CMIP6 Historical Simulations (1850–2014) With GISS-E2.1

Key Points:
- Tropospheric warming and ocean heat uptake by 2014 are smaller in GISS-E2.1 and closer to observed trends than in its CMIP5 predecessor
- GISS-E2.1 climate sensitivity is higher than in CMIP5 GISS-E2, but forcing by greenhouse gases is smaller
- Atmospheric trends vary among model configurations with the storage of heat beneath the thermocline

Supporting Information:
- Supporting Information S1

Correspondence to:
R. L. Miller,
ron.l.miller@nasa.gov

Ron L. Miller[1,2] , Gavin A. Schmidt[1] , Larissa S. Nazarenko[1,3] , Susanne E. Bauer[1] , Maxwell Kelley[1,4], Reto Ruedy[1,4] , Gary L. Russell[1] , Andrew S. Ackerman[1] , Igor Aleinov[1,3] , Michael Bauer[1,3], Rainer Bleck[5,6] , Vittorio Canuto[1] , Grégory Cesana[1,3] , Ye Cheng[1,3], Thomas L. Clune[7] , Ben I. Cook[1], Carlos A. Cruz[7,8], Anthony D. Del Genio[1] , Gregory S. Elsaesser[1,2] , Greg Faluvegi[1,3] , Nancy Y. Kiang[1] , Daehyun Kim[9], Andrew A. Lacis[1] , Anthony Leboissetier[1,4], Allegra N. LeGrande[1] , Ken K. Lo[1,4], John Marshall[10] , Elaine E. Matthews[1], Sonali McDermid[11] , Keren Mezuman[1,3], Lee T. Murray[12] , Valdar Oinas[1,4], Clara Orbe[1] , Carlos Pérez García-Pando[13,14] , Jan P. Perlwitz[1,15], Michael J. Puma[1,3], David Rind[1] , Anastasia Romanou[1], Drew T. Shindell[16] , Shan Sun[6], Nick Tausnev[1,4], Kostas Tsigaridis[1,3] , George Tselioudis[1] , Ensheng Weng[1,3] , Jingbo Wu[1,2], and Mao-Sung Yao[1,4]

[1]NASA Goddard Institute for Space Studies, New York, NY, USA, [2]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA, [3]Center for Climate Systems Research, Earth Institute, Columbia University, New York, NY, USA, [4]SciSpace LLC, New York, NY, USA, [5]CIRES, University of Colorado Boulder, Boulder, CO, USA, [6]NOAA/ESRL/Global Systems Laboratory, Boulder, CO, USA, [7]Goddard Space Flight Center, Greenbelt, MD, USA, [8]SSAI, Greenbelt, MD, USA, [9]Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA, [10]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge MA, USA, [11]Department of Environmental Studies, New York University, New York, NY, USA, [12]Department of Earth and Environmental Sciences, University of Rochester, Rochester, NY, USA, [13]Barcelona Supercomputing Center, Barcelona, Spain, [14]ICREA, Catalan Institution for Research and Advanced Studies, Barcelona, Spain, [15]Climate, Aerosol, and Pollution Research, LLC, Bronx, NY, USA, [16]Nicholas School of the Environment, Duke University, Durham, NC, USA

Abstract Simulations of the CMIP6 historical period 1850–2014, characterized by the emergence of anthropogenic climate drivers like greenhouse gases, are presented for different configurations of the NASA Goddard Institute for Space Studies (GISS) Earth System ModelE2.1. The GISS-E2.1 ensembles are more sensitive to greenhouse gas forcing than their CMIP5 predecessors (GISS-E2) but warm less during recent decades due to a forcing reduction that is attributed to greater longwave opacity in the GISS-E2.1 pre-industrial simulations. This results in an atmosphere less sensitive to increases in opacity from rising greenhouse gas concentrations, demonstrating the importance of the base climatology to forcing and forced climate trends. Most model versions match observed temperature trends since 1979 from the ocean to the stratosphere. The choice of ocean model is important to the transient climate response, as found previously in CMIP5 GISS-E2: the model that more efficiently exports heat to the deep ocean shows a smaller rise in tropospheric temperature. Model sea level rise over the historical period is traced to excessive drawdown of aquifers to meet irrigation demand with a smaller contribution from thermal expansion. This shows how fully coupled models can provide indirect observational constraints upon forcing, in this case, constraining irrigation rates with observed sea level changes. The overall agreement of GISS-E2.1 with observed trends is familiar from evaluation of its predecessors, as is the conclusion that these trends are almost entirely anthropogenic in origin.

Plain Language Summary Measurements show clear evidence of warming over the twentieth century and up to the present day. Our anticipation of future change comes from computer models of climate. These are based upon well-established physical principles like Newton's laws of motion and radiative transfer theory; the models are closely related to those used for weather forecasting. We can never predict the weather on a particular day, 50 years in the future, but we can calculate whether that future decade will be warmer than our present climate. Part of our confidence in such a forecast comes from testing a climate model's ability to reproduce warming and other changes measured over the past century. We use observations of atmospheric composition and the sunlight received by our planet to calculate how the model responds to their changes. The climate model of the NASA Goddard Institute for Space Studies,

GISS-E2.1, closely follows changes measured in the ocean and atmosphere as the concentrations of greenhouse gases and other pollutants rise. This agreement suggests that future warming by greenhouse gases will be reliably predicted by GISS-E2.1. This suggests that the warming we already experience is due to our consumption of fossil fuels that has led to the increase of carbon dioxide and other greenhouse gases in the atmosphere over the past two centuries.

1. Introduction

In an era of climate change and the frequent occurrence of record-setting measurements, there is intense interest in anticipating future changes. The Coupled Model Intercomparison Project (CMIP) was conceived to facilitate coordinated experiments of past and future climate among modeling groups (Meehl et al., 1997). Contrasting behavior between climate simulations would be attributed to model differences rather than differences in experimental design. The coordinated experiments allow models from different research centers to be meaningfully compared to each other and observations, while providing benchmarks for the comparison of successive generations of models.

A key element of CMIP Phase 6 (CMIP6 Eyring et al., 2016) is the simulation of climate change observed during the so-called historical period from 1850 to 2014. The comparison of observed changes to simulated trends driven by changes in atmospheric composition and other climate drivers is an incomplete but nonetheless useful test of model climate sensitivity that is important for assessing the likelihood of future changes projected by the models. Historical simulations have been performed with multiple generations of the NASA Goddard Institute for Space Studies (GISS) Earth system model (Hansen, Sato, Ruedy, Lacis, et al., 1997; Hansen et al., 2002, 2007, 1988; Miller et al., 2014). Over previous CMIP generations, GISS has assessed robustness of the modeled response by simulating the historical period with different combinations of model components. Since CMIP3, we have carried out experiments with two ocean versions coupled to the same atmospheric model, and for CMIP5, GISS analyzed the impact of different treatments of atmospheric composition and the aerosol indirect effect. For CMIP6, we continue to use multiple versions of the coupled model. The use of a common atmosphere allows the effect of ocean treatments to be highlighted, while different treatments of atmospheric composition test robustness of certain interactions and forcings.

Characterization of forcing is the basis for understanding the contribution of individual drivers to future changes as well as attributing past and currently observed changes to the forcings (e.g., Hansen et al., 2005; Marvel et al., 2015). For each CMIP generation, the forcings are better characterized, the number of forcings recognized as important to observed trends is increased, and the representation of physical processes impacted by the forcings is expanded. Interactions between expanded physical representations within the model can reveal limitations and the need for refinement that were obscure prior to their coupling.

The GISS contribution to CMIP6 includes varying ocean components, composition interactivity, vertical and horizontal resolution, carbon cycle, and forcings. We report here on a subset of this contribution based upon GISS-E2.1, an updated and more skillful version of the GISS-E2 model used in CMIP5 (Miller et al., 2014; Nazarenko et al., 2015; Schmidt et al., 2014). The seasonal climatology of GISS-E2.1 during the satellite period (1979–2014) is evaluated elsewhere by Kelley et al. (2020). The present article is a complementary analysis of climate trends during the CMIP6 historical period (1850–2014) simulated by GISS-E2.1 for comparison to observed changes.

In section 2, we summarize the model versions and describe the format of the historical experiments. In section 3, we document forcings that perturb the model's pre-industrial climate, noting differences with respect to the GISS CMIP5 simulations that impact the climate response. In section 4, we evaluate model climate trends as coherently expressed across disparate climate variables and the trend dependence upon the forcing along with the atmosphere and ocean components. In section 5, we present a summary of results and implications.

2. Models and Experiments

A full description of GISS-E2.1, including development since CMIP5, is given by Kelley et al. (2020). Here we summarize the main features of the atmosphere and ocean components, while describing how the historical ensembles arise from the pre-industrial experiments.
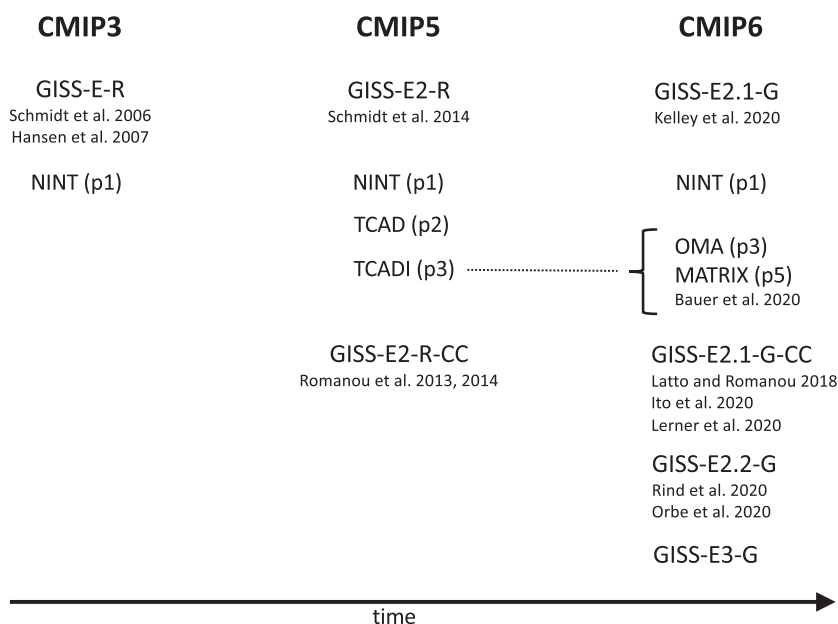
| CMIP3 | CMIP5 | CMIP6 |

GISS-E-R
Schmidt et al. 2006
Hansen et al. 2007

GISS-E2-R
Schmidt et al. 2014

GISS-E2.1-G
Kelley et al. 2020

NINT (p1)　　　NINT (p1)　　　NINT (p1)

TCAD (p2)

TCADI (p3)　················　OMA (p3)
MATRIX (p5)
Bauer et al. 2020

GISS-E2-R-CC
Romanou et al. 2013, 2014

GISS-E2.1-G-CC
Latto and Romanou 2018
Ito et al. 2020
Lerner et al. 2020

GISS-E2.2-G
Rind et al. 2020
Orbe et al. 2020

GISS-E3-G

time

Figure 1. Development of ModelE coupled to the GISS Ocean Model v.1 (née the Russell Ocean Model) from CMIP3 to the present. There is parallel development of ModelE coupled to HYCOM (Sun & Bleck, 2006), where "R" and "G" are replaced by "H." The suffix "CC" refers to an interactive carbon cycle (Ito et al., 2020; Latto & Romanou, 2018; Lerner et al., 2020; Romanou et al., 2013, 2014). The domain of GISS-E2.2 includes the mesosphere and has a better stratospheric circulation (Orbe et al., 2020; Rind et al., 2020), while GISS-E3 incorporates a new cubed-sphere grid and improved model physics.

2.1. Atmospheric Model

The horizontal and vertical resolution of the atmospheric component of GISS-E2.1 is unchanged since CMIP5: 2° latitude by 2.5° longitude with 40 vertical layers extending from the surface to 0.1 hPa in the lower mesosphere. Nonetheless, GISS-E2.1 shows substantial improvement in its seasonal climatology, along with variability like the Madden-Julian Oscillation (Del Genio et al., 2015; Kim et al., 2012) and teleconnection patterns associated with El Niño and the Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (Kelley et al., 2020).

Three versions of the GISS-E2.1 atmospheric component have been submitted to CMIP6. Atmospheric composition is prescribed in the first version, denoted here as NINT (for "non-interactive") and in the CMIP6 archive as `physics-version=1` ("p1"). In two other versions, aerosols and ozone are calculated prognostically using either the One-Moment Aerosol (OMA) model or else the MATRIX model ("Multiconfiguration Aerosol TRacker of mIXing state" Bauer et al., 2008, 2020). The OMA version is designated as `physics-version=3` ("p3") in the CMIP6 archive to signify continuity with its CMIP5 predecessor, the TCADI model from which OMA was derived. (The relation of different CMIP generations of ModelE is summarized in Figure 1) The MATRIX version of GISS-E2.1 is designated by `physics-version=5` ("p5") in the CMIP6 archive.

Aerosols and ozone in the NINT model are prescribed from "AMIP-style" OMA simulations of the historical period with forcings described in section 3 (Bauer et al., 2020). Sea surface temperature (SST) and sea ice fraction are available from Rayner (2003) after 1870. Prior to that, these variables are specified using an average from the decade 1876–1885, a period chosen for its spatial coverage of measurements prior to significant warming. The use of the OMA model to prescribe NINT composition gives greater consistency between the two model versions. For example, ozone variations are in phase with the solar forcing in all the GISS CMIP6 ensembles. This consistency allows differences in behavior between NINT and OMA to be more directly attributed to feedbacks between composition and climate variations. In addition, the OMA version of GISS-E2.1 provides atmospheric composition for NINT based upon a state-of-the-art model, in contrast to the CMIP5 NINT version, where some constituents like ozone were calculated with the previous-generation (CMIP3) version of ModelE. OMA aerosol species provided to NINT include sulfate, nitrate, ammonium, organic carbon and three types of black carbon (BC). These species are primarily anthropogenic in origin,

although there are non-anthropogenic sources from biomass burning and lightning. OMA also provides NINT with sea salt and dust aerosols that vary temporally through climate variables like wind speed. Anthropogenic dust sources are not represented in OMA, even though high dust concentration is observed in regions of cultivation (e.g., Ginoux et al., 2012; Mahowald et al., 2004; Tegen et al., 1996). Nonetheless, some anthropogenic trends in dust concentration can be driven indirectly by irrigation that increases soil moisture (Bauer et al., 2020), as well as by the formation of sulfate and nitrate coatings on the dust particle surface that increases the rate of wet removal (Bauer & Koch, 2005; Bauer et al., 2004).

Atmospheric composition from AMIP-style OMA and MATRIX experiments is described by Bauer et al. (2020). Aerosol concentrations are determined by simulating an extensive series of chemical reactions in the atmosphere, which includes prognostic budgets for other radiatively active constituents like methane, nitrous oxide and chlorofluorocarbons (CFCs). The surface concentration of each of these gases is prescribed during the historical period using measurements (Meinshausen et al., 2017), with their vertical and spatial distributions determined by a suite of physical and chemical processes. Photochemical oxidation of methane is a source of stratospheric water vapor that alters the column longwave opacity. Ozone in the troposphere and stratosphere is calculated prognostically based upon the prescribed concentration of constituents like halogens or emission of tropospheric pollutants (Hoesly et al., 2018; Meinshausen et al., 2017), photochemistry and heterogeneous reactions on aerosols and polar stratospheric cloud droplets. Photolysis reactions associated with ozone are modulated by variations in the solar photon flux as described in section 3.4 (Matthes et al., 2017), although at present this flux does not vary with the changing concentration of radiatively active constituents like volcanic aerosols or ozone itself.

Aerosol extinction is linearly related to aerosol mass. Solar extinction by BC is increased by a factor of 1.5 relative to its mass to represent enhanced absorption by sulfate and nitrate coatings that increase internal reflection (Bauer et al., 2007), while dust LW extinction is increased by a factor of 1.3 to represent scattering (Dufresne et al., 2002; Schmidt et al., 2006).

Along with their direct radiative effect, aerosols perturb climate through microphysical adjustments of cloud properties and the associated cloud radiative fluxes: the aerosol indirect effect (AIE; e.g., Boucher et al., 2013). In the NINT model, cloud fraction is increased according to the number concentration of sulfate, nitrate and carbonaceous aerosols above their pre-industrial levels: a heuristic representation of changes to cloud radiative forcing by aerosol nucleation of cloud droplets (Hansen et al., 2005). The relation between cloud fraction and aerosol number is logarithmic to represent the declining effect of adding anthropogenic aerosols to highly polluted air where nucleation sites are already abundant. This heuristic representation of the AIE within the NINT model is tuned to give roughly $-1\,\mathrm{W\,m^{-2}}$ of forcing over the historical period, consistent with more physically sophisticated models constrained by observations (Boucher et al., 2013).

The NINT AIE directly relates anthropogenic aerosol mass to a change in cloud cover. Radiative properties of observed clouds depend upon cloud droplet number (CDN), which depends upon the concentration of both anthropogenic and natural aerosols, but additionally upon droplet size and their distribution with height within the cloud. The OMA and MATRIX models incorporate this more comprehensive view by relating aerosol mass only to the CDN at cloud base, and simulating cloud microphysical processes that convert vapor to liquid water to determine droplet size and the cloud optical depth. CDN is derived from aerosol number and updraft speed using empirical relations for convective clouds (Menon & Rotstayn, 2006, their Figure 1) and for stratiform clouds (Lohmann et al., 2007, their Equation 1). These relations are derived from case studies of observed cloud properties simulated by more physically complex microphysical models (e.g., Segal et al., 2004). In MATRIX, aerosols serve as droplet nucleation sites depending upon their mass and number along with the cloud updraft velocity following Abdul-Razzak et al. (1998) and Abdul-Razzak and Ghan (2000).

The OMA and MATRIX versions of GISS-E2.1 represent only the first indirect effect relating aerosol number to cloud optical thickness (Twomey, 1977). Development of the second indirect effect relating the influence of aerosols upon the growth of cloud droplets to precipitable size (Albrecht, 1989) is underway in GISS-E3. In contrast to the NINT model, the OMA and MATRIX calculations of cloud radiative effect are not adjusted to give a specific indirect forcing.

For the NINT pre-industrial control simulations (denoted as "`piControl`" in the CMIP6 archive), aerosol and ozone concentrations vary with calendar month based upon a decadal average of the OMA AMIP

Table 1
Model Configurations Used for piControl Experiments With Their CMIP6 Identifier and Digital Object Identifier (doi) Along With the First and Last Years and Duration of the piControl Segment Used to Define Anomalies During the Historical Period

Model	CMIP archive tag	doi	First year	Last year	Duration
GISS-E2.1-G NINT f1	piControl_r1i1p1f1	10.22033/ESGF/CMIP6.7380	4150	5000	851
GISS-E2.1-G NINT f2	piControl_r1i1p1f2	10.22033/ESGF/CMIP6.7380	7550	8000	451
GISS-E2.1-G OMA	piControl_r1i1p3f1	10.22033/ESGF/CMIP6.7380	2000	2450	451
GISS-E2.1-H NINT f1	piControl_r1i1p1f1	10.22033/ESGF/CMIP6.7381	3180	3980	801
GISS-E2.1-H NINT f2	piControl_r1i1p1f2	10.22033/ESGF/CMIP6.7381	4000	4450	451
GISS-E2.1-H OMA	piControl_r1i1p3f1	10.22033/ESGF/CMIP6.7381	2000	2450	451

Note. Aerosols and ozone in the p1 (NINT) "f1" and "f2" simulations are taken from a four-member ensemble average of a prior OMA AMIP model version and a five-member ensemble average of the corrected OMA AMIP model, respectively. The doi of the latter ensemble is 10.22033/ESGF/CMIP6.6984, and its output in the CMIP6 archive is identified by `amip_r[1-5]i1p3f1`. The OMA experiments identified by "f1" in the CMIP6 archive use the corrected model. The first year of each piControl experiment is unrelated to the length of its coupled spin-up.

simulation between 1850 and 1859. For NINT experiments during the CMIP6 historical period, the prescribed concentration is updated using OMA AMIP values that vary by month and year.

2.2. Ocean Models

Each of the three physics versions of the GISS-E2.1 atmospheric component is coupled to one of two ocean general circulation models (OGCMs).

The coupled model denoted "`GISS-E2-1-G`" in the CMIP6 archive (and "E2.1-G" here) uses the GISS Ocean version 1 (GO1), a descendent of the Russell OGCM used in CMIP5 (whose coupled version was denoted "`GISS-E2-R`" or else "E2-R"). Both GO1 and the Russell OGCM have horizontal resolution of 1° latitude by 1.25° longitude, so that four ocean grid boxes nest within a single AGCM grid box. GO1 has 40 vertical layers, 8 more than its CMIP5 predecessor. The additional layers improve vertical resolution in the upper ocean, including the thermocline. GO1 has 24 layers above 1,400 m depth compared to 15 in the CMIP5 version. Above 200 m depth, GO1 has 11 layers compared to just 6 in its predecessor. This approximately doubles vertical resolution in shallow, marginal seas where GO1 shows reduced salinity biases. The strongest mixing is now directed along isopycnals (Redi, 1982), eliminating unintended cross-isopycnal diffusion present in the GISS CMIP5 simulations.

The HYCOM model calculates the ocean circulation on 32 isopycnal layers, an increase of 6 over its CMIP5 precursor. These are replaced by height coordinates in regions of weak stratification. Meridional grid spacing is 1° at the equator and decreases poleward with the cosine of latitude. Zonal resolution is a uniform 1° outside of the Northern Hemisphere (NH) high latitudes, where a separate bi-polar grid eliminates the singularity of spherical coordinates. Because of the varying horizontal resolution associated with the NH polar grid and varying meridional spacing elsewhere, output from the coupled HYCOM model ("`GISS-E2-1-H`" or "E2.1-H") is archived on a uniform 1° by 1° grid at 32 depths extending from the surface to 5500 m.

The combination of the three physics versions of the GISS-E2.1 AGCM and the two ocean models results in six distinct coupled models that have been submitted to CMIP6. At the time of writing, less model output was available for the MATRIX coupled models, so we present results only for the NINT and OMA ensembles.

2.3. Spin-Up of Pre-Industrial Control Runs and Branching

Each historical simulation departs from a control simulation that is near equilibrium with respect to pre-industrial atmospheric composition.

The pre-industrial control runs resulting from the different combinations of atmospheric model physics versions and ocean models are listed in Table 1.

During integration of the NINT historical experiments, we identified a coding error in the OMA model that caused an underestimate of stratospheric ozone destruction by heterogeneous reactions on the surface of volcanic aerosols. (The surface area of volcanic aerosols that host these reactions was too small.) This error had a small effect upon the pre-industrial climate, but restricted ozone destruction by halogens following
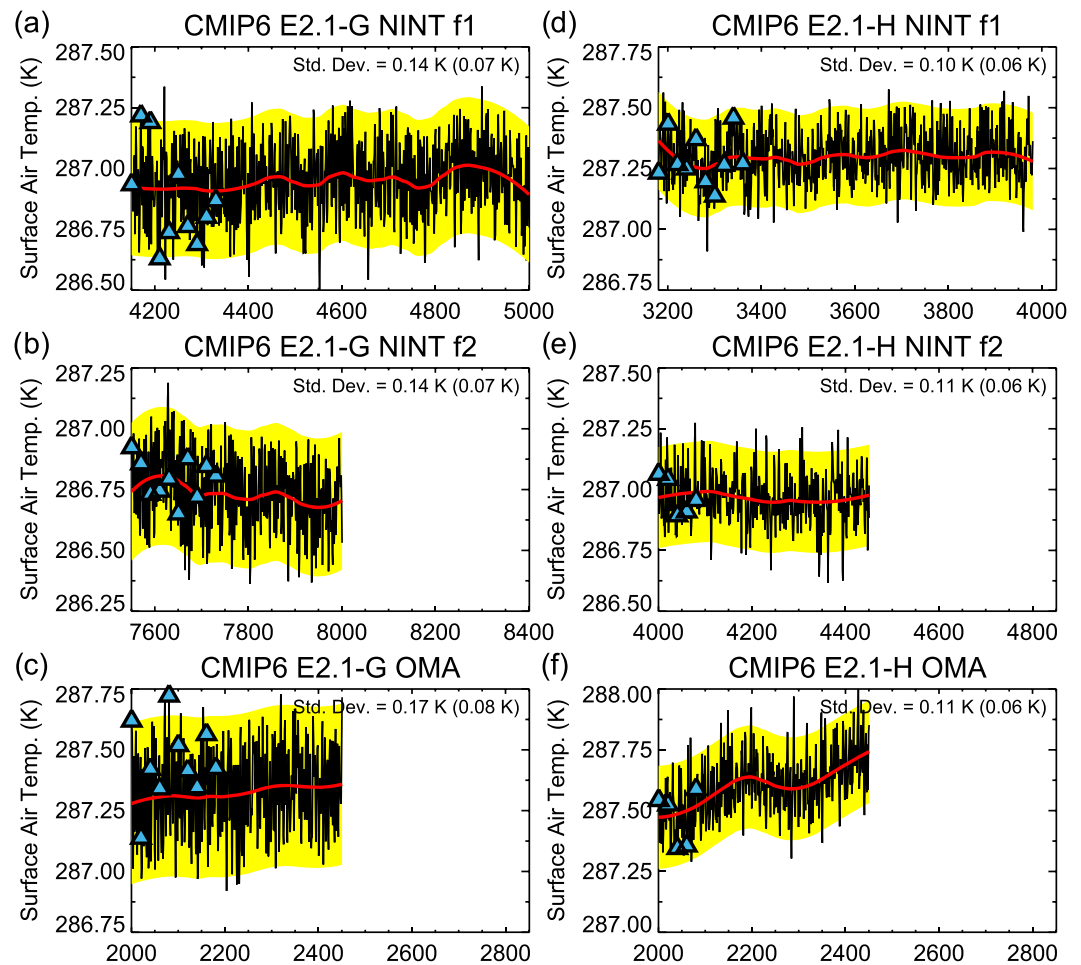
Figure 2. Global and annual surface air temperature (black) during the piControl simulation. In each panel, the red line is calculated using loess smoothing over a 200-year window. The standard deviation with respect to the loess fit is shown in the upper right corner of each panel; the value in parentheses is the standard deviation of the corresponding GISS CMIP5 model. The yellow bar marks two standard deviations above and below the smoothed temperature. The blue triangles show the year of branching for each historical ensemble member.

a volcanic eruption (Muthers et al., 2015; Solomon, 1999), leading to an ozone increase following Pinatubo in contradiction to observations (Hofmann & Oltmans, 1993). The AMIP OMA simulations were rerun to generate new ozone and aerosol inputs for a second set of NINT pre-industrial control and historical runs. The NINT model output with the original and corrected composition is distinguished by "f1" and "f2," respectively, in the CMIP6 `ripf` archive code. The atmospheric component of GISS-E2.1 used in these two NINT simulations is identical; all that varies is the prescribed ozone and aerosol composition. In this article, we emphasize the results of the NINT f2 and OMA f1 historical ensembles, both of which are based upon the corrected model. In the troposphere, NINT f1 and f2 variables that we have examined are statistically indistinguishable, so we show the NINT f1 results for stratospheric variables, mostly deferring tropospheric diagnostics of this configuration to the Supporting Information.

The E2.1-G NINT f1 pre-industrial ocean was started from rest using a present-day climatology of temperature and salinity (Levitus & Boyer, 1994; Levitus, Burgett, et al., 1994), and run for two millennia with pre-industrial atmospheric composition before a small correction was made to the sea ice calculation. The model was integrated for another 200 years before a small fix to lake ice was applied. After an additional 50 years, net radiation at top of atmosphere (TOA) was roughly 0.1 W m^{-2} (with an interannual standard deviation of 0.35 W m^{-2}). This was considered close enough to equilibrium to start the piControl simulation and branching of the first historical ensemble member. After nearly three millennia of additional integration, the NINT f1 piControl was used to initialize the hundred-year spin-up of the corresponding (E2.1-G)
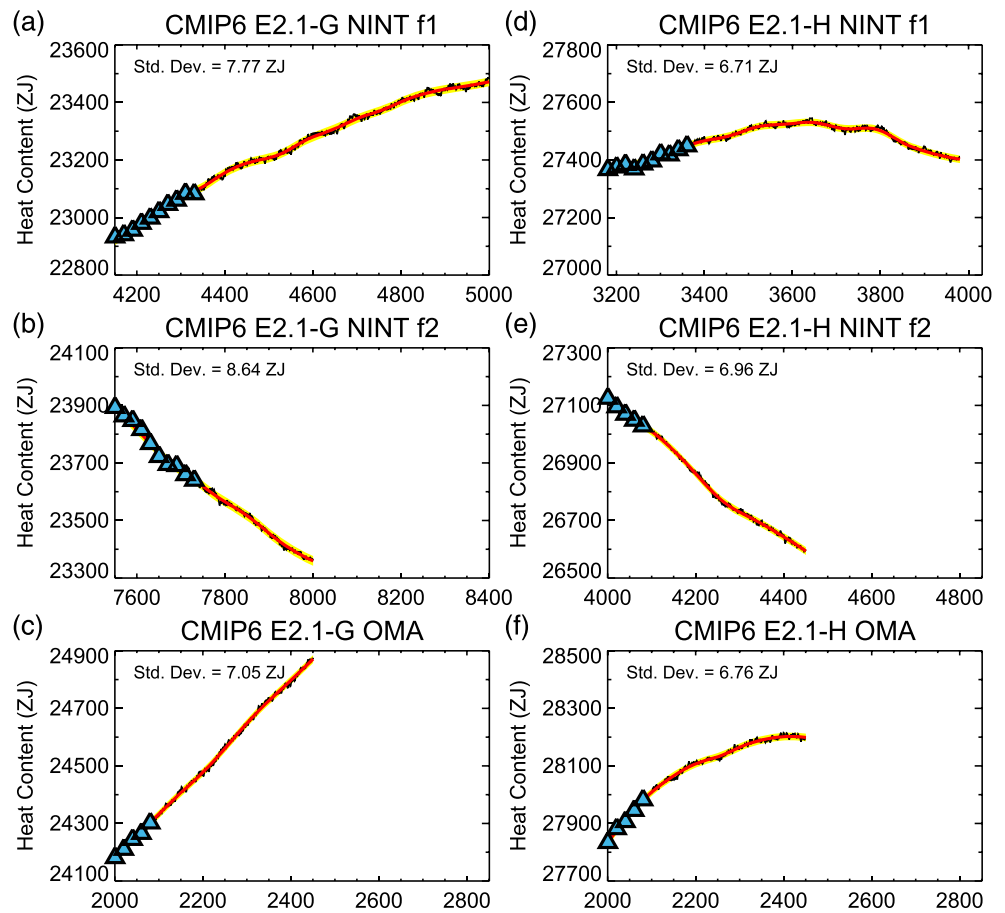
Figure 3. Same as Figure 2 but for ocean heat content (ZJ; 1 ZJ equals 10^{21} J). Heat content is calculated relative to a reference temperature of 273.15 K. The span of both axes is the same in each panel (850 years and 800 ZJ).

OMA model, after which the OMA piControl and historical ensembles were started. The E2.1-G NINT f2 model branched from the f1 piControl 500 years after the OMA model; after 50 years of spin-up, the f2 piControl and historical ensemble were initialized. The E2.1-H NINT and OMA coupled configurations followed a similar path toward equilibrium and initialization of the piControl and historical experiments.

The global and annual-mean surface air temperature for the NINT and OMA piControl experiments is shown in black in Figure 2, along with low-frequency variations in red that are identified using a loess fit (described below). For the NINT experiments, temperature variations are on the order of a tenth of a degree K, with higher variability in the E2.1-G coupled model than in E2.1-H. Unforced variability of surface air temperature in each piControl experiment is larger than in its CMIP5 predecessor. (The standard deviation about the loess fit is indicated in the upper right corner of each panel of Figure 2, with the CMIP5 GISS-E2 value in parentheses.) This contrast is associated with stronger ENSO variations in E2.1-G (Kelley et al., 2020). In general, surface trends are small for each model in Figure 2. Sub-surface trends are more apparent. Figure 3 shows drift in ocean heat content, consistent with the small radiation imbalance at TOA. For the NINT f2 model, the ocean heat content decreases by 550 ZJ over 450 years. This change is comparable in magnitude to an estimate of uptake since the pre-industrial (Zanna et al., 2019), suggesting that the model deep ocean is still adjusting toward equilibrium (Figure 3b; 1 ZJ equals 10^{21} J).

The first member of each historical ensemble branches at the start of its pre-industrial control run (Table 1). The remaining members are initialized every 20 years, a duration long enough so that the mid-latitude gyre circulations and tropical ENSO are decorrelated and independent between ensemble members (Blumenthal, 1991). (The dates of branching are shown by the blue triangles in Figure 2) To offset the larger unforced variability in E2.1-G compared to its CMIP5 E2-R predecessor, which can obscure forced variations, we chose

Table 2
Historical Ensembles

Model	CMIP archive tag	doi	Branch year	Ensemble size
GISS-E2.1-G NINT f1	historical_r1i1p1f1	10.22033/ESGF/CMIP6.7127	4150	10
GISS-E2.1-G NINT f2	historical_r1i1p1f2	10.22033/ESGF/CMIP6.7127	7550	10
GISS-E2.1-G OMA	historical_r1i1p3f1	10.22033/ESGF/CMIP6.7127	2000	10
GISS-E2.1-H NINT f1	historical_r1i1p1f1	10.22033/ESGF/CMIP6.7128	3180	10
GISS-E2.1-H NINT f2	historical_r1i1p1f2	10.22033/ESGF/CMIP6.7128	4000	5
GISS-E2.1-H OMA	historical_r1i1p3f1	10.22033/ESGF/CMIP6.7128	2000	5

Note. Each ensemble member begins in 1850 and runs through 2014. The branch year is the piControl date corresponding to 1850 of the first member of the historical ensemble. Subsequent ensemble members branch every 20 years.

a ten-member ensemble size, compared to five for E2-R and (CMIP6) E2.1-H, whose piControl variability is smaller.

Table 2 lists the historical ensembles including the branching date in the piControl simulation that corresponds to the start of the first member of each historical ensemble.

2.4. Definition of Anomalies During the Historical Period

The historical ensemble members are perturbed from the pre-industrial control experiment by the forcings described in section 3. However, each member initially retains some drift and low-frequency variability from the control run, in part because the latter may not have reached equilibrium. To isolate changes due to the forcing, we construct anomalies by subtracting from each historical member the corresponding low-frequency drift of the PI model. This drift is identified using a loess fit to the PI control experiment (Cleveland & Devlin, 1988), based upon the `loess` function in the R statistical (or `stats`) library (e.g., Teetor, 2011).

The loess fit derives from a local linear regression that adapts to a larger range of time scales in the piControl simulations than a simple linear or quadratic fit, similar to a low-pass filter. This is useful if the remaining approach to equilibrium is largest in the early part of the control run, so that minimal detrending is necessary thereafter. This is the case for the E2.1-H NINT model whose trend is largest in the first hundred years (Figure 2d). A disadvantage of loess or low-pass fits is that they also capture low-frequency unforced oscillations in a control experiment. Where forcing in the historical experiment drives oscillations out of phase with the control variability, there will be spurious variations in the anomaly that are due to this temporal mismatch and not the forcing.

We choose our smoothing duration near 200 years to capture drift at longer time scales while excluding higher frequency oscillations. Despite this, Figure 2a shows a temporary warming around Year 4900 whose origin between residual disequilibrium and internal variability cannot be distinguished. Pre-industrial control variations at the surface are on the order of a tenth of a degree K or less, small compared to the magnitude of the forced anomalies observed by the end of the historical period, so this distinction is not of practical importance. In general, we find that drift correction is important for ocean heat content and sea level, but does not affect atmospheric trends noticeably.

3. Forcing

We characterize perturbations to the pre-industrial climate using the effective radiative forcing (ERF), defined as the difference in TOA net radiation between two AMIP-style simulations with pre-industrial SST and sea ice, where one simulation contains the forcing agent: for example, an increase in greenhouse gas (GHG) concentration (Hansen et al., 2005). The difference is constructed from 30-year averages after a 1-year adjustment period, following the RFMIP-ERF protocol (Forster et al., 2016; Pincus et al., 2016; Smith et al., 2020). The ERF gives the radiative imbalance prior to the slow return by the ocean to equilibrium but after rapid adjustments in the atmosphere to the original forcing: for example, through stratospheric temperature or clouds, including their modification by aerosols.

Unforced variability in both simulations introduces uncertainty in the ERF. Within the 30-year averaging period, interannual variations of TOA net radiation are similar in magnitude for all forcings; we combine
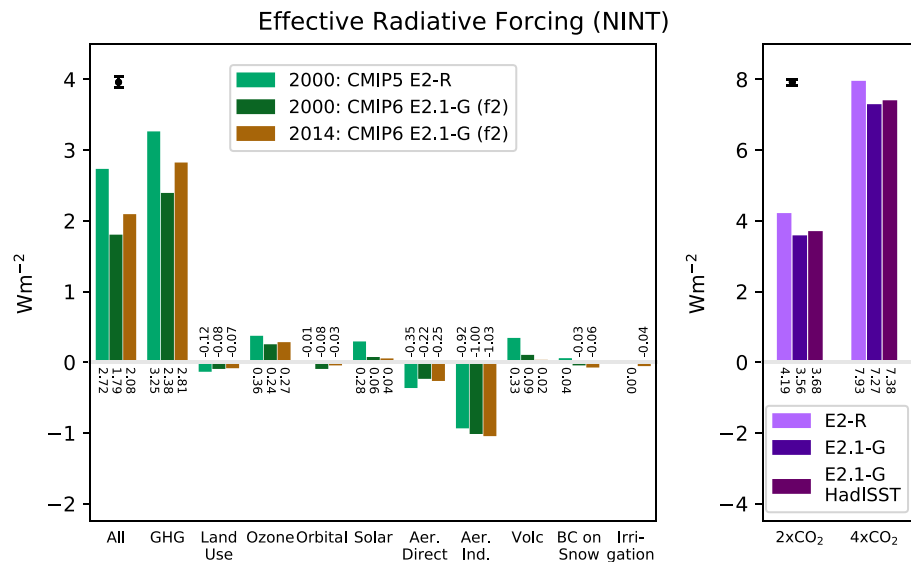
Figure 4. Effective radiative forcing for the E2.1-G NINT f2 model for the years 2000 (dark green) and 2014 (brown). For comparison, the year 2000 ERF of the CMIP5 predecessor model (E2-R NINT) is shown in light green. For 2xCO$_2$ and 4xCO$_2$, the E2-R and E2.1-G ERF values are in light and dark purple, respectively. SST and sea ice are prescribed for each calendar month by averaging the first 30 years of the pre-industrial simulation. For 2xCO2 and 4xCO$_2$, "E2.1-G HadISST" indicates that pre-industrial SST and sea ice are prescribed in E2.1-G using the Hadley observational analyses of Rayner (2003). The central 95% of the distribution of uncertainty for each individual ERF spans ± 0.08 W m^{-2}, marked by the error bar on the upper left of each panel.

variance across the individual forcings to compute a common standard deviation of 0.15 W m^{-2}, so that the standard error of the ERF, constructed from the difference of two 30-year average fluxes, is 0.04 W m^{-2}. Thus, drivers whose global ERFs have magnitudes less than 0.08 W m^{-2} cannot be distinguished statistically from 0 with more than 95% confidence without longer runs. (This same criterion applies when comparing ERF values computed from two forcing agents introduced in the same model. For an identical forcing agent applied to NINT and OMA, this range should be multiplied by $\sqrt{2}$, so that statistically distinct values are separated by at least 0.11 W m^{-2}.) This uncertainty is problematic for highly localized forcing agents like land use and irrigation, whose global ERFs are small. Regional ERF values are also obscured by unforced variations of TOA net radiation.

Below, we describe the forcings that perturb GISS-E2.1 during the historical period. Global ERF for the E2.1-G NINT f2 model version is shown for the years 2000 and 2014 in Figure 4. The year 2014 is chosen to characterize the present-day forcing; the year 2000 allows comparison to forcing computed with previous model versions. Comparison of forcing between the E2.1-G NINT f2 and OMA ensembles is given by Figure 5.

Forcing in the E2.1-G and E2.1-H configurations is expected to be similar, due to their common AGCM and pre-industrial climates that are only slightly different.

3.1. Long-Lived Greenhouse Gases

Long-lived greenhouse gases (GHGs) have nearly uniform concentrations as a result of atmospheric mixing. In the CMIP6 historical ensembles, concentrations of GHGs are annual averages of values taken from Meinshausen et al. (2017). We use "Option 2" from that study and calculate radiative forcing from prescribed concentrations of CO$_2$, CH$_4$, N$_2$O, CFC-11 (eq,) and CFC-12, where CFC-11 (eq.) includes CFC-11 plus a radiative equivalent of 38 other long-lived GHGs. Concentrations are uniform with height throughout the troposphere.

In the NINT ensembles, the NH concentrations of CH$_4$, N$_2$O and CFCs are increased compared to their Southern Hemisphere (SH) values to account for hemispheric contrasts of anthropogenic sources. In the OMA historical experiments, the vertical dependence of CH$_4$ is calculated prognostically, but its surface
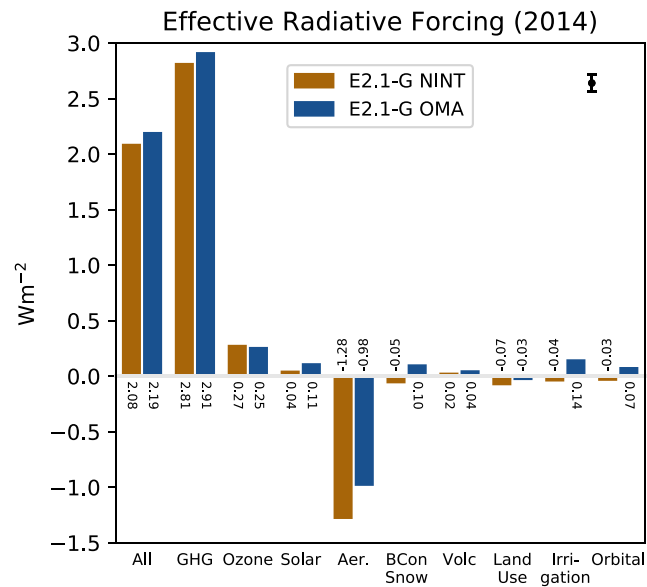
Figure 5. Effective radiative forcing in 2014 for the E2.1-G NINT f2 (brown) and OMA (blue) models. Only the combined aerosol direct and indirect forcing can be diagnosed for the OMA version; OMA forcing by BC on snow is diagnosed as the difference between two aerosol experiments. The central 95% of the uncertainty for each individual ERF spans $\pm 0.08 \, \mathrm{W \, m^{-2}}$, marked by the error bar on the upper right. ERF values corresponding to the same forcing agent applied to the OMA and NINT models are statistically distinct if they are separated by $0.11 \, \mathrm{W \, m^{-2}}$.

value is tethered to the prescribed NINT concentration and meridional contrast. The prognostic concentrations of N_2O and CFC calculated by the OMA model are replaced in the radiative calculation with the prescribed CMIP6 values used by the NINT model.

In the stratosphere, oxidation of methane is a source of water vapor (LeTexier et al., 1988) that is calculated prognostically in the OMA and MATRIX ensembles. For the NINT model, the source is implemented empirically by increasing stratospheric water according to the methane concentration at the surface, but lagged by 2 years to account for vertical transport (Miller et al., 2014). This corresponds to a forcing around $0.1 \, \mathrm{W \, m^{-2}}$ (Hansen et al., 2005).

Radiative forcing by GHGs is markedly smaller in E2.1-G than in the CMIP5 E2-R. Figure 4 shows that in Year 2000, ERF for the E2.1-G NINT configuration is $2.38 \, \mathrm{W \, m^{-2}}$ in contrast to the E2-R value of $3.25 \, \mathrm{W \, m^{-2}}$. This reduction by one-quarter is greater than the fractional contrast in forcing associated with the doubling or quadrupling of CO_2 (Figure 4), suggesting that other GHGs also contribute to the reduction.

The spatial distribution of the GHG ERF is shown in Figure 6 for both the CMIP6 and CMIP5 versions of the GISS NINT configuration. The forcing is largest mainly at low latitudes where outgoing longwave radiation is highest. The contrast is fairly uniform with latitude, although slightly higher in the SH mid-latitudes (Figure 6c).

Column water vapor (or "precipitable water") and longwave (LW) cloud radiative forcing are larger in the E2.1-G NINT pre-industrial climate, compared to their CMIP5 E2-R predecessors (Figure 7). The radiative effect of the precipitable water contrast is related to the percentage difference (Goody & Yung, 1989; Huang & Shahabadi, 2014), which is largest in the SH (Figure 7c). Contrasts in LW cloud radiative forcing vary more with latitude but show a smaller hemispheric contrast (Figure 7f). We interpret the smaller GHG forcing in the GISS CMIP6 ensembles as the result of a column that is more opaque to LW radiation due to higher concentrations of water vapor and increased LW cloud forcing, and therefore less sensitive to further increases in LW opacity by GHGs. The reduction of the E2.1-G ERF compared to the CMIP5 E2-R value brings the GHG forcing closer to that of the CMIP3 ModelE (Hansen et al., 2007).
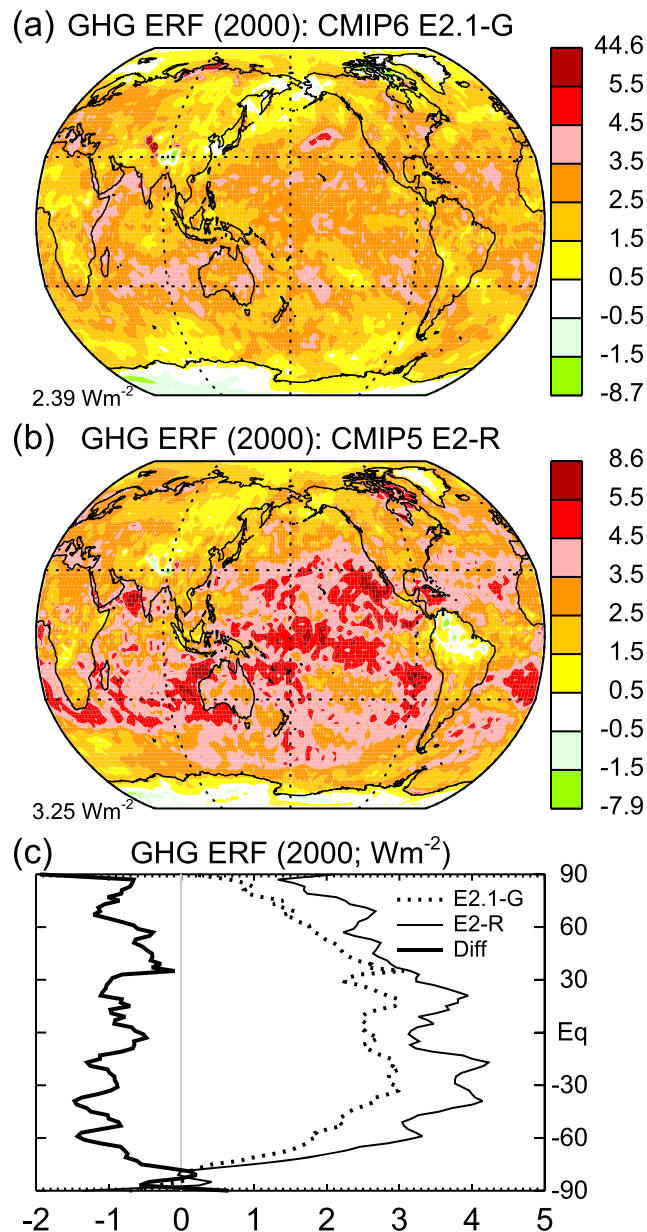
Figure 6. GHG ERF (W m^{-2}) in the (a) E2.1-G NINT f2 and (b) CMIP5 predecessor (E2-R NINT) models in the year 2000 relative to 1850. (c) difference.

3.2. Ozone

Skeie et al. (2020) show that GISS-E2.1 simulates the observed change in column ozone trend at spring-time high latitudes between the late 20th century and first decade of the 21st century. For Year 2014, the NINT ozone ERF is 0.27 W m^{-2} and statistically indistinguishable from the OMA value (Figures 4 and 5). During this year, the concentration of volcanic aerosols is small as is the forcing difference with respect to the uncorrected NINT f1 configuration that underestimates ozone destruction by heterogeneous chemistry (Figure S3).

This correction is nonetheless important to regional ozone concentration and forcing. The NINT f2 ERF for 2014 is generally positive at all latitudes (Figure 8e) except over Antarctica, where the stratospheric ozone hole caused by heterogeneous reactions with ozone depleting substances is apparent. In contrast, negative forcing over Antarctica is nearly absent in the f1 ensemble (Figure S7e).
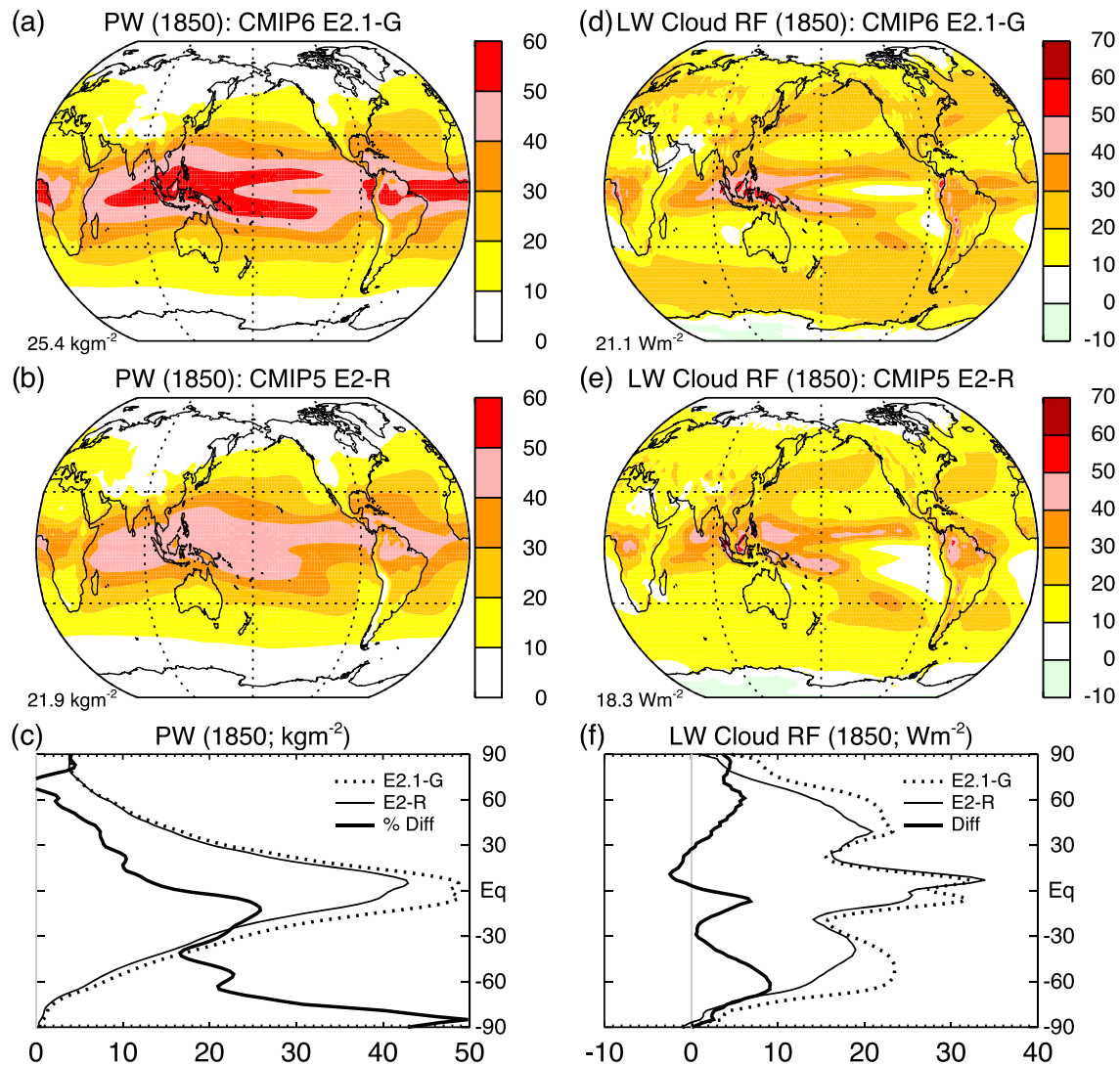
Figure 7. (left, a–c) Precipitable water vapor (PW) in pre-industrial AMIP-style simulations with the E2.1-G NINT f2 and CMIP5 E2-R NINT models. SST and sea ice are monthly varying 30-year averages from the respective coupled piControl simulations. The PW difference in the lowest panel is plotted as a percentage change, relative to the CMIP5 value. (right, d–f) Same but for LW cloud radiative forcing. The RF difference in (f) is in Wm^{-2}.

3.3. Aerosols

For Year 2000, the NINT f2 ERF corresponding to aerosol direct radiative forcing is $-0.35\,W\,m^{-2}$, more negative than the CMIP5 NINT ERF of $-0.22\,W\,m^{-2}$ (Figure 4). The rapid adjustment includes the semi-direct effect of aerosols (Hansen, Sato, & Ruedy, 1997). The ERF is largest over land and coastal regions, especially in polluted regions like East Asia and the Indo-Gangetic Plain along with regions of cultivation and fertilizer use like the North American Great Plains (Figure 9a). For OMA, emission rather than concentration is prescribed so that the latter is intertwined with cloud processes. Thus, the total aerosol forcing cannot be decomposed into separate direct and indirect contributions. However, the NINT aerosol concentration is derived from a prescribed-SST version of OMA, so the direct effect of the two models should be similar.

The ERF associated with the aerosol indirect effect is larger in magnitude than the direct forcing and dominates the total aerosol ERF, especially where aerosol emissions coincide with clouds, such as the Guinea coast of NH Africa and the Amazon basin (Figures 9b and 9c). For the NINT configuration, the magnitude of the AIE is imposed, as noted in section 2.1. However, for the OMA version, where the first indirect effect is calculated using a physically based parameterization without tuning, the magnitude of total aerosol forcing is only slightly reduced (Figure 9e).
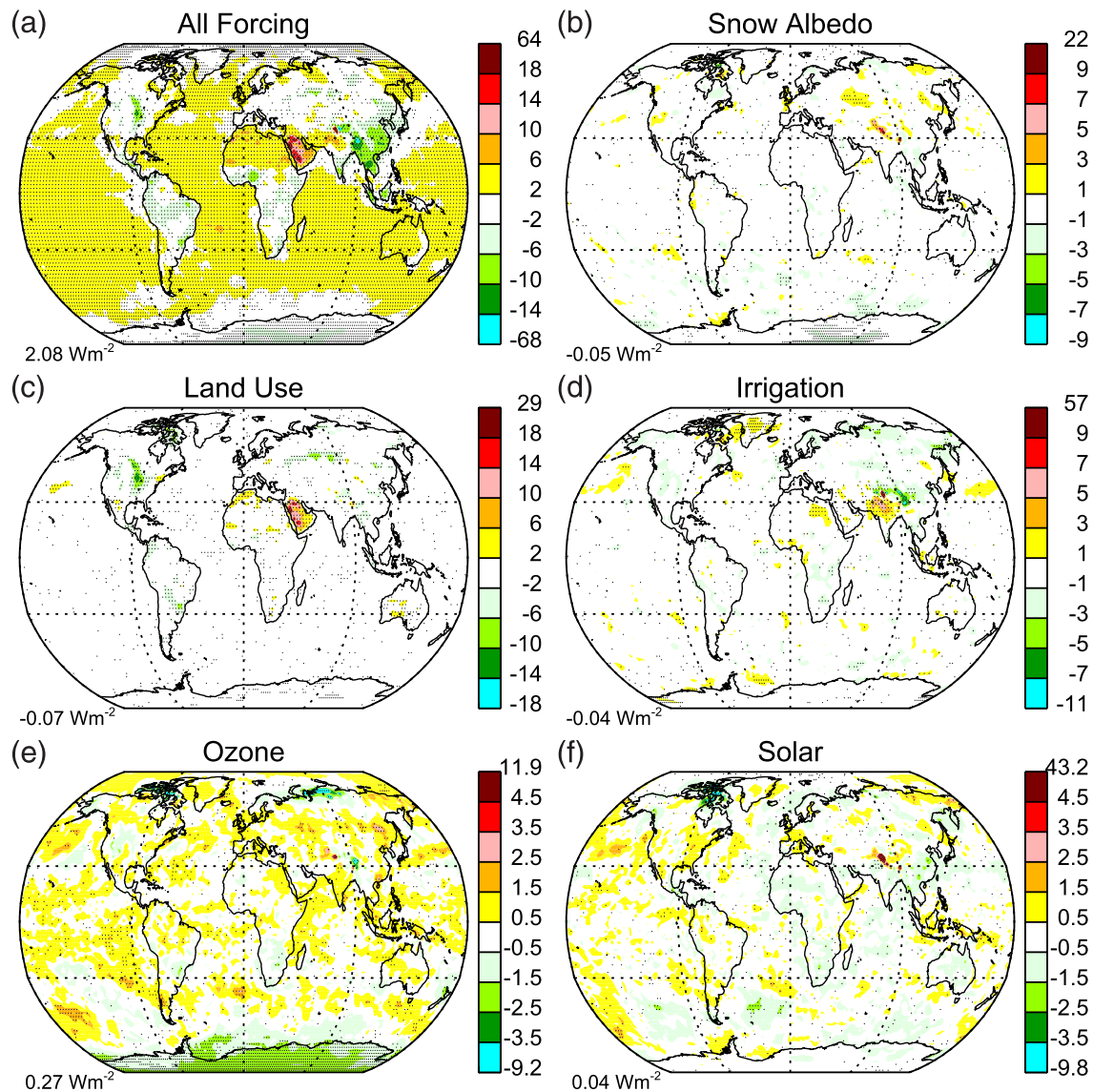
Figure 8. Effective radiative forcing in 2014 for the E2.1-G NINT f2 model (W m^{-2}) by (a) all forcing agents present simultaneously, (b) reduction of snow albedo through BC deposition, (c) land use, (d) irrigation, (e) ozone, and (f) total solar irradiance. Gray dots indicate where the forcing is distinct from 0 based upon a two-sided Student t test with a 95% central interval.

Aerosols perturb climate where their deposition upon snow and ice increases shortwave absorption (Flanner et al., 2007; Skiles et al., 2018; Warren & Wiscombe, 1985). In GISS-E2.1, surface albedo is reduced by BC deposition (Bauer et al., 2013; Hansen & Nazarenko, 2003). Forcing is locally positive, as seen in the Himalayas (Figure 8b) where prolific sources of BC are near the bright snowy surface. In the OMA configuration, the global ERF is slightly positive: 0.10 W m^{-2} (Figure 5). The NINT ERF is globally negative but not statistically distinct from 0 (Figure 4).

3.4. Solar

Incoming solar radiation at TOA is derived from the annual average of the spectrally resolved flux at 1 AU compiled by Matthes et al. (2017). The annual flux is derived from two empirical models that relate measurements of solar irradiance by satellite (Rottman, 2005) to proxies like sunspot number and facular darkening that are available farther back in time. The satellite measurements are themselves a source of uncertainty due to gaps in temporal coverage and unknown offsets between non-overlapping instruments (e.g., Fröhlich, 2006). The same CMIP6 data set provides the photon flux that is used to calculate photolysis rates.
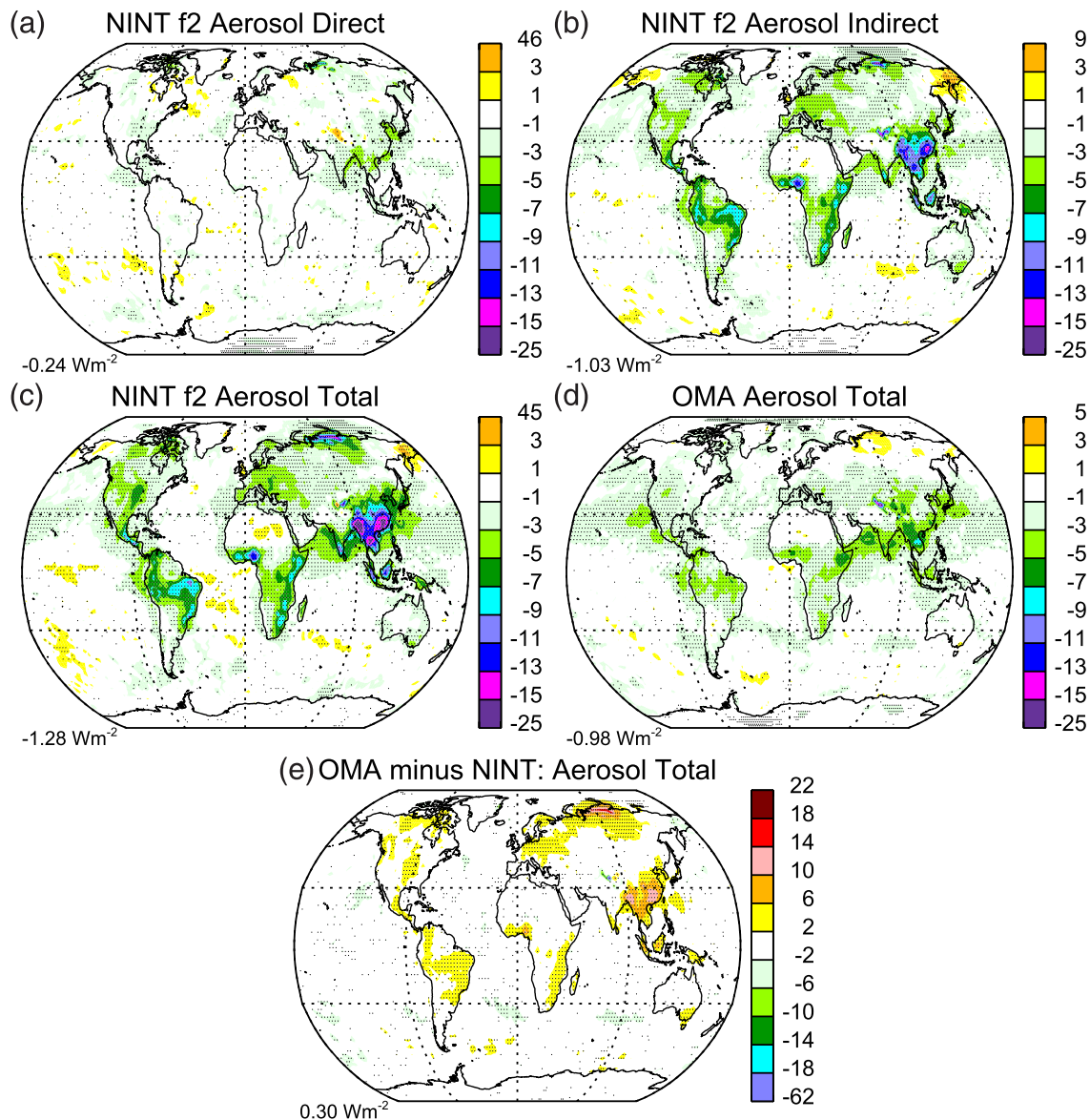
Figure 9. Effective radiative forcing in 2014. (a) the direct, (b) indirect, and (c) total radiative forcing of NINT f2 aerosols. (d) Total OMA aerosol forcing, and (e) difference of total aerosol forcing between OMA and NINT. Gray dots indicate where the forcing (or forcing difference) is distinct from 0 based upon a two-sided Student *t* test with a 95% central interval.

The NINT f2 global ERF for solar in 2000 is $0.06\,\mathrm{W\,m^{-2}}$, down from the CMIP5 NINT ERF of $0.28\,\mathrm{W\,m^{-2}}$ (Figure 4). For the year 2014, solar ERF for both CMIP6 NINT and OMA models is small and indistinguishable from 0 (Figure 5). Both of these years are near the maximum irradiance measured during their respectively solar cycles. The latest cycle (24) is unusually weak compared to recent predecessors (Jiang et al., 2015).

3.5. Volcanic Aerosols

We use the recommended CMIP6 data set for stratospheric aerosols created by volcanic eruptions (ftp://iacftp.ethz.ch/pub_read/luo/CMIP6/Readme_Data_Description.pdf), a change from prior GISS CMIP experiments, where volcanic aerosols were derived from an update of Sato et al. (1993).

The CMIP6 volcanic prescription is based upon a combination of direct measurements and model reconstructions. During the satellite era (after 1979), radiance measurements are used to retrieve the aerosol extinction and particle size as an evolving function of height and latitude (Thomason et al., 2018). This period includes major eruptions by El Chichon and Pinatubo.

Table 3
Major Volcanic Eruptions During the Historical Period (1850–2014)

Name	Location	Date
Makian	Indonesia 0.3°N	May 1861
Krakatoa	Indonesia 6°S	September 1883
Santa Maria	Guatamala 15°N	October 1902
Agung	Bali 8°S	June 1963
El Chichon	Mexico 17°N	April 1982
Pinatubo	Philippines 15°N	July 1991

Prior to the satellite era, stratospheric aerosol properties are derived using a chemical transport model (Arfeuille et al., 2014). The injection of sulfur dioxide following an eruption is estimated using measurements from multiple ice cores. Oxidation and conversion to sulfate aerosol is calculated by a model that simulates nucleation, condensation, coagulation, and deposition with transport prescribed from climatology as a function of latitude and height. Photometric extinction is used as a constraint when available. The model simulates extinction by sulfate aerosols for specific historical eruptions identified by the ice cores. During quiescent periods, aerosol properties are taken from satellite retrievals averaged between 1996 and 2005. Aerosol properties following eruptions are different after 1979 when retrievals replace model estimates.

The CMIP6 prescription moves one eruption identified by Sato et al. (1993) from the late 1850s to 1861 (Table 3). GISS prescribes total extinction at $0.55\,\mu m$ using the CMIP6 prescription, but specifies its own version of the remaining Mie parameters as described by Hansen et al. (2005). During the historical period, total extinction at $0.55\,\mu m$ is generally reduced compared to Sato et al. (1993). Extinction is now largest following the eruption of Pinatubo rather than Krakatau. Globally, the effective radius of volcanic aerosols is reduced.

During the piControl simulations, volcanic aerosol properties are held constant at a value equal to their average during the historical period. This is intended to minimize the climate trend resulting from successive eruptions during the historical period by requiring that the time-average volcanic forcing be 0. Negative forcing following eruptions after 1850 is punctuated by positive forcing during the intervening quiescent periods (Figure 10). For example, in 2000, during the quiescent period following Pinatubo, CMIP6 volcanic forcing is $0.09\,W\,m^{-2}$ (Figure 4). The assumption of zero forcing during the historical period results in a forcing discontinuity at 1850, when the constant pre-industrial concentration of volcanic aerosols is replaced by a smaller value until the Makian eruption in 1861 (Table 3).

3.6. Orbital

Orbital forcing results from variations in the eccentricity of Earth's orbit, the obliquity of its rotation axis compared to the orbital plane and the date of perihelion. As for CMIP5, these parameters are prescribed following Berger (1978). During the historical period, orbital variations in TOA insolation are small and are calculated only to allow continuity with Last Millennium simulations where changes are larger.

3.7. Land Use

The expansion of cultivation and grazing with human population over the historical period reduces the fraction of natural vegetation types within each land grid box. We inadvertently implemented land use in GISS-E2.1 using the same areal fractions of cropland and pasture as prescribed for the GISS CMIP5 model (Miller et al., 2014). Departures from the CMIP6 protocol are described by Ito et al. (2020). After 2005, the end of the CMIP5 historical period, land use is held constant.

Land use affects surface albedo. Forcing is negative where crops replace darker forests; forcing is small where crops expand into grassland and positive where bright bare soils are planted. Land use alters transpiration, which increases where the leaf area index is increased. Reforestation is not included in our prescription. The "pasture" category includes both managed pastures and unmanaged rangeland. The latter is found within natural vegetation categories and its surface properties are generally unmodified by grazing, so its inclusion within the pasture category leads to an excessive prescription of cultivated area. This overestimate is largest in grid boxes with abundant natural grasslands and arid shrublands where rangelands are identified. The inclusion of unmanaged rangeland in the cultivated category nearly doubles the area of simulated land use.
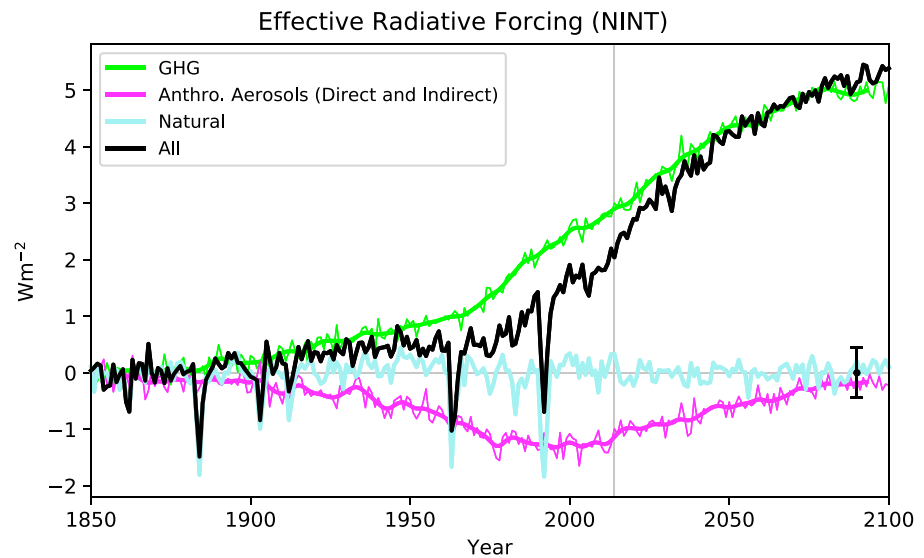
Figure 10. Effective radiative forcing for the E2.1-G NINT f2 model during the historical period extended to 2100 using Shared Socioeconomic Pathway (SSP) 2-4.5 (Riahi et al., 2017). The standard deviation of interannual variations of the ERF is $0.22\,\mathrm{W\,m^{-2}}$. The error bar near the right edge marks plus and minus two standard deviations. For GHGs and aerosols, whose concentrations rise gradually compared to volcanic forcing, we apply a 15-term low-pass filter that reduces variations shorter than 30 years. The vertical gray line marks 2014, the end of the historical period.

Despite the overestimation of land use area, neither NINT or OMA global forcing in 2014 are distinct from 0 (Figure 5). Nonetheless, negative values of ERF are significant over the Great Plains of North America (Figure 8c), consistent with the GISS CMIP5 ERF, due to replacement of forest by lighter crops. Figure 8c shows large positive forcing over the Arabian Peninsula, where cultivation expands to cover bright bare soil.

3.8. Irrigation

Irrigation was implemented within ModelE after CMIP5 (Cook et al., 2011, 2014; Puma & Cook, 2010; Shukla et al., 2014). Water demand for irrigation in GISS-E2.1 was calculated between 1900 and 2005 as described by Wada et al. (2014) using irrigation areal extent from Siebert et al. (2015). Nineteenth century irrigation is estimated by extrapolating backwards from the trend between 1900 and 1929. Irrigation is held constant after 2005 using the flux from 2004.

The irrigated flux is taken from lakes and runoff within the same grid box and deposited into the soil column (Puma & Cook, 2010). When lakes and runoff are insufficient to meet demand, water is assumed to be supplied by a aquifer of unlimited volume. The aquifer draw represents an addition of water to ModelE.

Irrigated water may run off or evaporate, but some is diverted into deep soil layers that have a long return time to the ocean and atmosphere. As a consequence, irrigation initially has the potential to sequester water in the soil column and reduce sea level by transferring moisture from the ocean to the land. Eventually enough water accumulates in the deep layers that a large fraction of the irrigated flux remains near the surface where it is subject to evapotranspiration and runoff. Crops and vegetation share the same soil column in GISS-E2.1. Thus, saturation of the deep soil layers through the irrigation flux takes longer than if water were applied solely to the crop fraction of the soil column.

Global radiative forcing by irrigation is statistically indistinguishable from 0 in the NINT model but slightly positive in the OMA model (Figure 5). Forcing is positive over the Indo-Gangetic Plain (Figure 8d), a site of extensive irrigation during the twentieth century (Puma & Cook, 2010). Moistening of the atmosphere by irrigation increases LW opacity, which contributes to positive forcing. The surface can nonetheless cool, as moisture supplied by irrigation shifts the compensation of solar heating away from the turbulent sensible heat flux toward efficient cooling by evaporation.

3.9. All Forcing

For Year 2014, the NINT ERF for the addition of all forcing agents simultaneously is $2.08\,\mathrm{W\,m^{-2}}$ and is dominated by the contribution from rising greenhouse gas concentrations (Figures 4 and 10). For the OMA
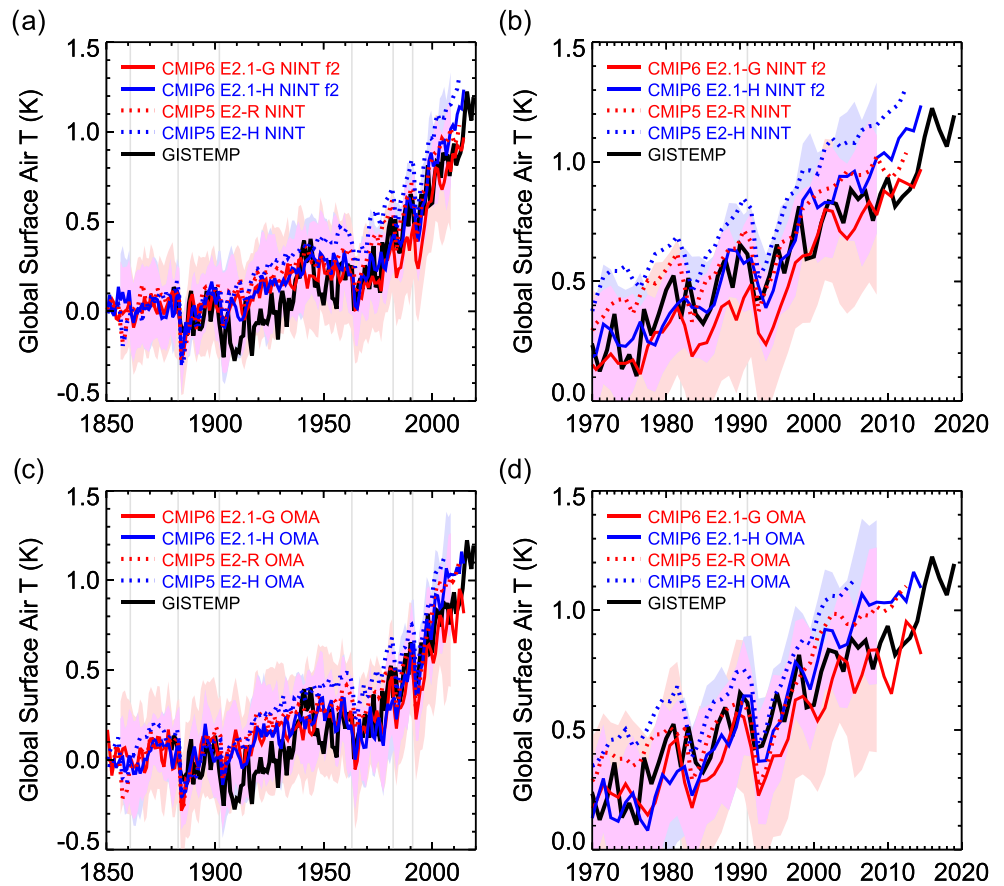
Figure 11. Ensemble, global, and annual-average surface air temperature (K) during the historical simulations (1850–2014). (a) NINT f2 and (c) OMA models; (b and d) same but emphasizing the period after 1970. Intra-ensemble variability at plus and minus two standard deviations is marked by pink and light blue shading for the E2.1-G and E2.1-H models, respectively, with light purple marking their overlap. This variability is smoothed using an 11-year window that damps periods less than a decade. Temperatures from the corresponding CMIP5 models (E2-R and E2-H) are plotted as dotted lines. The observed GISTEMP record is in black. Temperatures are anomalies, defined relative to their late-nineteenth century average (1880–1900). Vertical gray lines show major volcanic eruptions listed in Table 3.

configuration, the all-forcings ERF is 2.19 W m^{-2} (Figure 5), marginally larger than the NINT value, with aerosols making the largest contribution to the contrast.

For Year 2000, the CMIP6 NINT ERF is 1.79 W m^{-2}, nearly 1 W m^{-2} smaller than the CMIP5 NINT value of 2.72 W m^{-2}. Most of this change is due to a reduction of GHG forcing in the CMIP6 model. The total aerosol forcing is nearly unchanged between the two generations of NINT experiments, but solar and volcanic forcing combine for an additional reduction of 0.46 W m^{-2}.

The spatial distribution of the ERF is generally positive as a result of increasing LW opacity due to rising GHG concentrations (Figure 8a). However, the imprint of more regionalized forcings is apparent: for example, land use over the North American Great Plains and western Arabian Peninsula, aerosols over southeast Asia, the Guinea coast and Amazon basin, BC deposited upon snow in the Himalayas, irrigation over the Indo-Gangetic Plain, and the near offset of GHG forcing over SH high latitudes by stratospheric ozone loss.

4. Historical Response

4.1. Temperature

Surface air temperature for the historical ensembles is plotted in Figure 11 along with the GISTEMP Land-Ocean Temperature Index (LOTI; Lenssen et al., 2019). (Table 4 summarizes the observations used in this study.) Anomalies are defined relative to the start of the GISTEMP record in the late nineteenth century (1880–1900), when the forcing is still small.

Table 4
Observations Used in Model Evaluation

Variable	Observation	Source	Citation
tas	GISS surface temperature analysis (GISTEMP)	data.giss.nasa.gov/gistemp v.4, accessed 2020-07-29	Lenssen et al. (2019)
msu tls, tmt	MSU/AMSU brightness temperature Channel TLS, TMT	Remote Sensing Systems v.4.0 www.remss.com/measurements/upper-air-temperature	Mears and Wentz (2016)
siconca	Sea ice area	National Snow and Ice Data Center Hemispheric Sea Ice Area Index, v.3 nsidc.org/data/g02135	Fetterer et al. (2017)
	Sea ice extent	doi. org/10.5281/zenodo.3717240	Brennan et al. (2020)
ohc	Ocean heat content: column	laurezanna.github.io/post/ohc_updated_data	Zanna et al. (2019)
	Ocean heat content: upper 700 m	www.cmar.csiro.au/sealevel/thermal_expansion_ocean_heat_timeseries.html	Church et al. (2011)
msftmz	Atlantic Meridional Overturning	RAPID-MOCHA array (26°N) doi.org/10/c72s	Smeed et al. (2019)
zos	Sea level	research.csiro.au/slrwavescoast/sea-level	Church et al. (2011)
zostoga	Thermosteric sea level	research.csiro.au/slrwavescoast/sea-level	Church et al. (2011)
pr	Precipitation	Global Precipitation Climatology Project (GPCP) v.2.3 www.esrl.noaa.gov/psd/data/gridded/data.gpcp.html	Adler et al. (2018)

The GISS ensembles closely track the GISTEMP LOTI throughout the historical period, warming by 1 K and remaining within two standard deviations of intra-ensemble variability. For the NINT ensembles, two standard deviations is roughly 0.3 K (For GISTEMP, the 95% central uncertainty at 1880 is 0.15 K, decreasing to 0.05 K in the present day Lenssen et al., 2019).

Over the historical period, the GISS-E2.1 models warm by 0.1–0.2 K less than their GISS CMIP5 counterparts. The reduced warming is in spite of a higher transient climate response (TCR) for the GISS-E2.1 models compared to GISS-E2 (Table 5, excerpted from Table 6 of Kelley et al., 2020), suggesting that the CMIP6 reduction is a consequence of smaller total forcing (Figure 4). The E2.1-H version has a higher TCR than E2.1-G, consistent with its higher warming in both the NINT and OMA versions. (Table 5).

Because E2.1-G and E2.1-H share a common atmospheric model, the warming and TCR contrasts suggest that HYCOM exports heat from the surface to the deep ocean more slowly (as shown in section 4.3).

Regional warming is shown in Figure 12 as the temperature difference between the start of the GISTEMP LOTI (1880–1899) and the end of the historical period (2005–2014). Both the models and observations show their largest warming at high latitudes, especially in the NH, a robust feature of coupled models (Dai et al., 2019; Holland & Bitz, 2003). In E2.1-G, there is cooling in the region of North Atlantic deepwater formation near the outlet of the Labrador Sea that matches the location of the observed warming minimum (Figures 12a and 12b). This minimum is not reproduced by the E2.1-H ensembles (Figures 12d and 12h).

Despite the broad hemispheric agreement of model temperature with the observations, regional disagreements are apparent. The extratropical continents in both hemispheres warm less than observed in both the NINT and OMA simulations. The contrast is especially large over the Indo-Gangetic Plain and East Asia, regions cooled by irrigation and strong negative forcing by anthropogenic aerosols (Figures 8 and 9).

Table 5
Transient Climate Response From Kelley et al. (2020)

Model	Physics version	CMIP6 (K)	CMIP5 (K)
E2.1-G	NINT	1.8	1.4
E2.1-G	OMA	1.6	1.6
E2.1-H	NINT	1.9	1.7
E2.1-H	OMA	2.0	1.8

Note. The CMIP5 antecedent to the GISS-E2.1-G model is GISS-E2-R.
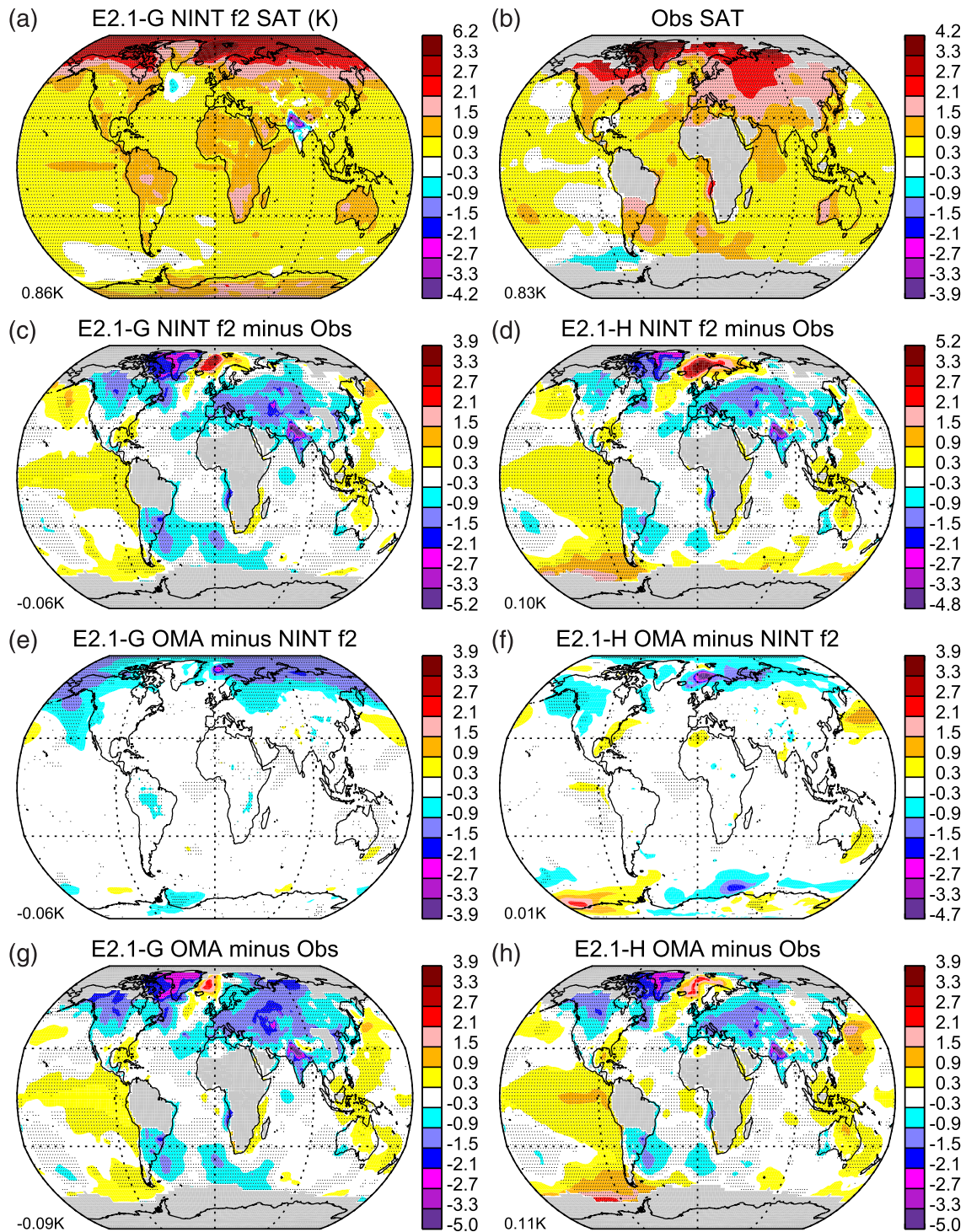
Figure 12. Change in ensemble and annual-average surface air temperature (K) between 1880–1899 and 2005–2014. (a) E2.1-G NINT f2, (b) GISTEMP LOTI, (c) E2.1-G NINT f2 minus GISTEMP, (d) E2.1-H NINT f2 minus GISTEMP, (e) E2.1-G OMA minus NINT f2, (f) E2.1-H OMA minus NINT f2, (g) E2.1-G OMA minus GISTEMP, (h) E2.1-H OMA minus GISTEMP. Observed temperature is plotted only at locations where measurements are available during 90% of each averaging period. Gray dots indicate where the (a) ensemble-mean or (b) observed change or (e, f) difference between the ensemble-mean changes is distinct from 0 based upon a two-sided Student t test with a 95% central interval. (c, d, g, h) Gray dots indicate where the observed value is outside the central 95% of a normal distribution constructed from the individual ensemble members.
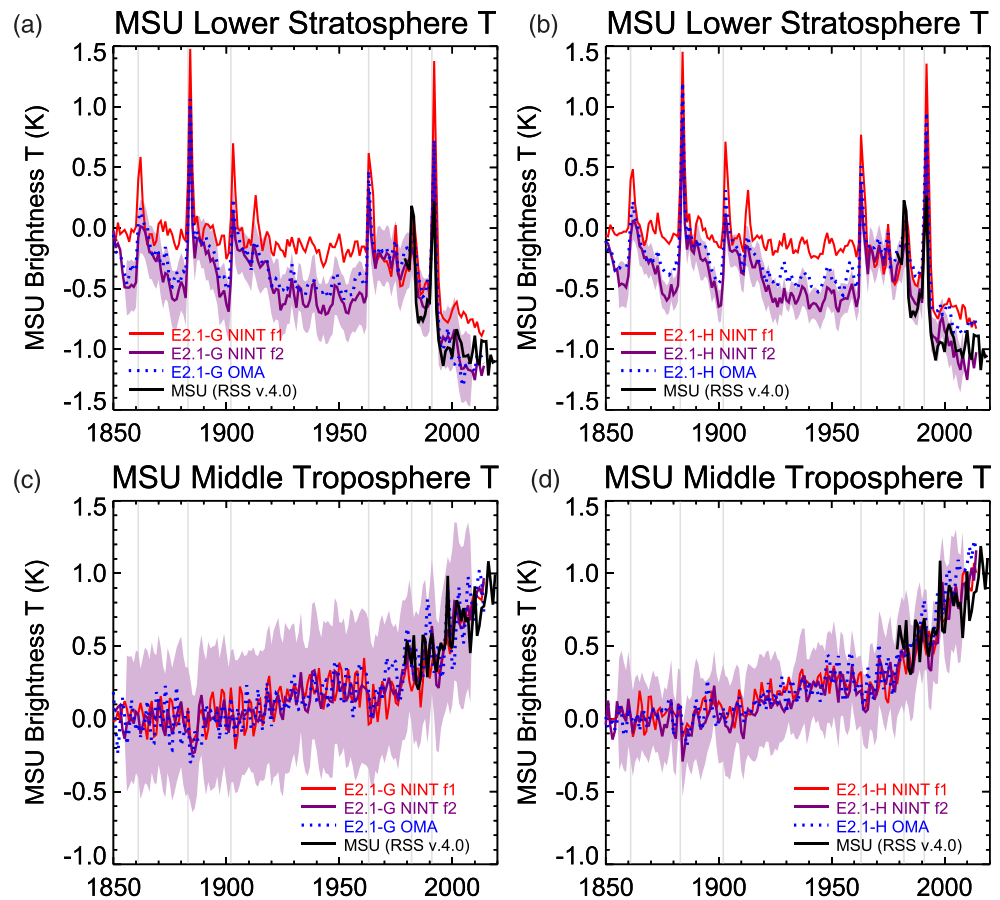
Figure 13. Global and annual temperature retrieved from the Microwave Sounding Unit (MSU) and Advanced Microwave Sounding Unit (AMSU) compared to model values. (top) Temperature of the lower stratosphere (TLS), (bottom) temperature of the middle troposphere (TMT); (left) E2.1-G, (right) E2.1-H. Intra-ensemble variability (corresponding to plus and minus two standard deviations) for the NINT f2 model is shaded in purple. The retrievals (black) are from RSS v.4.0 and are offset to have the same time average as the NINT f2 ensemble during the period 1979–2014. Vertical gray lines show major volcanic eruptions.

In contrast, there is excessive warming by all model versions in the Greenland and Norwegian Seas. In the Tropics, model warming is "ENSO-like" with greater magnitude in the equatorial Eastern Pacific compared to observations (Cane et al., 1997).

Differences between the OMA and NINT ensembles are largest at NH high latitudes, where meridional amplification of warming by OMA is smaller. This moderation is most prominent in the vicinity of the Norwegian Sea in both OMA models (Figures 12e and 12f).

Upper air temperatures retrieved from the Microwave Sounding Unit (MSU) and Advanced Microwave Sounding Unit (AMSU) are plotted in Figure 13. Retrievals of lower stratosphere (TLS) and middle tropo-sphere (TMT) temperature, representing weighted averages with respect to height, are provided by Remote Sensing Systems (Mears & Wentz, 2016, RSS Version 4.0). TLS is weighted most heavily near 20 km; TMT has its largest contribution from radiation originating near 10 km, but has significant weight extending into the lower stratosphere. The weighting with respect to height varies with location and season. For compari-son to the retrievals, we average model temperature over a range of altitude that is globally and temporarily invariant (Hansen et al., 1998), an approximation that has been shown to have a small effect upon annual trends compared to other uncertainties, including differences among retrieval algorithms (Santer et al., 1999; Thorne et al., 2011).

The TLS retrievals show abrupt warming of the lower stratosphere following a volcanic eruption. Recovery is initially rapid followed by a slower decadal cooling and return to pre-eruption values (Figure 13a and 13b). In the last half-century, this cycle is superimposed upon a multi-decadal cooling trend that has been
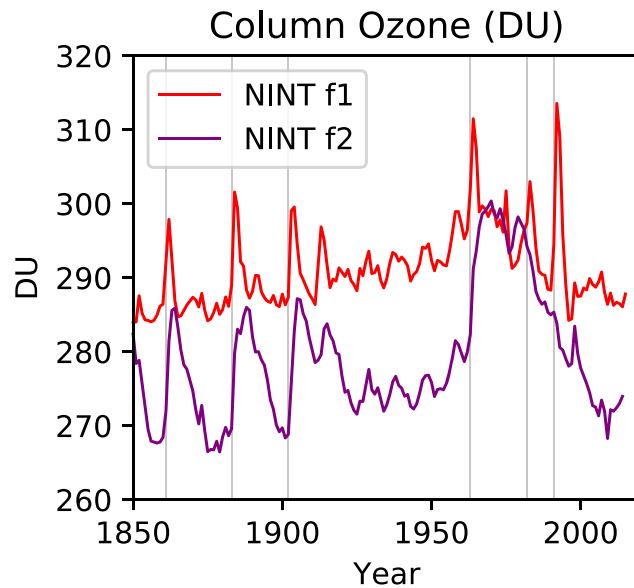
Figure 14. Global column ozone (DU) prescribed in the NINT f1 and f2 ensembles. Vertical gray lines show major volcanic eruptions.

attributed to increasing concentrations of ozone depleting substances and greenhouse gases (Gillett et al., 2011; Manabe & Weatherald, 1967). Each of the GISS-E2.1 ensembles reproduces these transient warmings and recent cooling. Throughout the historical period, TLS from the NINT f2 ensemble closely tracks the OMA value, reflecting the prescription of atmospheric composition in the former model from an AMIP version of the latter.

At the start of the historical period, the lower stratosphere in the OMA and NINT f2 models cools abruptly, prior to any volcanic eruption (Figures 13a and 13b). The anomaly is outside two standard deviations of intra-ensemble variability, and is significantly different from the initial pre-industrial temperature. We interpret this cooling as the result of a discontinuous removal of volcanic aerosols. As noted in section 3.5, aerosol concentration during the pre-industrial is constant and equal to its average during the historical period. At the start of the historical period, concentration drops abruptly and remains low until the Makian eruption in 1861. The removal of volcanic aerosols causes cooling in the OMA and NINT f2 ensembles by the same mechanism that the injection of aerosols into the stratosphere following an eruption causes rapid warming through longwave absorption (McCormick et al., 1995).

In contrast, this cooling is absent in the NINT f1 ensemble that also generally lacks the slow decadal recovery following each eruption that is present in the other model versions. This absence is associated with a consistent warm offset of the NINT f1 TLS during quiescent periods. Until the introduction of CFCs into the stratosphere in the mid-twentieth century, TLS is tightly coupled to column-integrated ozone in both NINT models (Figure 14). At the start of the historical period, column ozone drops in the NINT f2 ensemble, associated with a reduced concentration of volcanic aerosols that host heterogeneous chemical reactions on the aerosol surface. These reactions compete for nitrogen and reduce the concentration of NO_x and the associated catalytic destruction of ozone (Muthers et al., 2015; Tie & Brasseur, 1995). The abrupt reduction of volcanic aerosols at the start of the NINT f2 and OMA ensembles causes catalytic ozone destruction to increase, associated with cooling of the lower stratosphere. In contrast, heterogeneous reactions are muted in the NINT f1 model by the misspecification of surface area for volcanic aerosols. Ozone destruction in NINT f1 is nearly unchanged by the removal of volcanic aerosols in 1850, resulting in a steady concentration and temperature of the lower stratosphere.

By the late twentieth century, CFC concentrations have increased and the effect of heterogeneous reactions upon ozone is now dominated by destruction through the presence of chlorine. Figure 14 shows that column ozone decreases following the Pinatubo eruption in the NINT f2 and OMA ensembles, as observed (Hofmann & Oltmans, 1993), and in contrast to the increase in the NINT f1 model. The contrasting relation of TLS and ozone in the NINT f2 model between nineteenth and late-twentieth century eruptions shows that
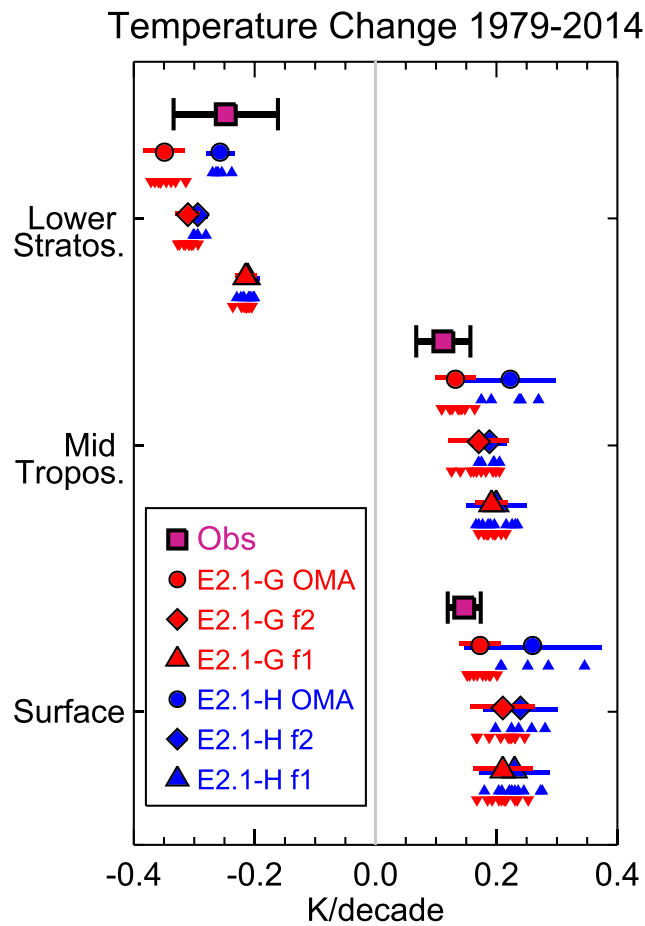
Figure 15. Trends between 1979 and 2014 for GISTEMP LOTI and RSS v.4.0 MSU TMT and TLS. Ensemble means for NINT f1, NINT f2, and OMA are marked by triangles, diamonds, and circles, respectively. E2.1-G is marked in red; E2.1-H is in blue. Trends for individual ensemble members are marked by small triangles. The red and blue horizontal lines indicate plus and minus two standard deviations of the trend distribution within each ensemble. Observational uncertainty reflects interannual variations but not instrument or retrieval uncertainty.

the representation of heterogeneous chemistry in OMA is flexible and able to reproduce the distinct limits dominated by chlorine and nitrogen associated with high and low background concentrations of CFCs (Tie & Brasseur, 1995).

Retrievals of TMT show a warming trend starting in the late twentieth century punctuated by abrupt warming associated with El Niño events and cooling following major volcanic eruptions (Figure 13c and 13d). (TMT trends also include an offsetting contribution from cooling in the lower stratosphere.) The warming trend and post-volcanic cooling are reproduced by each ensemble.

Temperature trends between 1979 and 2014 at the surface and MSU retrieval layers are shown in Figure 15. This period was chosen for overlap between the MSU retrievals and historical ensembles. Observational uncertainty reflects interannual variations, but not instrument or retrieval uncertainty. For surface temperature, this effect is on the order of 0.01 K per decade and relatively unimportant. Retrieval uncertainty for MSU is higher: 0.05 K per decade (Thorne et al., 2011).

The figure shows surface and tropospheric warming but stratospheric cooling, a contrast that is consistent with anthropogenic forcing, especially rising greenhouse gas concentrations and ozone depletion (e.g., Gillett et al., 2011; Manabe & Weatherald, 1967). The NINT f1 and f2 ensembles have tropospheric trends that are indistinguishable, showing the limited effect of the erroneous post-eruption ozone chemistry in the original ensemble (Table 6). The difference is larger and significant in the stratosphere.

Table 6
Modeled and Observed Trends in Surface, Middle Troposphere, and Lower Stratospheric Global Temperature Between 1979 and 2014 (in K per Decade)

	Surface	Middle troposphere	Lower stratosphere
E2.1-G NINT f1	0.21 ± 0.05	0.19 ± 0.03	-0.21 ± 0.02
E2.1-G NINT f2	0.21 ± 0.05	0.17 ± 0.05	-0.31 ± 0.02
E2.1-G OMA	0.17 ± 0.03	0.13 ± 0.03	-0.35 ± 0.03
E2.1-H NINT f1	0.23 ± 0.06	0.20 ± 0.05	-0.21 ± 0.02
E2.1-H NINT f2	0.24 ± 0.06	0.19 ± 0.03	-0.29 ± 0.02
E2.1-H OMA	0.26 ± 0.11	0.22 ± 0.08	-0.26 ± 0.02
Obs	0.15 ± 0.03	0.11 ± 0.05	-0.25 ± 0.09
Obs (1979–2018)	0.17 ± 0.03	0.13 ± 0.04	-0.22 ± 0.10

Note. Model trends are calculated from the ensemble mean, while their uncertainty range bounds the central 95th percentile of the distribution of ensemble member trends. Observed trends are based upon GISTEMP (surface) and RSS v.4.0 MSU channels TMT (middle troposphere) and TLS (lower stratosphere). Uncertainty of observed trends is related to interannual variability but does not include measurement or retrieval error. Different methods of MSU retrieval contribute additional TMT trend uncertainty near 0.05 K per decade (cf. Thorne et al., 2011).

For the lower stratosphere, the observed cooling is consistent with the distribution of model trends for each ensemble. In the mid-troposphere and surface, most ensembles have mean warming trends that are excessive compared to observations, but each E2.1-G ensemble has members whose trend is close to the observed value.

We recomputed trends with the observational record extended to 2018, which introduces an unusually large El Niño event. The observed surface and tropospheric trends increase by 0.02 K per decade, which reduces the model discrepancy (Table 6). This extension has little effect upon agreement in the lower stratosphere.

At the surface and mid-troposphere, E2.1-H ensembles with the HYCOM ocean model show larger warming than the E2.1-G ensembles. This contrast is largest in the OMA ensembles, consistent with the greater contrast in TCR (Table 5.)

4.2. Sea Ice

Sea ice area is shown for each ensemble as an annual hemispheric percentage in Figure 16, along with the percentage retrieved by the National Snow and Ice Data Center (NSIDC Fetterer et al., 2017). Sea ice area is greater in both hemispheres in the GISS-E2.1 ensembles compared to their GISS CMIP5 predecessors. This represents an improvement in the SH, where the E2.1-G ensemble means are in good agreement with the NSIDC retrievals at the start of the satellite period. Observed SH sea ice trends are obscured by interannual variability (Figure 16). E2.1-G trends are weak, but SH ice loss in the E2.1-H ensembles is comparable to NH trends and unrealistically large. Freshwater from melting ice sheets is one source of sea ice variability not represented in the models (Golledge et al., 2019; Rye et al., 2014). Anomalous meltwater increases surface stratification, reducing vertical mixing and the upward transport of ocean heat that leads to sea ice loss.

In the NH, ensemble-mean area is up to 30% higher than the NSIDC retrieval, with a larger bias than their GISS CMIP5 counterparts. Kelley et al. (2020) attribute this increase to excessive radiative loss at high latitudes related to brighter clouds. Figure 16 includes an estimate of sea ice extent that extends prior to the satellite period (Brennan et al., 2020). The estimate addresses the sparsity of sea ice observations prior to the mid-twentieth century by assimilating comparatively ubiquitous temperature measurements into an ESM to infer sea ice extent, defined as the total area of grid cells whose areal fraction of ice exceeds 15%. By excluding open water within a grid cell, ice area is generally smaller. For the NSIDC retrieval grid, ice extent is larger than area by about 20%. To facilitate comparison with model area, we give ice extent the same average between 1979 and 2013 as the NSIDC area.

NH ice extent shows a decline in the early twentieth century interrupted by a brief recovery following the eruption of Mount Agung before a steeper decline during the satellite period. Throughout the historical
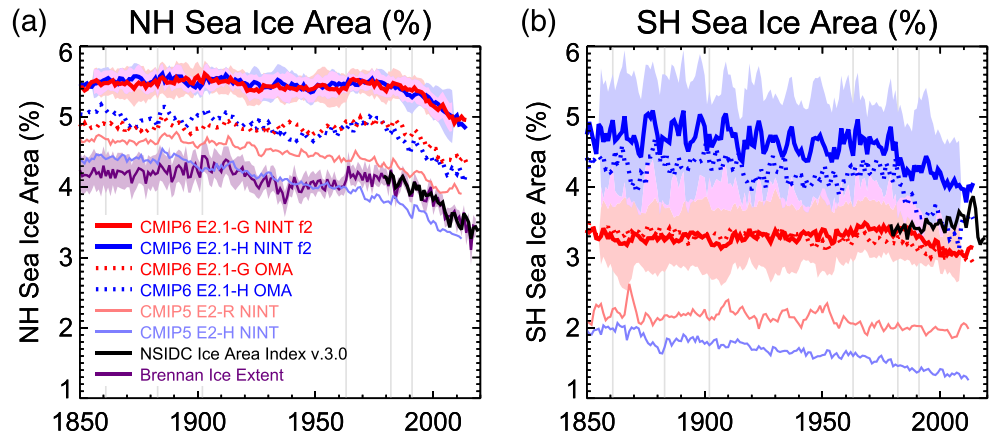
Figure 16. Ensemble and annual sea ice area as a hemispheric percentage during the historical simulations (1850–2014). For comparison to observations, total ice area is shown instead of a detrended anomaly. Unforced piControl trends are small compared to recent historical trends (Figures S1 and S2). The pink and light blue shading marks plus and minus two standard deviations of intra-ensemble variability for the E2.1-G and E2.1-H NINT f2 models, respectively, with light purple marking their overlap. Sea ice from the corresponding CMIP5 models (E2-R and E2-H) are plotted as thinner lines. The observed hemispheric percentage derived from the NSIDC sea ice area index (v.3.0) is in black. NH sea ice extent calculated from assimilated temperature measurements is shown in purple with shading to denote plus and minus two standard deviations of uncertainty (Brennan et al., 2020). Ice extent is assigned the NSIDC area average between 1979 and 2013 to reduce their offset.

period, the GISS-E2.1 ensembles have too much NH sea ice compared to their GISS CMIP5 predecessors or the Brennan et al. (2020) assimilation. The GISS-E2.1 ensembles also underestimate the early twentieth-century decrease and partial recovery with the E2.1-H OMA ensemble showing the best correlation. The GISS CMIP5 ensembles start from a more realistic baseline and show a declining twentieth-century trend as observed, although they lack the post-Agung recovery.

Over the satellite period, the observed decline in hemispheric ice area is generally within the range simulated by the GISS-E2.1 models (Figure 17). The GISS-E2.1 ensembles have greater success reproducing the decline of the sea ice maximum in March compared to the September minimum.

Regional sea ice trends are shown in Figure 18. The largest NH areal loss is at the seasonal ice margin in the Greenland and Norwegian Seas. This loss coincides with localized regions of atmospheric warming that is excessive compared to observations (Figures 12c, 12d, 12g, and 12h), suggesting that the removal of sea ice allows the ocean to warm the atmosphere. Sea ice differences between the NINT and OMA ensembles are small, even though the latter is colder over an extensive region at high latitudes (Figure 12e and 12f). Over most of this region, absolute temperatures are far below the melting point of sea ice, which is thus insensitive to small temperature contrasts.
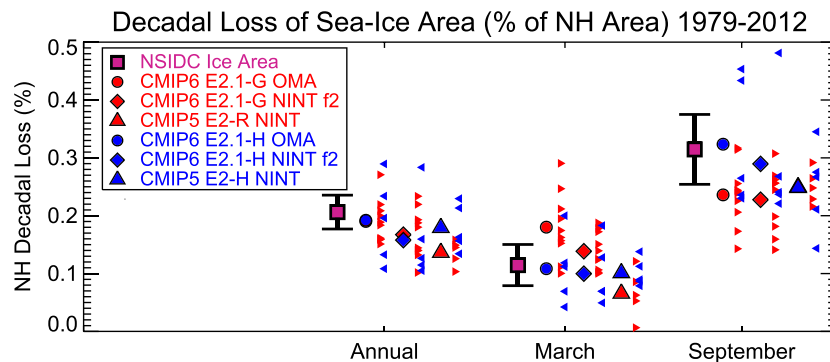

Figure 17. Trends in NH sea ice percentage between 1979 and 2012. The small triangles show trends for individual ensemble members. The NSIDC retrieved area is in purple; the uncertainty bar represents plus and minus two standard deviations related to interannual variability.
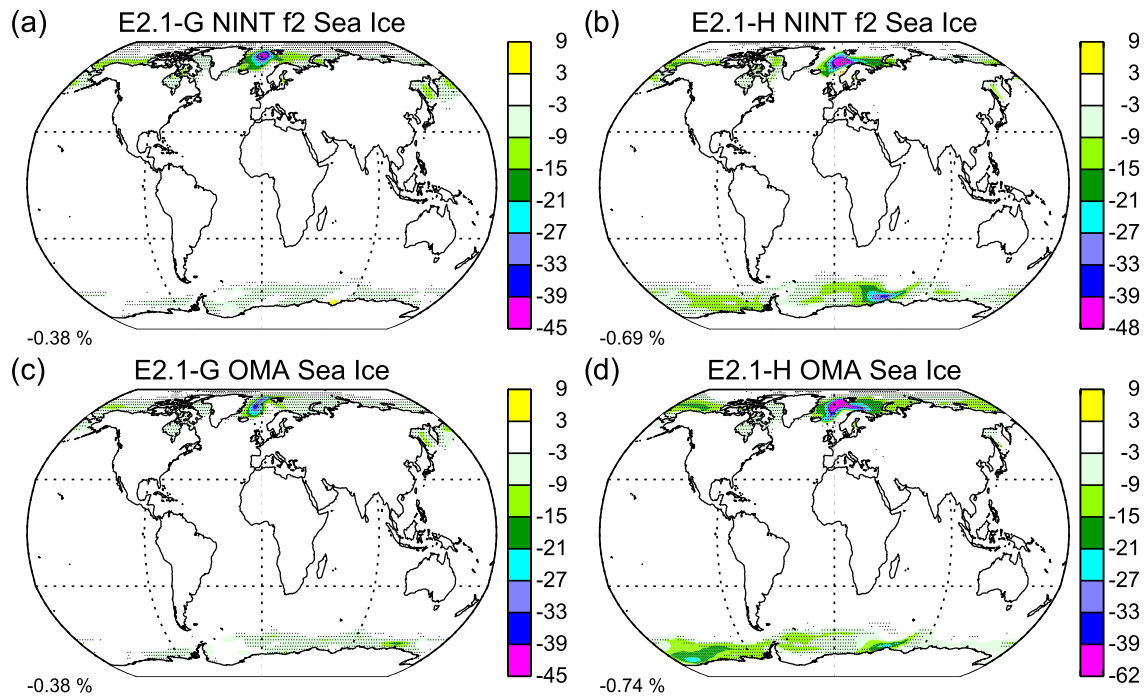
Figure 18. Change in sea ice concentration (percent) between the first and last decades of the E2.1-G and E2.1-H NINT f2 and OMA historical ensembles. (top) NINT f2, (bottom) OMA, (left) E2.1-G, (right) E2.1-H.

In the SH, melting is largest along the Antarctic coast rather than at the seasonal ice margin.

4.3. Ocean Heat Content

Ocean heat content (OHC) is proportional to temperature, but only potential temperature is available in the CMIP6 archive for all GISS-E2.1 models. We approximate OHC as the column integral of the product of mass, potential temperature and the specific heat of seawater. Potential temperature includes the effect of compression due to the weight of the overlying water column, so departures from temperature are smallest near the surface, increasing on the order of 0.1 K per kilometer (Pawlowicz, 2013). During the historical period, temperature anomalies are largely confined to the upper ocean so fractional differences in trends should be small. We use a constant value of specific heat that is independent of temperature and salinity, but chosen to minimize errors in heat content (Griffies et al., 2016; McDougall, 2003).

Ocean heat uptake by models and observations during the historical period is a few hundred ZJ (Figure 19). Direct measurements of global temperature are available only since the mid-twentieth century and mainly from above the thermocline. To evaluate the model at earlier times and over the entire column, we plot an assimilation model estimate of heat uptake that uses observed SST to infer sub-surface temperature anomalies (Zanna et al., 2019). Uptake increases steadily over the historical period with a brief hiatus following major volcanic eruptions. For most of the ensembles, this increase is relative to a non-negligible trend during the pre-industrial control experiment (Figure 3). For the E2.1-G NINT f2 ensemble, for example, the magnitude of the pre-industrial trend is half the increase during historical period, demonstrating the importance of subtracting OHC drift. Column uptake since the nineteenth century is estimated near 450 ZJ (Zanna et al., 2019), almost identical to uptake by the E2.1-H OMA ensemble (Figure 19c). Uptake for the NINT ensembles is smaller, between 350 and 380 ZJ (Figure 19a), although all GISS-E2.1 ensembles show a stronger trend in recent decades than observed.

Heat uptake in the GISS CMIP6 models is smaller and closer to the Zanna et al. (2019) estimate than the CMIP5 predecessors to E2.1-G and E2.1-H whose column uptake was 800 and 700 ZJ, respectively (Miller et al., 2014). This is consistent with the smaller total ERF for GISS-E2.1 that results from the reduction of GHG forcing (Figure 4).

Figures 19a and 19c show GISS-E2.1 model heat uptake above roughly 700 and 2,000 m. (Vertical layering in the GISS and HYCOM oceans is slightly offset from these depths.) Heat uptake within a particular
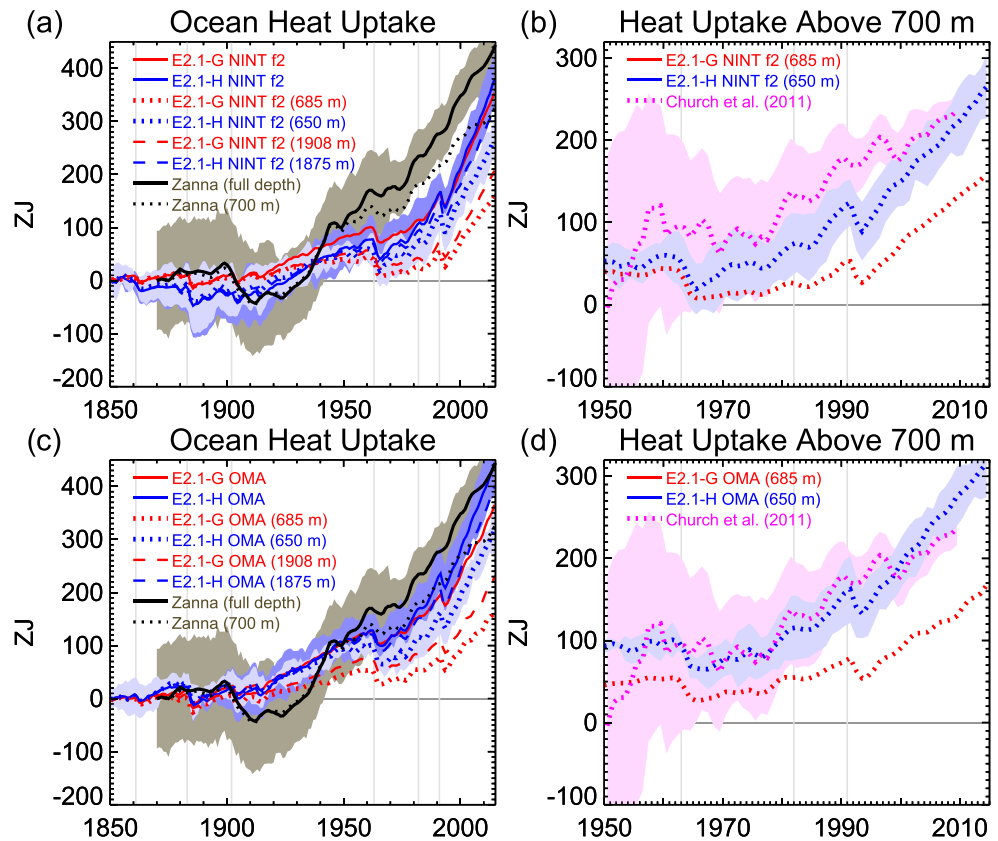
Figure 19. (top) NINT f2 and (below) OMA global ocean heat uptake in ZJ. Total column (left) and upper ocean heat uptake (right) are compared to observational estimates by Zanna et al. (2019) and Church et al. (2011), respectively. Heat uptake is defined as an anomaly relative to the start of each record. Observed upper ocean uptake is defined to have the same 1950–2013 average as the E2.1-H OMA ensemble. Shading represents plus and minus two standard deviations. Vertical gray lines show major volcanic eruptions.

depth range is given by the uptake difference between the two depths. For example, for the E2.1-G NINT f2 ensemble in 2014, uptake between 685 and 1,908 m is the difference between 160 and 210 ZJ: roughly 50 ZJ (Figure 19a). The difference between uptake by the entire column and that part above 1,908 m is about 140 ZJ, which corresponds to the uptake below this depth.

The E2.1-H ocean sequesters a larger fraction of heat near the surface compared to the E2.1-G configuration. Nearly 270 ZJ are concentrated within the upper 650 m of E2.1-H NINT f2 with an additional 60 ZJ down to roughly 2,000 m and 50 ZJ below this depth. In contrast, only about half the total column uptake within E2.1-G is within roughly 700 m of the surface. The remainder is mostly below 2,000 m, with little uptake in the intermediate layer.

Figures 19b and 19d evaluate model heat uptake above a depth near 700 m, where temperature measurements are most plentiful, based upon an update to the analysis of Domingues et al. (2008) by Church et al. (2011). Upper ocean heat uptake by 1950 at the start of the observational record is not known, so the time average of uptake between 1950 and 2009 is given the same value as the E2.1-H OMA ensemble: only observed and modeled trends are compared. Upper ocean uptake by the E2.1-H ensembles is largest and close to the observed trend; E2.1-G trends are weaker.

The regional distribution of ocean heat anomalies is shown in Figure 20 for the full column as well as the layers above roughly 700 m and below roughly 2,000 m. In the Pacific, heat uptake within E2.1-G NINT f2 is largest above the thermocline of the subtropical gyre where there is downward Ekman pumping (Figure 20c). In the North Atlantic, upper ocean heat uptake is generally positive except at the outlet of the Labrador Sea, a region of deepwater formation that is coincident with an anomalous reduction in surface air temperature (Figure 12a). We interpret cooling in this region as a consequence of weakening of
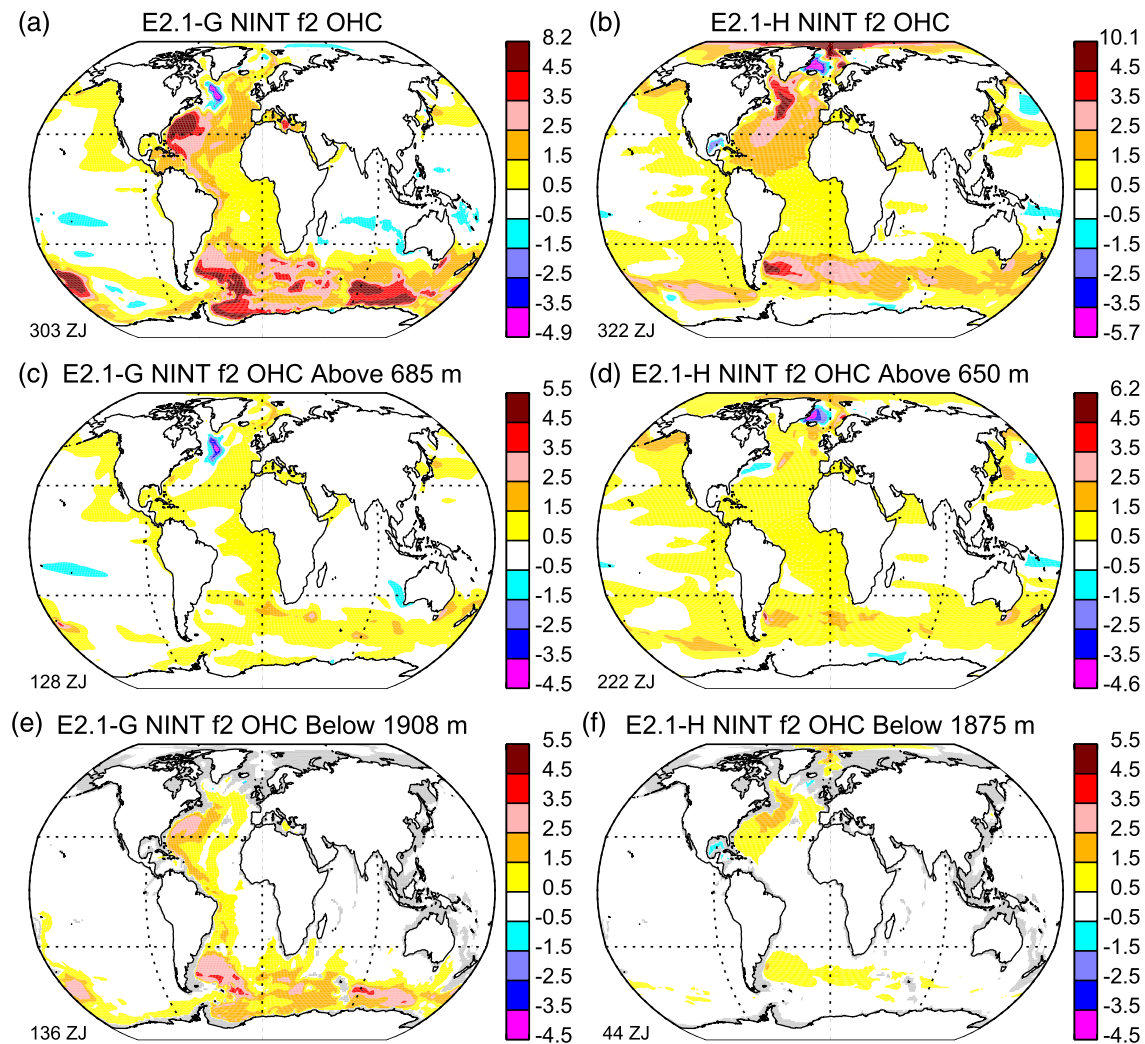
Figure 20. Change in ensemble and annual average ocean heat content ($\times 10^9\,\mathrm{J\,m^{-2}}$). between the first and last decades of the historical simulations (1850–2014) for NINT f2 models E2.1-G and E2.1-H. (a, b) Full column, (c, d) above 685 m (E2.1-G) or 650 m (E2.1-H), and (e, f) below 1,908 m (E2.1-G) or 1,850 m (E2.1-H). The value at the lower left is a global sum for that layer.

meridional overturning that begins near the end of the historical period (Figure 21), reducing the convergence of ocean heat into regions of deepwater formation. Despite this weakening, overturning continues to deliver heat below 2,000 m as seen by uptake along the deep western boundary currents adjacent to North and South America (Figure 20e). E2.1-G NINT f2 also exhibits large heat uptake below 2,000 m in the Southern Ocean.

E2.1-H NINT f2 shows more extensive warming within the upper ocean (Figure 20d). Atlantic overturning in this model shows almost no weakening during the historical period (Figure 21), consistent with positive heat uptake near the Labrador Sea and the absence of atmospheric cooling (Figures 12d and 20d). Below 2,000 m, anomalous heat is seen along the western boundary of the deep Atlantic, equatorward of the Labrador Sea, but uptake is smaller than in E2.1-G and elsewhere there is little heat in this layer.

The greater heat uptake by E2.1-H and its confinement to the upper ocean is consistent with the model's greater TCR (Table 5). Lerner et al. (2020) show that the mixed layer in E2.1-G is too deep in the North Atlantic, suggesting that there is excessive transport to the deep ocean. The greater fraction of heat storage below 700 m in E2.1-G is consistent with its CMIP5 E2-R predecessor, despite an incorrect implementation of isopycnal diffusion in the earlier model that resulted in excessive downward transport across isopycnals (Miller et al., 2014). The contrasting vertical extent of heat invasion common to both the CMIP5 and CMIP6
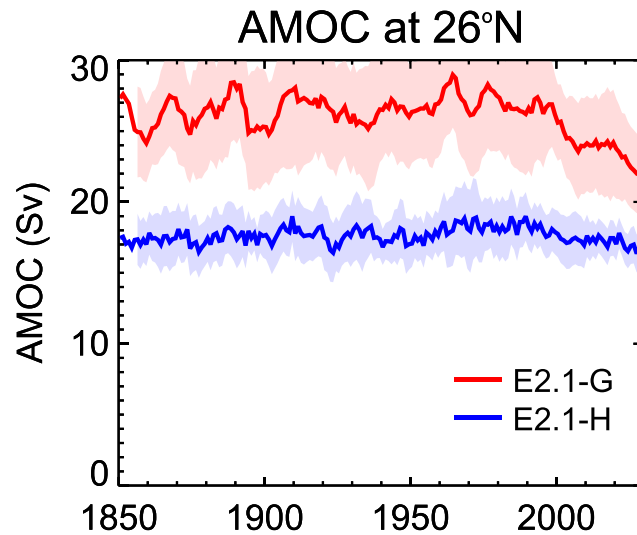
Figure 21. The Atlantic Meridional Overturning Circulation at 26°N (Sv) equal to the maximum value of the NINT f2 ensemble-mean meridional stream function. Measurements by the RAPID-MOCHA array indicate overturning of 18 Sv for the period 2004–2018 (Smeed et al., 2019).

generations of GISS ocean models suggests that this error was not central; the contrasting depths of heat uptake reflect a more fundamental difference in the response of each ocean model to radiative forcing.

4.4. Sea Level

Sea level rises both through steric and eustatic effects. Steric rise results from ocean expansion, equivalent to a decrease in seawater density, either through warming or freshening. The latter occurs through melting of sea ice, for example. Eustatic sea level rise results from an increase of ocean mass as through the melting of continental ice. Conversely, eustatic fall results from an onshore transport of water vapor and sequestration of precipitation in lakes, soil and snow fields. E2.1-G has a mass-conserving ocean whose sea level trends are diagnosed exactly. The E2.1-H ocean is volume conserving so sea level is computed diagnostically through linearization of the seawater equation of state and the excess of precipitation and runoff over evaporation. (The E2.1-H eustatic value was incorrectly diagnosed, so only the steric value is available.)

Since posting output from the historical ensembles to the CMIP6 archive, we discovered an error in the initialization of lake volume that caused a spurious sea level decline. We corrected the initialization and created an extra (eleventh) member of the E2.1-G NINT f2 ensemble. (Sea level shows little variation among members, so we considered a single corrected realization to be sufficient to diagnose historical changes.)

The total change in sea level during the historical period for the corrected model is shown in Figures 22a and 22b. The model and observed sea level are assumed to have identical averages during their period of overlap, so only changes are comparable. Model sea level is fairly uniform until the early twentieth century; thereafter it rises 26 cm by the end of the historical period. Over the entire historical period, the total model rise is close to the observed rise estimated since 1880 (Church & White, 2011). However, the recent model trend exceeds that computed from a combination of tide gauges and satellite altimetry (Domingues et al., 2008; Church et al., 2011), and the attribution is different as described below.

The thermosteric contribution to sea level rise is shown in Figures 22c and 22d. Because this is related to thermal changes during the historical period, we assume that the thermosteric trend from the ensembles with incorrect lake initialization remain informative. For E2.1-G, the thermosteric contribution is 2 cm over the historical period; this contribution is 2 times larger for both the E2.1-H NINT f2 and OMA ensembles, a contrast that is consistent with each model's greater surface warming and ocean heat uptake (Figures 11, 19a, and 19c). (The E2.1-H value is taken from the total steric value with the assumption that salinity effects are secondary.)

The model thermosteric contribution over the entire historical period is comparable to that estimated from observations since 1961 by Church et al. (2011). The observed thermosteric rise contributes slightly less than
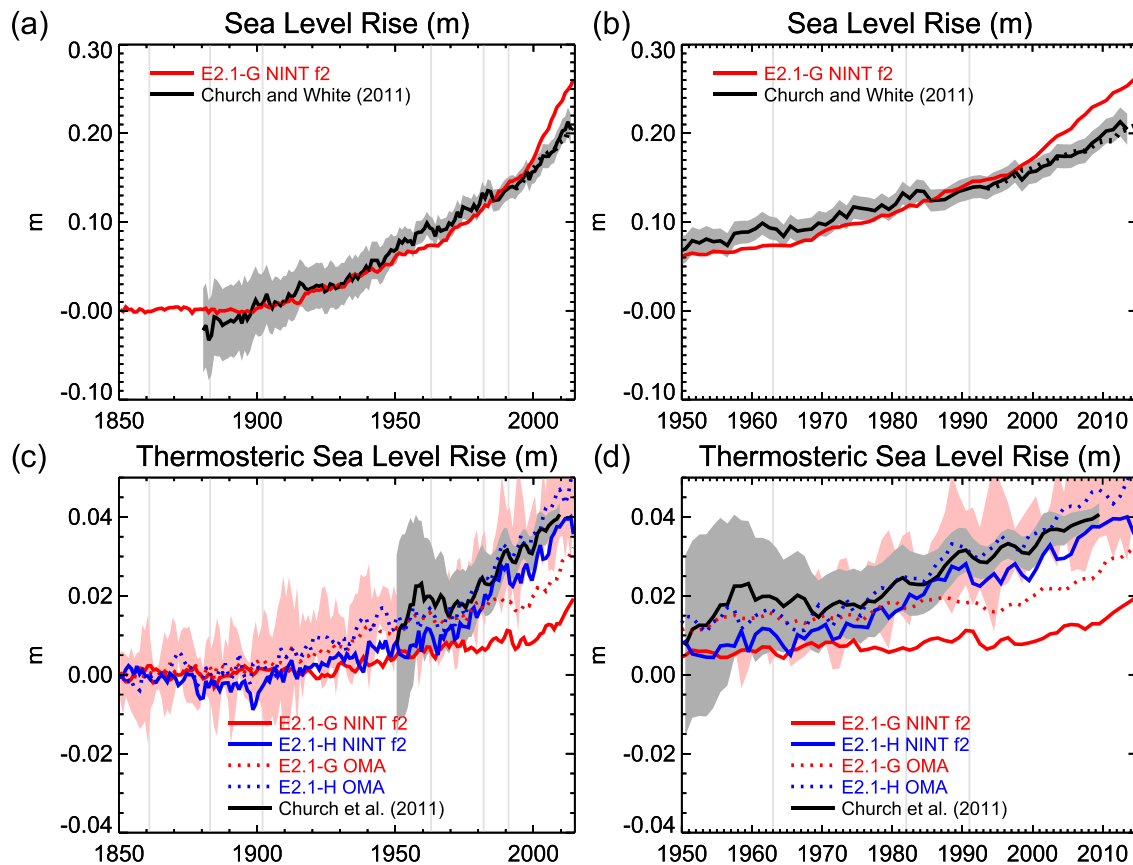
Figure 22. (top) Change in global annual sea level (m) for an added (eleventh) member of the E2.1-G NINT f2 historical ensemble. (Sea level in the original members of the GISS-E2.1 ensembles is corrupted by incorrect initialization of lake volume.) The observed record of Church and White is in black: gauge data are solid, altimeter data are dotted. The gauge data are offset to have the same average as the E2.1-G NINT f2 model during their period of overlap (1880–2013). The altimeter data are given the same offset to preserve its relation to the gauge measurements. (bottom) The thermosteric contribution to sea level rise in the NINT f2 and OMA ensembles. The gauge measurements are offset to match the E2.1-H OMA model during their period of overlap (1950–2009). Thermosteric rise in E2.1-H includes salinity effects that are assumed to be small. Vertical gray lines show major volcanic eruptions.

half of the total increase during recent decades. Most of the remaining observed rise is eustatic and attributed to melting of continental ice sheets like Greenland and Antarctica along with mountain glaciers.

Land ice mass is fixed in GISS-E2.1, so it cannot contribute to the eustatic rise in Figure 22. In the model, the main contribution to the total is eustatic (Figure 23a), but in contrast to the observations, the source of the eustatic rise is irrigation (Figure 23b). When there is insufficient lake and river water to supply the prescribed irrigation demand, water is drawn from an aquifer of unlimited volume. The aquifer draw increases the mass of total model water and causes a rise in sea level. Over the period 1980–2009, this rise is 7 cm, an order of magnitude larger than the observationally based estimate of Grogan et al. (2017). Figure 23b shows that other contributions to eustatic change like snowmelt in the warming climate or groundwater sequestration would be more important in comparison to a realistic aquifer draw.

4.5. Precipitation

The change in global precipitation over the historical period is shown in Figure 24. For all model ensembles, precipitation increases steadily during the final half century of the historical period except in the aftermath of a major volcanic eruption. Any observed trend is obscured by interannual variations in the Global Precipitation and Climatology Project (GPCP) analysis (Adler et al., 2003, 2018), where extrema track El Niño and La Niña events whose influence is diminished in the models by ensemble averaging. Nonetheless, Adler et al. (2017) find no significant GPCP trend after removing the influence of ENSO and volcanos. This discrepancy may result from a large unforced signal in the remaining variability or limitations of the models.
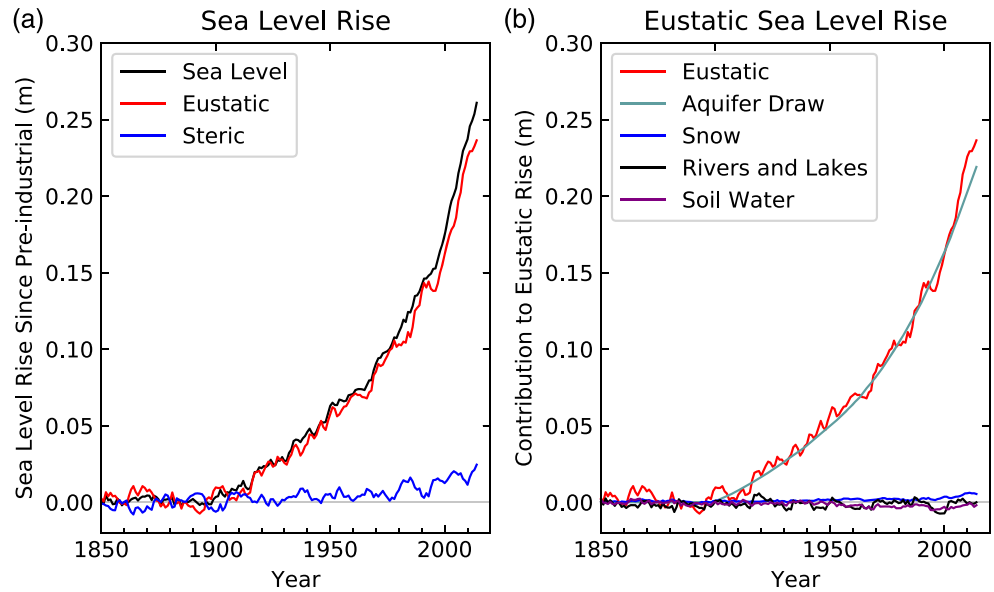
Figure 23. (a) Sea level rise in a supplementary (eleventh) member of the E2.1-G NINT f2 historical ensemble decomposed into eustatic and steric contributions. (b) Eustatic contributions from aquifer drawdown and snowmelt along with storage in rivers, lakes, and soil.

The largest model increase is associated with E2.1-H ensembles that show the greatest tropospheric warming (Figures 11 and 15). The precipitation increase results from offsetting changes by GHGs and aerosols. The former increase tropospheric radiative divergence (balanced mainly by additional latent heating associated with precipitation); the latter dim the surface and reduce evaporation (Allen & Ingram, 2002; Andrews et al., 2009; Miller et al., 2004; Samset et al., 2016). To isolate the effect of GHGs, we regressed precipitation with the temperature increase following an abrupt quadrupling of CO_2. The initial rapid adjustment of precipitation is negative but followed by a precipitation increase of 2.3% for each degree K of warming. Both the adjustment and the sensitivity of precipitation to warming are three-quarters of the corresponding values for the GISS CMIP5 ensembles. This is also the ratio of Year 2000 GHG forcing between the GISS CMIP6 and CMIP5 configurations (Figure 4). We have attributed the CMIP6 reduction of GHG forcing in GISS-E2.1
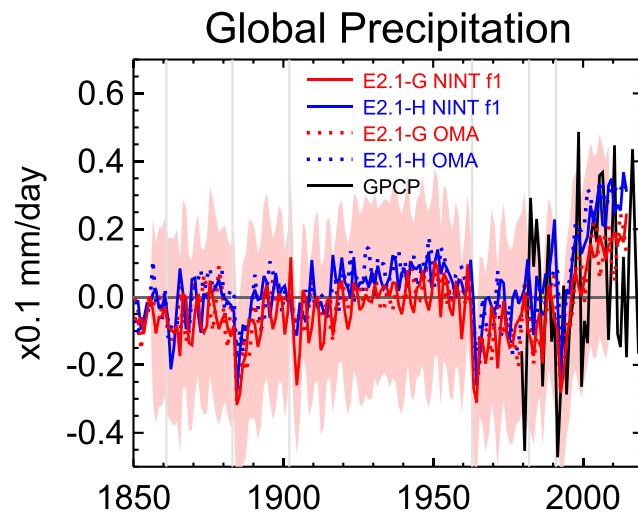

Figure 24. Ensemble and global annual precipitation ($\times 0.1$ mm day^{-1}) during the historical ensembles (1850–2014). Pink shading marks plus and minus two standard deviations of intra-ensemble variability of the E2.1-G NINT f2 configuration. This variability is smoothed using an 11-year window that damps periods less than a decade. The vertical gray lines mark volcanic eruptions. The GPCP anomaly is constructed to have the same time average as the E2.1-G NINT f2 ensemble between 1979 and 2014.

to its greater pre-industrial opacity to longwave radiation (Figure 7). The increase in opacity and decrease of forcing is also consistent with the reduction in the rapid adjustment of precipitation following the abrupt CO_2 increase and the subsequent increase of precipitation with warming.

5. Conclusions

We describe the GISS-E2.1 response to forcing during the CMIP6 historical period (1850–2014) using distinct atmospheric physics versions coupled to two different ocean models.

As observed, each model ensemble shows roughly 1 K of warming at the surface during the historical period with cooling in the stratosphere. The choice of ocean model is important to the simulated climate trends. For the specific period of 1979–2014, surface warming in the E2.1-G ensembles is consistent with observations, but warming in E2.1-H is marginally too rapid. The E2.1-G ensembles warm more slowly by exporting a greater fraction of ocean heat beneath the thermocline, a contrast that was also exhibited by the CMIP5 E2-R and E2-H configurations. While confinement of anomalous heat within the upper ocean in the E2.1-H ensembles is in better agreement with an estimate of OHC based upon historical SST (Zanna et al., 2019), more observations of heat uptake below the thermocline are needed to evaluate the models. Only the E2.1-G ensembles are able to reproduce a distinctive feature of the observed surface temperature: namely, a slight cooling in the North Atlantic that in the ensembles is associated with a recent reduction in the rate of meridional overturning.

Resemblance of the NINT trends to observed values is aided by the calibration of the aerosol indirect effect to $-1\,\mathrm{W\,m^{-2}}$, consistent with current estimates based upon models constrained by observations (Myhre et al., 2013). However, the OMA model versions warm similarly while calculating the first indirect effect with a physically based parameterization. Preliminary experiments with ModelE3 that include a physically based second indirect effect show that model climate sensitivity is only slightly increased and remains within the canonical range (Sherwood et al., 2020). The dominant indirect contribution to aerosol forcing shows the importance of continued efforts to represent interactions between aerosols and clouds mechanistically.

Warming of the GISS-E2.1 ensembles is reduced compared to the GISS CMIP5 simulations, despite the higher TCR of the former. This reduction is due to smaller total forcing in the GISS-E2.1 configurations that follows the reduction of GHG forcing by one-quarter. We attribute this reduction to greater longwave opacity of the GISS-E2.1 pre-industrial climate, which has higher water vapor concentration and greater cloud longwave forcing than the GISS CMIP5 model (Figure 7). This demonstrates the importance of the unperturbed climate to forcing and forced climate trends. The LW opacity increase and forcing decrease is also related to a lower sensitivity of precipitation to warming in GISS-E2.1 compared to the CMIP5 GISS-E2.

Comparison of multiple model configurations allows the importance of different climate processes to be examined. The NINT f2 ensemble corrects an error in the prescription of volcanic aerosol surface area available for heterogenous chemical reactions. Comparison with the NINT f1 ensembles shows the effect of these reactions upon ozone and the associated temperature variations in the lower stratosphere following a volcanic eruption (Figure 13). Comparison of the eruption response across the historical period demonstrates that whether ozone is destroyed or enhanced by these reactions depends upon the concentration of CFCs.

Climate forcing by irrigation was introduced into GISS-E2.1 after CMIP5. Irrigation is important for simulating regional trends, especially those associated with the Asian Summer Monsoon (Cook et al., 2011; Puma & Cook, 2010; Shukla et al., 2014). Aspects of the irrigation flux are not well-constrained by observations, including its decomposition into contributions from surface reservoirs like lakes as opposed to aquifers that are replenished over geologic durations that are long compared to the historical period. We show that sea level rise in the GISS historical ensembles is comparable to the observed change (Church & White, 2011), but results almost entirely from an unrealistically large draw upon aquifer fossil water. This shows how coupled models can reveal unexpected limitations of atmosphere-only representations, guiding model development and providing indirect observational constraints upon forcing, in this case, constraining irrigation rates with observed sea level changes.

Model comparison to observed climate trends is a first step to understanding these changes and assessing the reliability of predictions of future change. The overall agreement of GISS-E2.1 with observed trends is familiar from evaluation of its predecessors, as is the conclusion that these trends are almost entirely

anthropogenic in origin. This agreement points to larger climate change and disruption in the coming decades, even when assessed by a model with mid-range sensitivity to forcing like GISS-E2.1.

Data Availability Statement

CMIP6 standard variables analyzed in this study are available from the Earth System Grid Federation. Additional variables like MSU temperatures and OHC can be downloaded online (from https://portal.nccs. nasa.gov/GISS_modelE/E2.1). Table 4 provides links to the measurements and retrievals used to evaluate the models.

References

Abdul-Razzak, H., & Ghan, S. J. (2000). A parameterization of aerosol activation: 2. Multiple aerosol types. *Journal of Geophysical Research*, *105*(D5), 6837–6844. https://doi.org/10.1029/1999JD901161

Abdul-Razzak, H., Ghan, S. J., & Rivera-Carpio, C. (1998). A parameterization of aerosol activation: 1. Single aerosol type. *Journal of Geophysical Research*, *103*(D6), 6123–6131. https://doi.org/10.1029/97JD03735

Adler, R. F., Gu, G., Sapiano, M., Wang, J. J., & Huffman, G. J. (2017). Global precipitation: Means, variations and trends during the satellite era (1979–2014). *Surveys in Geophysics*, *38*(4), 679–699. https://doi.org/10.1007/s10712-017-9416-4

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P. P., Janowiak, J., & Nelkin, E.(2003). The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology*, *4*(6), 1147–1167. https://doi.org/10.1175/1525-7541(2003)004<1147:tvgpcp>2.0.co;2

Adler, R., Sapiano, M., Huffman, G., Wang, J. J., Gu, G., Bolvin, D., & Shin, D. B. (2018). The Global Precipitation Climatology Project (GPCP) monthly analysis (New Version 2.3) and a review of 2017 global precipitation. *Atmosphere*, *9*(4), 138. https://doi.org/10.3390/atmos9040138

Albrecht, B. A. (1989). Aerosols, cloud microphysics, and fractional cloudiness. *Science*, *245*, 1227–1230.

Allen, M. R., & Ingram, W. J. (2002). Constraints on future changes in climate and the hydrologic cycle. *Nature*, *419*, 224–232. https://doi.org/10.1038/nature01092

Andrews, T., Forster, P. M., & Gregory, J. M. (2009). A surface energy perspective on climate change. *Journal of Climate*, *22*(10), 2557–2570. https://doi.org/10.1175/2008JCLI2759.1

Arfeuille, F., Weisenstein, D., Mack, H., Rozanov, E., Peter, T., & Brönnimann, S. (2014). Volcanic forcing for climate modeling: A new microphysics-based data set covering years 1600–present. *Climate of the Past*, *10*(1), 359–375. https://doi.org/10.5194/cp-10-359-2014

Bauer, S. E., Balkanski, Y., Schulz, M., Hauglustaine, D. A., & Dentener, F. (2004). Global modelling of heterogeneous chemistry on mineral aerosol surfaces: The influence on tropospheric ozone chemistry and comparison to observations. *Journal of Geophysical Research*, *109*, D02304. https://doi.org/10.1029/2003JD003868

Bauer, S. E., Bausch, A., Nazarenko, L., Tsigaridis, K., Xu, B., Edwards, R., & McConnell, J. (2013). Historical and future black carbon deposition on the three ice caps: Ice core measurements and model simulations from 1850 to 2100. *Journal of Geophysical Research: Atmospheres*, *118*, 7948–7961. https://doi.org/10.1002/jgrd.50612

Bauer, S. E., & Koch, D. (2005). Impact of heterogeneous reactions between sulfate and mineral dust aerosols on climate in the GISS GCM. *Journal of Geophysical Research*, *110*, D17202. https://doi.org/10.1029/2005JD005870

Bauer, S. E., Mishchenko, M. I., Lacis, A. A., Zhang, S., Perlwitz, J., & Metzger, S. M. (2007). Do sulfate and nitrate coatings on mineral dust have important effects on radiative properties and climate modeling? *Journal of Geophysical Research*, *112*, D06307. https://doi.org/10.1029/2005JD006977

Bauer, S. E., Tsigaridis, K., Faluvegi, G., Kelley, M., Lo, K. K., Miller, R. L., & Wu, J. (2020). Historical (1850–2014) aerosol evolution and role on climate forcing using the GISS ModelE2.1 contribution to CMIP6. *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS001978. https://doi.org/10.1029/2019MS001978

Bauer, S. E., Wright, D. L., Koch, D., Lewis, E. R., McGraw, R., Chang, L. S., & Ruedy, R. (2008). MATRIX (Multiconfiguration Aerosol TRacker of mIXing state): An aerosol microphysical module for global atmospheric models. *Atmospheric Chemistry and Physics*, *8*, 6003–6035. https://doi.org/10.5194/acp-8-6003-2008

Berger, A. L. (1978). Long-term variations of daily insolation and quaternary climatic changes. *Journal of the Atmospheric Sciences*, *35*, 2362–2367.

Blumenthal, M. B. (1991). Predictability of a coupled ocean–atmosphere model. *Journal of Climate*, *4*(8), 766–784. https://doi.org/10.1175/1520-0442(1991)004<0766:poacom>2.0.co;2

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., & Zhang, X. (2013). Clouds and aerosols. In T. Stocker (Ed.), *Climate change 2013—The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 571–658). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. https://doi.org/10.1017/cbo9781107415324.016

Brennan, M. K., Hakim, G. J., & Blanchard-Wrigglesworth, E. (2020). Arctic sea-ice variability during the instrumental era. *Geophysical Research Letters*, *47*, e2019GL086843. https://doi.org/10.1029/2019GL086843

Cane, M. A., Clement, A. C., Kaplan, A., Kushnir, Y., Pozdnyakov, D., Seager, R., & Murtugudde, R. (1997). Twentieth-century sea surface temperature trends. *Science*, *275*(5302), 957–960. https://doi.org/10.1126/science.275.5302.957

Church, J. A., & White, N. (2011). Sea-level rise from the late 19th to the early 21st century. *Surveys in Geophysics*, *32*, 585–602. https://doi.org/10.1007/s10712-011-9119-1

Church, J. A., White, N. J., Konikow, L. F., Domingues, C. M., Cogley, J. G., Rignot, E., & Velicogna, I. (2011). Revisiting the Earth's sea-level and energy budgets from 1961 to 2008. *Geophysical Research Letters*, *38*, L18601. https://doi.org/10.1029/2011GL048794

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, *83*, 596–610. https://doi.org/10.2307/2289282

Cook, B., Puma, M., & Krakauer, N. (2011). Irrigation induced surface cooling in the context of modern and increased greenhouse gas forcing. *Climate Dynamics*, *37*, 1587–1600. https://doi.org/10.1007/s00382-010-0932-x

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

Dai, A., Luo, D., Song, M., & Liu, J. (2019). Arctic amplification is caused by sea-ice loss under increasing CO_2. *Nature Communications*, *10*(1), 121. https://doi.org/10.1038/s41467-018-07954-9

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

Domingues, C., Church, J., White, N., Gleckler, P., Wijffels, S., Barker, P., & Dunn, J.(2008). Improved estimates of upper-ocean warming and multi-decadal sea-level rise. *Nature*, *453*, 1090–1094. https://doi.org/10.1038/nature07080

Dufresne, J. L., Gautier, C., Ricchiazzi, P., & Fouquart, Y. (2002). Longwave scattering effects of mineral aerosols. *Journal of the Atmospheric Sciences*, *59*, 1959–1966.

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

Fetterer, F., Knowles, K., Meier, W., Savoie, M., & Windnagel, A. (2017). Sea Ice Index, Version 3. NSIDC, https://doi.org/10.7265/N5K072F8

Flanner, M. G., Zender, C. S., Randerson, J. T., & Rasch, P. J. (2007). Present-day climate forcing and response from black carbon in snow. *Journal of Geophysical Research*, *112*, D11202. https://doi.org/10.1029/2006JD008003

Forster, P. M., Richardson, T., Maycock, A. C., Smith, C. J., Samset, B. H., Myhre, G., & Schulz, M. (2016). Recommendations for diagnosing effective radiative forcing from climate models for CMIP6. *Journal of Geophysical Research*, *121*, 12,460–12,475. https://doi.org/10.1002/2016JD025320

Fröhlich, C. (2006). Solar irradiance variability since 1978. *Space Science Reviews*, *125*(1-4), 53–65. https://doi.org/10.1007/s11214-006-9046-5

Gillett, N. P., Akiyoshi, H., Bekki, S., Braesicke, P., Eyring, V., Garcia, R., & Shibata, K. (2011). Attribution of observed changes in stratospheric ozone and temperature. *Atmospheric Chemistry and Physics*, *11*(2), 599–609. https://doi.org/10.5194/acp-11-599-2011

Ginoux, P., Prospero, J. M., Gill, T. E., Hsu, N., & Zhao, M. (2012). Global scale attribution of anthropogenic and natural dust sources and their emission rates based on MODIS Deep Blue Aerosol Products. *Reviews of Geophysics*, *50*, RG3005. https://doi.org/10.1029/2012RG000388

Golledge, N. R., Keller, E. D., Gomez, N., Naughten, K. A., Bernales, J., Trusel, L. D., & Edwards, T. L. (2019). Global environmental consequences of twenty-first-century ice-sheet melt. *Nature*, *566*(7742), 65–72. https://doi.org/10.1038/s41586-019-0889-9

Goody, R. M., & Yung, Y. L. (1989). *Atmospheric radiation: Theoretical basis*. New York: Oxford Univ. Press.

Griffies, S. M., Danabasoglu, G., Durack, P. J., Adcroft, A. J., Balaji, V., Böning, C. W., & Yeager, S. G. (2016). OMIP contribution to CMIP6: Experimental and diagnostic protocol for the physical component of the Ocean Model Intercomparison Project. *Geoscientific Model Development*, *9*(9), 3231–3296. https://doi.org/10.5194/gmd-9-3231-2016

Grogan, D. S., Wisser, D., Prusevich, A., Lammers, R. B., & Frolking, S. (2017). The use and re-use of unsustainable groundwater for irrigation: A global budget. *Environmental Research Letters*, *12*(3), 034017. https://doi.org/10.1088/1748-9326/aa5fb2

Hansen, J., Fung, I., Lacis, A., Rind, D., Lebedeff, S., Ruedy, R., & Stone, P. (1988). Global climate changes as forecast by Goddard Institute for Space Studies three-dimensional model. *Journal of Geophysical Research*, *93*, 9341–9364. https://doi.org/10.1029/JD093id08p09341

Hansen, J., & Nazarenko, L. (2003). Soot climate forcing via snow and ice albedos. *Proceedings of the National Academy of Sciences*, *101*(2), 423–428. https://doi.org/10.1073/pnas.2237157100

Hansen, J. E., Sato, M., Nazarenko, L., Ruedy, R., Lacis, A., Koch, D., & Christy, J. (2002). Climate forcings in Goddard Institute for Space Studies SI2000 simulations. *Journal of Geophysical Research*, *107*(D18), 4347. https://doi.org/10.1029/2001JD001143

Hansen, J., Sato, M., & Ruedy, R. (1997). Radiative forcing and climate response. *Journal of Geophysical Research*, *102*, 6831–6864.

Hansen, J., Sato, M., Ruedy, R., Kharecha, P., Lacis, A., Miller, R., & Zhang, S. (2007). Climate simulations for 1880–2003 with GISS modelE. *Climate Dynamics*, *29*, 661–696. https://doi.org/10.1007/s00382-007-0255-8

Hansen, J. E., Sato, M., Ruedy, R., Lacis, A., Asamoah, K., Beckford, K., & Zawodny, J. (1997). Forcings and chaos in interannual to decadal climate change. *Journal of Geophysical Research*, *102*, 25,679–25,720.

Hansen, J. E., Sato, M., Ruedy, R., Lacis, A., & Glascoe, J. (1998). Global climate data and models: A reconciliation. *Science*, *281*, 930–932. https://doi.org/10.1126/science.281.5379.930

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., & Zhang, S. (2005). Efficacy of climate forcings. *Journal of Geophysical Research*, *110*, D18104. https://doi.org/10.1029/2005JD005776

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., & Zhang, Q. (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. https://doi.org/10.5194/gmd-11-369-2018

Hofmann, D. J., & Oltmans, S. J. (1993). Anomalous Antarctic ozone during 1992: Evidence for Pinatubo volcanic aerosol effects. *Journal of Geophysical Research*, *98*(D10), 18,555–18,561. https://doi.org/10.1029/93JD02092

Holland, M. M., & Bitz, C. M. (2003). Polar amplification of climate change in coupled models. *Climate Dynamics*, *21*, 221–232. https://doi.org/10.1007/s00382-003-0332-6

Huang, Y., & Shahabadi, M. B. (2014). Why logarithmic? A note on the dependence of radiative forcing on gas concentration. *Journal of Geophysical Research: Atmospheres*, *119*, 13,683–13,689. https://doi.org/10.1002/2014JD022466

Ito, G., Romanou, A., Kiang, N. Y., Faluvegi, G., Aleinov, I., Ruedy, R., & Lo, K. (2020). Global carbon cycle and climate feedbacks in the NASA GISS ModelE2.1. *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS002030. https://doi.org/10.1029/2019MS002030

Jiang, J., Cameron, R. H., & Schüssler, M. (2015). The cause of the weak solar cycle 24. *The Astrophysical Journal*, *808*(1), L28. https://doi.org/10.1088/2041-8205/808/1/l28

Kelley, M., Schmidt, G. A., Nazarenko, L. S., Bauer, S. E., Ruedy, R., Russell, G. L., & Yao, M. S. (2020). GISS-E2.1: Configurations and climatology. *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS002025. https://doi.org/10.1029/2019MS002025

Kim, D., Sobel, A. H., Genio, A. D. D., Chen, Y., Camargo, S. J., Yao, M. S., & Nazarenko, L. (2012). The tropical subseasonal variability simulated in the NASA GISS General Circulation Model. *Journal of Climate*, *25*(13), 4641–4659. https://doi.org/10.1175/JCLI-D-11-00447.1

Latto, R., & Romanou, A. (2018). The Ocean Carbon States Database: A proof-of-concept application of cluster analysis in the ocean carbon cycle. *Earth System Science Data*, *10*(1), 609–626. https://doi.org/10.5194/essd-10-609-2018

LeTexier, H., Solomon, S., & Garcia, R. (1988). The role of molecular-hydrogen and methane oxidation in the water-vapor budget of the stratosphere. *Quarterly Journal of the Royal Meteorological Society*, *114*, 281–295. https://doi.org/10.1002/qj.49711448002

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

Lerner, P., Romanou, A., Kelley, M., Romanski, J., Ruedy, R., & Russell, G. (2020). Drivers of air-sea CO_2 flux seasonality and its long-term changes in the NASA-GISS model CMIP6 submission. *Journal of Advances in Modeling Earth Systems*. https://doi.org/10.1029/2019MS002028

Levitus, S., & Boyer, T. P. (1994). World Ocean Atlas 1994, Volume 4: Temperature. NOAA Atlas NESDIS 4, Washington D. C., US Govt Printing Office.

Levitus, S., Burgett, R., & Boyer, T. P. (1994). World Ocean Atlas 1994, Volume 3: Salinity. NOAA Atlas NESDIS 3, Washington D. C., US Govt Printing Office.

Lohmann, U., Stier, P., Hoose, C., Ferrachat, S., Kloster, S., Roeckner, E., & Zhang, J. (2007). Cloud microphysics and aerosol indirect effects in the global climate model ECHAM5-HAM. *Atmospheric Chemistry and Physics*, *7*(13), 3425–3446. https://doi.org/10.5194/acp-7-3425-2007

Mahowald, N. M., Rivera, G. D. R., & Luo, C. (2004). Comment on "Relative importance of climate and land use in determining present and future global soil dust emission" by I. Tegen et al. *Geophysical Research Letters*, *31*, L24105. https://doi.org/10.1029/2004GL021272

Manabe, S., & Weatherald, R. T. (1967). Thermal equilibrium of the atmosphere with a given distribution of relative humidity. *Journal of the Atmospheric Sciences*, *24*, 241–259.

Marvel, K., Schmidt, G. A., Miller, R. L., & Nazarenko, L. S. (2015). Implications for climate sensitivity from the response to individual forcings. *Nature Climate Change*, *6*(4), 386–389. https://doi.org/10.1038/nclimate2888

Matthes, K., Funke, B., Andersson, M. E., Barnard, L., Beer, J., Charbonneau, P., & Versick, S. (2017). Solar forcing for CMIP6 (v3.2). *Geoscientific Model Development*, *10*(6), 2247–2302. https://doi.org/10.5194/gmd-10-2247-2017

McCormick, M. P., Thomason, L. W., & Trepte, C. R. (1995). Atmospheric effects of the Mt Pinatubo eruption. *Nature*, *373*(6513), 399–404. https://doi.org/10.1038/373399a0

McDougall, T. J. (2003). Potential enthalpy: A conservative oceanic variable for evaluating heat content and heat fluxes. *Journal of Physical Oceanography*, *33*(5), 945–963. https://doi.org/10.1175/1520-0485(2003)033<0945:peacov>2.0.co;2

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

Meehl, G. A., Boer, G. J., Covey, C., Latif, M., & Stouffer, R. J. (1997). Intercomparison makes for a better climate model. *Eos, Transactions American Geophysical Union*, *78*(41), 445. https://doi.org/10.1029/97EO00276

Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., & Weiss, R. (2017). Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geoscientific Model Development*, *10*(5), 2057–2116. https://doi.org/10.5194/gmd-10-2057-2017

Menon, S., & Rotstayn, L. (2006). The radiative influence of aerosol effects on liquid-phase cumulus and stratiform clouds based on sensitivity studies with two climate models. *Climate Dynamics*, *27*(4), 345–356. https://doi.org/10.1007/s00382-006-0139-3

Miller, R. L., Schmidt, G. A., Nazarenko, L., Tausnev, N., Ruedy, R., Kelley, M., & Zhang, J.(2014). CMIP5 historical simulations (1850–2012) with GISS ModelE2. *Journal of Advances in Modeling Earth Systems*, *6*, 441–477. https://doi.org/10.1002/2013MS000266

Miller, R. L., Tegen, I., & Perlwitz, J. (2004). Surface radiative forcing by soil dust aerosols and the hydrologic cycle. *Journal of Geophysical Research*, *109*, D04203. https://doi.org/10.1029/2003JD004085

Muthers, S., Arfeuille, F., Raible, C. C., & Rozanov, E. (2015). The impacts of volcanic aerosol on stratospheric ozone and the Northern Hemisphere polar vortex: Separating radiative-dynamical changes from direct effects due to enhanced aerosol heterogeneous chemistry. *Atmospheric Chemistry and Physics*, *15*(20), 11,461–11,476. https://doi.org/10.5194/acp-15-11461-2015

Myhre, G., Shindell, D., Breon, F. M., Collins, W., Fuglestvedt, J., Huang, J., & Zhang, H.(2013). Anthropogenic and natural radiative forcing. In T. Stocker (Ed.), *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 659–740). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.018

Nazarenko, L. S., Schmidt, G. A., Miller, R., Tausnev, N., Kelley, M., Ruedy, R., & Zhang, J. (2015). Future climate change under RCP emission scenarios with the GISS ModelE2. *Journal of Advances in Modeling Earth Systems*, *7*, 244–267. https://doi.org/10.1002/2014MS000403

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

Pawlowicz, R. (2013). Key physical variables in the ocean: Temperature, salinity, and density. *Nature Education Knowledge*, *4*, 13.

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

Puma, M. J., & Cook, B. I. (2010). Effects of irrigation on global climate during the 20th century. *Journal of Geophysical Research*, *115*, D16120. https://doi.org/10.1029/2010JD014122

Rayner, N. A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002JD002670

Redi, M. H. (1982). Oceanic isopycnal mixing by coordinate rotation. *Journal of Physical Oceanography*, *12*(10), 1154–1158. https://doi.org/10.1175/1520-0485(1982)012<1154:oimbcr>2.0.co;2

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., & Tavoni, M.(2017). The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, *42*, 153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Rind, D., Orbe, C., Jonas, J., Nazarenko, L., Zhou, T., Kelley, M., & Schmidt, G. (2020). GISS Model E2.2: A climate model optimized for the middle atmosphere—Model structure, climatology, variability, and climate sensitivity. *Journal of Geophysical Research: Atmospheres*, *125*, e2019JD032204. https://doi.org/10.1029/2019JD032204

Romanou, A., Gregg, W., Romanski, J., Kelley, M., Bleck, R., Healy, R., & Tausnev, N. (2013). Preindustrial control simulations of the ocean-atmosphere CO_2 exchanges in the NASA-GISS climate model: Sensitivity to physical ocean model. *Ocean Modelling*, *66*, 26–44. https://doi.org/10.1016/j.ocemod.2013.01.008

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

Rottman, G. (2005). The SORCE mission. *Solar Physics*, *230*(1–2), 7–25. https://doi.org/10.1007/s11207-005-8112-6

Rye, C. D., Garabato, A. C. N., Holland, P. R., Meredith, M. P., Nurser, A. J. G., Hughes, C. W., & Webb, D. J. (2014). Rapid sea-level rise along the Antarctic margins in response to increased glacial discharge. *Nature Geoscience*, *7*(10), 732–735. https://doi.org/10.1038/ngeo2230

Samset, B. H., Myhre, G., Forster, P. M., Hodnebrog, Ø., Andrews, T., Faluvegi, G., & Voulgarakis, A.(2016). Fast and slow precipitation responses to individual climate forcers: A PDRMIP multimodel study. *Geophysical Research Letters*, *43*, 2782–2791. https://doi.org/10.1002/2016GL068064

Santer, B., Hnilo, J., Wigley, T., Boyle, J., Doutriaux, C., Fiorino, M., & Taylor, K. (1999). Uncertainties in observationally based estimates of temperature change in the free atmosphere. *Journal of Geophysical Research*, *104*, 6305–6333. https://doi.org/10.1029/1998JD200096

Sato, M., Hansen, J., McCormick, M., & Pollack, J. (1993). Stratospheric aerosol optical depth, 1850-1990. *Journal of Geophysical Research*, *98*, 22,987–22,994. https://doi.org/10.1029/93JD02553

Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., & Zhang, J. (2014). Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *Journal of Advances in Modeling Earth Systems*, *6*, 141–184. https://doi.org/10.1002/2013MS000265

Schmidt, G. A., Ruedy, R., Hansen, J. E., Aleinov, I., Bell, N., Bauer, M., & Yao, M. S. (2006). Present day atmospheric simulations using GISS ModelE: Comparison to in-situ, satellite and reanalysis data. *Journal of Climate*, *19*, 153–192.

Segal, Y., Khain, A., Pinsky, M., & Sterkin, A. (2004). Sensitivity of raindrop formation in ascending cloud parcels to cloud condensation nuclei and thermodynamic conditions. *Quarterly Journal of the Royal Meteorological Society*, *130*(597), 561–581. https://doi.org/10.1256/qj.02.168

Sherwood, S., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., & Zelinka, M. D. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, *58*, e2019RG000678. https://doi.org/10.1029/2019RG000678

Shukla, S., Puma, M., & Cook, B. (2014). The response of the South Asian Summer Monsoon circulation to intensified irrigation in global climate model simulations. *Climate Dynamics*, *42*, 21–36. https://doi.org/10.1007/s00382-013-1786-9

Siebert, S., Kummu, M., Porkka, M., Döll, P., Ramankutty, N., & Scanlon, B. R. (2015). A global data set of the extent of irrigated land from 1900 to 2005. *Hydrology and Earth System Sciences*, *19*(3), 1521–1545. https://doi.org/10.5194/hess-19-1521-2015

Skeie, R. B., Myhre, G., Hodnebrog, Ø., Cameron-Smith, P. J., Deushi, M., Hegglin, M. I., & Wu, T.(2020). Historical total ozone radiative forcing derived from CMIP6 simulations. *npj Climate and Atmospheric Science*, *3*(1), 32. https://doi.org/10.1038/s41612-020-00131-0

Skiles, S. M., Flanner, M., Cook, J. M., Dumont, M., & Painter, T. H. (2018). Radiative forcing by light-absorbing particles in snow. *Nature Climate Change*, *8*(11), 964–971. https://doi.org/10.1038/s41558-018-0296-5

Smeed, D., Moat, B. I., Rayner, D., Johns, W. E., Baringer, M. O., Volkov, D. L., & Frajka-Williams, E. (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. British Oceanographic Data Centre, National Oceanography Centre, NERC, UK, https://doi.org/10.5285/8CD7E7BB-9A20-05D8-E053-6C86ABC012C2

Smith, C. J., Kramer, R. J., Myhre, G., Alterskjær, K., Collins, W., Sima, A., & Forster, P. M. (2020). Effective radiative forcing and adjustments in CMIP6 models. *Atmospheric Chemistry and Physics*, *20*(16), 9591–9618. https://doi.org/10.5194/acp-20-9591-2020

Solomon, S. (1999). Stratospheric ozone depletion: A review of concepts and history. *Reviews of Geophysics*, *37*(3), 275–316. https://doi.org/10.1029/1999RG900008

Sun, S., & Bleck, R. (2006). Multi-century simulations with the Coupled GISS-HYCOM Climate Model: Control experiments. *Climate Dynamics*, *26*, 407–428. https://doi.org/10.1007/s00382-005-0091-7

Teetor, P. (2011). *R cookbook*. Sebastopol, CA USA: O'Reilly Media. 438 pps.

Tegen, I., Lacis, A. A., & Fung, I. (1996). The influence on climate forcing of mineral aerosols from disturbed soils. *Nature*, *380*, 419–422.

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

Thorne, P. W., Lanzante, J. R., Peterson, T. C., Seidel, D. J., & Shine, K. P. (2011). Tropospheric temperature trends: History of an ongoing controversy. *WIREs Climate Change*, *2*, 66–88. https://doi.org/10.1002/wcc.80

Tie, X., & Brasseur, G. (1995). The response of stratospheric ozone to volcanic eruptions: Sensitivity to atmospheric chlorine loading. *Geophysical Research Letters*, *22*(22), 3035–3038. https://doi.org/10.1029/95GL03057

Twomey, S. (1977). The influence of pollution on the shortwave albedo of clouds. *Journal of the Atmospheric Sciences*, *34*, 1149–1152.

Wada, Y., Wisser, D., & Bierkens, M. F. P. (2014). Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. *Earth System Dynamics*, *5*(1), 15–40. https://doi.org/10.5194/esd-5-15-2014

Warren, S. G., & Wiscombe, W. J. (1985). Dirty snow after nuclear war. *Nature*, *313*(6002), 467–470. https://doi.org/10.1038/313467a0

Zanna, L., Khatiwala, S., Gregory, J. M., Ison, J., & Heimbach, P. (2019). Global reconstruction of historical ocean heat storage and transport. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(4), 1126–1131. https://doi.org/10.1073/pnas.1808838115