Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive

Gavin A. Schmidt[1], Max Kelley[1,2], Larissa Nazarenko[1,3], Reto Ruedy[1,2], Gary L. Russell[1], Igor Aleinov[1,3], Mike Bauer[1,4], Susanne E. Bauer[1,3], Maharaj K. Bhat[5], Rainer Bleck[1,4], Vittorio Canuto[1], Yong-Hua Chen[1,2], Ye Cheng[1,3], Thomas L. Clune[5], Anthony Del Genio[1], Rosalinda de Fainchtein[5], Greg Faluvegi[1,3], James E. Hansen[1,6], Richard J. Healy[1,3], Nancy Y. Kiang[1], Dorothy Koch[3,7], Andy A. Lacis[1], Allegra N. LeGrande[1], Jean Lerner[1,3], Ken K. Lo[1,2], Elaine E. Matthews[1], Surabi Menon[8], Ron L. Miller[1], Valdar Oinas[1,2], Amidu O. Oloso[5], Jan P. Perlwitz[1,4], Michael J. Puma[1,3], William M. Putman[5], David Rind[1], Anastasia Romanou[1,4], Makiko Sato[1,2,6], Drew T. Shindell[1], Shan Sun[1,2], Rahman A. Syed[5], Nick Tausnev[1,2], Kostas Tsigaridis[1,3], Nadine Unger[9], Apostolos Voulgarakis[1,3,10], Mao-Sung Yao[1,2], and Jinlun Zhang[11]

[1]NASA Goddard Institute for Space Studies and Center for Climate Systems Research, Columbia University, New York, New York, USA, [2]Trinnovim LLC, New York, New York, USA, [3]Center for Climate Systems Research, Columbia University, New York, New York, USA, [4]Department of Applied Physics and Applied Mathematics, Columbia University, 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 Office of Biological and Environmental Research, Department of Energy, Washington, D. C., Washington, USA, [8]Lawrence Berkeley National Laboratory, Berkeley, California, USA, [9]Yale University, New Haven, Connecticut, USA, [10]Now at Deptartment of Physics, Imperial College, London, UK, [11]University of Washington, Seattle, Washington, USA

Abstract We present a description of the ModelE2 version of the Goddard Institute for Space Studies (GISS) General Circulation Model (GCM) and the configurations used in the simulations performed for the Coupled Model Intercomparison Project Phase 5 (CMIP5). We use six variations related to the treatment of the atmospheric composition, the calculation of aerosol indirect effects, and ocean model component. Specifically, we test the difference between atmospheric models that have noninteractive composition, where radiatively important aerosols and ozone are prescribed from precomputed decadal averages, and interactive versions where atmospheric chemistry and aerosols are calculated given decadally varying emissions. The impact of the first aerosol indirect effect on clouds is either specified using a simple tuning, or parameterized using a cloud microphysics scheme. We also use two dynamic ocean components: the Russell and HYbrid Coordinate Ocean Model (HYCOM) which differ significantly in their basic formulations and grid. Results are presented for the climatological means over the satellite era (1980–2004) taken from transient simulations starting from the preindustrial (1850) driven by estimates of appropriate forcings over the 20th Century. Differences in base climate and variability related to the choice of ocean model are large, indicating an important structural uncertainty. The impact of interactive atmospheric composition on the climatology is relatively small except in regions such as the lower stratosphere, where ozone plays an important role, and the tropics, where aerosol changes affect the hydrological cycle and cloud cover. While key improvements over previous versions of the model are evident, these are not uniform across all metrics.

1. Introduction

Progress in general circulation modeling is continuous and incremental. However, it is important and necessary to take stock and assess progress, in scope as well as in results, every few years. The contributions of ModelE2 to the Coupled Model Intercomparison Project (Phase 5; CMIP5) database provide a significant milestone at which to do so.

The CMIP5 database is a coordinated set of numerical climate simulations with contributions from climate modeling institutes from around the world [*Taylor et al.*, 2009, 2012]. It builds from previous CMIP efforts, most notably CMIP3, and is intended (in part) to provide a basis of climate model simulations and studies

that can be assessed by the Intergovernmental Panel on Climate Change (IPCC). Analyses of the CMIP3 models played a large role in the Fourth Assessment Report (AR4) [*Solomon et al.*, 2007] and the preliminary analysis of CMIP5 was important for the Fifth Assessment Report (AR5), though the contributions to the database and the analysis will continue and expand in future.

The GISS climate model development process has been documented in papers stretching over 30 years [see *Schmidt et al.*, 2006, hereafter SEA06; *Hansen et al.*, 2000; *Hansen et al.*, 1983, and references therein]. This paper will describe changes to the basic model and all of its components that have occurred between the configurations used in the CMIP3 experiments (performed in 2004), and the CMIP5 experiments (performed in 2011/2012). These model versions are referred to within the CMIP nomenclature as GISS-ER, GISS-EH2 (CMIP3) and GISS-E2-R and GISS-E2-H (CMIP5; where R and H denote different ocean models, see section 2.10). For clarity, we will use E2-R/H and E-R/H2 in subsequent text.

We have used the opportunity provided by CMIP5 to update and expand our comprehensive approach to exploring model forcings and feedbacks (for instance, following on from *Hansen et al.* [1997, 2005, 2007]). The key issues are assessing the impacts of multiple forcings (together and singly) are their dependence on the modeled physics and the quantifying of fingerprints of climate change associated with different forcings, taking into account structural and other uncertainties.

The GISS CMIP5 contribution includes a greater variety of submitted model versions than in CMIP3 and it is worth expanding on the reasons behind this. Model simulations in CMIP3 were relatively uniform in their scope, that is, they included representations of the atmosphere, ocean, sea ice, land surface, etc., but did not include interactive chemistry, aerosols, a carbon cycle, dynamic vegetation, or dynamic ice sheets. At the time, these components were in early stages of development and were not generally mature enough to be part of the standard model intercomparisons, though early publications using these components had appeared. With CMIP5, however, that situation has changed and many of the modeling groups (including GISS) are using models that include many more of these "Earth System" components. Different groups have implemented different variations of those elements though no single-model simulation includes them all. Comparisons between models will therefore be more complicated since models will not have the same consistency of scope that existed in CMIP3, although analyses of the genealogy of the multimodel ensemble shed some light on this [*Knutti et al.*, 2013]. It is therefore incumbent on each of the groups to be explicit in what is included in their models so that appropriate comparisons can be made within and across groups since many groups have used similar structural variations [e.g., *Hallberg et al.*, 2013]. We have chosen to show the impact of the increasing levels of interactivity within CMIP5 to help understand the impact of each additional component (see Table 1 for the configurations).

Comparisons across the GISS models specifically give insight into the impact of more interactivity in atmospheric composition and different ocean models while also providing a traceable link to the behavior of the older CMIP3-like configurations. The GISS group has chosen to focus on the importance of interactive atmospheric composition because of its importance in attributing emissions to climate change [*Shindell et al.*, 2009], in responses to solar variability [*Shindell et al.*, 2006a], and in the complex interactions between air pollution and climate change [*Shindell et al.*, 2012]. The structural uncertainty in ocean models, whose long-term variability and mixing properties are hard to constrain from observations, are an important factor that was already explored in CMIP3. To the extent that ocean models have not converged, we feel it is incumbent on groups to continue to explore which aspects of climate change are and are not sensitive to the ocean model formulation.

This paper focuses on the modern climatology of the atmospheric and coupled models, including selected aspects of their intrinsic variability. Companion papers will discuss historical simulations of climate change since 1850 (R. L. Miller et al., CMIP5 historical simulations (1850–2012) with GISS ModelE2, submitted to *Journal of Advances in Modeling Earth Systems*, 2014), future projections (L. Nazarenko et al., Future climate

Table 1. Model Configurations Mentioned in the Text for the CMIP5 and CMIP3 Submissions

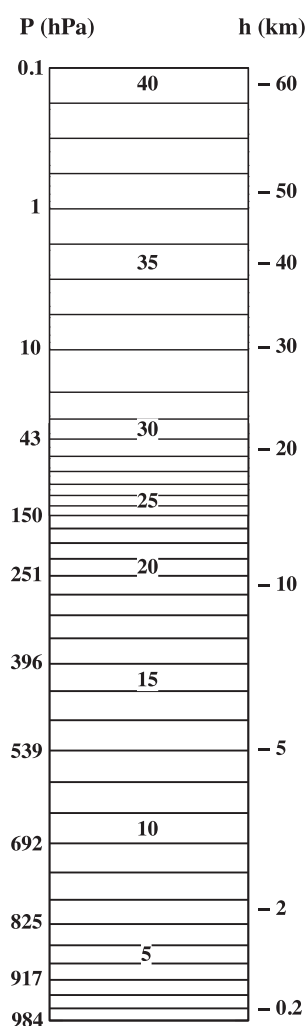| Model Configuration	Horizontal Resolution	Vertical Layers	Model Top (hPa)	Atmos. Composition Treatment
E2-R/H NINT	2° × 2.5°	40	0.1	Noninteractive, tuned AIE
E2-R/H TCAD	2° × 2.5°	40	0.1	Interactive chem/aerosol, tuned AIE
E2-R/H TCADI	2° × 2.5°	40	0.1	Interactive chem/aerosol and parameterized AIE
E-R/H2 (CMIP3)	4° × 5°	20	0.1	Noninteractive, tuned AIE

change under Representative Concentration Pathways (RCPs) simulations with GISS ModelE2, submitted to *Journal of Advances in Modeling Earth Systems*, 2014), performance of the chemistry simulation [*Shindell et al.*, 2013], and paleoclimate data/model comparisons. More model information, including access to online data, forcings, and descriptions are available at http://www.giss.nasa.gov. The ModelE2 source code (along with documentation) can be downloaded from http://www.giss.nasa.gov/tools/modelE. More complete diagnostics are available via the Earth System Grid Federation (ESGF) at http://esg.edu.

2. Model Physics

The model physics are predominantly based on the physics of the GISS ModelE (CMIP3 version) described in previous publications [*Schmidt et al.*, 2006, and references therein]. We will therefore only highlight the differences to the previously documented code here.

2.1. Configurations

The basic structure of the model is unchanged from CMIP3. The atmospheric model has a Cartesian grid point formulation for all quantities. The horizontal resolution is 2° × 2.5° latitude by longitude with coarser discretizations of 4° × 5° and 8° × 10° are available for historical and pedagogical reasons. The effective resolution for atmospheric tracer transport is greater than these nominal resolutions due to the nine higher-order moments that are carried along with the mean tracer values in each grid box [*Prather*, 1986]. The velocity points in the atmosphere are on the Arakawa-B grid and the vertical discretization follows a sigma coordinate to 150 hPa with constant pressure layers above. The standard vertical resolution in ModelE2 has 40 layers and a model top at 0.1 hPa (Figure 1), twice as many layers as in the CMIP3 versions. The dynamical core, atmospheric mixing, and boundary layer code are unchanged from SEA06. The model uses a 225 s time step for the dynamics and physics, are calculated every 30 min. The radiation code is called every five physics time steps (every 2.5 h).

The results described below are mainly from coupled ocean-atmosphere models with three different treatments of atmospheric composition and two different ocean models (Table 1). The atmospheric composition treatments (corresponding to physics-version = 1,2,3 in the CMIP5 archive) differ according to how atmospheric chemistry, aerosols, and their interactions with clouds are dealt with. The NonINTeractive ("NINT") version (physics-version = 1) is similar in conception to the CMIP3 version of the model (described in SEA06), with noninteractive (though decadally varying) fields of radiatively active components (ozone and multiple aerosol species) read in from previously calculated offline fields [*Koch et al.*, 2011; *Shindell et al.*, 2006a], and includes a tuned aerosol indirect effect (AIE) following *Hansen et al.* [2005]. The Tracers of Chemistry, Aerosols and their Direct effects ("TCAD") version (physics-version = 2) includes interactive chemistry, aerosols, and dust driven by emissions [*Koch et al.*,

Figure 1. Vertical layering for the 40 layer atmospheric model configuration (for graphical convenience, the vertical coordinate shown here is linear in pressure to 150 hPa, logarithmic above).

2011], with the same AIE as in NINT. The Tracers of Chemistry, Aerosols and their Direct and Indirect effects ("TCADI") version (physics-version = 3) have the same interactive components as TCAD, but the first indirect effect of aerosols is parameterized using calculated aerosol number concentrations (see details below) [*Menon et al.*, 2008, 2010]. We also include some results in section 6 from simulations driven by historical ocean temperatures, instead of a dynamic ocean that would calculate its own sea surface temperatures ("Atmospheric Model Intercomparison Project" (AMIP) runs).

Differences between the model versions are therefore a function of multiple factors: NINT and TCAD/TCADI have different aerosol distributions caused by different emission inventories and aerosol physics (TCAD and TCADI have updates to secondary organic aerosol chemistry and more absorbing black carbon that were not included in the *Koch et al.* [2011] simulations that provide the concentrations for the NINT runs) and changes to the base climate set by the ocean model version. The difference between TCAD and TCADI is only a function of the different AIE treatment, though this may have indirect impacts on composition as well. The details of the forcings including the treatments of the AIE are discussed in R. L. Miller et al. (submitted manuscript, 2013).

For the most appropriate comparison with recent climatological data, all the results described here are from the ensemble of transient 20th century simulations for the 25 year period 1980–2004. For each model configuration, there are five ensemble members, using initial conditions 20 years apart in the relevant preindustrial control run, giving a total of 30 individual simulations. For the CMIP3 (GISS-ER, GISS-EH2) models, we use the 1979–2003 average in the comparison.

While specific observational products used for comparisons may cover slightly different time periods, the differences between periods are much smaller than the differences between the observations and the model. The full historical simulations from which these model climatologies are drawn (R. L. Miller et al., submitted manuscript, 2013) include forcings from orbital variability, solar variability [*Wang et al.*, 2005b; *Lean*, 2009], volcanic aerosols (*Sato et al.* [1993], and updates), well-mixed greenhouse gas changes (including impacts on stratospheric water vapor from methane), stratospheric ozone depletion, tropospheric ozone changes, the direct and indirect sulfate, nitrate, black carbon, and organic aerosol changes (including the impact of black carbon on snow and ice albedo), and land use/land cover changes. The differences from the experimental design described in *Hansen et al.* [2007] are a reduction in the 20th century trend in solar forcing (by 55%) [*Wang et al.*, 2005b], updates to include more recent years, the inclusion of orbital forcing (which is a small effect on these time scales, but is included for consistency with longer simulations over the last millennium), and the switch toward simulations driven by emissions, rather than concentrations, of radiatively active atmospheric components (in TCAD and TCADI versions). Experiments including additional forcings (for instance, transient simulations of irrigation) will be described elsewhere (B. Cook et al, "Irrigation as an historical climate forcing", submitted to *Climate Dynamics*, 2014). Uncertainties in forcing fields in the 21st century do not significantly impact the climatological averages discussed in this paper.

2.2. Atmospheric Composition

Compared to the models described in SEA06, the interactive chemistry and aerosols (including dust) are new functionality (in TCAD and TCADI). Basic descriptions of the tracer code, chemical reactions, species, and model evaluation are given in *Shindell et al.* [2006b] and *Miller et al.* [2006a]. Updates are described in *Voulgarakis et al.* [2011], *Koch et al.* [2011], and *Shindell et al.* [2013]. Differences from these descriptions are highlighted below. All tracers follow the advective air mass fluxes in the dynamics and convection schemes, and there is an explicit dissolved tracer budget for all soluble species in atmospheric water/ice condensates. Tracers are advected with a flux limiter to avoid negative concentrations and with an adaptive time step to prevent numerical instability.

2.2.1. Gas-Phase Chemistry

Gas-phase chemistry in TCAD/TCADI is based upon the GISS model for Physical Understanding of Composition-Climate INteractions and Impacts (G-PUCCINI) documented in *Shindell et al.* [2006b]. The behavior of the chemical scheme in that model has been documented and extensively compared with observations, especially for the troposphere [e.g., *Dentener et al.*, 2006; *Shindell et al.*, 2006b, 2006c; *Stevenson et al.*, 2006]. Tropospheric chemistry includes basic NO_x-HO_x-O_x-CO-CH_4 chemistry as well as peroxyacyl nitrates and the hydrocarbons; terpenes, isoprene, alkyl nitrates, aldehydes, alkenes, and paraffins. The lumped hydrocarbon family scheme was derived from the Carbon Bond Mechanism-4 (CBM-4) and from the more extensive Regional Atmospheric Chemistry Model (RACM), following *Houweling et al.* [1998]. To represent stratospheric

chemistry, the model includes chlorine-containing and bromine-containing compounds and chlorofluorocarbons (CFC) and N_2O source gases. The chemistry used here is as documented previously, with a few additions: polar stratospheric cloud formation is now dependent upon the abundance of nitric acid, water vapor, and temperature [*Hanson and Mauersberger*, 1988], a reaction pathway for $HO_2 + NO$ to yield HNO_3 has been added [*Butkovskaya et al.*, 2007], and the heterogeneous hydrolysis of N_2O_5 on sulfate now follows *Kane et al.* [2001] and *Hallquist et al.* [2003]. Chemical calculations are performed seamlessly throughout the troposphere and stratosphere. The full scheme includes 156 chemical reactions among 51 species. We use a time step of 30 min. Photolysis rates are calculated using the Fast-J2 scheme [*Bian and Prather*, 2002], which takes into account the model distribution of clouds, aerosols, and ozone [*Bian et al.*, 2003], whereas other chemical reaction rate coefficients are from *Sander et al.* [2000]. Modeled aerosol optical depths are passed to Fast-J2 at every time step, while the tabulated optical properties required for different aerosol types (extinction efficiencies, single scattering albedos, scattering phase function expansion terms) have been calculated offline using input data consistent with the aerosol properties in the model's radiation code.

Relative to the experiments using the previous version of the model, modifications to the chemistry scheme and in particular improved stratospheric circulation have led to a substantially more realistic simulation of stratospheric ozone [*Shindell et al.*, 2013]. In particular, the seasonality of middle and high-latitude ozone, which was much too weak, is now fairly realistic. The current model's seasonal cycle of extratropical ozone columns is about 20% too large in the NH, consistent with a stratospheric Brewer-Dobson circulation that is stronger than observed [*Shindell et al.*, 2013]. The present-day (average of 2006–2010) model value for the chemical lifetime of methane in the troposphere is 8.8–9.3 years, in good agreement with constraints derived from methyl chloroform observations, which yield a tropospheric chemical lifetime of 9.6 ± 1.4 years [*Prather et al.*, 2001].

2.2.2. Interactive Aerosols
Aerosol species such as sulfate [*Koch et al.*, 2006, 2007], nitrate [*Bauer et al.*, 2007a], elemental and organic carbon, sea salt [*Koch et al.*, 2006], and dust [*Miller et al.*, 2006a] are interactively calculated and change the climate through direct [*Koch et al.*, 2006] and indirect effects [*Menon et al.*, 2008, 2010], and gas-phase chemistry by affecting photolysis rates [*Bian et al.*, 2003]. Sea salt, dust, and isoprene surface fluxes are calculated interactively while natural and anthropogenic decadal fluxes of other aerosols are provided by the CMIP5 emissions inventory [*Lamarque et al.*, 2010]. The model also includes heterogeneous chemistry on dust surfaces [*Bauer and Koch*, 2005; *Bauer et al.*, 2007b] and NO_x-dependent secondary organic aerosol production from isoprene and terpenes (K. Tsigaridis et al., manuscript in preparation), as described by *Tsigaridis and Kanakidou* [2007, and references therein].

There is a choice of aerosol modules in the model. The one used for these runs is the mass-based scheme, where aerosols are treated as externally mixed and have prescribed size and properties [e.g., *Koch et al.*, 2006, 2007], with the exception of sea salt that has two distinct size classes [*Tsigaridis et al.*, 2013], and dust that has four size classes [*Miller et al.*, 2006b] and can be coated by sulfate and nitrate aerosols [*Bauer and Koch*, 2005]. Results from other aerosol microphysics modules (MATRIX [*Bauer et al.*, 2008] and TOMAS [*Lee and Adams*, 2012]) are described in *Bauer et al.* [2013].

The dust aerosol distribution is calculated as described by *Miller et al.* [2006a]. The global emission of clay particles (with radii of less than 1 μm) and larger silt particles is chosen to optimize model agreement with a worldwide array of observations [*Cakmur et al.*, 2006]. Emissions depend upon the surface wind speed distribution calculated by the planetary boundary layer scheme every physics time step, also accounting for wind gusts [*Cakmur et al.*, 2004]. Solar absorption by dust particles is increased by roughly one half compared to the CMIP3 model *Miller et al.* [2006a].

2.2.3. Gas-Aerosol Interactions
In TCAD and TCADI model versions, gas and aerosol tracers are coupled in three direct ways: (a) gases affecting aerosol chemical production, (b) aerosols providing surfaces for the processing of gases in the atmosphere, and (c) aerosols affecting the photolysis of gases. Modeled oxidants, such as the hydroxyl radical (OH) and hydrogen peroxide drive the chemical production of sulfate aerosols [*Bell et al.*, 2005], while nitrate aerosols depend on gaseous ammonia and nitric acid [*Bauer et al.*, 2007b]. Secondary organic aerosol production, which is new to our simulations, depends on modeled isoprene and terpenes as oxidized by OH, ozone, and nitrate radicals [*Tsigaridis and Kanakidou*, 2007]. All these processes can have significant direct effects on aerosol and gas concentrations, as well as indirect impacts on radiation.

Dinitrogen pentoxide (N_2O_5) strongly affects the NO_x budget in the troposphere. In the ModelE2 version, the uptake coefficient for hydrolysis of N_2O_5 on sulfate has been updated to a temperature-dependent and humidity-dependent algorithm that assumes pure sulfuric acid aerosol droplets [*Kane et al.*, 2001; *Hallquist et al.*, 2003] and results in significantly lower values in the troposphere than the uniform value applied in the CMIP3 model. Another important tropospheric process occurring on simulated aerosol surfaces is the oxidation of SO_2 on dust surfaces by ozone to form sulfate-coated dust particles [*Bauer and Koch*, 2005; *Bauer et al.*, 2007a]. Gas processing on aerosol surfaces also occurs in the stratosphere, as described in *Shindell et al.* [2006a], while polar stratospheric cloud formation is now dependent upon the abundance of nitric acid, water vapor, and temperature [*Hanson and Mauersberger*, 1988].

We now include the impact of light attenuation from aerosols on the photolysis rates of gases, following *Bian et al.* [2003]. Modeled aerosol optical depths are passed to the Fast-J2 photolysis code [*Bian and Prather*, 2002] at every time step, while the tabulated optical properties required for different aerosol types (extinction efficiency, single scattering albedo, scattering phase function expansion terms) have been calculated offline, consistently what is used in the model's radiation code. The inclusion of this interaction leads to large modifications of oxidant (OH, ozone) concentrations over particular areas (e.g., East Asia, Central Africa), while the effects on global metrics are generally smaller [*Shindell et al.*, 2013].

2.2.4. Indirect Aerosol Effects

In the NINT and TCAD simulations, the indirect aerosol effect is crudely parameterized as in *Hansen et al.* [2005], based on the empirical effects of aerosols of cloud droplet number concentration (*CDNC*) on cloud cover [*Menon and Del Genio*, 2007]. The size of the effect is related to the logarithm of the concentration of soluble aerosols, with a scale factor set so that the fixed-SST Top of Atmosphere (TOA) radiative forcing with year 2000 conditions is approximately -1 W/m^2. This target was chosen to be consistent with the magnitude of AIE in *Hansen et al.* [2005]. We made a preliminary assessment of this using decadal AMIP simulations with and without the AIE with a year 2000 background. The forcing was calculated from the TOA imbalance and temperature change using $F = \Delta Q - \lambda \Delta T$ which gave -0.95 W/m^2 for the parameters chosen. Post hoc calculations of the actual AIE forcing in 2000 diagnosed from the historical runs holding the atmosphere fixed give values of approximately -0.66 W/m^2 (see R. L. Miller et al., submitted manuscript, 2013), with the difference related to the variation in background climate and the absence of fast feedbacks in the post hoc calculation. This difference illustrates the difficulty of specifying a priori the forcing magnitude for processes like the AIE that are part of a larger atmospheric adjustment. The difference also illustrates the ambiguity in the calculation of the forcing to which surface air temperature adjusts due to feedbacks that are fast compared to the adjustment time scale [*Hansen et al.*, 2005].

In the TCADI simulations, only the first AIE is represented but it is calculated using a prognostic treatment of *CDNC* from *Morrison and Gettelman* [2008] as a function of sources, including newly nucleated *CDNC*, and losses from autoconversion, contact nucleation, and via immersion freezing [*Menon et al.*, 2010]. The nucleation term (Q_{nucl}) [m^{-3} s^{-1}] for *CDNC* is based on *Lohmann et al.* [2007] and is given as:

$$Q_{nucl} = \max\left[\frac{1}{\Delta t}\left(0.1\left(\frac{N_a \omega}{\omega + \alpha N_a}\right)^{1.27} - CDNC_{old}\right), 0\right]$$

where N_a is the aerosol concentration obtained from the aerosol mass by assuming a lognormal distribution, ω is the vertical velocity obtained by taking into account model grid-mean velocity and subgrid turbulence,

$$\omega = \bar{\omega} + 1.33(TKE)^{0.5},$$

and $\alpha = 0.023$ cm^4 s^{-1} is a constant obtained from aircraft measurements. Δt is the time step in the model and $CDNC_{old}$ is the *CDNC* from the previous time step. Sulfates, sea salt in the submicron mode (0.01–1 μm), aged organic matter and black carbon from fossil/biofuel sources, and 80/60% of organic matter/black carbon from biomass sources are used to represent the hygroscopic mass fraction of aerosols that can participate in cloud nucleation processes.

2.3. Stratospheric and Gravity Wave (GW) Drag

In the middle atmosphere, we apply a GW drag scheme above 150 hPa [*Rind et al.*, 1988]. This is as described in SEA06 but for models with a model top at 0.1 hPa, it is modified to only include effects of mountain waves and deformation waves (neglecting the impact of gravity waves generated by convection

or shear). The Rayleigh drag near the top of the model is modified in the top four layers to improve the match to the more complicated GW drag scheme in higher model top versions. Table 2 shows the constants used in the current configurations.

2.4. Radiation

Structurally, the current ModelE2 radiation treatment is as described by SEA06, and as used in intercomparisons of GCM radiation models in *Collins et al.* [2006]. More recent improvements include updating and optimizing the correlated *k*-distribution absorption coefficient tables based on the HITRAN 2008 spectroscopic database [*Rothman et al.*, 2009]. The absorption coefficient optimization involves reproducing line-by-line calculated radiative fluxes and cooling rates to within a few tenths of a W/m^2 throughout the most of the atmosphere (below $\approx$50 km) for typical current climate absorber distributions. Modifications were also made to extend the lookup tables for the *k*-correlation approximation to line-by-line results to cover a wider range of CO_2 concentrations [*Schmidt et al.*, 2010]. Cloud optical depths now include the impact of precipitating hydrometeors in addition to cloud condensate.

The mean total solar irradiance (TSI) over the climatology period (1980–2004) is 1366.4 W/m^2. Note that recent corrections for instrument bias and testing have suggested that TSI is closer to 1361 W/m^2 [*Kopp and Lean*, 2011]. This new value was not used in the CMIP5 simulations. However, due to the constraint of quasi-stable energy balance in the preindustrial control runs, the impact of the lower TSI is negligible due to the needed compensating adjustment to cloud parameters [*Rind et al.*, 2014]. Variations in incoming solar forcing in the 20th Century simulation follows *Wang et al.* [2005a], and includes updated spectral solar irradiance from *Lean* [2009], which has higher spectral resolution and more variance in the UV bands than earlier versions [*Lean et al.*, 2002]. Orbital parameters that determine the seasonal and latitudinal variation in insolation are calculated each year (updated on 1st January), based on *Berger* [1978].

2.5. Cloud Processes

ModelE2 uses a mass flux cumulus parameterization originally described by *Del Genio and Yao* [1993], based on a cloud base neutral buoyancy flux closure with two entraining plumes sharing the mass flux. Stratiform clouds are based on a Sundqvist-type prognostic cloud water approach with diagnostic cloud fraction [*Del Genio et al.*, 1996]. Updates to both schemes for CMIP3 were described in SEA06. Improvements for CMIP5 are summarized below though more detail can be found in *Kim et al.* [2011, 2012] (including post-CMIP5 changes).

In moist convection, entrainment rates and cumulus updraft speeds are calculated interactively as a function of parcel buoyancy and updraft speed [*Gregory*, 2001; *Del Genio et al.*, 2007]. Convective condensate is partitioned at each level into precipitating, detrained, and vertically advected components by comparing the updraft speed to the fall speeds of different size hydrometeors of different phases and partitioning an assumed Marshall-Palmer particle size distribution (i.e., the part with larger particles that fall faster than the updraft speed precipitate, etc.). Thus convective condensate in small particles whose fall speeds are significantly less than the updraft speed is transported upward ("ice lofting"), while the portion of frozen condensate in the form of graupel extends up to a minimum temperature that depends on updraft speed. Instead of a single downdraft, multiple convective downdrafts can originate from different levels in the same time step; downdrafts can penetrate below cloud base if they are negatively buoyant there; if positively buoyant at some level they detrain a fraction of their mass at each succeeding lower level; downdrafts now entrain and transport momentum. The convective pressure gradient is assumed to reduce convective momentum transport as in *Gregory et al.* [1997]. The adjustment time for convection to adjust the cloud base to neutral buoyancy is increased to 1 h, twice the physics time step.

Table 2. Stratospheric Drag Treatment in the Current and Previous Configurations

Model Configuration	Top Layer Drag	Interior Drag	Simple Drag Coefficients	Full GWD Coefficients[a]
E2-R/H (CMIP5)	$\gamma = 0.1$, $\mu = 0.002$	Partial GWD	-	200, 4.5×10^{-5}, 0.2
E-R/H2 (CMIP3)	$\gamma = 0.1$, $\mu = 0.002$	Simple	$\gamma = 0$, $\mu = 0.0002$	-

[a]Tunable factors for defining the generation of gravity waves and their magnitude: pressure level above which gravity waves break (hPa), deformation threshold (s^{-1}), and mountain wave factor.

Various improvements have been made to the stratiform cloud parameterization. Clouds do not now form in subsaturated air below cloud top in the convective portion of the grid box or below the cloud base of a boundary layer convective cloud. The phase in which cloud forms is maintained until the cloud dissipates, unless supercooled liquid is glaciated by the Bergeron-Findeisen process; convective snow is no longer permitted to glaciate a supercooled stratiform cloud. The critical supersaturation for homogeneous nucleation of ice is now based on *Kärcher and Lohmann* [2002]. In ModelE, when the relative humidity in a cloudy grid box drops below the threshold relative humidity for cloud formation, cloud water is assumed to instantaneously precipitate out. In ModelE2, when such conditions occur, cloud erosion is assumed to occur by evaporating cloud water until either all condensate evaporates or the grid box reaches the threshold relative humidity. Any remaining cloud water is then precipitated. The threshold for autoconversion of cloud ice to snow has been significantly decreased to excessive cloud ice concentrations in the CMIP3 models [*Waliser et al.*, 2009]. We have replaced the use of threshold relative humidities for liquid and ice clouds (as in SEA06), with threshold relative humidities for free troposphere (U_a) and boundary layer clouds; for BL clouds the threshold is based on an assumed Gaussian distribution of saturation deficit as suggested by *Siebesma et al.* [2003] with a scaling parameter U_b, while for the free troposphere (above 850 hPa in the absence of moist convection) it assumes that clouds form at lower humidity when strong rising motion is present, with a scale-aware correction for layer thickness.

Note that the downdraft mass flux into the boundary layer is now used in the surface fluxes calculation to estimate a "gustiness" effect on surface fluxes and the optical thickness of precipitation is now seen by the radiation.

2.6. Land Surface
The land surface model is structured as in SEA06. No change has been made to the vegetation cover categories but improvements have been made to the soil biophysics and vegetation components.

Full documentation on leaf biophysics and phenology are provided in Y. Kim et al. (Interannual or spatial variability of phenology for 3 plant functional types in the Ent Terrestrial Biosphere Model: Leaf area and fluxes of water and carbon dioxide, in preparation 2014). The previous ModelE coupled leaf photosynthesis/stomatal conductance model of *Friend and Kiang* [2005] has been replaced by the Farquhar-von Caemmerer model of photosynthesis [*Farquhar and von Caemmerer*, 1982] and the Ball-Berry model of stomatal conductance [*Ball et al.*, 1987]. The leaf model is a well-tested approach that accounts for the response of transpiration to temperature, humidity, wind speed, light, and seasonal leaf area index, due to the coupling between transpiration and uptake of CO_2 by photosynthetic activity. In our implementation, the solution to the coupled photosynthesis-conductance equations utilizes leaf boundary layer conductances following the approach by *Collatz et al.* [1991] but with the boundary layer conductance derived from the canopy surface layer conductance of the ModelE2 land surface model; for the cubic equations of coupled photosynthesis/conductance, we have developed an analytical solution, solving for net assimilation of CO_2. Only C3 photosynthesis is simulated for all vegetation in the current version. The canopy radiative transfer and sunlit/shaded leaf partitioning are the same as in *Friend and Kiang* [2005].

To account for variation in parameters of the leaf biophysics by climate and seasonality of vegetation, we use the following functional responses: temperature dependence of leaf photosynthetic parameters is modeled with a Q10 function [*Hegarty*, 1973], with a reference temperature of 25°C. In addition, photosynthetic capacity is subject to cold-hardening in evergreen needleleaf trees, and radiation-sensitive seasonality in tropical rainforests. With these phenological behaviors, photosynthetic capacity declines in the winter or cloudy seasons, via a factor between 0 and 1 that modifies the maximum photosynthetic capacity. Cold-hardening (of "frost-hardiness") is a function of local air temperature trend, and is modeled based on the work of *Repo et al.* [1990]; *Hänninen and Kramer* [2007]; and *Mäkelä et al.* [2004, 2006]. Radiation sensitivity was developed by Kim et al. (submitted manuscript). Water stress is modeled after the approach of *Rodriguez-Iturbe et al.* [2001], in which stress is a factor between 0 (full stress) and 1 (unstressed), as a function of saturated soil fraction (volume water/pore volume). The factor modifies the leaf stomatal conductance, and different plant functional types have different critical soil moisture values for the onset of stress and wilting points. For mixed cover types like shrub and savanna, which include both grasses and woody plants, the photosynthetic parameters are a weighted average by cover fraction of values from the different types.

Because transpiration depends on the exposed leaf area index, and wet leaf fractions will not transpire, we added a scheme for canopy interception of precipitation following the algorithm presented in *Koster and Suarez* [1996] which is similar to the modified Shuttleworth scheme of *Wang et al.* [2007]. The scheme primarily involves modification of the wet fraction of the canopy by computing the drip rate taking into account the fraction of precipitation that falls onto a previously wetted portion of the canopy with the grid cell. This scheme also distinguished between stratiform and convective precipitation, because the stratiform type covers a greater portion of the grid cell that the convective type [e.g., *Koster and Suarez*, 1996; *Wang et al.*, 2007].

In SEA06, all surface conductance of water vapor for vegetated land was assigned to the vegetation, which meant that there was no soil evaporation under canopies. The rationale for this approach in *Abramopoulos et al.* [1988] was that vegetation was simply a surface with a particular albedo and conductance. The *Friend and Kiang* [2005] version had biophysics, but transpiration was tuned to total evapotranspiration measured by eddy covariance methods. Numerous studies have demonstrated the importance of soil evaporation as a component of evapotranspiration, especially for semiarid ecosystems [e.g., *Lawrence et al.*, 2007]. We incorporate a simple representation of soil evaporation based on the work of *Zeng et al.* [2005] and *Lawrence et al.* [2007], which involves parameterizing the relative contribution of transpiration and soil evaporation for vegetated areas.

2.7. Lakes

SEA06 described the basic lake module in the CMIP3 model. Two issues arose with long term use of that code. First, imbalances in net watershed precipitation and lake evaporation in closed basins (in the Andes, Central Africa, Central Asia) led to situations where water would pile up in small lakes without an outlet, and where evaporation could not keep up with the supply of runoff. In the real world, such lakes with excess accumulation will expand within topographic limits, with consequent increases in evaporation (due to a greater surface area) such that a new balance can be found. In order to account for this effect, we now allow for variable lake fractions.

The surface area of a lake is determined by the mean volume to area ratio determined from observations (where they exist), or an average ratio (where they do not), assuming that lakes are conical in shape. Once a day, the volume of water in a lake is determined and a new lake fraction is calculated. If the lake is expanding, there is a calculation of how much water would be needed to saturate soils in the same grid box and that is taken into account in assessing the new lake fraction. There is an exact compensation between the change in fraction of lake and the change in soil and vegetated land, and water, energy and tracer budgets are adjusted so that there is no net change in any conserved variable. For lakes that are shrinking, they leave behind water in the soils corresponding to saturated conditions. Lakes below 0.001% of the grid box area are removed with their conserved variables passed into the rivers.

A second change to the lake module was made necessary because of the increasing number of grid boxes that have near 100% lake cover in basins with no current outlet to the sea (particularly in the Caspian region). In these cases, any positive hydrological imbalances cannot be dealt with via lake expansion. Instead, we have implemented two relatively ad hoc procedures to allow for water to backup into an adjacent box (i.e., river flow from upstream is reduced dependent on the height of the water in the downstream box) and in the (very rare) event that these lakes nonetheless exceed 100 m over their original depth, there is a change in river directions so that the excess water spills into a neighboring watershed that does have an outlet. Previously, ad hoc limits on surface fluxes and adjustments of lake albedo, and an adjustment factor to the global river runoff were necessary to deal with these imbalances, but these are now no longer used.

2.8. Sea, Lake, and Land Ice

As before, sea ice and lake ice processes are considered together (while taking into account the salinity and ice dynamics over the ocean) though they cannot coexist in the same grid box. In the previous coupled models, however, the sea ice, particularly in the Arctic, exhibited far too little seasonal variation in area (although the variation in thickness was reasonable) and insufficient sensitivity in extent in the late 20th century [*Stroeve et al.*, 2007; *Gorodetskaya et al.*, 2008; *Rampal et al.*, 2011]. Investigation of this problem revealed that the vertical regridding during the thermodynamic calculations was imposing unphysical (and

excessive) heat diffusion in the ice. Thus temperature profiles in the ice were too convex and the excessively cold ice was becoming too thick to melt back sufficiently over the summer. This manifested itself mainly in the Arctic. In the current code, a reformulation of this procedure has led to a large reduction in the numerical diffusion and a much larger seasonal cycle.

The sea ice thermodynamics have been coded to allow for the use of brine-pocket thermodynamics [*Bitz and Lipscomb*, 1999], but this was not finalized in time for the simulations described here, and so the salinity, as in CMIP3, is treated as thermodynamically passive in the ice, though it does impact basal ice formation/melt and heat diffusion [*Schmidt et al.*, 2004]. Note that parameters associated with the sea ice were identical for each ocean model. Sea ice dynamics use a viscous-plastic formulation of *Zhang and Rothrock* [2000] as in SEA06 with a maximum ice pressure P^* of 2.75×10^4 Pa/m.

Land ice is treated as in the previous model, with the exception that the mass and energy associated with net snow accumulation over each hemisphere's ice sheets is used to adjust the implicit iceberg calving fluxes into the adjacent oceans over a 10 year relaxation time. This ensures that the model can reach a long-term equilibrium under changed climate forcings (unlike the previous model), but does not imply that we have a reasonable simulation of the response of ice sheets to climate. Efforts are ongoing to couple more realistic ice sheet, ice shelf, and under-ice shelf cavity dynamics into the model.

2.9. Ocean, Lake, Ice, and Land Surface Coupling

In the current model, a number of changes have been made to the coupling infrastructure, particularly between the atmospheric and (varying) oceanic grids that improve conservation and interpolation of the exchanged fluxes and fields. Surface flux calculations are the same as in SEA06, with the exception of the "gustiness" enhancement mentioned above. Fluxes of energy, mass, and momentum between the atmosphere and ocean are computed on the atmospheric grid and remapped via a flux coupler. River flow occurs on the atmospheric grid and is deposited into the ocean by the coupler. Ice-ocean fluxes involving melt, heat conduction, and momentum drag are computed on the ice grid, while ice formation in the ocean occurs on the ocean grid and is interpolated to the ice grid.

2.10. Ocean Models

As in previous configurations, multiple ocean treatments can be used with the atmospheric model. As described in *Hansen et al.* [2007], we can (a) simply use the observed transient SST and sea ice fields (i.e., an AMIP-style configuration), (b) use a q-flux model (as described in *Schmidt et al.* [2006]) that allows for changes in ocean temperature as a function of surface flux variations, (c) use a dynamic ocean model. We use two dynamic ocean models in CMIP5, E2-R uses the Russell ocean model [*Hansen et al.*, 2007], while E2-H uses the HYCOM [*Sun and Bleck*, 2006]. The structural differences in the two ocean models are described in Table 3.

The HYCOM tripolar grid is a standard equirectangular latitude-longitude grid from the South Pole to 57°N, where it merges with a rotated grid with poles in Greenland and Siberia. Compared to the HYCOM version used in CMIP3, the diapycnal diffusivity was increased, primarily to counteract the net production of excessively dense bottom waters in the model. Adjustments were also made to the calculation of equivalent salt fluxes in order to preserve global integrals more precisely.

Compared to CMIP3, the Russell model has four times the resolution in each horizontal direction, and over twice the resolution in the vertical. Changes to the vertical advection were made to reduce numerical diffusivity. Since the simulations were completed, we discovered an error in the implementation of the skew flux Gent-McWilliams parameterization which had the effect of making the eddy-related diffusion more horizontal than intended. We are currently exploring the impact of this error, and the results will be reported elsewhere.

The q-flux ocean is on the same horizontal grid as the atmospheric model, but is otherwise unchanged from the CMIP3 version. The dynamic ocean models have significantly higher resolution than the atmosphere and some adjustments have been made to the regridding of fluxes between the components accordingly.

2.11. Tuning Procedure

While individual parameterizations are calibrated to process-level data as much as possible, there remain a number of parameters that are not as strongly constrained but that nonetheless have large impacts on

Table 3. Summary of Differences in Structure Between the Two Ocean Models

	Russell Ocean (E2-R)	HYCOM (E2-H)
Horizontal grid	Regular latitude-longitude	Tripolar grid north of 58°N, mercator below
Horizontal resolution	$1° \times 1.25°$	$1° \times \cos(lat)$, additional refinement at equ.
Vertical grid	z^*	Combination of near-srfc. isobaric
Vertical resolution	32 layers, layer 1 is 12 m	Isopycnic at depth; 26 layers; ref. depth for σ is 2000 m
Conserved quantities	Water mass, salt mass, heat	Mass, heat, salt content
Heat variable	Potential enthalpy	Potential temperature
Surface water flux	Fresh water flux (+salt in sea ice)	Virtual salt flux
Advection of tracers	Linear Upstream Scheme, prognostic X-Y gradients, and mean centered differences in vertical	Second-order flux-corrected transport
Vertical Mixing Scheme	KPP: background diff. 0.1 cm^2/s, incl. nonlocal terms	Below ML: diapycnal coefficient N^{-1}
Horizontal moment diffusion	*Wajsowicz* [1993]	Combin. of constant, *Smagorinsky* [1996]
Mesoscale eddy-induced fluxes	Isoneutral diffusion, skew-flux GM	Used in isopycnal subdomain only

some emergent properties of the simulation. We use these additional degrees of freedom to tune the model for a small selection of metrics. Specifically, we use the parameters in the cloud schemes that control the threshold relative humidity (U_a, U_b) and the critical ice mass for condensate conversion (WMU_i) to achieve global radiative balance and a global mean albedo of between 29 and 30% in the AMIP simulations using preindustrial ocean conditions. For the models described here, the U_a is 0.54, 0.56, and 0.55 for NINT, TCAD, and TCADI, respectively (while in these cases U_b and WMU_i are equal). Additional tunings using the gravity wave drag are chosen to optimize the simulation to the lower stratospheric seasonal zonal wind field and the minimum tropopause temperature. This also affects the high-latitude sea level pressure diagnostics. Ocean model parameters are chosen to minimize drift from observations in ocean-only simulations as described in the CORE protocol [*Griffies et al.*, 2009].

Upon coupling the ocean and atmosphere models, there is an initial drift to a quasi-stable equilibrium which is judged on overall terms for realism, including the overall skill in the climatological metrics presented in section 3. For the configuration to be acceptable, drifts have to be relatively small and quasi-stable behavior of the North Atlantic meridional circulation and other ocean metrics are required. Further fine tuning, for instance for the global mean surface temperature, is effectively precluded by the long spin-up times and limited resources available. Hence, there is a spread in mean temperatures in the coupled controls (as can be inferred from Table 4). Our procedures in producing a final model are similar to those described in *Mauritsen et al.* [2012], though we differ in the (parameterization-specific) variables used and the judgments over which elements are more important. No tuning is done for climate sensitivity or for performance in a simulation with transient forcing or hindcasts.

Diagnostics associated with ocean variables or air-sea surface fluxes may still have a residual drift associated with the control run from which the transients described here were spun off. After many hundreds of years surface temperature drifts in the matching 25 year/100 year period of the control runs are small (<0.08/0.015°C/decade), but deep ocean temperatures and salinity are still adjusting. Any diagnostic affected by this (e.g., net energy flux at the ocean surface, ocean heat content change, ocean salinity) needs to have the suitable control period subtracted away from each ensemble member in order to isolate the perturbation associated with the climate of 1980–2004 relative to the preindustrial. We calculate the control run drift using a loess smooth of 500 years of control run smoothing variance at the multicentennial and longer time scales. The long-term behavior of the controls and reasoning for the drift correction are discussed more deeply in R. L. Miller et al. (submitted manuscript, 2013).

3. Climatology

Our atmospheric model evaluation here follows the comparison in SEA06, with updates in the data products to use longer averages and/or more accurate data. We replace the ERBE data for Top of Atmosphere (TOA) radiation fluxes with the more recent Clouds and the Earth's Radiant Energy System (CERES) products, and replace the use of European Centre for Medium-Range Weather Forecasts (ECMWF) 40 year Reanalysis (ERA-40) [*Simmons and Gibson*, 2000]) with ERA-Interim [*Dee et al.*, 2011]. Sea ice and ocean evaluations use basic quantities of ice extent and volume (from the National Snow and Ice Data Center (NSIDC)), temperature and salinity at the surface and through the basins (from World Ocean Atlas (WOA) and the Polar

Table 4. Global Annual Mean Model Features Over the Period 1980–2004 (1979–2003, 1979–2002 for the GISS-ER, GISS-EH2 Models, Respectively) and Key Diagnostics Compared to Observations or Best Estimates[a]

Field	GISS-E2-R			GISS-E2-H			GISS-ER	GISS-EH2	
	NINT	TCAD	TCADI	NINT	TCAD	TCADI	(NINT)	(NINT)	Obs.
Surface air temperature (°C)	14.9	14.9	15.1	15.6	15.6	15.9	14.3	15.4	14.0[b]
Planetary Albedo	29.9	29.7	29.6	29.7	29.5	29.3	29.8	29.3	30[c]/29.4[d]
Cloud cover (%)	62	61	61	62	61	61	59	57	68[e]
Precipitation (mm d^{-1})	3.17	3.18	3.17	3.21	3.22	3.22	2.95	2.99	3.0[d]
Snowfall (mm d^{-1})	0.19	0.19	0.19	0.17	0.17	0.16	0.22	0.19	0.18[f]/0.12[d]
Atmospheric water (mm)	23.8	23.7	24.4	24.0	24.1	24.7	24.9	26.3	24.5[g]
Energy fluxes (W m^{-2})									
TOA Absorbed SW	239.5	240.1	240.5	240.3	240.9	241.4	240.0	241.6	240.2[d]
TOA Outgoing LW	238.8	239.4	239.6	239.5	240.2	240.6	239.3	241.1	239.7[d]
Surface abs. SW	169.5	170.3	170.3	170.1	170.9	171.0	167.5	169.0	165[d]
Surface down. LW	341	340	341	344	344	345	334.9	340.3	345.6[d]
Surface net LW (up)	56.9	57.6	57.5	56.9	57.5	57.4	59.9	60.4	52.4[d]/63[h]
Sensible heat flux	19.3	19.4	19.6	19.0	19.1	19.1	20.8	21.0	24[d]/17[h]
Latent heat flux	91.9	92.0	91.7	92.8	93.1	93.1	85.3	86.4	88[d]/80[h]
TOA SW cloud forcing	−48.9	−47.9	−47.6	−48.5	−47.8	−47.5	–	–	−48.4[d]/−47.4[i]
TOA LW cloud forcing	18.8	18.8	18.8	19.0	19.1	19.1	–	–	31.1[c]/28.2[i]
TOA net. rad. imb.	0.66	0.57	0.76	0.62	0.49	0.61	0.40	0.34	0.4[j]/0.36[k]/0.26 ± 0.07[f]
Trop. lower strat. water									
Vapor minima (ppmv)	4.5	4.2	4.2	4.4	4.2	4.2	3.3	4.2	3.8 ± 0.3[l]
Zonal mean tropopause temperature (minimum, DJF) (°C)	−80	−81	−81	−80	−80	−80	−79	−78	−80
Hadley circulation (10^9 kg s^{-1}) (DJF)	206	208	204	208	210	204	193	163	170–238[m]

[a]Net. Rad. Imbalance is corrected for control run drift. Cloud cover is estimated based on clouds with optical thickness >0.1.
[b]*Jones et al.* [1999]
[c]ERBE [*Harrison et al.*, 1990]
[d]*Stephens et al.* [2012]
[e]*Stubenrauch et al.* [2013]
[f]*Levitus et al.* [2009]
[g]NVAP [*Randel et al.*, 1996]
[h]*Trenberth et al.* [2009]
[i]CERES [*Loeb et al.*, 2009]
[j]*Hansen et al.* [2011]
[k]*Church et al.* [2011]
[l]*Dessler* [1998]
[m]*Stachnik and Schumacher* [2011]

science center Hydrographic Climatology (PHC)). Since many individual products contain biases that, when combined, can imply unphysical global imbalances, we use the global synthesis of *Stephens et al.* [2012] for estimates of the mean global fluxes.

The diagnostics presented here inevitably provide an incomplete view of the model climate; however, they do outline the principal successes and continuing targets for improvement in the models. In the figures that follow (Figure 2 onward), we show the annual mean observed field, the difference of the NINT-R model ensemble to that observed field, the differences between the TCADI-R and NINT-R simulations, the difference between the TCAD-R and TCADI-R simulations and the difference between the TCADI versions of the two ocean models (TCADI-H minus TCADI-R). Where the colors in the top plots are the same as in the NINT-observations plot b, that implies that the variation makes the discrepancy with data worse, while opposing signs in plots b and either c, d, or e, imply a better fit to observations. Note however, that the scales for the color bars are not necessarily the same in each plot. Other differences can be inferred from the presented results. In the final plot, we show the zonal mean absolute values for all models and the observations. Figures for surface air temperature and sea level pressure show two seasons (January-July-August (JJA) and December-January-February (DJF)) instead of annual mean values and the zonal mean values are shown in a separate figure.

3.1. Global Mean Diagnostics

The global mean quantities described in Table 4 show that some elements of the simulations are remarkably robust to physics-version and resolution. Note that for these integrated quantities the variation across ensemble members is small. The net albedo and TOA radiation fluxes at the preindustrial are tuned for, and so it should be no surprise that they are similar across models and to observations. Global mean surface

TOA Absorbed Shortwave Radiation (W/m^2)
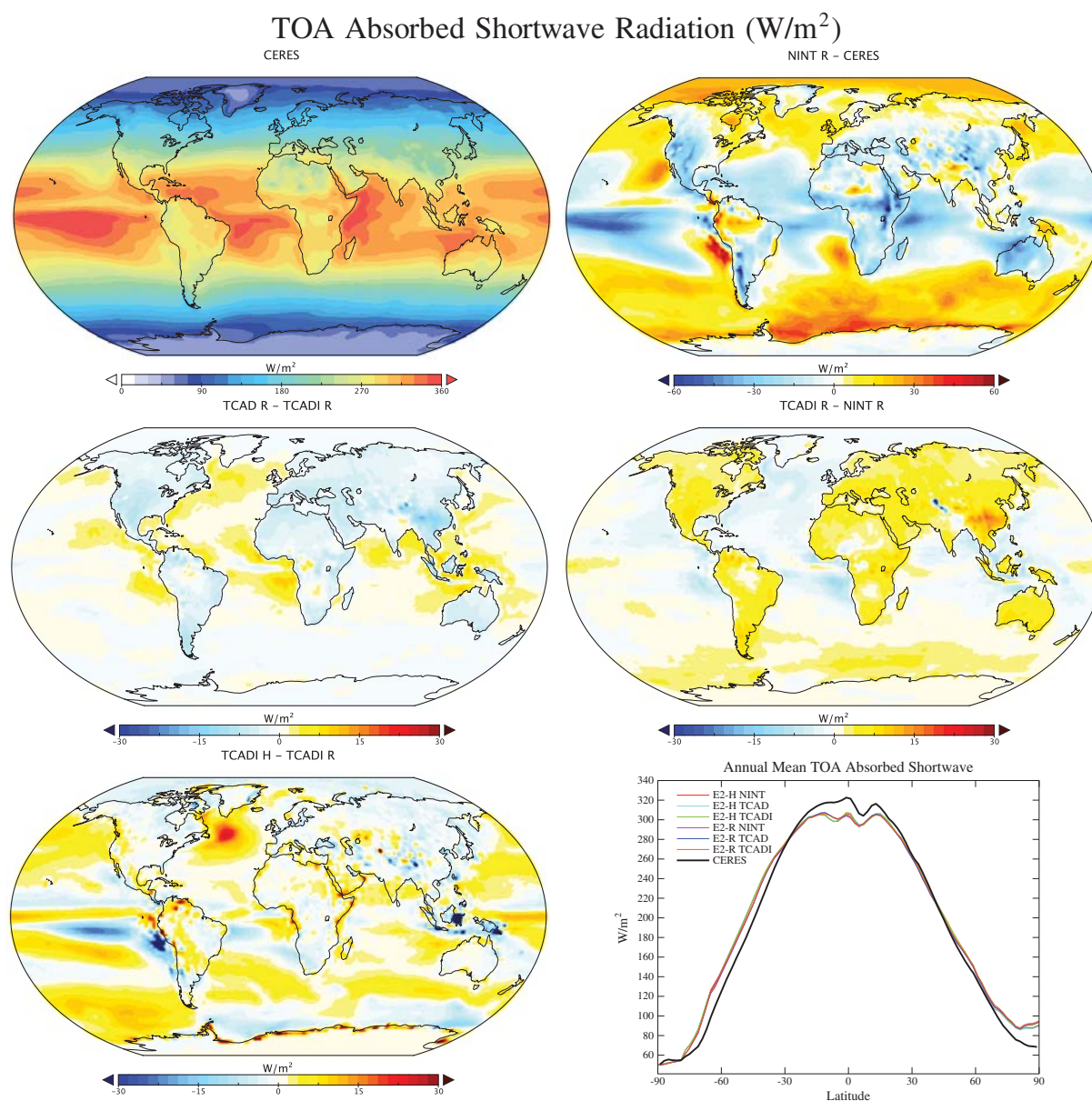
Figure 2. Annual mean absorbed shortwave radiation at TOA. (a) Observed data from CERES, (b) the difference between the ensemble mean of the NINT-R simulations and CERES, (c) the difference between TCAD-R and TCADI-R simulations, (d) the difference between TCADI-R and NINT-R simulations, (e) the difference between TCADI-H and TCADI-R simulations, and (f) the zonal mean absolute values for the observations and each simulation ensemble.

temperature is not specifically tuned for and is systematically warmer by up to 1°C in the coupled model using HYCOM. Consistent with this is a more active hydrological cycle (higher precipitation and evaporation rates). Precipitation and surface latent heat flux are uniformly high, particularly when compared to GPCP/CPC Merged Analysis of Precipitation (CMAP), around 2.6 mm/d [*Xie and Arkin*, 1997; *Huffman et al.*, 1997]. This is partially related to undercounting (perhaps 10%) in the remote sensing [*Stephens et al.*, 2012; *Trenberth et al.*, 2009] which is taken into account in the estimate used. The global Bowen ratio (sensible heat/latent heat) ≈21% matches some newer estimates ≈21% well [*Trenberth et al.*, 2009] but is less than others [*Stephens et al.*, 2012]. Total cloud cover is improved over ModelE-R, but still too low. Updates of the TOA radiation fluxes and cloud radiative forcing data from ERBE [*Harrison et al.*, 1990] to the CERES data [*Stephens et al.*, 2012; *Loeb et al.*, 2009] slightly improve the match to the model, but the LW CRF is

systematically too low. Across all model versions, absorbed solar and net LW are a few W/m^2 high, latent heat is too high and sensible heat too low, compared to the Stephens et al. synthesis. However, the surface energy budget terms are still uncertain (with estimates differing by $\approx$10 W/m^2 depending on the source and analysis method [*Stephens et al.*, 2012]).

There is a systematic radiative imbalance between incoming shortwave and outgoing longwave due to the rapid increase in radiative forcing in the late 20th Century in all simulations. The net surface heating and net ocean heat accumulation are of a similar magnitude. The net imbalance is a little smaller in the E2-H runs than in the E2-R runs consistent with a less effective ocean diffusion, and higher SST trends in the HYCOM. There is a systematic difference in the imbalance as a function of physics-version; TCAD models have the lowest imbalance, followed by NINT and then TCADI, though this is a function of both the differences in forcings and sensitivity. Over all 30 simulations, the net imbalance ranges from 0.42 to 0.92 W/m^2 (intra-ensemble standard deviation is about 0.1 W/m^2). An observational estimate is based on the rate of change of World Ozone Data Center (WODC) estimates of ocean heat content changes over 0–2000 m for 1980–2004, but this might be biased low due to data sparseness in the early period and lack of information from greater than 2000 m depth [*Levitus et al.*, 2009]. Another estimate (though for the period 1972–2008) is higher [*Church et al.*, 2011], but may suffer from similar data quality issues. A recent assessment, including nonocean heat content changes (+0.04 W/m^2), suggests an imbalance over this period of around 0.4 W/m^2 [*Hansen et al.*, 2011]. Increases in the expected imbalance over the CMIP3 runs are very likely related to a reduction in net aerosol forcings in the newer simulations.

The Hadley circulation strength is on the high end of estimates derived from the reanalyzes [*Stachnik and Schumacher*, 2011], and tropical tropopause temperatures are close to observed. Lower stratospheric water vapor minima are on the high side of observed estimates but accurate to within a few tenths of a ppmv.

3.2. Radiation Data

Estimates of the TOA radiation balance from the CERES measurements of the Earth's radiation budget (CERES, 2003–2010 [*Loeb et al.*, 2009]) are compared to the models in Figures 2 and 3.

The models are consistently biased low in absorbed solar radiation (implying a too high albedo) in the tropical regions and the continents, and high in the midlatitudes (too low an albedo). Note that the model tuning process ensures that the global mean bias is close to zero, thus the two biases are coupled. It is likely that this overall bias is driven by the model's poor performance in the Southern midlatitudes, where insufficient clouds are produced that are insufficiently reflective (and so for overall radiative balance, the tropical clouds are made too reflective). The marine stratus regions also stand out as areas of high biases (insufficient cloud). In the longwave, there is insufficient radiative cooling to space in areas of deep convection, and in the subsiding regions there is an excess of Outgoing Longwave Radiation (OLR). These two factors have implications for the implied total poleward heat transport (see section 3.9).

The difference introduced by TCAD physics is minor (a few W/m^2 improvement over the continents, a slight worsening in the Southern Oceans) but the difference in TCADI is larger (Figures 2c and 2d) There is a decrease in cloud cover and albedo over the continents—particularly in polluted regions of Asia, increasing the absorbed solar radiation by up to 10–15 W/m^2. This increase is compensated by increasing cloudiness (decreases absorbed solar) over the oceans—improving slightly the overall match to the observations.

The difference made by the ocean component is isolated to areas of the North Atlantic and tropical Pacific where the sea surface temperatures are most affected by the ocean model differences involving mixing and advection.

3.3. Cloud-Related Data

We use the International Satellite Cloud Climatology Project (ISCCP) [*Rossow and Schiffer*, 1999] and the ISCCP cloud simulators [*Klein and Jakob*, 1999; *Webb et al.*, 2001] built into the GCM to compare cloud-related fields. More sophisticated satellite simulators (for CloudSat and CALYPSO) will be assessed elsewhere in papers related to the Cloud Feedback Model Intercomparison Project (CFMIP) project. The ISCCP data have some clear biases, mainly over the Indian Ocean where polar orbiter data are used to infill between geostationary satellites. Other (smaller) biases can be found elsewhere on the fringes of the geostationary satellites' field of view where observer zenith angles are high. Over the polar regions ISCCP is biased low

TOA Outgoing Longwave Radiation (W/m^2)
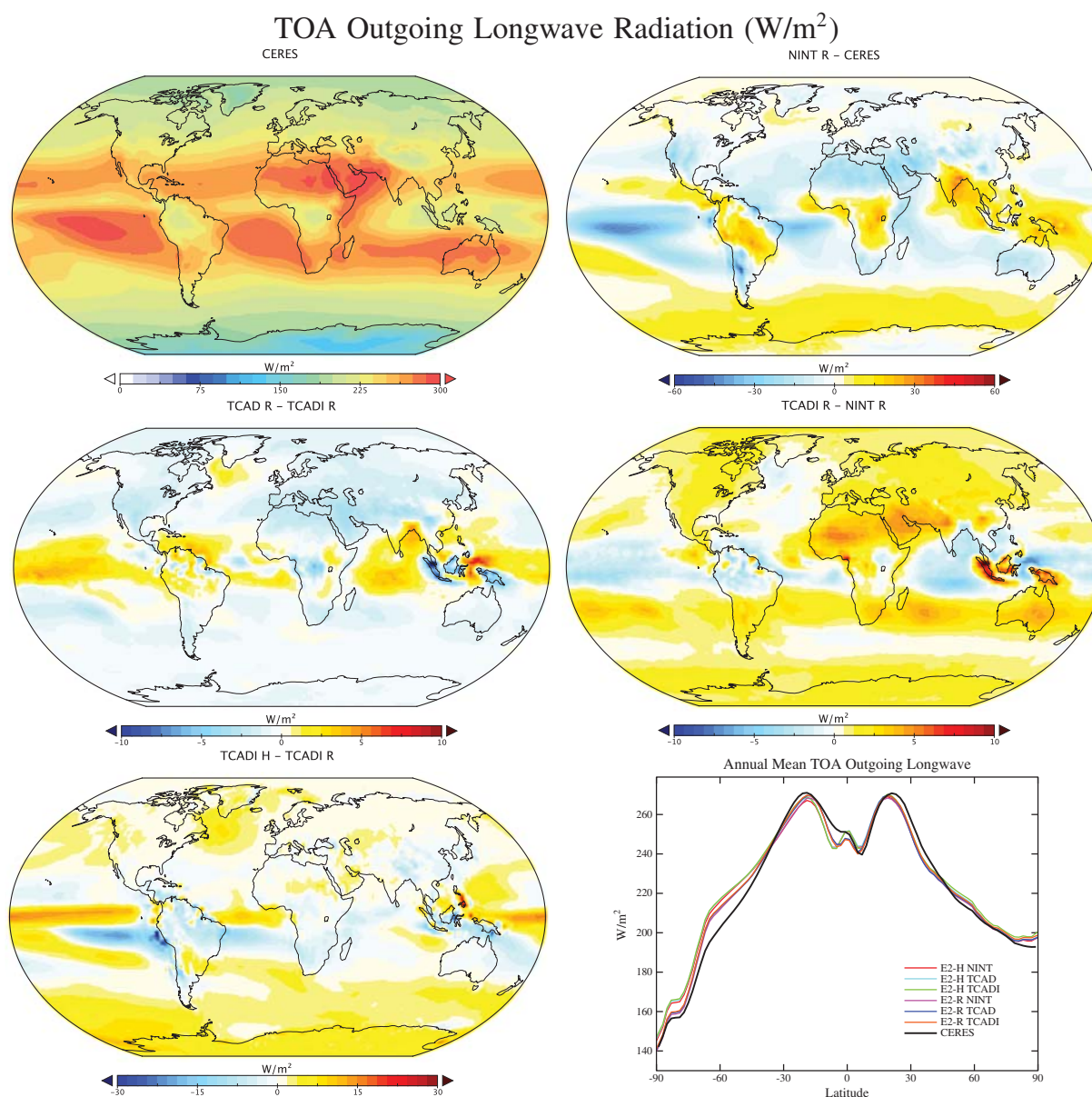
Figure 3. Annual mean outgoing longwave fluxes at TOA. Plots as in Figure 2.

because passive techniques cannot differentiate cloud from snow/sea ice in the visible and have trouble in the IR because inversions make the air radiating temperature too similar to the surface temperature.

For total cloud cover (Figure 4), even accounting for ISCCP biases, the differences to the model are stark. Over midlatitude ocean regions, the model does not produce enough cloud (consistent with the radiative property comparisons discussed above), while there is an excess of cloud along the equator. Over the continents, there is also a ≈10% high bias. Differences between model versions are significant, but much smaller, for instance, TCAD simulations have a couple of percent less cloud. The TCADI ensemble, with its mechanistic treatment of the AIE, shows reduced errors in low cloud cover over the Gulf of Alaska and North Atlantic (Figure 5), where anthropogenic aerosols from East Asia and North America, respectively, are transported into relatively pristine oceanic air. The TCADI ensemble also shows improved low cloud cover within the eastern equatorial Atlantic, downwind of a region of biomass burning, and general improvement of total cloud cover over land (Figure 4). With respect to the ocean models, there is a decrease in North Atlantic
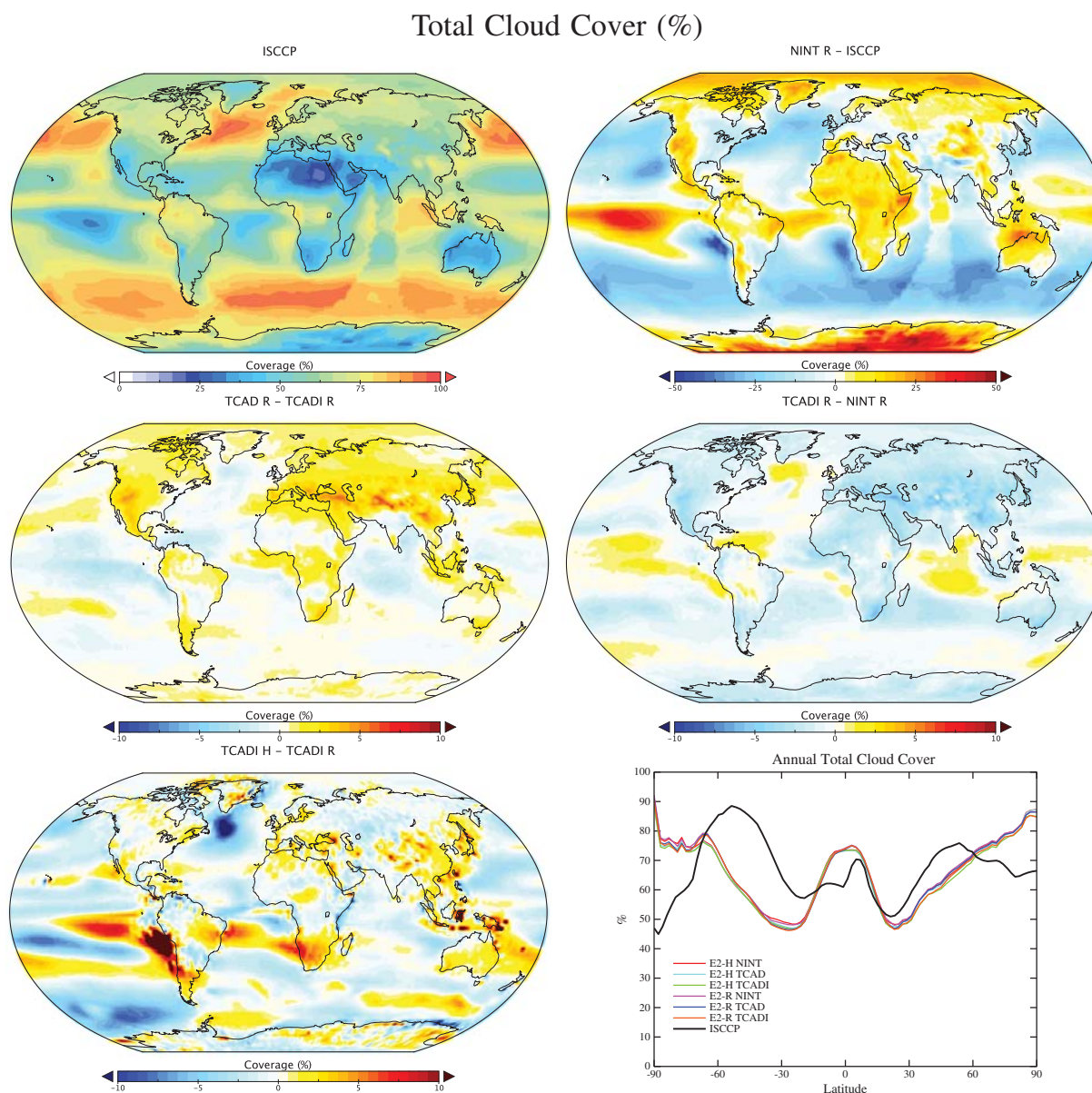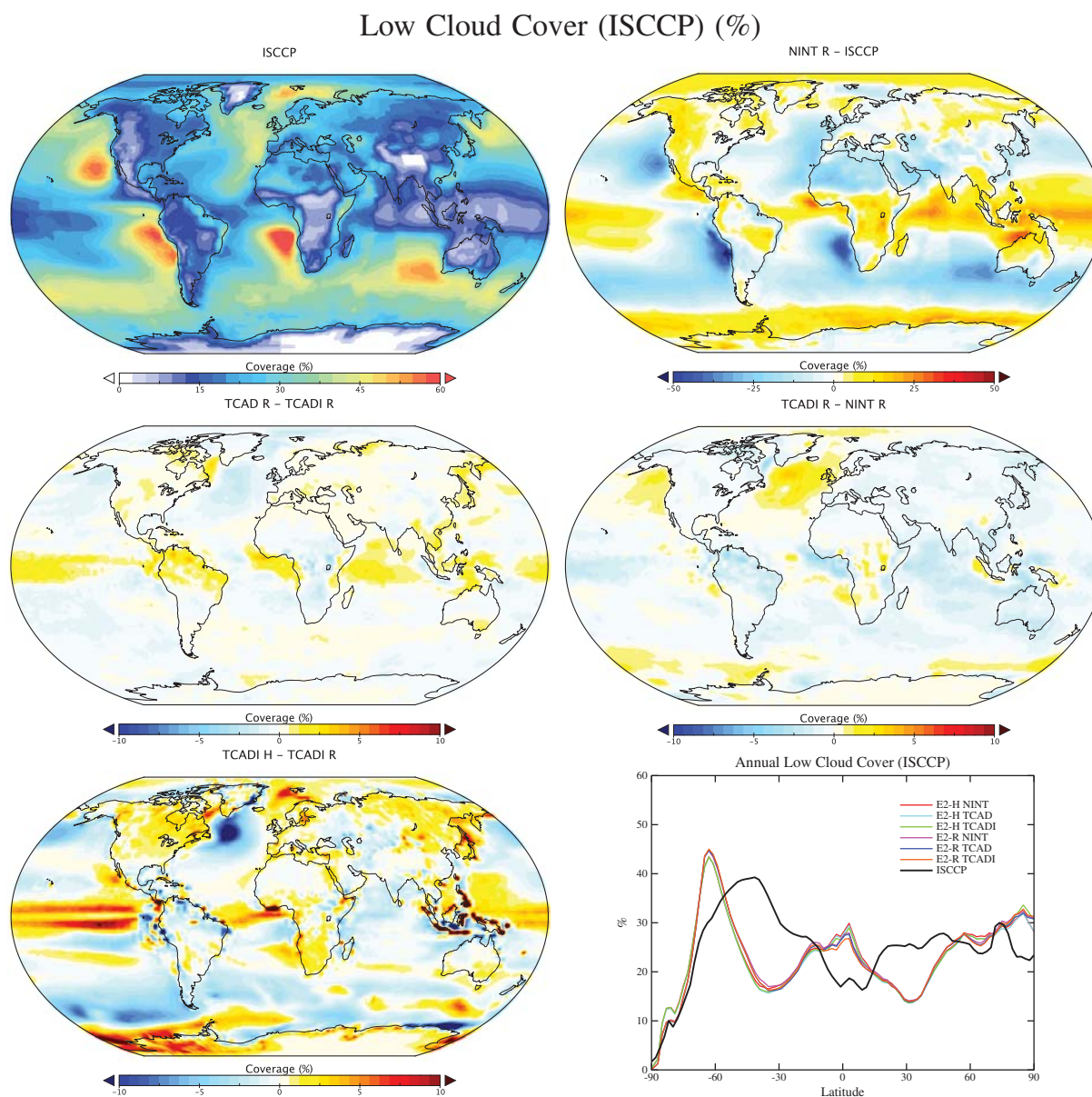
Figure 4. Annual mean total cloud cover (%) compared to ISCCP. Pots as in Figure 2.

cloudiness in response to the warmer SST with the HYCOMs, which makes the match to observations worse. There are big increases of cloud off the coasts of Peru and Southern Africa in the HYCOMs, almost removing the bias seen in the Russell models. However, looking specifically at the low clouds (Figure 5) reveals that the extra cloud is at middle to high levels and not reflective of an improved marine stratus deck. The slightly worse double-Intertropical Convergence Zone (ITCZ) problem in the HYCOMs is also apparent in the low cloud difference.

The ISCCP histograms of annual mean cloud top pressure/optical depth pairs (Figure 6) do not vary substantially across either atmospheric physics or ocean model, and so we only show one set for clarity. There are substantial differences from previous models (cf. Figure 9 in SEA06), and indeed from the observations. In the midlatitudes, we are missing relatively thin midlevel clouds, while tropical and subtropical clouds are at the right height, but slightly too optically thick. The Southern Hemisphere subtropics are noticeably deficient in midlevel cloud, consistent with the radiative biases discussed above.

Low Cloud Cover (ISCCP) (%)

Figure 5. Annual mean low cloud cover compared to ISCCP, using the internal model ISCCP cloud simulator. Plots as in Figure 2.

The cloud radiative forcing is again very similar across the models and, in the global mean, similar to the CERES analysis (Figures 7 and 8 and Table 4).

3.4. Hydrological Data

The precipitation patterns (Figure 9) show some specific deficiencies compared to the observed GPCP (1987–1998) patterns [*Huffman et al*., 1997]. In all models, there is a tendency toward an excessive "double ITCZ," a noted problem in many CMIP5 models [*Hwang and Frierson*, 2013]. Total precipitation is also high as discussed above. Differences across atmospheric composition treatments are small, but are significant in the Western Warm Pool, where the TCADI runs have less of a land-ocean contrast than observed. The differences due to ocean models follow the SST differences. Amazonian rainfall is too low across the board. Snowfall amounts (Figure 10) are consistent with satellite retrievals [*Liu*, 2008] in all models, but with a slight
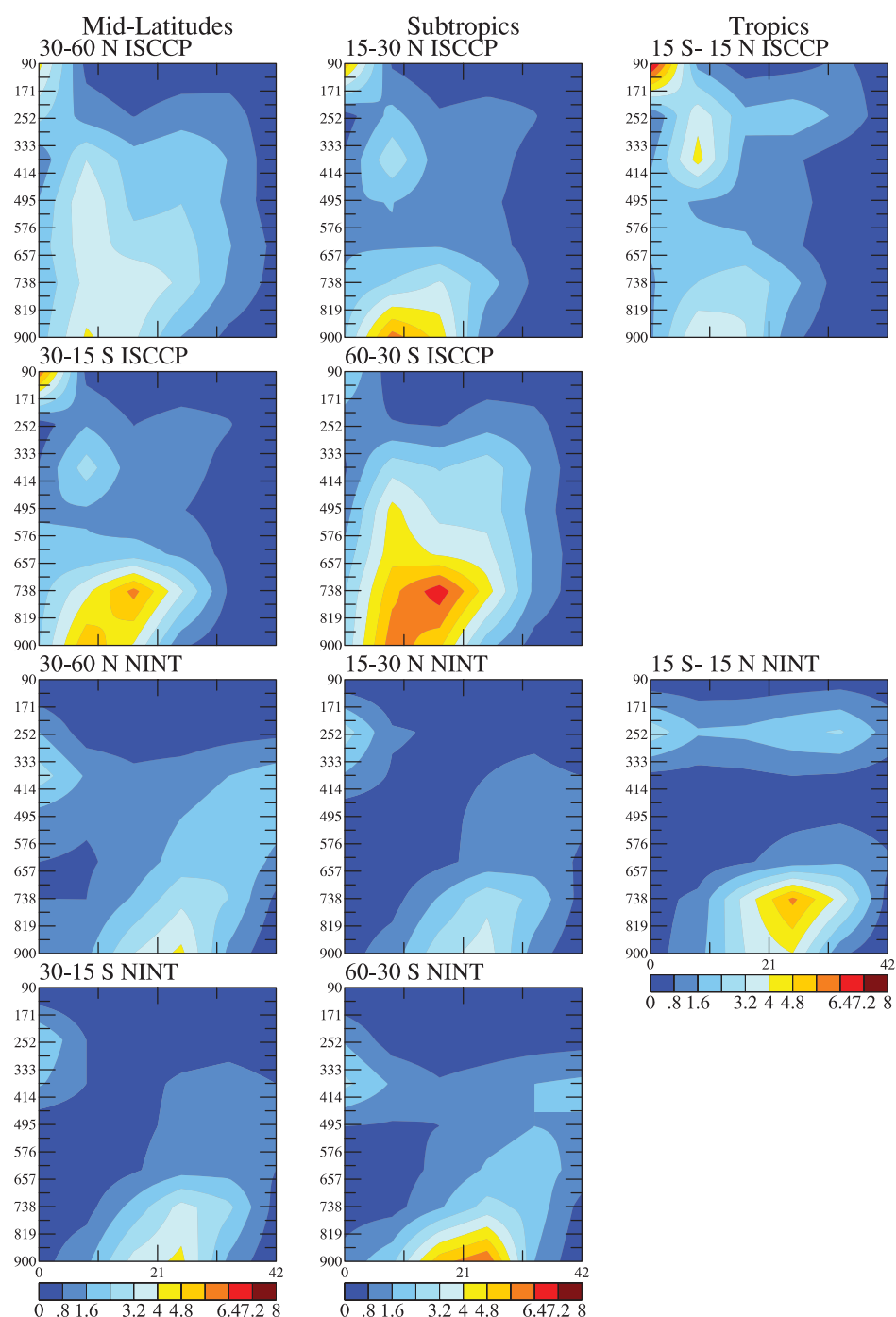
Figure 6. ISCCP optical depth/cloud top pressure histograms for different regions (NH Midlatitudes: 60°N–30°N, NH Subtropics: 30°N–15°N, Tropics: 15°N–15°S, etc.). Model output is only shown for E2-R (NINT). Other configurations are similar.

southward bias of snowfall peak amounts in the Southern Hemisphere, most likely associated with insufficient sea ice cover and too warm conditions in the Southern Ocean.

To deal with a notable dry bias in the tropics in the standard Atmospheric Infrared Sounder (AIRS) products [*Fetzer et al.*, 2006], we compare the model column water vapor to a blended product that uses SSMI over the ocean and a scaled value of the AIRS retrieval over land. The global scaling factor is given by the ratio of
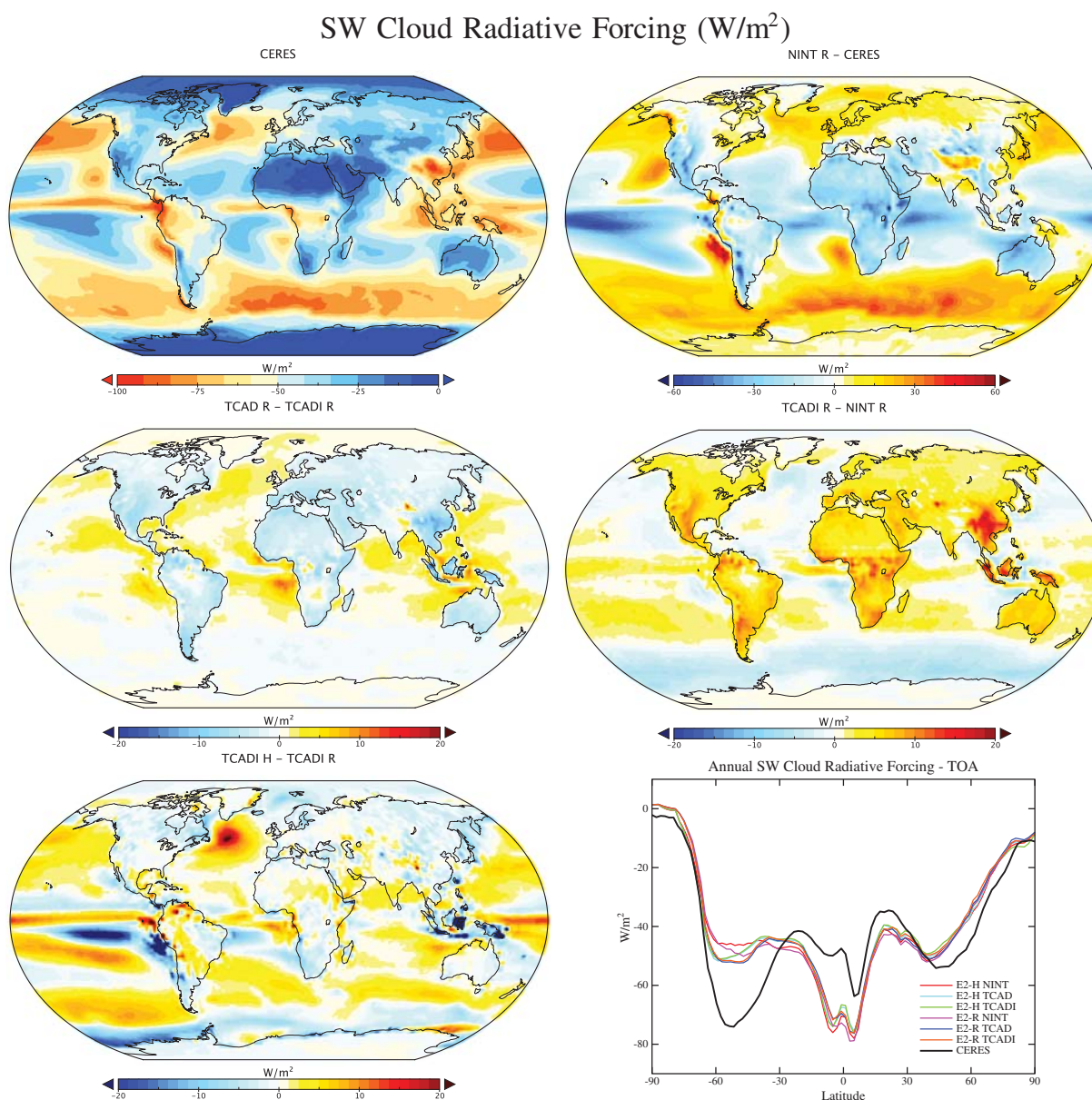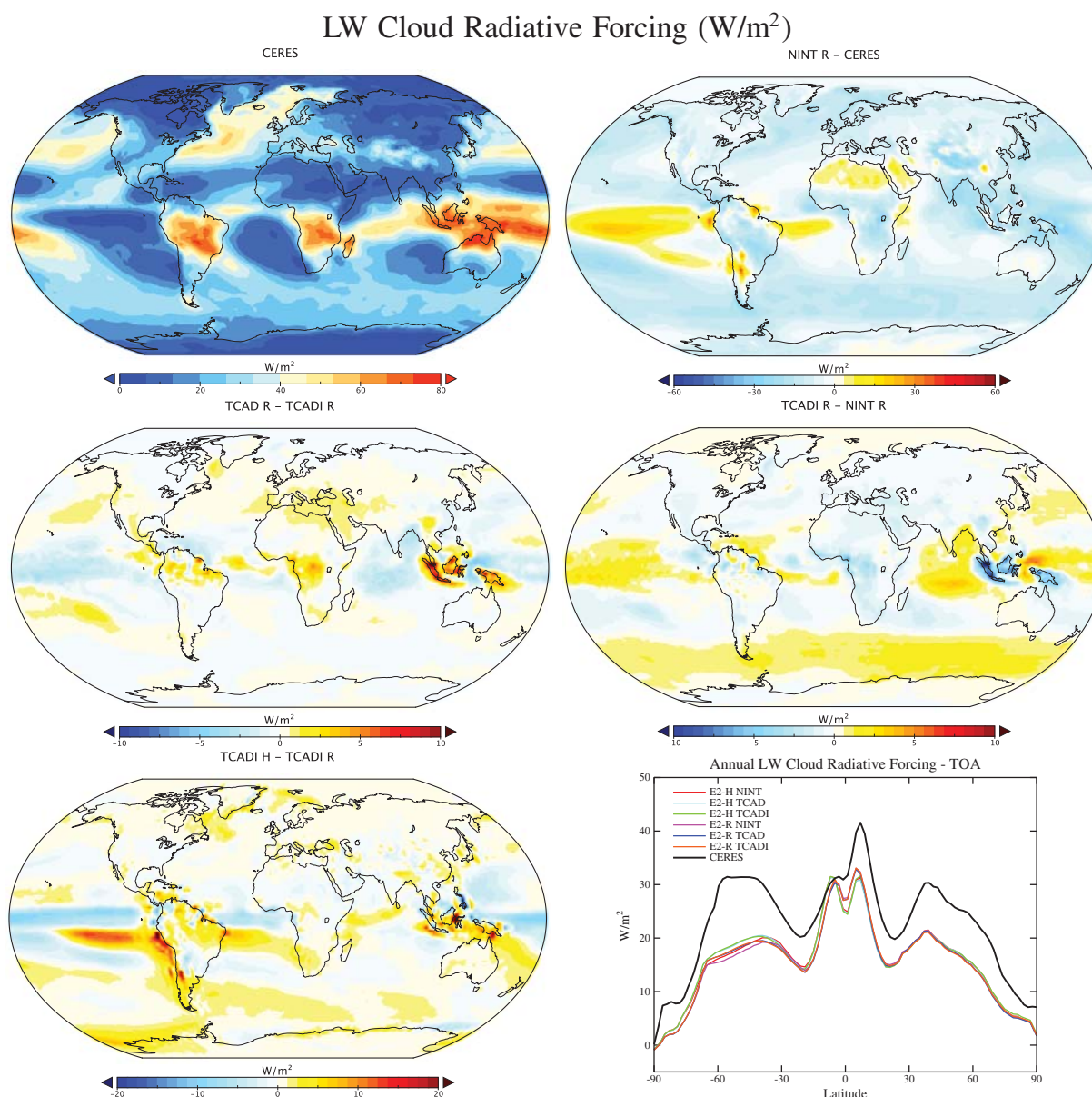
Figure 7. SW Cloud Radiative Forcing calculated within the models and from CERES estimates. Plots as in Figure 2.

the SSMI values divided by the AIRS values over the ocean. The column water vapor values (Figure 11) are well modeled in the zonal mean, but there are clear differences in the tropics, associated with an insufficiently cold Eastern Pacific and subtropical Atlantic Ocean. Overall, the model does not capture the slight tropical asymmetry seen in the retrieval. Vapor amounts over the Indian subcontinent are clearly deficient, and there is a noticeable dryness over the Northern Hemisphere mid to high latitudes. The impact of interactive composition in TCAD is small, but including the prognostic AIE in TCADI does generally increase atmospheric water amounts. As with precipitation, the water vapor differences across the ocean models follow the SST difference.

3.5. Zonal Mean Temperature and Wind Data
Comparisons of the models zonal wind and temperature output (Figure 12) are evaluated up through the stratopause against the CIRA data set [*Fleming et al.*, 1990] for DJF though some offsets are to be

LW Cloud Radiative Forcing (W/m^2)

Figure 8. LW Cloud Radiative Forcing calculated within the models and from CERES estimates. Plots as in Figure 2.

expected because of the development of the ozone hole in the later part of the climatology period. Differences due to the vertical resolution and gravity wave treatments compared to previous versions are seen in the stratosphere and higher. The models are a reasonable match in temperature except for the near-stratopause values (5°C too warm in the winter hemisphere, too cold in the summer) where the model top is damping the circulation. Above the stratopause in the summer hemisphere the interactive models are 10°C or so too warm. Zonal winds in this region are too weak (and insufficiently variable) in all cases. Peak winds in the winter hemisphere in the stratosphere are weaker than observed and have peak too close to the equator. Tropospheric jet speed maxima are about 10 m/s too fast. Differences as a function of atmospheric composition treatments are small except above the stratopause and regions of significant ozone variability, most notably in a dipole in the summer lower stratosphere. The impact of the ocean treatments is negligible.

Precipitation (mm/day)
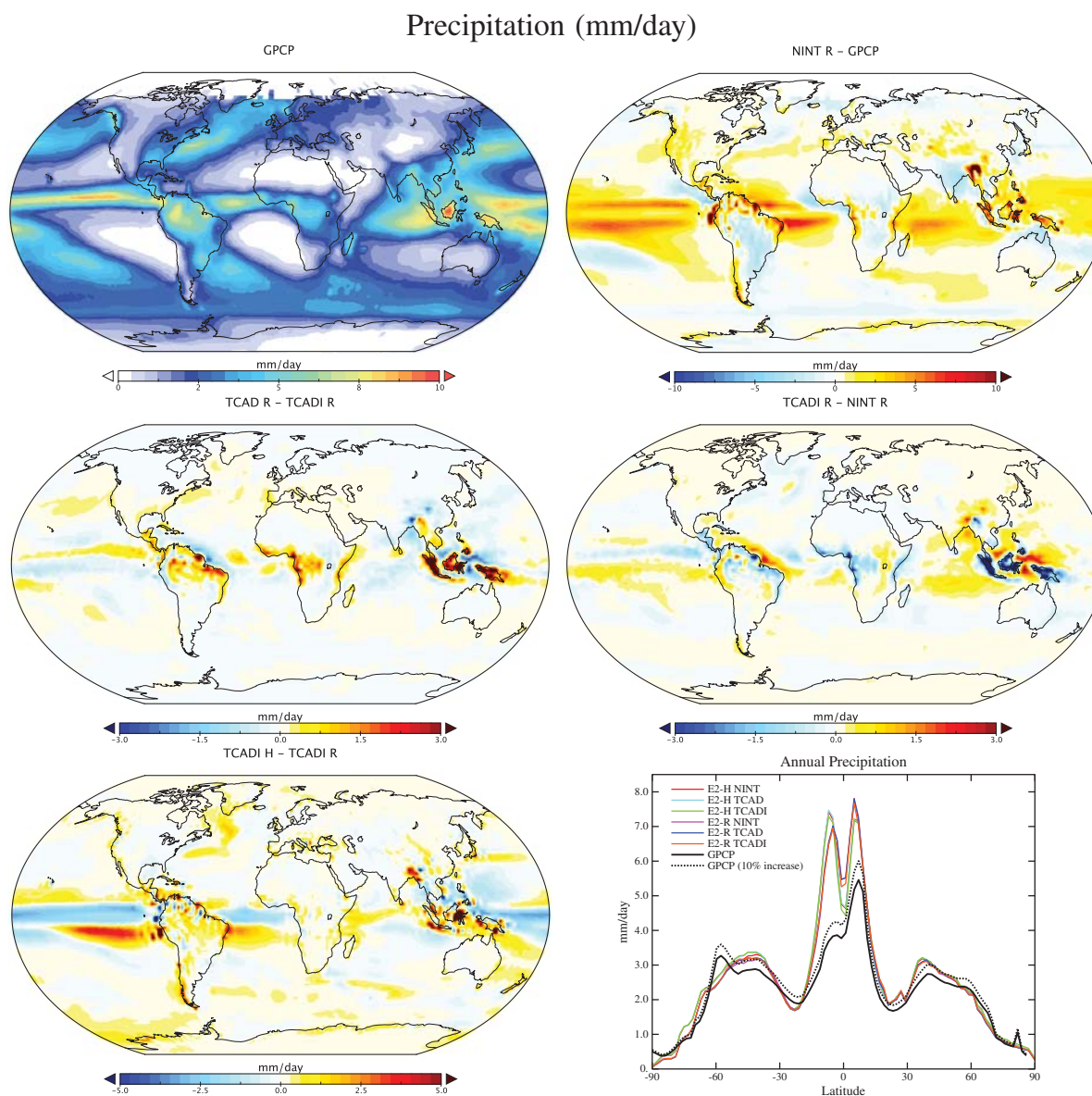
Figure 9. Precipitation (mm/d) compared to GPCP. Plots as in Figure 2. We additionally show the zonally averaged GPCP data enhanced by an additional 10% to account for potential undercounts [*Stephens et al.*, 2012].

### 3.6. Surface and Atmospheric Temperature Data

Surface air temperatures (SAT) (Figure 13) are shown in comparison to the ERA-Interim data (1979–2000). The NINT models have excessive warmth in the tropics (particularly so over land), along with too cool conditions in the northern midlatitudes in all seasons. Differences to TCAD are minimal, though the TCADI simulations have improvements in midlatitudes. Substantial differences are seen with different ocean models, with the HYCOM versions substantially warmer in the southern high latitudes and North Atlantic, associated with different ocean circulations.

Moving up through the atmosphere, we compare the models to the microwave sounding unit (MSU) 1978–2004 brightness temperature climatologies [*Mears et al.*, 2003] (Figures 14 and 15). We highlight results from the midtroposphere (TMT) and the lower stratosphere (TLS) which have global weightings centered on 600 and 70 hPa, respectively (though with substantial tails). We use a static weighting function to
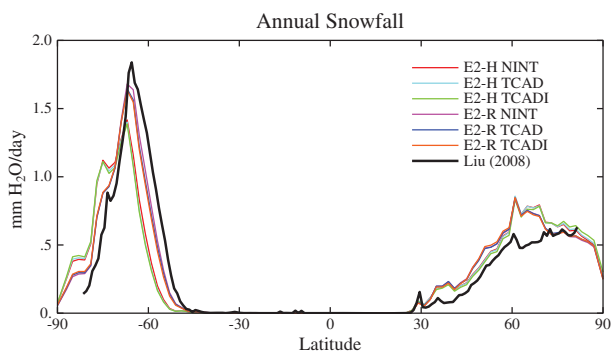
Figure 10. Zonal mean snowfall (mm/d) in the simulations compared to satellite retrievals (*Liu* [2008] and updates).

estimate the MSU channels, which though slightly less accurate than a radiative transfer calculation that takes into account surface emissivity, atmospheric water vapor, and temperature profiles [*Shah and Rind*, 1995], does not produce significantly different results. MSU-TMT shows a slight (1–2°C) warm bias in the tropics, and a similar cool bias in the midlatitudes. This is slightly improved in the TCAD and TCADI runs and reversed in the HYCOM

Column Water Vapor (mm H$_2$O)
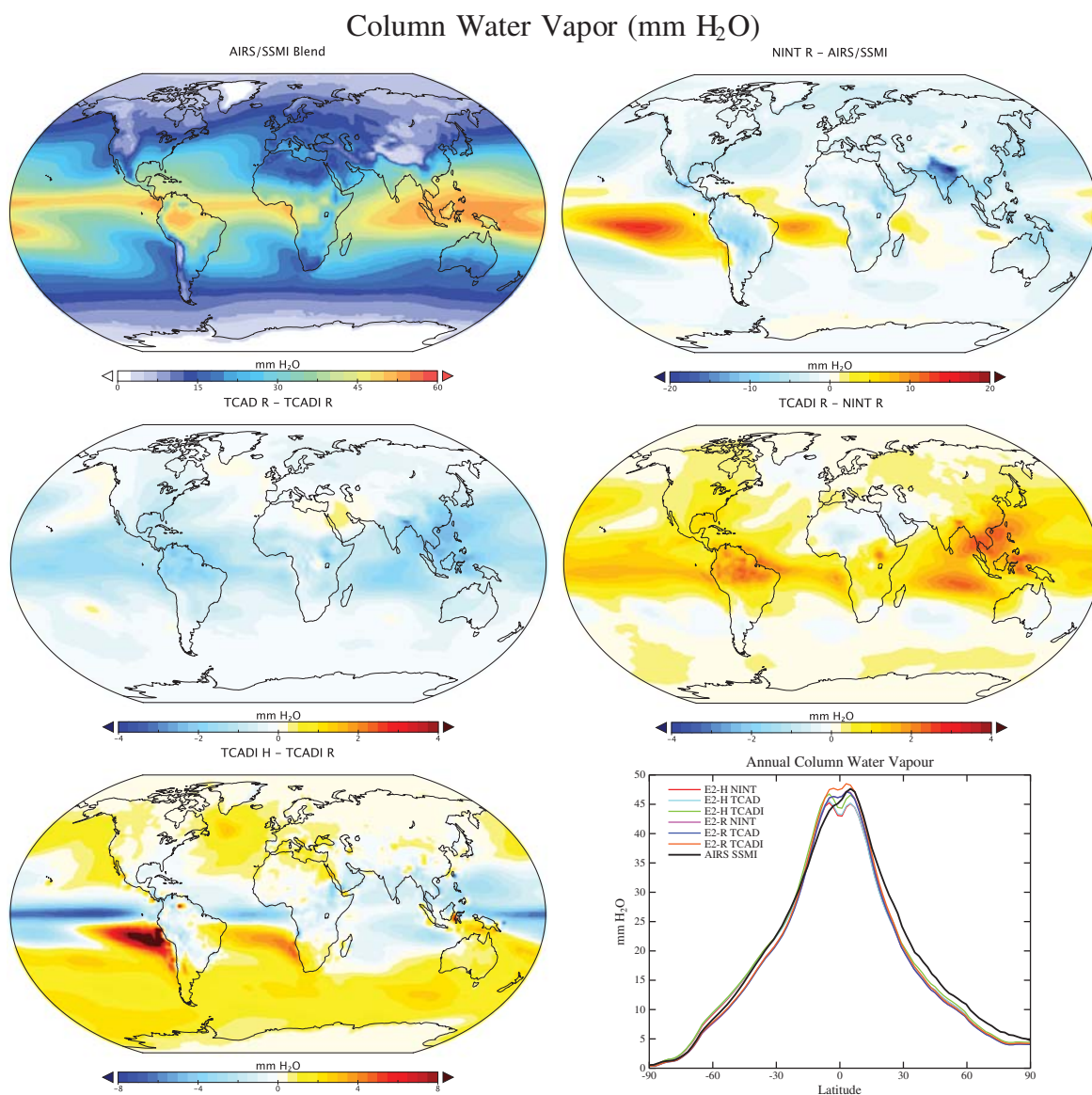
Figure 11. Column water vapor compared to a blended product from SSMI over the ocean and scaled AIRS values over land. Plots as in Figure 2.

Zonal Wind and Temperature
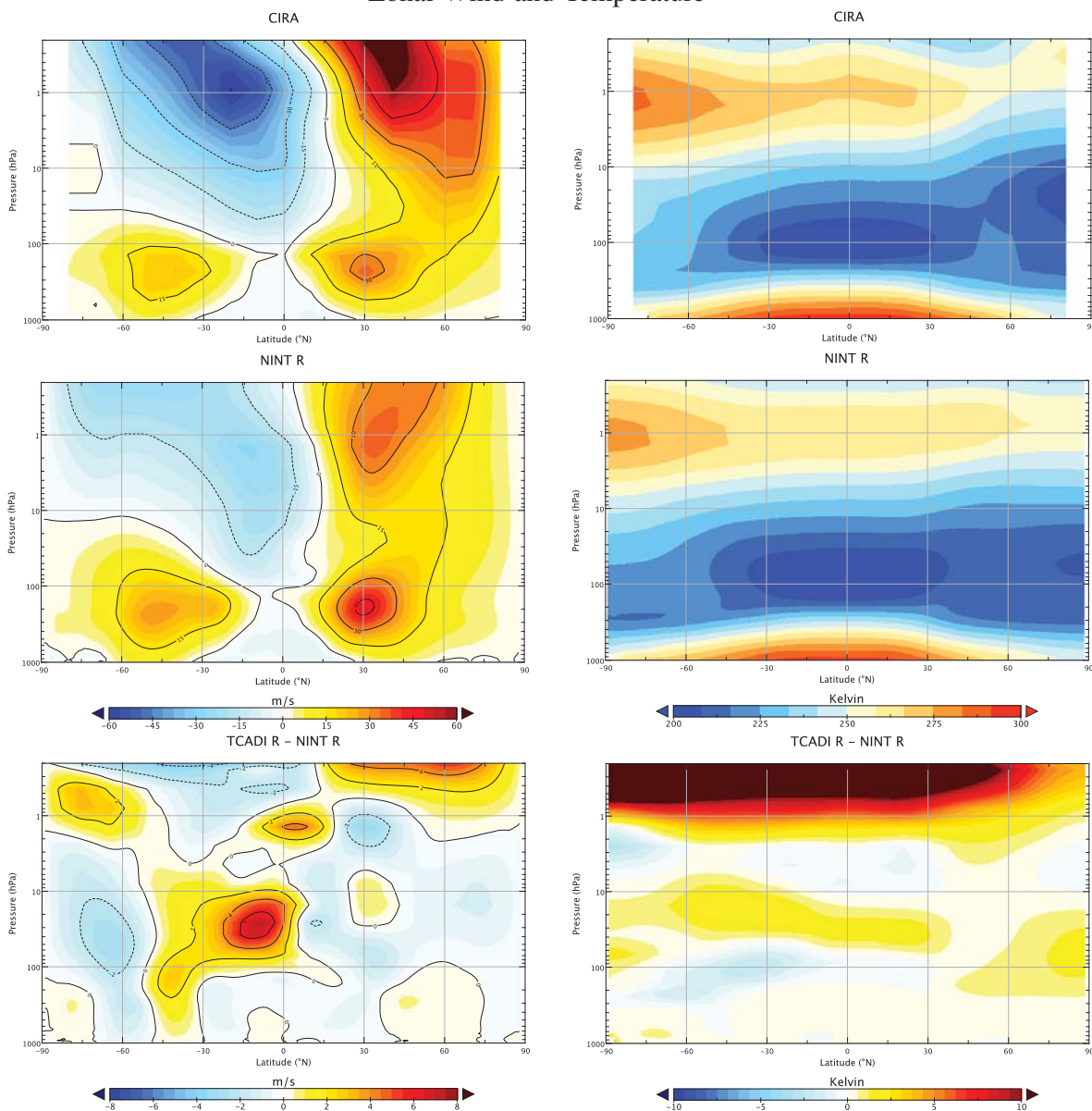
Figure 12. (left column) Ensemble mean DJF zonal wind and (right column) temperature compared to CIRA (*Fleming et al.* [1990] and updates) for E2-R NINT and TCADI. Differences to other versions are slight.

ocean runs. For MSU-TLS, there is an overall cool bias, and the clear differences in the Southern Hemisphere for the TCAD and TCADI runs are related to differences in timing of the imposed and calculated ozone depletion effects [*Shindell et al.*, 2013].

3.7. Surface Data

For sea level pressure (SLP) comparisons, unlike previous models, all models now have slightly too high SLP in the tropics (Figures 16 and 17). Differences among the models are partly a function of local responses to SST differences—particularly in the North Atlantic, but are most noticeable around the Antarctic, where SLP differences associated with Southern Annular Mode (SAM) are likely driven by the slightly different mean
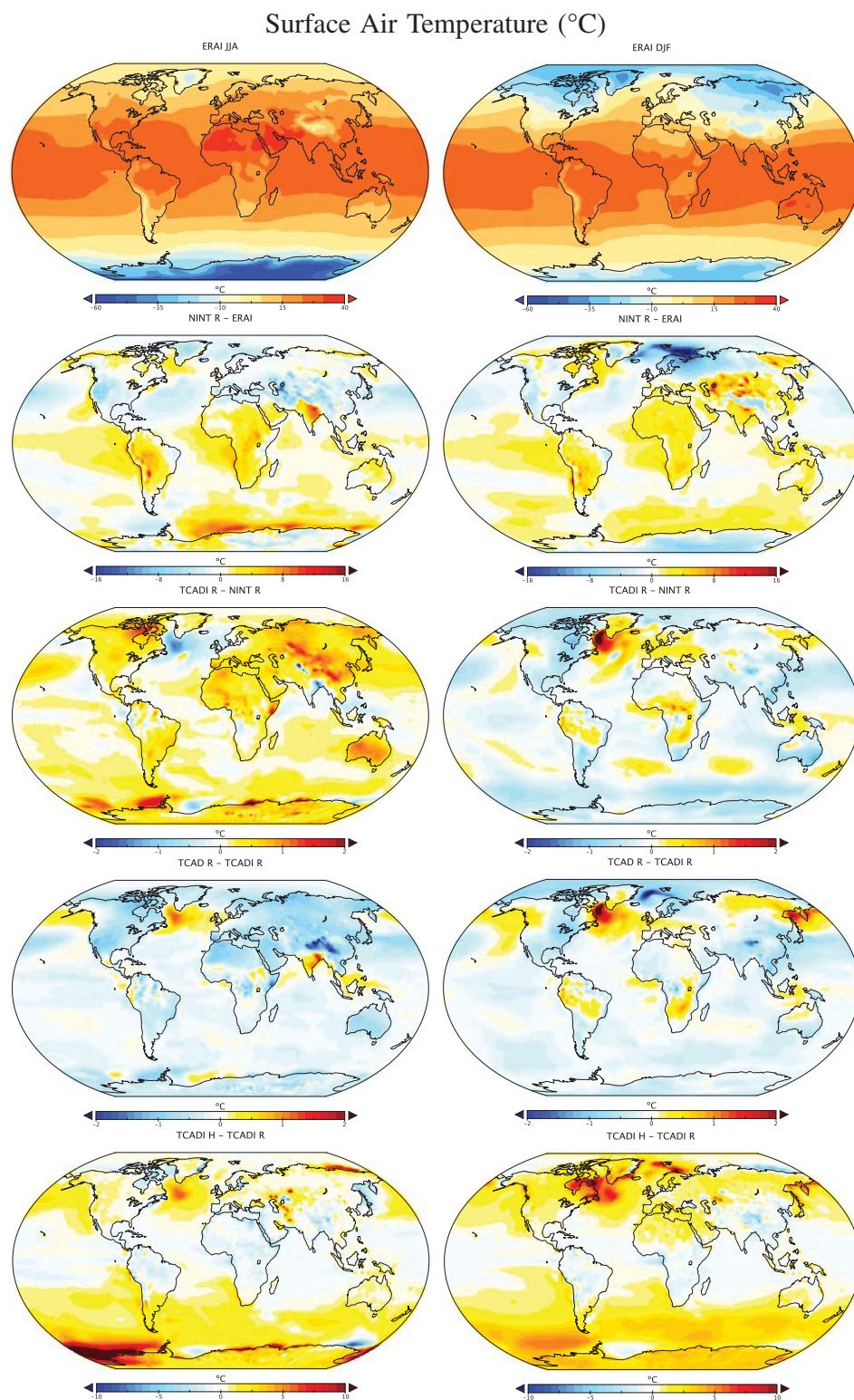
Figure 13. Seasonal surface air temperature from ERA-Interim (row 1). Differences between; NINT and ERA-Interim (row 2), TCADI and NINT (row 3), TCAD and TCADI (row 4), and TCADI-H and TCADI-R (row 5); (left) JJA, (right) DJF.

## Microwave Sounding Unit Temperature — Mid-Troposphere (°C)
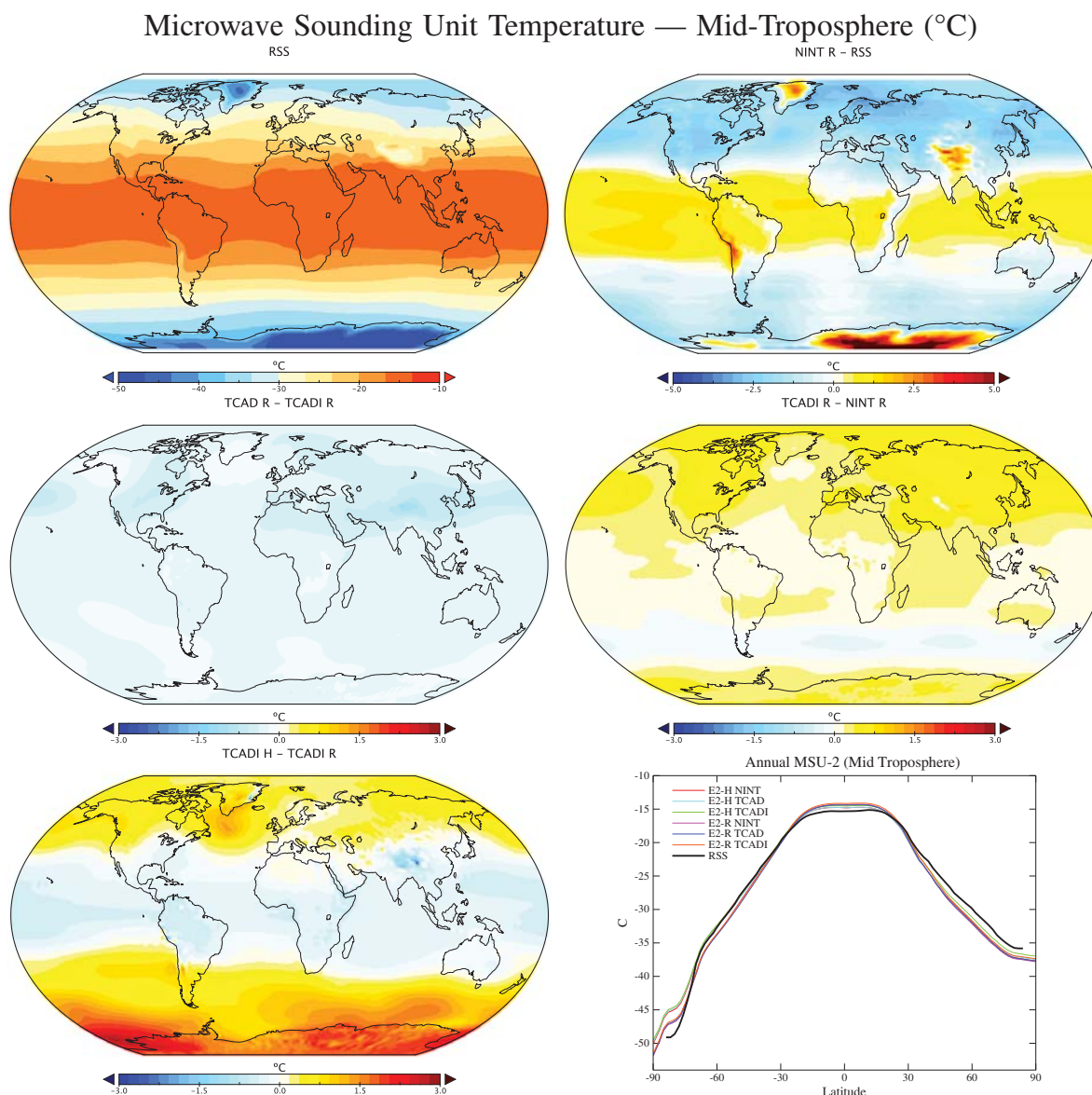
Figure 14. MSU Midtropospheric temperatures compared to RSS. Plots as in Figure 2.

state of the ozone hole in each model. Wind stresses are improved over the CMIP3 models (not shown), particularly in the Southern Hemisphere, but still slightly deficient in the North Atlantic storm track region.

3.8. Land Surface Hydrology

Runoff from the major rivers can be compared to observational data [*Milliman and Meade*, 1983] (Table 5). In the tropics, runoff is severely deficient in the Amazon basin and African rainforests (due to insufficient rainfall, and especially so in the TCADI runs). This is a decrease in skill compared to earlier model versions. High-latitude rivers are more consistently modeled.

The consequence of the new lake functionality is that there is now a seasonal cycle and variability in lake extent. However, while the extent is now more coherently coupled to the local hydrological balance, there will be a penalty in a less realistic spatial distribution of lakes at the annual average (since of course no model can do better than the previous method of simply imposing the observed fractions). Comparisons of

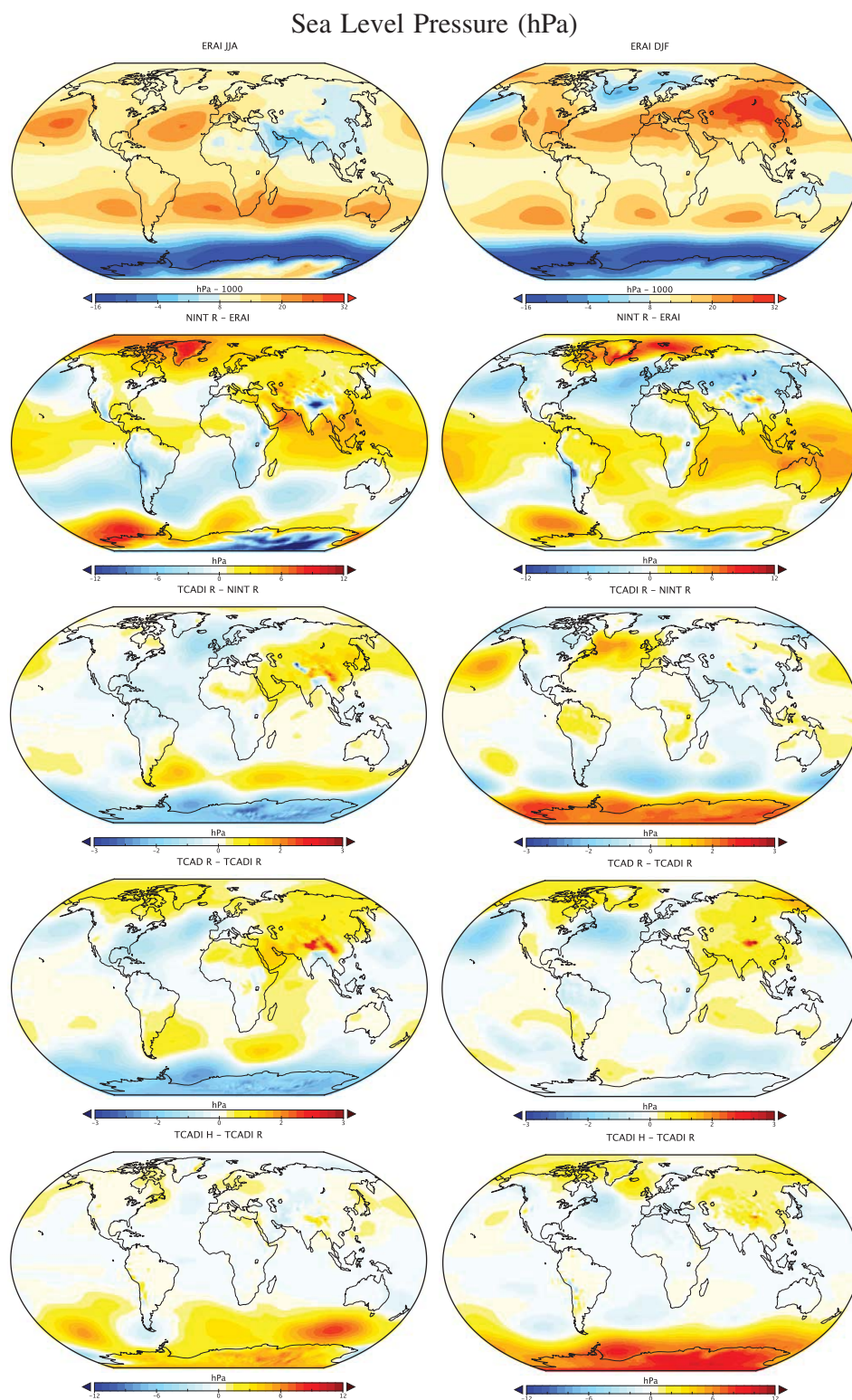## Microwave Sounding Unit Temperature — Lower-Stratosphere (°C)
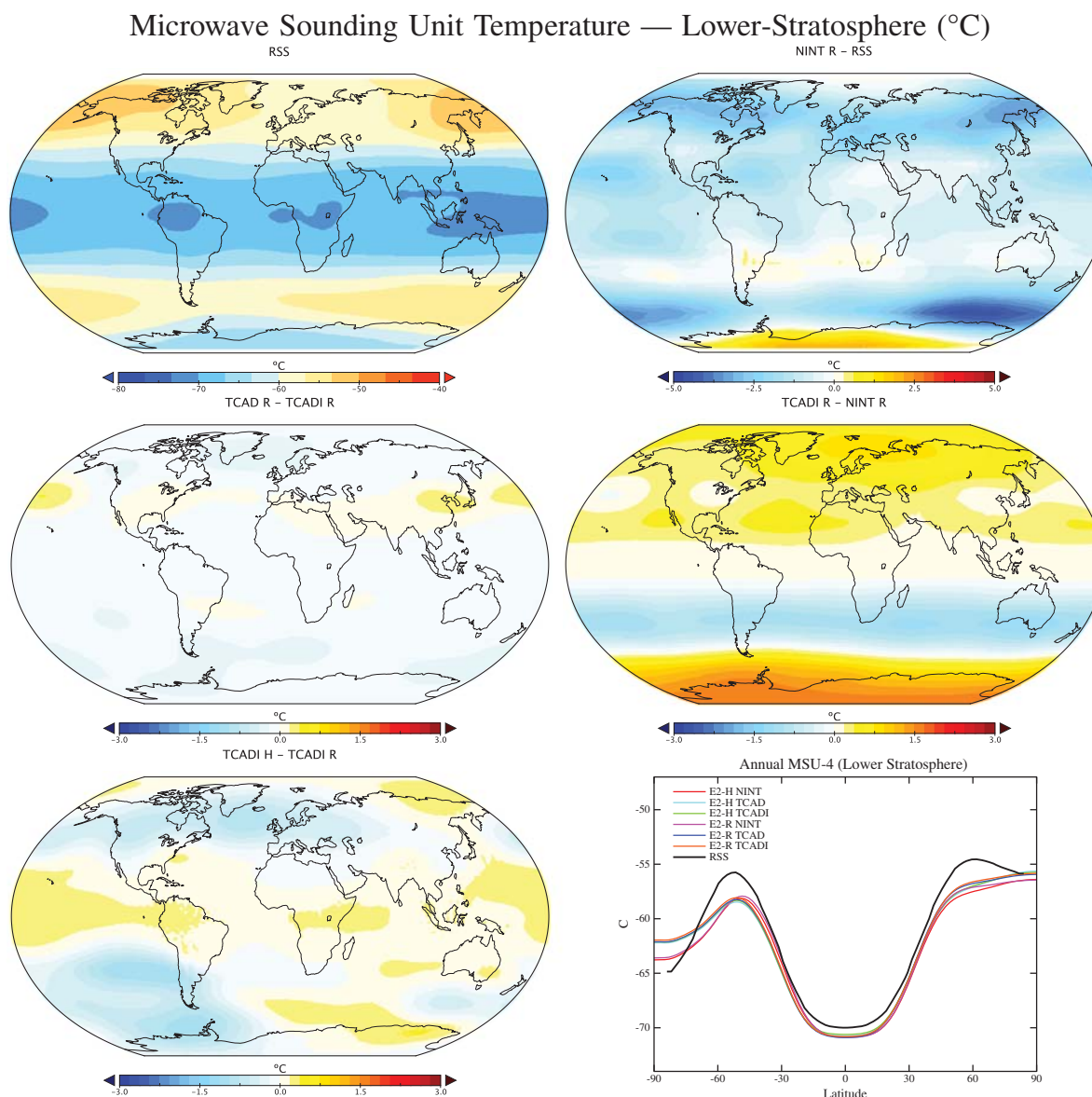
Figure 15. MSU Lower stratospheric temperatures compared to RSS. Plots as in Figure 2.

the open lake water area (i.e., the area of open water visible from space, not including any ice covered areas) to a satellite-derived data set of surface water [*Prigent et al.*, 2007; *Papa et al.*, 2010] are possible (Figure 18). Biases in the observed data set with respect to the model diagnostic are possible due to residual ambiguities related to vegetation cover and definition of "surface water." Overall, the model shows reasonable magnitudes of seasonality (60–30°N, 30–0°N) but the phasing is off by a month or two in the highest latitudes. There is an overestimate in the midlatitudes of the NH and an underestimate in the lower latitudes of both absolute amount and seasonality (60°N–30°N versus 30°N–0°N). Seasonal areas in the southern tropics (0°S–30°S) more closely follow observations, possibly because surface flooding in this region is primarily riverine.

Comparisons to lake databases of ice duration to the Global Lake and River Ice Phenology (GLRIP) database [*Benson and Magnuson*, 2000] in the Northern Hemisphere are greatly improved over results shown in SEA06, with reductions by about 20–50 days of iced-up conditions in the boreal regions, much closer to observed values, though still a little too long.

Sea Level Pressure (hPa)

Figure 16. Sea level pressure compared to ERA-Interim. Plots as in Figure 13.

Table 5. Annual Mean Discharge From Selected Rivers[a]

River	GISS-E2-R			GISS-E2-H			GISS-E	
	NINT	TCAD	TCADI	NINT	TCAD	TCADI	(NINT)	Obs.
Amazon	236	255	198	272	300	229	372	525
Congo	69	79	35	75	82	41	142	104
Brahmaputra-Ganges	75	68	86	119	110	140	118	81
Yangtze	93	85	100	200	191	210	265	74
Lena	33	32	34	30	29	31	33	42
St. Lawrence	54	55	53	28	28	27	35	37
Ob	52	50	47	88	89	80	28	32
Mackenzie	28	28	29	31	31	31	26	25

[a]All values are in $km^3\ month^{-1}$, observations from *Milliman and Meade* [1983].

3.9. Oceans

Ocean surface properties (temperature and salinity) are compared to compiled climatologies in Figures 19 and 20. For temperature, we use the World Ocean Atlas upper-most layer values [*Locarnini et al.*, 2010], while for salinity, we use the PHC 3.0 product which has more accurate salinities in the Arctic (updated from *Steele et al.* [2001]). Figure 19 shows that temperature biases are strongest on the eastern boundaries of the subtropical oceans, consistent with insufficient marine stratus clouds as inferred from Figures 5 and 7. The Southern Oceans are systematically warm in all model versions. Other notable errors are associated with a lack of Gulf Stream separation north of Cape Hatteras in the western North Atlantic. The impact of TCAD is small, while that of TCADI is larger, notably, increasing the overall bias in SST, except in the N. Atlantic.

As one might expect, the impact of the ocean model choice has a large effect (note the expanded color scale), though in comparison with the climatology (not shown) the basic patterns of the biases are similar everywhere except in the Southern Oceans, where the HYCOM warm biases are roughly twice as large. Root-mean-square (RMS) errors are 1.6°C in the Russell models (with little variation across atmospheric treatments) compared with 2.4°C in the HYCOMs, overall biases are also a little higher (1.5–1.7°C compared to 0.7–0.8°C).

The salinity maps Figure 20 demonstrate similarly related high-salinity biases in the marine stratus regions, but also highlight specific issues with the freshwater balance in many areas. Marginal seas have large biases—too fresh in Hudson Bay and the Black Sea, too saline in the Baltic and Red Sea. These biases arise through the difficulty in simulating the exchanges across straits unresolved by the standard grid. Other biases are associated with insufficient outflow from large tropical rivers- in the Arabian Sea, the Amazon, Gulf of Guinea and Rio Plata in South America, consistent with the direct river flow diagnostics in Table 5, and tropical land precipitation biases (Figure 9). The TCAD and TCADI models have systematic differences, but they are small compared to the overall biases. Again, HYCOM versions are strikingly different, with a 0.5–1.0 psu overall higher surface salinity than the Russell model versions. Both models have the largest biases in the Arctic, though in opposite directions (E2-H is too salty, E2-R, too fresh). While the mean salinity biases are larger in the HYCOMs (0.6 relative to 0.2 psu), the RMS error is slightly less (1.0 compared to 1.2 psu) (see also Table 7).

Climatological sea surface height patterns (not shown) are similar in the two models, though gradients are slightly steeper in E2-H. One interesting diagnostic of the whole model emergent behavior is the seasonality of ocean surface height. This is a function of net water mass transfers between ocean and land combined with the thermosteric effects of ocean temperature seasonality (there are also minor terms associated with sea ice growth and melt) and is only available with fully mass-conserving ocean models (like the Russell ocean). The sea level seasonal cycle can be inferred from the TOPEX and JASON altimeters and shows a near sinusoid with min to max amplitude of about 10 mm [*Nerem et al.*, 2010]. There is a significant asymmetry in the observations, with the minimum in March/April/May being less pronounced than the maximum peak in October. In the E2-R results, the phasing is accurate, but the seasonal amplitude is larger (by about 70%) and more symmetrical than observed (Figure 21). Differences in how the polar oceans are treated (not included in the observed data) might explain some of the difference, but the phasing and diagnostics of excessive precipitation suggest that the modeled ocean-to-land water mass flux is too large. Note that thermosteric effects alone have opposite phase, with a maximum in March.

Table 6. Selected Ocean Mass (Sv) and Heat (PW) Fluxes[a]

	GISS-E2-R			GISS-E2-H			GISS-ER	
	NINT	TCAD	TCADI	NINT	TCAD	TCADI	(NINT)	Obs.
North Atlantic MOC (Max)	27.2 ± 0.7	26.6 ± 0.5	24.5 ± 2.6	24.5 ± 0.8	25.2 ± 0.5	22.7 ± 0.6	26 ± 1	-
North Atlantic MOC (26°N)	18.4 ± 0.3	18.0 ± 0.4	17.2 ± 1	22.4 ± 0.6	23.1 ± 0.3	20.4 ± 0.5	19 ± 0.5	≈17 [Rayner et al., 2011] (estimate at 26°N)
Atlantic Heat 26°N	0.97 ± 0.01	0.96 ± 0.02	0.92 ± 0.07	0.99 ± 0.02	1.02 ± 0.01	0.99 ± 0.02	0.86	1.3 ± 0.4 [Johns et al., 2011]
ACC transport (Drakes Pass.)	254 ± 1	251 ± 1	254 ± 1	192 ± 2	192 ± 1	190 ± 1	225	≈130 [Petersen, 1988]
Gulf Stream	49 ± 1	47 ± 2	45 ± 2	39.8 ± 0.8	39.2 ± 0.8	41.0 ± 0.5	33 ± 1	≈35 [Rayner et al., 2011]
Kuroshio	64 ± 1	67 ± 2	64 ± 1	71.7 ± 0.5	72.1 ± 0.5	74.5 ± 1.4	43 ± 1	≈57 [Imawaki et al., 2001]
Bering Strait	0.16 ± 0.01	0.16 ± 0.01	0.16 ± 0.01	0.45 ± 0.01	0.46 ± 0.01	0.46 ± 0.01	0.05 ± 0.04	0.8 ± 0.2 [Woodgate et al., 2005]
Indonesian throughflow	11.5 ± 0.2	11.4 ± 0.2	11.7 ± 0.6	17.6 ± 0.3	18.5 ± 0.5	17.5 ± 0.5	6.2 ± 0.6	15 [Sprintall et al., 2009]

[a]Range is standard deviation of the 1980–2004 average from 5 ensemble members for each configuration.

Key diagnostics of the ocean circulation are given in Table 6. In the Atlantic, the Meridional Overturning Circulation (MOC) matches observations quite well (at the latitudes near 26°N where good repeat sections are available to constrain long term fluxes), giving mass fluxes of around 18–19 Sv. In both ocean models, the maximum overturning stream function value is higher (around 22–27 Sv) but this is not as well constrained from observations. Atlantic ocean northward heat fluxes are around 1 PW which is slightly less than observed (though see Figure 22 for the zonal context).

The barotropic mass fluxes carried by the two major NH western boundary currents are defined from the maximum in the horizontal mass stream function at around 30°N, prior to boundary separation, and in the Atlantic can be compared approximately to the mass flux through the Florida Strait combined with the wind-driven Ekman flux components [Rayner et al., 2011]. In the Pacific, observational estimates are drawn from altimeter-derived analyses [Imawaki et al., 2001]. Overall, the model results are comparable to estimates although a little high. Whether this is a definitional problem related to recirculations or symptomatic of model deficiencies requires further analysis.

The inter-ocean mass flux through the Bering Strait is small, but appears to be deficient in all simulations—more so in the E2-R simulations. The Indonesian throughflow is a challenge to model due to the small scale of the currents and the complexity of the topography, but the net fluxes are qualitatively reasonable—slightly too small in E2-R, slightly too large in E2-H.

In the Southern Ocean, there are more obvious systematic deficiencies associated with excessive transport in the Antarctic Circumpolar Current (up to a factor of 2 too large), and large SST and sea ice biases. These are coupled problems associated with structural deficiencies in mixing by unresolved eddies, insufficiently optically thick clouds and sea ice problems. They remain a challenge to model successfully in a fully coupled system.

Table 7. Arcsin-Mielke Scores for Selected Fields (See Figure 26 for Most Field Definitions)[a]

	Coupled Models								Atmos-Only
	GISS-E2-R			GISS-E2-H			GISS-ER	GISS-EH2	GISS-E2
Field	NINT	TCAD	TCADI	NINT	TCAD	TCADI	NINT	NINT	AVG
OTR	0.64	**0.64**	0.63	0.60	0.61	0.61	0.38	0.32	0.74
ASR	0.78	0.79	0.79	0.77	0.77	0.78	**0.80**	0.79	0.81
MSU-TMT	0.90	0.90	0.91	0.91	0.91	**0.91**	0.85	0.85	0.92
MSU-TLS	0.83	0.83	**0.84**	0.80	0.80	0.81	0.83	0.80	0.82
Tcld	0.17	0.17	0.17	0.13	0.13	0.14	**0.40**	0.35	0.25
Lcld	0.15	0.15	0.16	0.10	0.10	0.11	**0.40**	0.38	0.16
SLP (DJF)	**0.79**	0.78	0.79	0.72	0.71	0.70	0.70	0.49	0.82
SLP (JJA)	**0.80**	0.79	0.78	0.75	0.76	0.75	0.71	0.54	0.80
SAT (DJF)	**0.90**	0.89	0.89	0.87	0.86	0.86	0.89	0.87	(0.92)
SAT (JJA)	**0.89**	0.89	0.88	0.85	0.84	0.84	0.87	0.84	(0.92)
PRECIP	0.39	0.39	0.38	0.35	0.34	0.34	**0.45**	0.38	0.49
SST	**0.90**	0.90	0.90	0.84	0.83	0.83	-	-	-
SSS	**0.64**	0.64	0.64	0.55	0.55	0.55	-	-	-

[a]The highest score across the coupled models for each field is highlighted (taking account of precision not necessarily reflected in the table). The Atmos-only numbers are an average of the three AMIP versions which do not differ by more than 0.01. Note that the global SAT score is affected by the use of observed SST in the forcing in AMIP simulations.
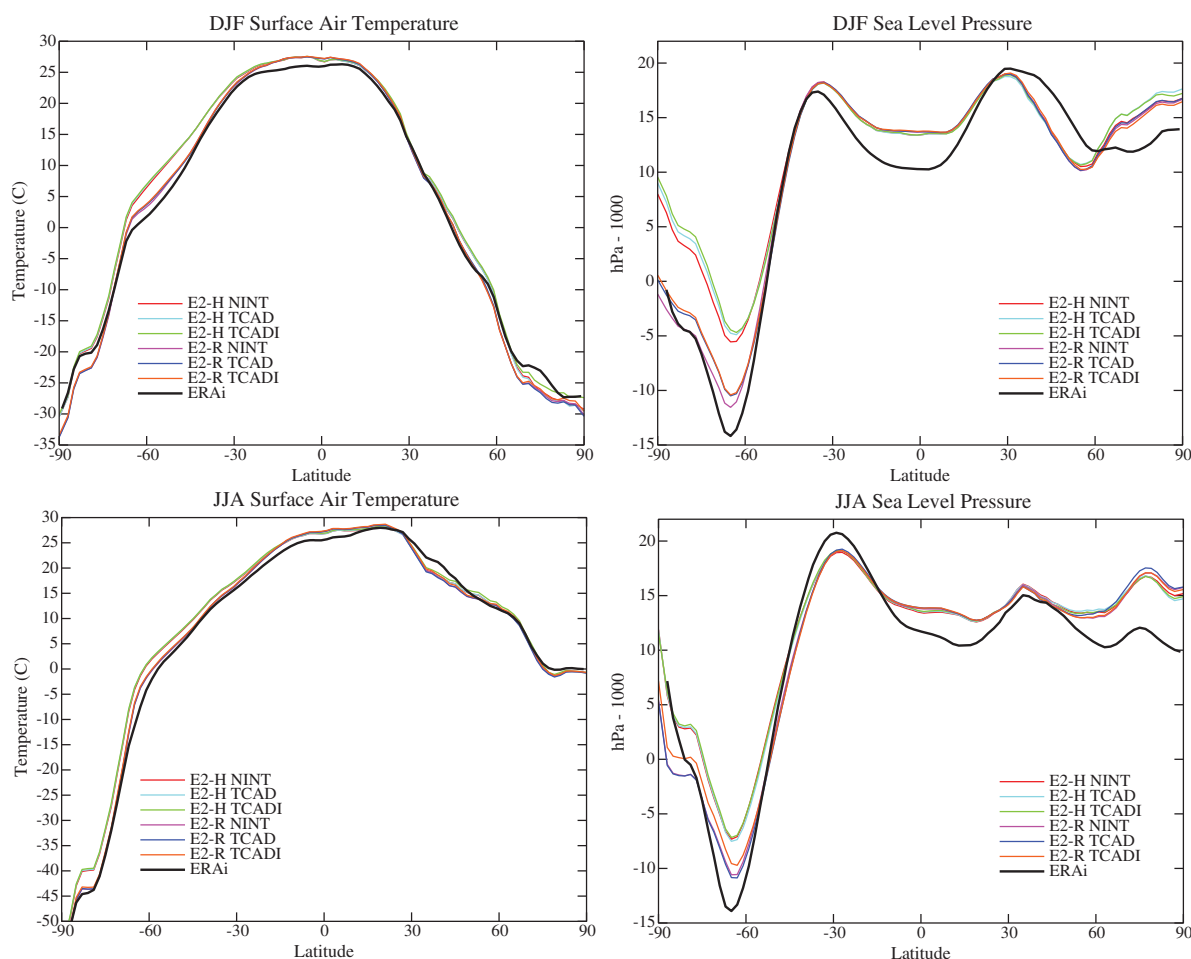
Figure 17. Zonal mean simulations of SAT and SLP for the DJF and JJA seasons compared to the ERA-Interim reanalysis.

Estimates of the implied poleward ocean heat transport can be determined from analyses of the atmospheric reanalyzes data [e.g., *Trenberth and Caron*, 2001] and ocean inverse calculations [*Ganachaud and Wunsch*, 2003]. Globally these can be compared to the total heat fluxes in the ocean models (incorporating the resolved flow and parameterized eddy heat fluxes). Variation among the simulations with the same ocean model is barely distinguishable (Figure 22). Differences to observations are clear in the peak Northern Hemisphere value, which is too small, deficient Southern Hemisphere subtropical flux and perhaps excessive poleward heat flux in the Southern oceans in E2-R. In the Atlantic, specifically the heat flux at the latitude of the RAPID array (26°N) (Table 6) is around 25% too low, though (just) within the uncertainty of the observations [*Johns et al.*, 2011].

3.10. Sea Ice

Compared to the simulations in *Hansen et al.* [2007], the sea ice climatology and seasonality is improved in all versions of the model (Figure 23). However, the models overestimate seasonality in the Arctic with a consistent September minimum that underpredicts (current) conditions. Antarctic ice is consistently low, especially in the E2-H simulations which is consistent with the warm biases seen in the SST in Figure 19. Impacts of different atmospheric composition treatments are small and inconsistent. In all models, minimum Antarctic ice in March is very close to zero, some 2 million km^2 too low (retaining summer ice only on the western edges of the Weddell and Ross Seas). During winter, warm SSTs and insufficient ice cover are associated with excessive vertical mixing of heat from below which is exacerbated during the summer when excessive
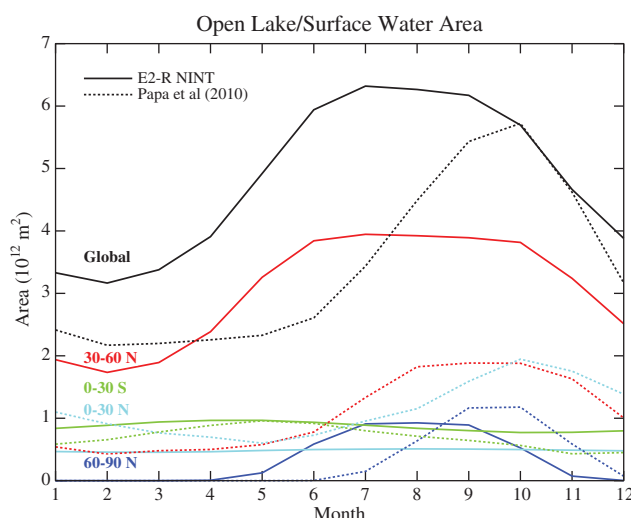
Figure 18. Seasonal cycle of surface water detected by satellite [*Prigent et al.*, 2007; *Papa et al.*, 2010] compared to open (nonice covered) lake in the E2-R NINT simulation. Each color refers to a distinct latitude band, with the observations (dashed) and model (solid). Seasonality is driven mainly by ice cover in the mid to high latitudes, and riverine inundation in the tropics and subtropics.

solar radiation from insufficient Southern Ocean clouds warms the mixed layer. Further investigation has isolated implementation issues in the mesoscale ocean eddy diffusion parameterization in E2-R and the reference density in E2-H as being important in controlling the Antarctic sea ice bias.

In the Arctic, the principal difference across the models is between the E2-R and E2-H versions; the former has substantially thicker ice in the Arctic winter, though only slightly larger ice area (Figure 24). This difference is associated with a difference in ice albedo (which is a function of snow thickness) which is about 0.1 too low throughout the summer season (compared to data from the Surface Heat Budget of the Arctic Ocean (SHEBA) project [*Perovich et al.*, 2002]) in the E2-R runs. We have subsequently found that this is due to two effects; too extensive melt-pond formation developing in the summer (around 10% more area than observed), and excessive (erroneous) rates of snow-to-ice conversion in dynamically convergent situations in E2-R. The formulation in E2-H had the ocean model allocating new ice formation across open water and ice covered areas uniformly which did not trigger the problem.

Ice velocities in all simulations (Figure 24e) show a clear Beaufort gyre and transpolar drift stream in the Arctic, with magnitudes up to 0.3 m/s in the Fram Strait, but with an average speed in the main basin between 0.05 and 0.1 m/s, a little faster than observed values, likely a function of insufficient thickness. Note that there are a small number of ocean boxes adjacent to Antarctica and Greenland where "landfast" sea ice is abnormally large in the HYCOMs. This was due to an error in the regridding interfering with the expected advection of thick ice away from the coast.

3.11. Land Ice

Net accumulation of snow over glaciated regions is similar in all model versions, around 700 Gt/yr in the Northern Hemisphere (Greenland and other small ice caps), and 1800 Gt/yr in the Antarctica. This is a significant reduction compared to E-R (a factor of two less in Antarctica), and much closer to observational estimates for Greenland and Antarctica (540 Gt/yr and 1800–2100 Gt/yr, respectively [*van de Berg et al.*, 2006; *Box et al.*, 2006]), consistent with more realistic values for total snowfall from Table 4. This is probably due to higher resolution in the models, reducing the amount of unphysical transport of water vapor to higher latitudes.

4. Variability

4.1. Annular Modes

The principal empirical orthogonal functions (EOF) for winter season monthly sea level pressure (SLP) in each hemisphere are often described as the Northern and Southern Annular Modes (NAM, and SAM, respectively) [*Miller et al.*, 2006a]. In observations, the first EOF explains 20% of the NH monthly variance for October-March, while in the models the same basic pattern explains between 21 and 24% with no significant differences between models, and almost identical monthly variance in total ($\approx$11 hPa). In the SH, observations show a more dominant EOF1 (explaining 33% of the variance Nov-Feb). While overall variance in SLP is similar in models and observations (around 7 hPa), the percentage explained by EOF1 is larger

Sea Surface Temperature (°C)
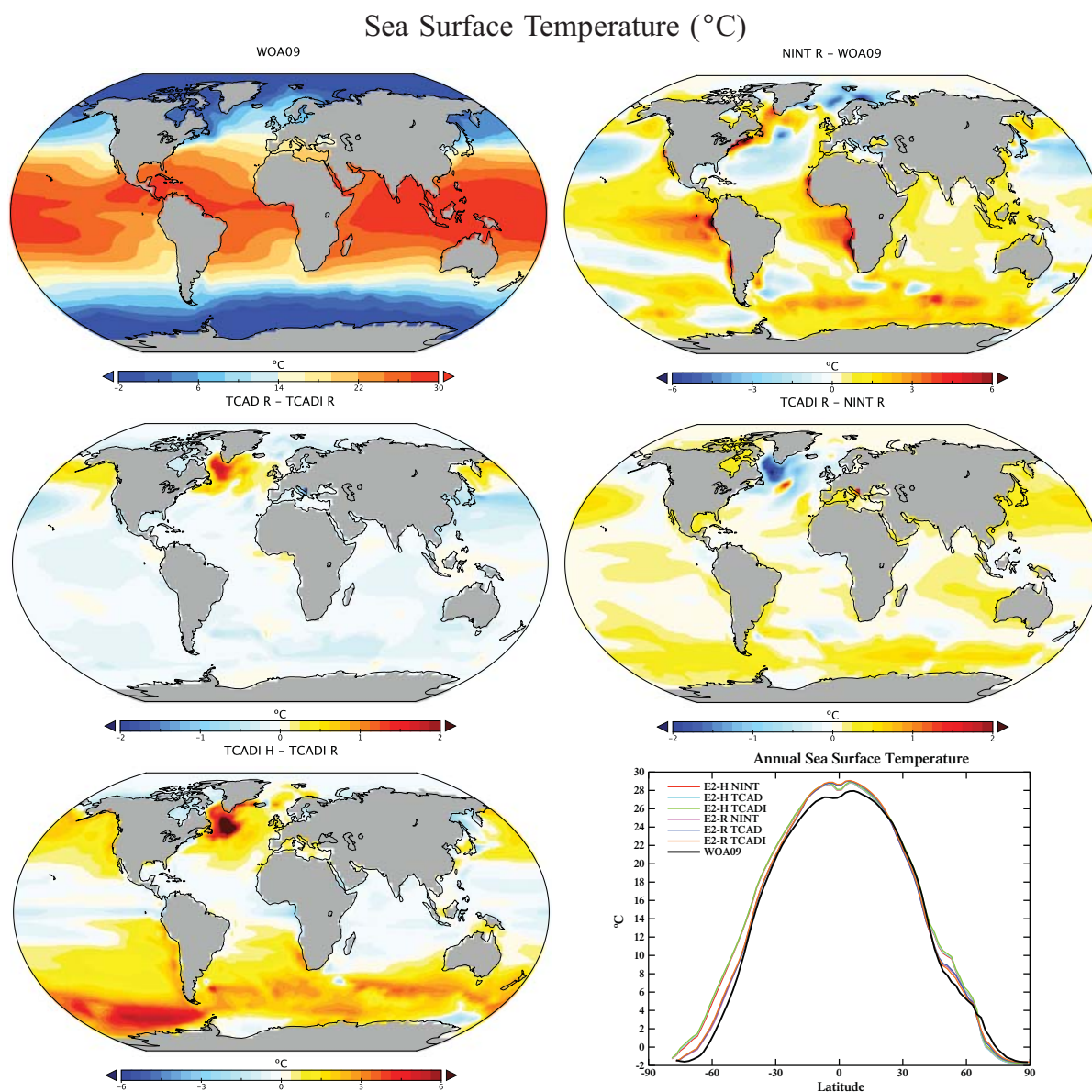
Figure 19. Sea surface temperatures compared to WOA09. Plots as in Figure 2.

between 39 and 46%. This may be connected to overall biases in the climatological SH jet stream that is endemic to many GCMs [*Simpson et al.*, 2013].

4.2. Interannual Tropical Variability

Compared to the CMIP3 GISS models, tropical variability is greatly enhanced, principally related to the increase in equatorial ocean resolution (particularly in E2-R). Variations as a function of atmospheric composition treatments are negligible. Between 1980 and 2004, the ensemble mean of the standard deviation of sea surface temperature in the Niño 3.4 region (between 150°W–90°W and 5°S–5°N) is 0.57 ± 0.07 K and 0.72 ± 0.06 K for the NINT E2-R and E2-H models, respectively. These values can be compared to the observed standard deviation of 0.93 K during this period, based upon the Extended Reynolds SST reconstruction [*Smith et al.*, 2008]. Unlike the observations which have a skew toward cold events, there is a near symmetry between cold and warm events in both model versions. The spectra of the NINO 3.4 index (Figure 25a) shows maximum power over a range of periods centered between 2 and 3 years in both ocean models
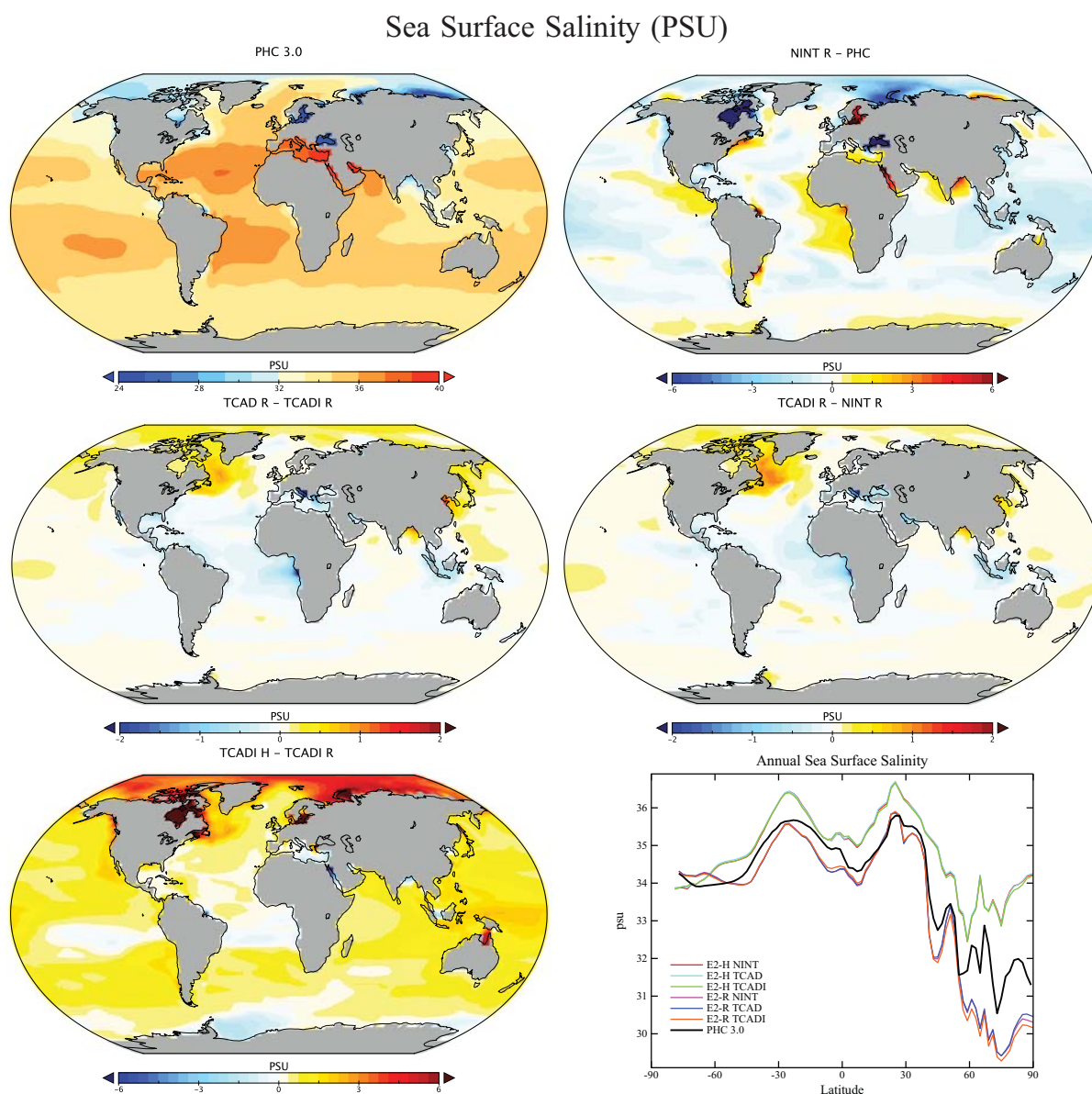
Figure 20. Sea surface salinity compared to PHC version 3. Plots as in Figure 2.

with higher spectral density in E2-H. Both ocean models underestimate the importance of the 3–7 year band relative to the real world.

4.3. North Atlantic Ocean Variability

The variations of the meridional overturning circulation (MOC) in the N. Atlantic are an important factor in decadal to multidecadal time periods, and this can be quantified from the control runs for each model configuration. There is some significant nonstationarity to the statistics of the MOC over the millennia in the control simulations, and so it is impossible to exactly specify the spectral characteristics. We analyze the maximum MOC stream function for a 1000 year period for each model and remove very low-frequency variability (>250 years) to minimize the impact of drifts. The spectral characteristics (Figure 25b) vary dramatically as a function of ocean model, but not obviously as a function of atmospheric treatment. The E2-R models show significant variability (annual standard deviation ≈1.3 Sv) in the MOC centered on 6–8 year period and with a broader peak over the 20–50 year period. In the E2-H simulations, MOC variation is larger

**Seasonal Cycle of Eustatic Sea Level**
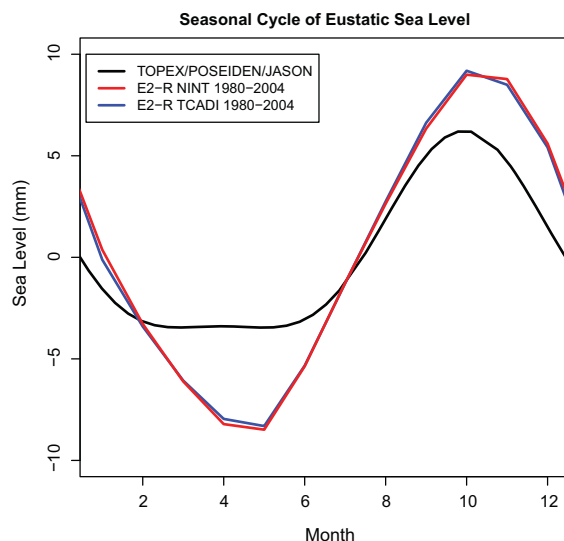
Figure 21. Seasonal cycle of eustatic sea level in the E2-R model in comparison with the inferred seasonal cycle from satellite altimeters [*Nerem et al.*, 2010]. This metric includes thermosteric and halosteric effects on seawater as a function of the seasonal cycle, but also the transfer of freshwater mass between the ocean, land, and ice sheets. The observed cycle is only for the ocean area 66°S–66°N, while the model diagnostic is fully global.

(with a annual standard deviation of ≈3.1 Sv), and with a broad spectral peak around the 10–20 year period.

5. Climate Sensitivity

The climate sensitivity [*Charney et al.*, 1979] for each model configuration is calculated using the *q*-flux model (with a maximum mixed layer depth of 65 m to reduce computation time) to estimate the climate response to $2 \times CO_2$. The NINT, TCAD and TCADI ModelE2 versions have sensitivities of 2.7°C, 2.7°C, and 2.9°C, respectively. Coincidentally, the ModelE sensitivity in the same configuration was also 2.7°C. Estimates of the coupled-ocean model sensitivity (which additionally allows for impacts of changes in ocean and sea ice dynamics) can be estimated from the transient behavior of the $4 \times CO_2$ instant forcing simulations [*Gregory et al.*, 2004]. Offline calculations using the preindustrial control climate give that the radiative forcing after fast stratospheric adjustment for $2 \times CO_2$ and $4 \times CO_2$ (i.e., $\mathbf{F}_a$ [*Hansen et al.*, 2005]) are 4.1 and 7.9 W/m^2, respectively. Scaling the long-term response to $4 \times CO_2$ to the $2 \times CO_2$ forcing, the estimated sensitivities are 2.3°C, 2.3°C, and 2.4°C and 2.5°C,

Ocean Heat Transports
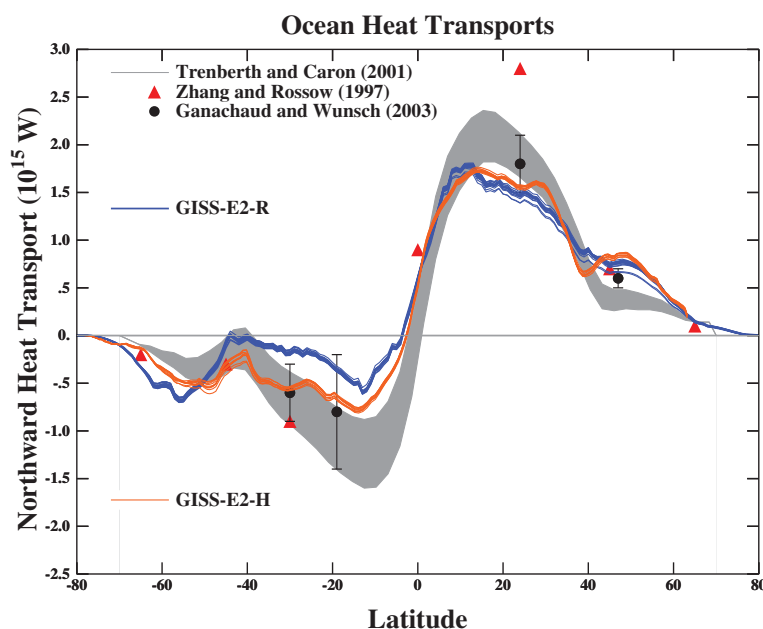
Figure 22. Annual mean poleward ocean heat transports. Comparisons of the models with residual calculations (with error bars) from the NCEP reanalysis [*Trenberth and Caron*, 2001], from the ISCCP remotely sensed fluxes [*Zhang and Rossow*, 1997], and from ocean inverse calculations [*Ganachaud and Wunsch*, 2003]. Each model simulation is plotted as an individual line and only the differences between ocean models are noted.
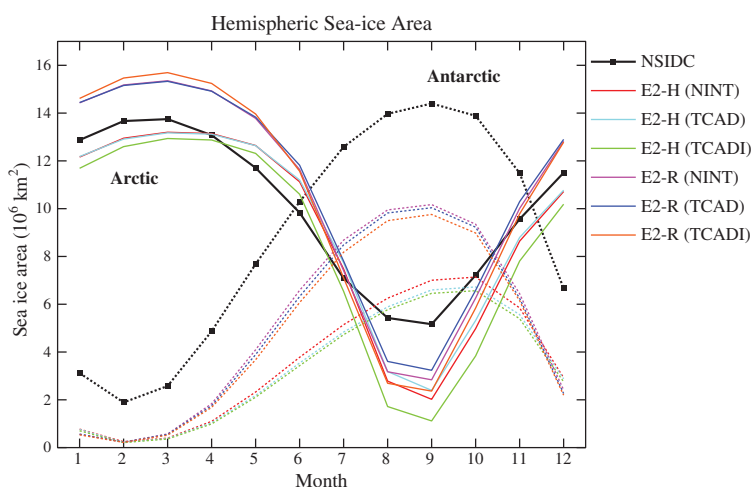
Figure 23. Annual climatology of sea ice area in both hemispheres. Observational data comes from NSIDC (1980–2004), after correction for the Arctic polar "hole" [*Fetterer et al.*, 2011]. The ensemble mean climatology is plotted for each model version (spread across ensemble members is small). Solid lines are for the Arctic, dashed for the Antarctic.

2.4°C, and 2.5°C, respectively for the three physics-versions and the two ocean models (R and H). There is a consistently slightly higher sensitivity for the TCADI runs (attributable to an increase in cloud feedbacks with changes in aerosols), and an increase in sensitivity when using the HYCOM, related either to the slightly different base climate or a different pattern of SST change, though the overall differences are small. All calculations include a sensitivity in the stomatal conductance to CO_2 level which causes reduces latent heat fluxes over land and relatively higher land/ocean contrast in surface temperature response, though the global mean impact is small ($<0.1°C$).

The TCAD/TCADI simulations for the future include a feedback in CH_4 (via wetlands) that would be an additional feedback to those considered in the q-flux models mentioned above. Given the sensitivity in the future scenarios of about 130 ppb $CH_4/°C$ of warming (L. Nazarenko et al., submitted manuscript, 2013), that translates to an additional radiative effect of 0.05 $W/m^2/°C$ and so an additional warming effect of about 0.1°C. Thus including this additional CH_4 as a feedback would lead to a q-flux sensitivity of 2.8°C and 3.0°C, respectively.

Using the methodology of *Andrews et al.* [2012] (who use a linear fit to the first 150 years of the instantaneous $4×CO_2$ results), the $4×CO_2$ effective forcings are diagnosed as 7.5, 7.0, and 7.2 W/m^2 (E2-R) and 7.7, 7.5, 7.4 W/m^2 (E2-H)—all significantly less than the stratospherically adjusted radiative forcing. The implied effective sensitivities are 2.1, 2.2, and 2.4°C and 2.3, 2.3, and 2.5°C, respectively, up to 0.3°C lower than the actual long-term sensitivity. The reason for these offsets are that the relationship between radiative imbalance and temperature is not perfectly linear in these models (it is slightly convex), and thus a linear fit underestimates the intercepts on both axes. This is related to a time dependence in effective sensitivity [e.g., *Williams et al.*, 2008; *Armour et al.*, 2013].

## 6. Comparison to Previous GISS Models

As in SEA06, we use a selection of well-observed metrics to assess the overall skill of the models in capturing the climatological coherence and spatial distribution. We use both Taylor diagrams [*Taylor*, 2001] and the Arcsin-Mielke (AM) score [*Watterson*, 1996] which additionally takes into account the mean bias. A perfect AM score (unity) corresponds to the observational point, while a score of zero corresponds to the (vertical) zero-correlation line.

Figure 26 shows comparisons among the selected models for the DJF and JJA ERA-Interim SAT, GPCP precipitation (60°S–60°N), RSS MSU-TLS and TMT, total cloud and low cloud amounts (calculated using the ISCCP simulator—60°S–60°N), the CERES TOA LW and SW fluxes, and the DJF/JJA oceanic sea level pressures (ERA-Interim). In each plot, different colors refer to different fields, while the symbols denote the six
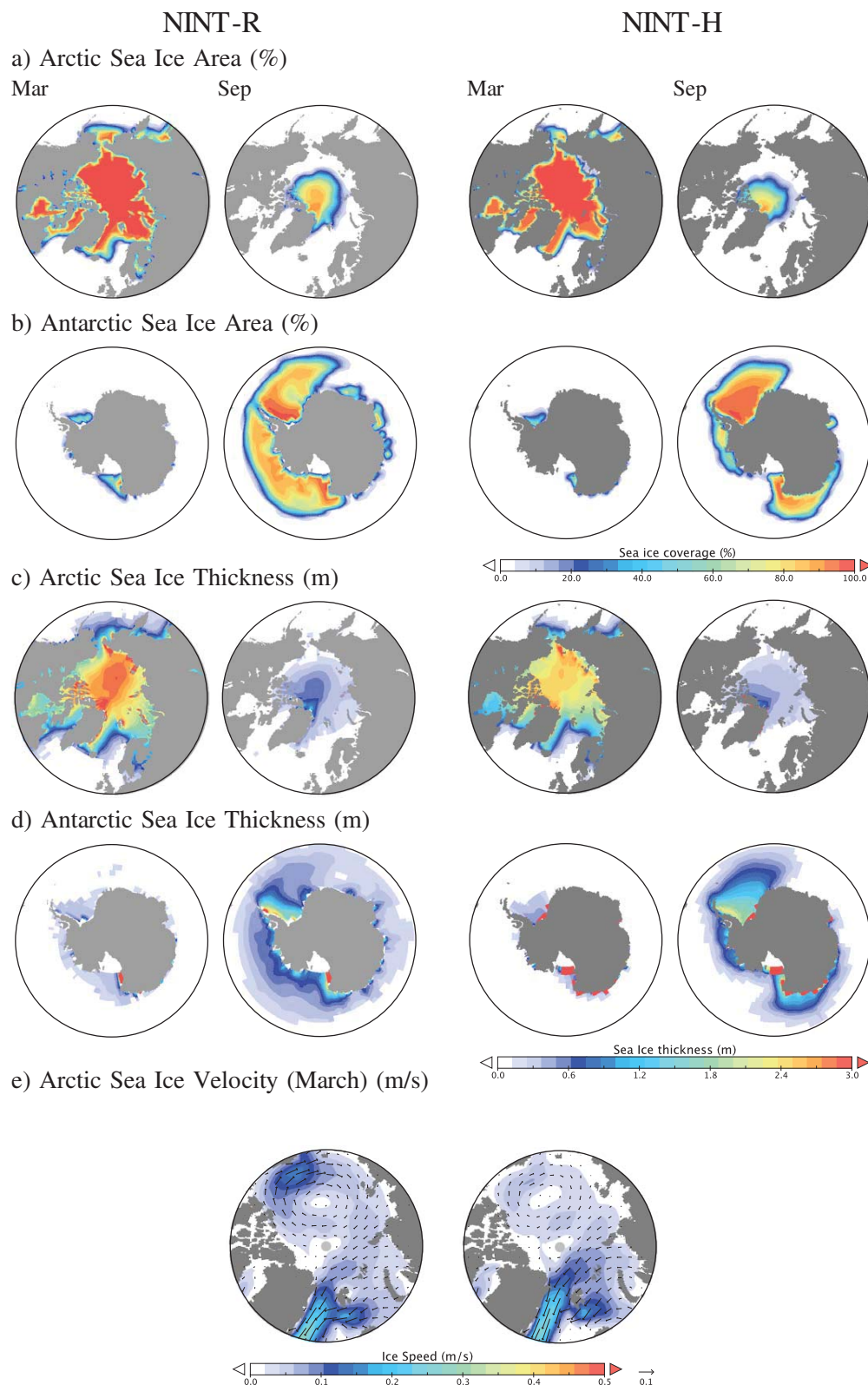
Figure 24. Sea ice diagnostics for March and September in the NINT-R and NINT-H simulations (columns). (rows a and b) Arctic and Antarctic sea ice area (%); (rows c and d) Arctic and Antarctic sea ice thickness (m); (row e) Arctic sea ice velocities and speed (m/s).
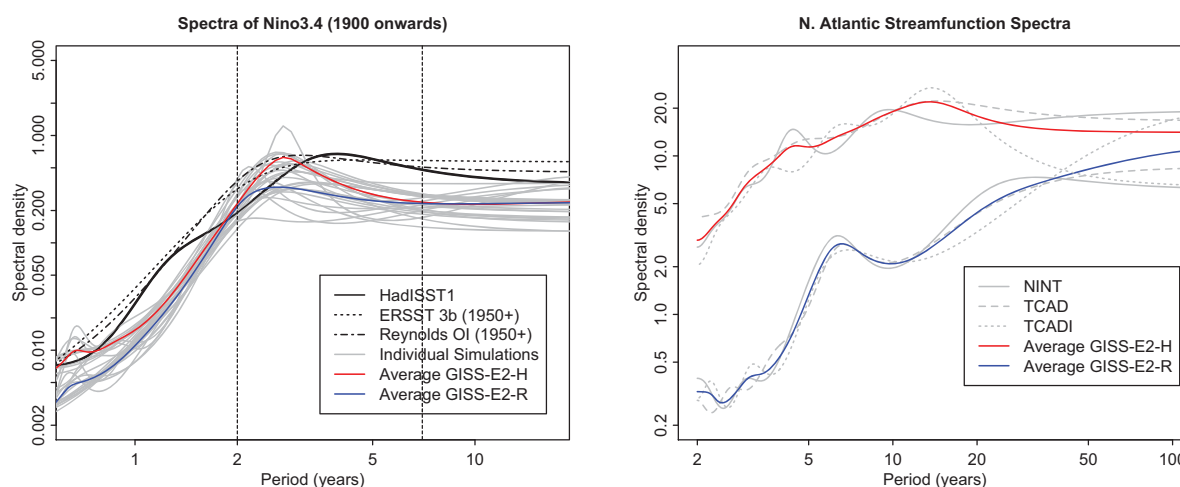
Figure 25. Spectra of ocean variability: (a) Tropical Pacific NINO 3.4 region: Tropical Pacific SST is deseasonalized and the spectra estimated using a maximum entropy method (30 poles) using model output in the historical simulations (1900–2005). Observational spectra are estimated from the HadISST1 data set [*Rayner et al.*, 2003], ERSST v3b (1950–2012) [*Smith et al.*, 2008] and the Reynolds OI data (1950–2012) [*Reynolds et al.*, 2002]; (b) North Atlantic Stream function spectra from 1000 year control simulations with the lowest frequencies (>250 years) removed using a Lowess fit. Gray lines are spectra from individual simulations, while red and blue lines are the average spectra for all simulations using the same ocean model.

model simulations presented here, the average of three AMIP simulations and the two previous CMIP3 versions (GISS-ER and GISS-EH2, which both have NINT physics).

The Taylor diagrams provide a good insight into the main conclusions from this study. Differences in climatology between the models as a function of atmospheric interactivity are small (and only in the case of precipitation is there a noticeable difference between NINT and TCAD/TCADI simulations). Differences as a function of ocean model are more noticeable in all cases, although the variation is relatively minor, with the E2-R models performing slightly better overall in climatology, though not in variability. Compared to the CMIP3 versions though, the differences are stark: for SLP, SAT, and the MSU diagnostics, the skill scores are noticeably improved. However, for precipitation or cloud diagnostics, they are substantially worse. Further analysis (not shown) indicates that the poor relative performance in clouds is almost all due to the Southern Hemisphere (as is clear from Figure 4 as well). The impression gained from the Taylor plots is mirrored in the AM score (Table 7) (which also assesses mean bias): very similar scores for the atmospheric versions, slightly different scores for the ocean model variations, and a large contrast with the CMIP3 model versions.

The role of SST biases in the ocean models can be assessed by comparison with the AMIP simulation results. There are substantially higher correlations to observations for the AMIP precipitation and outgoing longwave. Cloud biases are affected by the ocean errors, but via a reduction in the spatial variation, rather than through a large difference in correlation. Overall skill in sea level pressure and temperature diagnostics are not strongly affected.

7. Discussion and Future Directions

The importance of large-scale coordinated model experiments such as CMIP5 lies mainly with the opportunity to assess the robustness and range of model simulations across a diverse set of modeling approaches and implementations. Different model groups from across the world—each with their own history, assumptions, errors, foci and experience—produce a suite of simulations that allow for many tests of model structural uncertainty. However, this multimodel ensemble is an "ensemble of opportunity" that has not been designed with any specific goal of completeness or with the attribution of model differences as a high priority.

Outside of CMIP, the increasing number of perturbed physics ensembles (PPEs) in GCMs [e.g., *Stainforth et al.*, 2005; *Shiogama et al.*, 2012] or via emulators [*Lee et al.*, 2011] is beginning to explore another dimension of structural uncertainty within specific models. Comparisons of the two approaches have demonstrated that the methodologies are complementary [*Yokohata et al.*, 2013] in that a single-model PPE does not generally span the phase space of the MME, but does provide a more clearly process-based

understanding of model differences. It is thus increasingly clear that a merged approach will be required to better explore model structural diversity.

Some authors have stressed the difference between model variations accessible via a change in existing but uncertain parameter values in particular parameterizations ("parameter uncertainty") and variations in the model structure itself, i.e., using one parameterization over another, changing the scope of the model calculation, switching model components, etc. ("structural uncertainty") [e.g., *Rougier*, 2013]. However, in theory all such structural perturbations are accessible via a suitably designed parameterization (despite this rarely being done in practice), so the distinction is a little arbitrary. It is nonetheless true that most existing PPEs have only sampled the parameter space within a set of fixed parameterizations.

The GISS contribution to CMIP5 is an attempt to add some further dimensions to the discussion of structural uncertainty. The changes to the interactivity in atmospheric composition, aerosol indirect effects and ocean models are structural changes that go beyond what has yet been tested in PPEs. The advantage of this over the much large structural variation across other groups in CMIP5 is that many aspects of the model versions are identical (i.e., dynamical core, radiation, moist convection, etc.), allowing for a clearer association between changes and responses. Other groups have made analogous contributions, though this has not been a coordinated activity [e.g., *Hallberg et al.*, 2013]. Changes in scope can however make comparisons problematic. For instance, there are inevitable differences in effective forcings between NINT and TCAD. The former is driven by concentrations of ozone and aerosols derived from the previous generation of interactive composition models using inventories of precursors that have subsequently been amended, while TCAD is driven by up-to-date emissions and updated composition modeling as well as being interactive with the simulated climate change. Thus baselines and trends in ozone, sulfates, nitrates, etc. are different in the two versions. Differences between TCAD and TCADI are easier to interpret (since the code and emission data sets are the same), with only the indirect effects of aerosols on clouds varying (though since that impacts the base climatology, composition will differ slightly too).

This paper summarizes a wide range but partial suite of available diagnostics from the climate models related to some common available observational products for a 25 year climatology. Differences between ensemble members for any specific configuration are small for these averaged diagnostics. Diagnostics for the historical period (time series, responses to events) are more sensitive to both initial conditions and to structural variations and are discussed in R. L. Miller et al. (submitted manuscript, 2013). An even wider range of diagnostics based on specific processes, or sensitivity over time, is being published as part of the community focus on the CMIP5 archive.

The results presented here show that the CMIP5 versions of the coupled model are improved in some key respects over the results in CMIP3—in particular, the increased resolution has lead to improvements in sea level pressure, surface temperatures, many land surface diagnostics, sea ice and better ENSO variability. Specific improvements in physics (and composition modeling) have improved stratospheric circulation and temperatures. Some degradations are seen in precipitation and cloud metrics, and these are currently the focus of renewed development effort (e.g., as described in *Kim et al.* [2012]).

We have not shied away for showing metrics where model-data mismatches are poor, nor have we gone to great lengths to find matches that are good. Most of the quantities here are simply standard outputs. It is striking to what extent biases are common to all the variations examined here. Systematic errors in cloud amount, or sea surface temperatures are not greatly affected by the interactivity of the composition, nor by the treatment of the indirect aerosol effect. This is not true for all variables, with TCADI simulations showing significant differences in clouds, column water vapor and in the stratosphere (mainly via interactive ozone). Unsurprisingly, climatological differences associated with different ocean models are very significant, particularly in the tropical ocean SST biases and variability and for the polar sea ice where there are important differences in feedbacks that relate to the baseline climatology.

The impact of the structural changes on the equilibrium climate sensitivity is clearer and varies systematically across the models, though within a relatively narrow range, with the (most interactive) TCADI versions showing a slightly higher sensitivity. Differences in internal variability are apparent, but not readily attributable to any specific difference and require further investigation. There are larger and more significant differences between configurations visible in time series over the historical period (see R. L. Miller et al., submitted manuscript, 2013) and in future simulations (L. Nazarenko et al., submitted manuscript, 2013).
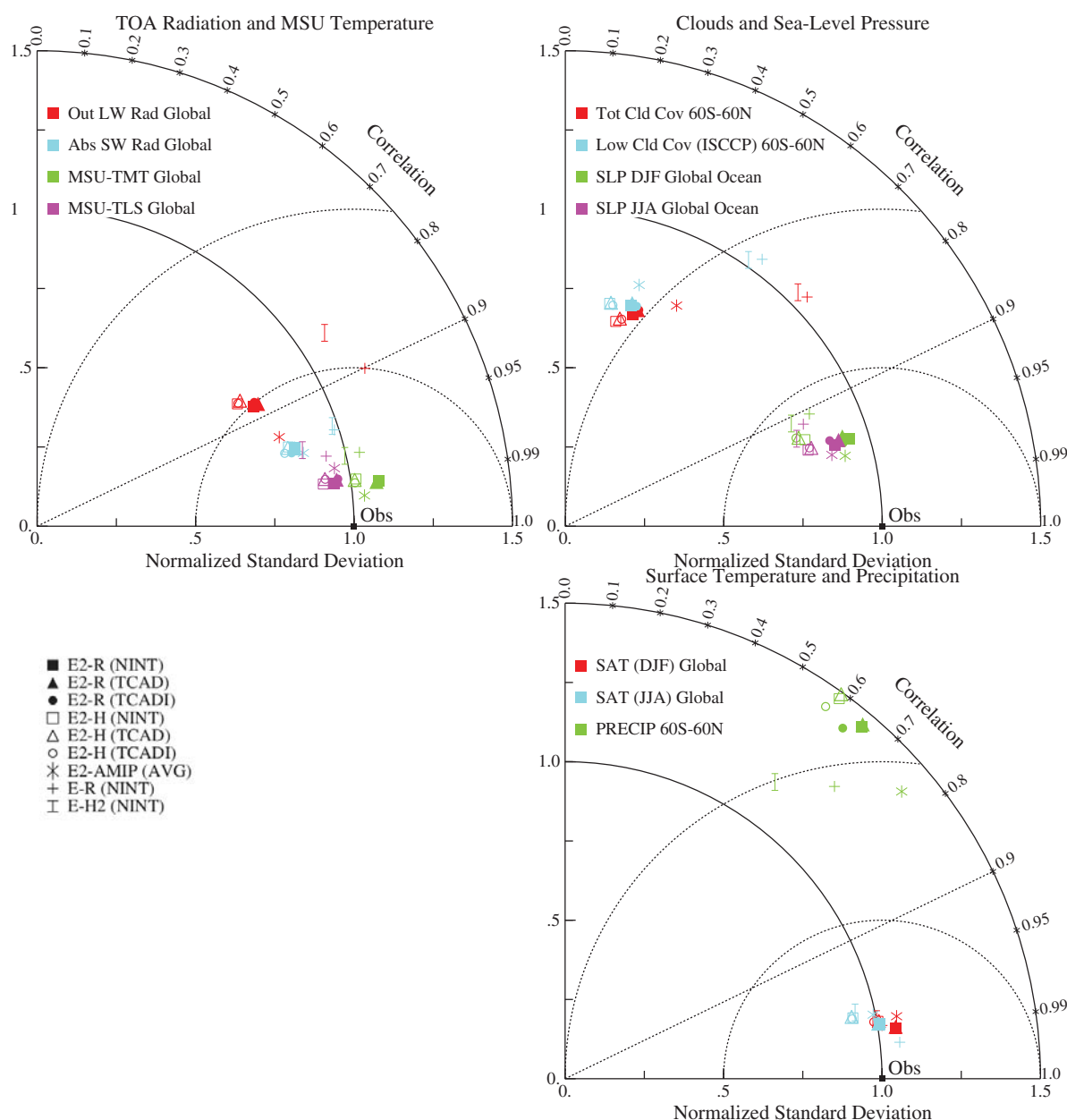
Figure 26. Taylor diagrams for selected quantities showing the difference in performance for different model configurations. Different colors are for different fields, while different symbols are used for each model configuration. (a) CERES: Outgoing Longwave Radiation and Absorbed Solar Radiation (60°S–60°N); RSS MSU-TMT and MSU-TLS; (b) ISCCP Total Cloud Cover, Low Cloud Cover (60°S–60°N) and oceanic Sea Level Pressure (DJF and JJA); (c) ERA-Interim SAT (DJF and JJA) and GPCP Precipitation.

This highlights the fact that model skill in matching climatologies (similar in these cases) is not necessarily directly correlated to sensitivity (which is quite different).

Analyses of a wide suite of model experiments and diagnostics are useful for setting the stage for future development and for the assessment of the utility of models to answer key questions. Our experience in developing the model for CMIP5 purposes has revealed important sensitivities, and new targets for model development—for instance, improvements in ocean mixing, sea ice parameterizations, and moist convection are now being tested in response to the model-observation mismatches described here and to the inevitable discovery and correction of coding errors over time. The extent to which models improve over time is however dependent on the specific diagnostics being examined, some of which have benefited

from greater resolution, improved parameterizations, or perhaps simply greater attention during the development process.

Further development will continue with the TCADI model versions, while the functionality of the NINT version will be retained. The results from TCAD are not distinctive enough to pursue further, though the need for an intermediate version between NINT and TCADI has enabled us to evaluate the impacts of composition interactivity and AIE separately. The large sensitivity to the ocean model both in climatology, variability and sensitivity (as seen also in *Hallberg et al*. [2013]) suggests that more exploration of this structural uncertainty is warranted, for instance, via coupled PPEs (as in *Shiogama et al*. [2012]).

References

Abramopoulos, F., C. Rosenzweig, and B. Choudhury (1988), Improved ground hydrology calculations for Global Climate Models (GCMs): Soil water movement and evapotranspiration, *J. Clim.*, *1*, 921–941.

Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor (2012), Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models, *Geophys. Res. Lett.*, *39*, L09712, doi:10.1029/2012GL051607.

Armour, K. C., C. M. Bitz, and G. H. Roe (2013), Time-varying climate sensitivity from regional feedbacks, *J. Clim.*, *26*, 4518–4534, doi: 10.1175/JCLI-D-12-00544.1.

Ball, J. T., I. E. Woodrow, and J. A. Berry (1987), A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions, in *Progress in Photosynthesis Research*, vol. *4*, edited by J. Biggins, pp. 110–112, Nijhoff, Dordrecht, Netherlands.

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

Bauer, S. E., D. Koch, N. Unger, S. M. Metzger, D. T. Shindell, and D. G. Streets (2007a), Nitrate aerosols today and in 2030: Importance relative to other aerosol species and tropospheric ozone, *Atmos. Chem. Phys.*, *7*, 5043–5059, doi:10.5194/acp-7-5043-2007.

Bauer, S. E., M. I. Mishchenko, A. A. Lacis, S. Zhang, J. Perlwitz, and S. M. Metzger (2007b), 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., D. L. Wright, D. Koch, E. R. Lewis, R. McGraw, L.-S. Chang, S. E. Schwartz, and R. Ruedy (2008), MATRIX (Multiconfiguration Aerosol TRacker of mIXing state): An aerosol microphysical module for global atmospheric models, *Atmos. Chem. Phys.*, *8*, 6003–6035.

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

Bell, N., D. Koch, and D. 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.

Benson, B., and J. Magnuson (2000), Global lake and river ice phenology database, in *Digital Media*, Natl. Snow and Ice Data Cent./World Data Cent. for Glaciol., Boulder, Colo.

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

Bian, H., and M. Prather (2002), Fast-J2: Accurate simulations of photolysis in global climate models, *J. Atmos. Chem.*, *41*, 281–296.

Bian, H., M. J. Prather, and T. Takemura (2003), Tropospheric aerosol impacts on trace gas budgets through photolysis, *J. Geophys. Res.*, *108*(D8), 4242, doi:10.1029/2002JD002743.

Bitz, C. M., and W. H. Lipscomb (1999), An energy-conserving thermodynamic model of sea ice, *J. Geophys. Res.*, *104*, 15,669–15,677.

Box, J. E., D. H. Bromwich, B. A. Veenhuis, L.-S. Bai, J. Stroeve, J. C. Rogersey, K. Steffen, T. Haran, and S.-W. Wang (2006), Greenland ice sheet surface mass balance variability (1988–2004) from calibrated polar MM5 output, *J. Clim.*, *19*, 2783–2800, doi:10.1175/JCLI3738.1.

Butkovskaya, N., A. Kukui, and G. Le Bras (2007), HNO3 forming channel of the $HO_2 + NO$ reaction as a function of pressure and temperature in the ranges of 72–600 Torr and 223–323 K, *J. Phys. Chem. A*, *111*, 9047–9053.

Cakmur, R., R. L. Miller, and O. Torres (2004), Incorporating the effect of small scale circulations upon dust emission in an AGCM, *J. Geophys. Res.*, *109*, D07201, doi:10.1029/2003JD004067.

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

Charney, J., A. Arakawa, D. J. Baker, B. Bolin, R. E. Dickenson, R. M. Goody, C. E. Leith, H. M. Stommel, and C. I. Wunsch (1979), Carbon dioxide and climate: A scientific assessment, technical report, Natl. Acad. of Sci., Woods Hole, Mass.

Church, J. A., et al. (2011), Revisiting the Earth's sea-level and energy budgets from 1961 to 2008, *Geophys. Res. Lett.*, *38*, L18601, doi: 10.1029/2011GL048794.

Collatz, G. J., J. T. Ball, C. Grivet, and J. A. Berry (1991), Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: A model that includes a laminar boundary layer, *Agric. For. Meteorol.*, *54*, 107–136, doi:10.1016/0168-1923(91)90002-8.

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.

Dee, D. P., et al. (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*(656), 553–597, doi:10.1002/qj.828.

Del Genio, A. D., and M.-S. Yao (1993), Efficient cumulus parameterization for long-term climate studies: The GISS scheme, in *The Representation of Cumulus Convection in Numerical Models, AMS Meteorol. Monogr.*, vol. *46*, edited by K. A. Emanuel and D. A. Raymond, pp. 181–184, Am. Meteorol. Soc., Washington D.C.

Del Genio, A. D., M. S. Yao, W. Kovari, and K. K. Lo (1996), A prognostic cloud water parameterization for general circulation models, *J. Clim.*, *9*, 270–304.

Del Genio, A. D., M.-S. Yao, and J. Jonas (2007), Will moist convection be stronger in a warmer climate?, *Geophys. Res. Lett.*, *34*, L16703, doi: 10.1029/2007GL030525.

Dentener, F., et al. (2006), Nitrogen and sulfur deposition on regional and global scales: A multimodel evaluation, *Global Biogeochem. Cycles*, *20*, GB4003, doi:10.1029/2005GB002672.

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

Farquhar, G. D., and S. von Caemmerer (1982), Modelling photosynthetic response to environmental conditions, in *Encyclopedia of Plant Physiology*, vol. *12B*, edited by H. Z. O. L. Lange, C. B. Osmond, pp. 549–587, Springer, Berlin.

Fetterer, F., K. Knowles, W. Meier, and M. Savoie (2011), Sea ice index, Natl. Snow and Ice Data Cent., Boulder, Colo., doi: 10.7265/N5QJ7F7W.

Fetzer, E. J., B. H. Lambrigtsen, A. Eldering, H. H. Aumann, and M. T. Chahine (2006), Biases in total precipitable water vapor climatologies from Atmospheric Infrared Sounder and Advanced Microwave Scanning Radiometer, *J. Geophys. Res.*, *111*, D09S16, doi:10.1029/2005JD006598.

Fleming, E. L., S. Chandra, J. J. Barnett, and M. Corney (1990), Zonal mean temperature, pressure, zonal wind and geopotential height as function of latitude, COSPAR International Reference Atmosphere: 1986, Part II: Middle Atmosphere Models, *Adv. Space Res.*, *12*, 11–59.

Friend, A., and N. Kiang (2005), Land-surface model development for the GISS GCM: Effects of improved canopy physiology, *J. Clim.*, *18*, 2883–2902, doi:10.1175/JCLI3425.1.

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

Gorodetskaya, I. V., L. B. Tremblay, B. Liepert, M. A. Cane, and R. I. Cullather (2008), The influence of cloud and surface properties on the Arctic Ocean shortwave radiation budget in coupled models, *J. Clim.*, *21*, 866–882, doi:10.1175/2007JCLI1614.1.

Gregory, D. (2001), Estimation of entrainment rate in simple models of convective clouds, *Q. J. R. Metrol. Soc.*, *127*, 53–72.

Gregory, D., R. Kershaw, and P. Inness (1997), Parametrization of momentum transport by convection. II: Tests in single column and general circulation models, *Q. J. R. Metrol. Soc.*, *123*, 1153–1183.

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.

Griffies, S. M., et al. (2009), Coordinated Ocean-ice Reference Experiments (COREs), *Ocean Modell.*, *26*, 1–46, doi:10.1016/j.ocemod.2008.08.007.

Hallberg, R., A. Adcroft, J. P. Dunne, J. P. Krasting, and R. J. Stouffer (2013), 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.

Hallquist, M., D. J. Stewart, S. K. Stephenson, and R. A. Cox (2003), Hydrolysis of N_2O_5 on sub-micron sulfate aerosols, *Phys. Chem. Chem. Phys.*, *5*, 3453–3463.

Hänninen, H., and K. Kramer (2007), A framework for modelling the annual cycle of trees in boreal and temperate regions, *Silva Fennica*, *41*, 167–205.

Hansen, J., R. Ruedy, A. Lacis, M. Sato, L. Nazarenko, N. Tausnev, I. Tegen, and D. Koch (2000), Climate modeling in the global warming debate, in *General Circulation Model Development*, edited by D. Randall, pp. 127–164, Academic Press, San Diego, Calif.

Hansen, J., M. Sato, P. Kharecha, and K. von Schuckmann (2011), Earth's energy imbalance and implications, *Atmos. Chem. Phys.*, *11*, 13,421–13,449, doi:10.5194/acp-11-13421-2011.

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., G. L. Russell, D. Rind, P. Stone, A. Lacis, R. Ruedy, and L. Travis (1983), Efficient three-dimensional models for climatic studies, *Mon. Weather Rev.*, *111*, 609–662.

Hansen, J. E., et al. (1997), Forcings and chaos in interannual to decadal climate change, *J. Geophys. Res.*, *102*, 25,679–25,720.

Hansen, J. E., et al. (2005), Efficacy of climate forcings, *J. Geophys. Res.*, *110*, D18104, doi:10.1029/2005JD005776.

Hanson, D., and K. Mauersberger (1988), Laboratory studies of the nitric acid tridydrate: Implications for the south polar stratosphere, *Geophys. Res. Lett.*, *15*, 855–858.

Harrison, E. F., P. Minnis, B. R. Barkstrom, V. Ramanathan, R. D. Cess, and G. G. Gibson (1990), Seasonal variation of cloud radiative forcing derived from the Earth Radiation Budget Experiment, *J. Geophys. Res.*, *95*, 18,687–18,703.

Hegarty, T. W. (1973), Temperature coefficient (Q_{10}), seed germination and other biological processes, *Nature*, *243*, 305–306, doi:10.1038/243305a0.

Houweling, S., F. Dentener, and J. Lelieveld (1998), The impact of non-methane hydrocarbon compounds on tropospheric photochemistry, *J. Geophys. Res.*, *103*, 10,673–10,696.

Huffman, G. J., R. F. Adler, P. Arkin, A. Chang, R. Ferraro, A. Gruber, J. Janowiak, A. McNab, B. Rudolf, and U. Schneider (1997), The Global Precipitation Climatology Project (GPCP) combined precipitation data set, *Bull. Am. Meteorol. Soc.*, *78*, 5–20.

Hwang, Y.-T., and D. M. W. Frierson (2013), Link between the double-Intertropical Convergence Zone problem and cloud biases over the Southern Ocean, *Proc. Natl. Acad. Sci.*, *110*, 4935–4940, doi:10.1073/pnas.1213302110.

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

Johns, W. E., et al. (2011), Continuous, array-based estimates of Atlantic Ocean heat transport at 26.5°N, *J. Clim.*, *24*, 2429–2449, doi:10.1175/2010JCLI3997.1.

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

Kane, S. M., F. Caloz, and M.-T. Leu (2001), Heterogeneous uptake of gaseous N_2O_5 by $(NH_4)_2SO_4$, NH_4HSO_4, and H_2SO_4 aerosols, *J. Phys. Chem. A*, *105*, 6465–6470.

Kärcher, B., and U. Lohmann (2002), A parameterization of cirrus cloud formation: Homogeneous freezing of supercooled aerosols, *J. Geophys. Res.*, *107*(D2), 4010, doi:10.1029/2001JD000470.

Kim, D., A. Del Genio, and M.-S. Yao (2011), Moist convection scheme in ModelE2, technical report, NASA Goddard Inst. for Space Stud. arXiv:1312.7496.

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

Klein, S. A., and C. Jakob (1999), Validation and sensitivities of frontal clouds simulated by the ECMWF model, *Mon. Weather Rev.*, *127*, 2514–2531.

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

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., 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.

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

Koster, R., and M. Suarez (1996), Energy and water balance calculations in the Mosaic LSM, *Tech. Rep. 104606*, NASA Tech. Memor., 76 pp., Greenbelt, Md.

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.

Lawrence, D. M., P. E. Thornton, K. W. Oleson, and G. B. Bonan (2007), The partitioning of evapotranspiration into transpiration, soil evaporation, and canopy evaporation in a GCM: Impacts on land-atmosphere interaction, *J. Hydrometeorol.*, *8*, 862–880.

Lean, J. (2009), Calculations of Solar Irradiance. [Available at http://www.geo.fu-berlin.de/en/met/ag/strat/forschung/SOLARIS/Input_data/Calculations_of_Solar_Irradiance.pdf.]

Lean, J., Y.-M. Wang, and N. R. Sheeley Jr. (2002), The effect of increasing solar activity on the Sun's total and open magnetic flux during multiple cycles: Implications for solar forcing of climate, *Geophys. Res. Lett.*, *29*(24), 224, doi:10.1029/2002GL015880.

Lee, L. A., K. S. Carslaw, K. J. Pringle, G. W. Mann, and D. V. Spracklen (2011), Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters, *Atmos. Chem. Phys.*, *11*, 12,253–12,273, doi:10.5194/acp-11-12253-2011.

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

Levitus, S., J. I. Antonov, T. P. Boyer, R. A. Locarnini, H. E. Garcia, and A. V. Mishonov (2009), Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems, *Geophys. Res. Lett.*, *36*, L07608, doi:10.1029/2008GL037155.

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

Locarnini, R. A., A. V. Mishonov, J. I. Antonov, T. P. Boyer, H. E. Garcia, O. K. Baranova, M. M. Zweng, and D. R. Johnson (2010), *World Ocean Atlas 2009*, vol. *1*, *Temperature, Atlas NESDIS 68*, edited by S. Levitus, NOAA, U.S. Gov. Print. Off., Washington, D. C.

Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong (2009), Toward optimal closure of the Earth's Top-of-Atmosphere radiation budget, *J. Clim.*, *22*, 748–766, doi:10.1175/2008JCLI2637.1.

Lohmann, U., P. Stier, C. Hoose, S. Ferrrachat, S. Kloster, E. Roeckner, and J. Zhang (2007), Cloud microphysics and aerosol indirect effects in the global climate model ECHAM5-HAM, *Atmos. Chem. Phys.*, *7*, 3425–3446.

Mäkelä, A., P. Hari, F. Berninger, H. Hänninen, and E. Nikinmaa (2004), Acclimation of photosynthetic capacity in Scots pine to the annual cycle of temperature, *Tree Physiol.*, *24*(4), 369–376.

Mäkelä, A., P. Kolari, J. Karimäki, E. Nikinmaa, M. Perämäki, and P. Hari (2006), Modelling five years of weather-driven variation of GPP in a boreal forest, *Agric. For. Meteorol.*, *139*(3–4), 382–398, doi:10.1016/j.agrformet.2006.08.017.

Mauritsen, T., et al. (2012), Tuning the climate of a global model, *J. Adv. Model. Earth Syst.*, *4*, M00A01, doi:10.1029/2012MS000154.

Mears, C. A., M. Schabel, and F. J. Wentz (2003), A reanalysis of the MSU Channel 2 tropospheric temperature record, *J. Clim.*, *16*, 3650–3664.

Menon, S., and A. D. Del Genio (2007), Evaluating the impacts of carbonaceous aerosols on clouds and climate, in *Human-Induced Climate Change: An Interdisciplinary Assessment*, edited by M. E. Schlesinger et al., pp. 34–48, Cambridge Univ. Press, New York.

Menon, S., N. Unger, D. Koch, J. Francis, T. Garrett, I. Sednev, D. Shindell, and D. Streets (2008), Aerosol climate effects and air quality impacts from 1980 to 2030, *Environ. Res. Lett.*, *3*, 024004, doi:10.1088/1748-9326/3/2/024004.

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.

Miller, R. L., G. A. Schmidt, and D. T. Shindell (2006a), Forced annular variations in the 20th century Intergovernmental Panel on Climate Change Fourth Assessment Report models, *J. Geophys. Res.*, *111*, D18101, doi:10.1029/2005JD006323.

Miller, R. L., et al. (2006b), Mineral dust aerosols in the NASA Goddard Institute for Space Sciences ModelE atmospheric general circulation model, *J. Geophys. Res.*, *111*, D06208, doi:10.1029/2005JD005796.

Milliman, J. D., and R. H. Meade (1983), World-wide delivery of river sediment to the oceans, *J. Geol.*, *91*, 1–21.

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.

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*, 435–446, doi:10.1080/01490419.2010.491031.

Papa, F., C. Prigent, F. Aires, C. Jimenez, W. Rossow, and E. Matthews (2010), Interannual variability of surface water extent at global scale, 1993–2004, *J. Geophys. Res.*, *115*, D12111, doi:10.1029/2009JD012674.

Perovich, D. K., T. C. Grenfell, B. Light, and P. V. Hobbs (2002), Seasonal evolution of the albedo of multiyear Arctic sea ice, *J. Geophys. Res.*, *107*(C10), 8044, doi:10.1029/2000JC000438.

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

Prather, M., et al. (2001), Atmospheric chemistry and greenhouse gases, in *Climate Change 2001: The Scientific Basis*, edited by J. T. Houghton et al., pp. 881, Cambridge Univ. Press, New York.

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

Prigent, C., F. Papa, F. Aires, W. B. Rossow, and E. Matthews (2007), Global inundation dynamics inferred from multiple satellite observations, 1993–2000, *J. Geophys. Res.*, *112*, D12107, doi:10.1029/2006JD007847.

Rampal, P., J. Weiss, C. Dubois, and J.-M. Campin (2011), IPCC climate models do not capture arctic sea ice drift acceleration: Consequences in terms of projected sea ice thinning and decline, *J. Geophys. Res.*, *116*, C00D07, doi:10.1029/2011JC007110.

Randel, D. L., T. H. Vonder Haar, M. A. Ringerud, G. L. Stephens, T. J. Greenwald, and C. L. Combs (1996), A new global water vapor dataset, *Bull. Am. Meteorol. Soc.*, *77*, 1233–1246.

Rayner, D. J., et al. (2011), Monitoring the Atlantic Meridional Overturning Circulation, *Deep Sea Res., Part I*, *48*, 1744–1753, doi:10.1016/j.dsr2.2010.10.056.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan (2003), Global analyses of SST, sea ice and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, *108*(D14), 4407, doi:10.1029/2002JD002670.

Repo, T., A. Mäkelä, and H. Hänninen (1990), Modelling frost resistance of trees, *Silva Carelica*, *15*, 61–74.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Q. Wang (2002), An improved in situ and satellite SST analysis for climate, *J. Clim.*, *15*, 1609–1625.

Rind, D., R. Suozzo, N. K. Balachandran, A. Lacis, and G. Russell (1988), The GISS Global Climate-Middle Atmosphere Model. Part I: Model structure and climatology, *J. Atmos. Sci.*, *45*, 329–370.

Rind, D. H., J. L. Lean, and J. Jonas (2014), The impact of different absolute solar irradiance values on current climate model simulations, *J. Clim.*, doi:10.1175/JCLI-D-13-00136.1, in press.

Rodriguez-Iturbe, I., A. Porporato, F. Laio, and L. Ridolfi (2001), Plants in water-controlled ecosystems: Active role in hydrologic processes and responses to water stress. I. Scope and general outline, *Adv. Water Resour.*, *24*, 695–705.

Rossow, W. B., and R. A. Schiffer (1999), Advances in understanding clouds from ISCCP, *Bull. Am. Meteorol. Soc.*, *80*, 2261–2288.

Rothman, L. S., 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.

Rougier, J. (2013), "intractable and unsolved": Some thoughts on statistical data assimilation with uncertain static parameters, *Philos. Trans. R. Soc. A*, *371*, 20120297, doi:10.1098/rsta.2012.0297.

Sander, S. P., et al. (2000), Chemical kinetics and photochemical data for use in stratospheric modeling, *Eval. 13, JPL Publ. 00-003*, Jet Propul. Lab, Pasadena, Calif.

Sato, M., J. E. Hansen, M. P. McCormick, and J. B. Pollack (1993), Stratospheric aerosol optical depths, 1850–1990, *J. Geophys. Res.*, *98*, 22,987–22,994.

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

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, doi:10.1175/JCLI3612.1.

Schmidt, G. A., R. Ruedy, R. L. Miller, and A. A. Lacis (2010), The attribution of the present-day total greenhouse effect, *J. Geophys. Res.*, *115*, D20106, doi:10.1029/2010JD014287.

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

Shindell, D., et al. (2012), Simultaneously mitigating near-term climate change and improving human health and food security, *Science*, *335*, 183–189, doi:10.1126/science.1210026.

Shindell, D. T., G. Faluvegi, R. L. Miller, G. A. Schmidt, J. E. Hansen, and S. Sun (2006a), Solar and anthropogenic forcing of tropical hydrology, *Geophys. Res. Lett.*, *33*, L24706, doi:10.1029/2006GL027468.

Shindell, D. T., G. Faluvegi, N. Unger, E. Aguilar, G. A. Schmidt, D. M. Koch, S. E. Bauer, and R. L. Miller (2006b), Simulations of preindustrial, present-day, and 2100 conditions in the NASA GISS composition and climate model G-PUCCINI, *Atmos. Chem. Phys.*, *6*, 4427–4459.

Shindell, D. T., et al. (2006c), Multi-model simulations of carbon monoxide: Comparison with observations and projected near-future changes, *J. Geophys. Res.*, *111*, D19306, doi:10.1029/2006JD007100.

Shindell, D. T., G. Faluvegi, D. M. Koch, G. A. Schmidt, N. Unger, and S. E. Bauer (2009), Improved attribution of climate forcing to emissions, *Science*, *326*, 716–718, doi:10.1126/science.1174760.

Shindell, D. T., et al. (2013), Interactive ozone and methane chemistry in GISS-E2 historical and future simulations, *Atmos. Chem. Phys.*, *13*, 2653–2689, doi:10.5194/acp-13–2653-2013.

Shiogama, H., et al. (2012), Perturbed physics ensemble using the MIROC5 coupled atmosphere-ocean GCM without flux corrections: Experimental design and initial results, *Clim. Dyn.*, *39*, 3041–3056, doi:10.1007/s00382-012-1441-x.

Siebesma, A. P., et al. (2003), A large eddy simulation intercomparison study of shallow cumulus convection, *J. Atmos. Sci.*, *60*, 1201–1219.

Simmons, A. J., and J. K. Gibson (2000), The ERA-40 project plan, *Tech. Rep. ERA-40, Proj. Rep. Ser. 1*, Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.

Simpson, I. R., P. Hitchcock, T. G. Shepherd, and J. F. Scinocca (2013), Southern Annular Mode dynamics in observations and models. Part 1: The influence of climatological zonal wind biases in a comprehensive GCM, *J. Clim.*, *26*, 3953–3967, doi:10.1175/JCLI-D-12–00348.1.

Smagorinsky, J. (1996), Some historical remarks on the use of non-linear viscosities, in *Large Eddy Simulation of Complex Engineering and Geophysical Flows*, edited by B. Galperin and S. A. Orszag, pp. 3–36, Cambridge Univ. Press, New York.

Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore (2008), Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006), *J. Clim.*, *21*, 2283–2296.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller (Eds.) (2007), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge Univ. Press, Cambridge, U. K.

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

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

Stainforth, D. A., et al. (2005), Uncertainty in predictions of the climate response to rising levels of greenhouse gases, *Nature*, *433*, 403–406.

Steele, M., R. Morley, and W. Ermold (2001), PHC: A global ocean hydrography with a high quality Arctic Ocean, *J. Clim.*, *14*, 2079–2087.

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.

Stevenson, D. S., et al. (2006), Multimodel ensemble simulations of present-day and near-future tropospheric ozone, *J. Geophys. Res.*, *111*, D08301, doi:10.1029/2005JD006338.

Stroeve, J., M. M. Holland, W. Meier, T. Scambos, and M. Serreze (2007), Arctic sea ice decline: Faster than forecast, *Geophys. Res. Lett.*, *34*, L09501, doi:10.1029/2007GL029703.

Stubenrauch, C. J., et al. (2013), Assessment of global cloud datasets from satellites: Project and database initiated by the GEWEX radiation panel, *Bull. Am. Meteorol. Soc.*, *94*, 1031–1049, doi:10.1175/BAMS-D-12-00117.1.

Sun, S., and R. Bleck (2006), Multi-century simulations with the coupled GISS-HYCOM climate model: Control experiments, *Clim. Dyn.*, *26*, 407–428, doi:10.1007/s00382-005-0091-7.

Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, *106*, 7183–7192.

Taylor, K. E., R. Stouffer, and G. Meehl (2009), *A summary of the CMIP5 Experimental Design*.

Taylor, K. E., R. Stouffer, and G. 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.

Trenberth, K. E., and J. M. Caron (2001), Estimates of meridional atmosphere and ocean heat transports, *J. Clim.*, *14*, 3433–3443.

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

Tsigaridis, K., and M. Kanakidou (2007), Secondary organic aerosol importance in the future atmosphere, *Atmos. Environ.*, *41*, 4682–4692.

Tsigaridis, K., D. Koch, and S. Menon (2013), Uncertainties and importance of sea spray composition on aerosol direct and indirect effects, *J. Geophys. Res.*, *118*, 220–235, doi:10.1029/2012JD018165.

van de Berg, W. J., M. R. van den Broeke, C. H. Reijmer, and E. van Meijgaard (2006), Reassessment of the Antarctic surface mass balance using calibrated output of a regional atmospheric climate model, *J. Goephys. Res.*, *111*, D11104, doi:10.1029/2005JD006495.

Voulgarakis, A., P. J. Telford, A. M. Aghedo, P. Braesicke, G. Faluvegi, N. L. Abraham, K. W. Bowman, J. A. Pyle, and D. T. Shindell (2011), Global multi-year O_3-CO correlation patterns from models and TES satellite observations, *Atmos. Chem. Phys.*, *11*, 5819–5838, doi: 10.5194/acp-11-5819-2011.

Wajsowicz, R. C. (1993), A consistent formulation of the anisotropic stress tensor for use in models of the large-scale ocean circulation, *J. Comput. Phys.*, *105*, 333–338.

Waliser, D. E., et al. (2009), Cloud ice: A climate model challenge with signs and expectations of progress, *J. Geophys. Res.*, *114*, D00A21, doi: 10.1029/2008JD010015.

Wang, D., G. L. Wang, and E. N. Anagnostou (2007), Evaluation of canopy interception schemes in land surface models, *J. Hydrol.*, *347*, 308–318.

Wang, Y., H. Cheng, R. L. Edwards, Y. He, X. Kong, Z. An, J. Wu, M. J. Kelly, C. A. Dykoski, and X. Li (2005a), The Holocene Asian Monsoon: Links to solar changes and North Atlantic climate, *Science*, *308*, 854–857.

Wang, Y.-M., J. L. Lean, and N. R. Sheeley Jr. (2005b), Modeling the Sun's magnetic field and irradiance since 1713, *Astrophys. J.*, *625*, 522–538, doi:10.1086/429689.

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

Webb, M., C. Senior, S. Bony, and J. J. Morcrette (2001), Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models, *Clim. Dyn.*, *17*, 905–922.

Williams, K. D., W. J. Ingram, and J. M. Gregory (2008), Time variation of effective climate sensitivity in GCMs, *J. Clim.*, *21*, 5076–5090, doi: 10.1175/2008JCLI2371.1.

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

Xie, P., and A. Arkin (1997), Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, *Bull. Am. Meteorol. Soc.*, *78*, 2539–2558.

Yokohata, T., J. D. Annan, M. Collins, C. S. Jackson, M. Tobis, M. J. Webb, and J. C. Hargreaves (2013), Reliability of multi-model and structurally different single-model ensembles, *Clim. Dyn.*, *39*, 599–616, doi:10.1007/s00382-011-1203-1.

Zeng, X., M. Barlage, R. E. Dickinson, Y. Dai, G. Wang, and K. Oleson (2005), Treatment of undercanopy turbulence in land models, *J. Clim.*, *18*, 5086–5094.

Zhang, J., and D. Rothrock (2000), Modeling Arctic sea ice with an efficient plastic solution, *J. Geophys. Res.*, *105*, 3325–3338.

Zhang, Y.-C., and W. Rossow (1997), Estimating meridional energy transports by the atmospheric and oceanic general circulations using boundary flux data, *J. Clim.*, *10*, 2358–2373.