AGU PUBLICATIONS

Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE

10.1002/2013MS000266

CMIP5 historical simulations (1850–2012) with GISS ModelE2

Ron L. Miller[1,2], Gavin A. Schmidt[1,3], Larissa S. Nazarenko[1,3], Nick Tausnev[1,4], Susanne E. Bauer[1,3], Anthony D. DelGenio[1], Max Kelley[1,4], Ken K. Lo[1,4], Reto Ruedy[1,4], Drew T. Shindell[1,3], Igor Aleinov[1,3], Mike Bauer[1,2], Rainer Bleck[1,2], Vittorio Canuto[1], Yonghua Chen[1,4], Ye Cheng[1,3], Thomas L. Clune[1,5], Greg Faluvegi[1,3], James E. Hansen[1,6], Richard J. Healy[1,3], Nancy Y. Kiang[1], Dorothy Koch[1,7], Andy A. Lacis[1], Allegra N. LeGrande[1,3], Jean Lerner[1,3], Surabi Menon[8], Valdar Oinas[1,4], Carlos Pérez García-Pando[1,2], Jan P. Perlwitz[1,2], Michael J. Puma[1,3], David Rind[1], Anastasia Romanou[1,2], Gary L. Russell[1], Makiko Sato[1,6], Shan Sun[1,9], Kostas Tsigaridis[1,3], Nadine Unger[1,10], Apostolos Voulgarakis[1,11], Mao-Sung Yao[1,4], and Jinlun Zhang[12]

[1]NASA Goddard Institute for Space Studies, New York, New York, USA, [2]Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York, USA, [3]Center for Climate Systems Research, Columbia University, New York, New York, USA, [4]Trinnovim LLC, New York, New York, USA, [5]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA, [6]Now at Earth Institute, Columbia University, New York, New York, USA, [7]Now at the Department of Energy, Washington, D. C., USA, [8]Lawrence Berkeley National Laboratory, Berkeley, California, USA, [9]Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, [10]Now at School of Forestry and Environmental Studies, Yale University, New Haven, Connecticut, USA, [11]Now at Department of Physics, Imperial College, London, UK, [12]Polar Science Center, Applied Physics Laboratory, University of Washington, Seattle, Washington, USA

Key Points:
- The model simulates many observed climate trends during the twentieth century
- Contrasting warming trends among different model configurations can be related to differences in climate sensitivity, aerosol and ozone forcing, and ocean heat uptake
- Total forcing at the end of the CMIP5 historical period is larger than forcing within the previous (CMIP3) version of the model, due to greater greenhouse gas forcing with a smaller aerosol offset

Correspondence to:
R. L. Miller,
Ron L. Miller@nasa.gov

Citation:
Miller, R. L., et al. (2014), CMIP5 historical simulations (1850–2012) with GISS ModelE2, *J. Adv. Model. Earth Syst.*, *XX*, doi:10.1002/2013MS000266.

Received 12 SEP 2013
Accepted 9 FEB 2014
Accepted article online 14 FEB 2014

Abstract Observations of climate change during the CMIP5 extended historical period (1850–2012) are compared to trends simulated by six versions of the NASA Goddard Institute for Space Studies ModelE2 Earth System Model. The six models are constructed from three versions of the ModelE2 atmospheric general circulation model, distinguished by their treatment of atmospheric composition and the aerosol indirect effect, combined with two ocean general circulation models, HYCOM and Russell. Forcings that perturb the model climate during the historical period are described. Five-member ensemble averages from each of the six versions of ModelE2 simulate trends of surface air temperature, atmospheric temperature, sea ice and ocean heat content that are in general agreement with observed trends, although simulated warming is slightly excessive within the past decade. Only simulations that include increasing concentrations of long-lived greenhouse gases match the warming observed during the twentieth century. Differences in twentieth-century warming among the six model versions can be attributed to differences in climate sensitivity, aerosol and ozone forcing, and heat uptake by the deep ocean. Coupled models with HYCOM export less heat to the deep ocean, associated with reduced surface warming in regions of deepwater formation, but greater warming elsewhere at high latitudes along with reduced sea ice. All ensembles show twentieth-century annular trends toward reduced surface pressure at southern high latitudes and a poleward shift of the midlatitude westerlies, consistent with observations.

1. Introduction

Confidence in projections of future climate depends upon our ability to simulate and understand observed changes. The two centuries since the start of the industrial revolution offer an increasingly extensive set of observations of a warming climate that is apparent in trends of temperature (both at the surface and aloft), ocean heat content, and cryosphere.

The period of 1850 to the present has been chosen for coordinated climate modeling experiments comprising Phase 5 of the Coupled Model Intercomparison Project (CMIP5) [*Taylor et al.*, 2012]. This is designated as the "historical period", whose beginning is chosen prior to significant anthropogenic forcing and climate change, but sufficiently recent to contain observations that allow extensive evaluation of models. The historical experiments are part of a larger set of CMIP5 coordinated experiments that have informed the fifth and latest assessment report (AR5) by the Intergovernmental Panel on Climate Change (IPCC), who summarize our understanding of the changing climate.

The main perturbation to climate over the historical period is the change in atmospheric composition: primarily the increasing concentration of greenhouse gases and aerosols. Long-lived greenhouse gases like carbon dioxide show small geographic variations in concentration and can thus be characterized by measurements at a few sites. However, aerosols and short-lived greenhouse gases like ozone show much larger spatial contrasts, challenging our ability to measure their concentration with sufficient spatial and temporal detail. The perturbation to the radiative budget by aerosols opposes the effect of long-lived greenhouse gases, but the precise offset remains uncertain. Imprecise knowledge of aerosol forcing allows models with a range of climate sensitivity to reproduce the warming already observed, despite divergent projections of future warming [*Kiehl*, 2007]. The rate of atmospheric warming is tied to the response of the upper ocean (whose depth is defined by the extent of rapid vertical mixing of heat, typically a hundred meters or so). Thus, the forced surface response can be delayed by the export of heat to the deep ocean whose temperature is not directly tied to that of the overlying atmosphere. This export is not well constrained by observations so a model's excessive climate sensitivity can be masked during simulation of the historical period by overestimating heat uptake by the deep ocean. In summary, successful simulation of the observed climate is not a guarantee of reliable future projections because the agreement may result spuriously from cancellation of errors in the prescribed forcing, model sensitivity, and heat uptake by the deep ocean. Despite these challenges, the ability to simulate observed trends during the historical period remains an important test of a climate model. Historical simulations provide the greatest opportunity for constraining model physics and weighing model projections for the future, not only by comparison to observed trends, but also observed climate processes [*Sherwood et al*., 2014].

ModelE2, the latest generation of the NASA Goddard Institute for Space Sciences (GISS) coupled general circulation model (CGCM), was prepared for participation in CMIP5. This article, a description of the model response to forcing during the historical period, is part of a trilogy. *Schmidt et al*. [2014a] compare the climatology of ModelE2 to observations during the satellite era that spans the most recent decades. Projections of future warming by ModelE2 in response to anticipated changes in atmospheric composition will be described elsewhere.

This article has three objectives. First, we document the experimental design of the historical simulations and the forcings that perturb the climate (sections 2 and 3, respectively). This allows comparable experiments to be carried out during future model development. In addition, our description of the forcings provides a quantitative measure of the radiative imbalance leading to the simulated climate change. Second, we compare the model in section 4 to standard global data sets that extend across several decades and characterize climate change since the preindustrial period. We emphasize connections between trends of different quantities like temperature, ocean heat transport, and sea ice. This integrated comparison is intended to provide context for subsequent studies that are more detailed but more narrow in scope. In section 4, we address our third goal of identifying how the model response varies as a result of different treatments of the ocean circulation and the aerosol indirect effect. Our conclusions are presented in section 5.

2. Model Description

Here we summarize the main features of ModelE2, referring the reader to *Schmidt et al*. [2014a] for greater detail.

2.1. Atmospheric Model

The ModelE2 atmospheric general circulation model (AGCM) has resolution of 2.5° longitude by 2° latitude and 40 vertical layers. The hydrostatic approximation is introduced through the use of pressure as a vertical coordinate, with terrain-following sigma coordinates in the 23 layers below 150 mb. Above this pressure level, the layer thickness is horizontally uniform to the model top at 0.1 mb. Vertical resolution is enhanced near the lower boundary, with six layers below 825 mb, and the tropopause, with nine layers between 251 and 43 mb.

Three versions of ModelE2 were used for the CMIP5 experiments, distinguished by their treatment of atmospheric composition and the aerosol indirect effect (AIE). Aerosols perturb radiative fluxes indirectly by acting as cloud condensation nuclei (CCN), altering cloud brightness and lifetime. Specifically, aerosols reduce

cloud droplet size by distributing condensed water over a greater number of nucleation sites (the "Twomey" or first AIE) [e.g., *Twomey*, 1977], and these smaller droplets increase the number of collisions required for the droplets to grow to precipitable size (the "Albrecht" or second AIE) [e.g., *Albrecht*, 1989].

In the first version of the CMIP5 ModelE2, designated as the NINT model version (for its "noninteractive" atmospheric composition) and identified as "p1" (physics-version = 1) in the CMIP5 archive, the mass of radiatively active atmospheric constituents is prescribed, based upon separate calculations of their spatial extent and decadal variation since the preindustrial era [*Shindell et al*., 2003; *Miller et al*., 2006a; *Koch et al*., 2011]. The magnitude of the aerosol indirect effect is prescribed using an empirical relation between aerosol mass and low cloud cover [*Hansen et al*., 2005]. The NINT version is intended for comparison to the previous generation (CMIP3) version of ModelE described by *Schmidt et al*. [2006] and *Hansen et al*. [2007].

The second version of the CMIP5 ModelE2 calculates the atmospheric chemical and aerosol composition prognostically as described in section 3, allowing the resulting climate perturbation to feed back upon atmospheric composition. The aerosol indirect effect is again prescribed empirically in terms of the aerosol mass, as in the NINT experiment. Because of the prognostic tracers of chemistry, aerosols, and their direct radiative effect, this model is referred to as TCAD and identified as "p2" (physics-version=2) in the CMIP5 archive.

The third model version is identical to the TCAD version, except that the first AIE is parameterized according to a simple budget of unactivated cloud condensation nuclei, described in section 3.4.2. There is no representation of the second AIE. This model version is referred to as TCADI for its inclusion of the first indirect effect, and is identified as "p3" (physics-version = 3) in the CMIP5 archive.

The contrast between the NINT and TCAD models shows the effect of prognostic aerosols and chemistry, while the TCAD and TCADI contrast shows the importance of a mechanistic versus empirical treatments of the aerosol indirect effect.

The equilibrium climate sensitivity is 2.7, 2.7, and 2.9 K for the NINT, TCAD, and TCADI versions of ModelE2, respectively [*Schmidt et al*., 2014a]. These sensitivities are calculated relative to the preindustrial climate, after an instantaneous doubling of CO_2, using a mixed-layer ocean with prescribed ocean heat transport. For comparison, the previous generation (CMIP3) version of ModelE had a sensitivity of 2.7 K [*Schmidt et al*., 2006].

2.2. Ocean Models

For each of the three versions of the ModelE2 AGCM, we couple one of two ocean GCMs (OGCMs).

The "Russell" OGCM, originally described by *Russell et al*. [1995], and more recently by *Hansen et al*. [2007], is a mass-conserving model whose vertical coordinate is height. Conservation of mass ensures that water added as precipitation and runoff is conserved exactly before returning to the atmosphere through evaporation and sublimation, so that salt is conserved as well. The Russell OGCM has horizontal resolution of 1.25° longitude by 1° latitude, with four OGCM grid boxes nested within each AGCM grid box. There are 32 vertical layers extending to a depth of 4990 m. Tracer advection is calculated using a linear upstream scheme that updates both the tracer value and its slope within the grid box. The additional calculation of the slope maintains tighter gradients against numerical diffusion. Mesoscale mixing is parameterized according to the Gent-McWilliams scheme, although the along-isopycnal flux was misaligned resulting in excessive cross-isopycnal diffusion.

The Hybrid Coordinate Ocean Model (HYCOM) merges layers of constant depth near the surface with isopycnal layers below [*Sun and Bleck*, 2006]. The isopycnal coordinate allows cross-isopycnal diffusion of tracers beneath the surface mixed layer to be kept small compared to along-isopycnal diffusion, as generally observed. The total number of constant depth and isopycnal layers is 26. HYCOM has zonal resolution of 1° longitude with meridional resolution of 1/3° at the equator to resolve equatorial waves. The meridional grid spacing increases to nearly 1° outside of the equatorial waveguide before decreasing poleward as the cosine of latitude. A conventional spherical coordinate grid extends from the Antarctic coast to near 57°N, where it merges with a separate rectilinear grid that represents the Arctic Ocean without a polar singularity. (In the CMIP5 archive, HYCOM ocean variables are mapped onto the 1° longitude by 1° latitude Levitus grid.)

The ModelE2 AGCM coupled to the HYCOM or Russell OGCM is designated within the CMIP5 archive as the GISS-E2-H or GISS-E2-R model, respectively. (For brevity, we delete "GISS" in the remainder of this article.) Combined with the NINT, TCAD, and TCADI physics versions of the ModelE2 AGCM, the two OGCMs comprise six distinct coupled models.

2.3. Spin-Up of PI Control Runs and Branching

The historical experiments from 1850 to the present are initialized from a preindustrial control experiment. In this subsection, we describe the spin-up of the control experiment and subsequent branching of the historical experiments.

For each of the six coupled models, the ocean is initialized from rest using the Levitus climatological distribution of temperature and salinity [*Levitus and Boyer*, 1994; *Levitus et al*., 1994]. The models are integrated toward equilibrium with atmospheric composition and solar irradiance held constant at 1850 values. Greenhouse gas concentrations are prescribed using ice-core measurements [cf. *Schmidt et al*., 2011]. Since the initialization from the Levitus climatology, minor changes were occasionally introduced to the model physics. The six integrations proceeded without further change for roughly 500 years before the start of the preindustrial control runs. This 500 year duration is longer than the few decades needed to bring the upper ocean into balance, but insufficient for complete equilibration of the deep ocean, which requires a few thousand years. During the preindustrial control experiment, the top-of-atmosphere (TOA) energy imbalance in the E2-R model is small and among the top quintile of the CMIP5 models surveyed by *Forster et al*. [2013]. Nonetheless, there is slow climate drift throughout this period, as described in section 2.4.

Figure 1 shows that control-run variations of global surface air temperature are small compared to warming of approximately 1 K observed over the twentieth century [e.g., *Jones et al*., 2012]. In practice, smaller drift can be achieved only at the considerable cost of integrating the control runs an additional few thousand years until the deep ocean comes into equilibrium.

Figure 1 shows the segment of each control run that overlaps with the historical and Representative Concentration Pathway (RCP) experiments, with the segment's first and last years listed in Table 1. (These years are measured relative to the date of the Levitus initialization, and do not indicate the degree of equilibration, due to intervening changes in model physics as described above. We use this dating to be consistent with the model output archived in the piControl CMIP5 directory.) Variations in other control-run quantities like sea ice are similarly small (not shown), compared to recent observed trends.

For each of the six coupled models, we carry out an ensemble of five perturbation experiments, starting from the segment of the preindustrial control experiment listed in Table 1, and applying various forcing agents as described below to simulate the changing climate of the late-nineteenth and twentieth centuries. (Certain ensembles in the CMIP5 historical archive also include a sixth member with diagnostics at submonthly resolution that we do not analyze here.) Individual members of each historical perturbation ensemble are branched from the control run at 20 year intervals, with 80 years separating the first and last members. Twenty years is an estimate of the decorrelation time for coupled experiments with slightly different ocean conditions, and is chosen to reduce the dependence of the ensemble mean upon the initial state of the preindustrial ocean, which is sparsely observed, especially below the surface. The branch points of the individual ensemble members are marked by blue triangles in Figure 1. In total, we have 30 historical experiments of the observed climate between 1850 and the early twenty-first century, with additional ensembles perturbed by subsets of the forcing.

Each of the six control-run segments listed in Table 1 has a duration exceeding 531 years. This threshold consists of the 81 years required to initialize the five historical ensemble members, plus 451 years for comparison to the historical and RCP perturbation experiments that extend from calendar years 1850 to 2300.

2.4. Definition of Anomalies During the Historical Period

We define the climate response to forcing since 1850 by subtracting from each historical experiment a smoothed, reference time series from the corresponding preindustrial control run whose slow variation we attribute to drift. A common approach is to define the slowly evolving control climate by fitting a linear trend [e.g., *Forster et al*., 2013]. While this fit might be acceptable over the first century or so comprising the historical period, Figure 1 shows that a single, linear trend would fit poorly at several intervals within the

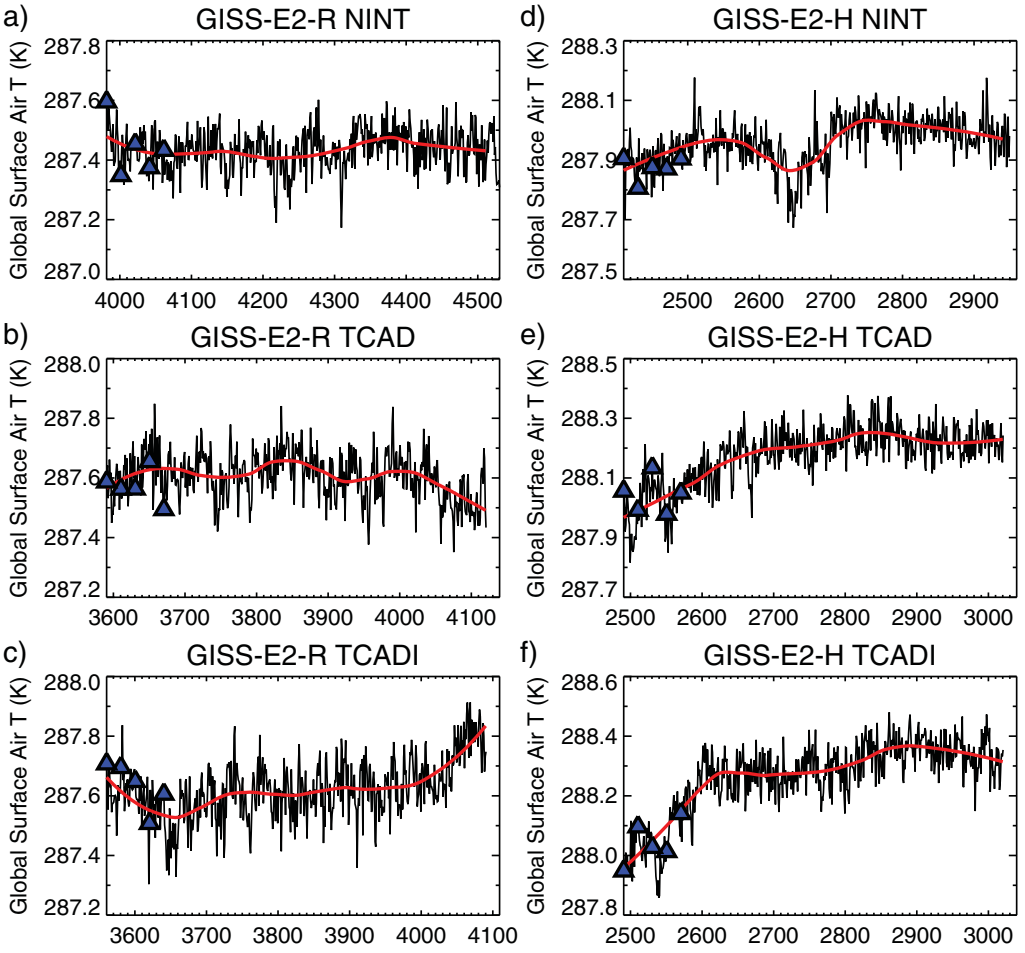

Figure 1. Global surface air temperature (black) in the preindustrial control simulations that are used to initialize the perturbation experiments. The five ensemble members for each perturbation experiment branch from the control experiment at the date marked by the blue triangles. Forced anomalies for each perturbation experiment are constructed by removing a smoothed loess fit constructed from the control (red).

entire 531 year segment, introducing spurious variability into the perturbation time series. For example, drift within the E2-H TCAD control run is rapid during the initial two centuries, but smaller thereafter. As a more flexible alternative to a linear fit, we define a slowly evolving control run using a loess fit [*Cleveland and Devlin*, 1988], calculated using the loess function of the R statistical (or stats) library [e.g., *Teetor*, 2011]. We smooth across a span of roughly 200 years to retain slow centennial-scale drift in each control. The smoothed evolution of global surface air temperature in the control experiments is shown by the red curve in Figure 1, which shows centennial-scale variability along with temporary changes in the sign of the drift within certain control runs (e.g., GISS E2-H NINT).

Compared to linear detrending, a loess fit offers greater flexibility in defining a slowly varying reference time series, when centennial-scale oscillations are present (e.g., Figure 1b). However, it is not clear for how long these oscillations remain in unison between the control and historical experiments, and thus whether they should be included within the slowly evolving reference time series that is subtracted to define the forced response (Oscillations that are out of phase between the control and forced experiments introduce variability into the perturbation that result from chaotic divergence rather than the forcing). We could reduce oscillations in the reference time series by increasing the temporal width of the loess smoothing window. However, this reduces the ability of the reference time series to match the changing rate of drift within each control. Accounting for drift depends upon finding an optimal smoothing duration that simultaneously captures the initial drift while minimizing longer period fluctuations. In any case, the smoothed reference time series in Figure 1 shows only small oscillations during the initial 156 years corresponding to the

Table 1. First and Last Year of the piControl-Run Segment Used to Construct Anomalies During the Historical Period[a]

Model	CMIP5 Archive Tag	GISS Model Name	First Year	Last Year (Duration)
E2-R NINT	piControl_r1i1p1	E135F40oQ32	3981	4530 (550)
E2-R TCAD	piControl_r1i1p2	E133TcadF40oQ32	3590	4120 (531)
E2-R TCADI	piControl_r1i1p3	E134TcadiF40oQ32	3560	4090 (531)
E2-H NINT	piControl_r1i1p1	Eh135	2410	2949 (540)
E2-H TCAD	piControl_r1i1p2	Eh137Tcad	2490	3020 (531)
E2-H TCADI	piControl_r1i1p3	Eh134Tcadi	2490	3020 (531)

[a]The first year corresponds to 1850 in the first member of the historical run ensembles, with subsequent ensemble members branching every 20 years.

historical experiments. Thereafter, the forced signal from the longer RCP projections into future centuries dwarfs any oscillations in the control experiments.

Although the 156 year historical period originally defined by CMIP5 ends in 2005 [*Taylor et al*., 2012], extensions to the end of 2012 were performed to allow comparison to observations up until the present day. In response, we extended the E2-R TCADI ensemble, along with the NINT ensembles for both ocean models. To distinguish the effect of anthropogenic and natural forcing, we carried out additional experiments for the historical period with only subsets of the forcings applied. The ensembles of the historical experiments analyzed in this article are summarized in Table 2.

3. Radiative Forcing During the Historical Period

Radiative forcing is designed to allow quantitative comparison of the climate response expected from incommensurate forcing agents, such as changing insolation and atmospheric composition, through the agents' common effect upon the radiative budget. Forcing measures the energy imbalance of the planet before the climate adjusts to restore equilibrium. This imbalance is removed over multiple time scales as the climate adjusts. Different forcing definitions have been proposed that account for rapid, initial adjustment in order to better anticipate the magnitude of the final surface response. Historically, radiative forcing has been characterized by the initial (or "instantaneous") radiative perturbation at the tropopause [e.g., *Ramaswamy et al*., 2001; *Forster et al*., 2007]. The use of the tropopause (rather than the top of atmosphere or TOA) approximates the effect of stratospheric adjustment that occurs long before the surface temperature reaches equilibrium [*Hansen et al*., 1981], a duration of a decade or so that is related to the thermal inertia of the upper ocean. The instantaneous tropopause forcing omits the effect of other fast responses, involving tropospheric clouds and water vapor, for example [e.g., *Koch et al*., 2009]. The surface temperature responds to the radiative imbalance after these fast adjustments have occurred, and recent definitions of "adjusted" or "effective" radiative forcing (ERF) have been proposed to include these fast changes and their modification of the initial radiative imbalance [e.g., *Hansen et al*., 2005; *Myhre et al*., 2014]. The ERF is typically calculated by fixing the forcing agent, along with ocean temperature and sea ice and diagnosing the radiative imbalance after the atmospheric column returns to equilibrium, a calculation that must be

Table 2. Perturbation Experiments During the Historical Period That Started From a Preindustrial Control Run in 1850[a]

Model	CMIP5 Archive Tag	GISS Model Name	First Year	Last Year
All Forcings				
E2-R NINT	historical_r1i1p1	E135f9[a-e]F40oQ32	1850	2012
E2-R TCAD	historical_r1i1p2	E137Tcadf9[a-e]F40oQ32	1850	2005
E2-R TCADI	historical_r1i1p3	E134Tcadif9[a-e]F40oQ32	1850	2012
E2-H NINT	historical_r1i1p1	Eh135f9[a-e]	1850	2012
E2-H TCAD	historical_r1i1p2	Eh137Tcadf9[a-e]	1850	2005
E2-H TCADI	historical_r1i1p3	Eh134Tcadif9[a-e]	1850	2005
Anthropogenic Forcings Only				
E2-R NINT	historicalMisc_r1i1p109	E135fa[a-e]F40oQ32	1850	2012
E2-H NINT	historicalMisc_r1i1p109	Eh135fa[a-e]	1850	2012
Natural Forcings Only				
E2-R NINT	historicalNat_r1i1p1	E135fn[a-e]F40oQ32	1850	2012
E2-H NINT	historicalNat_r1i1p1	Eh135fn[a-e]	1850	2012

[a]Each experiment is an ensemble of five members.

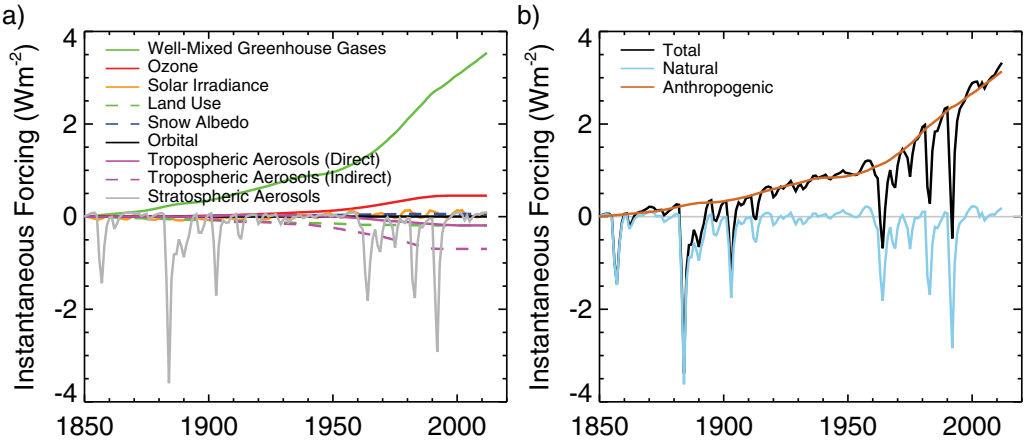

Figure 2. Instantaneous global radiative forcing at the tropopause (Wm^{-2}) in the E2-R NINT ensemble. (a) Individual forcings and (b) Total forcing, along with the separate sum of natural (solar plus volcanic aerosol plus orbital) and anthropogenic forcings.

continually repeated as the forcing agent changes with time (e.g. through varying concentration). Alternatively, the ERF is calculated by regressing temperature against the TOA radiative imbalance [*Gregory et al*., 2004]. In practice, the ERF includes a spurious offset caused by the adjustment of temperature over land that is rapid compared to the ocean response. This offset is difficult to identify and remove on a regional scale [*Shindell et al*., 2013a]. Thus, the IPCC continues to use both the instantaneous tropopause value (their "Radiative Forcing") along with the ERF [*Myhre et al*., 2014]. Because our focus is the response of climate to changing forcing agents since the preindustrial, we will generally characterize the forcing in terms of its (time-evolving) instantaneous tropopause value. However, to characterize the present-day forcing, along with the effect of aerosols upon cloud radiances, we will make use of the ERF, as described below.

Figure 2 shows the evolution of the global-average instantaneous forcing at the tropopause, calculated for the E2-R NINT ensemble. The tropopause is defined using the standard lapse-rate definition [*W. M. O*., 1957]. Inputs to the radiative calculation such as temperature, water vapor, cloud cover, and cloud optical thickness are saved from the E2-R NINT control experiment every two-and-a-half hours (the interval between successive radiative calculations in ModelE2) during the single year of 1850, before the model has been perturbed by any forcing agent. The forcing is calculated as the perturbation to the net radiative flux at the tropopause following a model change such as increased greenhouse gas concentration. The radiative forcing evolves in time with the evolution of the forcing agent. Calculation of forcing using the atmospheric state during 1850 neglects the effect of interannual variations in temperature or cloud cover, but we expect the resulting forcing uncertainty to be small.

In the remainder of this section, we discuss forcing by individual agents. Data used to prescribe or calculate each forcing agent are summarized in Table 3.

Table 3. Summary of Inputs Used to Prescribe (NINT) or Calculate (TCAD and TCADI) Atmospheric Composition and Forcing

Forcing Agent	NINT	TCAD and TCADI
Long-lived GHG		
CO_2	Observations [update from *Hansen et al*., 2007]	
CH_4, N_2O, CFC	Observations [*Hansen et al*., 2007]	Observations (surface only)
Ozone	Seasonal cycle from *Shindell et al*. [2003] with linear decrease of SH stratospheric values from 1979–1997, thereafter constant to 2005 [*Hansen et al*., 2007].	1850–2000: precursor emissions from *Lamarque et al*. [2010], and thereafter from RCP4.5 [*Lamarque et al*., 2011], with volcanic aerosols from update of *Sato et al*. [1993].
Stratospheric H_2O	Augmented in proportion to (prescribed) surface CH_4 [*Hansen et al*., 2007].	Augmented using prognostic CH_4 oxidation [*Shindell et al*., 2013b].
Solar irradiance	TSI from an updated version of *Wang et al*. [2005] with spectral variations from *Lean* [2009]. For the years 2000–2005 along with 2011 and 2012, TSI is prescribed using a recent solar cycle.	
Land use	1850–1900: transition from *Pongratz et al*. [2008] to HYDE3.0 (used thereafter).	
Orbital	None	*Berger* [1978]
Tropospheric aerosols	1850–2000: concentration from *Miller et al*. [2006a] and *Koch et al*. [2011], and thereafter using year-2000 concentration.	1850–2000: emission from *Lamarque et al*. [2010], and thereafter from RCP4.5 [*Lamarque et al*., 2011].
Snow albedo	From black carbon aerosol deposition.	
Volcanic aerosols	Aerosol concentration and size from update of *Sato et al*. [1993].	

3.1. Long-Lived Greenhouse Gases

The long-lived greenhouse gases CO_2, CH_4, N_2O, and chlorofluorocarbons (CFCs) have concentrations that are almost regionally invariant due to their long lifetime and small annual increment. Regional variations in monthly averaged concentration of CO_2 retrieved from satellite are 20 ppm or less compared to the present global value near 400 ppm [*Chahine et al*., 2008]. Routine in situ measurements for CO_2 are available after 1958 and only more recently for the other long-lived greenhouse gases. Prior to direct atmospheric sampling, past concentration is measured from ice core bubbles [e.g., *Etheridge et al*., 1996].

For the historical experiments described here, we specify the CO_2 concentration based upon measurements. [Simulations with specified emission are described by *Romanou et al*., 2013.] We approximate the concentration of CO_2 as globally uniform in the absence of regional information prior to the most recent decades. Values measured in situ and from the Antarctic ice cores are converted to a global average using climatological spatial variations revealed by the present-day global network. Construction of the global average concentration is described in the supporting information for *Hansen and Sato* [2004], while updated values are tabulated at http://data.giss.nasa.gov/modelforce/ghgases/. In the NINT ensembles, concentrations of the other long-lived species CH_4, N_2O, and CFCs are prescribed to be slightly larger in the Northern Hemisphere (NH), based upon analysis of surface measurements by *Minschwaner et al*. [1998]. Prescribed concentrations of these species are vertically uniform within the troposphere, but fall off exponentially above the tropopause with a scale height of a few tens of kilometers [*Hansen et al*., 2005]. In contrast, the CO_2 concentration is assumed to be uniform throughout the entire model column. The regional and vertical distribution of all greenhouse gases are held fixed as their concentrations rise.

For the TCAD and TCADI ensembles, the CO_2 concentration is prescribed as in the NINT experiments. Concentrations of other long-lived greenhouse gases (CH_4, N_2O, and CFC) are prescribed from observations only at the surface. Their concentration above is calculated as a result of gas-phase chemistry and transport [*Shindell et al*., 2013b].

Radiative forcing by long-lived greenhouse gases was calculated with the previous (CMIP3) version of ModelE and compared to fluxes derived from line-by-line (LBL) models with substantially greater spectral resolution [*Collins et al*., 2006]. Different LBL models give virtually identical results, suggesting that they can serve as a benchmark for assessing the accuracy of GCM radiation codes. With forcing defined as the flux difference between the present day and preindustrial, the ModelE and LBL longwave forcing differ by only 7% at the tropopause for the example of a midlatitude atmospheric column, while the net forcing difference is less than 1% due to offsetting differences in the shortwave and longwave flux.

The CMIP5 forcing associated with long-lived greenhouse gases is larger by nearly 15% compared to the CMIP3 model; most of this difference is outside the Tropics (Figure 3). At present, we have not identified the origin of this contrast. The CMIP5 ModelE2 radiation module has been updated using the latest version of the HITRAN spectroscopic database [*Rothman et al*., 2009]. However, forcing associated with a doubling of CO_2 is nearly identical between the CMIP3 and CMIP5 models [*Hansen et al*., 2005; *Schmidt et al*., 2014a]. This suggests that the discrepancy is due to the other greenhouse gases or their spectral overlap with CO_2, with a possible contribution from the different reference climates used to calculate the forcing.

3.2. Ozone

Tropospheric ozone has increased during the past century due to anthropogenic emission of ozone precursors like methane, nitrogen oxides, carbon monoxide, and volatile organic compounds [*Seinfeld and Pandis*, 2006]. Seasonal variations of tropospheric ozone are prescribed in the NINT ensembles based upon offline calculations by *Shindell et al*. [2003]. Decadal changes result from time-dependent emissions that are updated every quarter-century between 1850 and 1950 and every decade thereafter through the 1990s. The NINT distribution of tropospheric ozone is identical to that used in the CMIP3 simulations, and its temporal trends as a function of latitude, altitude, and season are shown in Figure 2 of *Hansen et al*. [2007]. Ozone loss in the polar stratospheric vortex due to ozone-depleting chemicals like CFCs is prescribed as a decreasing linear trend between 1979 and 1997 based upon measurements [*Randel and Wu*, 1999]. Thereafter, stratospheric ozone is held constant until 2005 before linear restoration to its 1979 value by 2050. The prescribed trend is largest in the stratosphere but extends into the troposphere at high latitudes.

In the TCAD and TCADI ensembles, ozone is fully prognostic [*Shindell et al*., 2013b]. Emission of ozone precursors is prescribed using *Lamarque et al*. [2010] until 2000, and the RCP4.5 scenario thereafter [*Lamarque et al*., 2011]. Stratospheric ozone is destroyed by the addition of chlorine radicals derived from

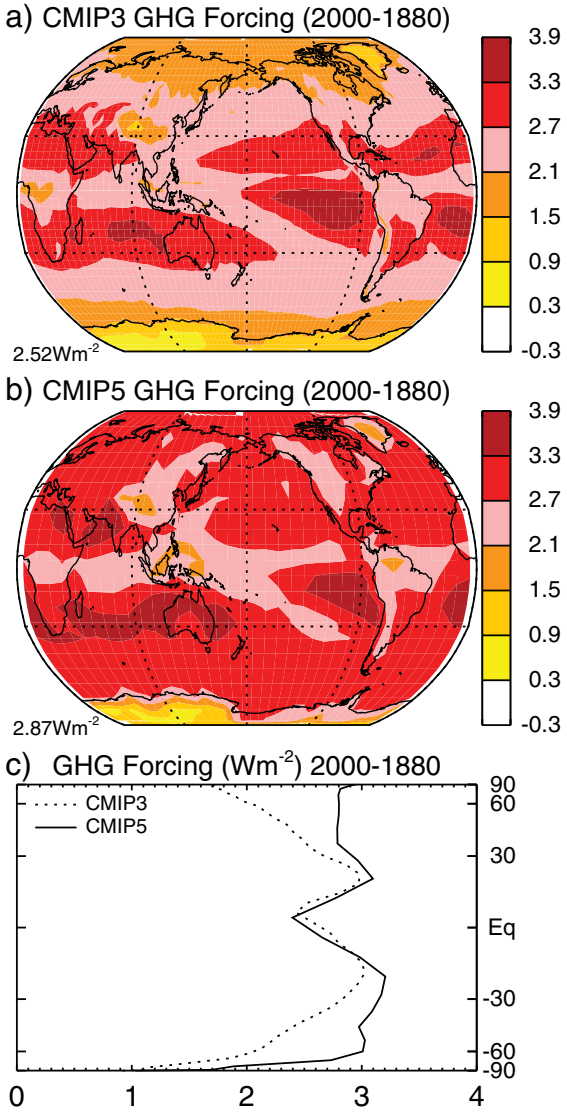

Figure 3. Instantaneous radiative forcing at the tropopause (Wm^{-2}) due to long-lived greenhouse gases for the period 1880–2000 in the CMIP5 (E2-R NINT) and CMIP3 (GISS-ER) [*Hansen et al.*, 2007] versions of ModelE. (a) CMIP3 forcing (calculated using the 1880 climate as reference), (b) CMIP5 forcing (calculated using the 1850 climate as reference), and (c) zonal-mean forcing.

anthropogenic emission of CFCs. This destruction is exacerbated after volcanic eruptions when heterogeneous chemical reactions on the surface of volcanic aerosols increase the concentration of chlorine radicals. During solar cycles, when UV irradiance temporarily increases, the ozone concentration increases in both the stratosphere and upper troposphere [*Shindell et al.*, 2006b]. Both the volcanic and solar perturbations to the ozone concentration and their attendant climate response are present only in the TCAD and TCADI ensembles.

Ozone absorbs both solar and thermal radiation, warming the stratosphere and troposphere. The global radiative forcing is positive during the past century (Figure 2a) because the tropospheric increases overcome the effect of stratospheric ozone loss. Figure 4a shows the regional distribution of instantaneous radiative forcing at the tropopause for ozone in the E2-R NINT ensemble. (The forcing will be modified quantitatively due to shortwave absorption during stratospheric adjustment, but the regional forcing distribution shown in the figure will remain approximately the same.) Ozone forcing exhibits a strong interhemispheric contrast with positive forcing associated with the large NH increase of tropospheric ozone concentration due to anthropogenic emission of precursors. This is offset by negative forcing from the expanding ozone hole in the Southern Hemisphere, where the winter stratosphere is thermally isolated and primed for ozone destruction [*Shindell et al.*, 2013b].

The global radiative forcing by ozone prescribed for the E2-R NINT ensemble is 0.45 Wm^{-2} in the year 2000 (Figure 4a). SH stratospheric ozone depletion is slightly larger in the E2-R TCADI ensemble compared to the observed trend used to specify the NINT concentration, consistent with the smaller TCADI ozone forcing of 0.39 Wm^{-2}. This contrast of the NINT and TCADI instantaneous forcings will be amplified during stratospheric adjustment, due to initial radiative cooling in the latter ensemble that is nearly twice as large (cf. Pyle et al., 2005; Box 1.3).

3.3. Stratospheric Water Vapor From Methane Oxidation

Oxidation of methane contributes significantly to stratospheric water vapor [e.g., *LeTexier et al.*, 1988; *Trenberth et al.*, 2007], more than doubling its concentration in the upper stratosphere, according to calculations by *Hansen et al.* [2007]. In the TCAD and TCADI ensembles, methane is a prognostic variable and oxidation is calculated explicitly. In the NINT ensembles, this contribution is implemented empirically, as in the CMIP3 model [*Hansen et al.*, 2007]. Stratospheric water vapor is augmented according to the surface methane concentration, with a prescribed 2 year lag that accounts for transport to the stratosphere and subsequent

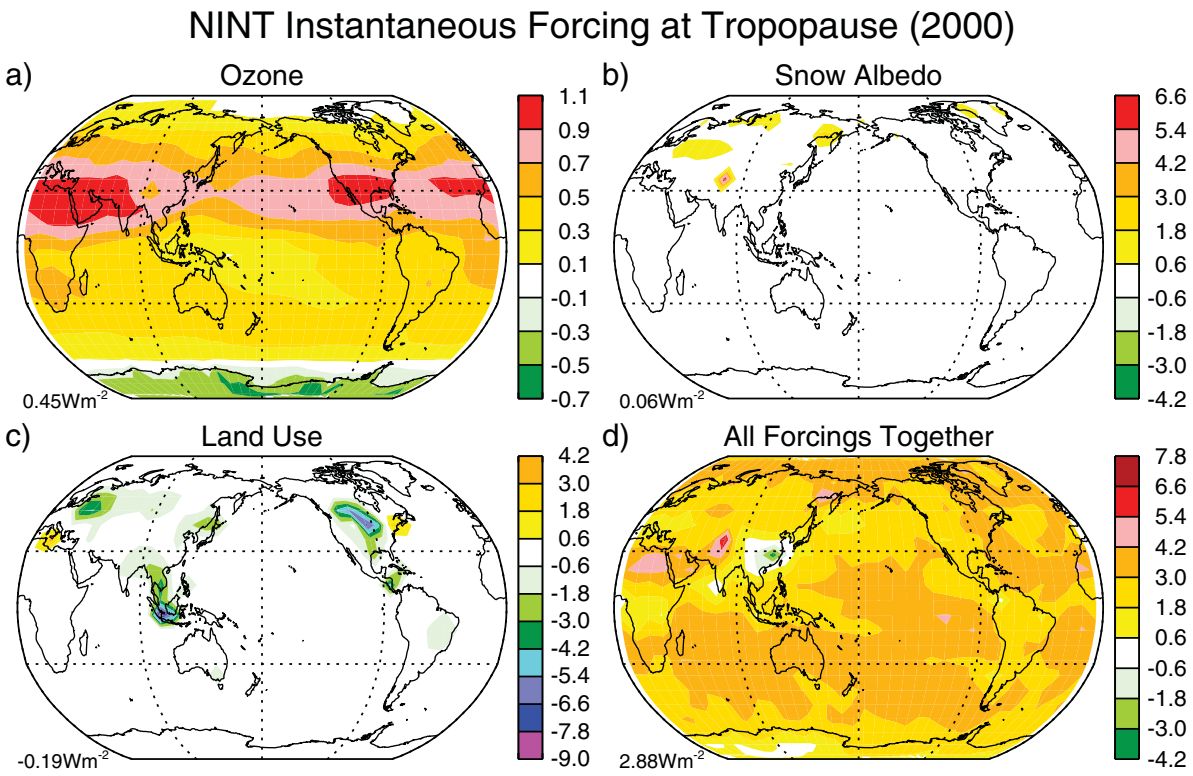

Figure 4. Regional distribution of instantaneous radiative forcing (Wm^{-2}) at the tropopause for the E2-R NINT ensemble by (a) ozone, (b) aerosols on snow and ice, (c) land use, and (d) all forcings simultaneously present. All values are for the year 2000 compared to 1850. Forcing is calculated at 8° latitude by 10° longitude resolution for computational efficiency so some regions of land-use forcing appear to extend offshore.

photochemistry. To prescribe the magnitude along with the height and latitude dependence of this increase, the equilibrium response of stratospheric water vapor was calculated using a stratospheric transport model [*Hall and Prather*, 1995], given a tropospheric methane concentration of 1740 ppbv. Oxidation and conversion to water vapor were modeled as a linear relaxation of the methane concentration, whose rate was taken from a zonally averaged model with more inclusive stratospheric photochemistry [*Fleming et al*., 1999].

Calculation of the forcing due to oxidation of methane is complicated, because several years elapse before the water vapor originating as surface methane circulates throughout the stratosphere. The forcing by stratospheric water vapor is not simply related to the current tropospheric methane concentration, but is a convolution of methane during the previous decade. *Hansen et al*. [2005] use the CMIP3 version of ModelE to calculate the equilibrium forcing by stratospheric water vapor following an abrupt increase in the surface methane concentration, estimating that the twentieth-century forcing is below 0.1 Wm^{-2}. We could similarly estimate this forcing in the (CMIP5) ModelE2, but because our methane concentration in the E2-R NINT ensemble is identical, we assume that the forcing remains small.

Solomon et al. [2010] propose that reduced surface warming observed during the recent decade is partly a consequence of a declining water vapor trend in the lower stratosphere, where methane oxidation is small, and water vapor upwells through the tropical cold trap. This feedback shows that our methane-derived forcing is an incomplete measure of the total climate effect of stratospheric water vapor.

3.4. Tropospheric Aerosols

3.4.1. Direct Radiative Forcing

Tropospheric aerosols in ModelE2 consist of sulfate, nitrate, organic carbon, black carbon, sea salt, and soil (or "mineral") dust. In the NINT ensembles, dust is prescribed using the climatological seasonal cycle calculated by *Miller et al*. [2006a]. The spatial and temporal distributions of the other aerosol species are prescribed from a separate transient simulation of the late nineteenth and twentieth century with the CMIP3

Table 4. Comparison of Forcing for the Year 2000 Within the E2-R NINT, TCAD and TCADI Ensembles Calculated Relative to 1850[a]

	NINT	TCAD	TCADI
*Ozone**			
	0.45	–	0.39
Tropospheric Aerosol			
Direct			
TOA	0.00	−0.06	−0.05
Tropopause	−0.19	−0.19	−0.19
Surface	−0.77	−0.65	−0.69
Indirect			
TOA	−0.66	−0.64	−1.05
Tropopause	−0.67	−0.66	–
Surface	−0.42	−0.40	–
Effective RF			
TOA	–	–	−1.10

[a]The TCADI ERF is provided by *Shindell et al*. [2013a].
*Ozone forcing represents the instantaneous value at the tropopause.

version of the GISS ModelE-R [*Koch et al*., 2011]. This simulation included prognostic budgets of atmospheric chemical species and aerosols coupled by gas-phase chemistry and heterogeneous reactions on dust particles between sulfate and nitrate precursors and ozone [*Bell et al*., 2005; *Bauer and Koch*, 2005; *Shindell et al*., 2006a; *Koch et al*., 2006; *Koch et al*., 2007; *Bauer et al*., 2007]. Emission was prescribed and updated once each decade starting in the 1890s using community databases [e.g., *van Aardenne et al*., 2001], so that direct and indirect forcing by anthropogenic species are zero prior to this date. The CMIP3 model has 4° latitude by 5° longitude horizontal resolution and 20 vertical layers. The profile in each CMIP3 column was used directly without horizontal interpolation for each of the four CMIP5 model columns nested within. Remapping to the higher CMIP5 vertical resolution was conservative, assuming that the concentration is piecewise constant within each CMIP3 layer.

In the TCAD and TCADI ensembles, concentrations of aerosols and atmospheric chemical species are calculated prognostically and differ from the prescribed NINT values for several reasons. First, an updated emission inventory is used for anthropogenic aerosols and biomass burning [*Lamarque et al*., 2010]. Second, aerosol physics is more extensive in the current version of ModelE2, with the introduction of secondary organic aerosols, for example. Third, the climate model itself has changed, most notably with doubled horizontal and vertical resolution since the calculations used to prescribe the NINT concentrations. In the TCAD and TCADI models, nitrate mass is multiplied by 0.2 in the radiative calculation to account for concentrations that are believed to be excessive. Black carbon mass is multiplied by 1.5 to account for coatings of sulfate and nitrate that increase shortwave absorption through internal reflection [*Bauer et al*., 2007].

In the E2-R NINT ensemble, direct radiative forcing by aerosols in the year 2000 is −0.19 and 0.00 Wm^{-2} at the tropopause and TOA, respectively (Table 4). (Our definition of aerosol forcing includes only anthropogenic species, and excludes dust and sea salt.) Direct forcing at TOA in the E2-R TCAD and TCADI ensembles is slightly more negative at −0.06 and −0.05 Wm^{-2}, respectively. Aerosol forcing is sensitive to the reference climate used to calculate the radiative fluxes. For example, *Shindell et al*. [2013a] calculated the shortwave component of aerosol direct forcing at TOA in the E2-R TCADI ensemble as −0.49 Wm^{-2}. Their forcing is calculated prognostically using the "double call" method, where the flux difference including and omitting aerosols is calculated using the instantaneous climate. That is, aerosol forcing during the year 2000 is calculated relative to the climate at that same point in time. In contrast, we calculate our forcing relative to the start of the historical period in 1850 (while including both shortwave and longwave contributions). Table 5 shows the result of additional calculations that were carried out to reconcile the difference in TCADI aerosol direct forcing between our method and that of *Shindell et al*. [2013a]. Recalculating our net flux difference using the year-2000 climate reduced the TOA forcing to −0.41 Wm^{-2}, while the shortwave contribution alone was −0.48 Wm^{-2}, nearly identical to the value calculated by *Shindell et al*. [2013a]. *Koch et al*. [2011] similarly found that NINT aerosols in the year 2000 result in TOA direct forcing of −0.40 Wm^{-2} when using the double-call method (compared to our value of 0.00 Wm^{-2} based upon the 1850 climate). Table 5 shows that aerosol direct forcing is highly sensitive to the climate used to calculate the radiative fluxes. The same aerosol distribution that would barely perturb the TOA flux in 1850 results in a much larger forcing magnitude by the end of the historical period. The aspect of

Table 5. TCADI Global-Mean Direct Aerosol Forcing at TOA and Its Dependence Upon the Climate Used to Calculate the Radiative Fluxes[a]

Experiment	Forcing (Wm^{-2})
Net relative to 1850 climate (Figure 5b)	−0.05
Net relative to 2000 climate	−0.41
Shortwave only, relative to 2000 climate	−0.48
Shindell et al. [2013a] shortwave	−0.49

[a]Additional differences in the global-mean forcing between Figure 5b and *Shindell et al*. [2013a] result from differences in temporal or ensemble averaging. Figure 5b is calculated using aerosol concentration from a five-member ensemble mean, while *Shindell et al*. [2013a] uses a single ensemble member averaged over an 11 year period centered about the year 2000.

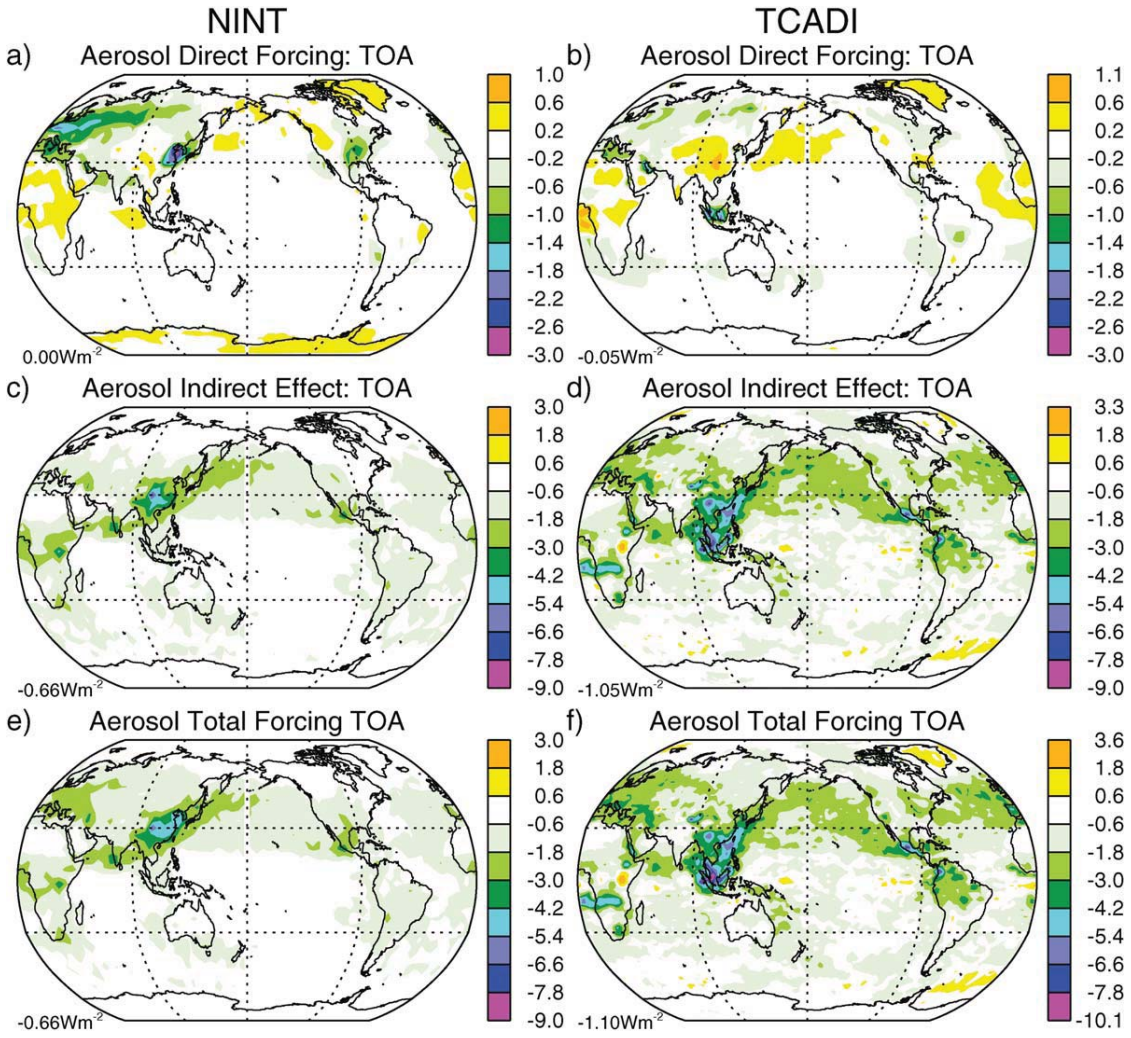

Figure 5. Regional comparison of aerosol radiative forcing (Wm^{-2}) at TOA between the GISS E2-R NINT and TCADI ensembles. (a and b) direct radiative forcing, (c and d) indirect effect, (e and f) total aerosol forcing. For the TCADI ensemble, the total aerosol forcing is taken from the effective radiative forcing (ERF) [*Shindell et al.*, 2013a], and the indirect effect is diagnosed by subtracting the direct radiative forcing from the ERF. All values are for the year 2000 compared to 1850. Forcing is calculated at 4° latitude by 5° longitude resolution, while the TCADI ERF is calculated at twice this resolution.

climate change that causes this increase in forcing magnitude deserves further investigation.

Accounting for differences in the method of calculation (based upon Table 5), our aerosol direct forcing for all E2-R ensembles is within the range of −0.58 and −0.02 Wm^{-2} calculated for Phase II of the Aerosol Comparisons between Observations and Models ("AeroCom") Project [*Myhre et al.*, 2013]. Direct forcing in this project includes only the shortwave component and was calculated using meteorological variables from 2006. Our forcing is also consistent with the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP) best estimate of −0.33 to −0.50 Wm^{-2} [*Shindell et al.*, 2013a].

The regional contrast between the E2-R NINT and TCADI direct forcing is shown in Figures 5a and 5b. Because of large spatial contrasts in aerosol concentration, we calculate the forcing at higher resolution (4° latitude by 5° longitude) compared to the other forcing agents. Although the global-mean forcing is similar in the two models, there are large regional differences. In the NINT ensemble, large negative forcing occurs in populous and industrialized regions such as Europe along with eastern North America and Asia. In contrast, the TCADI ensemble has positive forcing in these latter two regions. The TCADI model shows large negative forcing near Indonesia that has no counterpart in the NINT ensemble. Aerosol forcing by the TCADI ensemble along with the other ACCMIP models was compared to an extensive set of observations by

Shindell et al. [2013a]. These models generally reproduce the observed total aerosol optical thickness, despite its systematic underestimation over Asia.

The deposition of black carbon aerosols upon snow decreases the surface albedo and increases solar absorption. Its global effect (referred to as "Snow Albedo" in Figure 2) is small in both the E2-R NINT and TCADI ensembles, but can dominate the total forcing in specific regions like the Himalayas and Tibetan Plateau (Figures 4b and 4d), where there is a bright, snowy surface, high solar zenith angle, and a significant nearby source of anthropogenic aerosols. *Bauer et al*. [2013] provided additional description of black carbon deposition upon snow in the TCADI ensembles and comparison to measurements.

None of the ensembles allow temporal variations in dust sources. A recent attribution based upon satellite retrievals suggest that one-quarter of the present-day dust load originates from anthropogenic sources created by cultivation and grazing [*Ginoux et al*., 2012]. This suggests that dust source extent has varied markedly since the preindustrial era. *Mahowald et al*. [2010] used ice cores and other deposition records to estimate that the dust load may have doubled during the twentieth century. For comparison, the global dust load in the TCAD and TCADI ensembles shows significantly smaller variations (on the order of 10%) resulting from climate variations that influence dust emission, transport and removal.

The emission inventories used to prescribe or calculate aerosol concentration in the historical simulations extend only to the end of the twentieth century [*van Aardenne et al*., 2001; *Lamarque et al*., 2010]. Between the years 2000 and 2005, the emission inventories in the TCAD and TCADI experiments are taken from the RCP4.5 scenario [*Lamarque et al*., 2011]. For the NINT experiments, aerosol concentrations are held constant after the year 2000.

3.4.2. Aerosol Indirect Effect

In the NINT and TCAD ensembles, the magnitude of the AIE is tuned using an empirical relation between low cloud cover and the logarithm of aerosol number concentration [*Hansen et al*., 2005, 2007]. The logarithmic dependence results in a stronger indirect effect in regions of pristine air. Aerosol number concentration is derived from aerosol mass using prescribed particle size and solubility. Dust and salt aerosols do not contribute to the AIE due to their presumed small anthropogenic component. Aerosols modify only low cloud cover, and only where these clouds would have formed in the absence of aerosols. (This prevents aerosols from creating clouds where humidity is low.) Because the NINT ensembles were designed to be compared with the corresponding CMIP3 experiments, we intended the AIE at the tropopause to be near the value of -1 Wm^{-2} used by *Hansen et al*. [2007]. However, the precise magnitude of this forcing cannot be specified a priori because it depends upon the simulated climate, especially the cloud distribution. For the E2-R NINT ensemble in the year 2000, the actual instantaneous AIE at the tropopause is -0.67 Wm^{-2} (Table 4), with its regional dependence shown in Figure 5c. For the E2-R TCAD ensemble, this forcing is slightly smaller at -0.66 Wm^{-2}, with a nearly identical regional distribution (not shown). Note that while the TCAD direct forcing is calculated using prognostic aerosol concentration, the TCAD AIE is calculated using the same offline aerosol concentration prescribed in the NINT ensembles.

In the TCADI ensembles, the first AIE is represented mechanistically using a prognostic budget for the number of unactivated cloud condensation nuclei. This number is a balance between aerosol supply and loss by wet deposition, among other processes [*Morrison and Gettelman*, 2008; *Menon et al*., 2010]. Only the first indirect effect is represented: the autoconversion rate depends upon only the cloud water content, and is not reduced with decreasing droplet size. At present, larger aerosols like dust do not nucleate ice crystals.

For the E2-R TCADI ensemble, an ERF of -1.10 Wm^{-2} is calculated by fixing ocean temperature (at its 1850 value) and allowing the model to come into equilibrium with the emission of anthropogenic aerosols prescribed for the year 2000 [*Shindell et al*., 2013a]. The ERF also includes the semidirect effect [*Hansen et al*., 1997; *Perlwitz and Miller*, 2010], along with changes to other variables such as water vapor and the lapse-rate in response to the introduction of aerosols. We define the TCADI AIE as the difference between the ERF and the direct radiative forcing by the aerosols. This contrast is somewhat inconsistent because the aerosol concentration used to calculate the ERF differs slightly from that used for the direct forcing, since the former is based upon prescribed year-2000 *emission* but a climate consistent with 1850 ocean temperatures. (The concentration used to calculate direct forcing results from both emission and climate taken from 2000.) In any case, the resulting AIE at TOA is -1.05 Wm^{-2}, whose magnitude is almost twice the E2-R NINT value (Table 4).

Because of its inclusion of other feedbacks, the ERF of the TCADI model is only approximately comparable to the total forcing of the NINT and TCAD ensembles that is the sum of the direct forcing and the AIE. With these caveats, the aerosol ERF and total forcing are most negative in the TCADI ensemble, but all are within the large range of -0.9 ± 1.0 Wm^{-2} compiled from observational estimates and other CMIP5 models [*Myhre et al*., 2014].

In the E2-R NINT and TCAD ensembles, the magnitude of AIE is largest over eastern Asia, collocated with large direct radiative forcing (Figure 5c). (Because of the similarity of the regional distributions of NINT and TCAD forcing, only the former are shown in the figure.) The AIE of the E2-R TCADI ensemble, calculated from the difference of the ERF and direct forcing, is largest over Indonesia (Figure 5d). This corresponds to a region where the magnitude of direct forcing is large, but the AIE is also large within the North Pacific and North Atlantic storm tracks that are downwind of large direct forcing. Similarly, the TCADI AIE is large just offshore of the region of biomass burning over West Africa. The TCADI experiments show reduced errors in cloud fraction in these regions [*Schmidt et al*., 2014a], although confident attribution of cloud anomalies to the presence of aerosols requires a more careful analysis.

Because it is the ERF rather than the AIE that is directly calculated for the TCADI model, we show the former along with the total forcing for the NINT model in Figures 5e and 5f. Because of the small magnitude of the direct effect in both models (at least when calculated relative to the 1850 climate), the ERF and total forcing are strongly influenced by the interaction of aerosols with clouds.

3.5. Volcanic Aerosols

Forcing by volcanic aerosols in the stratosphere depends upon the optical thickness and effective radius of the sulfate droplets created by oxidation of sulfur dioxide injected above the tropopause. For all ensembles, we specify volcanic aerosol properties in the stratosphere following major eruptions, along with a time-invariant background state that represents the effect of ubiquitous smaller eruptions that maintain a minimum aerosol concentration between major events. For both components, the stratospheric aerosol optical thickness and effective radius as functions of latitude and height are prescribed using a compilation updated from *Sato et al*. [1993], available at http://data.giss.nasa.gov/modelforce/strataer/. The aerosol effective radius is typically 0.2 μm but can grow as large as 0.6 μm in the Tropics in the first year following a major tropical eruption before relaxing to smaller values. Aerosol properties for eruptions within the satellite era are constrained by measurement, but values from prior eruptions are characterized by extrapolating from measurements following El Chichón and Pinatubo. *Hansen et al*. [2007] estimates the forcing uncertainty as 20% for Pinatubo and as large as 50% for Krakatau. At the time we carried out the historical integrations, we had not updated the volcanic aerosol compilation after the year 2000. For subsequent years through 2005, we held aerosol properties constant at their year-2000 value.

3.6. Solar Irradiance

Total solar irradiance (TSI) varies with an approximate 11 year cycle, but longer term variations in the cycle amplitude and minimum value (or "background") are known less precisely and still debated [*Schmidt et al*., 2011, 2012]. Variations in irradiance over the cycle are nearly in phase at all wavelengths, with the largest variations at shorter wavelengths near the ultraviolet [*Lean*, 2009].

While TSI has been measured from space since 1978, radiance measurements were less reliable between mid-1989 and the fall of 1991, and there are multiple reconstructions of a continuous multidecadal satellite record of TSI. Prior to the satellite measurements, TSI and its spectral dependence must be inferred from proxies like visible images and counts of sunspots.

For the historical period starting in 1850, we use the TSI reconstruction of *Wang et al*. [2005] as recommended for the CMIP5 experiments (http://sparcsolaris.gfz-potsdam.de/cmip5.php). This postulates smaller changes in the background over the historical period, and thus smaller solar forcing, compared to *Lean* [2000] that was used for the GISS CMIP3 simulations [*Hansen et al*., 2007]. Between 2000 and 2005, TSI is prescribed by repeating the previous 11 year cycle instead of using measurements. Spectral variations are taken from *Lean* [2009]. Recent observations of out-of-phase behavior [*Harder et al*., 2009] remain to be confirmed [*Lean and Deland*, 2012], and were not used in this study.

During the last decade of the twentieth century, global forcing by TSI relative to 1850 ranges between zero and 0.14 Wm^{-2} (Figure 2a). The forcing trend during the entire historical period is slightly positive, but

Table 6. Instantaneous Radiative Forcing at the Tropopause (Wm^{-2}) in 2003 Relative to 1880 for the CMIP5 E2-R NINT Ensemble Compared to Forcing in the Previous Generation (CMIP3) GISS ModelE[a]

Forcing	ModelE-R (CMIP3)	ModelE2-R NINT (CMIP5)
Well-mixed greenhouse gases	2.62	2.99
Ozone	0.44	0.45
Solar irradiance	0.23	0.11
Land use	(−0.09)	−0.14
Snow albedo	0.05	0.06
Orbital	–	<0.003
Tropospheric aerosols (direct)	−0.41	−0.19
Tropospheric aerosols (indirect)	−1.00	−0.67
Volcanic aerosols	0.00	−0.14
Sum	1.84	2.47

[a]The period of 2003 to 1880 is chosen to allow comparison to the CMIP3 values in Table 1 of *Hansen et al.* [2007]. Due to the omission of the CMIP3 instantaneous forcing for land use, we use the ERF value (denoted as F_S in that study). For stratospheric water vapor from methane oxidation, we assume that the CMIP5 forcing is comparable to the value of 0.1 Wm^{-2} calculated for the CMIP3 model.

remains small compared to that of long-lived greenhouse gases, a comparison that remains robust when forcing is calculated from other TSI reconstructions.

3.7. Orbital Variations

Orbital parameters including eccentricity, obliquity, and the date of perihelion vary over the course of the TCAD and TCADI historical ensembles according to *Berger* [1978]. (We inadvertently omitted orbital forcing from the NINT ensembles, so the value in Table 6 represents the importance of this neglected forcing.) The associated forcing since 1850 is small (with a global value of less than 0.003 Wm^{-2}, and seasonal differences are less than 1 Wm^{-2} in magnitude at any latitude. Nonetheless, we include orbital forcing in the historical runs to allow a coherent comparison to CMIP5 runs for the last millennium where regional variations of insolation compared to the present day are as large as 5–10 Wm^{-2} in magnitude.

3.8. Land Use

After 1900, we specify the location of cropland and pasture using the HYDE3.0 land-use and land-cover data base [*Klein Goldewijk and van Drecht*, 2006]. This data base was not available in time for the CMIP5 ModelE2 simulation of the last millennium, which instead uses *Pongratz et al.* [2008]. Although HYDE3.0 is available back to the beginning of the historical period, we blend these two data sets for the period between 1850 and 1900 using linear weighting with respect to time for continuity of forcing between the last millennium and historical experiments.

Land-use forcing due to changes in surface albedo is as large in some cultivated regions as that due to increasing greenhouse gas concentrations (Figure 4c). The forcing is negative, because fields cleared for cultivation and grazing are typically brighter than the natural vegetation they replace. Figure 4c shows clearing of the rain forests in Indonesia and Central America, along with contraction of the boreal forests in the western plains of Canada. Conversely, there is positive forcing in the northeastern United States where farms abandoned during the last century are being reclaimed by forest succession. Global average land-use forcing in the E2-R NINT ensemble is −0.19 Wm^{-2} at the tropopause by the start of the twenty-first century, comparable to the direct forcing by tropospheric aerosols.

The historical experiments do not include prescribed changes in irrigation that have been subsequently introduced into ModelE2, and that reduce surface warming by increasing evaporation over agricultural regions (B. Cook et al., Irrigation as an historical climate forcing, submitted in 2013 to *Climate Dynamics*).

3.9. Extension of Historical Runs From 2005 to Present Day (2012)

We extended both of the NINT ensembles and the E2-R TCADI ensemble to allow their comparison to observations up until the present day. When the ensemble extensions were carried out, forcing observations were available only through 2010. For the years 2011 and 2012, we used RCP4.5 to prescribe concentrations of well-mixed greenhouse gases. For these years, solar irradiance was taken from 1998 and 1999: analogous years in a recent solar cycle. (The effect of taking values from previous cycles in lieu of observations resulted in a slight overestimate of solar forcing since 2000.) We used an updated atlas of volcanic aerosol properties between 2006 and 2010 [cf. *Sato et al.*, 1993]. No volcanic eruptions were prescribed for 2011 and 2012, when volcanic aerosol concentration was limited to a small background value.

In the NINT extensions, tropospheric aerosols subsequent to 2005 were fixed at year-2000 concentrations, the last year in the *Koch et al.* [2011] simulations. Stratospheric ozone, held constant since 1997, began a

slow recovery after 2005 [*Hansen et al*., 2007]. For the TCADI extension, tropospheric aerosols and ozone were calculated using emission inventories from RCP4.5 [*Lamarque et al*., 2011].

Because observed solar irradiance and volcanic aerosol concentrations were not used between 2000 and 2005, there are small discontinuities in the forcing at the beginning of the extension runs in 2006.

3.10. Total Forcing and Comparison Between CMIP3 and CMIP5

The total instantaneous tropopause forcing for the E2-R NINT ensemble (Figure 2b) is dominated by the effect of increasing greenhouse gas concentrations, offset by aerosol forcing with intermittent contributions by volcanic aerosols following eruptions. Figure 4d shows that the effect of the combined forcing agents is nearly everywhere positive. There is negative forcing only in a few limited regions, like eastern Asia as a result of tropospheric aerosols, and Indonesia due to land use.

Table 6 compares the E2-R NINT forcing to that calculated with the previous (CMIP3) version of ModelE [*Hansen et al*., 2007, Table 1]. The comparison is based upon the difference in the instantaneous tropopause values between 2003 and 1880. [While the forcing is computed as the difference over the same period for both models, the year 1850 represents the unperturbed climate for the calculation of the E2-R NINT forcing, compared to the year 1880 used by *Hansen et al*., 2007.] Greenhouse forcing is 15% larger in the E2-R NINT ensemble, as described previously (Figure 3). The negative offset by the aerosol forcing in the E2-R NINT ensemble is smaller in magnitude compared to the CMIP3 value, due to both direct and indirect effects. Solar forcing is smaller in the E2-R NINT ensemble. The sum of the forcings for the E2-R NINT ensemble is larger than the CMIP3 ModelE-R sum by about 0.6 Wm^{-2}.

Note that the forcing in Table 6 is calculated relative to its value in 1880, by which time there was already forcing due to rising greenhouse gas concentrations (Figure 2a). The present-day NINT forcing relative to the (1850) preindustrial climate is larger; 2.88 and 3.32 Wm^{-2} by 2000 and 2012, respectively (Figures 2b and 4d).

Figure 2b shows the separate contributions to the sum of the forcings by natural and anthropogenic effects. The former combines forcing by solar and orbital variations, along with volcanic aerosols, while the latter is the sum over the remainder of the individual forcings. Natural forcing is dominated by episodic increases in the volcanic aerosol concentration following major tropical eruptions, with a residual trend from solar irradiance that is positive but small over the historical period. In contrast, the pronounced upward trend of the total forcing is almost perfectly matched by the slow but persistent increase of the anthropogenic component.

Computation of the total instantaneous forcing for each ensemble is expensive. As an alternative, the ERF calculated according to *Gregory et al*. [2004] is shown in Figure 6. The relation between surface air temperature and the TOA radiative imbalance is calculated using regression within the first 150 years of an experiment after the concentration of CO_2 is abruptly quadrupled. This regression coefficient is then multiplied by temperature variations over the historical period and combined with the radiative imbalance to derive the ERF [*Forster and Taylor* 2006]. (We do not attempt any correction resulting from rapid adjustment of the land temperature.) For comparison, the instantaneous tropopause forcing (denoted by "iRF") from the E2-R NINT ensemble is shown. The E2-R NINT ERF exceeds the iRF early in the historical period, but is up to 0.3 Wm^{-2} smaller by the most recent decade. This may be the result of rapid adjustments to the forcing, or nonlinearities in the relation between temperature and the TOA imbalance that we do not account for. Among the historical experiments, the two TCAD ensembles have the lowest ERF. The deficit increases with time and is as large as 0.5 Wm^{-2} by the present day.

Forster et al. [2013] derived the ERF for several CMIP5 models during the historical period. The GISS models had the largest forcing when perturbed only by greenhouse gases, and were in the top quartile of total forcing.

4. Results

4.1. Surface Air Temperature

Globally averaged surface air temperature during the historical period is shown in Figure 7 for all forcings along with anomalies from separate simulations perturbed by either anthropogenic or natural forcings. (Table 2 presents a complete list of the historical experiments.) Each model is represented by its ensemble mean. For comparison, the GISTEMP analysis of global surface air temperature is shown [*Hansen et al*.,

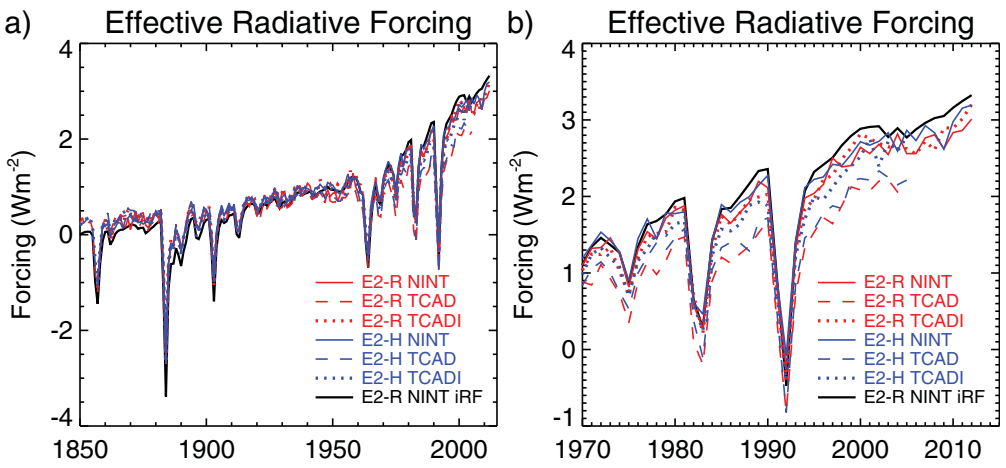

Figure 6. (a) Effective radiative forcing at TOA (Wm^{-2}) relative to 1850, and calculated according to *Forster and Taylor* [2006]. The black line shows instantaneous radiative forcing at the tropopause ("iRF") for the E2-R NINT ensemble. (b) excerpts the ERF since 1970.

2006]. Other observational analyses document similar warming [e.g., *Jones et al*., 2012; *Rohde et al*., 2013]. The uncertainty of the GISTEMP global temperature anomaly (corresponding to twice the standard error) is ±0.05 K in recent decades and twice that in the earliest part of the record [*Hansen et al*., 2006]. This uncertainty includes the effect of measurement error and incomplete global coverage by the observational network. *Morice et al*. [2012] estimate roughly twice this uncertainty for the HadCRUT4 temperature analyses, but in either case the errors are small compared to the observed warming of roughly 1 K over the course of the twentieth century.

The model anomalies are perturbations from the unforced preindustrial climate, and indicate the effect of forcing. The forced response in the GISTEMP analysis is harder to isolate, because the global record starts in the late-nineteenth century when the response to increases in greenhouse gas concentrations may have already begun. In Figure 7, we define the model and observational anomalies to have zero average between 1880 and 1900, the earliest part of the observational record.

The models warm during the twentieth century at a slightly greater rate than observed, with the exception of the TCAD simulations. The range of values spanned by individual members within the E2-R NINT ensemble is indicated by gray shading that corresponds to two standard deviations of the intraensemble anomalies. (The standard deviation is similar for other ensembles.) The E2-R ensembles and the GISTEMP analysis agree within two standard deviations until the last decade, when the NINT and TCADI simulations are roughly 0.1 K too warm compared to observations by 2012 (Figure 7b). Warming is larger in the E2-H simulations, which is associated with relatively small heat uptake by the deep ocean compared to the E2-R ensembles (section 4.4).

The NINT and TCADI ensembles warm at similar rates during the twentieth century, even though the latter has a slightly larger climate sensitivity. This sensitivity contrast is offset by the greater magnitude of TCADI aerosol forcing, mainly due to aerosol effects upon clouds (compare Figures 5c and 5d). The forcing contrast is augmented by the smaller TCADI ozone forcing. The TCAD ensembles also have prognostic ozone, which contributes to their slower rate of warming compared to the NINT ensembles with which they share a low climate sensitivity relative to the TCADI model. Prognostic TCAD ozone would contribute to these ensembles' lower effective forcing in Figure 6. We do not have an explanation for the similarity of the NINT and TCADI forcings in this figure. Calculation of the ERF assumes that climate sensitivity, which for the figure was calculated with respect to quadrupling CO_2, is the same for all forcing agents. This may not be valid [*Hansen et al*., 2005], and would cause inaccuracies in the derived ERF, where the forcing agents are varying in time and relative importance.

Each ensemble warming trend is consistent with the observed warming of the surface between 1979 and 2005 (Figure 8), as are the trends of many of the individual ensemble members. We chose the beginning of this period to coincide with the starting date typically used to calculate trends in the Microwave Sounding

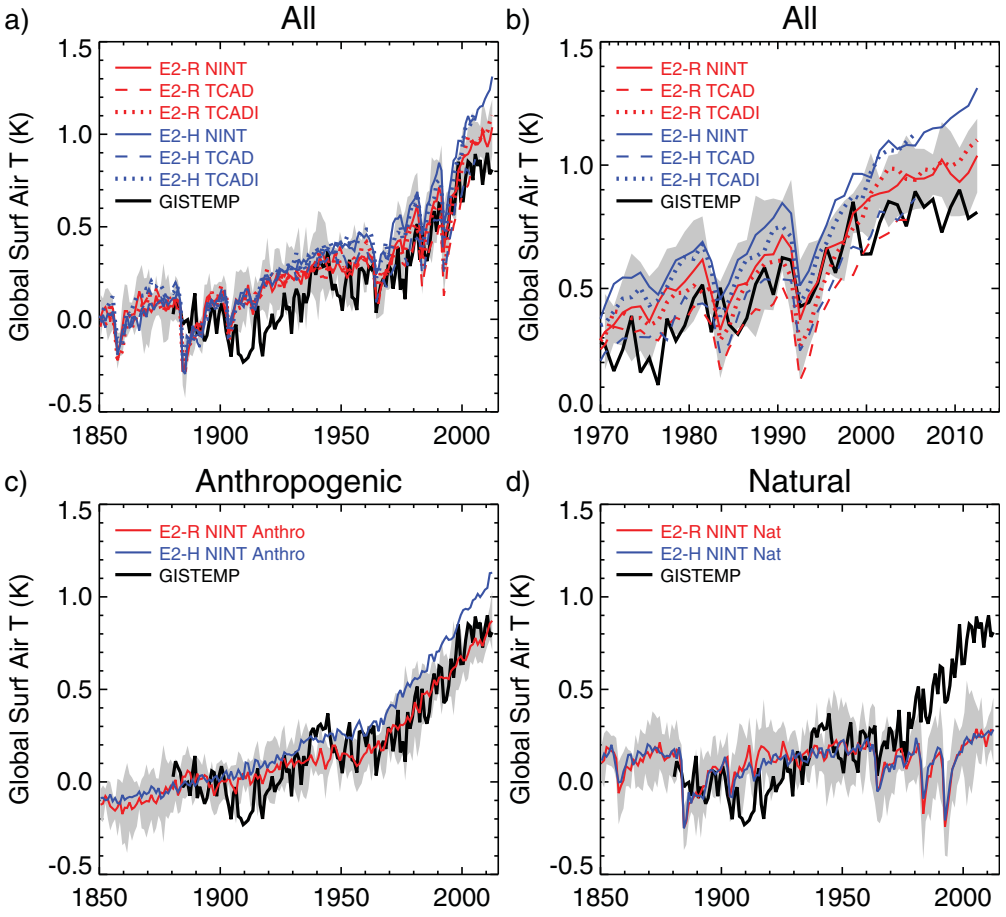

Figure 7. Global surface air temperature anomalies (K) for the historical experiments relative to a smoothed version of the control run. Gray shading shows plus or minus two standard deviations for the five members of the E2-R NINT ensemble. For comparison, the observed GISTEMP anomaly is plotted. All time series are plotted relative to their 1880–1900 average. The historical experiments are perturbed by (a and b) all, (c) only anthropogenic, and (d) only natural forcings. Figure 7b excerpts the response to all forcings since 1970.

Unit temperature retrievals described in section 4.2. The end of the period marks the conclusion of the historical experiments, after which not all ensembles are available. Table 7 gives numerical values for the trend of the observations along with each ensemble mean over this period, showing agreement within the uncertainty due to interannual variations.

In the last decade, the observed rate of warming has slowed, but the historical ensembles extended past 2005 have generally maintained their upward trend of temperature (Figure 7b). Table 7 compares the observed and modeled trends over the period 1979–2012 using only the extended ensembles. Only the E2-R NINT trend shows any overlap with the observed warming; the other two ensembles show excessive warming. This overestimate could result from excessive climate sensitivity, insufficient heat uptake by the deep ocean, or an overestimate of the total climate forcing. We have noted that the effective forcing by the NINT ensembles is among the larger values associated with the CMIP5 models [*Forster et al*., 2013]. Moreover, recent observations suggest that the CMIP5 forcing protocols (defined prior to the conclusion of the extended simulations) are a slight overestimate [*Schmidt et al*., 2014b]. Part of the warming discrepancy may also be the temporary result of fluctuations in deep ocean heat uptake that are uncorrelated between the models and observations [*Meehl et al*., 2011]. Alternatively, the model climate sensitivity may be too large due to unrealistic features that we have yet to identify.

Experiments including anthropogenic forcing can reproduce the warming observed during the twentieth century (Figure 7c), consistent with previous generations of coupled models surveyed by the IPCC [e.g.,

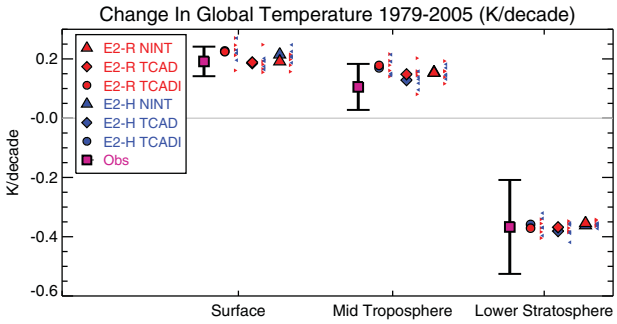

Figure 8. Change in modeled and observed global average temperature between 1979 and 2005 at the surface, middle troposphere, and lower stratosphere. The observed values are from GISTEMP (surface), and MSU channels TMT (middle troposphere) and TLS (lower stratosphere). The observed value is in purple with its uncertainty in black (corresponding to plus or minus two standard deviations of interannual variations). Ensemble average change is shown as triangles (NINT), diamonds (TCAD), and circles (TCADI), with change for individual ensemble members plotted as small triangles. Change is expressed per decade.

Hegerl et al., 2007]. In contrast, models with only natural forcings like solar irradiance and volcanic aerosols show no sustained warming trend (Figure 7d). The most discernible response to natural forcing is the cooling that follows volcanic eruptions. This temporary cooling is apparent in the GISTEMP analysis, except following the Krakatau eruption in 1883, when the absence of cooling may be a consequence of limited ocean measurements. (Following Krakatau, land stations cool at a magnitude consistent with our ensembles) [*Hansen et al*. [2007, Figure S1].

The latitudinal dependence of surface warming is shown in Figure 9. Anomalies are defined relative to their time average between 1961 and 1990, a period chosen because it is the earliest that the GISTEMP analyses are available at all latitudes, including those containing Antarctica. Because the average temperature of each time series is constrained to be identical late in the record, contrasting rates of warming between the ensembles are most apparent in the earlier decades. Volcanic eruptions are marked in the figure by green dashed lines, and are generally followed by cooling as seen in the global average temperature of Figure 7. (Table 8 identifies the eruptions marked in this and subsequent figures.) For this and subsequent plots of the temporal evolution of zonal means, dots mark locations where the anomaly (or difference) is outside the 95% confidence interval that results from intraensemble variability. For comparison between the model and observations, this interval is proportional to the intraensemble standard deviation; dots indicate where the observations are significantly outside the range of ensemble members and thus inconsistent with the model. For comparison between model ensemble means, this interval is proportional to the intraensemble standard error, and dots indicate where the ensembles are statistically distinct.

Warming within the E2-R NINT ensemble is largest over the Arctic (Figure 9a). This "Arctic amplification" is robust among models [*Holland and Bitz*, 2003], and partly the result of the surface-albedo feedback, although other feedbacks involving clouds along with atmosphere and ocean heat transport contribute [*Kay et al*., 2012]. Warming is only slightly enhanced at high SH latitudes, where the Antarctica surface albedo is buffered by a deep layer of snow and ice, and there is negative forcing by stratospheric ozone loss (Figure 4a) along with substantial export of heat to the deep ocean.

Comparison of model warming to the observed trend is obscured by interdecadal variability. The largest discrepancy occurs prior to the midtwentieth century when temporary warming observed at NH high latitudes is not simulated by the E2-R NINT ensemble (Figures 9b and 9c). Individual members of the ensemble show warming

Table 7. Modeled and Observed Trends in Surface, Middle Troposphere, and Lower Stratospheric Global Average Temperature (K per decade) for all Historical Ensembles (1979–2005) and Only the Extended Ensembles (1979–2012)[a]

	Surface	Middle Troposphere	Lower Stratosphere
		1979–2005	
Obs	0.192 ± 0.050	0.106 ± 0.078	−0.367 ± 0.159
E2-R NINT	0.192 ± 0.057	0.153 ± 0.060	−0.354 ± 0.198
E2-R TCAD	0.188 ± 0.065	0.148 ± 0.077	−0.368 ± 0.211
E2-R TCADI	0.224 ± 0.072	0.178 ± 0.091	−0.371 ± 0.209
E2-H NINT	0.215 ± 0.057	0.155 ± 0.058	−0.362 ± 0.201
E2-H TCAD	0.186 ± 0.061	0.128 ± 0.073	−0.381 ± 0.207
E2-H TCADI	0.227 ± 0.066	0.170 ± 0.079	−0.359 ± 0.217
		1979–2012	
Obs	0.168 ± 0.033	0.080 ± 0.053	−0.309 ± 0.101
E2-R NINT	0.191 ± 0.036	0.168 ± 0.038	−0.303 ± 0.123
E2-R TCADI	0.234 ± 0.045	0.217 ± 0.058	−0.300 ± 0.133
E2-H NINT	0.232 ± 0.036	0.190 ± 0.039	−0.307 ± 0.125

[a]Model trends represent the ensemble mean. Observed trends are calculated from GISTEMP (surface) and MSU channels TMT (middle troposphere) and TLS (lower stratosphere). Also shown is the uncertainty represented by two standard deviations of the trend estimate (related to interannual variability). Different methods of MSU TMT retrieval contribute additional trend uncertainty near 0.05 K per decade [cf. *Thorne et al*., 2011, Figure 10].

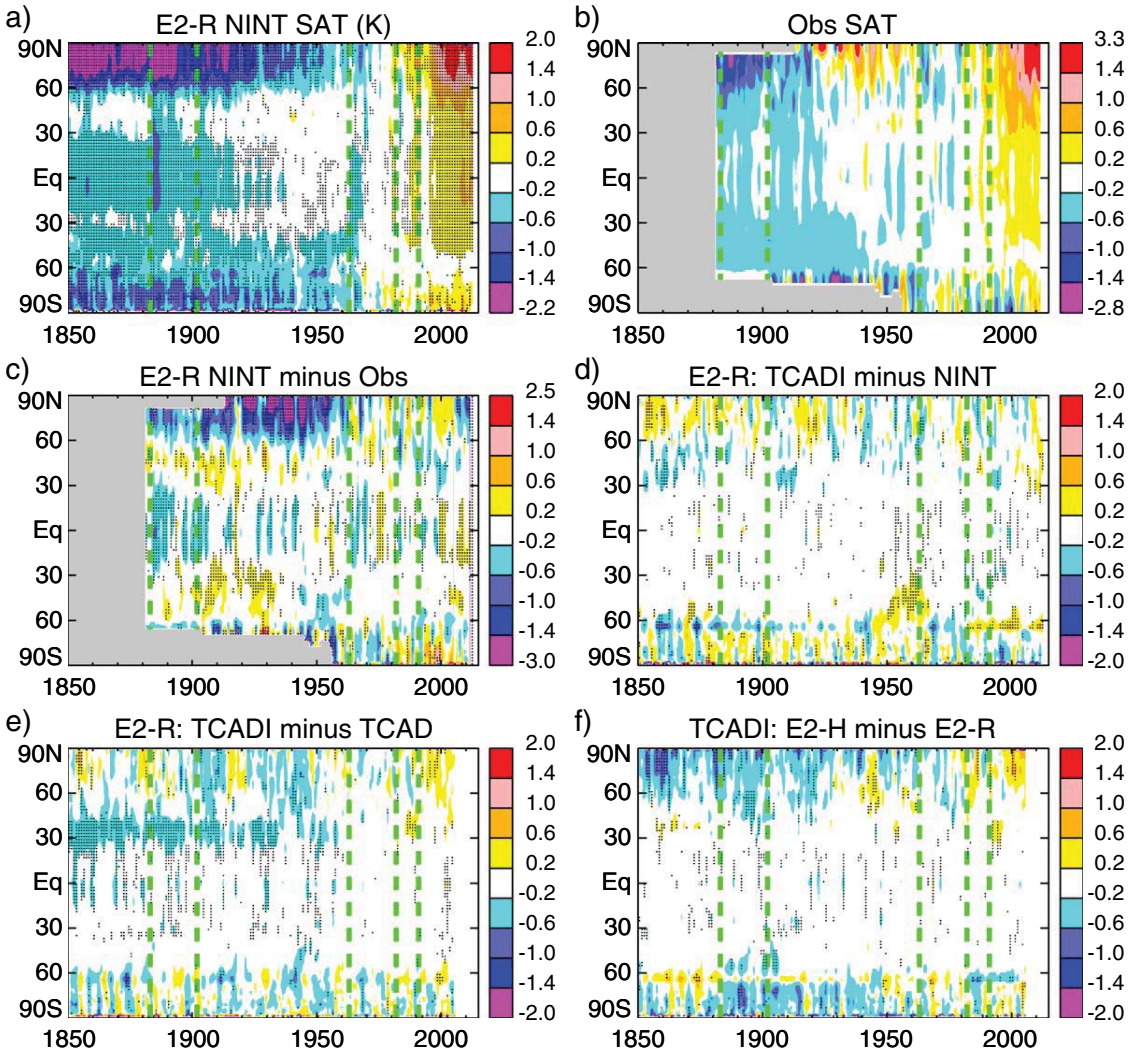

Figure 9. Zonal average of surface air temperature anomalies (K) for the model ensemble means and the GISTEMP analyses. The average anomaly at each latitude is defined to be zero between 1961 and 1990. Dots mark locations where the anomaly (or difference) is outside the 95% confidence interval that results from intraensemble variability. The dates of major volcanic eruptions (listed in Table 8) are marked by green dashed lines.

closer to the observed value, but the intraensemble standard deviation is around 0.5 K, less than the 1–3 K discrepancy of the ensemble mean with respect to observations during this period.

Zonal-mean warming is similar in the E2-R NINT, and TCADI ensembles (Figure 9d), with the largest temperature differences associated with uncorrelated interdecadal variability at high latitudes. The similar twentieth century trends at each latitude show the offsetting effect of climate sensitivity and tropospheric aerosol forcing, even though the latter is confined to specific latitudes. The E2-R TCAD and TCADI ensembles show similar temperature trends, except within the NH midlatitudes where the latter ensemble warms by roughly an extra 0.5 K over the entire historical period (Figure 9e). (This trend contrast is statistically distinct from zero.) This extra warming is contrary to expectation. The two ensembles are distinguished by their different treatment of the indirect effect, and the magnitude of the TCADI AIE is roughly twice the

Table 8. Major Volcanic Eruptions Marked in Figures[a]

Name	Location	Date
Krakatoa	Indonesia 6°S	Sep 1883
Santa Maria	Guatemala 15°N	Oct 1902
Agung	Bali 8°S	Jun 1963
El Chichon	Mexico 17°N	Apr 1982
Pinatubo	Philippines 15°N	Jul 1991

[a]The date indicates the month that zonal-average column optical thickness first exceeds 0.1 according to *Sato et al.* [1993].

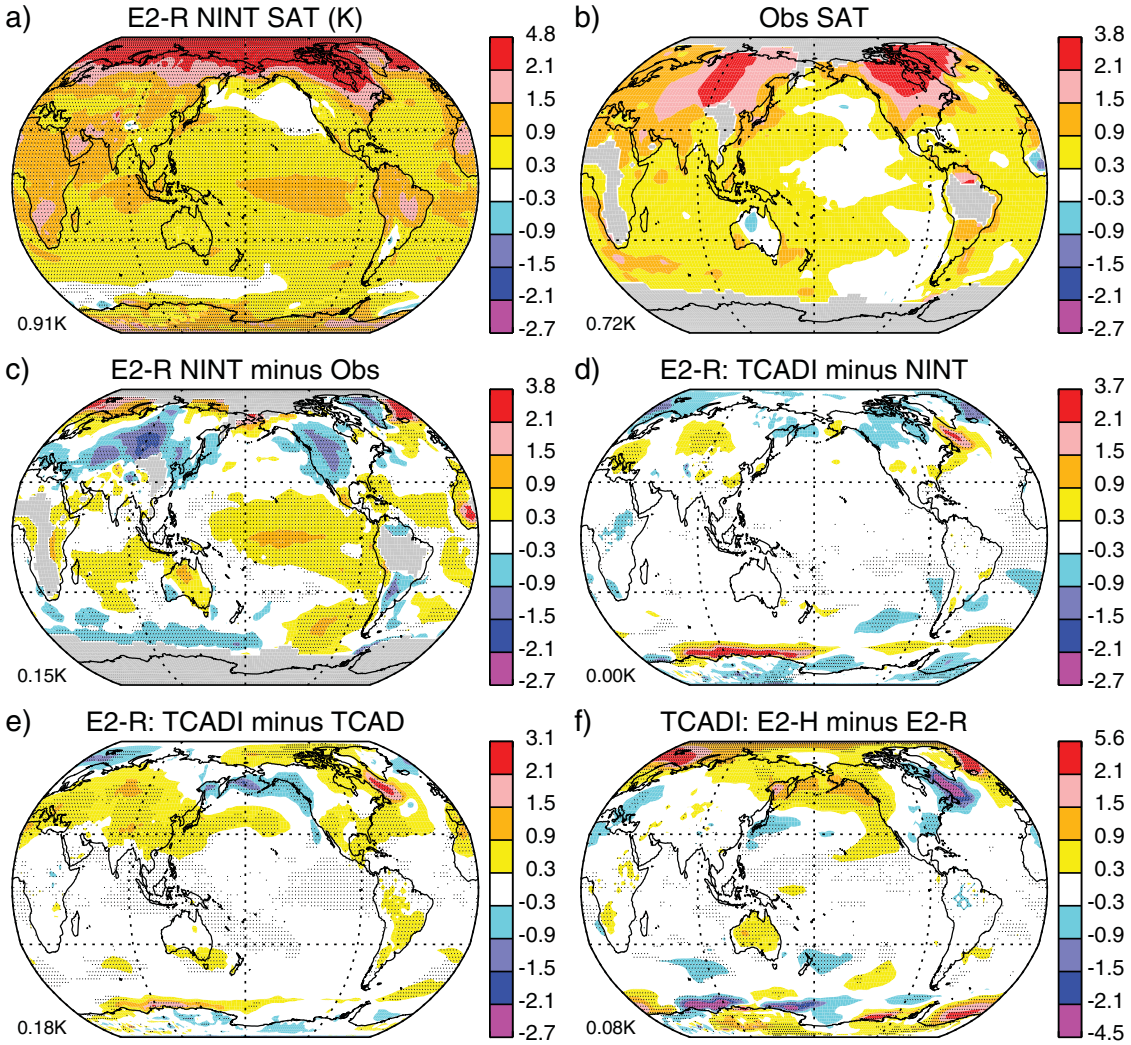

Figure 10. Contrast of decadal averages (1996–2005 minus 1880–1889) of surface air temperature (K). The number at the lower right of each panel is the global average (subject to data availability). Dots mark locations where the anomaly (or difference) is outside the 95% confidence interval that results from intraensemble variability.

TCAD value in the zonal average. (The TCAD AIE has a nearly identical regional distribution and global average compared to the NINT ensemble in Figure 5c.) This forcing contrast would ostensibly result in a weaker warming trend within the TCADI ensemble as aerosol concentrations increase throughout the twentieth century, contrary to the model behavior (Figure 9e).

The regional pattern of warming over the historical period is shown in Figure 10, which contrasts temperature during a recent decade (1996–2005) with the first decade of the GISTEMP analyses (1880–1889). For this and subsequent plots of regional trends, dots mark locations where the anomaly (or difference) is statistically distinct from zero, as in Figure 9. The E2-R NINT ensemble exhibits enhanced warming over high latitudes as described above, but warming is also enhanced over the continents compared to the ocean (Figure 10a), as expected from the small heat capacity of the land surface and its limited ability to accommodate changes in the incident radiative flux with evaporation. Differences between the E2-R NINT ensemble and the GISTEMP analysis are generally largest over elevated regions like the Tibetan Plateau, the North American Rockies, and Greenland, where the model shows less warming than observed (Figure 10c). In contrast, the model warming is excessive over the North Atlantic and Norwegian Sea.

The Tropics in the E2-R NINT ensemble warm slightly faster than observed (Figure 9c). The zonal contrast in model temperature across the equatorial Pacific is reduced, reminiscent of changes during an El Niño event

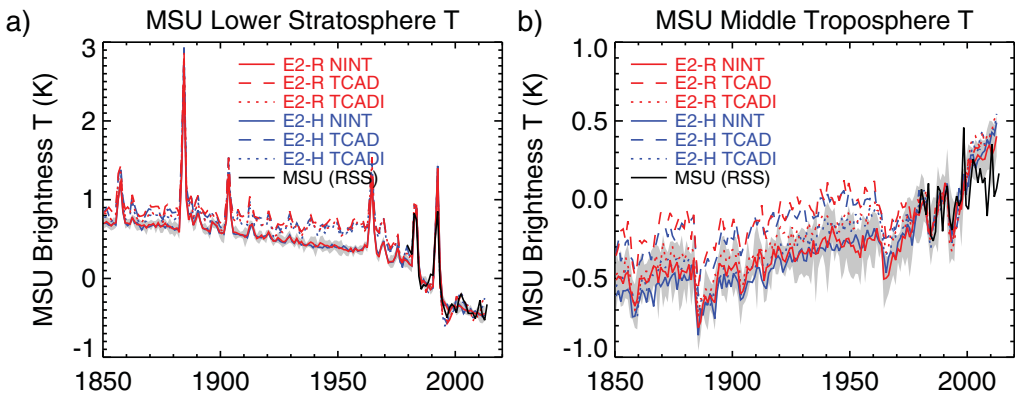

Figure 11. (a) Middle Troposphere and (b) Lower Stratosphere brightness temperature (K) from the Microwave Sounding Unit. Global and annual average anomalies are defined relative to the period 1979–2005. Gray shading shows variations among members within the E2-R NINT ensemble (plus or minus two standard deviations).

(Figure 10a). In contrast, GISTEMP shows an increase in the zonal contrast and a minimum in warming within the Central Pacific resembling a La Niña pattern (Figure 10b).

The greater globally averaged warming of the E2-H ensembles, compared to those of E2-R (Figure 7a), originates within the NH high latitudes (Figure 9f). Ocean heat content trends in the E2-H ensembles (section 4.4.) suggest that the excessive surface warming represents accumulated heat that is slow to be exported into the deep ocean. This interpretation is consistent with the relatively warm Arctic and Norwegian Sea in the E2-H TCADI ensemble, as well as the relative cooling in the vicinity of the Labrador Sea (Figure 10f). The latter is a region of Atlantic deepwater formation, where reduced heat export to the deep ocean is associated with decreasing heat convergence.

4.2. Upper Air Temperature

Greenhouse warming has a characteristic vertical structure with tropospheric warming accompanied by stratospheric cooling [*Manabe and Weatherald*, 1967]. We compare model warming above the surface to brightness temperatures retrieved from the Microwave Sounding Unit and Advanced Microwave Sounding Unit by Remote Sensing Systems (RSS) [*Mears and Wentz*, 2009, Version 3.3]. Discrepancies between trends derived from different upper air data sets are larger than differences of estimated trends of surface air temperature [*Thorne et al*., 2011]. The RSS trends are intermediate compared to two other MSU data sets. We use two retrieved variables: TLS ("Temperature of the Lower Stratosphere"), that is influenced by radiation originating from a layer centered near 20 km, and TMT ("Temperature of the Middle Troposphere"), whose layer of influence is centered near 10 km but extends into the lower stratosphere. Although TLS and TMT nominally distinguish between temperature of the stratosphere and troposphere, both retrievals are influenced by radiation over a broader depth and depend upon temperature at common levels.

We construct a model brightness temperature using vertical weights that are globally uniform and invariant with time, neglecting variations of surface emissivity and the vertical distribution of microwave absorbers, among other properties [*Hansen et al*., 1998]. The effect of this approximation has been calculated by *Santer et al*. [1999], who conclude that the error introduced is no more than a quarter of the global, seasonal trend of the midtropospheric TMT retrieval, and far less for the lower stratospheric TLS retrieval.

Global averages of retrieved and modeled brightness temperatures for the lower stratosphere and middle troposphere are shown in Figure 11, which shows the expected signature of increasing greenhouse gas concentrations with a warming troposphere accompanied by cooling within the stratosphere. The models show stratospheric cooling that closely matches the observed trend of -0.37 ± 0.16 K per decade between 1979 and 2005 (Table 7), with all ensemble means along with the individual members well within the observed trend uncertainty related to interannual variability (Figure 8). This variability primarily reflects the large but temporary response following the El Chichón and Pinatubo eruptions that injected volcanic aerosols above the tropopause, increasing the absorption of solar radiation and interrupting the cooling of the lower stratosphere (Figure 11a). The observed increase of the middle troposphere brightness temperature

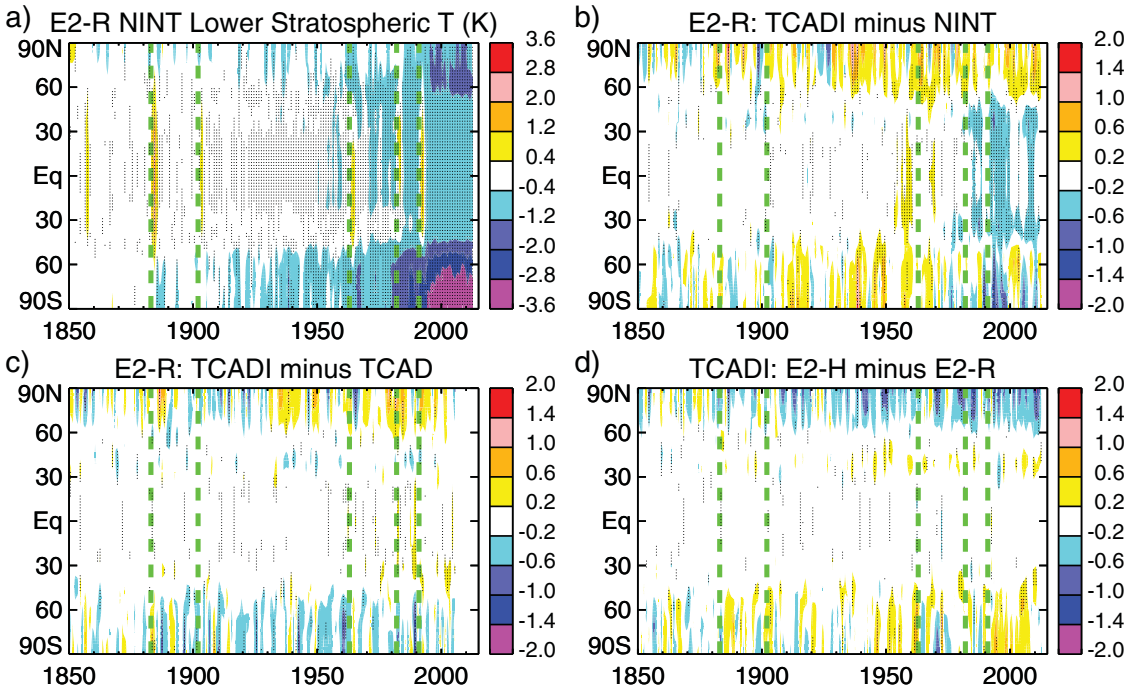

Figure 12. Zonal average of lower stratosphere brightness temperature (K). The average anomaly at each latitude is defined to be zero between 1850 and 1880. Dots mark locations where the anomaly (or difference) is outside the 95% confidence interval that results from intraensemble variability. The dates of major volcanic eruptions (listed in Table 8) are marked by green dashed lines.

of 0.106 ± 0.078 K per decade is smaller than warming at the surface (Figure 8 and Table 7), at least in part because the TMT retrieval includes a contribution from the cooling stratosphere [*Mears and Wentz*, 2009]. When the stratospheric contribution is removed, the surface and middle-tropospheric values are comparable [*Fu et al*., 2004].

Model warming of the midtroposphere is slightly larger than the retrieved value between 1979 and 2005 (Figure 11b), although each ensemble-mean trend is within the observed uncertainty that results from interannual variations, as are the trends of most ensemble members (Figure 8). Different methods of MSU TMT retrieval contribute additional uncertainty of roughly 0.05 K per decade to the observed trend [cf. *Thorne et al*., 2011, Figure 10]. The NINT and TCADI ensembles show the most rapid warming, consistent with their trends at the surface (Figure 7). The warming observed within the middle troposphere has slowed in recent years, coincident with the surface trend, and for the longer period of 1979–2012, the extended ensembles overestimate the MSU trend. Unusually warm years in the MSU TMT record like 1998 and 2010 follow El Niño events [cf. *Yulaeva and Wallace*, 1994]. While the models exhibit interannual variability in sea surface temperature in the equatorial East Pacific that resembles El Niño [*Schmidt et al*., 2014a], these events are primarily unforced and thus temporally uncorrelated with observed events. Simulations with prescribed SST would allow a more detailed comparison of the model to observed temporal variations of upper air temperature.

Century-scale variations in lower stratospheric temperature are shown as zonal averages of the model ensemble means in Figure 12. The E2-R NINT ensemble shows stratospheric cooling that begins by the mid-twentieth century and is largest at high latitudes, especially within the SH (Figure 12a). The exceptional cooling over Antarctica represents the combined effect of increasing greenhouse gas concentrations and stratospheric ozone depletion [*Shine et al*., 2003]. The E2-R TCADI ensemble shows recent tropical cooling that is slightly larger compared to the corresponding NINT ensemble (Figure 12b), consistent with greater ozone depletion in the former [*Shindell et al*., 2013b]. The TCADI ensemble also shows additional cooling over Antarctica following the Pinatubo eruption (Figure 12b), as volcanic aerosols increase the concentration of ozone-depleting substances, temporarily amplifying the ozone hole and cooling the lower

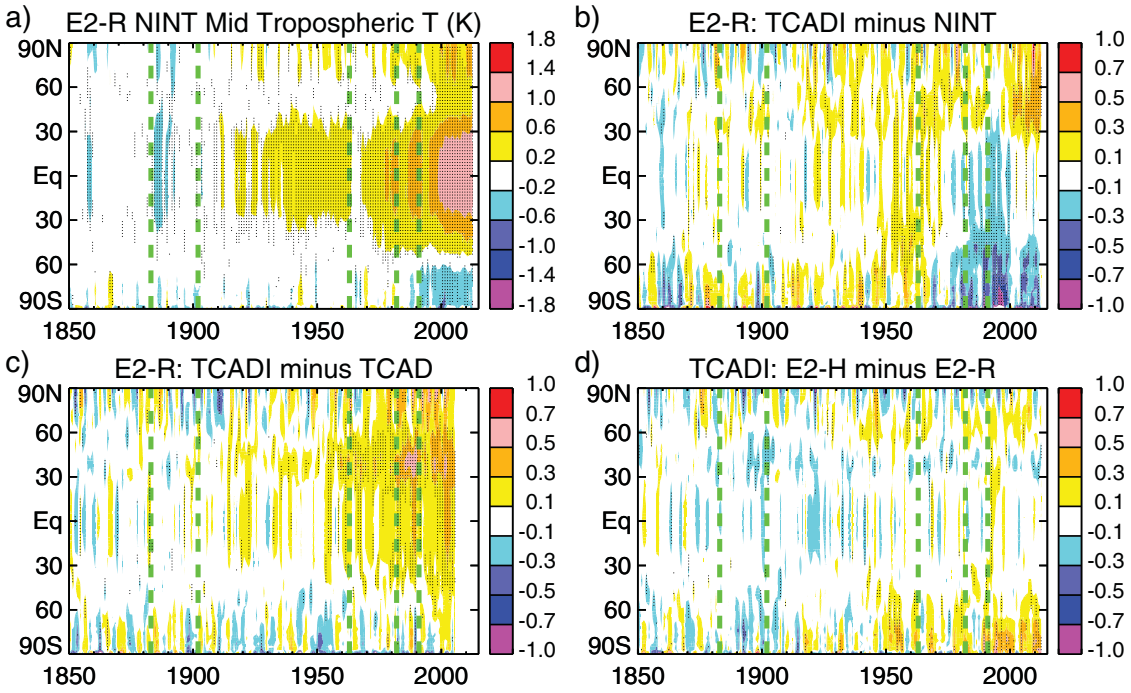

Figure 13. As in Figure 12 but for middle troposphere brightness temperature.

stratosphere. (Post-eruption reductions in ozone are not prescribed in the NINT ensembles.) Differences in stratospheric temperature trends among the other ensembles are small compared to interdecadal variations (Figures 12c and 12d).

The corresponding zonal-average trends for the midtroposphere brightness temperature are shown in Figure 13. At all latitudes, this brightness temperature represents the competing influence of proximal warming at the surface with less strongly weighted but larger stratospheric cooling (Table 7). The E2-R NINT ensemble shows warming at most latitudes that is temporarily interrupted following major volcanic eruptions (Figure 13a). The exception to warming is at high SH latitudes, where decadal trends of ozone depletion augment stratospheric cooling. The E2-R TCADI ensemble shows even greater SH cooling (Figure 13b), consistent with its TLS trend (Figure 12b). Brightness trends resulting from the treatment of the indirect effect and ocean resemble differences in the surface air temperature (Figures 9c, 9d, 13c, and 13d).

The regional distribution of trends in lower stratospheric brightness temperature is shown in Figure 14. Trends are defined by linear regression at each location between 1979 and 2005. In general, stratospheric cooling within the E2-R NINT ensemble is more zonally symmetric than the MSU retrieval (Figures 14b and 14c). The observed zonal asymmetry exceeds two standard deviations of the individual ensemble members, so its contrast with the model is not simply the result of ensemble averaging. Differences among the ensembles due to their treatment of atmospheric composition, the indirect effect, and ocean are smaller. (Figures 14d–14f).

Corresponding regional trends in middle troposphere brightness temperature are shown in Figure 15. The E2-R NINT ensemble generally reproduces the observed contrast between Antarctic cooling and warming at all other latitudes. However, the NINT ensemble overestimates the observed tropical warming, especially in the Central Pacific, as at the surface (Figure 10c), while underestimating the NH high latitude warming, especially over the Norwegian Sea and Arctic (Figure 15c). Differences between the ensembles are smaller and seem to be unsystematic, exceeding intraensemble variations in only a few locations.

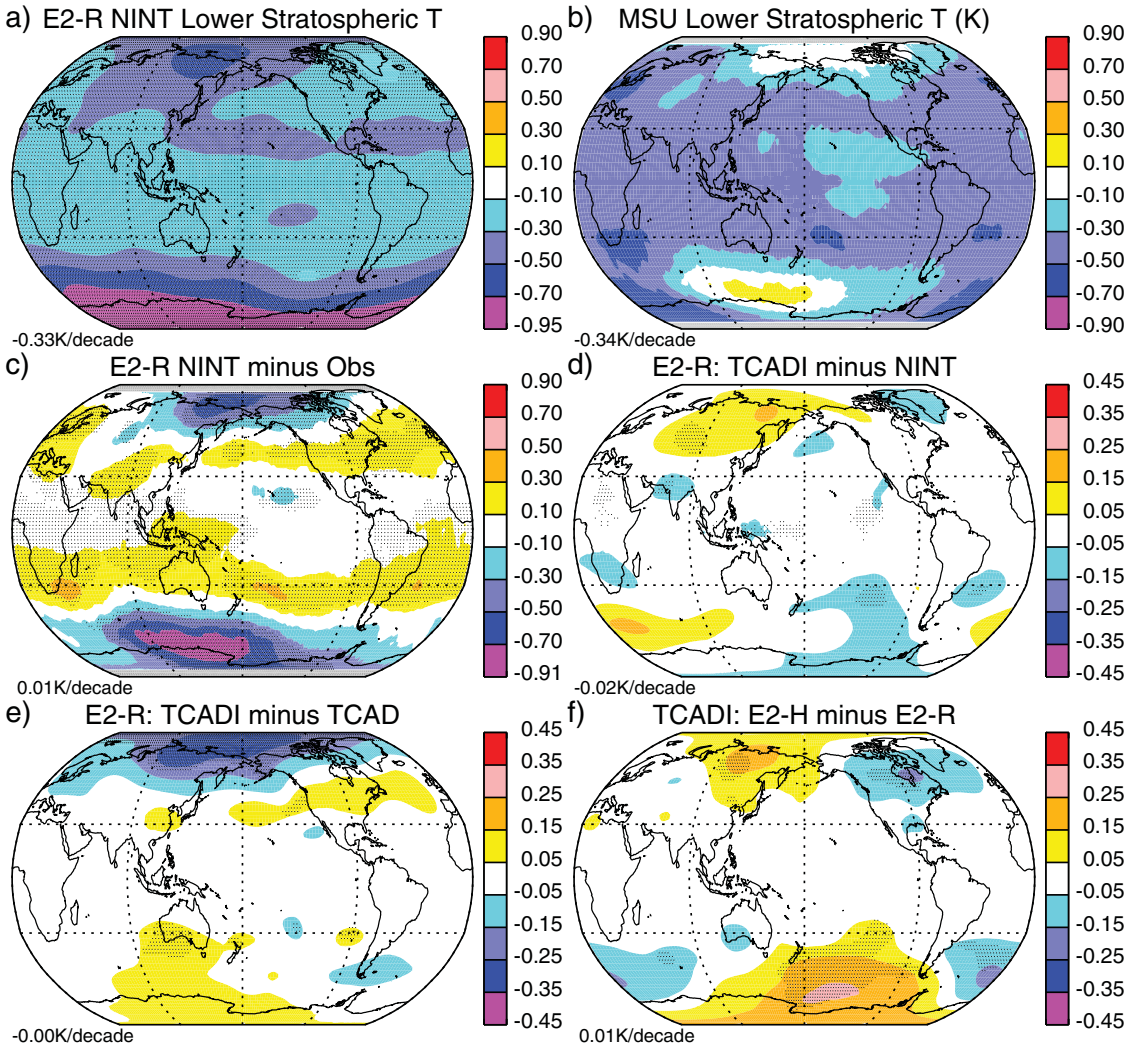

Figure 14. Change in lower stratosphere brightness temperatures (K per decade) between 1979 and 2005. The number at the lower right of each panel is the global average (subject to data availability). Dots mark locations where the anomaly (or difference) is outside the 95% confidence interval that results from intraensemble variability.

4.3. Sea Ice

The annual-average sea-ice area is shown in Figure 16 as a percentage of the total hemispheric area. For comparison, sea-ice area distributed by the National Snow and Ice Data Center (NSIDC) is plotted [*Fetterer et al.*, 2011], including a contribution from the unobserved northernmost latitudes of the Arctic Ocean (due to the satellites' inclined orbit) that assumes this region is always ice-covered.

The E2-R ensembles show excellent agreement with the Northern Hemisphere (NH) NSIDC annual average (Figure 16a). This performance is a clear improvement over the CMIP3 ModelE-R that overestimated the observed NH annual value [*Schmidt et al.*, 2006; *Hansen et al.*, 2007]. The agreement in Figure 16a results partly from compensating errors with respect to the observed seasonal cycle. All ModelE2 ensembles have too little sea-ice area at the September annual minimum, while the E2-R ensembles have too much at the March maximum [*Schmidt et al.*, 2014a]. The E2-H ensembles underestimate the NSIDC NH annual value, consistent with their warmer temperatures at high latitudes (Figure 10f).

Despite biases in the seasonal cycle, almost all ModelE2 ensembles reproduce the observed rate of decrease of NH ice area, both in the annual average and at the annual minimum and maximum (Figure 17). The exception is the E2-R TCAD ensemble that exhibits slow trends of both melting and warming in the late-

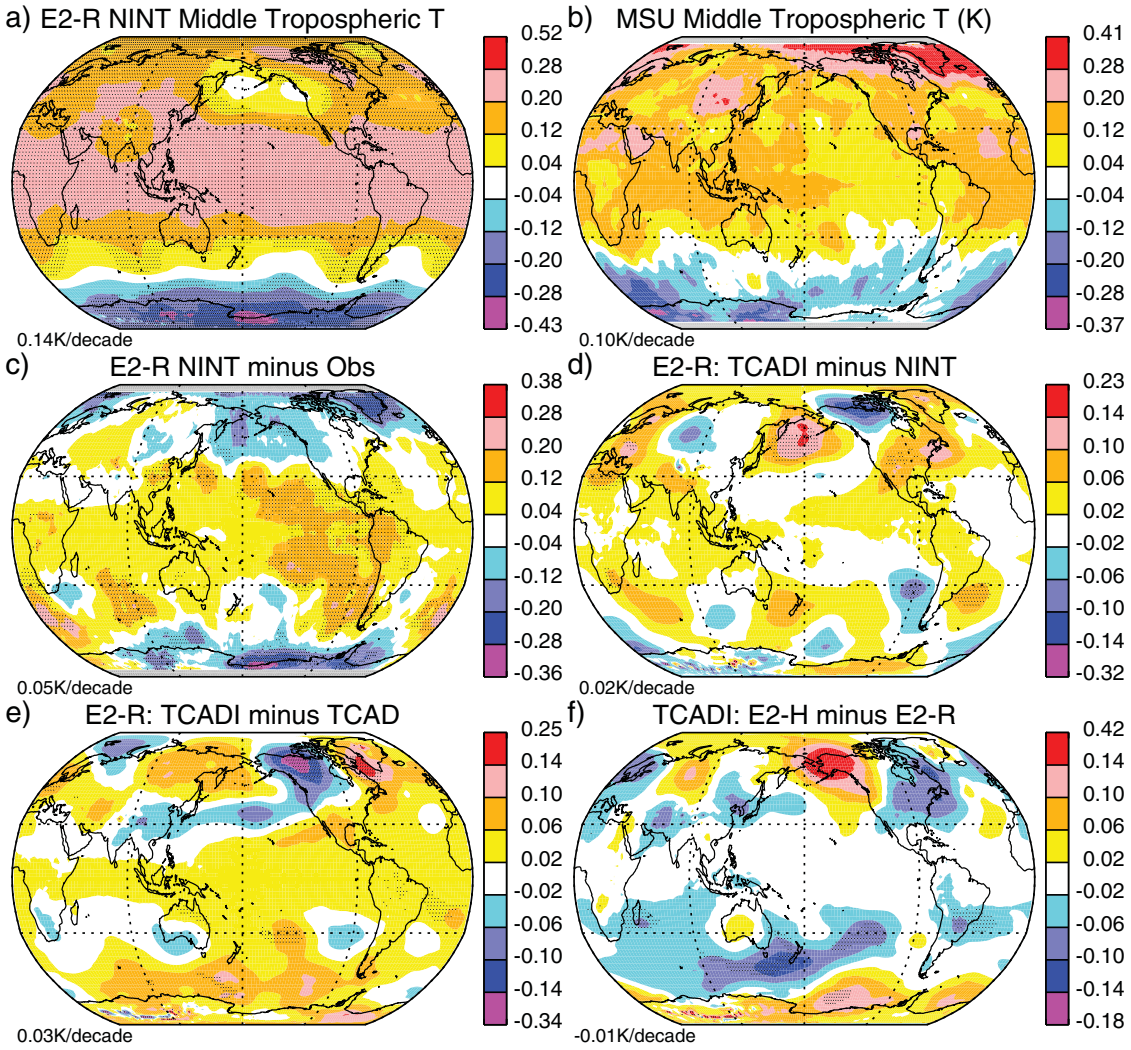

Figure 15. As in Figure 14 but for middle troposphere brightness temperature.

twentieth century (Figure 7). The E2-H ensembles reproduce the observed rate of melting, but due to insufficient present-day ice cover lead to estimates of an ice-free Arctic that are among the earliest of the CMIP5 models [*Stroeve et al*., 2012].

All ensembles underestimate Antarctic sea-ice area, simulating about half the retrieved annual average, while failing to reproduce the slight upward trend in the observations (Figure 16b). In this region, both ocean models exhibit excessive vertical mixing of warm water to the surface, an error that is exacerbated during the summer by insufficient low clouds and excessive surface shortwave absorption [*Schmidt et al*., 2014a].

4.4. Ocean Heat Content

The total forcing in Figure 2b represents an inferred energy imbalance at TOA that must be compensated at equilibrium, for example, by an increase in longwave radiation to space. Outgoing longwave radiation (OLR) increases with the atmospheric temperature, which at equilibrium is tightly coupled to the temperature of the upper ocean. This approach to equilibrium takes at least several centuries because the deep ocean removes heat from the upper ocean, slowing the rate at which the atmosphere warms and OLR increases. The incomplete compensation of the forcing is shown in Figure 18, which shows the net radiative flux at TOA for each ensemble along with the instantaneous forcing from Figure 2b. The net flux is nearly identical

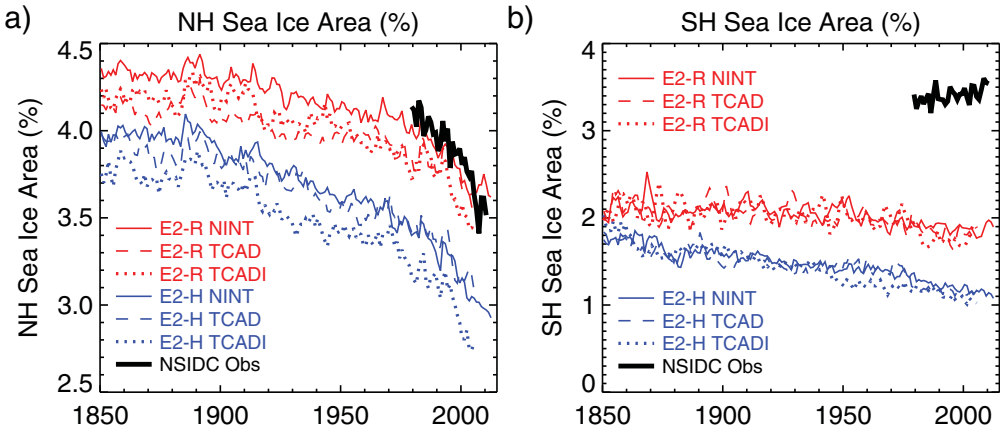

Figure 16. Model sea-ice area compared to passive microwave retrievals from the National Snow & Ice Data Center for the (a) Northern Hemisphere and (b) Southern Hemisphere. Plotted area is an annual average and hemispheric sum expressed as a percentage of the total hemispheric area.

for all the ensembles, suggesting similar OLR, despite different rates of surface warming. This similarity could result from compensation between the lapse-rate and water vapor feedbacks that is a consistent feature of climate models [*Soden and Held* 2006; *Held and Shell* 2012].

Warming of the ocean is indicated by an increase in the ocean heat content, shown for the historical period in Figure 19a. This heat content is a sum over all depths, and increases gradually with the forcing for all ensembles, despite temporary interruptions following major volcanic eruptions. Shading shows the range of variability among the members of the E2-R NINT ensemble, measured here by plus or minus twice the standard deviation of the intraensemble variations. The heat content of the E2-H ensembles fall below this range, indicating that ocean heat uptake is significantly smaller in these ensembles.

While the entire ocean column in the E2-R ensembles takes up more heat compared to the E2-H ensembles, the latter store more heat in the upper ocean (Figure 19a), the layer extending to depths of 776 and 700 m in the E2-R and E2-H ensembles, respectively. This depth is roughly the extent of the gyre circulations, and heat penetrates below this level mainly in regions of deepwater formation by ocean deep convection. Thus, the E2-H TCADI ensemble accumulates more heat in the upper ocean compared to the corresponding E2-R model, despite taking up less heat when summed over all depths. This is consistent with the more rapid warming at the surface in the E2-H TCADI ensemble (Figure 7a). It is also consistent with the reduced rate of warming in the Labrador Sea in this ensemble compared to E2-R (Figure 10f) that indicates a reduction of

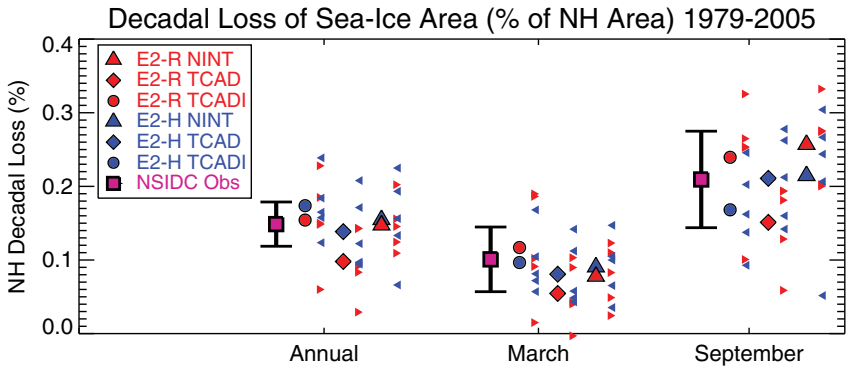

Figure 17. Change in NH sea-ice area (per decade and as a percentage of total hemispheric area) between 1979 and 2005. The observed value from the NSIDC is in purple with its uncertainty in black (corresponding to two standard deviations of interannual variability). Ensemble average change is shown as triangles (NINT), diamonds (TCAD), and circles (TCADI), with change for individual ensemble members plotted as small triangles.

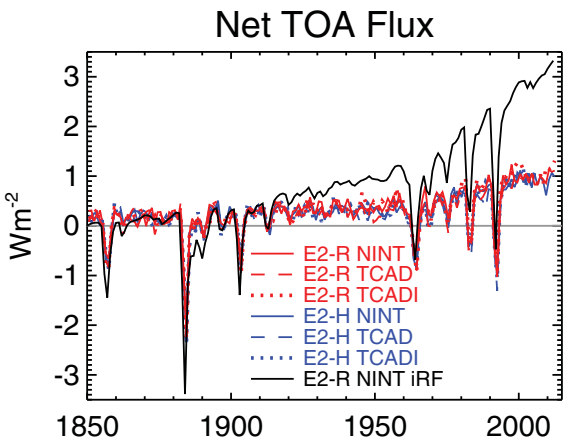

Figure 18. Net radiative flux at TOA (Wm^{-2}) for the ensemble averages. The black line shows instantaneous radiative forcing at the tropopause ("iRF") for the E2-R NINT ensemble. Ensemble values are global and annual averages.

heat convergence associated with sinking of surface water into the deep ocean. In contrast, the E2-R ensembles transfer heat into the deep ocean to a greater extent, slowing the warming at the surface.

Observations of the global heat content of the upper ocean extend back to the mid-twentieth century. Figure 19b shows the heat content of the upper 700 m, according to analysis of subsurface temperature measurements by *Domingues et al.* [2008], extended to the present by *Church et al.* [2011]. The uncertainty, indicated by purple shading, is largest early in the record, and arises from a combination of measurement error and limited spatial coverage, especially within the Southern Ocean. Other global estimates of heat content within the upper ocean are based upon different treatments of the measurement limitations, but yield trends that are consistent, given the uncertainty [e.g., *Levitus et al.*, 2005]. During the past decade, global coverage of the ocean has been achieved with more accurate profilers, resulting in more confident estimates of the observed global trend. Figure 19b shows the estimate during the past two decades by *Lyman et al.* [2010], with gray shading denoting error bars that are small compared to estimates during previous decades. The warming associated with the E2-R ensembles is in agreement with the trend inferred by both observational analyses. In contrast, the E2-H ensembles warm excessively, possibly because of insufficient heat export to the deep ocean below 700 m.

The reduction of heat export to the deep ocean when an isopycnal OGCM is replaced with a model based upon height coordinates is consistent with behavior exhibited by the Geophysical Fluid Dynamics Laboratory coupled model [*Hallberg et al.*, 2012]. Measuring heat export and the sensitivity of model export to subgrid mixing remains an important challenge.

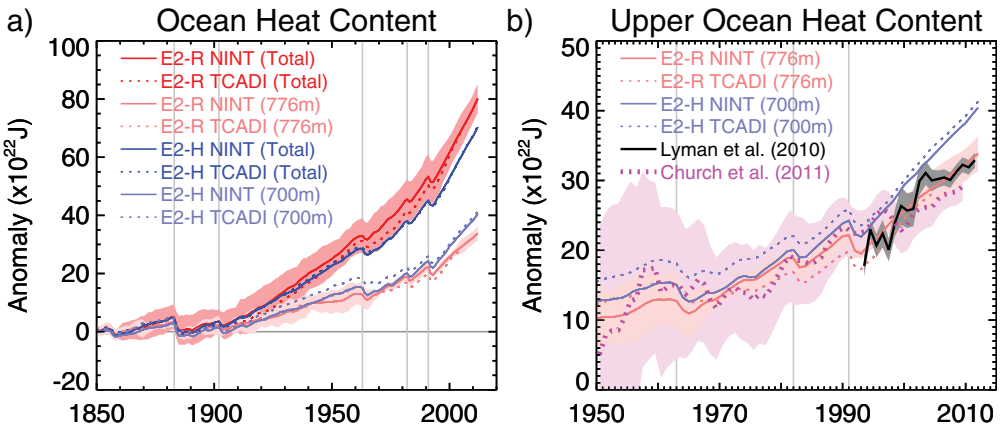

Figure 19. (a) Increase in ocean heat content since 1850 in the NINT and TCADI ensembles. Red and blue lines indicate heat content summed over all depths; light red and blue lines indicate heat within the upper 776 and 700 m of the ocean for the E2-R and E2-H ensembles, respectively. Pink and light pink shading shows the intraensemble uncertainty in the E2-R NINT ensemble (equal to plus or minus two standard deviations) for the total and upper ocean heat content. (b) Trends in upper ocean heat content compared to measurements in the upper 700 m analyzed by *Church et al.* [2011] and *Lyman et al.* [2010], whose uncertainties are denoted by purple and gray shading, respectively. Observed anomalies are adjusted so that their time average matches the E2-R NINT average over the same period. Vertical lines mark major eruptions of tropical volcanoes.

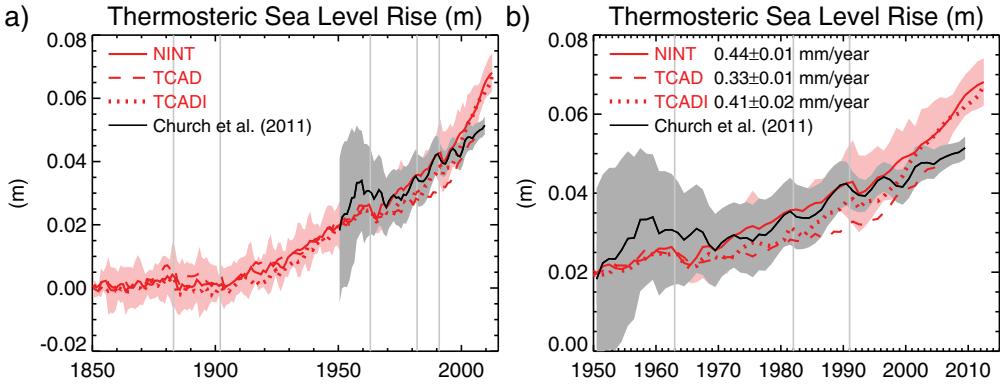

Figure 20. (a) Global thermosteric sea-level rise (m) since the preindustrial in the E2-R ensembles, along with the observational estimate from density changes in the upper 700 m by *Church et al.* [2011]. The observations are offset to have the same time-average as the NINT ensemble over their common period, so that trends can be more easily compared. Pink shading shows variations among the NINT ensemble members (plus or minus two standard deviations), while the gray shading represents twice the standard error of the observational estimate of *Church et al.* [2011] based upon a combination of tide gauges and satellite altimetry retrievals. Ensemble-mean trends are computed between 1900 and 1990. Vertical lines mark major eruptions of tropical volcanoes. (b) excerpt since 1950.

4.5. Sea Level

Warming during the historical period is associated with an expansion of ocean volume. Density changes (including the effect of salinity anomalies) contribute a little under half of the observed rise in sea level during the past four decades [*Church et al.*, 2011]. Melting of glaciers and the continental ice sheets of Greenland and Antarctica supply most of the remainder. The observed rise in sea level is slightly reduced by an increase in terrestrial water storage by river dams offset by groundwater depletion. The model does not simulate the dynamical disintegration of land ice and has unphysical trends in lake storage, so we present in Figure 20 only the density-related or thermosteric changes in E2-R sea level. (Values for the E2-H model are unavailable due to a diagnostic error.) For comparison, the observed thermosteric sea-level rise limited to temperature and salinity changes above 700 m is also plotted [*Church et al.*, 2011]. Density changes below this depth are omitted due to infrequent measurements prior to the last decade. (To facilitate comparison of trends, the observations are given the same time-averaged value as the E2-R NINT ensemble over their common period.)

Figure 20 shows that thermosteric sea level is reduced in both the model and observations following major volcanic eruptions. The model trend has increased in the past two decades (Figure 20b), consistent with measurements of total sea level by both tide gauges and satellite altimeters [*Church et al.*, 2011; *Nerem et al.*, 2010]. Thermosteric trends in each ensemble after 1990 are three times the trend earlier in the century. Between 1993 and 2009, the modeled thermosteric contribution to sea level increases by just over 1 mm per year in each ensemble: twice the observed trend of 0.65 $\pm$ 0.12 mm per year. The omission of the observed thermosteric change below 700 m makes an unknown contribution to this discrepancy. Excessive warming during the past decade by the NINT and TCADI ensembles also contributes (Figure 7b).

Since the preindustrial, the total sea level rise is estimated near 210 mm [*Church and White*, 2011]. The thermosteric change of the NINT and TCADI ensembles since 1850 is 65 $\pm$ 15 mm, about one-third of the total observed value.

4.6. Precipitation

Climatological precipitation varies over a smaller spatial scale than temperature, making precipitation anomalies more difficult to detect with present-day in situ measurements. Satellites offer extensive spatial coverage (at least within the Tropics) but retrieving rain rates from measured radiances is not straightforward. Evaluation of climatological precipitation in each ensemble during the satellite era is described by *Schmidt et al.* [2014a].

Precipitation changes are related to the tropospheric radiative divergence anomaly [*Allen and Ingram*, 2002]. The warming associated with rising greenhouse gas concentrations, for example, increases this

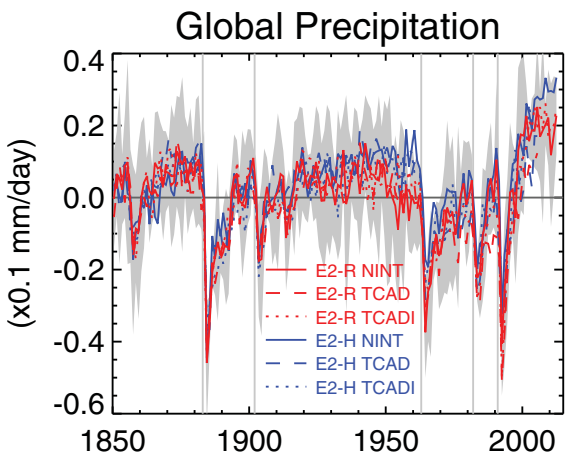

Figure 21. Global precipitation anomaly (in tenths of a mm day^{-1}). Shading represents variations among ensemble members within the E2-R NINT ensemble (plus or minus two standard deviations). Major volcanic eruptions are marked with vertical lines.

divergence, requiring greater latent heating (and therefore precipitation) to maintain energy balance. In contrast, many aerosol species reduce this divergence, and thus precipitation.

Global precipitation increases over the historical period in each ensemble (Figure 21), coincident with the increasing temperature. The total increase since the preindustrial is on the order of 0.02 mm day^{-1} compared to the observed global annual mean near 3 mm day^{-1} [*Stephens et al*., 2012]. The largest absolute increase of precipitation by the end of the historical period is associated with the E2-H NINT ensemble (Figure 21), whose warming is also largest (Figure 7a and 11b).

For each ensemble, global precipitation changes during the twentieth century by roughly 0.6% (E2-R) or 0.7% (E2-H) per degree K of warming. Precise estimation of the fractional increase of global precipitation is difficult because variability is large compared to trends over the historical period. Global precipitation decreases following Pinatubo [*Trenberth and Dai* 2007]. The model decrease following this eruption is over twice as large as the simulated increase over the entire twentieth century, and comparable reductions occur following other major eruptions (Figure 21). Comparison of the model fractional increase over the twentieth century to changes in other models compiled by *Held and Soden* [2006, Figure 2d] suggests that the ModelE2 sensitivity during this period is on the low side, although a few of these other models show a decrease in twentieth-century precipitation.

Precipitation sensitivity during the historical period is influenced by aerosol forcing that varies among models to a greater extent than forcing by long-lived greenhouse gases [e.g., *Shindell et al*., 2013a]. As a more robust comparison of sensitivity with other models, we calculate the fractional change of precipitation in an idealized CMIP5 experiment where aerosol forcing is absent and the climate is perturbed solely by a 1% per year increase of the CO_2 concentration until quadrupling. Over the initial 150 years of this simulation, the fractional precipitation increase is 1.3% (E2-R) or 1.4% (E2-H) per degree K of warming. This is virtually identical to the mean increase calculated among the CMIP5 models by *Pendergrass and Hartmann* [2014]. For comparison, the fractional increase for the CMIP3 counterpart to the E2-R NINT model is over twice as large [*Lambert and Webb*, 2008], for reasons that bear further investigation. The higher precipitation increase during the experiment forced only by increasing CO_2 compared to the historical experiments illustrates the ability of aerosols to reduce precipitation.

Zonal-average changes in precipitation during the historical period are shown in Figure 22. A 5 year low-pass filter is applied to remove residual El Niño variability that is not completely damped by averaging over ensemble members. Without filtering, this variability obscures multidecadal trends along with differences between the ensembles. In the E2-R NINT ensemble, precipitation increases in the deep Tropics, while decreasing in the subtropics (Figure 22a). This spatial pattern is consistent with trends calculated from Global Precipitation Climatology Project retrievals [*Allan et al*., 2010], and theoretical expectations of precipitation sensitivity to increasing temperature and net radiation into the surface [*Held and Soden*, 2006]. The precipitation increase is also consistent with the "upped-ante" mechanism where tropical precipitation intensifies and contracts areally where it is already large. (This is because latent heat release increases the upper tropospheric temperature, increasing the static stability throughout the entire Tropics, including the edge of the convecting region that originally was only marginally unstable) [*Chou and Neelin* [2004]. The ensemble also shows increased precipitation within the Northern Hemisphere extratropical storm track as well as near the Antarctic coast (Figure 22a). Differences between the models are relatively small compared to the E2-R NINT trend (Figures 22b–22d).

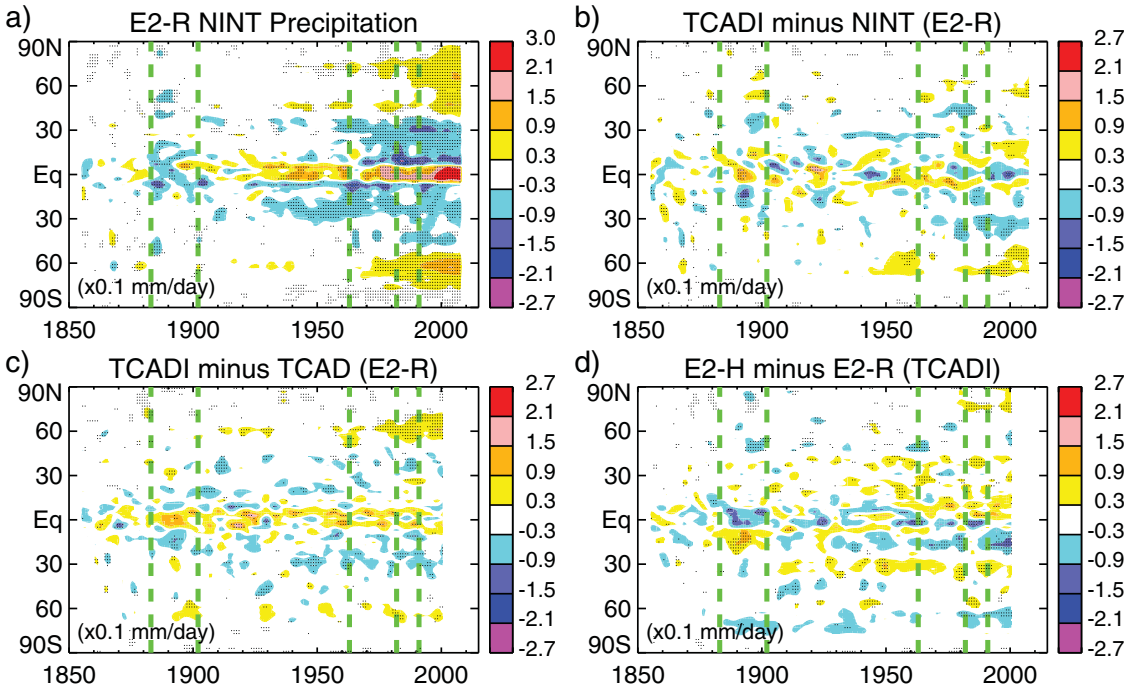

Figure 22. Zonal average precipitation anomaly compared to the control experiment (in tenths of a mm day^{-1}). The average anomaly at each latitude is defined to be zero between 1850 and 1880. The dates of major volcanic eruptions (listed in Table 8) are marked by green dashed lines. To remove uncorrelated El Niño variability between the simulations that is not completely damped by ensemble averaging, a 5 year low-pass filter is applied. Dots mark locations where the anomaly (or difference) is outside the 95% confidence interval that results from intraensemble variability.

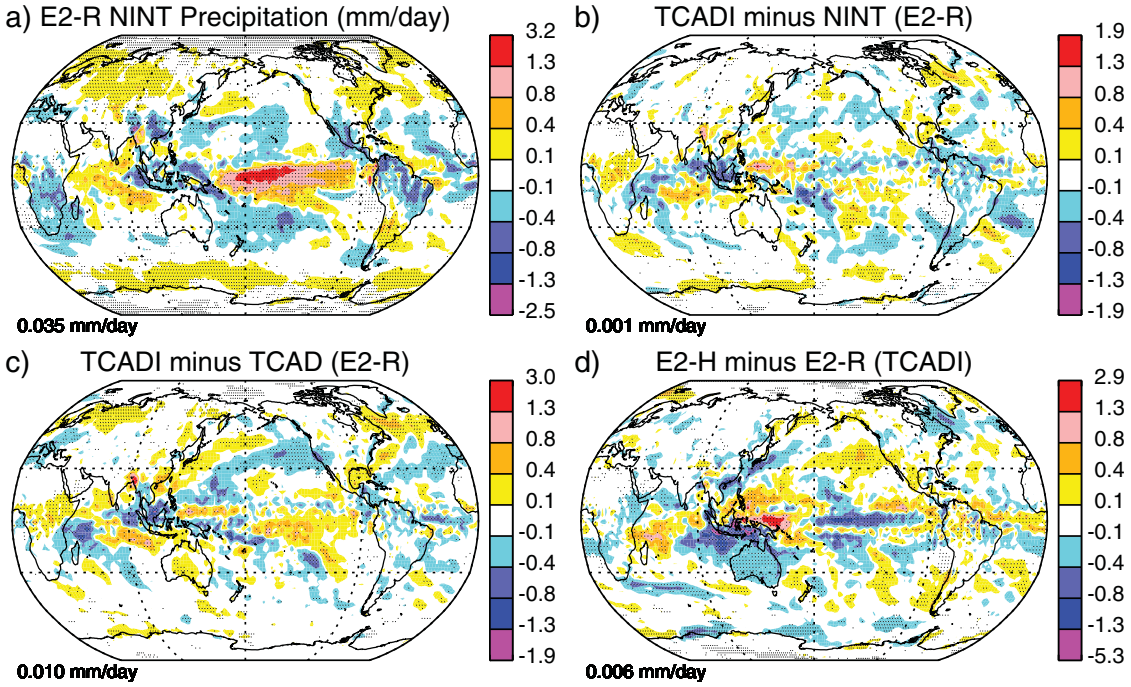

Figure 23. Contrast of decadal averages of precipitation (1996–2005 minus 1880–1889 in mm/day). The number at the lower right of each panel is the global average. Dots mark locations where the anomaly (or difference) is outside the 95% confidence interval that results from intraensemble variability.

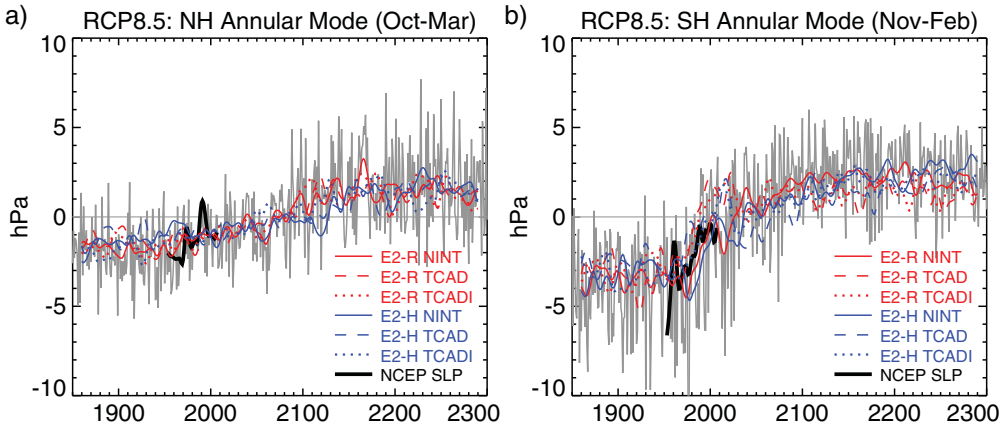

Figure 24. Leading principal component (PC) of sea-level pressure (hPa) during the historical period and RCP8.5 scenario. The PC represents the pressure anomaly corresponding to the annular mode, averaged poleward of 60° for (a) the NH (October-March) and (b) the SH (November–February). The colored lines represent the PC of each ensemble after filtering to remove subdecadal frequencies. To allow comparison of long-term trends with interannual variations, the gray line shows the unfiltered seasonal average PC for the E2-R NINT ensemble. The black line shows the corresponding NCEP PC between 1948 and 2012, with only periods shorter than 5 years filtered to preserve PC values at the beginning and end. The model PCs are constructed to average zero over their duration. To facilitate comparison of trends, the observed PCs are offset to have the same time average as the E2-R NINT PCs during their period of overlap.

Regional differences in decadal-mean rainfall between the late nineteenth century (1880–1889) and the end of the twentieth century (1996–2005) are shown in Figure 23. In the NINT ensemble, the most prominent feature is the displacement of rainfall from the West Pacific and subtropics toward the Central Pacific (Figure 23a), consistent with the El Niño-like warming in Figure 10a. Despite this warming, the ensemble predicts drying over Mexico and the American Southwest, in contrast to the canonical El Niño precipitation anomaly in these regions [*Ropelewski and Halpert*, 1996; *Seager et al*., 2005]. There is an increase over the interior of the Eurasian continent, eastern North America, and within the Atlantic and Pacific storm tracks. In contrast, precipitation is reduced over the Mediterranean and the Sahel. In the E2-H TCADI ensemble, precipitation increases over the entire equatorial Pacific, with a maximum west of the dateline (Figure 23d).

4.7. Circulation Trends

The climate anomalies described above are coupled to trends in the atmospheric and ocean circulation. Annular modes are the leading form of atmospheric variability in the extratropics of both hemispheres [*Thompson and Wallace*, 2000]. By convention, the positive phase of each mode represents an equatorward displacement of mass from high latitudes to the subtropics, with a poleward shift of the westerlies within the troposphere and at the surface. Previous studies have shown that both stratospheric ozone loss and increasing concentrations of greenhouse gases drive positive annular trends [e.g., *Shindell and Schmidt*, 2004], although not all models simulate this response [*Miller et al*., 2006b].

Figure 24 shows a positive annular trend within each hemisphere between the start of the historical period and the end of the twenty-third century, assuming the RCP8.5 scenario for future atmospheric composition, whose radiative forcing increases to approximately 8.5 and 12 Wm^{-2} by 2100 and 2300, respectively [*van Vuuren et al*., 2011]. (We include model output past the end of the historical period to distinguish forced trends from natural variability.) Ensembles based upon the RCP4.5 scenario with weaker anthropogenic forcing also show positive, albeit smaller, annular trends. Annular trends are represented by the first principal component (PC) of sea-level pressure (SLP) in each hemisphere, rescaled to represent the average pressure anomaly poleward of 60° in hPa. The PC is calculated from monthly anomalies between October and March in the NH, and November through February in the SH for comparison to trends in the CMIP3 models calculated by *Miller et al*. [2006b]. The colored lines in Figure 24 show the PC for each ensemble after filtering subdecadal time scales. To allow comparison of the trend to interannual variations, the gray line shows unfiltered seasonal averages of the PC for the E2-R NINT ensemble. Both hemispheres show a positive trend, and by association, a poleward displacement of the westerlies. The trend is apparent first in the SH, by the end of the twentieth century, due to stratospheric ozone depletion that augments the effect of greenhouse

gases. The SH trend is seen first in the TCAD and TCADI ensembles, where prognostic ozone loss is slightly stronger than the observed depletion used to prescribe the NINT ozone concentration [*Shindell et al*., 2013b]. Because of their excessive depletion, the TCAD and TCADI ensembles exhibit greater ozone recovery, consistent with their smaller PC in later centuries compared to the NINT ensemble. For comparison, the figure shows the annular PC calculated from National Center for Environmental Prediction (NCEP) SLP between 1948 and 2012 [*Kalnay et al*., 1996]. Because each time-averaged PC is originally zero by construction, only the ensemble trends (and not the individual PC values) can be meaningfully compared between the reanalysis value. (To facilitate comparison of trends, we shifted the observed PC in Figure 24 so that its time-mean equals that of the E2-R NINT PC during the observing period.) In the SH, late twentieth-century trends in the ensembles are similar to the reanalyzed PC trend, although the comparison remains imprecise due to decadal variability. (Only periods shorter than 5 years are filtered from the NCEP PC to preserve its values at the beginning and end of the reanalysis period.) In the NH, ensemble trends are not distinguishable from decadal variations until almost the end of the twenty-first century, consistent with the indistinct trend in the present-day reanalysis PC.

5. Conclusions

We have described the forcing and response in the NASA GISS ModelE2 CGCM for the CMIP5 historical period between 1850 and 2005, with extensions of some ensembles through 2012. There are three versions of the atmospheric model distinguished by their treatment of atmospheric constituents and the aerosol indirect effect. In the NINT model, atmospheric composition is specified based upon previous offline calculations, while the indirect effect consists of an empirical relation between aerosol mass and low cloud cover whose global magnitude is tuned [*Hansen et al*., 2005]. In the TCADI model, ozone and aerosol concentrations are calculated interactively as a function of atmospheric chemistry and climate, while the concentrations of all long-lived greenhouse gases other than CO_2 are prescribed at the surface but calculated above [*Shindell et al*., 2013b]. The first indirect effect is implemented as a prognostic equation for the number of available CCN that balances supply by aerosols and removal by wet deposition. The TCAD model is a hybrid, sharing prognostic atmospheric constituents with TCADI but an empirical AIE with NINT. Each of the three versions of the ModelE2 AGCM is coupled to either the HYCOM or the Russell OGCM, resulting in six versions of the coupled ModelE2, each integrated as a five-member ensemble.

The total forcing by the end of the historical period is in the upper quartile of CMIP5 models [*Forster et al*., 2013]. Compared to the previous (CMIP3) generation of ModelE, year-2003 forcing in the E2-R NINT ensemble is higher by 0.6 Wm^{-2}. This difference is mainly attributable to larger forcing by long-lived greenhouse gases, along with reduced direct and indirect forcing by aerosols (cf. Table 6). Total forcing is smaller in the E2-R TCADI ensemble due to greater stratospheric ozone depletion and more negative aerosol forcing.

The E2-R NINT ensemble reproduces the observed warming of the surface and troposphere along with stratospheric cooling until the end of the historical period in 2005 (Figure 8). However, the ensemble continues to warm at a steady rate without the slowdown observed during the last decade. The warming discrepancy is larger than unforced differences among the individual ensemble members, pointing to possible errors related to forcing or some aspect of the model that we have yet to identify. The E2-R NINT ensemble matches the observed upward trend of heat content in the upper 700 m of the ocean. (We are investigating whether this agreement remains after correcting an error in the subgrid mixing parameterization.) The ensemble reproduces the annual loss of NH sea ice observed during recent decades, along with observed trends of maximum and minimum area (Figure 17). However, sea-ice area in the SH decreases in contrast to the small observed increase. Precipitation increases in the deep Tropics during the late twentieth century at the expense of the subtropics, associated with an eastward displacement of precipitation toward the equatorial Central Pacific that resembles El Niño.

Trends in the other ensembles with respect to NINT are influenced by their different climate sensitivity, forcing, and ocean heat uptake. Temperature and sea-ice trends are similar in the E2-R TCADI ensemble with its greater climate sensitivity offset by smaller ozone forcing and larger aerosol-cloud interactions. The TCAD ensembles warm the least, sharing the relatively low NINT climate sensitivity, but smaller ozone forcing with the TCADI ensembles. Warming is largest in the E2-H ensembles, where heat accumulates at the surface due to small export to the deep ocean (Figure 19). By the end of the twenty-first century, when forcing

by long-lived greenhouse gases dominates the total forcing, the E2-H TCADI ensemble is the warmest, due to its elevated climate sensitivity and slow heat uptake by the deep ocean, as will be reported elsewhere.

The contrast between the TCADI and TCAD climates shows the effect of a mechanistic treatment of the AIE. *Schmidt et al*. [2014a] shows that cloud fraction errors are reduced in the TCADI ensembles downwind of regions of large aerosol emission like the North Pacific, the Labrador Sea, and Gulf of Guinea. Attribution of reduced errors to a mechanistic AIE treatment is challenging due to the complexity of cloud formation, but more comprehensive studies have found a similar reduction [*Ekman*, 2014]. The NH midlatitudes warm faster in the TCADI ensembles (Figure 9e), even though the aerosol forcing is more negative. This unanticipated contrast of trends shows the limitation of simplified, empirical treatments of the AIE that cannot anticipate the behavior emerging from more mechanistic representations.

All ensembles show positive annular trends corresponding to a decrease in Arctic and Antarctic surface pressure with a corresponding poleward shift of the midlatitude westerlies. Annular trends are apparent in the SH by the end of the twentieth century, associated with stratospheric ozone depletion that augments the effect of increasing greenhouse gas concentrations. The trend appears first in the TCAD and TCADI ensembles with prognostic ozone, whose loss slightly exceeds the observed value used to prescribe depletion in the NINT ensembles. In contrast, annular trends in the NH, where stratospheric ozone loss is negligible, cannot be distinguished from unforced variability until the late twenty-first century.

Particular aspects of model behavior deserve further investigation: for example, the increase in long-lived greenhouse gas forcing and fractional precipitation sensitivity compared to the previous generation (CMIP3) model. It is also of interest what aspect of climate change between the preindustrial and present day causes direct forcing by the same aerosol distribution to become substantially more negative (Table 5).

We are preparing new TCADI simulations that correct deficiencies while incorporating updated forcing values that were not available at the time of the original simulations. Preliminary analysis suggests that the prescribed forcing was biased upward, and that corrected values from observations will reduce the simulated warming trend toward the observed value [*Schmidt et al*., 2014b]. Given the remaining uncertainty associated with aerosol forcing and deep ocean heat uptake, reproduction of the observed warming is not a demonstration that the model has the correct climate sensitivity. Previous generation (CMIP3) models simulated the twentieth-century warming despite a wide range of sensitivity [*Kiehl*, 2007]. However, in the latest generation of models, sensitivity and forcing are nearly uncorrelated [*Forster et al*., 2013; *Shindell et al*., 2013a], so reproduction of the observed climate is a more stringent test. Historical simulations remain an important opportunity for comparison to observed climate trends and processes that helps to assess projections of future climate change [*Otto et al*., 2013; *Sherwood et al*., 2014].

Acknowledgments

We thank two anonymous reviewers for their substantial effort to improve this article. CMIP5 simulations with the GISS ModelE2 were made possible by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center. Development of ModelE2 was supported by the NASA Modeling, Analysis, and Prediction (MAP) Program with additional support from the National Science Foundation, the National Oceanic and Atmospheric Administration, and the Department of Energy. We thank Ellen Salmon and the NCCS staff for hosting and providing convenient access to the model output. MSU data are made available by Remote Sensing Systems and supported by the NOAA Climate and Global Change Program and the NASA Earth Science Division. Sea-ice area is provided by the National Snow and Ice Data Center. The Division of Marine and Atmospheric Research within the Commonwealth Scientific and Industrial Research Organisation provides analyses of ocean heat content and thermosteric sea level, while the NOAA Pacific Marine Environmental Laboratory makes available additional ocean heat content data. NCEP Reanalysis sea-level pressure is provided by the Physical Science Division of the NOAA Earth Science Research Laboratory. Model output analyzed in this study is available from the Earth System Grid Federation.

References

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

Allan, R. P., B. J. Soden, V. O. John, W. Ingram, and P. Good (2010), Current changes in tropical precipitation, *Env. Res. Lett.*, *5*, 025205, doi:10.1088/1748–9326/5/2/025205.

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

Bauer, S. E., and D. Koch (2005), Impact of heterogeneous reactions between sulfate and mineral dust aerosols on climate in the GISS GCM, *J. Geophys. Res.*, *110*, D17202, doi:10.1029/2005JD005870.

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

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

Bell, N., D. Koch, and D. T. Shindell (2005), Impacts of chemistry-aerosol coupling on tropospheric ozone and sulfate simulations in a general circulation model, *J. Geophys. Res.*, *110*, D14305, doi:10.1029/2004JD005538.

Berger, A. L. (1978), Long-term variations of daily insolation and quaternary climatic changes, *J. Atmos. Sci.*, *35*, 2362–2367.

Chahine, M. T., L. Chen, P. Dimotakis, X. Jiang, Q. Li, E. T. Olsen, T. Pagano, J. Randerson, and Y. L. Yung (2008), Satellite remote sounding of mid-tropospheric CO2, *Geophys. Res. Lett.*, *35*, L17807, doi:10.1029/2008GL035022.

Chou, C., and J. D. Neelin (2004), Mechanisms of global warming impacts on regional tropical precipitation, *J. Clim.*, *17*, 2688–2701, doi:10.1175/1520-0442(2004)017¡2688:MOGWIO¿2.0.CO;2.

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

Church, J. A., N. J. White, L. F. Konikow, C. M. Domingues, J. G. Cogley, E. Rignot, J. M. Gregory, M. R. van den Broeke, A. J. Monaghan, and I. Velicogna (2011), Revisiting the Earth's sea-level and energy budgets from 1961 to 2008, *Geophys. Res. Lett.*, *38*, L18601, doi:10.1029/2011GL048794.

Cleveland, W. S., and S. J. Devlin (1988), Locally weighted regression: An approach to regression analysis by local fitting, *J. Am. Stat. Assoc.*, *83*, 596–610, doi:10.2307/2289282.
Collins, W. D., et al. (2006), Radiative forcing by well-mixed greenhouse gases: Estimates from climate models in the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4), *J. Geophys. Res.*, *111*, D14317, doi:10.1029/2005JD006713.
Domingues, C., J. Church, N. White, P. Gleckler, S. Wijffels, P. Barker, and J. Dunn (2008), Improved estimates of upper-ocean warming and multi-decadal sea-level rise, *Nature*, *453*, 1090–1094, doi:10.1038/nature07080.
Ekman, A. M. L. (2014), Do sophisticated parameterizations of aerosol-cloud interactions in CMIP5 models improve the representation of recent observed temperature trends?, *J. Geophys. Res. Atmos*, *119*, doi:10.1002/2013JD020511, in press.
Etheridge, D., L. Steele, R. Langenfelds, R. Francey, J. Barnola, and V. Morgan (1996), Natural and anthropogenic changes in atmospheric CO_2 over the last 1000 years from air in Antarctic ice and firn, *J. Geophys. Res.*, *101*, 4115–4128.
Fetterer, F., K. Knowles, W. Meier, and M. Savoie (2011), *Sea Ice Index* [Digital media], Natl. Snow and Ice Data Cent., Boulder, Colo.
Fleming, E. L., C. H. Jackman, R. S. Stolarski, and D. B. Considine (1999), Simulation of stratospheric tracers using an improved empirically based two-dimensional model transport formulation, *J. Geophys. Res.*, *104*, 23,911–23,934, doi:10.1029/1999JD900332.
Forster, P., et al. (2007), Changes in atmospheric constituents and in radiative forcing, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., chap. 2, 129–234, Cambridge Univ. Press, Cambridge, U. K.
Forster, P. M., T. Andrews, P. Good, J. M. Gregory, L. S. Jackson, and M. Zelinka (2013), Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models, *J. Geophys. Res.*, *118*, 1139–1150, doi:10.1002/jgrd.50174.
Forster, P. M. d. F., and K. E. Taylor (2006), Climate forcings and climate sensitivities diagnosed from coupled climate model integrations, *J. Clim.*, *19*, 6181–6194, doi:10.1175/JCLI3974.1.
Fu, Q., C. M. Johanson, S. G. Warren, and D. J. Seidel (2004), Contribution of stratospheric cooling to satellite-inferred tropospheric temperature trends, *Nature*, *429*, 55–58, doi:10.1038/nature02524.
Ginoux, P., J. M. Prospero, T. E. Gill, N. Hsu, and M. Zhao (2012), Global scale attribution of anthropogenic and natural dust sources and their emission rates based on MODIS Deep Blue aerosol products, *Rev. Geophys.*, *50*, RG3005, doi:10.1029/2012RG000388.
Gregory, J. M., W. J. Ingram, M. A. Palmer, G. S. Jones, P. A. Stott, R. B. Thorpe, J. A. Lowe, T. C. Johns, and K. D. Williams (2004), A new method for diagnosing radiative forcing and climate sensitivity, *Geophys. Res. Lett.*, *31*, L03205, doi:10.1029/2003GL018747.
Hall, T. J., and M. J. Prather (1995), Seasonal evolutions of N_2O, O_3, and CO_2: Three-dimensional simulations of stratospheric correlations, *J. Geophys. Res.*, *100*, 16,699–16,720, doi:10.1029/94JD03300.
Hallberg, R., A. Adcroft, J. P. Dunne, J. P. Krasting, and R. J. Stouffer (2012), Sensitivity of 21st century global-mean steric sea level rise to ocean model formulation, *J. Clim.*, *26*, 2947–2956, doi:10.1175/JCLI-D-12–00506.1.
Hansen, J., D. Johnson, A. Lacis, S. Lebedeff, P. Lee, D. Rind, and G. Russell (1981), Climate impact of increasing atmospheric carbon dioxide, *Science*, *213*, 957–966, doi:10.1126/science.213.4511.957.
Hansen, J., M. Sato, and R. Ruedy (1997), Radiative forcing and climate response, *J. Geophys. Res.*, *102*, 6831–6864.
Hansen, J., et al. (2005), Efficacy of climate forcings, *J. Geophys. Res.*, *110*, D18104, doi:10.1029/2005JD005776.
Hansen, J., M. Sato, R. Ruedy, K. Lo, D. W. Lea, and M. Medina-Elizade (2006), Global temperature change, *Proc. Natl. Acad. Sci. U. S. A.*, *103*, 14,288–14,293, doi:10.1073/pnas.0606291103.
Hansen, J., et al. (2007), Climate simulations for 1880–2003 with GISS modelE, *Clim. Dyn.*, *29*, 661–696, doi:10.1007/s00382-007-0255-8.
Hansen, J. E., and M. Sato (2004), Greenhouse gas growth rates, *Proc. Natl. Acad. Sci. U. S. A.*, *101*, 16,109–16,114, doi:10.1073/pnas.0406982101.
Hansen, J. E., M. Sato, R. Ruedy, A. Lacis, and J. Glascoe (1998), Global climate data and models: A reconciliation, *Science*, *281*, 930–932, doi:10.1126/science.281.5379.930.
Harder, J. W., J. M. Fontenla, P. Pilewskie, E. C. Richard, and T. N. Woods (2009), Trends in solar spectral irradiance variability in the visible and infrared, *Geophys. Res. Lett.*, *36*, L07801, doi:10.1029/2008GL036797.
Hegerl, G., F. W. Zwiers, P. Braconnot, N. Gillett, Y. Luo, J. M. Orsini, N. Nicholls, J. Penner, and P. Stott (2007), Understanding and attributing climate change, in *Climate Change 2007: The Physical Science Basis*, edited by S. Solomon et al., chap. 9, 663–746, Cambridge Univ. Press, Cambridge, U. K.
Held, I. M., and K. M. Shell (2012), Using relative humidity as a state variable in climate feedback analysis, *J. Clim.*, *25*, 2578–2582, doi:10.1175/JCLI-D-11–00721.1.
Held, I. M., and B. J. Soden (2006), Robust responses of the hydrological cycle to global warming, *J. Clim.*, *19*, 568–5699, doi:10.1175/JCLI3990.1.
Holland, M. M., and C. M. Bitz (2003), Polar amplification of climate change in coupled models, *Clim. Dyn.*, *21*, 221–232, doi:10.1007/s00382-003-0332-6.
Jones, P. D., D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice (2012), Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010, *J. Geophys. Res.*, *117*, D05127, doi:10.1029/2011JD017139.
Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*, 437–471.
Kay, J. E., M. M. Holland, C. M. Bitz, E. Blanchard-Wrigglesworth, A. Gettelman, A. Conley, and D. Bailey (2012), The influence of local feedbacks and northward heat transport on the equilibrium arctic climate response to increased greenhouse gas forcing, *J. Clim.*, *25*, 5433–5450, doi:10.1175/JCLI-D-11–00622.1.
Kiehl, J. T. (2007), Twentieth century climate model response and climate sensitivity, *Geophys. Res. Lett.*, *34*, L22710, doi:10.1029/2007GL031383.
Klein Goldewijk, K., and G. van Drecht (2006), HYDE3: Current and historical population and land cover, in *Integrated Modelling of Global Environmental Change, An Overview of IMAGE 2.4*, edited by A. F. Bouwman, T. Kram, and K. K. Goldewijk, chap. 6, 93–112, Neth. Environ. Assess. Agency (MNP), Bilthoven, Netherlands.
Koch, D., G. A. Schmidt, and C. V. Field (2006), Sulfur, sea salt, and radionuclide aerosols in GISS ModelE, *J. Geophys. Res.*, *111*, D06206, doi:10.1029/2004JD005550.
Koch, D., T. C. Bond, D. Streets, N. Unger, and G. R. van der Werf (2007), Global impacts of aerosols from particular source regions and sectors, *J. Geophys. Res.*, *112*, D02205, doi:10.1029/2005JD007024.
Koch, D., S. Menon, A. D. Genio, R. Ruedy, I. Aleinov, and G. Schmidt (2009), Distinguishing aerosol impacts on climate over the past century, *J. Clim.*, *22*, 2659–2677, doi:10.1175/2008JCLI2573.1.
Koch, D., et al. (2011), Coupled aerosol-chemistry-climate twentieth century transient model investigation: Trends in short-lived species and climate responses, *J. Clim.*, *24*, 2693–2714, doi:10.1175/2011JCLI3582.1.

Lamarque, J.-F., et al. (2010), Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: Methodology and application, *Atmos. Chem. Phys.*, *10*, 7017–7039, doi:10.5194/acp-10–7017-2010.
Lamarque, J. F., K. G. Page, M. Meinshausen, K. Riahi, S. S., D. P. van Vuuren, C. A. J., and F. Vitt (2011), Global and regional evolution of short-lived radiatively-active gases and aerosols in the Representative Concentration Pathways, *Clim. Change*, *109*, 191–212, doi: 10.1007/s10584-011-0155-0.
Lambert, F. H., and M. J. Webb (2008), Dependency of global mean precipitation on surface temperature, *Geophys. Res. Lett.*, *35*, L16706, doi:10.1029/2008GL034838.
Lean, J. (2000), Evolution of the Sun's spectral irradiance since the Maunder Minimum, *Geophys. Res. Lett.*, *27*, 2425–2428, doi:10.1029/2000GL000043.
Lean, J. (2009), Calculations of solar irradiance: Monthly means from 1882 to 2008, annual means from 1610 to 2008. [Available at http://www.geo.fu-berlin.de/en/met/ag/strat/forschung/SOLARIS/Input_data/Calculations_of_Solar_Irradiance.pdf.]
Lean, J., and M. T. Deland (2012), How does the sun's spectrum vary?, *J. Clim.*, *25*, 2555–2560, doi:10.1175/JCLI-D-11–00571.1.
LeTexier, H., S. Solomon, and R. Garcia (1988), The role of molecular-hydrogen and methane oxidation in the water-vapor budget of the stratosphere, *Q. J. R. Meteorol. Soc.*, *114*, 281–295.
Levitus, S., and T. P. Boyer (1994), World ocean atlas 1994, vol. 4: Temperature, NOAA Atlas NESDIS 4, pp. 177, US Govt Printing Office, Washington D. C.
Levitus, S., R. Burgett, and T. P. Boyer (1994), World ocean atlas 1994, vol. 3: Salinity, NOAA Atlas NESDIS 3, pp. 99, US Govt Printing Office, Washington D. C.
Levitus, S., J. Antonov, and T. T. Boyer (2005), Warming of the world ocean, 1955–2003, *Geophys. Res. Lett.*, *32*, L02604, doi:10.1029/2004GL021592.
Lyman, J. M., S. A. Good, V. V. Gouretski, M. Ishii, G. C. Johnson, M. D. Palmer, D. M. Smith, and J. K. Willis (2010), Robust warming of the global upper ocean, *Nature*, *465*, 334–337, doi:10.1038/nature09043.
Mahowald, N. M., et al. (2010), Observed 20th century desert dust variability: Impact on climate and biogeochemistry, *Atmos. Chem. Phys.*, *10*, 10,875–10,893, doi:10.5194/acp-10–10875-2010.
Manabe, S., and R. T. Weatherald (1967), Thermal equilibrium of the atmosphere with a given distribution of relative humidity, *J. Atmos. Sci.*, *24*, 241–259.
Mears, C. A., and F. J. Wentz (2009), Construction of the Remote Sensing Systems v3.2 atmospheric temperature records from the MSU and AMSU microwave sounders, *J. Atmos. Oceanic Technol.*, *26*, 1040–1056, doi:10.1175/2008JTECHA1176.1.
Meehl, G. A., J. M. Arblaster, J. T. Fasullo, A. Hu, and K. E. Trenberth (2011), Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods, *Nat. Clim. Change*, *1*, 360–364, doi:10.1038/nclimate1229.
Menon, S., D. Koch, G. Beig, S. Sahu, J. Fasullo, and D. Orlikowski (2010), Black carbon aerosols and the third polar ice cap, *Atmos. Chem. Phys.*, *10*, 4559–4571, doi:10.5194/acp-10–4559-2010.
Miller, R. L., et al. (2006a), Mineral dust aerosols in the NASA Goddard Institute for Space Studies ModelE AGCM, *J. Geophys. Res.*, *111*, D06208, doi:10.1029/2005JD005796.
Miller, R. L., G. A. Schmidt, and D. T. Shindell (2006b), Forced annular variations in the 20th century Intergovernmental Panel Climate Change Fourth Assessment Report models, *J. Geophys. Res.*, *111*, D18101, doi:10.1029/2005JD006323.
Minschwaner, K., R. W. Carver, B. P. Briegleb, and A. E. Roche (1998), Infrared radiative forcing and atmospheric lifetimes of trace species based on observations from UARS, *J. Geophys. Res.*, *103*, 23,243–23,253, doi:10.1029/98JD02116.
Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones (2012), Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, *J. Geophys. Res.*, *117*, D08101, doi:10.1029/2011JD017187.
Morrison, H., and A. Gettelman (2008), A new two-moment bulk stratiform cloud microphysics scheme in the Community Atmosphere Model, Version 3 (CAM3). Part I: Description and numerical tests, *J. Clim.*, *21*, 3642–3659, doi:10.1175/2008JCLI2105.1.
Myhre, G., et al. (2013), Radiative forcing of the direct aerosol effect from AeroCom Phase II simulations, *Atmos. Chem. Phys.*, *13*(4), 1853–1877, doi:10.5194/acp-13–1853-2013.
Myhre, G., et al. (2014), Anthropogenic and natural radiative forcing, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker et al., chap. 8, 659–740, Cambridge Univ. Press, Cambridge, U. K.
Nerem, R. S., D. Chambers, C. Choe, and G. T. Mitchum (2010), Estimating mean sea level change from the TOPEX and Jason altimeter missions, *Mar. Geod.*, *33*, 435, doi:10.1080/01490419.2010.491031.
Otto, A., et al. (2013), Energy budget constraints on climate response, *Nat. Geosci.*, *6*, 415–416, doi:10.1038/ngeo1836.
Pendergrass, A. G., and D. L. Hartmann (2014), The atmospheric energy constraint on global-mean precipitation change, *J. Clim.*, *27*, 757–768, doi:10.1175/JCLI-D-13–00163.1.
Perlwitz, J., and R. Miller (2010), Cloud cover increase with increasing aerosol absorptivity—a counterexample to the conventional semi-direct aerosol effect, *J. Geophys. Res.*, *115*, D08203, doi:10.1029/2009JD012637.
Pongratz, J., C. Reick, T. Raddatz, and M. Claussen (2008), A reconstruction of global agricultural areas and land cover for the last millennium, *Global Biogeochem. Cycles*, *22*, GB3018, doi:10.1029/2007GB003153.
Pyle, J., T. Shepherd, G. Bodeker, P. Canziani, M. Dameris, P. Forster, A. Gruzdev, R. Muller, N. J. Muthama, G. Pitari, and W. Randel (2005), Ozone and climate: A review of interconnections, in *Safeguarding the Ozone Layer and the Global Climate System: Special Report of the Intergovernmental Panel on Climate Change*, edited by S. Andersen, L. Kuijpers, J. Pons, S. Solomon, O. Davidson, and B. Metz, chap. 1, pp. 83–132, Cambridge Univercity Press, Cambridge, U. K.
Ramaswamy, V., et al. (2001), Radiative forcing of climate change, in *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, edited by J. Houghton et al., chap. 6, 881 pp., Cambridge Univ. Press, Cambridge, U. K.
Randel, W., and F. Wu (1999), A stratospheric ozone trends data set for global modeling studies, *Geophys. Res. Lett.*, *26*, 3089–3092.
Rohde, R., R. A. Muller, R. Jacobsen, E. Muller, S. Perlmutter, A. Rosenfeld, J. Wurtele, D. Groom, and C. Wickham (2013), A new estimate of the average earth surface land temperature spanning 1753 to 2011, *Geoinfor. Geostat.*, *1*, doi:10.4172/gigs.1000101.
Romanou, A., W. Gregg, J. Romanski, M. Kelley, R. Bleck, R. Healy, L. Nazarenko, G. Russell, G. Schmidt, S. Sun, and N. Tausnev (2013), Preindustrial control simulations of the ocean-atmosphere CO_2 exchanges in the NASA-GISS climate model: Sensitivity to physical ocean model, *Ocean Model.*, pp. *26–44*, doi:10.1016/j.ocemod.2013.01.008.
Ropelewski, C. F., and M. S. Halpert (1996), Quantifying Southern Oscillation-precipitation relationships, *J. Clim.*, *9*, 1043–1059.
Rothman, L., et al. (2009), The HITRAN 2008 molecular spectroscopic database, *J. Quant. Spectrosc. Radiat. Transfer*, *110*, 533–572, doi: 10.1016/j.jqsrt.2009.02.013.

Russell, G. L., J. R. Miller, and D. Rind (1995), A coupled atmosphere-ocean model for transient climate change, *Atmos.Ocean*, *33*(4), 683–730.
Santer, B., J. Hnilo, T. Wigley, J. Boyle, C. Doutriaux, M. Fiorino, D. Parker, and K. Taylor (1999), Uncertainties in observationally based estimates of temperature change in the free atmosphere, *J. Geophys. Res.*, *104*(D6), 6305–6333, doi:10.1029/1998JD200096.
Sato, M., J. Hansen, M. McCormick, and J. Pollack (1993), Stratospheric aerosol optical depth, 1850–1990, *J. Geophys. Res.*, *98*, 22,987–22,994.
Schmidt, G. A., et al. (2006), Present day atmospheric simulations using GISS modelE: Comparison to in-situ, satellite and reanalysis data, *J. Clim.*, *19*, 153–192.
Schmidt, G. A., et al. (2011), Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0), *Geosci. Model Dev.*, *4*, 33–45, doi:10.5194/gmd-4–33-2011.
Schmidt, G. A., et al. (2012), Climate forcing reconstructions for use in PMIP simulations of the Last Millennium (v1.1), *Geosci. Model Dev.*, *5*, 185–191, doi:10.5194/gmd-5–185-2012.
Schmidt, G. A., et al. (2014a), Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, *J. Adv. Model. Earth Syst.*, doi:10.1002/2013MS000265, in press. [Available at http://pubs.giss.nasa.gov/authors/gschmidt.html.].
Schmidt, G. A., D. T. Shindell, and K. Tsigaridis (2014b), Reconciling warming trends, *Nat. Geosci.*, *7*, 158–160, doi:10.1038/ngeo2105.
Seager, R., N. Harnik, W. A. Robinson, Y. Kushnir, M. Ting, H.-P. Huang, and J. Velez (2005), Mechanisms of ENSO-forcing of hemispherically symmetric precipitation variability, *Q. J. R. Meteorol. Soc.*, *131*, 1501–1527, doi:10.1256/qj.04.96.
Seinfeld, J. H., and S. N. Pandis (2006), *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, 2nd ed., 1232 pp., Wiley-Interscience, Hoboken, N. J.
Sherwood, S. C., S. Bony, and J.-L. Dufresne (2014), Spread in model climate sensitivity traced to atmospheric convective mixing, *Nature*, *505*, 37–42, doi:10.1038/nature12829.
Shindell, D., G. Faluvegi, N. Unger, E. Aguilar, G. Schmidt, D. Koch, S. Bauer, and R. Miller (2006a), Simulations of preindustrial, present-day, and 2100 conditions in the NASA GISS composition and climate model G-PUCCINI, *Atmos. Chem. Phys.*, *6*, 4427–4459, doi:10.5194/acp-6–4427-2006.
Shindell, D. T., and G. A. Schmidt (2004), Southern Hemisphere climate response to ozone changes and greenhouse gas increases, *Geophys. Res. Lett.*, *31*, L18209, doi:10.1029/2004GL020724.
Shindell, D. T., G. Faluvegi, and N. Bell (2003), Preindustrial-to-present-day radiative forcing by tropospheric ozone from improved simulations with the GISS chemistry-climate GCM, *Atmos. Chem. Phys.*, *3*, 1675–1702, doi:10.5194/acp-3–1675-2003.
Shindell, D. T., G. Faluvegi, R. L. Miller, G. A. Schmidt, J. E. Hansen, and S. Sun (2006b), Solar and anthropogenic forcing of tropical hydrology, *Geophys. Res. Lett.*, *33*, L24706, doi:10.1029/2006GL027468.
Shindell, D. T., et al. (2013a), Radiative forcing in the ACCMIP historical and future climate simulations, *Atmos. Chem. Phys.*, *13*, 2939–2974, doi:10.5194/acp-13–2939-2013.
Shindell, D. T., et al. (2013b), Interactive ozone and methane chemistry in GISS-E2 historical and future climate simulations, *Atmos. Chem. Phys.*, *13*, 2653–2689, doi:10.5194/acp-13–2653-2013.
Shine, K., et al. (2003), A comparison of model-simulated trends in stratospheric temperatures, *Q. J. R. Meteorol. Soc.*, *129*, 1565–1588, doi:10.1256/qj.02.186.
Soden, B. J., and I. M. Held (2006), An assessment of climate feedbacks in coupled ocean-atmosphere models, *J. Clim.*, *19*, 3354–3360, doi:10.1175/JCLI3799.1.
Solomon, S., K. H. Rosenlof, R. W. Portmann, J. S. Daniel, S. M. Davis, T. J. Sanford, and G.-K. Plattner (2010), Contributions of stratospheric water vapor to decadal changes in the rate of global warming, *Science*, *327*, 1219–1223, doi:10.1126/science.1182488.
Stephens, G. L., J. Li, M. Wild, C. A. Clayson, N. Loeb, S. Kato, T. L'Ecuyer, P. W. Stackhouse Jr., M. Lebsock, and T. Andrews (2012), An update on earth's energy balance in light of the latest global observations, *Nat. Geosci.*, *5*, 691–696, doi:10.1038/ngeo1580.
Stroeve, J. C., V. Kattsov, A. Barrett, M. Serreze, T. Pavlova, M. Holland, and W. N. Meier (2012), Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations, *Geophys. Res. Lett.*, *39*, L16502, doi:10.1029/2012GL052676.
Sun, S., and R. Bleck (2006), Multi-century simulations with the coupled GISS-HYCOM climate model: Control experiments, *Clim. Dyn.*, *26*, 407–428, doi:10.1007/s00382-005-0091-7.
Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of CMIP5 and the experiment design, *Bull. Am. Meteorol. Soc.*, *93*, 485–498, doi:10.1175/BAMS-D-11–00094.1.
Teetor, P. (2011), *R Cookbook*, O'Reilly Media, 438 pp, Sebastopol CA, USA.
Thompson, D. W. J., and J. M. Wallace (2000), Annular modes in the extratropical circulation. Part I: Month-to-month variability, *J. Clim.*, *13*, 1000–1016.
Thorne, P. W., J. R. Lanzante, T. C. Peterson, D. J. Seidel, and K. P. Shine (2011), Tropospheric temperature trends: History of an ongoing controversy, *WIREs Clim. Change*, *2*, 66–88, doi:10.1002/wcc.80.
Trenberth, K., et al. (2007), Observations: Surface and atmospheric climate change, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., chap. 3, 235–336, Cambridge Univ. Press, Cambridge, U. K.
Trenberth, K. E., and A. Dai (2007), Effects of Mount Pinatubo volcanic eruption on the hydrological cycle as an analog of geoengineering, *Geophys. Res. Lett.*, *34*, L15702, doi:10.1029/2007GL030524.
Twomey, S. (1977), The influence of pollution on the shortwave albedo of clouds, *J. Atmos. Sci.*, *34*, 1149–1152.
van Aardenne, J. A., F. J. Dentener, J. G. J. Olivier, C. G. M. K. Goldewijk, and J. Lelieveld (2001), A $1^\circ \times 1^\circ$ resolution data set of historical anthropogenic trace gas emissions for the period 1890–1990, *Global Biogeochem. Cycles*, *15*, 909–928, doi:10.1029/2000GB001265.
van Vuuren, D., et al. (2011), The representative concentration pathways: An overview, *Clim. Change*, *109*, 5–31, doi:10.1007/s10584-011-0148-z.
W. M. O. (1957), Definition of the tropopause, *World Meteorol. Organ. Bull.*, *6*, 136.
Wang, Y.-M., J. L. Lean, and N. R. Sheeley, Jr. (2005), Modeling the Sun's Magnetic Field and Irradiance since 1713, *Astrophys. J.*, *625*, 522–538, doi:10.1086/429689.
Yulaeva, E., and J. M. Wallace (1994), The signature of ENSO in global temperature and precipitation fields derived from the microwave sounding unit, *J. Clim.*, *7*, 1719–1736, doi:10.1175/1520-0442(1994)007¡1719:TSOEIG¿2.0.CO;2.