, M. Forrest, S. Hantson, N. M. Kehrwald, W. Knorr, G. Lasslop, F. Li, S. Mangeon, C. Yue, J. W. Kaiser, and G. R. van der Werf (2017), Historic global biomass burning emissions for CMIP6 (BB4cmip) based on merging satellite observations with proxies and fire models (1750–2015), *Geoscientific Model Development*, *10*(9), 3329–3357, doi:10.5194/gmd-10-3329-2017.

Watterson, I. G. (1996), Non-dimensional measures of climate model performance, *Int. J. Climatol.*, *16*, 379–391.

Wentz, F. J., and M. Schabel (2000), Precise climate monitoring using complementary satellite data sets, *Nature*, *403*(6768), 414–416, doi:10.1038/35000184.

Wentz, F. J., L. Ricciardulli, K. Hilburn, and C. Mears (2007), How much more rain will global warming bring?, *Science*, *317*(5835), 233–235, doi:10.1126/science.1140746.

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

Wisser, D., B. M. Fekete, C. J. Vörösmarty, and A. H. Schumann (2010), Reconstructing 20th century global hydrography: a contribution to the Global Terrestrial Network-Hydrology (GTN-h), *Hydrology and Earth System Sciences*, *14*(1), 1–24, doi:10.5194/hess-14-1-2010.

Woodgate, R. A., K. Aagaard, and T. J. Weingartner (2005), Monthly temperature, salinity, and transport variability of the Bering Strait throughflow, *Geophys. Res. Lett.*, *32*, L04601, doi:10.1029/2004GL021880.

Yang, W. Z., B. Tan, D. Huang, M. Rautiainen, N. V. Shabanov, Y. Wang, J. L. Privette, K. F. Huemmrich, R. Fensholt, I. Sandholt, M. Weiss, D. E. Ahl, S. T. Gower, R. R. Nemani, Y. Knyazikhin, and R. B. Myneni (2006), MODIS leaf area index products: From validation to algorithm improvement, *IEEE Transactions on Geoscience and Remote Sensing*, *44*(7), 1885–1898, doi:10.1109/tgrs.2006.871215.

Yao, M.-S., and Y. Cheng (2012), Cloud simulations in response to turbulence parameterizations in the GISS ModelE GCM, *Journal of Climate*, *25*(14), 4963–4974, doi:

10.1175/jcli-d-11-00399.1.

Young, A. H., K. R. Knapp, A. Inamdar, W. Hankins, and W. B. Rossow (2018), The International Satellite Cloud Climatology Project H-Series climate data record product, *Earth System Science Data*, *10*(1), 583–593, doi:10.5194/essd-10-583-2018.

Zeng, X., M. Zhao, and R. E. Dickinson (1998), Intercomparison of bulk aerodynamic algorithms for the computation of sea surface fluxes using TOGA COARE and TAO data, *Journal of Climate*, *11*(10), 2628–2644, doi:10.1175/1520-0442(1998)011<2628:iobaaf>2.0.co;2.

Zou, C.-Z., and H. Qian (2016), Stratospheric temperature climate data record from merged SSU and AMSU-A observations, *Journal of Atmospheric and Oceanic Technology*, *33*(9), 1967–1984, doi:10.1175/jtech-d-16-0018.1.