RESEARCH ARTICLE

The GFDL Global Atmosphere and Land Model AM4.0/LM4.0: 1. Simulation Characteristics With Prescribed SSTs

M. Zhao[1], J.-C. Golaz[1,2], I. M. Held[1], H. Guo[3], V. Balaji[4], R. Benson[1], J.-H. Chen[1], X. Chen[4], L. J. Donner[1], J. P. Dunne[1], K. Dunne[5], J. Durachta[1], S.-M. Fan[1], S. M. Freidenreich[1], S. T. Garner[1], P. Ginoux[1], L. M. Harris[1], L. W. Horowitz[1], J. P. Krasting[1], A. R. Langenhorst[1,8], Z. Liang[1], P. Lin[1], S.-J. Lin[1], S. L. Malyshev[1], E. Mason[6], P. C. D. Milly[1,5], Y. Ming[1], V. Naik[1], F. Paulot[4], D. Paynter[1], P. Phillipps[1], A. Radhakrishnan[6], V. Ramaswamy[1], T. Robinson[6], D. Schwarzkopf[1], C. J. Seman[1], E. Shevliakova[1], Z. Shen[4], H. Shin[4], L. G. Silvers[4], J. R. Wilson[7], M. Winton[1], A. T. Wittenberg[1], B. Wyman[1], and B. Xiang[3]

[1]GFDL/NOAA, Princeton, NJ, USA, [2]Lawrence Livermore National Laboratory, Livermore, CA, USA, [3]UCAR/GFDL, Princeton, NJ, USA, [4]Princeton University/GFDL, Princeton, NJ, USA, [5]U.S. Geological Survey, Princeton, NJ, USA, [6]Engility Corporation/GFDL, NJ, USA, [7]NASA Ames Research Center, Moffett Field, CA, USA, [8]Deceased 31 March 2016

Abstract In this two-part paper, a description is provided of a version of the AM4.0/LM4.0 atmosphere/land model that will serve as a base for a new set of climate and Earth system models (CM4 and ESM4) under development at NOAA's Geophysical Fluid Dynamics Laboratory (GFDL). This version, with roughly 100 km horizontal resolution and 33 levels in the vertical, contains an aerosol model that generates aerosol fields from emissions and a "light" chemistry mechanism designed to support the aerosol model but with prescribed ozone. In Part 1, the quality of the simulation in AMIP (Atmospheric Model Intercomparison Project) mode—with prescribed sea surface temperatures (SSTs) and sea-ice distribution—is described and compared with previous GFDL models and with the CMIP5 archive of AMIP simulations. The model's Cess sensitivity (response in the top-of-atmosphere radiative flux to uniform warming of SSTs) and effective radiative forcing are also presented. In Part 2, the model formulation is described more fully and key sensitivities to aspects of the model formulation are discussed, along with the approach to model tuning.

1. Introduction

Documentation of comprehensive climate models is challenging and can take various forms and provide varying levels of detail. Our goal here is to provide enough detail to explain the key choices made in developing the AM4.0/LM4.0 atmosphere/land model that will serve as a base for a new set of climate and Earth system models at NOAA's Geophysical Fluid Dynamics Laboratory (GFDL). We focus specifically on model performance in AMIP mode. We mention but do not emphasize the various features that this model shares with previous GFDL models AM2 (GFDL-GAMDT, 2004), AM3 (Donner et al., 2011), HiRAM (Zhao et al., 2009), and LM3 (Milly et al., 2014). We have extensively studied a version of this model coupled to both 0.5° and 0.25° versions of the new MOM6 ocean model and the new SIS2 sea-ice model. These coupled simulations will be described elsewhere, but reference to our experience with the coupled model will be needed to motivate some of the choices made in the atmospheric model.

The atmospheric model grid has cubed-sphere topology with 96×96 grid boxes per cube face (approximately 100 km grid size); hence, it is referred to as having C96 horizontal resolution. This represents an increased horizontal resolution compared to AM2 and AM3, which have horizontal grid-spacing of $\sim$200 km. The number of vertical levels is set at 33, with relatively crude resolution of the stratosphere with model top at 1 hPa, but with a sponge layer extending down to 8 hPa. The vertical level structure in the model is provided in supporting information S1. The vertical resolution is similar to HIRAM and AM3 in the troposphere except an additional layer is placed near the surface so that the model's lowest layer is closer to surface. We have also studied versions of the model with an increased horizontal (50 km) or vertical (49 and 63 levels) resolution and they will be documented in a separate paper.

Table 1
A Description of the Observational and Reanalysis Data Sets Used in This Paper

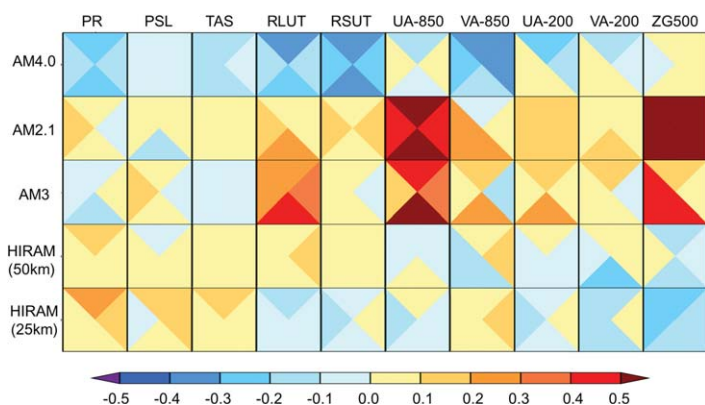| Data	Description
Precipitation (1980–2014)	*Abbreviation*: GPCP-v2.3 *Full name*: Global Precipitation Climatology Project, version 2.3 *Reference*: Adler et al. (2003) *Url*: https://www.esrl.noaa.gov/psd/data/gridded/data.gpcp.html
TOA radiative fluxes (03/2000 to 02/2015)	*Abbreviation*: CERES-EBAF-ed2.8 *Full name*: Clouds & Earth's Radiant Energy Systems Energy Balanced & Filled data, edition 2.8 *Reference*: Loeb et al. (2009); Smith et al. (2011) *Url*: http://ceres.larc.nasa.gov/documents/DQ_summaries/CERES_EBAF_Ed2.8_DQS.pdf
Land surface temperature (1980–2014)	*Abbreviation*: CRU-TS-v4.01 *Full name*: Climatic Research Unit TS data-set version 4.01 *Reference*: Harris et al. (2014) *url*: https://crudata.uea.ac.uk/cru/data/hrg/#superseded
Monthly atmospheric temperature, winds, sea level pressure, geopotential (1980–2014)	*Abbreviation*: ERA-Interim *Full name*: European Reanalysis INTERIM *Reference*: Dee et al. (2011) *Url*: https://www.ecmwf.int/en/research/climate-reanalysis/era-interim
Daily stratosphere temperature and winds for computing SSW (1957–2002)	*Abbreviation*: ERA-40 *Full name*: European Reanalysis 40 *Reference*: Uppala et al. (2005) *Url*: http://apps.ecmwf.int/datasets/data/era40-daily/levtype=sfc/
Daily atmospheric temperature, winds for computing eddy fluxes (1980–2014)	*Abbreviation*: MERRA *Full name*: Modern-Era Retrospective Analysis for Research and Applications *Reference*: Rienecker et al. (2011) *Url*: https://gmao.gsfc.nasa.gov/reanalysis/MERRA/
Tropical cyclone tracks (1980–2014)	*Abbreviation*: IBTRACS-v03r09 *Full name*: International Best Track Archive for Climate Stewardship version v03r09 *Reference*: Knapp et al. (2010) *Url*: https://www.ncdc.noaa.gov/ibtracs/index.php?name=ibtracs-data-access/
Daily OLR for computing tropical waves and MJO (01/1979 to 12/2011)	*Abbreviation*: NOAA-AVHRR *Full name*: NOAA Advanced Very High Resolution Radiometer *Reference*: Liebmann and Smith (1996) *Url*: https://www.esrl.noaa.gov/psd/data/gridded/data.interp_OLR.html
Aerosol optical depth (2000–2010)	*Abbreviation*: AERONET-v2 *Full name*: Aerosol Robotic Network version 2 *Reference*: Holben et al. (1998); Smirnov et al. (2000) *Url*: https://aeronet.gsfc.nasa.gov/
Aerosol optical depth (2000–2014)	*Abbreviation*: MISR *Full name*: Multi-angle Imaging SpectroRadiometer *Reference*: Kahn et al. (2009) *Url*: https://eosweb.larc.nasa.gov/project/misr/cgas_table
Aerosol optical depth (2000–2014)	*Abbreviation*: MODIS *Full name*: Moderate Resolution Imaging Spectroradiometer Collection 6 *Reference*: Levy et al. (2015) *Url*: https://modis-atmos.gsfc.nasa.gov/MOD08_M3/index.html
Cloud amount, optical depth (2003–2010) Cloud droplet number (2003–2015)	*Abbreviation*: MODIS *Full name*: Moderate Resolution Imaging Spectroradiometer *Reference*: Platnick et al. (2003); Pincus et al. (2012); Bennartz and Rausch (2017) *Url*: https://modis-atmos.gsfc.nasa.gov/MOD06_L2/index.html *Url*: ftp://ladsweb.nascom.nasa.gov/NetCDF/L3_Monthly/V02/
Cloud amount and cloud optical depth (1983–2008)	*Abbreviation*: ISCCP *Full name*: International Satellite Cloud Climatology Project *Reference*: Rossow and Schiffer (1991); Pincus et al. (2012) *Url*: https://isccp.giss.nasa.gov/ *Url*: ftp://ftp.climserv.ipsl.polytechnique.fr/cfmip/ISCCP/

Figure 1. Comparison of AM4.0 AMIP simulation with the previous GFDL models using the PCMDI metrics package version 1.1.1 portrait plot (Gleckler et al., 2016). The color scale portrays a model's RMSE as a relative error (unitless) by normalizing the result by the median error of all models results shown in the figure. For example, a value of 0.20 indicates that a model's RMSE is 20% larger than the median error for that variable across all simulations on the figure. The triangles in each grid square show results for four individual seasons (from left rotating clockwise are respectively for JJA, SON, DJF, and MAM seasons) from multiple global fields (PR, precipitation; PSL, sea level pressure; TAS, surface air temperature; RLUT, outgoing LW radiation at TOA; RSUT, reflected SW radiation at TOA; UA-850, 850 hPa zonal wind; VA-850, 850 hPa meridional wind; UA-200, 200 hPa zonal wind; VA-200, 200 hPa meridional wind; ZG500, 500 hPa geopotential height).

The model described here has aerosol physics based in large part on that in AM3, but with a simplified chemistry, retaining a minimal chemical mechanism designed to simulate aerosols from emissions while prescribing ozone concentrations. We have also studied a version of this model with a level structure and a more comprehensive troposphere/stratosphere chemistry mechanism following that in AM3 that predicts ozone concentrations. We refer to the latter chemistry mechanism as the "full" chemistry module and the truncated chemical mechanism utilized in the model described here as "light" chemistry. In addition to the prognostic variables horizontal velocity, temperature, specific humidity, liquid and frozen water, cloud fraction, and surface pressure, AM4.0 also includes a subset of the tracers in AM3—water droplet number plus 17 aerosol and 4 gas phase tracers. As in AM3, the model simulates the indirect aerosol effect through a predictive model for cloud droplet numbers. To avoid the need for long spin-ups needed to equilibrate the vegetation state, the land model utilized here is run with static rather than dynamic vegetation.

We start in Part 1 by providing an overview of the model performance in AMIP (1980–2014) mode after a brief introduction to the model formulation. This paper highlights some improvements from previous efforts and some more recalcitrant problems. In some cases, we compare to the quality of all available CMIP5 AMIP simulations, as well as to past GFDL model simulations. We include some discussion of climate sensitivity as measured by the "Cess sensitivity"—increasing SSTs everywhere by 2K and examining the response in the top-of-atmosphere (TOA) fluxes. We also refer to computations of the "effective radiative forcing" or "radiative flux perturbation," RFP, computed by holding SSTs and sea ice fixed and comparing TOA fluxes given preindustrial and present-day forcing agents.

The formulation of key model components is discussed much more fully in Zhao et al. (2018) (below referred to as Part 2), where the sensitivity of the simulation to some of the choices within these model components is also discussed, with particular attention to the choices in the convection, cloud, and orographic gravity wave drag schemes. In Part 2, we also make a special effort to describe our optimization strategy for the simulation of clouds and TOA fluxes. And we try to make clear the extent to which we have tuned toward values of RFP and Cess sensitivity.

In AMIP mode, the model is driven by time-varying boundary conditions, and natural and anthropogenic forcings developed in support of CMIP6 (Eyring et al., 2016) which are archived by the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and served by the Earth System Grid Federation (https://esgf-node.llnl.gov/search/input4mips/). Details are provided in Appendix A.

We have utilized a variety of observational and reanalysis data sets for evaluating AM4.0/LM4.0 performance in simulating the observed long-term mean climate as well as statistics of atmospheric transients. Table 1 provides a detailed description of the data sets, including their short and full names, references, URL as well as the period used for computing observed mean climatology and transients statistics. The short-names of the data sets will be used throughout the paper when referring to the data.

2. Brief Introduction to the Model Formulation

Each of the following topics is discussed in substantial detail in Part 2. We include this list here to provide some context for the following description of the AMIP simulation. The AM4.0 atmospheric model includes the following:

1. the hydrostatic version of the FV3 finite-volume cubed-sphere dynamical core, with minor modifications from the version used in AM3;
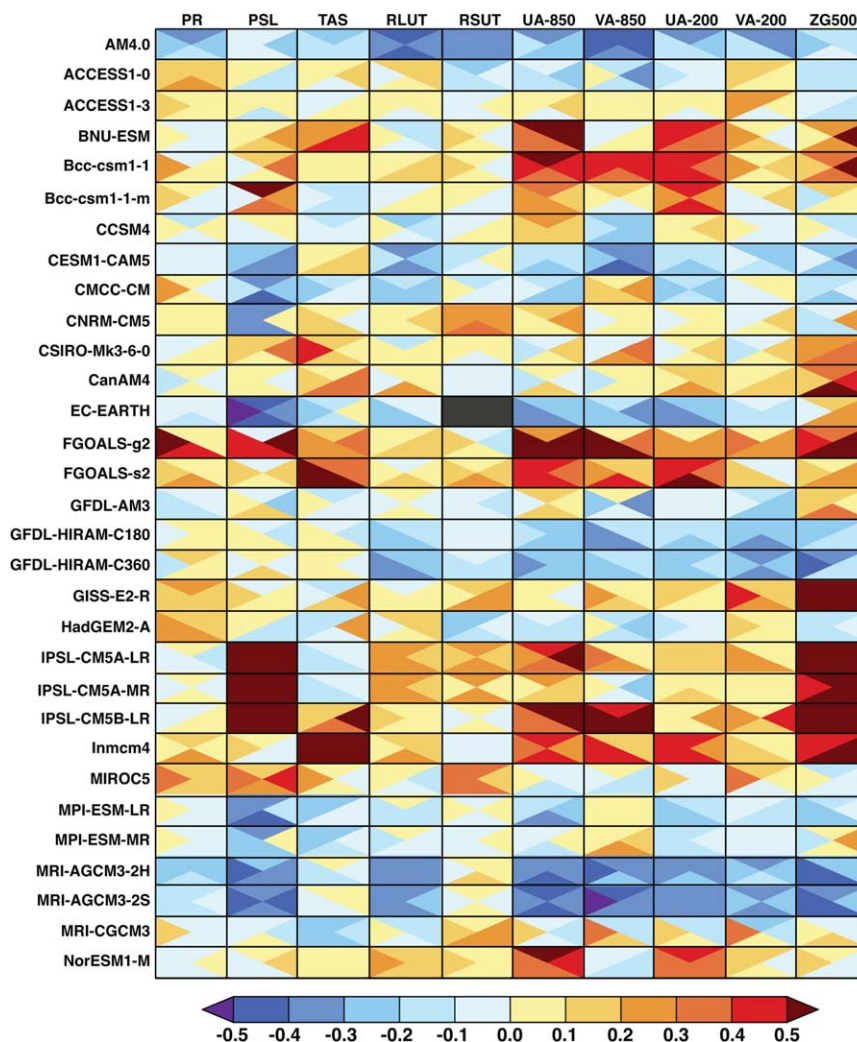
Figure 2. As in Figure 1 except compared to the CMIP5 ensemble of AMIP simulations, with RMS errors normalized by the median error in the 30 CMIP5 models available from the PCMDI metrics package version 1.1.1.

2. a substantially updated version of the GFDL radiative transfer code, with refitting to line-by-line simulations using the latest spectroscopy and adding 10 μm CO_2 band among other changes; this increases the TOA radiative forcing due to quadrupling of CO_2 by $\sim$10%;

3. an alternate topographic gravity wave drag formulation that handles the anisotropy of the subgrid topography naturally and treats the blocked versus wave components of the drag in a distinctive way;

4. a double-plume model representing shallow and deep convection, with stronger/weaker lateral mixing rate for the shallow/deep plume, with convective inhibition closure for the shallow plume mass flux and cloud work function relaxation closure for the deep plume, and with the deep plume strongly constrained by the environmental relative humidity;

5. a "light" chemistry mechanism using AM3's chemistry as a starting point, retaining the ability to generate aerosol distributions from emissions, but with ozone and other oxidants prescribed;

6. an aerosol module similar in structure to that in AM3, but with significant modification to wet removal by convection and by frozen precipitation;

7. single-moment cloud microphysics as in AM3 but with substantial changes in the droplet simulation and strongly reduced aerosol indirect effect;

8. updated surface flux formulation over oceans;

9. cloud fraction prediction, boundary layer formulation, and nonorographic gravity wave scheme essentially as in AM3, with minor retuning.
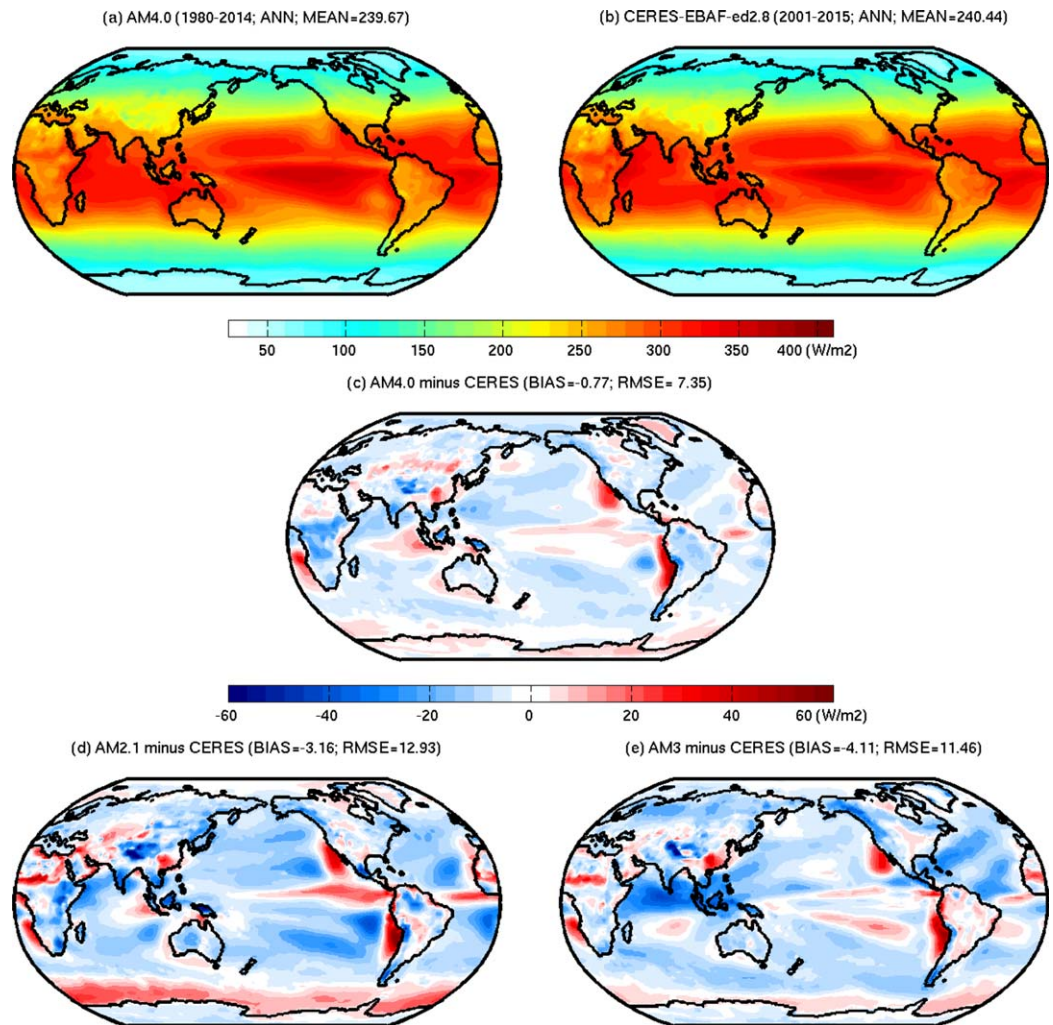
Figure 3. Long-term annual mean TOA net SW downward (absorption) radiative flux in W m^{-2} from (a) AM4.0 AMIP simulation and (b) observational estimate based on CERES-EBAF-ed2.8, averaged for the 2001–2015 period. (c) Model biases (AM4.0 minus CERES). (d) As in Figure 3c except for AM2.1. (e) As in Figure 3c except for AM3. Titles of Figures 3a and 3b show global mean values. Titles of Figures 3c–3e show global mean biases and RMS errors.

In addition, the land model LM4.0 is only incrementally modified from LM3.0, when run with static vegetation (representative of 1981 conditions) as in the simulations described here.

3. Portrait Plot Overview

During the model development, we have routinely used a variety of automatic analysis packages to evaluate model performance against the observations and reanalysis (see Table 1 for some of the data sets). Most of the analysis focused on spatial distribution of meteorological fields such as winds, temperature, humidity, clouds, atmospheric tracers as well as the energy fluxes. In addition, we have also recently utilized the portrait plots described by Gleckler et al. (2008, 2016) to monitor aspects of simulation quality along our development trajectory. In Figures 1 and 2, we compare features of the seasonal mean large-scale climate of AM4.0/LM4.0 to that of other GFDL models and to the AMIP simulations in the CMIP5 archive. The different models are displayed along the vertical axis (AM4.0/LM4.0 is the top line in both figures) while different large-scale fields (as described in the caption) define the horizontal axis. Each square is divided into four triangles representing different seasons. (The triangles from left rotating clockwise are respectively for the JJA, SON, DJF, and MAM seasons.) Blue colors indicate that the spatial RMS error versus an observational data
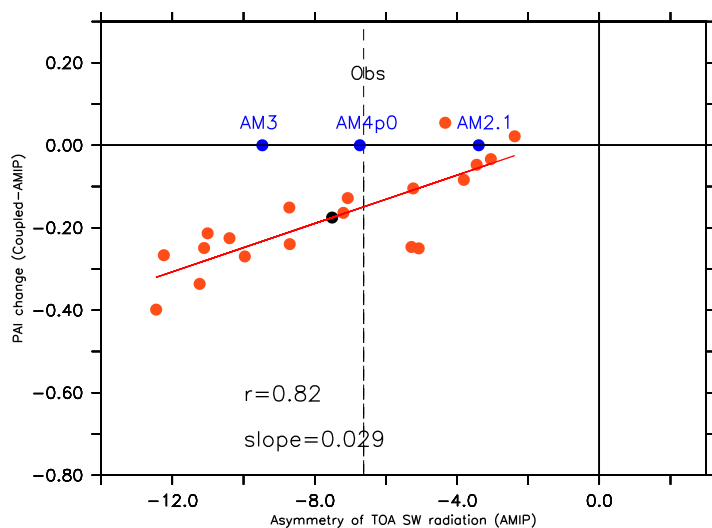
Figure 4. Scatterplot of the change in precipitation asymmetry index (PAI: unitless, as defined in Xiang et al. (2017)) from an AMIP to coupled simulation versus the asymmetry in tropical TOA net downward SW radiative flux (unit: W m^{-2}) of the corresponding AMIP simulation. Red dots include 19 CMIP5 models which provided both coupled and AMIP simulations. The black dot shows the 19-model mean. The asymmetry in tropical TOA net downward SW radiative flux computed from the CERES-EBAF-ed2.8 observations (dashed line) and the AM4.0, AM3, and AM2.1 AMIP simulations (blue dots) are denoted along the zero horizontal line. The PAI is computed as the difference in climatological precipitation between 0°N–20°N and 0°S–20°S normalized by the tropical (20°S–20°N) mean precipitation. The TOA SW asymmetry is defined as the difference between the average over 0°N–38°N and that over 0°S–38°S.

set is small relative to that of the other models in the comparison set, while red colors indicate larger RMS errors. AM4.0 compares well overall with previous GFDL models, especially AM2.1 and AM3, but with a few circulation metrics comparable to or inferior to the higher resolution HiRAM models. The comparison with the CMIP5 models is also favorable overall, with the high resolution (20–60 km) MRI AGCMs (Mizuta et al., 2012) evidently being the best examples of CMIP5 AMIP simulations of comparable overall quality to AM4.0. Fields in which AM4.0 has larger biases than the best of these CMIP5 models include sea level pressure and 500 hPa geopotential. The latter is in part due to an overall tropical tropospheric cold bias (discussed below) rather than circulation-related features.

4. TOA Radiative Fluxes

We start our more detailed evaluation of AM4.0 performance with the TOA radiative fluxes for two reasons. One is that they are among the best observed climate fields. More importantly, we consider the quality of the simulation of the TOA radiative fluxes as especially important for coupled atmosphere-ocean simulations. These portrait plots indicate that AM4.0's bias in TOA fluxes is particularly small compared to the biases in other models, especially in the reflected solar. Figure 3 shows the bias pattern in the annual mean net downward shortwave flux at the TOA in AM4.0 compared with two other recent GFDL models, AM2.1 and AM3. (Similar plots for the four different seasons are provided in supporting information Figures S1–S4). The observational estimate of the TOA radiative fluxes is based on CERES-EBAF-ed2.8. This data set is used for observational estimates of TOA and surface radiative flux throughout the paper unless otherwise noted. A bias toward excessive reflection over most ocean regions in earlier models is largely ameliorated. The excessive SW absorption over the southern ocean (∼60°S) in AM2.1 is greatly reduced in both AM3 and AM4.0 with AM4.0 in particular producing the least overall bias over the Southern Hemisphere oceans, both north and south of ∼60°S. Remaining biases are due to insufficient low cloud cover in the subtropical stratocumulus regions off the west coasts of North and South America and Africa, and excessive reflection from convective cloudiness over sub-Saharan Africa, North Indian Ocean, and the western tropical Pacific.

We consider the quality of the simulation of this field as especially important for coupled atmosphere-ocean simulations. For example, coupled models are typically biased toward excessive tropical precipitation in the Southern Hemisphere, associated with the well-known double ITCZ bias. A quantitative measure of this bias is the tropical precipitation asymmetry index (PAI), defined as difference in climatological precipitation between 0°N–20°N and 0°S–20°S (e.g., Hwang & Frierson, 2013; Xiang et al., 2017). Xiang et al. (2017) has shown, using the CMIP5 archive, that the PAI in an AMIP simulation is not a good predictor of the PAI in the corresponding coupled model; however, the north-south asymmetry in the reflected shortwave at TOA is a very good predictor of the change in the PAI from its value in an AMIP simulation to its value in the corresponding coupled model. Following Xiang et al. (2017), Figure 4 shows this relationship across 19 CMIP5 models. This TOA SW asymmetry is computed as the difference between the average over the Northern tropics/subtropics (0°N–38°N) and the value for the corresponding region in the Southern Hemisphere (0°S–38°S); (See Xiang et al., 2017 for the rationale for this choice of averaging region.) The observational estimate of the TOA SW asymmetry is also indicated along with the results from AM4.0, AM3, and AM2 simulations.

AM4.0 produces a TOA SW asymmetry very close to the observations, suggesting that development of a PAI bias in coupled mode from unrealistic asymmetry in the SW forcing provided by AM4.0 should be minimal. But there are evidently other sources of this bias, possibly oceanic, shared by most models, since the CMIP5 model regression line has a negative value where it intersects the observational estimate.
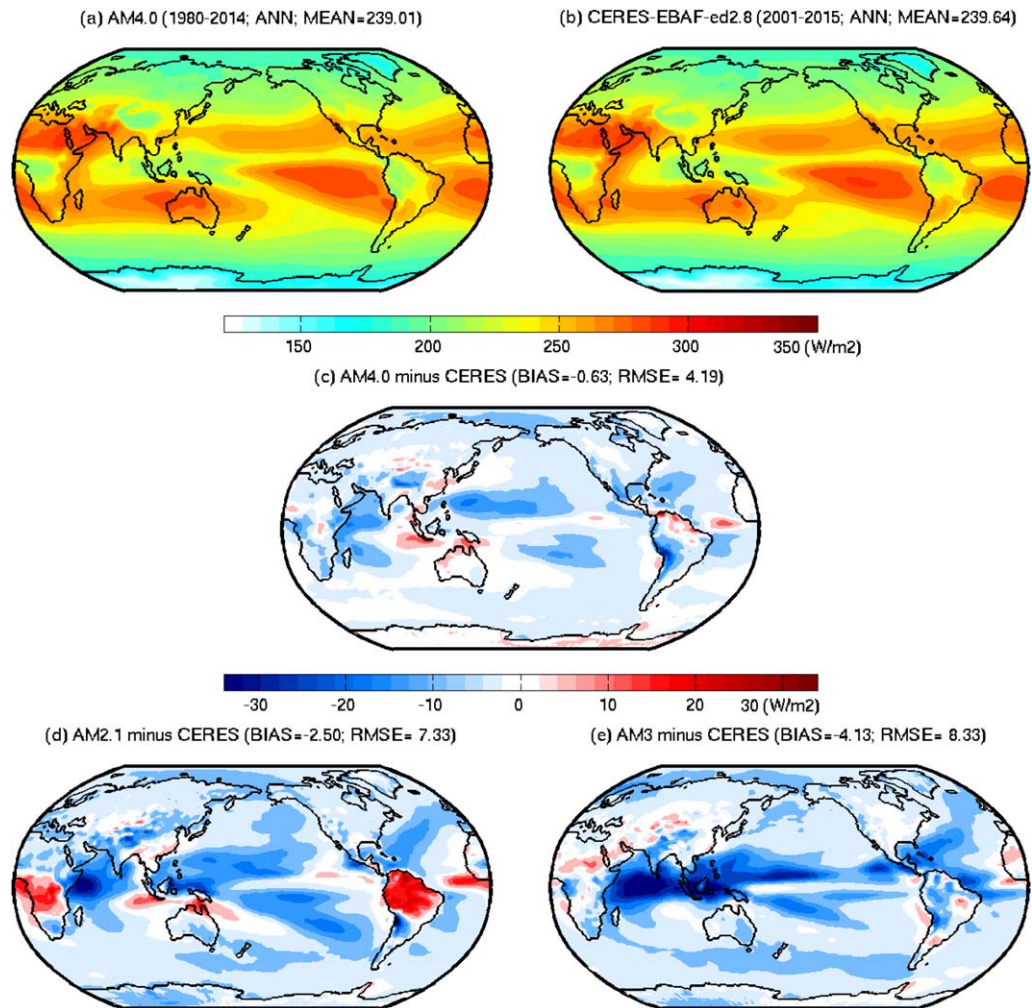
Figure 5. As in Figure 3 except for OLR.

The corresponding annual mean spatial bias plot for the outgoing longwave radiation (OLR) is shown in Figure 5. The OLR pattern is clearly improved as compared to previous GFDL models: global RMS error is reduced by nearly half from AM2 and AM3 to AM4.0. Much of the reduction in OLR error is due to the improvement in AM4.0 simulated precipitation climatology (described in the following) and associated high cloud cover. For example, the reduction of OLR error over much of the south-east Pacific in AM4.0 as compared to AM2.1 is due to less precipitation in that region (less of a double ITCZ; see Figure 11). The improvement in OLR over the Amazonian region seems to be due to a reduction of the Amazon dry bias in AM2.1. However, the picture is not consistent between the African tropics and the Amazon, with the OLR in the former region improving with little reduction in precipitation bias. Comparing AM4.0 with both AM2.1 and AM3, the reduction in OLR bias in the tropics is due to a mix of reduced bias in precipitation and in the OLR response to precipitation. We emphasize especially that the reduction of global mean bias in OLR from AM2.1 ($-2.5\,\mathrm{W\,m^{-2}}$) and AM3 ($-4.1\,\mathrm{W\,m^{-2}}$) to AM4.0 ($-0.6\,\mathrm{W\,m^{-2}}$) is due to the explicit tuning of AM4.0 OLR toward the value ($239.6\,\mathrm{W\,m^{-2}}$) in CERES-EBAF-ed2.8. Supporting information Figures S5–S8 show the corresponding OLR bias plots for individual seasons.

The bias in net TOA radiative flux is shown in Figure 6. AM4.0 shows clear improvement from previous GFDL models with the global RMS error reduced from $11.9\,\mathrm{W\,m^{-2}}$ in AM2.1, $8.7\,\mathrm{W\,m^{-2}}$ in AM3 to $6.5\,\mathrm{W\,m^{-2}}$ in AM4.0. Regionally, the improvement includes the Southern Ocean, the ITCZ, the broad subtropical and middle latitude ocean as well as modest improvements in the coastal stratocumulus regions.

Motivated by Tsushima and Manabe (2013), we show in Figure 7 the seasonal cycle of global mean TOA fluxes in AM4.0 and in observations for the total sky, clear sky, and the difference between total sky and
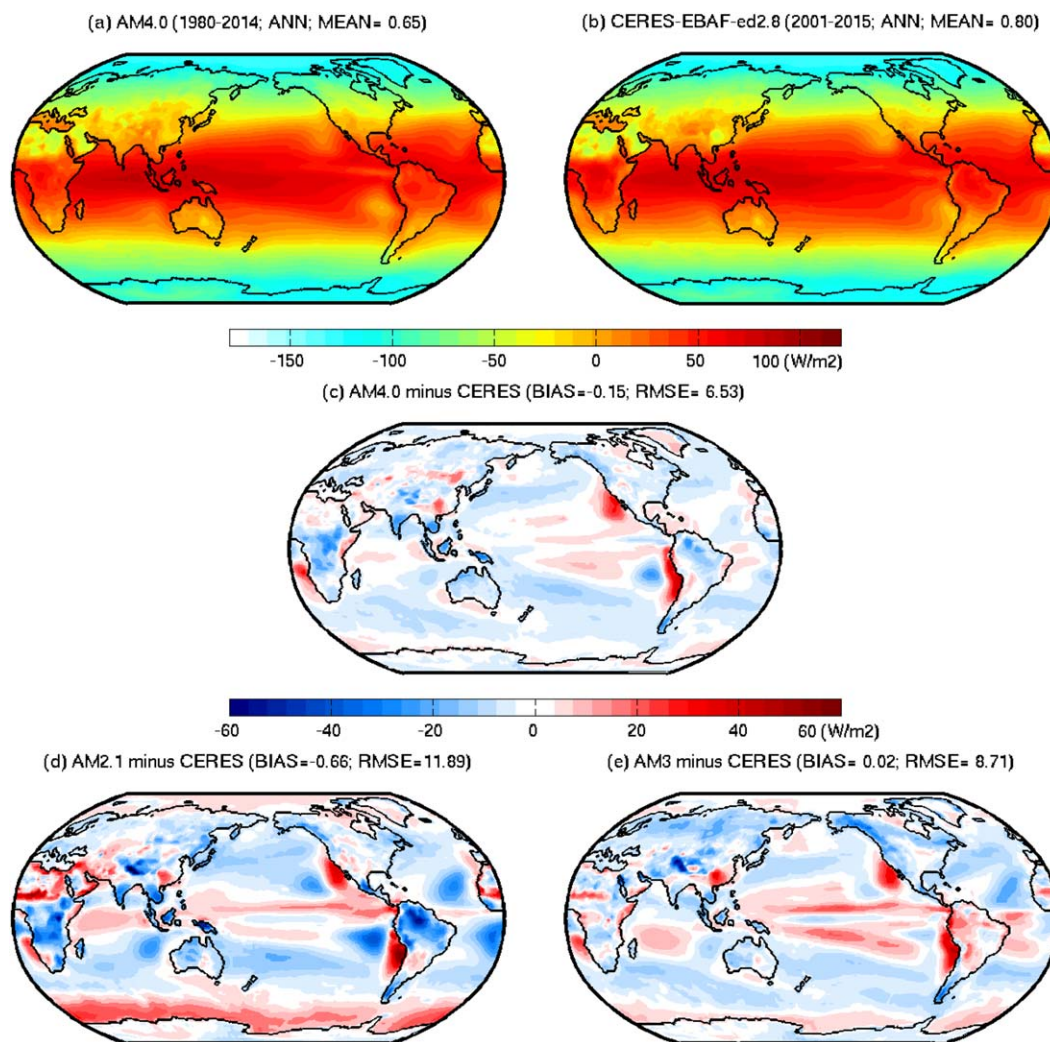
Figure 6. As in Figure 3 except for TOA net radiative flux (positive values indicate a downward flux).

clear sky, the cloud radiative effect (CRE). The amplitudes of net, shortwave (SW), and longwave (LW) total-sky fluxes are well simulated overall, but there is a modest but significant bias in Northern summer and autumn, with the model producing more SW reflection in July, August, and September, but less reflection from November to December compared to the observations. This error is partly due to the clear-sky component and partly to clouds. In particular, the model produces roughly $\sim 1.5\,W\;m^{-2}$ more negative SW CRE in July and August and $\sim 1.5\,W\;m^{-2}$ less negative SW CRE in November–December compared to the CERES observations. The geographical pattern of this bias, as can be seen in supporting information Figure S2, is not straightforward, as there is cancellation of errors in the global mean from biases in the subtropical stratus deck regions and the regions of deep convection.

To evaluate the model's ability in simulating interannual variability of global TOA radiative fluxes, Figure 8 compares deseasonalized 12 month running mean time series of the global mean TOA net radiative fluxes between the AM4.0 simulation and the CERES observations for the 2000–2014 period. The model captures reasonably well the temporal evolution of the global mean TOA flux with a correlation coefficient of roughly 0.8. We have generated eight AMIP simulations of this period and the shading in the figure indicates the total spread of this ensemble. One would expect the single observational realization to be outside the spread of an eight-member ensemble about 23% of the time for a perfect model, so there is little evidence here of systematic bias. The strong correlation comes from both LW and SW components. The correlation in the LW component comes primarily from the clear sky while the correlation in the SW component results
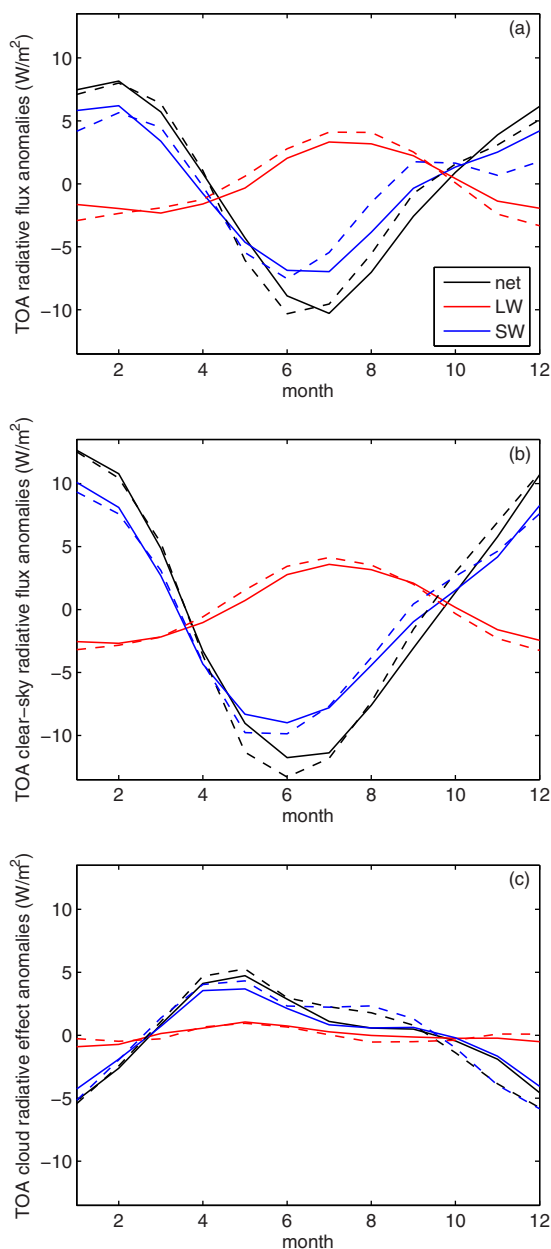
Figure 7. (a) Seasonal cycle of TOA net radiative flux (black) and its partitioning into the SW (blue) and LW (red) components from the AM4.0 AMIP simulation and the observations. Solid lines: AM4.0; dashed line: observational estimates based on CERES-EBAF-ed2.8; unit: W m^{-2}. Positive values indicate downward flux anomalies. (b) As in Figure 7a except for clear sky. (c) As in Figure 7a except for cloud radiative effects.

from the SW CRE as shown in Figure 8d. Compared to the SW CRE, the model simulated LW CRE shows significantly lower correlation ($\sim$0.33) with the CERES observations.

The simulated in-cloud liquid droplet number concentrations are compared with Moderate Resolution Imaging Spectroradiometer (MODIS) retrievals (Bennartz & Rausch, 2017) in Figure 9. To facilitate the comparison, the model results are averages between 850 hPa and the surface as the retrievals are weighted toward low-level clouds (R. Bennartz, personal communication, 2017). AM4.0 simulated cloud droplet number concentrations are respectively 66.3, 60.5, 80.4, 74.6, and 59.6 cm^{-3} averaged over the entire globe, the global ocean, the global land, the northern and Southern Hemisphere. Compared to the MODIS retrieval, AM4.0 appears to significantly underestimate the averaged droplet number over the ocean ($\sim$20%) although the observational uncertainty is quite large (Bennartz & Rausch, 2017). Despite this, AM4.0 roughly captures the spatial distribution of the droplet number. The midlatitudes generally see more droplets due to anthropogenic aerosols (sulfate and organic carbon) in the Northern Hemisphere and wind-driven sea salt aerosol in the Southern Hemisphere. In the tropics, droplet number concentration decreases gradually while moving through different cloud regimes (stratocumulus, trade cumulus, and deep cumulus), a pattern shared by the model and retrievals. AM4.0, however, underestimates droplet number concentrations downwind of the aerosol emission sources (e.g., East Asia and North America) and over the stratocumulus regions (e.g., Southeast Pacific).

To assess AM4.0 simulated cloud properties, Figures 10a and 10b show the climatological distribution of global mean cloud fraction as a function of cloud top pressure (p_t) and cloud optical thickness (τ) generated by the ISCCP (International Satellite Cloud Climatology Project) and the MODIS simulators in AM4.0. (Both simulators were implemented in AM4.0 as part of the CFMIP [Cloud Feedback Model Intercomparison Project] Observation Simulator Package; Bodas-Salcedo et al., 2011). For comparison, the ISCCP and the MODIS observational retrievals are presented in Figures 10c and 10d, respectively. The two satellite products differ significantly especially for the optically thin ($0.3 < \tau < 3.6$) clouds due to their differences in cloud masking and retrieving methods (Pincus et al., 2012). The pattern correlation between the two products is 0.47 for $\tau > 0.3$. We ignore clouds with optical thickness smaller than 0.3 for the correlation calculation here and below because it is not detectable by ISCCP and MODIS remote sensors. Despite the large differences there are also some similarities between the two satellite products such as the peak of low cloud fraction at cloud top below 680 hPa and $3.6 < \tau < 23$. In contrast to the observations, the model results from the two simulators tend to show more agreement with a pattern correlation about 0.74.

Compared to both ISCCP and MODIS observations, AM4.0 reasonably well captures the low cloud fraction peak below 680 hPa with intermediate optical thickness ($3.6 < \tau < 23$). However, AM4.0 tends to underestimate the optically thin low clouds and overestimate the optically thick low clouds especially when compared to ISCCP. This is consistent with the "too-few too-bright" low cloud bias typically seen in a global climate model (GCM). In addition, AM4.0 tends to produce less mid-top clouds (310 hPa $< p_t <$ 680 hPa) for most intermediate cloud optical thickness ($1.3 < \tau < 23$). The lack of midlevel cloudiness is also a typical bias seen in most GCMs. Thus, the ISCCP and MODIS observations indicate fewer low-level and middle level
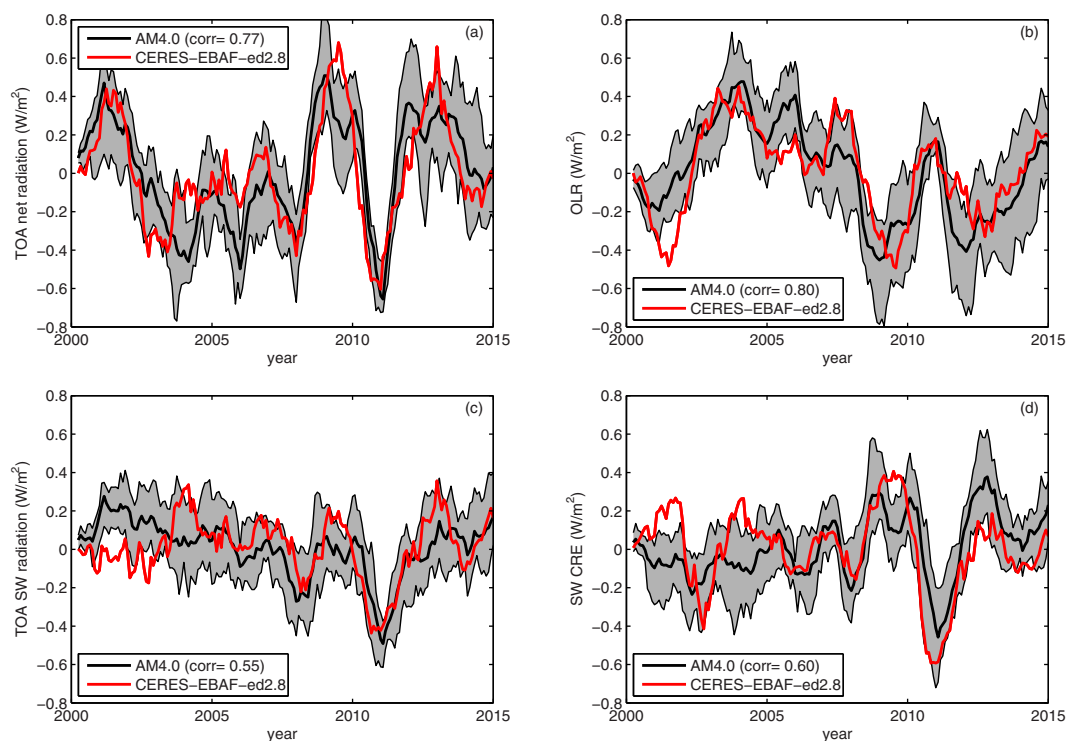
Figure 8. Deseasonalized time series of global mean TOA (a) net radiative fluxes and its partitions into (b) LW, (c) SW components from eight AM4.0 AMIP simulations and the observations based on CERES-EBAF-ed2.8. Black line shows eight-member mean; shading shows the maximum and minimum values of the ensemble. (d) As in Figure 8c except for SW CRE. All time series are first deseasonalized and smoothed by 12 month running mean before taking the ensemble average, maximum, and minimum values.

clouds in AM4.0. Globally, AM4.0 appears to underestimate the total cloud fraction by roughly 5% for both simulators. The MODIS observation has 14% lower total cloud fraction than the ISCCP because of a more stringent test for cloudy pixels (Pincus et al., 2012). In general, the large differences between ISCCP and MODIS products also make it difficult to constrain cloud biases quantitatively. For example, compared to the ISCCP, AM4.0 appears to underestimate the optically thin ($0.3 < \tau < 3.6$) clouds at most levels. But this bias does not show up clearly in the MODIS results. The pattern correlations between AM4.0 and the observations are respectively 0.56 and 0.74 for the ISCCP and MODIS, which are higher than the correlation between the two satellite observations (0.47). As discussed in Part 2, because of the large observational uncertainty in individual cloud properties (i.e., cloud amount, liquid/ice water content, and cloud droplet number), we have considered an improvement of the TOA radiative flux as a higher priority during the AM4.0 development. Nevertheless, limited observational evaluations of the individual cloud properties suggest AM4.0 performs at least as well as the previous GFDL models.

5. Precipitation

Since a model's ability in simulating the global distribution of precipitation given observed SST and sea-ice distribution is important to the model's overall quality in simulating the atmospheric general circulation, temperature, humidity, and clouds, we have considered it as a high priority during model development and evaluation. Annual mean precipitation in AM4.0 is compared to the observational estimates based on the GPCP-v2.3 and to the corresponding AM2.1 and AM3 simulations in Figure 11. The global mean precipitation in all of the models is higher than in GPCP-v2.3 by about 10%. The reliability of this observational estimate remains a subject of controversy (e.g., Gehne et al., 2016). The ERA-Interim reanalysis produces global mean of 2.92 mm d^{-1}, close to the AM4.0 value of 2.96 mm d^{-1}. From experience with different versions of this model, we conclude that it would be very difficult to create a version with as weak a global mean hydrological cycle as in the GPCP-v2.3 data set. The most natural way of reducing the strength of the
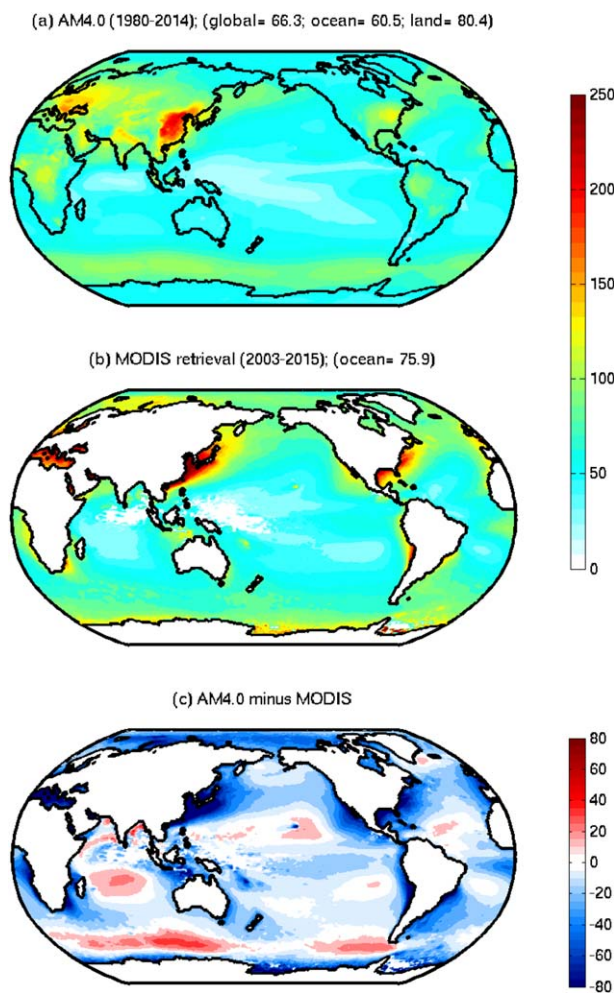
Figure 9. (a) AM4.0 simulated cloud liquid drop number obtained by cloud amount weighted average below 850 hPa (unit: cm^{-3}). (b) Observational estimate of the spatial distribution of cloud liquid drop number based on MODIS retrieval (Bennartz & Rausch, 2017). (c) The difference between model and the MODIS retrieval.

hydrological cycle without major changes in the net TOA balance is to increase atmospheric absorption of solar radiation at the expense of surface absorption, but we have no observational justification or plausible physical mechanism for doing so.

If we normalize AM4.0's precipitation to equal that of GPCP-v2.3, in the global mean, it would still have excessive precipitation in the West Pacific, sometimes referred to as the "Philippines hotspot" bias. This bias is sensitive to the model's convective parameterization. Reducing the lateral mixing of the parameterized convection can ameliorate this bias, but with consequences for other aspects of the model simulation, as described in Part 2. In some cases at least, this bias is reduced when the model is coupled to a dynamic ocean (as in, e.g., Stan & Xu, 2014). We also have suggestions that coupled versions of this model can be negatively affected if too much precipitation is moved from the West Pacific into the eastern Pacific ITCZ in AMIP simulations. Therefore we have not made it a priority to reduce this bias. AM4.0's precipitation biases are still modest compared to that in the bulk of the CMIP5 AMIP models (see PR column in Figure 2). In particular, AM4.0 simulated climatological PAI discussed in section 4 is 0.23, which is very close to the observed value 0.2 (Xiang et al., 2017).

The DJF and JJA seasonal mean precipitation corresponding to this annual mean plot are provided in supporting information Figures S9 and S10. The positive global mean bias compared to GPCP-v2.3 is present in all seasons. An interesting bias in DJF is excess precipitation in the North Pacific storm track, which is present in AM3 as well. We do not have a diagnosis of the cause of this bias. The western Pacific rain bias,
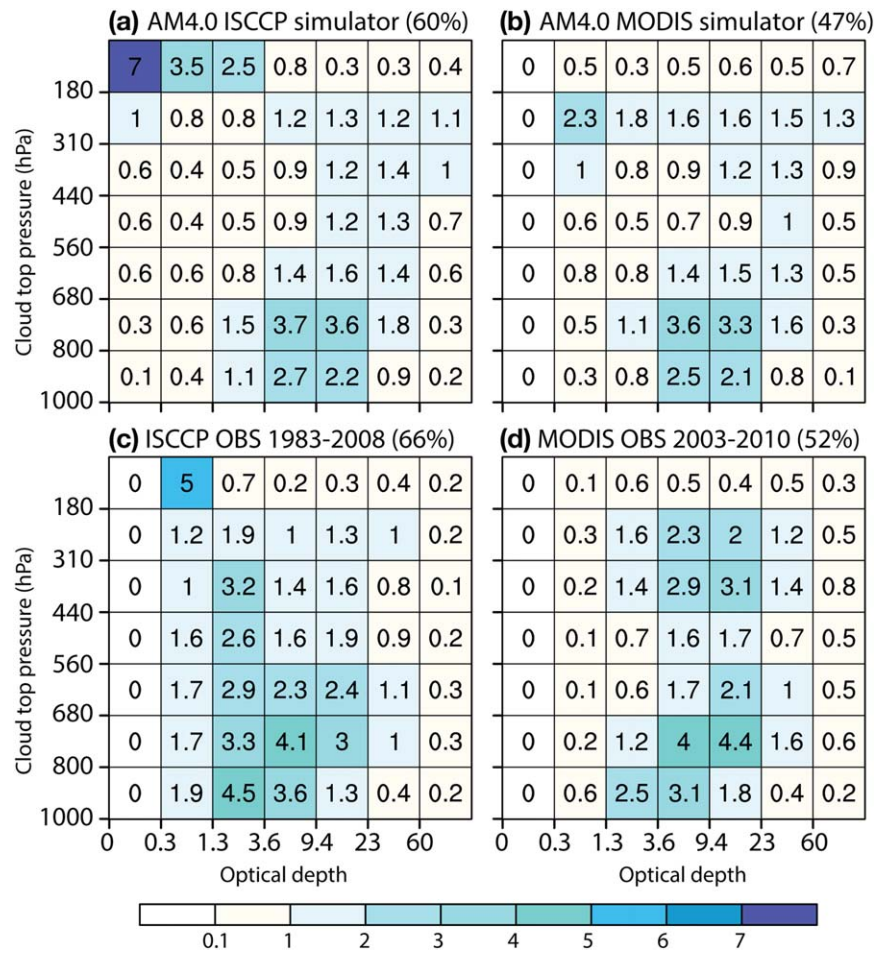
Figure 10. Climatological distribution of global mean cloud fraction as a function of cloud top pressure p_t (ordinate) and cloud optical thickness τ (abscissa). (a) AM4.0 climatology generated from the ISCCP simulator from the 1980–2014 AMIP simulation. (b) As in Figure 10a except from the MODIS simulator. The satellite observations from the (c) ISCCP (1983–2008) and (d) MODIS (2003–2010) climatology. The global mean cloud fraction (sum of each figure) is shown in the parentheses. High topped and optically thick clouds are represented in the top right corner, while low topped and optically thin clouds are represented in the bottom left corner.

especially the Philippines hotspot, is most apparent in JJA, and there is no obvious improvement over AM3 in monsoonal rainfall. The Eastern Pacific precipitation is improved (e.g., less double ITCZ) in both DJF and JJA seasons in AM4.0, while a bias toward too wet a winter season in Western North America (more apparent with a different choice of contour interval) persists.

In the tropical Atlantic and south America sector, AM4.0 significantly improves the precipitation dry bias over the South America Amazonian region as well as the summertime Atlantic ITCZ location and intensity compared to AM2.1 and AM3. During boreal summer, most GCMs tend to underestimate the northward shift of the tropical Atlantic rain belt, leading to deficient precipitation over land and an anomalous precipitation maximum over the west Atlantic ocean (Siongco et al., 2015). AM4.0 performs quite well in this aspect with little west Atlantic precipitation biases. The dry bias in central North America, which improved in AM3 over AM2, is somewhat more severe in AM4.0 for reasons that are unclear at present. When AM4.0 is coupled with an ocean, this central North America dry bias tends to be largely reduced, while the Amazonian region gets drier.

Runoff is an informative indicator of water and energy balances of both land and atmosphere. In the absence of long-term storage change, runoff equals the long-term difference between precipitation and evapotranspiration. The global pattern of basin-average runoff, as inferred from long-term discharge measurements, is reproduced by the model (Figure 12). Errors in the large tropical basins (Amazon, Parana, and
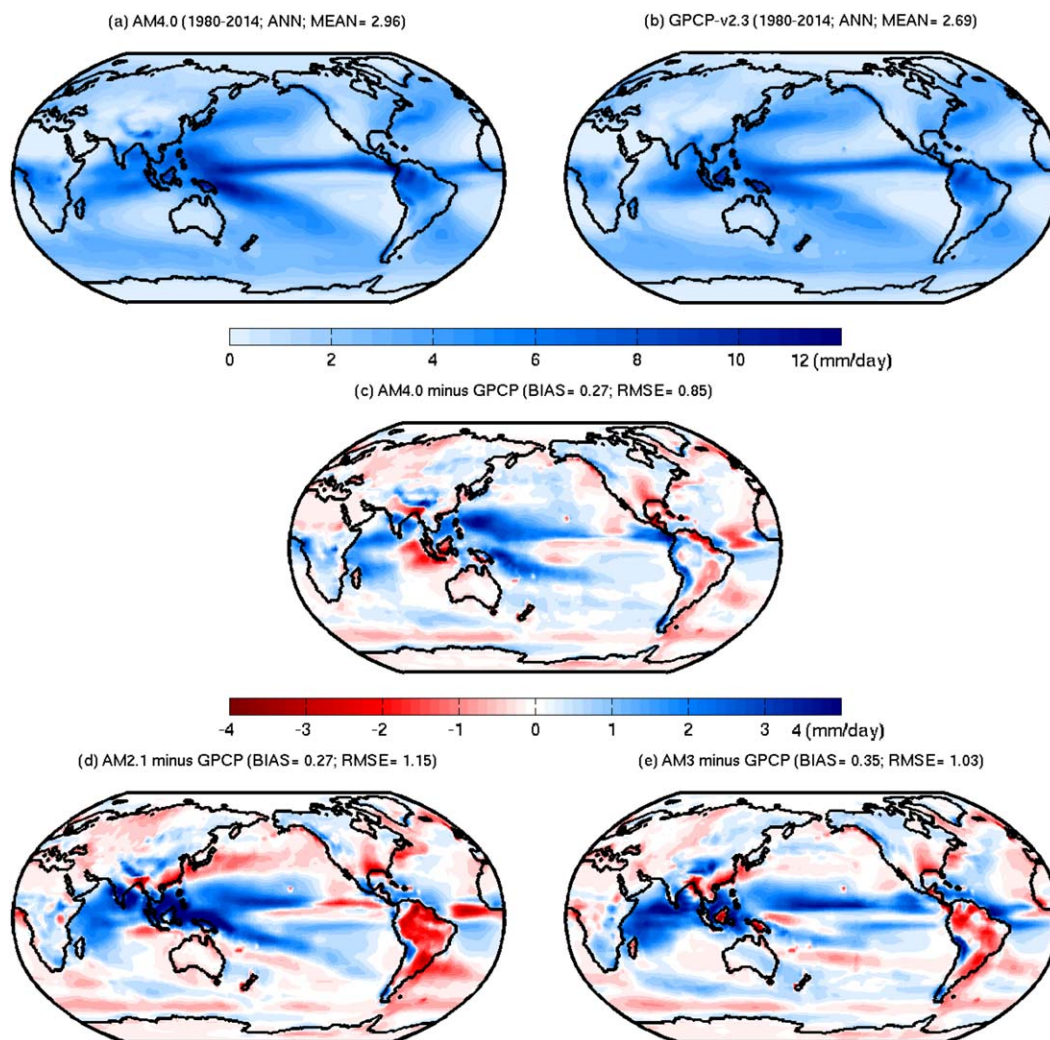
Figure 11. Long-term annual mean precipitation in mm d^{-1} from (a) AM4.0 AMIP simulation and (b) observational estimate based on GPCP-v2.3, averaged over the 1980–2014 period. (c) Model biases (AM4.0 minus GPCP-v2.3). (d) As in Figure 11c except for AM2.1. (e) As in Figure 11c except for AM3. Titles of Figures 11a and 11b show global mean values. Titles of Figures 11c–11e show global mean biases and RMS errors.

Congo) appear generally consistent with annual precipitation errors (Figure 11) in the model. In the northern high latitudes, however, the partitioning of precipitation appears to be biased away from runoff toward evapotranspiration. This has been a persistent bias in the model and is the subject of ongoing investigation. Note that runoff provides a particularly sensitive measure of water-balance error, since it is often the small difference between relatively large values of precipitation and evapotranspiration.

A persistent bias in the diurnal cycle of precipitation over land remains in AM4.0, with convection peaking near local noon, much earlier than the observations. Figure 13 shows an example of this bias over the United States. While the observed precipitation often peaks in late afternoon or early evening, AM4.0 produces a precipitation peak near local noon. This phase-lock of convective rainfall to local noon has been a longstanding problem in GCMs (e.g., Dai, 2006). Global high (mesoscale) resolution models with explicit convection and GCMs with superparameterized convection often show significant improvement in the phase of the diurnal cycle of precipitation, but biases in other aspects of the precipitation simulation can remain large in such models, including the mean rainfall distribution and the magnitude of the diurnal cycle (e.g., Dirmeyer et al., 2012). So it is unclear what the ramifications of this bias are for other aspects of the simulation. One can speculate, however, that SW cloud feedbacks could be affected by diurnal biases. We
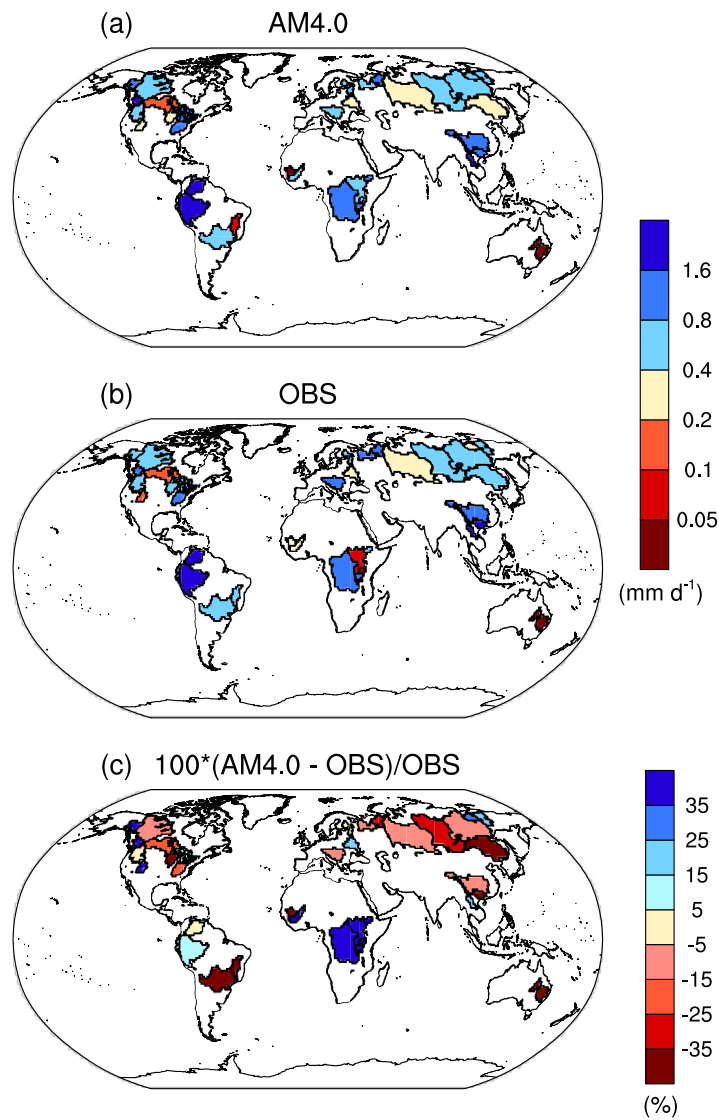
Figure 12. (a) AM4.0 modeled and (b) measured long-term, annual mean basin-mean runoff (mm d^{-1}) from gaged basins (Milly & Dunne, 2002). Note logarithmic color scale. (c) Estimated bias (%) in AM4.0 modeled runoff.

have postponed attempts to create a better diurnal cycle while maintaining a high quality mean precipitation distribution to future development, likely starting with higher resolution models.

6. Temperature

Since the annual mean surface air temperature tends to obscure important seasonal biases, we show the DJF and JJA seasonal mean 2 m temperature biases in Figures 14 and 15, using CRU-TS-v4.01 for the observations. (The annual mean figure can be found in supporting information Figure S11.) The RMS bias has improved somewhat in AM4.0 compared to the earlier models in both seasons, but the geographical pattern of bias is quite similar in all three models. In particular, in winter there is a large warm bias in Siberia and a warm bias in North America in a region arcing from Alaska to the U.S. upper midwest. Some aspects of these biases have been found to be sensitive to the details of the masking of vegetation by snow and to the choice of static vegetation and land use in these runs.

Assumptions concerning snow masking of vegetation and the associated albedo prescription for snow covered land areas can be important for albedo feedback in global warming simulations. In particular, Hall and
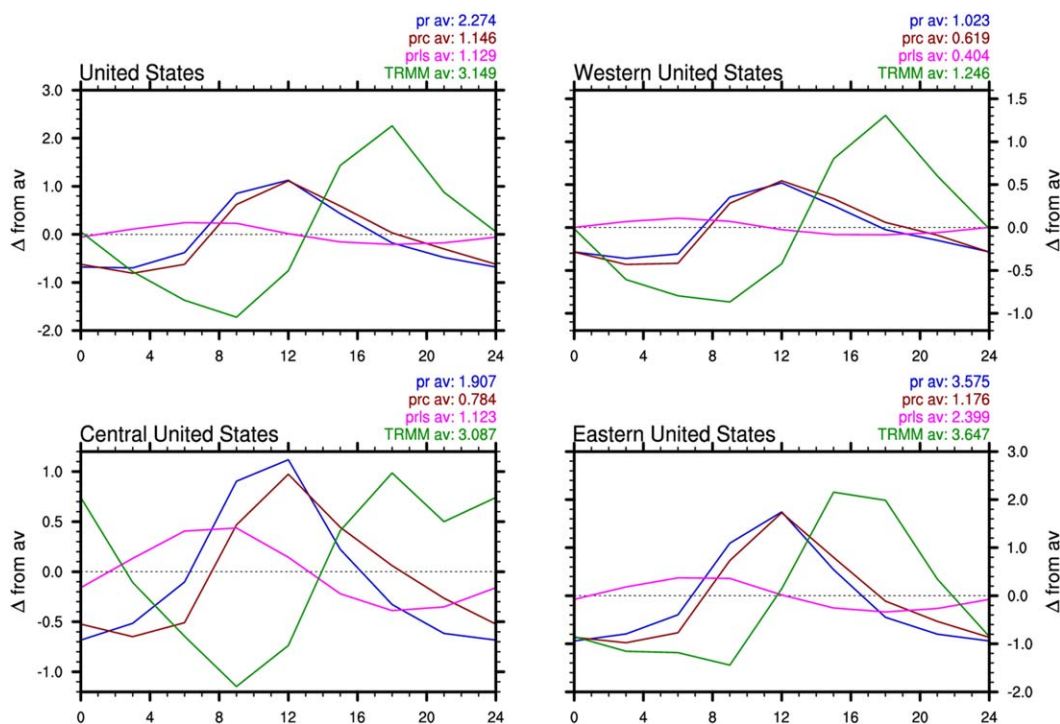
Figure 13. Comparison of AM4.0 simulated JJA seasonal mean diurnal cycle of precipitation over the United States with the observational estimates (green) from the TRMM data. Precipitation anomalies from daily mean (unit: mm d^{-1}) are plotted against local time. Model total precipitation (blue) is decomposed into convective (brown) and large-scale (pink) components. Legends show daily means from both model and observations. The US and the Western, Central, and Eastern U.S. are defined respectively as the land cover regions from 50°W–130°W, 10°N–50°N; 103°W–130°W, 25°N–50°N; 85°W–103°W, 25°N–50°N; 50°W–85°W, 25°N–50°N.

Qu (2006) have suggested that aspects of the seasonal cycle of land albedo and temperatures can be used as an "emergent constraint," a feature of the observed climate that is predictive of an aspect of climate change of interest—in this case, the strength of future albedo feedback over land. In particular, they define the constraint to be the difference between April and May in land surface albedo divided by the same difference in near-surface air temperature, both averaged from 30°N to 90°N, and find model values ranging over a factor of 3. The value in AM4.0 is −1.01%/K, as compared to the 95% confidence interval on the observational estimate of −1.01 to −1.13%/K in Hall and Qu (2006). The value in AM2.1 is very similar to that in AM4.0. (We caution that the Hall and Qu (2006) results [their Figure 3] are obtained from coupled model values for this albedo-temperature ratio, while we are using an AMIP simulation.) Accepting the value of this constraint, we would not expect the land albedo feedback in this model to be strongly biased.

The model has a relatively uniform cold bias at 2 m over the oceans of about 0.3 K, compared to ERA-Interim reanalyses, that is sensitive to the placement of the lowest level on which temperature is a prognostic variable (∼18 m in AM4.0, see supporting information Table S1 for AM4.0 vertical levels). The interpolation from this lowest model level to 2 m uses Monin-Obukhov similarity which assumes that this layer is embedded in the constant flux layer—the lower the first model level the better is this approximation. The height of this lowest model level was reduced during the development process, adding one additional layer and adjusting the remaining layer heights, to help ameliorate this bias.

The zonal mean temperature as a function of pressure and latitude is displayed in Figures 16 and 17 for DJF and JJA season. The observational estimates of the zonal mean atmospheric temperature are based on the ERA-Interim. There are smaller cold biases in the subpolar upper troposphere and lower stratosphere in all seasons in AM4.0 compared to AM2.1 and AM3, both of which have lower horizontal resolution. However, there is more of a cold bias in the tropical upper troposphere than in earlier models. This bias is sensitive to the convection parameterization; as described in Part 2, attempts to ameliorate this bias produce other problems in the model. It is also worth noting that for a fixed model physics, the tropical upper tropospheric
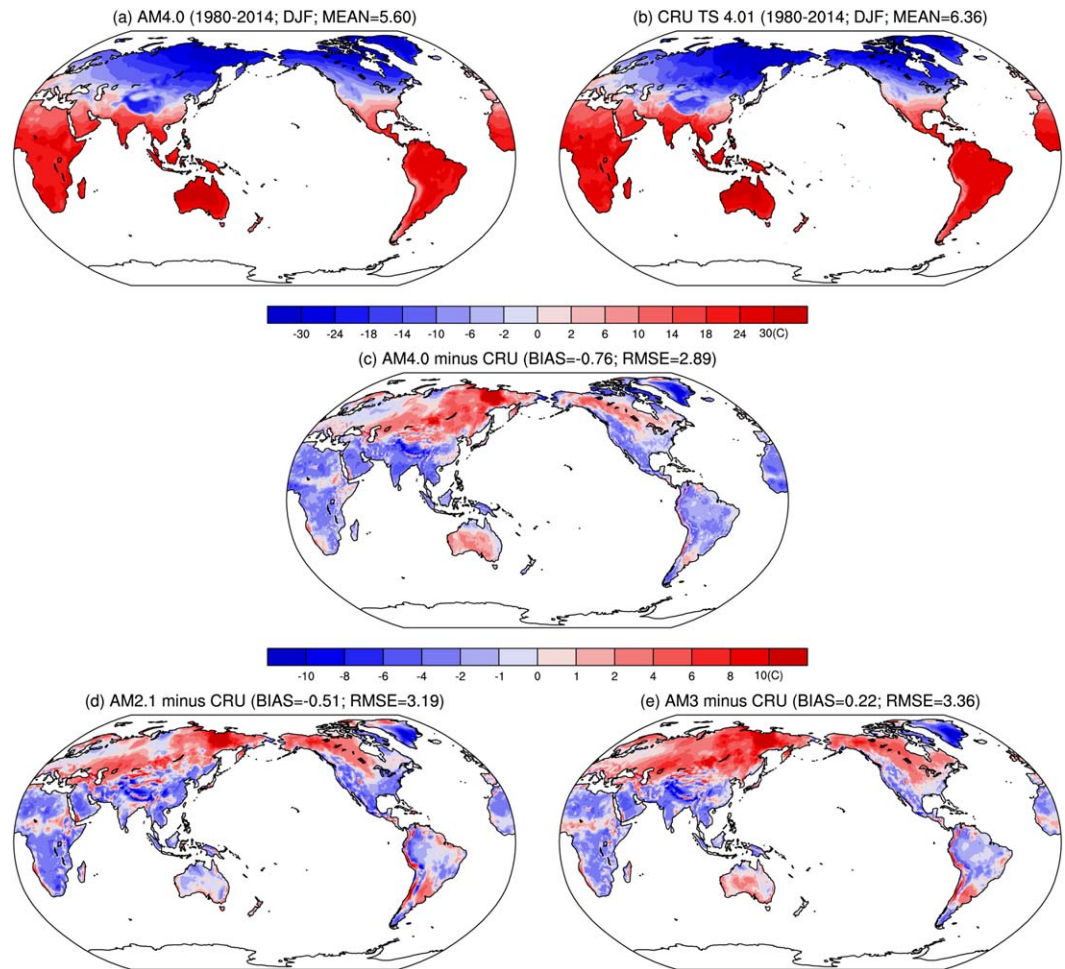
Figure 14. Long-term DJF seasonal mean 2 m temperature (°C) over land from (a) AM4.0 AMIP simulation and (b) observational estimate based on CRU-TS-v4.01, averaged over the 1980–2014 period. (c) Model biases (AM4.0 minus CRU). (d) As in Figure 14c except for AM2.1. (e) As in Figure 14c except for AM3. Titles of Figures 14a and 14b show the global mean values over land. Titles of Figures 14c–14e show global mean biases and RMS errors.

temperature tends to increase as the model's horizontal resolution increases. The warm spot in DJF near 200 hPa in AM2.1 and AM3 is related to the vertical distribution of the orographic gravity wave drag.

7. Mean Circulation

Similar to the temperature field, we show DJF and JJA seasonal mean zonal mean zonal wind in Figures 18 and 19. (Annual mean zonal mean zonal wind is shown in supporting information Figure S14.) Some of the improvements over recent GFDL models are presumably due to horizontal resolution. A common deficiency in GCMs is equatorward displacement of midlatitude westerlies, specifically in austral summer. We see no significant deficiency of this type in AM4.0, although it can recur in coupled versions of the model due to SST drifts.

A resilient bias in GFDL models has been excessive easterlies in the lower troposphere (800–900 hPa) over the equator, especially in JJA. Regionally, this bias occurs predominately in the central Pacific. In AM4.0, this bias does not strongly impact the surface stresses and, it is hoped, has little effect on ENSO dynamics in the coupled model. The bias is sensitive to the parameterized convective momentum transport, but alternative formulations of the convective momentum transport have not removed this bias, and can make it worse by extending the bias to the surface.
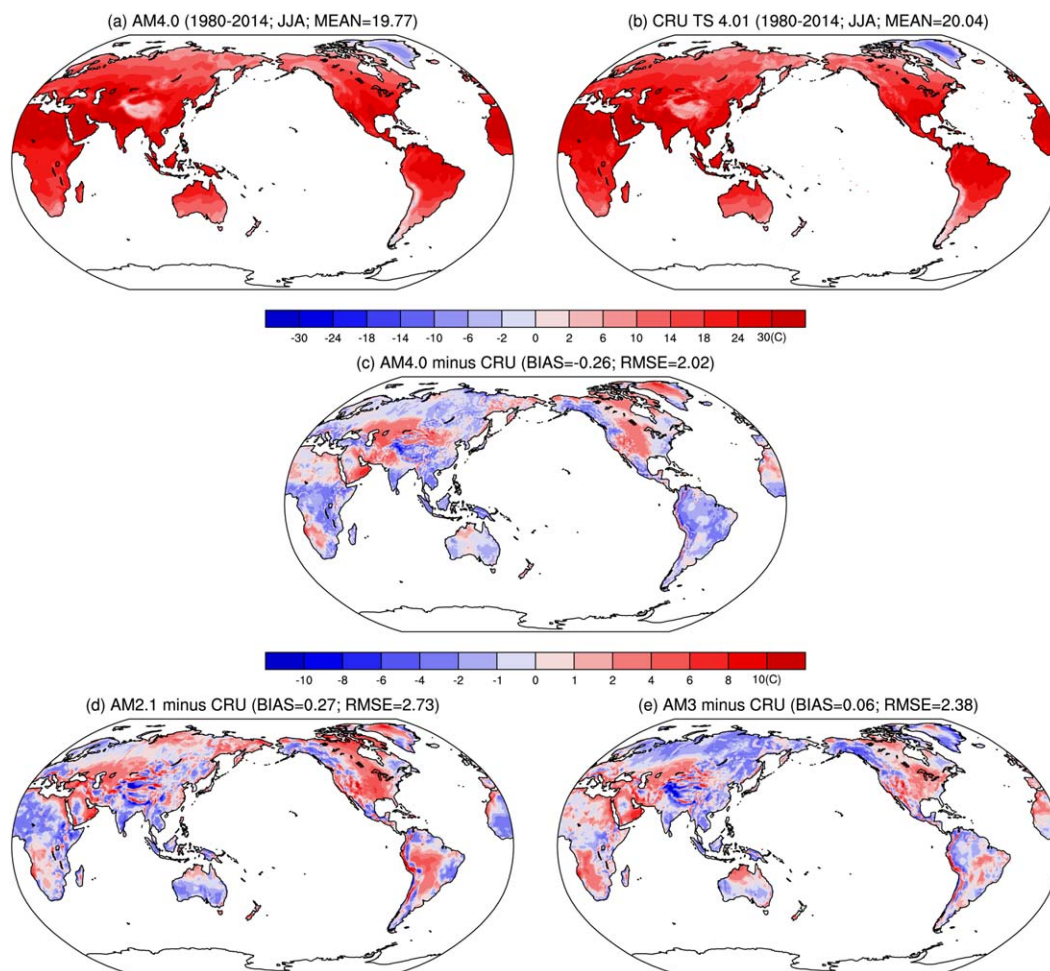
Figure 15. As in Figure 14 except for the JJA season.

AM4.0 also has a modest bias immediately above the midlatitude jet maximum in DJF, associated with the gravity wave drag, implying that the decomposition between low-level deposition due to blocked flow and the deposition immediately above the jet due to critical level absorption of mountain waves may be too heavily weighted to the latter. We did not further reduce the linear mountain drag in AM4.0 because it tends to increase the tropical lower tropospheric easterlies which are already too strong. The simulations of stratospheric wind and temperature can be found in supporting information Figures S15–S22.

Figure 20 compares the annual mean zonal wind stress exerted by the atmosphere on the surface over the Pacific Ocean, the Atlantic Ocean, and the land in AM4.0 with two reanalyses and with the spread of AMIP simulations in CMIP5. The bias and large spread in the CMIP5 ensemble in the location of the maximum positive stress over the Southern Hemisphere oceans is most evident in the Pacific sector, where AM4.0 agrees well with the reanalyses. Larger differences with the reanalyses appear in individual seasons (see supporting information Figures S23–S26), especially a bias toward too strong stress below the easterlies in the winter hemisphere. Despite this bias, in the annual mean the stress due to the trade winds is weaker than most of the CMIP5 models in the Pacific, and in good agreement with the reanalysis.

The Northern Hemisphere wintertime sea level pressure field is shown in Figure 21, and is one field in which AM4.0 shows little improvement over AM3. The figure is drawn to avoid regions in which a significant extrapolation to sea level from an elevated surface is required. Similar figures for other seasons (JJA, MAM, and SON) do show significant improvement (especially MAM and SON, see supporting information Figures S27–S29). Comparisons with an alternative reanalysis product (MERRA) provide a consistent picture. Our interpretation of the DJF bias is that it is in part related to the teleconnections from tropical precipitation
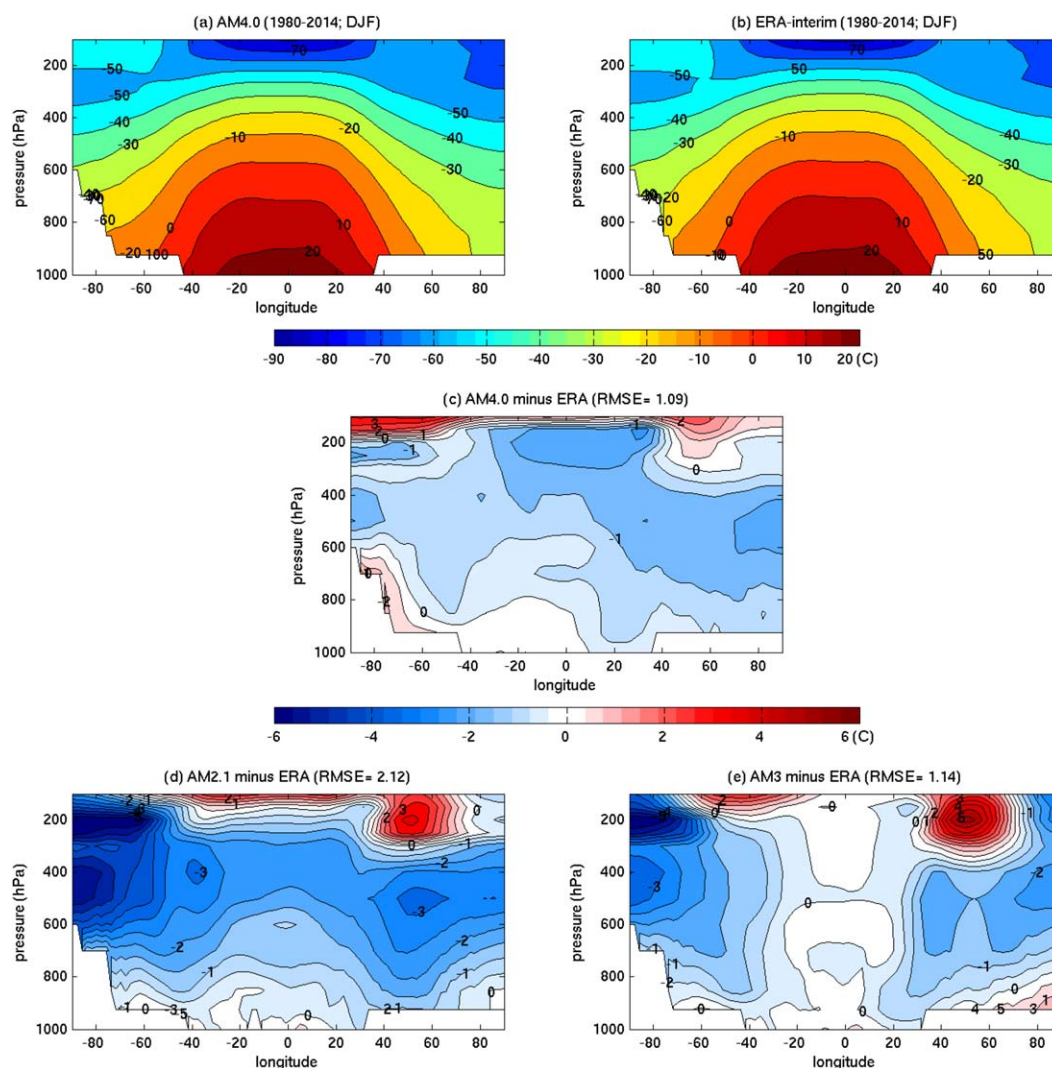
Figure 16. Long-term DJF season zonal mean troposphere temperature in Celsius from (a) AM4.0 AMIP simulation and (b) ERA-Interim (1980–2014). (c) Model biases (AM4.0 minus ERA-Interim). (d) As in Figure 16c except for AM2.1. (e) As in Figure 16c except for AM3. Titles of Figures 16c–16e show mass-weighted RMS errors. Model and reanalysis data are first interpolated to standard pressure levels (1,000 925 850 700 600 500 300 250 200 150 100 hPa) before taking the differences.

biases. In coupled simulations performed as part of the development process, the DJF sea level pressure simulation often improves, at least marginally, as does the precipitation bias in the western Pacific, consistent with the effects of tropical precipitation biases teleconnecting most strongly to the extratropical stationary wave pattern in winter.

A consistent picture also emerges from inspection of the Northern Hemisphere 500 hPa stationary wave patterns in the different seasons (see supporting information Figures S30–S33). AM4.0's 500 mb stationary wave patterns improve in all seasons compared to earlier GFDL models, except in winter. While this pattern is consistent across different realizations of AMIP runs of AM4.0, given the large interannual variability of stationary waves in winter we must keep in mind that the observations provide only a single realization.

8. Midlatitude Storm-Track Transients

Given the prominent role of the midlatitude baroclinic eddies to meridional transport of momentum, heat, and moisture and their tight interactions with climate variability and changes, we evaluate AM4.0's
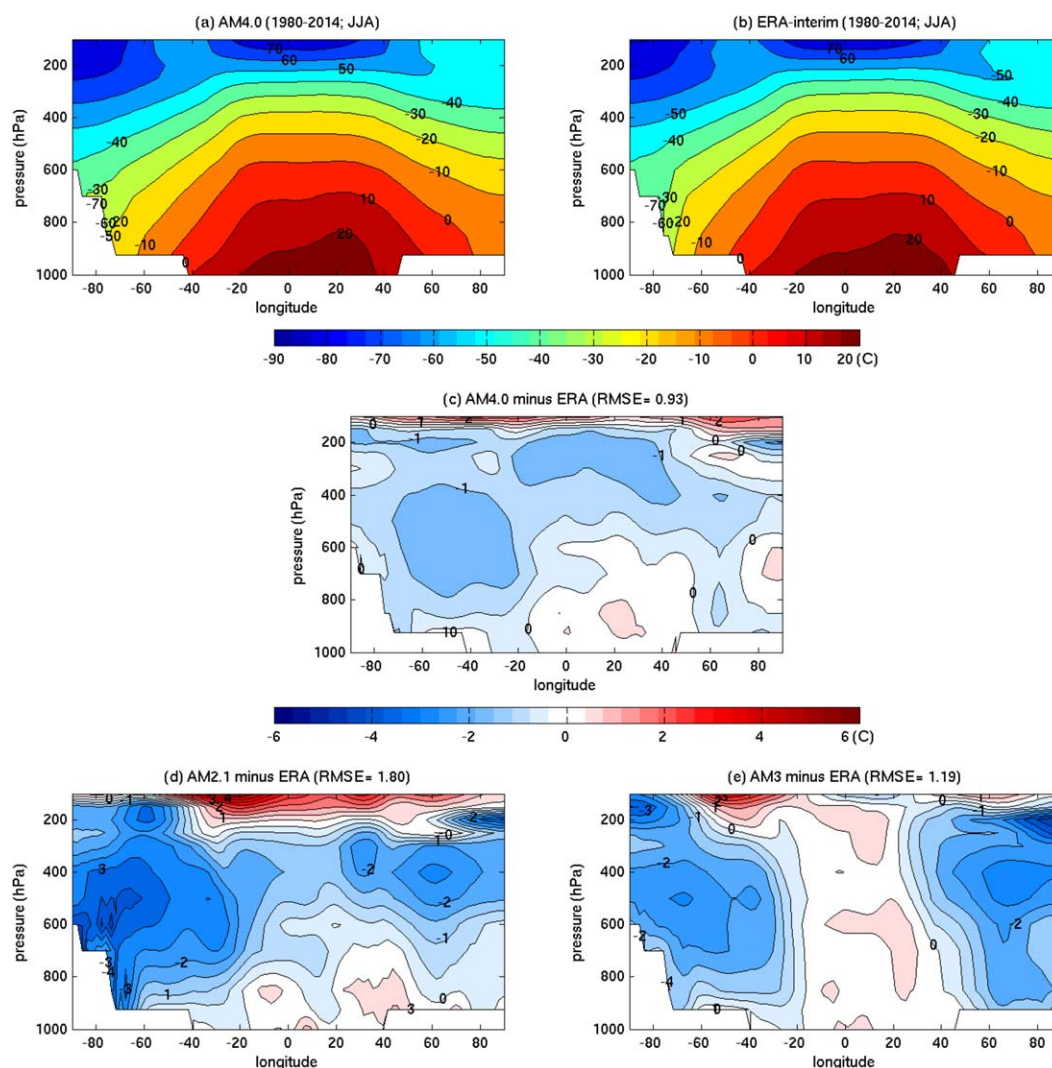
Figure 17. As in Figure 16 except for the JJA season.

simulation of the midlatitude storm track in Figures 22 and 23. The figures show the transient eddy momentum and heat fluxes in DJF and JJA seasons. The observational data are based on MERRA reanalysis. The velocities and temperatures are band passed using a Lanczos filter with half-power points at roughly 2 and 7 days, as is traditional when displaying storm-track eddies (e.g., Blackmon, 1976). (The corresponding unfiltered fields are displayed in supporting information Figures S34 and S35.) The fluxes are displayed at the pressure level near which they take on maximum values (250 hPa for the momentum fluxes and 850 hPa for the heat fluxes).

The basic pattern of the eddy momentum fluxes in the Northern Hemisphere (NH) winter is accurately simulated, with the poleward flux overestimated somewhat over the central North Atlantic and underestimated in Northern Europe. The strength of this band-passed poleward flux in DJF in the Southern Hemisphere (SH) is overestimated by as much as 10% as compared to the reanalyses. Biases in JJA are more subtle.

The eddy heat fluxes show a subtle equatorward shift in the North Atlantic in DJF, and once again no obvious systematic biases in JJA. Comparison with Southern Hemisphere reanalysis requires more detailed study to isolate any robust biases. We are hoping to examine these storm-track statistics, and other higher order moments and extremes, as a function of horizontal resolution in a follow-up paper. Our initial indications are that the improvements in these variance fields as one moves to higher resolution are smaller than the deterioration as one moves to lower resolution. But we presume that higher resolution is more significant for higher order moments and extremes.
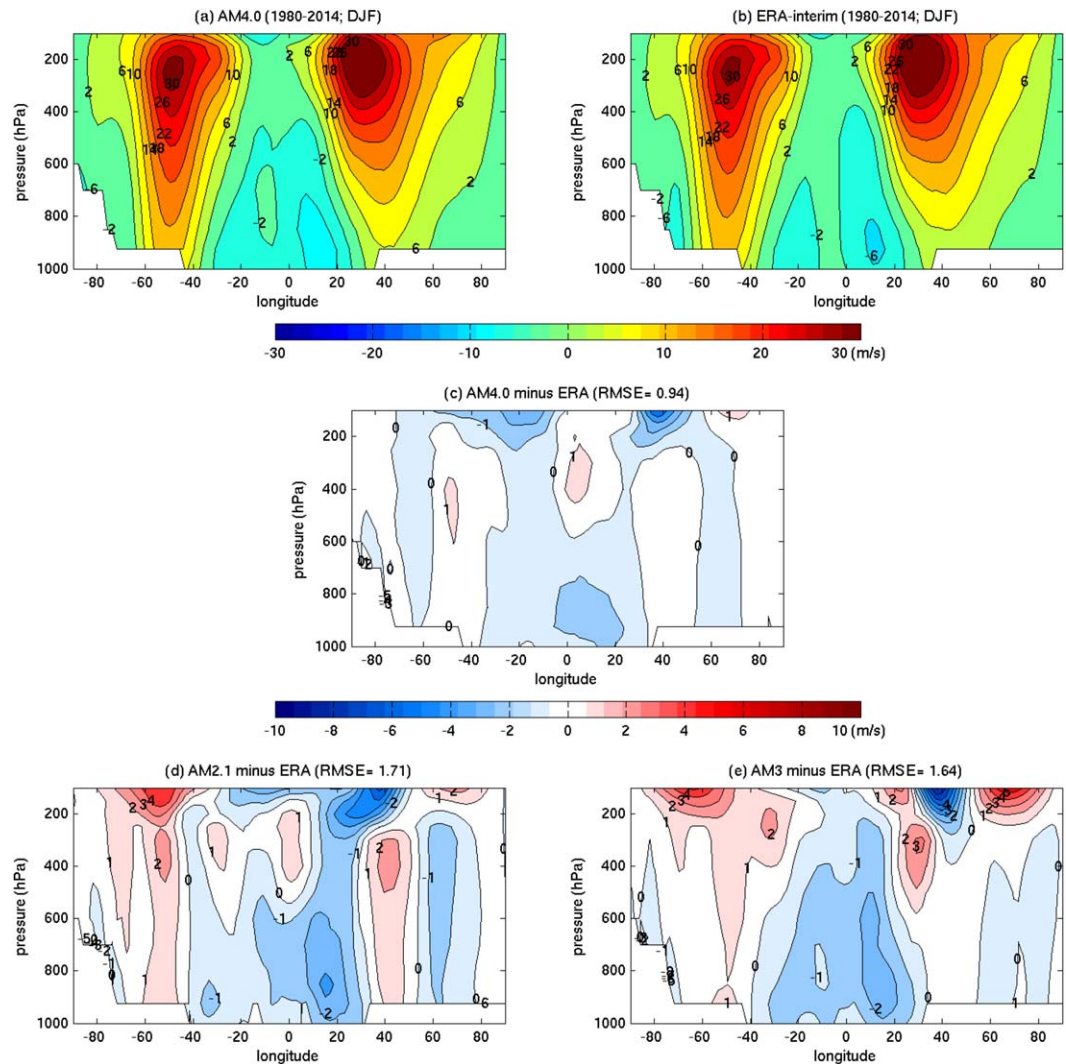
Figure 18. As in Figure 16 except for zonal mean zonal wind in m s^{-1}.

9. Tropical Transient Activity

Coarse resolution GCMs typically perform poorly in simulating tropical transient activities. GFDL AM2 and AM3 are no exceptions (Donner et al., 2011; GFDL-GAMDT, 2004). Given the improved horizontal resolution and model physics in AM4.0, we hope to improve the model's representation of tropical transient activity. We evaluate the tropical transient activity by focusing on two phenomena: tropical cyclones and the Madden-Julian oscillation (MJO; Madden & Julian, 1971).

Tropical cyclones in AM4.0 are detected using the same algorithm as that described in Zhao et al. (2009, Appendix B) except with two modifications for adjustment to a lower resolution GCM. In Zhao et al. (2009), a storm trajectory must last at least 3 days and have a maximum surface wind speed greater than 17 m s^{-1} during at least 3 days (not necessarily consecutive). We modified this wind speed threshold to be 14 m s^{-1}. In addition, the maximum surface wind speed over the entire trajectory must exceed 16 m s^{-1} to qualify as a model storm. The use of a lower wind speed threshold is recommended for this low resolution GCM by Walsh et al. (2007), but it should be recognized that the total storm counts displayed here are sensitive to these criteria. However, the spatial pattern of genesis as well as the seasonal distribution of counts within each ocean basin are less sensitive to the storm counting algorithm.

Figure 24 shows the distribution of annual tropical cyclone genesis frequency from the average of eight AM4.0 AMIP simulations for the 1980–2014 period. For comparison, the observations from the IBTRACS
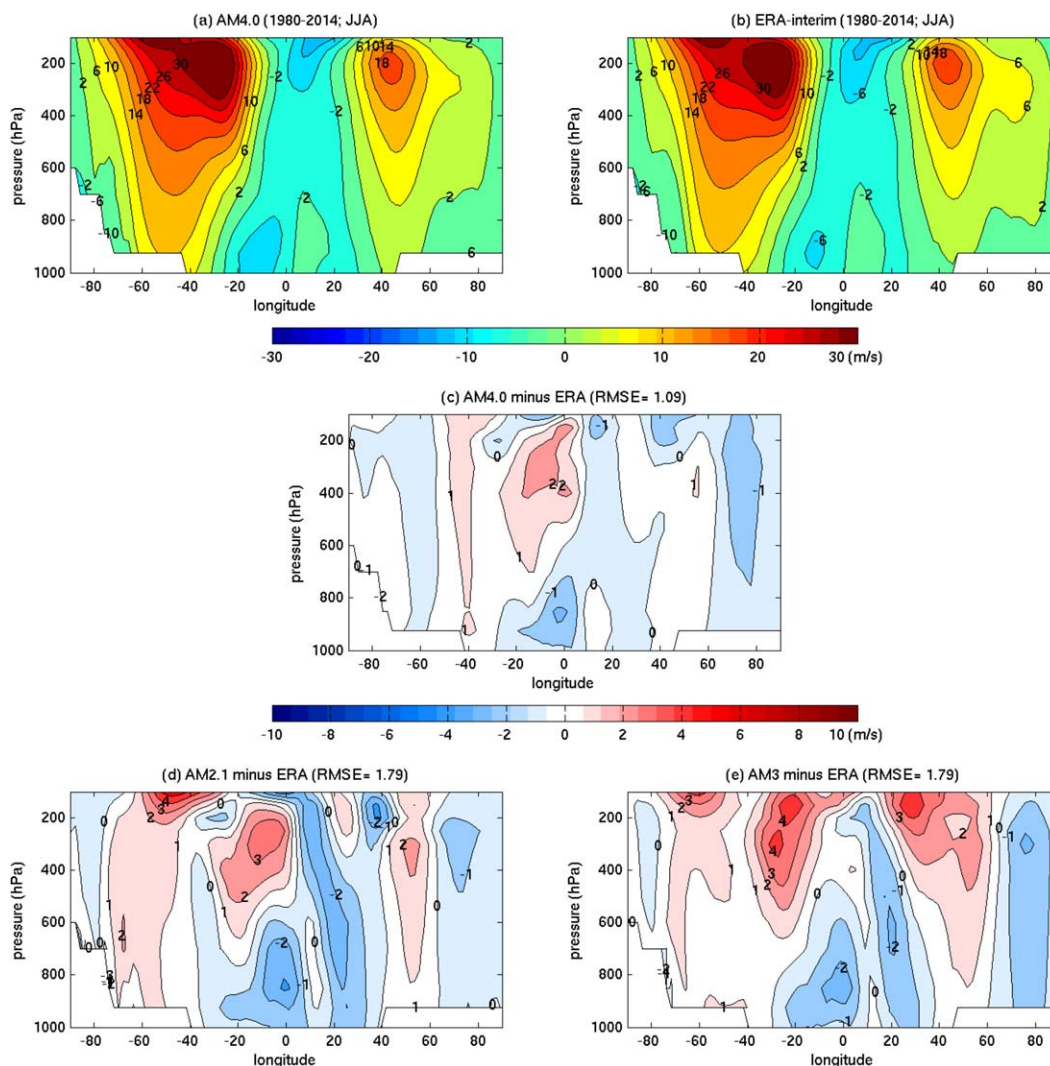
Figure 19. As in Figure 18 except for the JJA season.

(International Best Track Archive for Climate Stewardship) data set (Knapp et al., 2010) are also plotted. AM4.0 captures the broad geographical distribution of TC genesis frequency over different ocean basins. Compared to the IBTRACS, AM4.0 tends to overproduce TC frequency over the western Pacific and underestimate storms in the eastern Pacific and North Atlantic basins, a bias typical to most low resolution (i.e., ~100 km) GCMs (Walsh et al., 2015). AM4.0 captures the TC frequency over the Indian Ocean and South Pacific rather well.

AM4.0 also captures the seasonal cycle of TC frequency over each ocean basin reasonably well, as can be seen in Figure 25. In this figure the results are normalized to 1 for both the model and the observations in the annual mean in each basin. In the NH ocean basins, AM4.0 tends to produce a phase shift with excessive late season storm frequency and insufficient genesis in the early and peak seasons. The problem appears to be more severe over the West Pacific and the North Atlantic. The cause of the phase shift in NH tropical cyclone seasonal cycle is not yet clear.

Figure 26 further shows interannual variability of the annual tropical cyclone count over the North Atlantic, the East and the West Pacific. The model also reproduces some of the observed interannual variability with a correlation coefficient between the model ensemble mean and observation of 0.59, 0.58, and 0.49 for the North Atlantic, the East, and West Pacific, respectively. These correlations are significantly lower than some of earlier studies using higher resolution GFDL models (e.g., Chen & Lin, 2013; Murakami et al., 2015; Zhao
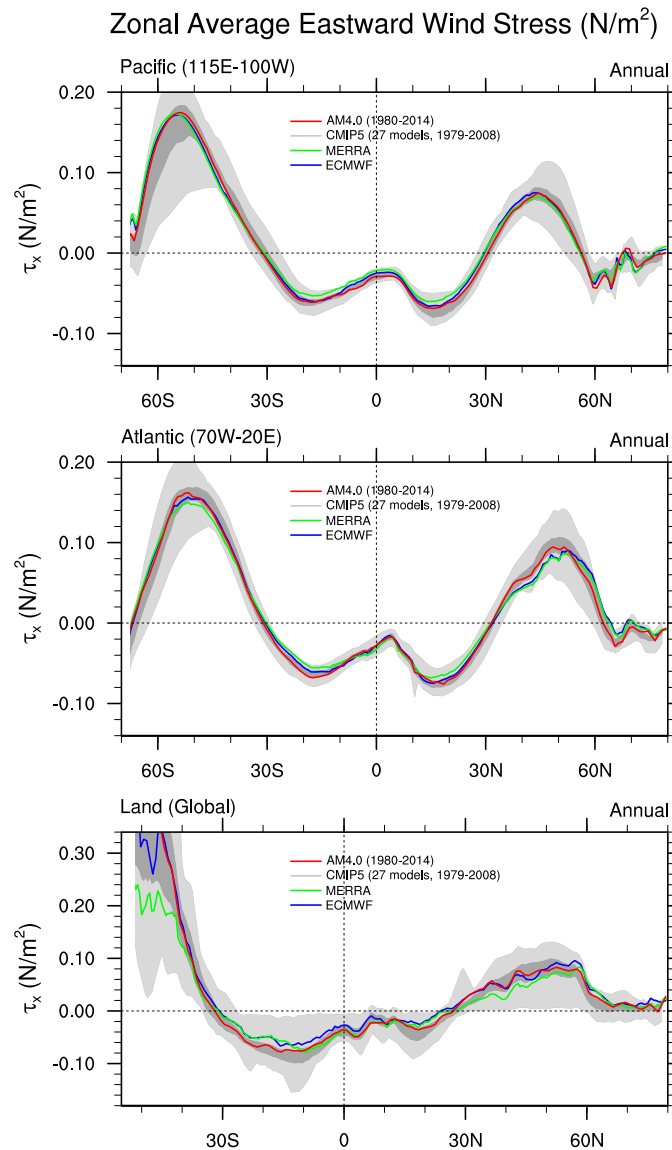
Figure 20. Comparison of annual mean zonal mean zonal wind stress over (top) the Pacific, (middle) the Atlantic oceans, and (bottom) the global land region. Colored lines show AM4.0 simulation and the observational estimates based on ERA-Interim and MERRA data (see legends). The shading shows ensemble spread from 27 CMIP5 models. The full spread of the CMIP5 ensemble is in light gray and the results between the 25 and 75 percentiles are in darker gray.

et al., 2009) but are better than most GCMs participated in the US CLIVAR Hurricane Working Group (Shaevitz et al., 2014).

Our overall impression is that the TCs in this model are physically meaningful and may provide useful input into questions regarding responses of TC genesis frequency to climate variability and change, despite the model's limited resolution. However, our impression is also that the TCs are too weak for there to be much useful information in the model's intensity distribution (unlike the 50 km HiRAM model for example, in which it does appear to be possible to partially adjust for the model intensity biases with quantile-quantile regression; Zhao & Held, 2010). Higher resolution versions of AM4.0 under development will hopefully be more suitable in this regard.

We evaluate the AM4.0 tropical wave spectrum in the format of Wheeler and Kiladis (1999). We show the tropical symmetric (15°S–15°N) power spectrum, with background removed following Wheeler and Kiladis (1999), from both the upper (200 hPa) troposphere zonal wind (U200) and OLR. We analyze both the AMIP
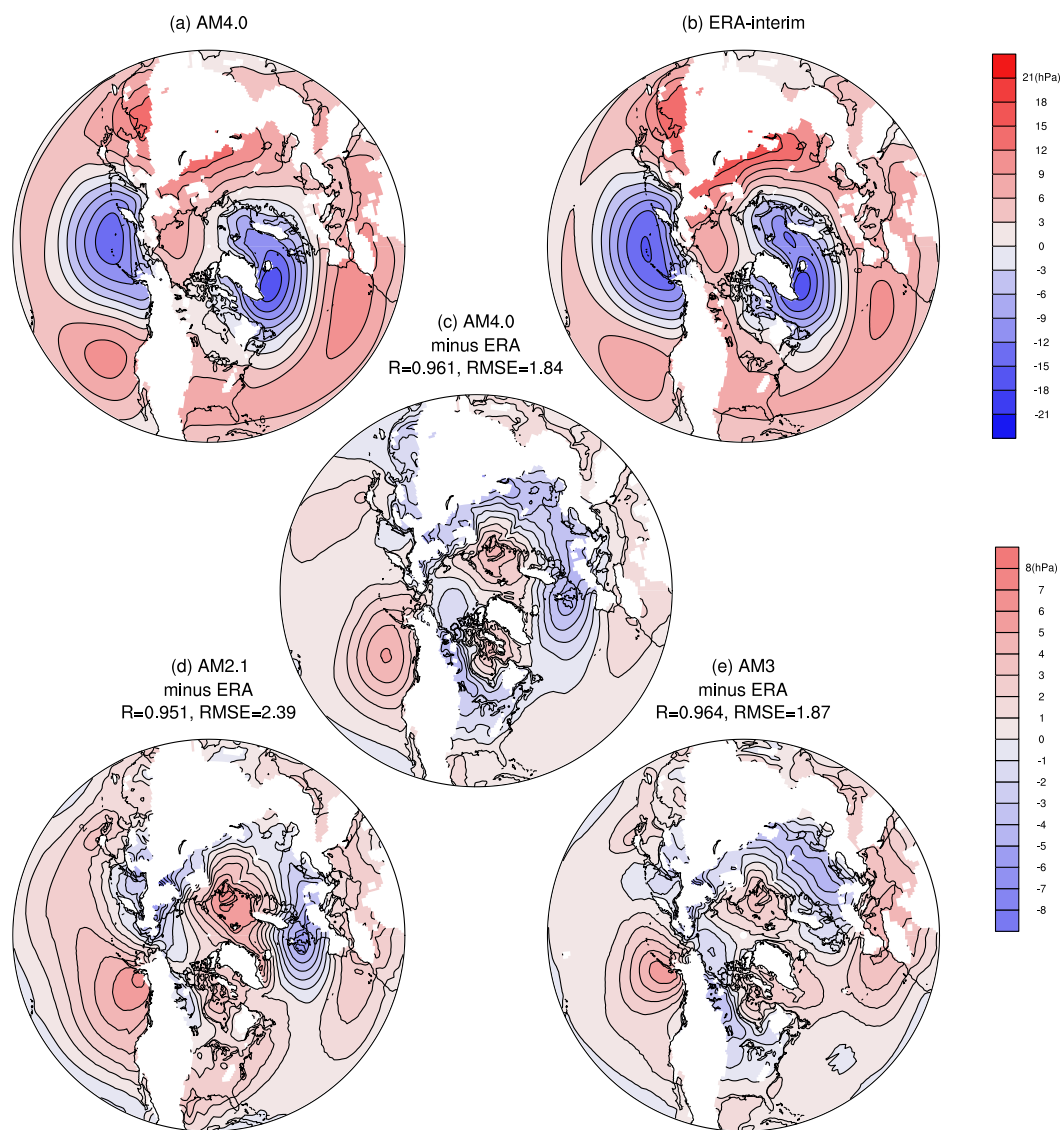
Figure 21. Long-term Northern Hemisphere DJF seasonal mean sea level pressure minus 1,013.25 hPa (contour intervals: 3 hPa) from (a) AM4.0 AMIP (1980–2014) simulation and (b) ERA-Interim (1980–2014). (c) Model bias (AM4.0 minus ERA-Interim; contour intervals: 1 hPa). (d) As in Figure 21c except for AM2.1. (e) As in Figure 21c except for AM3. Titles of Figures 21c–21e show spatial correlations and RMS errors. Sea level pressure is masked out where surface pressure is less than 950 hPa.

version of the model and a coupled version. Consistent with the literature (e.g., Waliser et al., 1999) coupling enhances tropical wave activity, especially the MJO (frequency = 0.025 day^{-1} and zonal wave number = 1–3) in the model. Whether there is any sensitivity to the near-surface mixing within the ocean model is not yet determined, but because this enhancement is found in all of our coupled simulations, we feel it is useful to show a coupled version here as well, as we do not feel that AMIP simulations alone are a reliable measure of the simulation of the MJO.

Figure 27 shows that AM4.0 captures very well the observed U200 power spectrum based on the ERA-Interim, including the Kelvin wave, the MJO and the westward moving variance, including some westward propagating inertial gravity waves. However, AM4.0 produces much weaker Kelvin wave signal and misses entirely the inertial gravity wave in the OLR spectrum, suggesting deficiency in coupling convective clouds and precipitation with the wind field for these waves. The coupled model enhances the OLR signal more than that in U200. In particular, the coupled version of AM4.0 captures quite well the observed MJO in the
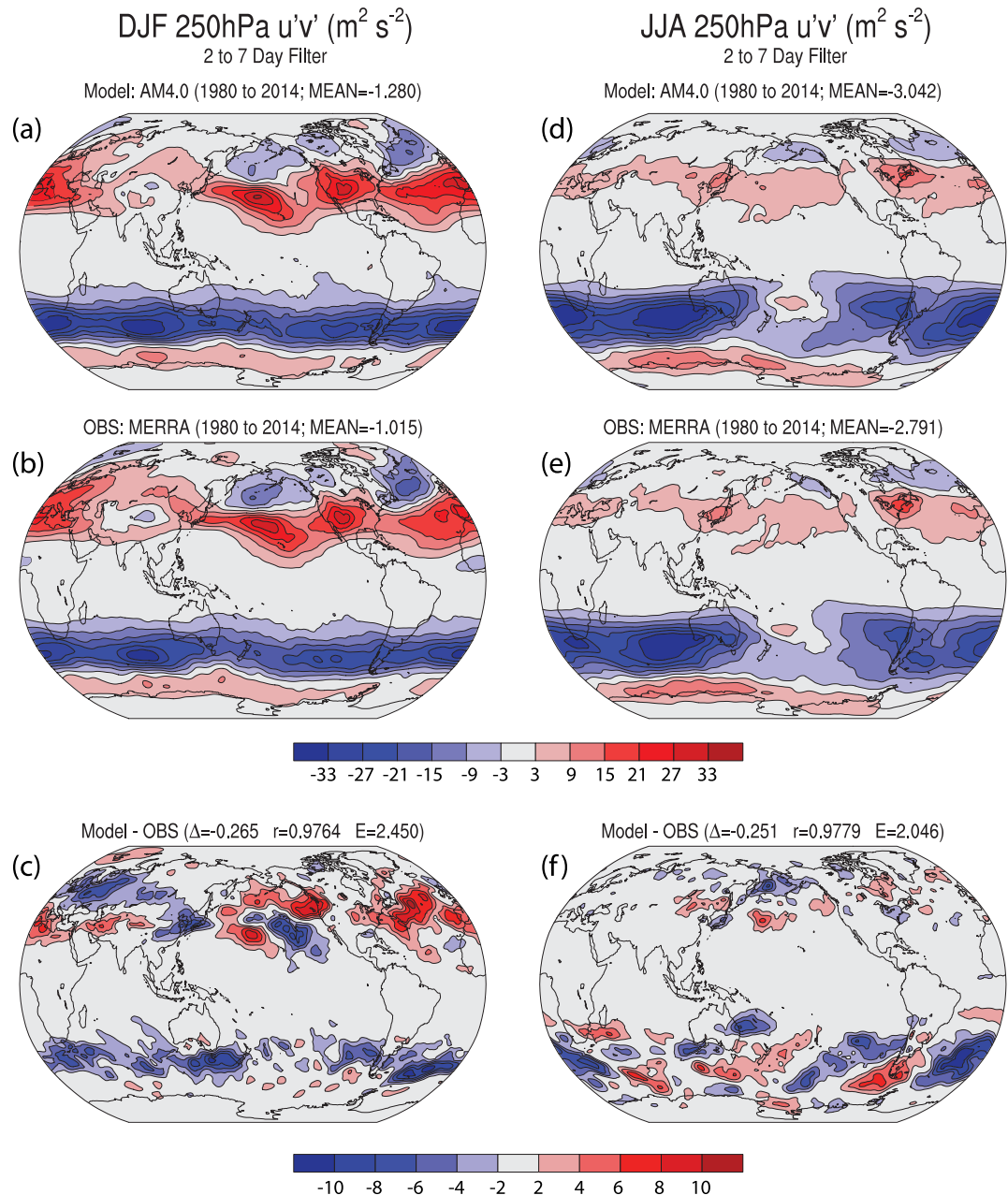
Figure 22. (a–c) Comparison of DJF season 250 hPa 2–7 day filtered eddy momentum flux u'v' (unit: m² s⁻²) between (a) AM4.0 AMIP simulation and (b) the MERRA reanalysis; (c) The differences (AM4.0 minus MERRA). The legend in Figure 22c shows the global mean bias (Δ), spatial correlations (r), and RMS error (E). (d–f) As in Figures 22a–22c except for the JJA season.

OLR spectrum. In addition, Figure 27 also shows that AM4.0 reasonably well reproduces the westward equatorial Rossby wave and the Rossby-Haurwitz wave signals which appear only in the wind field.

To illustrate the models' MJO further, Figure 28 shows the lag correlation between time series of central Indian ocean upper level zonal wind and OLR and the corresponding winds or OLR at all other longitudes. Propagation is clearly enhanced in the coupled simulation. In the zonal winds, AM4.0 shows well-defined rapid propagation east of the dateline, indicative of the excitation of dry Kelvin waves. Compared to the observation, the eastward propagation of the OLR signal is weaker as it passes through the Maritime continent even in the coupled simulation, suggesting a stronger barrier effect from the Maritime continent in
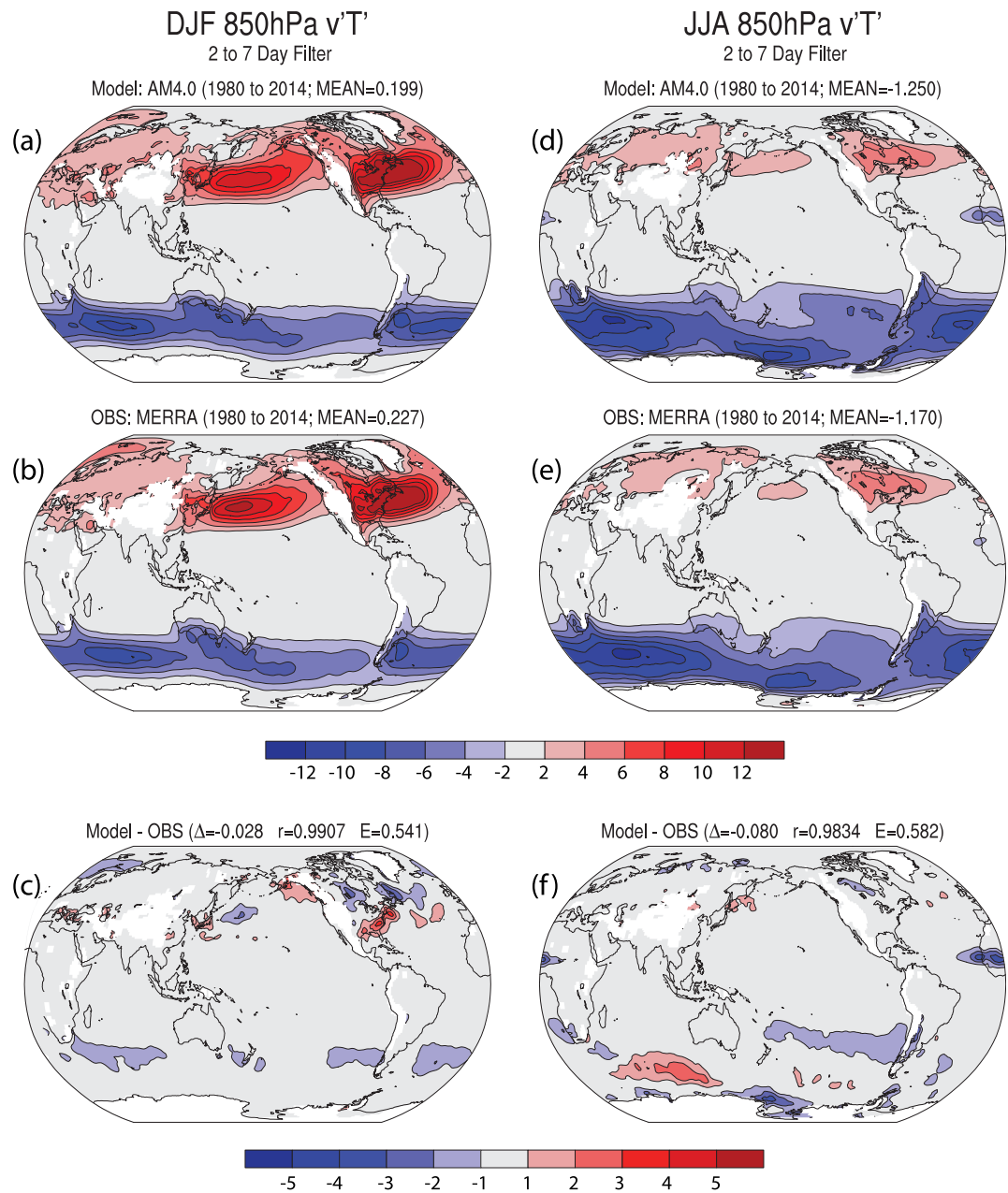
Figure 23. As in Figure 22 except for 850 hPa 2–7 day filtered eddy heat flux v'T' (unit: m s^{-1} K).

AM4.0 than in reality. In general, the coupled simulation tends to enhance the eastward propagation of the OLR signal more than that in U200, consistent with Figure 27.

10. Atmospheric Response to Prescribed ENSO SSTs

A key test for an atmospheric component of a climate model is its response to the SST anomalies (SSTAs) associated with the El Niño-Southern Oscillation (ENSO). Figure 29 illustrates the spatial shifts of rainfall associated with ENSO, for the observations and SST-forced AGCM simulations during 1980–2014. The observed tropical Pacific rainfall response to El Niño is an eastward and equatorward shift of deep convection, with enhanced rainfall east of the dateline (especially north of the equator) and reduced rainfall near the Maritime continent. The AM2.1 simulation produces a stronger rainfall response than observed, with more drying near Indonesia and over South America, and a larger boost in rain near the dateline (Figure
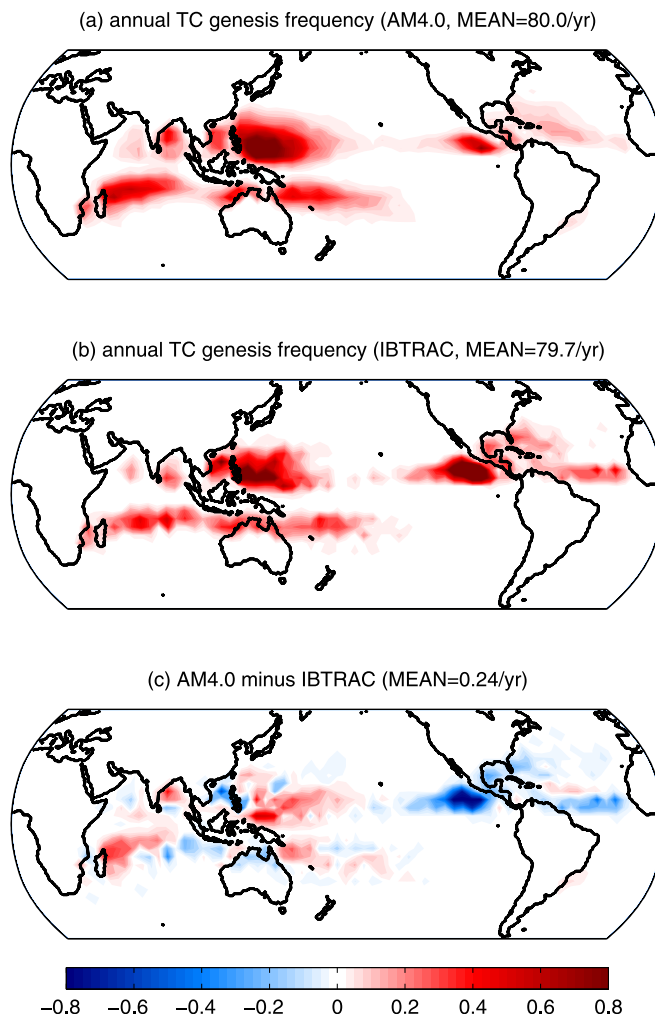
Figure 24. Geographical distribution of the annual tropical cyclone genesis frequency from (a) the average of eight AM4.0 AMIP simulations for the 1980–2014 period and (b) the observations based on the IBTRACS data set. (c) The model biases (AM4.0 minus IBTRACS). Unit: frequency of occurrence per year per 4 × 5 (lat-lon) area. The global mean numbers and their difference are denoted at the top of each figure.

29d). AM2.1's eastward shift of Pacific convection also falls short of the observed shift, giving less of a rainfall increase over the far eastern equatorial Pacific. Compared to AM2.1, AM3 shows improved eastward and equatorward penetration of the Pacific convective zones during El Niño but fails to sufficiently dry the west Pacific region near the Maritime continent and Philippines (Figure 29e). AM4.0, however, shows a much better rainfall response during El Niño—with more realistic eastward and equatorward penetration of Pacific convection, and realistic drying over Indonesia and South America (Figures 29a and 29c). This improved pattern of anomalous atmospheric latent heating would be expected to improve ENSO's remote teleconnections, particularly over Asia, the Pacific, and North America (Delworth et al., 2012; Jia et al., 2015; Krishnamurthy et al., 2015). AM4.0's rainfall response to ENSO does remain slightly too strong; this is likely due to the excessive rainfall in the model's climatological convective regimes (ITCZ and SPCZ), which when shifted then produce excessive local rainfall anomalies.

Figure 30 shows the observed and simulated surface zonal wind stress response to ENSO, which is a critical control on both the amplitude and period of ENSO in coupled models (Capotondi et al., 2006; Kim et al., 2008). Simulating the spatial structure of the zonal wind stress response to ENSO has long been a challenge for atmospheric and especially coupled GCMs, due to biases in the simulated climatological patterns of tropical Pacific convection (Capotondi et al., 2015; Choi et al., 2015). Figure 30d shows that in AM2.1, the equatorial westerly wind stress response to El Niño is too strong and too far west, reflecting a similar
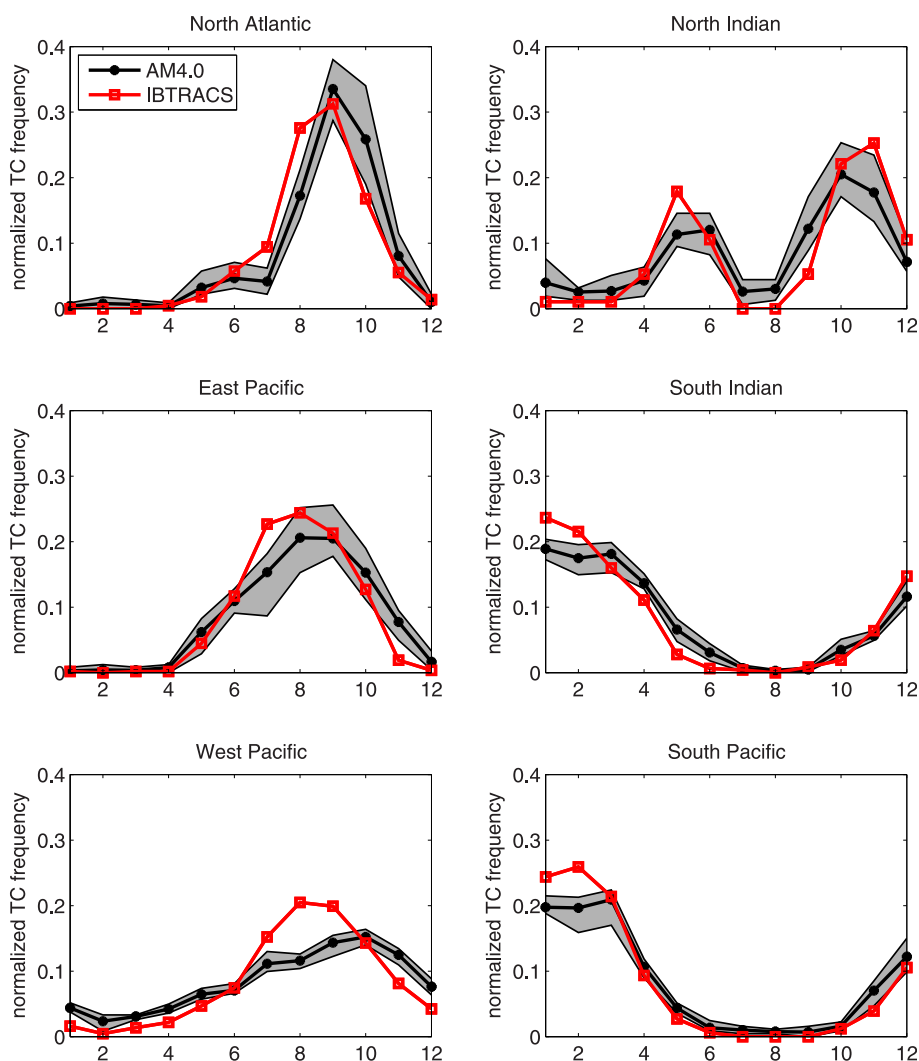
Figure 25. Seasonal cycle of tropical cyclone genesis frequency over different ocean basins from eight AM4.0 AMIP simulations. Black line shows eight-member mean; shading shows the maximum and minimum values of the ensemble. Model and observed total frequencies are normalized to 1 for each ocean basin. The definition of ocean basins follows Zhao et al. (2009).

excessive intensity and westward displacement of its rainfall and tropospheric latent heating anomalies (Figure 29d). AM2.1's off-equatorial easterly anomalies are also excessive poleward of 10°S and 5°SN, giving too much cyclonic curl and poleward Sverdrup discharge within the upper ocean near the equator. Compared to AM2.1, AM3 reduces these equatorial biases in the strength and zonal position of the westerly anomalies; yet its off-equatorial easterlies and cyclonic curl remain too strong (Figure 30e).

AM4.0, however, shows a much improved wind stress response to ENSO—with a realistic zonal and meridional structure for the westerly wind anomalies, and greatly reduced off-equatorial easterly biases (Figures 30a and 30c). In a coupled system, the reduced cyclonic curl and poleward Sverdrup transport associated with AM4.0's wind response could be expected to help lengthen the period of ENSO, by slowing the poleward discharge of equatorial ocean heat content during El Niño. AM4.0 does, however, still show excessive easterly wind anomalies in the eastern equatorial Pacific. In a coupled system, this remaining wind stress bias would likely weaken SSTAs in the cold tongue region during El Niño—both by boosting the evaporative damping of local SSTAs, and by weakening the local SSTA growth due to Bjerknes feedbacks (associated with local anomalous downwelling, and an anomalously depressed equatorial thermocline in the east).
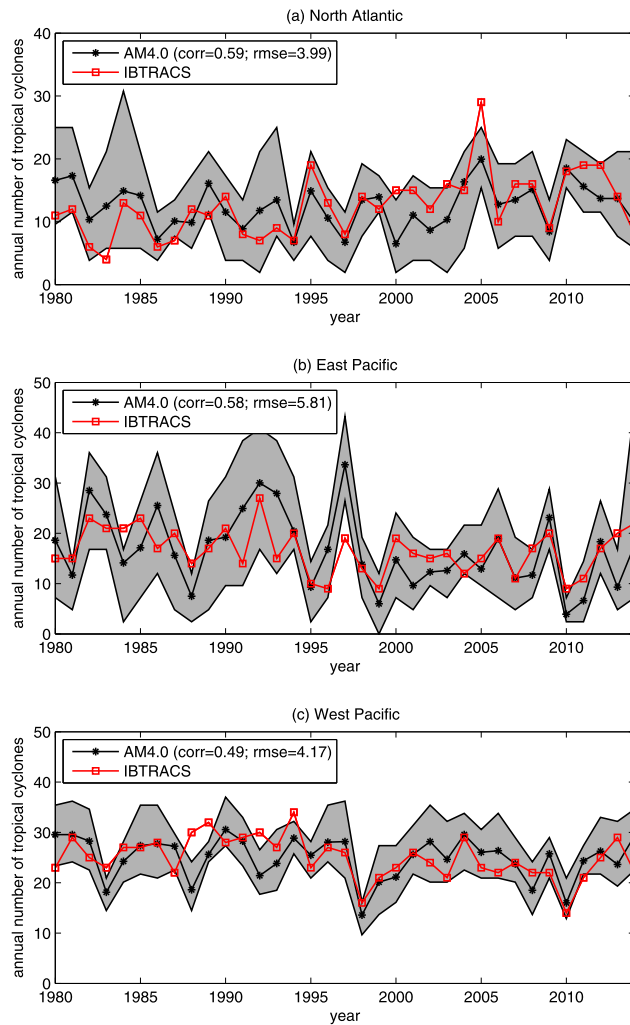
Figure 26. Time series of basin-wide annual tropical cyclone frequency for (a) North Atlantic, (b) East Pacific, and (c) West Pacific. Black: ensemble mean from eight AM4.0 AMIP simulations; shading: minimum and maximum values from the eight members. Red: observational estimates from IBTRACS. The model time series are first normalized to the observational values by multiplying a constant value for each basin before taking the ensemble mean, minimum, and maximum values. Legends show the correlation coefficients and root mean square errors of the normalized ensemble mean time series.

Figure 31 shows the anomalous net surface heat flux response to ENSO. Compared to the observational estimates, AM2.1 (Figure 31d) shows less net heat flux damping of SSTAs in the central/western equatorial Pacific and around the Maritime continent. This stems from two key biases in the AM2.1 response to El Niño: (1) AM2.1's westward-displaced equatorial westerly wind anomalies (Figure 30d), which during El Niño overly weaken the trade winds and evaporative cooling west of 170°W, leading to insufficient evaporative damping of its warm SSTAs in that region and (2) AM2.1's failure to sufficiently reduce its high cloudiness (and associated cloud shading) near the Maritime continent during El Niño. East of 170°W, AM2.1's anomalous easterly wind bias actually leads to an excessive evaporative response and SSTA damping; but this is trumped by too little shortwave damping of SSTAs in the eastern equatorial Pacific, due to an insufficient eastward and equatorward shift of deep convective cloud during El Niño. Compared to AM2.1, AM3's biases in the surface heat flux response to ENSO (Figure 31e) are qualitatively similar but with somewhat larger magnitude.

Compared to AM2.1 and AM3, AM4.0 shows a significantly improved net surface heat flux response to ENSO (Figures 31a and 31c). This is especially the case in the equatorial Pacific west of 170°W and near the Maritime continent, where AM4.0's patterns of anomalous winds and evaporation have greatly improved.
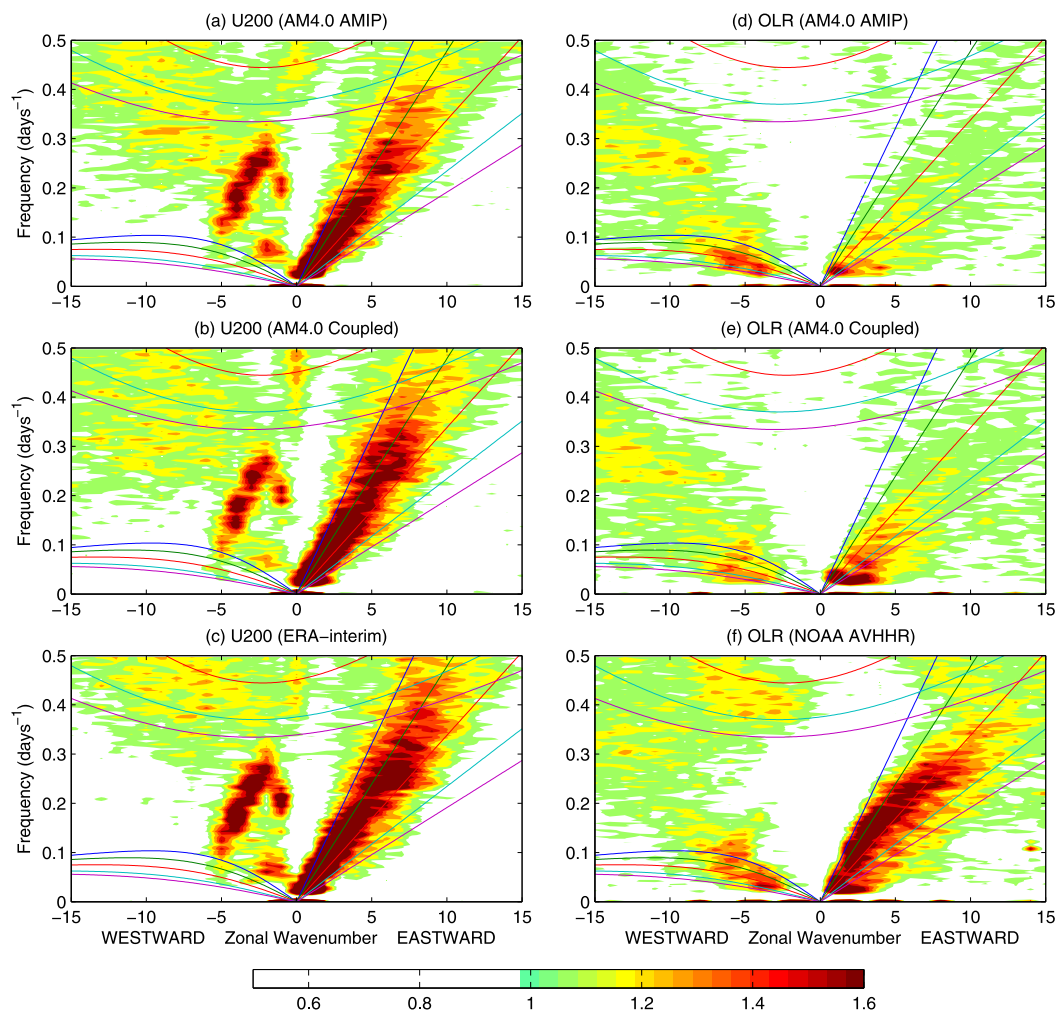
Figure 27. Normalized tropical (15°S–15°N) symmetric upper level (200 hPa) zonal wind (U200) wave number-frequency power spectrum from (a) AM4.0 AMIP simulation, (b) AM4.0 coupled simulation, and (c) observational estimates based on ERA-Interim. Colored shading shows power associated with MJO, Kelvin, and other tropical convective waves that are significantly above an approximately red-noise background power spectra. The colored lines represent various equatorial wave dispersion curves labeled for five different equivalent depths (8, 12, 25, 50, and 90 m). (d–f) As Figures 27a–27c except for OLR. The OLR observation is based on NOAA AVHRR (see Table 1 for details).

(AM4.0's shortwave response pattern is also improved, though not as dramatically as for the evaporative cooling.) If AM4.0's stronger and more realistic surface heat flux response pattern were retained in coupled mode, it would likely help to weaken ENSO. And given the overly strong ENSO in both the AM2.1-based CM2.1 and ESM2M coupled models (Dunne et al., 2012; Wittenberg, 2009; Wittenberg et al., 2006) and the AM3-based CM3 coupled model (Chen et al., 2017), this would be a welcome change in CM4.

To the extent that these improved patterns in AM4.0 are retained upon coupling, they may improve the amplitude, period, dynamics, and remote impacts of the simulated ENSO in CM4. However, as a note of caution, our early experience to date with ENSO simulations in coupled models using AM4.0 is that a range of different ENSO behavior can be generated, depending on the oceanic resolution and physics.

11. Stratospheric Variability

The wind and temperature anomalies associated with the stratospheric sudden warmings (SSWs) can propagate downward into the troposphere and significantly alter the weather at the surface (e.g., Baldwin & Dunkerton, 2001; Scaife et al., 2005). Since several studies have reported a general lack of stratospheric
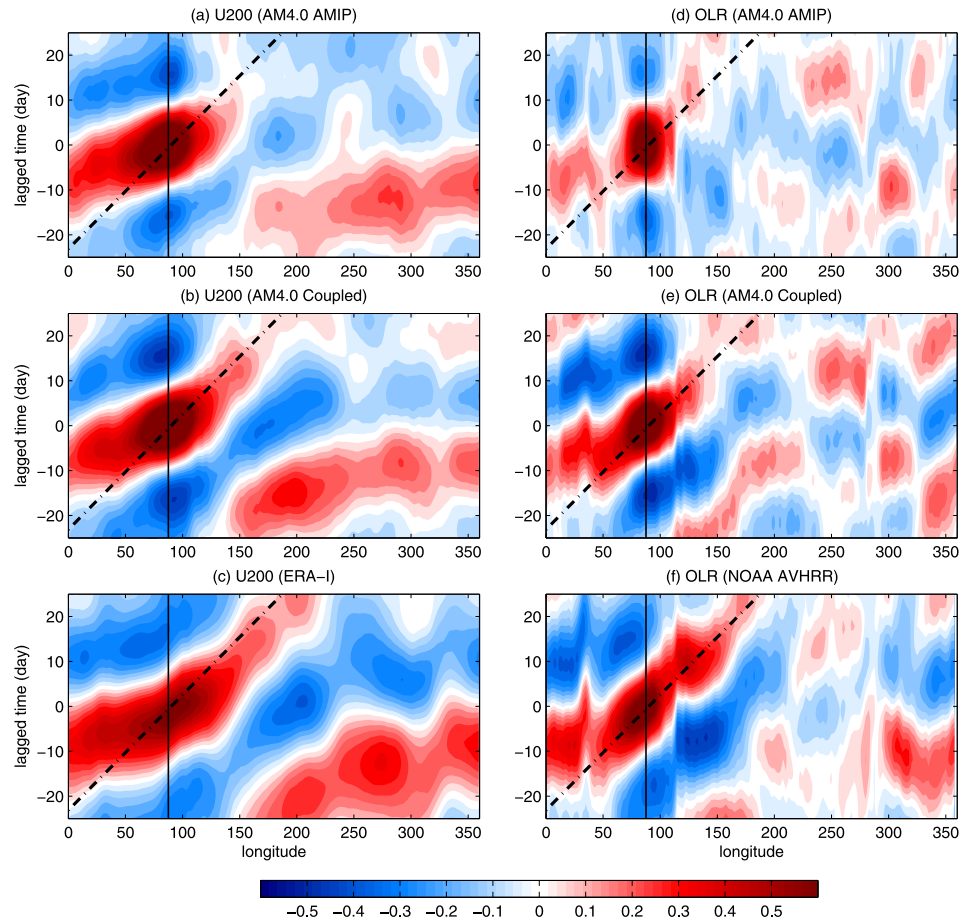
Figure 28. (a–c) Lag correlation between central Indian ocean (10°S–5°N, 75°E–100°E) time series of 200 hPa zonal wind (U200) and the associated near-equatorial (5°S–5°N) U200 at all longitudes. Lag correlations are computed using 20–100 day band-pass filtered data for winter season. (a) AM4.0 AMIP simulation, (b) AM4.0 coupled simulation, and (c) observational estimates based on ERA-Interim (1980–2014). Dashed lines show 5 m s^{-1} eastward propagation phase speed. Vertical lines show 87.5°E longitude. (d–f) As in Figures 28a–28c except for OLR.

variability in the low-top models such as AM4.0 (Charlton-Perez et al., 2013; Hardiman et al., 2012; Shaw & Perlwitz, 2010), it is of interest to see the extent to which this choice has compromised model performance in this regard. These results also serve as a benchmark for versions of AM4 with higher tops and finer stratospheric resolution. Here, we evaluate the key characteristics of SSWs in the AM4.0 long (1870–2014) AMIP simulation, and compare them to ERA40 reanalysis. Despite the low model top, we find that AM4.0 simulates fairly realistic SSW frequencies, but biased toward occurring later in the winter season, and with relatively weak tropospheric signatures following SSW events.

Figure 32 shows the frequency of SSWs per year as well as its distribution by month. The SSW events are diagnosed as the reversal of stratospheric zonal wind at 10 hPa following Charlton and Polvani (2007). The annual mean SSW frequency of AM4.0 is statistically indistinguishable from that of ERA40 at the 95% confidence level, though the distribution of SSWs in AM4.0 is skewed toward late winter compared to observations. We find that the annual mean SSW frequency is sensitive to the parameterization for the convective gravity wave drag, with more SSWs when the drag is increased. (The strength of polar vortex jet is weakened with increased convective gravity wave drag. Convective gravity wave drag is tuned to provide a better agreement of the polar jet with the ERA-Interim data compared to AM2.1 and AM3, see supporting information Figure S15.) However, the monthly distribution is relatively insensitive to the parameterized drag.

The amplitude of SSW is measured by the polar cap temperature anomalies at 10 hPa during SSW and the change in zonal wind at 60°N 10 hPa following the SSW onset (Charlton & Polvani, 2007). On average,
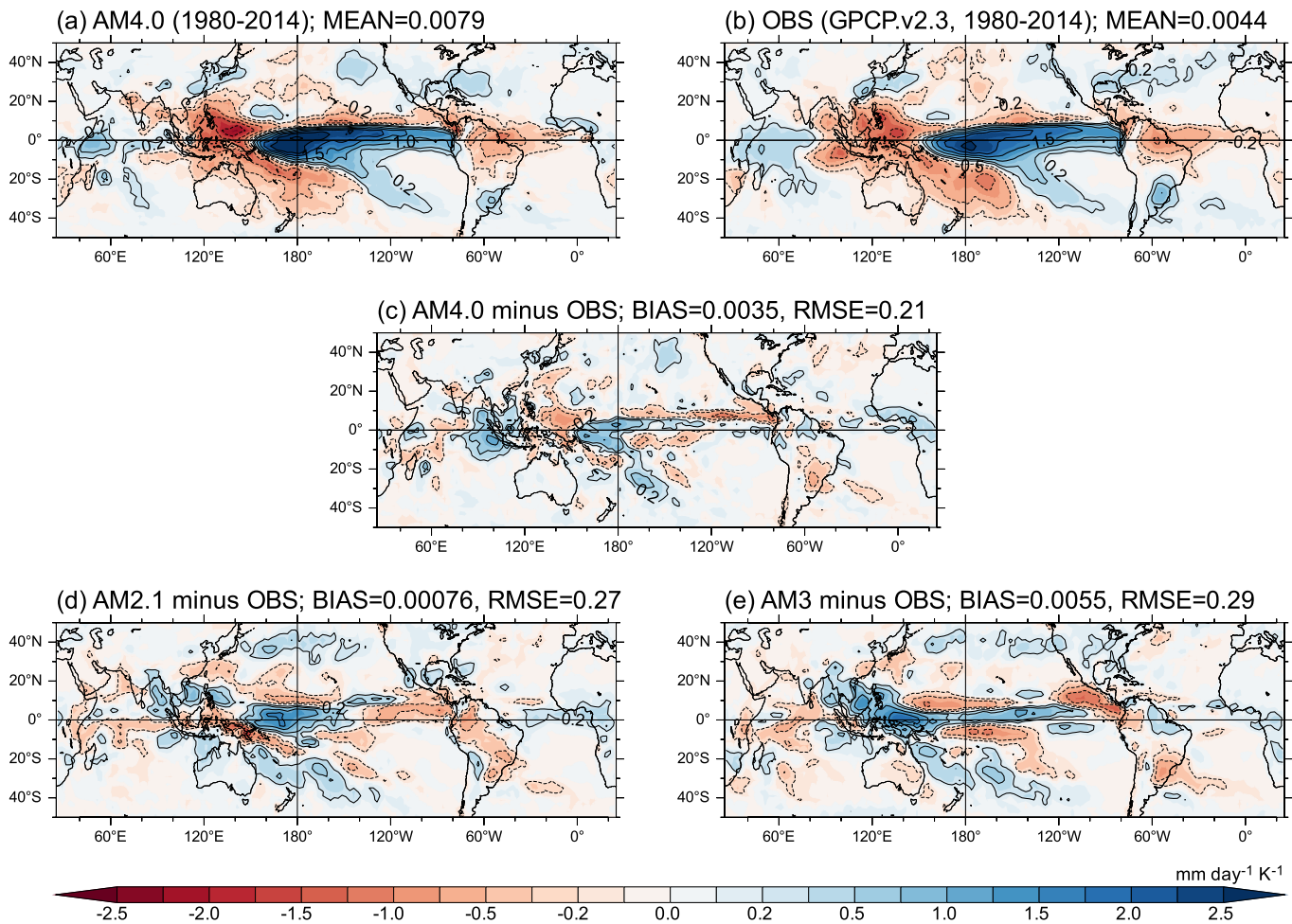
Figure 29. (a) Regression of tropical monthly rainfall anomalies onto monthly SST anomalies (SSTA) averaged over the Niño-3 region (150°W–90°W, 5°S–5°N) (units: mm d^{-1} of local rainfall anomaly per K of Niño-3 SSTA) from (a) the AM4.0 AMIP simulation and (b) the observations based on GPCP-v2.3 and the NOAA Extended Reconstructed Sea Surface Temperature analysis (ERSST-v4, Huang et al., 2014). (c) AM4.0 model biases (AM4.0 minus OBS). (d) As in Figure 29c but for AM2.1. (e) As in Figure 29c but for AM3. All data sets are first averaged onto a uniform 2.5° longitude by 2° latitude grid comparable to that of AM2.1, before the regression is computed. Titles of Figures 29a and 29b show spatial-mean regression values over the region shown. Titles of Figures 29c–29e show spatial-mean regression biases and RMS errors over the region shown.

AM4.0 simulates 7.2 K warming at the polar cap and 27.7 m s^{-1} decrease in zonal wind, which agree well with the benchmark values suggested by Charlton and Polvani (2007). We also compare the simulated influence of SSW at the surface with reanalysis following Hardiman et al. (2012). Figure 33 shows the mean sea level pressure (SLP) anomalies in the month following a SSW from AM4.0 and ERA40. Both model and reanalysis show an anomalously negative Northern Annular Mode (NAM)-like pattern following SSW, though the anomalies are weaker in the model by more than a factor of two, a bias that we estimate to be statistically significant at the 95% level. While this bias could be directly related to the inability of a model with low stratospheric resolution to transmit this signal, it could also be related to the fact that most of the model SSWs occur late in the season, providing little time to set up a tropospheric response before the seasonal wind conditions favorable for coupling disappear.

A related and important consequence of stratospheric-tropospheric coupling is the response of the SH summertime near-surface winds to the development of the Antarctic ozone hole. Figure 34 shows the evolution in time of the latitude of the maximum of the westerlies in the SH summer at 850 hPa, comparing the ERA-Interim data with AM4.0 simulations. The spread across eight AMIP simulations is shown along with the ensemble mean. In addition to showing that AM4.0 has very modest bias in the mean, one can compare the trends toward a more poleward position in the westerlies over this period. This trend has generally
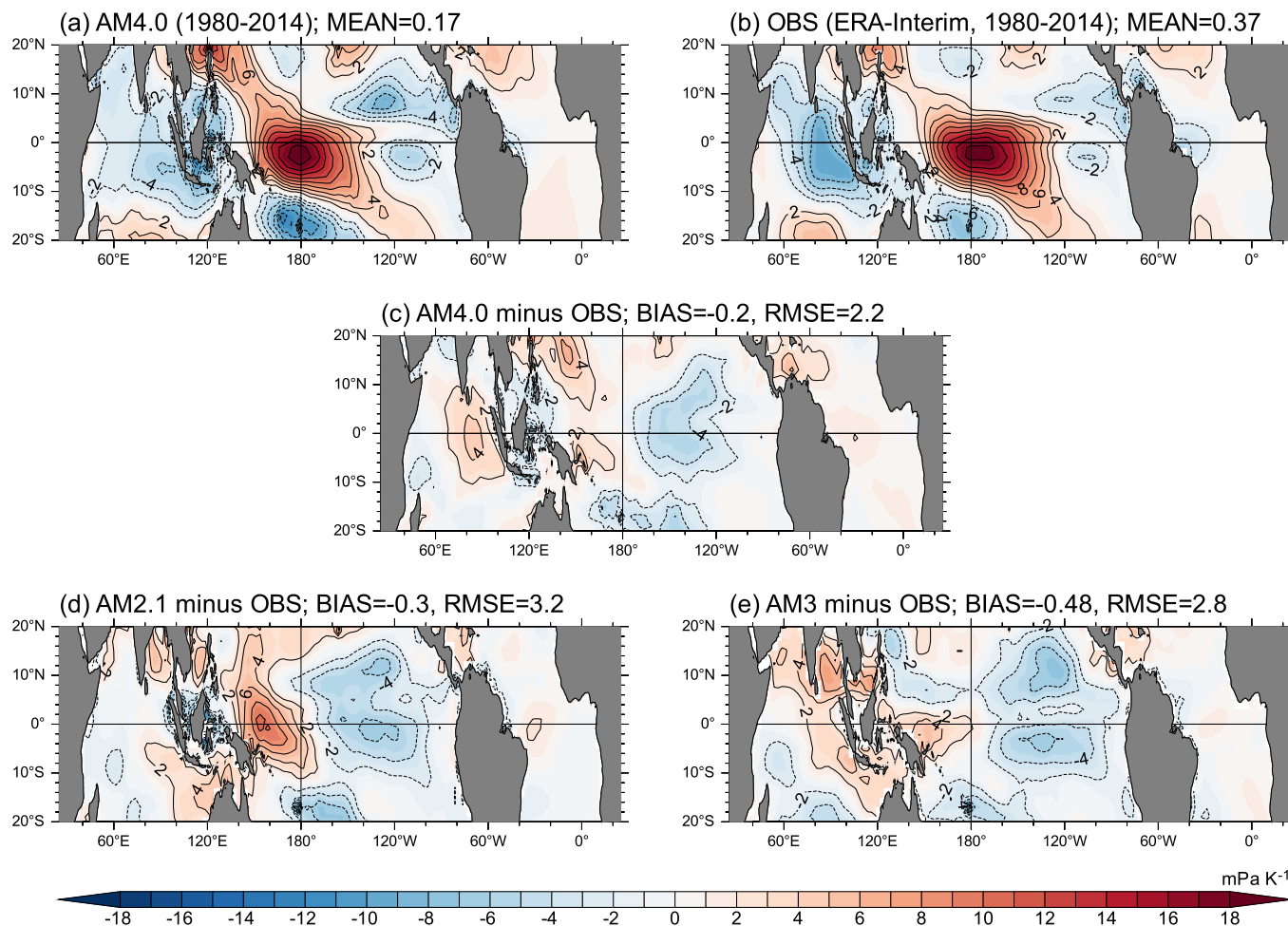
Figure 30. As in Figure 29, but for zonal wind stress anomalies (mPa) regressed onto Niño-3 SSTAs (K). Observed wind stress anomalies in (b) are from the ERA-Interim reanalysis. Positive values correspond to westerly surface winds and westerly stress on the ocean.

been attributed to the stratospheric ozone hole, since Thompson and Solomon (2002), but recent work, such as Garfinkel et al. (2015) and Seviour et al. (2017) have emphasized the difficulty in quantitative attribution due to large internal variability. While the modeled trend appears by eye to be on the weak side compared to the reanalysis, computing trends over the 1980–2000 period during which ozone depletion is the strongest, the reanalysis trend and the trend averaged over an eight-member AMIP ensemble are nearly identical (model ensemble mean is $-0.063°$/yr while ERA-Interim is $-0.078°$/yr). The reanalysis trend lies within the envelope of the eight model realizations, with four out of the eight model realizations having larger poleward trends than the reanalysis (trends computed with the nonparametric Theil-Sen algorithm; Theil, 1950a, 1950b, 1950c).

AM4.0 has no QBO (Quasi-Biennial Oscillation) in the tropical stratosphere. A QBO-like oscillation does appear when the vertical resolution of the model is increased in the lower stratosphere.

12. Aerosols

Given the important interactions between aerosols and the Earth climate system, one of our goals of the model development is to improve the model's simulations of global distribution of aerosols and their radiative effects. For this evaluation, AM4.0 simulated climatological aerosol optical depths (AOD) are compared with their direct measurement by the AERONET sunphotometer network (Holben et al., 1998) in Figure 35. Here we used the quality assured and cloud screened level 2 version 2 AOD data (Smirnov et al., 2000). For
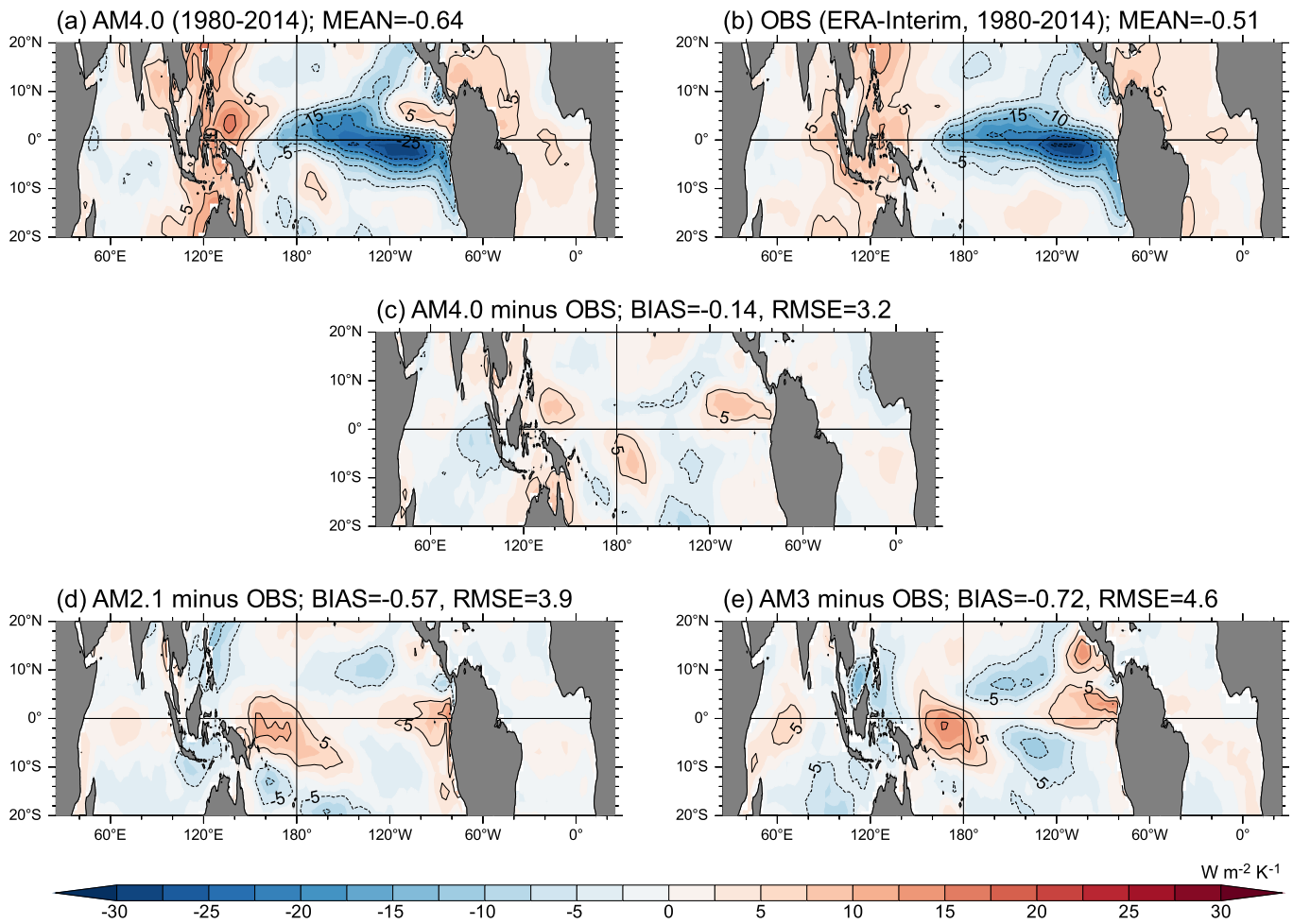
Figure 31. As in Figure 30, but for net surface heat flux anomalies (W m^{-2}) regressed onto Niño-3 SSTAs (K). Positive values indicate a heating of the ocean.

comparison we also show the results from AM3 and AM2.1. AM4.0 significantly improves the simulated AOD over the Maritime continent, Gulf of Mexico, Caribbean Sea, and the west coast of Europe with the overall correlation coefficient between model and observations improved from 0.72/0.82 in AM2.1/AM3 to 0.93 in AM4.0. Table 2 provides the AOD averaged over the entire globe, the ocean and land surface, and the two hemisphere from the models and two satellite measurement.

Since satellite measurement provides a better global and seasonal coverage of the AOD Figure 36 compares by regions the monthly mean AOD simulated by AM3 and AM4.0 with satellite measurement by MODIS (Levy et al., 2007) and MISR (Kahn et al., 2009) instruments. Each plotted value is the surface area weighted average of the simulated and observed value within the region. For all regions, AM4.0 outperforms AM3 in simulating the seasonal variation, as indicated by higher correlation coefficients with MODIS and MISR data. In terms of amplitude, AM4.0 results show less overshooting and undershooting during summer and winter months, respectively. This is particularly pronounced in industrialized regions. The spring maximum over East Asia and the North Pacific is much better captured with AM4.0.

Figure 37 shows that AM4.0 captures the observed seasonality of sulfate concentrations at four polar sites much better than AM3. AM4.0's simulation of black carbon in the wintertime Arctic (not shown) is also much improved as well, consistent with the study of Shen et al. (2017) using a version of AM3 that includes all of the aerosol model modifications in AM4.0. The model development that results in these improvements in aerosol simulation is described in Part 2.
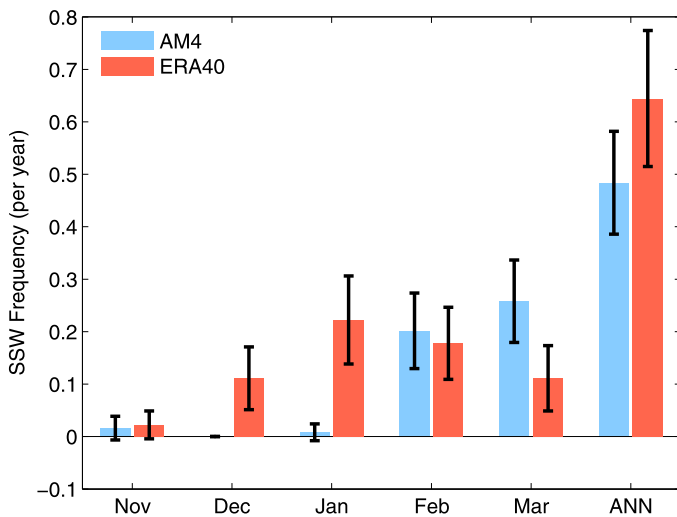
Figure 32. Monthly and annual (ANN) stratospheric sudden warming (SSW) frequency for 1870–2014 from AM4.0 and 1957–2002 from ERA40. SSW is defined as in Charlton and Polvani (2007). Error bars indicate the 95% confidence interval (the statistical test of the SSW frequency is calculated as in Charlton et al. (2007)).

13. Sensitivities to Greenhouse Gases, Aerosols, and SST Perturbations

Given fixed preindustrial (PI) and present-day (PD) forcing agents we can utilize a set of atmospheric only experiments to derive important model properties which measure a model's effective radiative forcing and/or feedbacks (e.g., Andrews, 2014; Pincus et al., 2016; Sherwood et al., 2015). These properties are important to the model's simulation of historical temperature trend as well as future projections of various climate change scenarios. For this purpose, we have conducted a set of idealized experiments to assess AM4.0 sensitivity to perturbations of atmospheric greenhouse gases (GHGs, including ozone) and aerosol emissions. The control experiment is forced by monthly varying climatological SSTs and sea-ice concentration averaged for the 1981–2014 period using the data recommended for the CMIP6 AMIP simulations. The GHG concentrations and aerosol emissions are fixed at 2010 conditions while the volcanic aerosols are prescribed at the average value of the 2000–2014 period. The CO_2 concentration for vegetation photosynthesis is also specified at the 2010 value. This specification of GHGs and aerosol emissions corresponds to that used in the CMIP6 PD control simulations. Below we refer to this experiment as 2010RAD.

The first perturbation experiment is based on 2010RAD except using 1850 GHGs, aerosol emissions, CO_2 concentration for photosynthesis, and the volcanic aerosols averaged over the 1850–2014 period (1850RAD below). This specification of GHGs and aerosols corresponds to that used in the CMIP6 PI control simulations. The change in global TOA net radiative flux between 2010RAD and 1850RAD provides an estimate of the total TOA radiative flux perturbation (RFP) or effective radiative forcing, due to changes in GHGs, aerosols and volcanoes from PI to 2010 conditions. The second perturbation is based on 1850RAD except that only the aerosol emissions are replaced by 2010 conditions (called 2010AER). The difference in TOA flux between 2010AER and 1850RAD provides an estimate of total (nonvolcanic) aerosol RFP. The third perturbation is also based on 1850RAD except that only the GHGs are replaced by 2010 conditions (called 2010GHG). The difference in TOA flux between 2010GHG and 1850RAD provides an estimate of the GHG RFP. Finally, the difference between the total RFP and the sum of aerosol and GHG
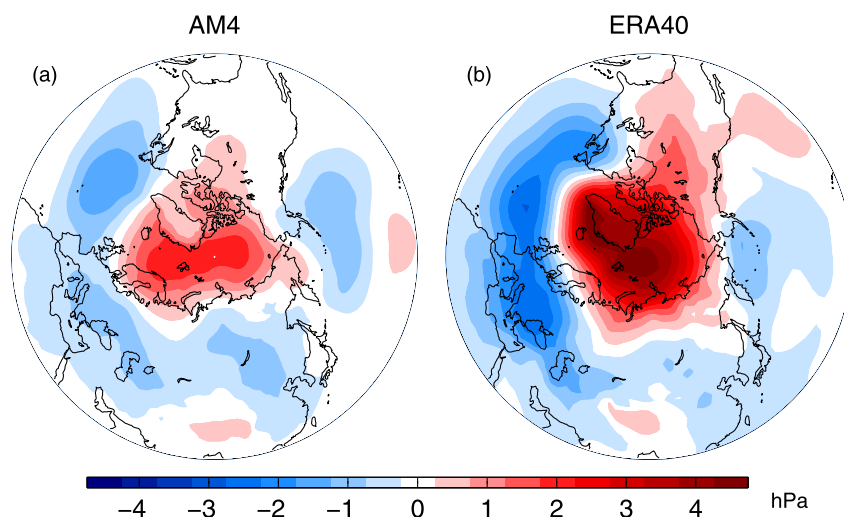

Figure 33. The mean sea level pressure anomalies averaged over the month following an SSW in (a) AM4.0 and (b) ERA40.
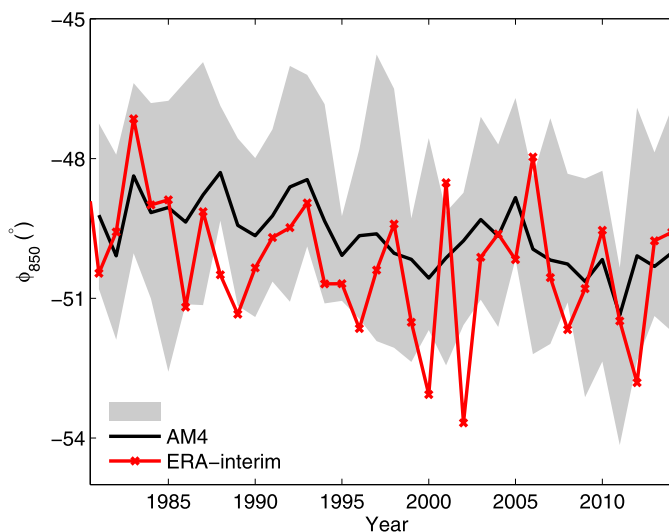
Figure 34. Time series of AM4.0 simulated Southern Hemisphere jet location, defined as the latitude of the maximum westerlies at 850 hPa in austral summer DJF. The black lines are eight-member ensemble mean and shading shows the spread based on the minimum and maximum values from the eight members. Red lines show the observational estimates based on ERA-Interim.

RFPs is derived as a residue, roughly accounting for the effects of model difference in volcanoes and total solar irradiance.

To assess AM4.0 feedbacks to changes in global mean SST, we also conducted an idealized global warming experiment based on 2010RAD except with SSTs uniformly increased by 2K (P2K below) following Cess et al. (1990). The change in TOA net flux measures the total feedback to warming; a Cess climate sensitivity can be derived by dividing the global mean surface temperature increase by this total TOA change (Cess et al., 1990). Here we define Cess sensitivity from these simulations using the 2K ocean warming in the numerator rather than the global mean surface air temperature, to be consistent with the published AM2 and AM3 results (Zhao, 2014; Zhao et al., 2016). Using the global mean in AM4.0 results in a sensitivity ∼11% higher.

The control and each perturbation experiment are integrated for 31 years with the output from the last 30 years used to compute the model statistics. The global mean surface air temperature and TOA radiative fluxes from the 2010RAD and each of the perturbation experiments are listed in Table 3. Figure 38 shows that the total RFP from 1850RAD to 2010RAD in AM4.0 is estimated as 2.66 W m^{-2} with –0.74 W m^{-2} aerosol RFP and 3.24 W m^{-2} GHG RFP. The residue of 0.17 W m^{-2} results primarily from the weaker than historically averaged volcanoes specified for the 2010 condition.

A point of confusion when comparing these numbers to those quoted for AM2.1 and AM3 in the literature is that the AM2.1 and AM3 values were derived previously based on a different set of experiments where the present-day conditions were specified at 1990 conditions (using CMIP3 specifications for AM2.1 and CMIP5 for AM3; but the differences in specification of GHGs and aerosol emissions for the same year are not key to these differences) and with volcanoes turned off in both the control and the perturbation experiments. To provide a clean comparison, we conducted an additional set of AM4.0 simulations using the 1990 GHGs and aerosol emissions for the "present-day" control simulation (1990RAD). In addition, both the 1990 control and its corresponding perturbation runs used the same volcanoes. The RFPs derived from this 1990RAD-based set of experiments are also shown in Figure 38 along with the results from AM2.1 and AM3. Table 4 provides the numbers shown in Figure 38.

With the 1990 conditions, AM4.0 produces a total RFP of roughly 1.64 W m^{-2}, which is smaller than AM2.1 (1.98 W m^{-2}) but much larger than AM3 (0.94 W m^{-2}). Nearly all of the increase in total RFP from AM3 to AM4.0 is due to a reduction in the magnitude of aerosol RFP (from –1.69 to –0.96 W m^{-2}). Both AM4.0 and AM3 produce a significant (∼0.5 W m^{-2}) increase of GHG RFP compared to AM2.1. The increase of GHG RFP from AM2.1 to AM4.0 comes from two sources. One is the change in the radiative transfer code described in
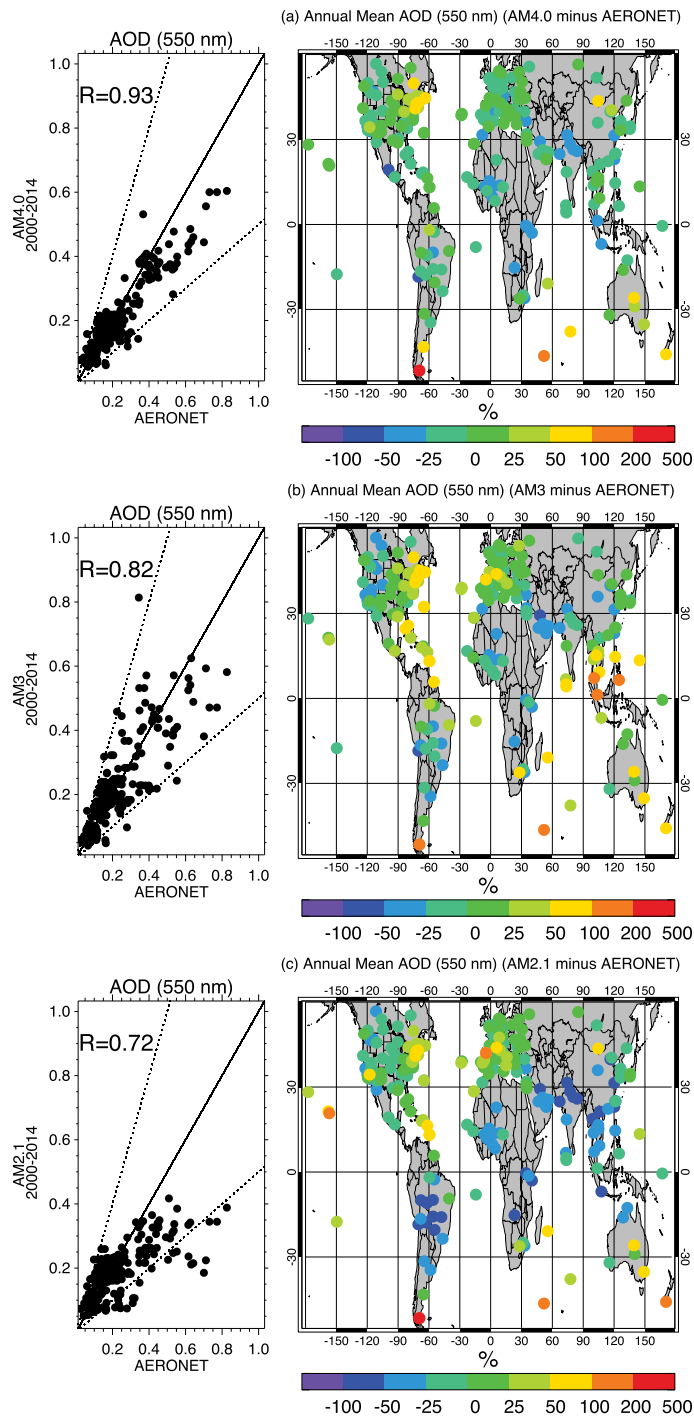
Figure 35. Comparison of model simulated climatological (2000–2014) aerosol optical depths (550 nm) with AERONET for (a) AM4.0, (b) AM3, and (c) AM2.1 AMIP simulation. Dashed lines in left panels denote slopes of 0.5 and 2. Color in right figures shows the percentage difference between model and AERONET (i.e., 100× [model-AERONET]/AERONET).

Part 2, and the other is due to changes in ozone specification. In contrast, the increase of GHG RFP from AM2.1 to AM3 is primarily due to the interactive ozone in AM3. To summarize, when evaluated with 1990 conditions, AM4.0/CMIP6 produces a larger magnitude in both GHG and aerosol RFPs than AM2.1/CMIP3 with a modest (−0.3 W m⁻²) decrease of total RFP. In contrast, compared to AM3/CMIP5, AM4.0/CMIP6 produces a substantially smaller aerosol RFP with a similar GHG RFP so that the total RFP increases by

Table 2

Annual Mean (2000–2014) Aerosol Optical Depth at 550 nm Simulated From AM4.0, AM3, and AM2.1, and Retrieved From MODIS and MISR Satellite Measurements

Model	Global	Ocean	Land	NH	SH
AM4.0	0.148	0.133	0.183	0.186	0.113
AM3	0.161	0.156	0.182	0.201	0.122
AM2.1	0.137	0.134	0.149	0.170	0.104
MODIS	0.179	0.163	0.222	0.219	0.135
MISR	0.167	0.158	0.190	0.193	0.140

Note. The numbers are averaged respectively over the entire globe, the global ocean and land surface, the Northern Hemisphere (NH), and the Southern Hemisphere (SH).

$0.7 \, \mathrm{W \, m^{-2}}$. The aerosol RFP in AM4 and AM3 contains both the direct effect and the indirect effects due to aerosol cloud interactions while AM2.1 contains only the aerosol direct effect. The reduction of aerosol RFP from AM3 to AM4 is mostly due to the reduction of aerosol indirect effect while the smaller aerosol RFP in AM2.1 is mostly due to its lack of aerosol indirect effects.

The smaller aerosol RFP in AM4.0 compared to AM3 is an important feature of this model, and requires closer examination. More information can be inferred from the time evolution of aerosol RFP obtained by
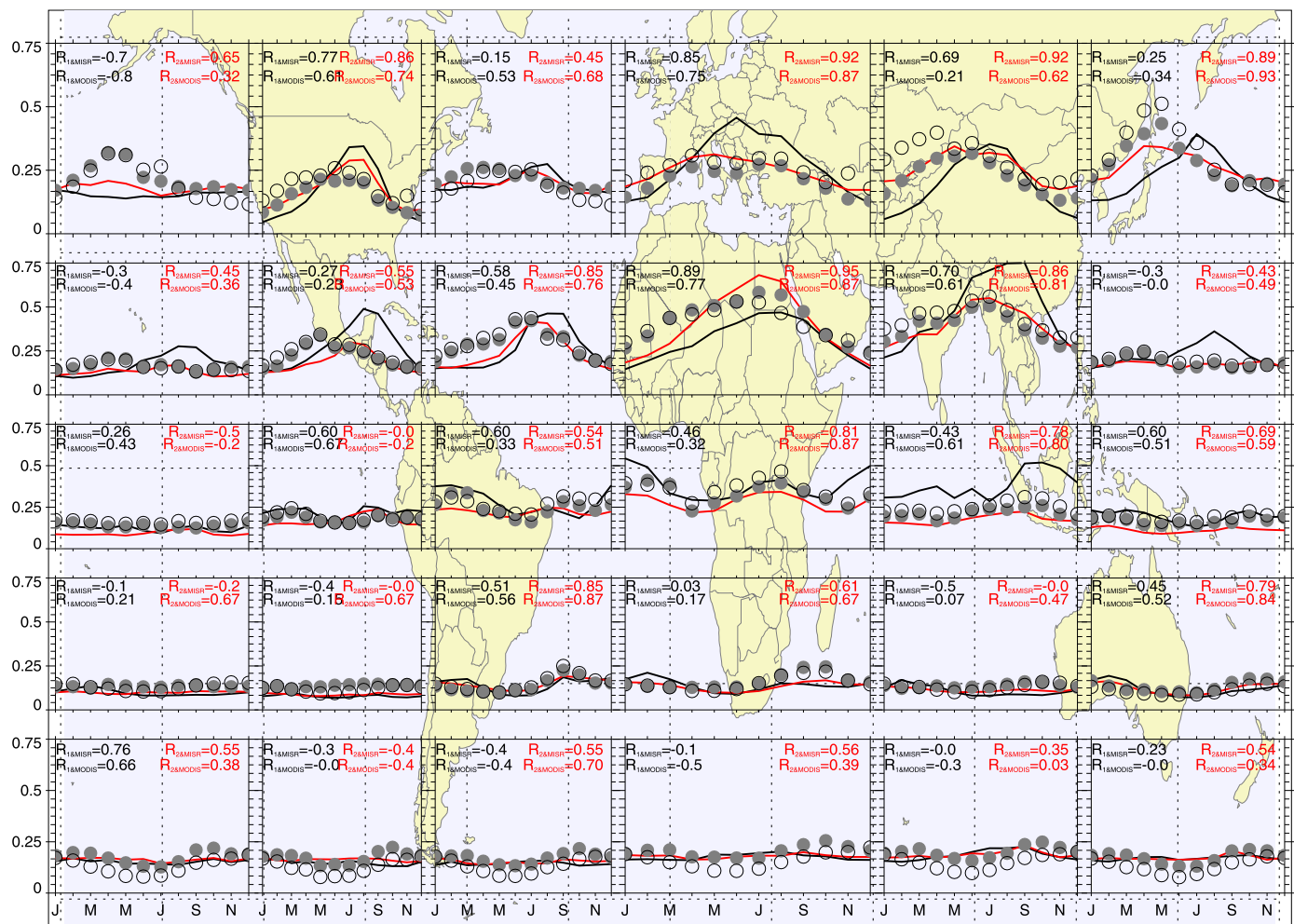

Figure 36. Monthly climatology (2000–2004) of aerosol optical depth (unitless) simulated by AM3 (black lines) and AM4.0 (red lines) and measured by MODIS (open circles) and MISR (filled circles) satellite instruments over 30 regions. The numbers in each box show the correlation coefficients compared to MODIS and MISR (black: AM3, red: AM4.0).
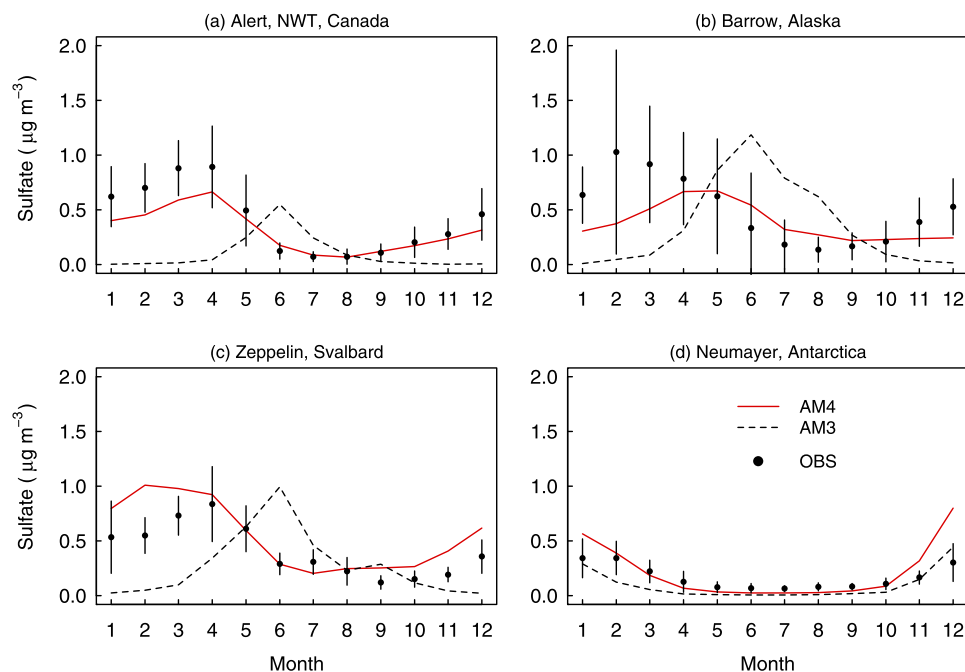
Figure 37. Average monthly sulfate concentrations (unit: $\mu g\,m^{-3}$ at standard temperature pressure) in submicron aerosols at long-term observation sites compared to model results. (a) Alert (82.5°N, 62.3°W, 1998–2006) (Siorois & Barrie, 1999), (b) Barrow (71°N, 156.6°W, 1998–2008) (Quinn et al., 2000), (c) Zeppelin (78.9°N, 11.9°E, 474 m above sea level, 1998–2009) (Udisti et al., 2016), (d) Neumayer (70.7°S, 8.3°W, 2004–2007) (Weller et al., 2011). Measurements are indicated by symbols with vertical lines for standard deviations. Model results are averaged for the period of 2000–2009 (AM4, solid red line; AM3 dashed black line).

differencing the TOA fluxes in a pair of long (1870–2014) AMIP simulations, both with observed time-evolving SSTs and sea ice using CMIP6 specifications for the boundary conditions. In the control simulation preindustrial forcing agents are fixed through time, while the perturbed simulation includes time-evolving aerosol and aerosol-precursor emissions with all other forcings held at preindustrial values. The AM3 simulations here use CMIP5 emissions while AM4.0 uses CMIP6. The result is shown for both AM3 and AM4.0 in Figure 39 and indicates a difference in the relationship between the two models prior to and after 1980.

Table 3

Global Averaged 2 m Surface Air Temperature (TAS, Unit: K) and TOA Radiative Fluxes (Unit: W m^{-2}) From the AM4.0 Present-Day Climatological Simulation (2010RAD) and Each of the Perturbation Experiments (1850RAD, 2010AER, 2010GHG, and P2K)

Exp	TAS	OLR	SWABS	NETRAD	LW CRE	SW CRE	Total CRE
2010RAD	287.16	238.54	240.23	1.69	23.68	−48.54	−24.86
1850RAD	287.06	241.16	240.18	−0.98	24.53	−48.81	−24.28
2010AER	287.01	241.09	239.36	−1.72	24.48	−48.99	−24.51
2010GHG	287.19	238.55	240.81	2.26	23.74	−48.31	−24.57
P2K	289.43	242.91	241.07	−1.85	23.89	−48.37	−24.48
Total RFP 2010RAD-1850RAD	0.12	−2.61	0.05	2.66	−0.85	0.28	−0.58
Aerosol RFP 2010AER-1850RAD	−0.05	−0.07	−0.82	−0.74	−0.05	−0.18	−0.23
GHG RFP 2010GHG-1850RAD	0.13	−2.61	0.63	3.24	−0.79	0.50	−0.29
Residue (Total-Aero-GHG) RFP	0.04	0.07	0.24	0.16	−0.01	−0.04	−0.06
Cess feedback P2K-2010RAD	2.25	4.37	0.84	−3.53	0.21	0.17	0.39

Note. The total RFP and RFPs due to aerosols, GHGs, and other changes can be derived from the differences between the corresponding experiments. The differences between P2K and 2010RAD are also listed as Cess feedback to show the TOA radiative feedback to idealized 2K uniform SST warming. See text for a description of the experiments. TAS, surface air temperature; OLR, TOA outgoing LW radiation; SWABS, TOA net downward SW radiation; NETRAD, TOA net radiation (positive: downward). LW, SW, and total CREs are respectively for the LW, SW, and total cloud radiative effects.
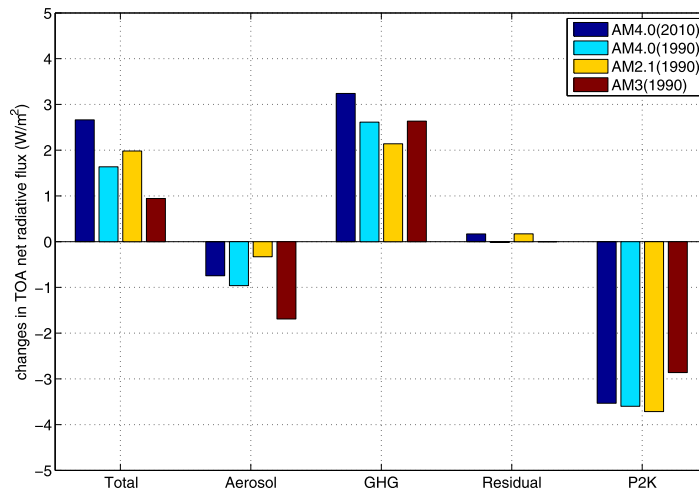
Figure 38. Changes in global TOA net radiative flux in W m^{-2} between various perturbation experiments described in the text. AM4.0 (2010) denotes the radiative flux perturbations (RFPs) derived using 2010 as the present-day (PD) condition for GHGs and aerosols. AM4.0 (1990) denotes the RFPs derived using 1990 as the PD condition for GHGs and aerosols. All simulations use prescribed climatological SSTs and sea-ice concentration averaged for the 1981–2014 period from CMIP6. Along the abscissa, "Total," "Aerosol," and "GHG" show respectively the total, aerosol, and GHG RFPs. "Residual" shows the difference between total and the sum of aerosol and GHG RFPs. P2K denotes the difference in TOA net radiative flux between a PD control climatological simulation and its corresponding P2K (uniform 2K increase of SSTs) simulation. For comparison, the RFP values (derived using 1990 GHGs and aerosols specified from previous CMIPs) from AM2.1 and AM3 are also shown. The climatological SSTs and sea-ice concentration used for AM2.1 and AM3 are also based on previous CMIP specifications and averaged for the 1981–2000 period.

There is a ratio of 0.72 between the global RFPs in the two models that is nearly constant in time before 1980. After this date, the two RFPs diverge further, with the magnitude of the aerosol forcing in AM4.0 decreasing rapidly, unlike the stationary value in AM3. As addressed in Part 2, this appears to be related to the shift in emissions from the NH midlatitudes to the tropics after 1980. Our understanding of this behavior is described in Part 2.

Figure 38 also shows that as global mean SST increases uniformly by 2 K, AM4.0 produces 3.5–3.6 W m^{-2} reduction in TOA net downward radiative flux. This yields a Cess sensitivity of 0.56–0.57 K W^{-1} m^2. This magnitude of net radiative restoring strength and Cess sensitivity values are fairly similar to AM2.1 (0.54 K W^{-1} m^2). In contrast, AM3 produces significantly weaker net radiative restoring and larger Cess sensitivity. Most of the difference is due to different cloud feedbacks, which can be traced to the models' representation of convective clouds (Zhao, 2014; Zhao et al., 2016).

The implications of these values of Cess sensitivity and aerosol RFP for coupled simulations, and the extent to which these values have been tuned in the development process, are discussed in Part 2.

Table 4

A Tabular List of Model Simulated Total Top-Of-Atmosphere Radiative Flux Perturbation (RFP) due to Changes in Radiative Forcing Agents From Preindustrial to Present-Day Conditions and Its Partition Into Individual Components (Aerosols and Greenhouse Gases)

Model	Total RFP (W/m^2)	Aerosol RFP (W/m^2)	GHG RFP (W/m^2)	Residue (W/m^2)	Cess feedback (W/m^2)
AM4.0 (PD = 2010)	2.66	−0.74	3.24	0.16	−3.53
AM4.0 (PD = 1990)	1.64	−0.96	2.61	−0.01	−3.59
AM3 (PD = 1990)	0.94	−1.69	2.63	0.00	−2.86
AM2.1 (PD = 1990)	1.98	−0.33	2.14	0.17	−3.72

Note. (PD = 2010) indicates the RFP values using 2010 as the present-day (PD) conditions while (PD = 1990) using 1990 as the PD conditions. See text for the description of the control and perturbation experiments and the derivation of the RFPs and Cess feedback.
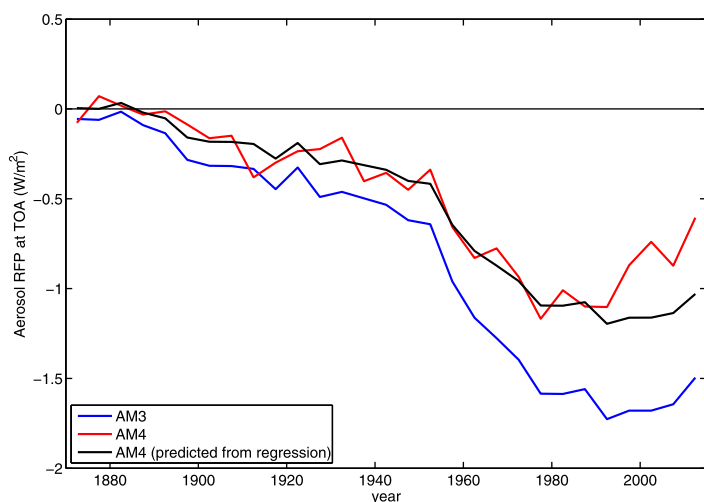
Figure 39. Time series of the aerosol RFP (i.e., net change in radiative flux at TOA) derived from a long (1870–2014) AMIP simulation with a constant radiative forcing fixed at 1850 condition and an additional identical run except with the fixed aerosol emissions replaced by time-varying aerosol emissions. Both long AMIP runs are forced by observed time-varying SSTs and sea-ice concentrations. Time series are computed by averaging over 5 year periods. Red line is from AM4.0 simulations using the CMIP6 specifications of aerosol emissions. Blue line is from AM3 simulations using the CMIP5 specifications of aerosol emission. Black line is a predicted AM4.0 aerosol RFP based on a linear regression of AM4.0 and AM3 aerosol RFPs for the 1870–1980 period. The regression slope is 0.72 with correlation coefficient 0.98.

14. Summary

The development of the AM4.0/LM4.0 model described here was determined by a mix of motivations, ranging from addressing weaknesses in previous models developed at GFDL to specific interests and expertise of members of the model development team. We have described the model's performance in AMIP mode, although in a few instances (such as the MJO, aspects of the tropical precipitation climatology) we have referred to results when AM4.0/LM4.0 was coupled to prototypes of new ocean and sea-ice models.

We have focused here on a few measures of simulation quality. There are many others that we have not discussed, and many that we have not even examined, which we hope will be critically assessed by the community.

We have described the simulation of TOA fluxes from a variety of perspectives because of the central role that it plays in the analysis of climate change. The model in which this simulation was obtained is described in Part 2 and is strongly dependent on the treatment of clouds and convection. This component of the model is certainly not constructed entirely from first principles but is a mix of basic physics and empiricism, as well as experience regarding which parameters influence which aspects of model behavior, experience gained not only in this cycle of model development but in previous cycles with related models. Our confidence in this aspect of the resulting model is based on the low spatial biases throughout the seasonal cycle, but also the simulation of the seasonal cycle and interannual variability of global means of the TOA fluxes (Figures 7 and 8).

The precipitation has important biases, especially regarding the excessive precipitation maximum in the West Pacific (the Philippines hotspot) but biases throughout the seasonal cycle are still comparable to the best AMIP simulations in the CMIP5 archive. AM4.0 significantly improves the precipitation dry bias over the south America Amazonian region as well as the summertime Atlantic ITCZ location and intensity compared to AM2.1 and AM3. Importantly, the response to ENSO SSTs in precipitation, wind stresses, and surface fluxes agree well with observations and recent reanalyses (Figures 29–31). Despite the relatively low resolution, the spatial distribution and frequency of tropical storms is realistic though not quite of the quality we have seen in higher resolution models, especially regarding the simulation of the interannual variability of genesis in the North Atlantic. In addition, the simulated MJO has reasonable amplitude and propagation characteristics, but only when coupled to an ocean model. A clear weakness of the convection scheme is evident in the diurnal cycle over land, which has the common bias of peaking too close to local noon.

The more complete chemistry mechanism in AM3 has been replaced by a light chemistry primarily meant to provide a minimal framework needed to generate sulfate and secondary organic aerosol from emission. The aerosol simulation in the model is substantially improved over that in AM3, particularly for the seasonal variations in Northern high latitudes and in highly polluted midlatitude regions (Figures 36 and 37).

The aerosol RFP is substantially lower in AM4.0 than in AM3, and the Cess sensitivity is lower as well. When the aerosol RFP is examined over the full twentieth century, we see a uniform in time reduction in RFP moving from AM3 to AM4.0 of 28% until roughly 1980, after which AM4 aerosol forcing, using CMIP6 emissions, weakens more rapidly than AM3 with CMIP5 emissions. The important question of the extent of tuning that resulting in these sensitivities will also be described in Part 2.

We have chosen to start with the C96 (roughly 100 km) horizontal resolution model and a vertical grid with 33 levels with a low top (and prescribed ozone), thereby putting aside model development focusing on the stratosphere in this initial stage. We do see a reasonable tropospheric wind response to the stratospheric ozone hole in the Southern Hemisphere summer, and sudden warmings have a realistic frequency, although the tropospheric response following stratospheric sudden warmings is relatively weak. In the near

future, we hope to document a version of AM4.0 with 49 levels, with a higher top and more complete stratospheric and tropospheric chemistry (with predictive ozone).

We also have C192 (roughly 50 km) simulations in hand which produce superior tropical cyclone climatology, and coastal wind and cloud biases are somewhat muted when coupled to an ocean model, as compared to C96. In preliminary results we find, however, that Cess sensitivity, aerosol RFP, precipitation biases, and MJO variability are similar to first order to those of the C96 model. All of these features are strongly affected by the model convection and cloud parameterizations, justifying a strategy of starting the development process at the lower resolution, given that the design of the convection scheme and cloud optimization are a major focus.

The value of this new atmosphere/land model, and these other extensions of the model, will only emerge as it is used in attempts to improve our understanding of climate variability and change and when its simulations are compared to other models. To this end, we provide the AM4.0/LM4.0 code and selected model results for an AMIP simulation at http://data1.gfdl.noaa.gov/nomads/forms/am4.0/.

Appendix A: Boundary Conditions and Forcings

Observed gridded SST and sea-ice concentration boundary conditions for driving the AMIP simulation (1979–2014) are taken from the reconstructions of Taylor et al. (2000). Historical reconstructions of monthly solar irradiances are from Matthes et al. (2017). Global monthly mean concentrations of greenhouse gases (GHGs), including carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O), and ozone depleting substances (ODSs, including CFC-11, CFC-12, CFC-113, and HCFC-22) are from Meinshausen et al. (2017). Annually varying time series of monthly anthropogenic and biomass burning emissions of carbonaceous aerosols and sulfur dioxide (SO_2) precursor to sulfate aerosols, are from the Community Emissions Data System (CEDS; Hoesly et al., 2017) and the data set of van Marle et al. (2017), respectively. (One point is worth clarifying here. These simulations use a version of historical emissions available from CMIP6 as of 3 January 2017. The finalized emissions became available after this date, and we have redone some of the figures that might plausibly be affected by this change, especially the long 1870–2014 AMIP simulation used for the computations of aerosol RFP as a function of time in section 13, and find no changes of physical significance.) Vertical distribution of biomass burning emissions for carbonaceous aerosols is treated similarly to that in AM3 (Donner et al., 2011; Naik et al., 2013).

As in AM3, direct injection of SO_2 from volcanic eruptions and emissions of carbonyl sulfide (COS) are not considered in AM4. Instead, we specify time series of stratospheric aerosol optical properties, which includes not only the volcanic contribution to stratospheric aerosol abundance but also other natural and anthropogenic contributions. The contribution to tropospheric SO_2 from continuously degassing and explosive volcanoes is treated in the same way as in AM3 (Donner et al., 2011).

The fixed land vegetation in these simulations is determined by running the land model, uncoupled from the atmosphere and forced with observed near-surface temperatures, winds and humidities, as well as precipitation (Sheffield et al., 2006). This weather is taken from the year 1981, while the land use is taken from the same year in Hurtt et al. (2011).

References

Adler, R., Huffman, G., Chang, A., Ferraro, R., Xie, P., Janowiak, J., et al. (2003). The version 2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *Journal of Hydrometeorology*, 4, 1147–1167.

Andrews, T. (2014). Using an agcm to diagnose historical effective radiative forcing and mechanisms of recent decadal climate change. *Journal of Climate*, 27, 1193–1209. https://doi.org/10.1175/JCLI-D-13-00336.1

Baldwin, M. P., & Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous weather regimes. *Science*, 294, 581–584. https://doi.org/10.1126/science.1063315

Bennartz, R., & Rausch, J. (2017). Global and regional estimates of warm cloud droplet number concentration based on 13 years of AQUA-MODIS observations. *Atmospheric Chemistry and Physics*, 17, 9815–9836. https://doi.org/10.5194/acp-17-9815-2017

Blackmon, M. L. (1976). A climatological spectral study of the 500 mb geopotential height of the northern hemispher. *Journal of the Atmospheric Sciences*, 33, 1607–1623.

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J.-L., Klein, S. A., et al. (2011). Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, 92, 1023–1043. https://doi.org/10.1175/2011BAMS2856.1

Capotondi, A., Ham, Y.-G., Wittenberg, A. T., & Kug, J.-S. (2015). Climate model biases and El Niño Southern Oscillation (ENSO) simulation. *U.S. CLIVAR Variations*, 13, 21–25.

Capotondi, A., Wittenberg, A., & Masina, S. (2006). Spatial and temporal structure of tropical Pacific interannual variability in 20th century coupled simulations. *Ocean Modelling*, *15*, 274–298. https://doi.org/10.1016/j.ocemod.2006.02.004

Cess, R., Potter, G. L., Blanchet, J. P., Boer, G. J., Del Genio, A. D., DéQué, M., et al. (1990). Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models. *Journal of Geophysical Research*, *95*, 16601–16615.

Charlton, A., & Polvani, L. M. (2007). A new look at stratospheric sudden warmings. Part I: Climatology and modeling benchmarks. *Journal of Climate*, *20*, 449–469.

Charlton, A., Polvani, L. M., Perlwitz, J., Sassi, F., Manzini, E., Shibata, K., et al. (2007). A new look at stratospheric sudden warmings. Part II: Evaluation of numerical model simulations. *Journal of Climate*, *20*, 470–488.

Charlton-Perez, A. J., Baldwin, M. P., Birner, T., Black, R. X., Butler, A. H., Calvo, N., et al. (2013). On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models. *Journal of Geophysical Research: Atmospheres*, *118*, 2494–2505. https://doi.org/10.1002/jgrd.50125

Chen, C., Cane, M. A., Wittenberg, A. T., & Chen, D. (2017). ENSO in the CMIP5 simulations: Life cycles, diversity, and responses to climate change. *Journal of Climate*, *30*, 775–801.

Chen, J.-H., & Lin, S.-J. (2013). Seasonal predictions of tropical cyclones using a 25-km resolution general circulation model. *Journal of Climate*, *26*, 380–398. https://doi.org/10.1175/JCLI-D-12-00061.1

Choi, K.-Y., Vecchi, G. A., & Wittenberg, A. T. (2015). Nonlinear zonal wind response to ENSO in the CMIP5 models: Roles of the zonal and meridional shift of the ITCZ/SPCZ and the simulated climatological precipitation. *Journal of Climate*, *28*, 8556–8573. https://doi.org/10.1175/JCLI-D-15-0211.1

Dai, A. (2006). Precipitation characteristics in 18 coupled climate models. *Journal of Climate*, *9*, 4605–4630.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*, 553–597. https://doi.org/10.1002/qj.828

Delworth, T., Rosati, A., Anderson, W., Adcroft, A. J., Balaji, V., Benson, R., et al. (2012). Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *Journal of Climate*, *25*, 2755–2781. https://doi.org/10.1175/JCLI-D-11-00316.1

Dirmeyer, P. A., Cash, B. A., Kinter, J. L., III, Jung, T., Marx, L., Satoh, M., et al. (2012). Simulating the diurnal cycle of rainfall in global climate models: Resolution versus parameterization. *Climate Dynamics*, *39*, 399–418. https://doi.org/10.1007/s00382-011-1127-9

Donner, L. J., Wyman, B. L., Hemler, R., Horowitz, L. W., & Ming, Y., Zhao, M., et al. (2011). The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3. *Journal of Climate*, *24*, 3484–3519.

Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., et al. (2012). GFDL's ESM2 global coupled climate-carbon Earth System Models, Part I: Physical formulation and baseline simulation characteristics. *Journal of Climate*, *25*, 6646–6665. https://doi.org/10.1175/JCLI-D-11-00560.1

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

Garfinkel, C. I., Waugh, D., & Polvani, L. (2015). Recent Hadley cell expansion: The role of internal atmospheric variability in reconciling modeled and observed trends. *Geophysical Research Letters*, *42*, 10824–10831.

Gehne, M., Hamill, T. M., Kiladis, G. N., & Trenberth, K. E. (2016). Comparison of global precipitation estimates across a range of temporal and spatial scales. *Journal of Climate*, *29*, 7773–7795.

GFDL-GAMDT. (2004). The new GFDL global atmosphere and land model AM2/LM2: Evaluation with prescribed SST simulations. *Journal of Climate*, *17*, 4641–4673.

Gleckler, P. J., Doutriaux, C., Durack, P. J., Taylor, K. E., Zhang, Y., Williams, D. N., et al. (2016). A more powerful reality test for climate models. *Eos, Transactions, American Geophysical Union*, *97*, https://doi.org/10.1029/2016EO051663

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research*, *113*, D06104. https://doi.org/10.1029/2007JD008972

Hall, A., & Qu, X. (2006). Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophysical Research Letters*, *33*, L03502. https://doi.org/10.1029/2005GL025127

Hardiman, S. C., Butchart, N., Hinton, T. J., Osprey, S. M., & Gray, L. J. (2012). The effect of a well-resolved stratosphere on surface climate: Differences between CMIP5 simulations with high and low top versions of the Met Office Climate Model. *Journal of Climate*, *25*, 7083–7099.

Harris, I., Jones, P., Osborn, T., & Lister, D. (2014). Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 dataset. *International Journal of Climatology*, *34*, 623–642.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., et al. (2017). Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the community emissions data system (CEDS). *Geoscientific Model Development*, *11*, 369–408. https://doi.org/10.5194/gmd-11-369-2018

Holben, B., Eck, T., Slutsker, I., Tanre, D., Buis, J., Setzer, A., et al. (1998). AERONET—A federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment*, *66*, 1–6.

Huang, B., Banzon, V., Freeman, E., Lawrimore, J., Liu, W., Peterson, T., et al. (2014). Extended reconstructed sea surface temperature version 4 (ERSST.v4): Part I. *Journal of Climate*, *28*, 911–930. https://doi.org/10.1175/JCLI-D-14-00006.1

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., et al. (2011). Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic Change*, *109*, 117–161. https://doi.org/10.1007/s10584-011-0153-2

Hwang, Y.-T., & Frierson, D. M. W. (2013). Link between the double-intertropical convergence zone problem and cloud biases over the Southern Ocean. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 4935–4940. https://doi.org/10.1073/pnas.1213302110

Jia, L., Yang, X., & Vecchi, G. (2015). Improved seasonal prediction of temperature and precipitation over land in a high-resolution GFDL climate model. *Journal of Climate*, *28*, 2044–2062. https://doi.org/10.1175/JCLI-D-14-00112.1

Kahn, R. A., Nelson, D. L., Garay, M. J., Levy, R. C., Bull, M. A., Diner, D. J., et al. (2009). MISR aerosol product attributes and statistical comparisons with MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, *47*, 4095–4114. https://doi.org/10.1109/TGRS.2009.2023115

Kim, D., Kug, J.-S., Kang, I.-S., Jin, F.-F., & Wittenberg, A. T. (2008). Tropical Pacific impacts of convective momentum transport in the SNU coupled GCM. *Climate Dynamics*, *31*, 213–226. https://doi.org/10.1007/s00382-007-0348-4

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010). The International Best Track Archive for Climate Stewardship (IBTrACS)—Unifying tropical cyclone best track data. *Bulletin of the American Meteorological Society*, *91*, 363–376.

Krishnamurthy, L., Vecchi, G., Msadek, R., Wittenberg, A., Delworth, T., & Zeng, F. (2015). The seasonality of the Great Plains Low-Level Jet and ENSO relationship. *Journal of Climate*, *28*, 4525–4544. https://doi.org/10.1175/JCLI-D-14-00590.1

Levy, R., Hsu, C., et al. (2015). *MODIS Atmosphere L2 Aerosol Product. NASA MODIS Adaptive Processing System*, Goddard Space Flight Center, USA. https://doi.org/10.5067/MODIS/MOD04_L2.006

Levy, R. C., Remer, L., Mattoo, S., Vermote, E. F., & Kaufman, Y. J. (2007). Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of moderate resolution imaging spectroradiometer spectral reflectance. *Journal of Geophysical Research*, *112*, D13211. https://doi.org/10.1029/2006JD007811

Liebmann, B., & Smith, C. A. (1996). Description of a complete (interpolated) Outgoing Longwave Radiation dataset. *Bulletin of the American Meteorological Society*, *77*, 1275–1277.

Loeb, N. G., Wielicki, B. A., Doellingast, D. R., Smith, G. L., Keyes, D. F., Kato, S., et al. (2009). Toward optimal closure of the Earth's top-of-atmosphere radiation budget. *Journal of Climate*, *22*, 748–766. https://doi.org/10.1175/2008JCLI2637.1

Madden, R., & Julian, P. (1971). Detection of a 40–50 day oscillation in the zonal wind in the tropical pacific. *Journal of the Atmospheric Sciences*, *28*, 702–708.

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

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

Milly, P., & Dunne, K. A. (2002). Macroscale water fluxes: 1. Quantifying errors in the estimation of basin mean precipitation. *Water Resources Research*, *38*(10), 1205. https://doi.org/10.1029/2001WR000759

Milly, P., Malyshev, S., Shevliakova, E., Dunne, K. A., Findell, K. L., Gleeson, T., et al. (2014). An enhanced model of land water and energy for global hydrologic and Earth-system studies. *Journal of Hydrometeorology*, *15*, 1739–1761.

Mizuta, R., Yoshimura, H., Murakami, H., Matsueda, M., Endo, H., Ose, T., et al. (2012). Climate simulations using MRI-AGCM3.2 with 20-km grid. *Journal of the Meteorological Society of Japan*, *90*, 233–258.

Murakami, H., Vecchi, G., Underwood, S., Delworth, T., Wittenberg, A., Anderson, W., et al. (2015). Simulation and prediction of category 4 and 5 hurricanes in the high-resolution GFDL HiFLOR coupled climate model. *Journal of Climate*, *28*, 9058–9079. https://doi.org/10.1175/JCLI-D-15-0216.1

Naik, V., Horowitz, L., Fiore, A., Ginoux, P., Mao, J., Aghedo, A. M., et al. (2013). Impact of preindustrial to present-day changes in short-lived pollutant emissions on atmospheric composition and climate forcing. *Journal of Geophysical Research: Atomspheres*, *118*, 8086–8110. https://doi.org/10.1002/jgrd.50608

Pincus, R., Forster, P. M., & Stevens, B. (2016). The radiative forcing model intercomparison project (RFMIP): Experimental protocol for CMIP6. *Geoscientific Model Development*, *9*, 3447–3460. https://doi.org/10.5194/gmd-9-3447-2016

Pincus, R., Platnick, S., Ackerman, S. A., Hemler, R. S., & Hofmann, R. J. P. (2012). Reconciling simulated and observed views of clouds: MODIS, ISCCP, and the limits of instrument simulators. *Journal of Climate*, *25*, 4699–4720. https://doi.org/10.1175/JCLI-D-11-00267.1

Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riedi, J. C., et al. (2003). The MODIS cloud products: Algorithms and examples from Terra. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 459–473. https://doi.org/10.1109/TGRS.2002.808301

Quinn, P. K., Bates, T., Miller, T., Coffman, D., Johnson, J., Harris, J., et al. (2000). Surface submicron aerosol chemical composition: What fraction is not sulfate? *Journal of Geophysical Research*, *116*, 6785–6805. https://doi.org/10.1029/1999JD901034

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., et al. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of Climate*, *24*, 3624–3648.

Rossow, W. B., & Schiffer, R. A. (1991). ISCCP cloud data products. *Bulletin of the American Meteorological Society*, *72*, 2–20.

Scaife, A. A., Knight, J. R., Vallis, G. K., & Folland, C. K. (2005). A stratospheric influence on the winter NAO and North Atlantic surface climate. *Geophysical Research Letters*, *32*, L18715. https://doi.org/10.1029/2005GL023226

Seviour, W., Waugh, D., Polvani, L., Correa, G., & Garfinkel, C. (2017). Robustness of the simulated tropospheric response to ozone depletion. *Journal of Climate*, *30*, 2577–2585.

Shaevitz, D. A., Camargo, S. J., Sobel, A. H., Jonas, J. A., Kim, D., Kumar, A., et al. (2014). Characteristics of tropical cyclones in high-resolution models in the present climate. *Journal of Advances in Modeling Earth Systems*, *6*, 1154–1172. https://doi.org/10.1002/2014MS000372

Shaw, T. A., & Perlwitz, J. (2010). The impact of stratospheric model configuration on planetary-scale waves in Northern Hemisphere winter. *Journal of Climate*, *23*, 3369–3389.

Sheffield, J., & Goteti, G., & Wood, E. F. (2006). Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate*, *19*, 3088–3111.

Shen, Z., Ming, Y., Horowitz, L. W., & Ramaswamy, V. (2017). On the seasonality of Arctic black carbon. *Journal of Climate*, *30*, 4429–4441. https://doi.org/10.1175/JCLI-D-16-0580.1

Sherwood, S. C., Bony, S., Boucher, O., Bretherton, C., Forster, P. M., Gregory, J. M., et al. (2015). Adjustments in the forcing-feedback framework for understanding climate change. *Bulletin of the American Meteorological Society*, *96*, 217–228. https://doi.org/10.1175/BAMS-D-13-00167.1

Siongco, A. C., Hohenegger, C., & Stevens, B. (2015). The Atlantic ITCZ bias in CMIP5 models. *Climate Dynamics*, *45*, 1169–1180. https://doi.org/10.1007/s00382-010-0742-1

Siorois, A., & Barrie, L. A. (1999). Arctic lower tropospheric aerosol trends and composition at Alert, Canada: 1980–1995. *Journal of Geophysical Research*, *104*, 11599–11618.

Smirnov, A., Holben, B., Eck, T., Dubovik, O., & Slutsker, I. (2000). Cloud-screening and quality control algorithms for the AERONET database. *Remote Sensing of Environment*, *73*, 337–349.

Smith, G. L., Priestley, K. J., Loeb, N. G., Wielicki, B. A., Charlock, T. P., Minnis, P., et al. (2011). Clouds and Earth Radiant Energy System (CERES), a review: Past, present and future. *Advances in Space Research*, *48*, 254–263. https://doi.org/10.1016/j.asr.2011.03.009

Stan, C., & Xu, L. (2014). Climate simulations and projections with a super-parameterized climate model. *Environmental Modelling and Software*, *60*, 134–152.

Taylor, K., Williamson, D., & Zwiers, F. (2000). *The sea surface temperature and sea ice concentration boundary conditions for AMIP II simulations* (PCMDI Rep. 60, pp. 1–25). Livermore, CA: Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory.

Theil, H. (1950a). A rank-invariant method of linear and polynomial regression analysis I, *Nederlandse Akademie van Wetenschappen Proceedings*, *53*, 386–392.

Theil, H. (1950b). A rank-invariant method of linear and polynomial regression analysis II, *Nederlandse Akademie van Wetenschappen Proceedings*, *53*, 521–525.

Theil, H. (1950c). A rank-invariant method of linear and polynomial regression analysis III, *Nederlandse Akademie van Wetenschappen Proceedings*, *53*, 1397–1412.

Thompson, D. W. J., & Solomon, S. (2002). Interpretation of recent Southern Hemisphere climate change. *Science*, *296*, 895–899. https://doi.org/10.1126/science.1069270

Tsushima, Y., & Manabe, S. (2013). Assessment of radiative feedback in climate models using satellite observations of annual flux variation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 7568–7573. https://doi.org/10.1073/pnas.1216174110

Udisti, R., Bazzano, A., Becagli, S., Bolzacchini, E., Caiazzo, L., Cappelletti, D., et al. (2016). Sulfate source apportionment in the Ny-Alesund (Svalbard Islands) Arctic aerosol. *Reniconti Lincei*, *27*, 85–94.

Uppala, S., Kållberg, P., Simmons, A., Andrae, U., da Costa Bechtold, V., Fiorino, M., et al. (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, *131*, 2961–3012.

van Marle, M. J. E., Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A.-L., Field, R. D., et al. (2017). Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fire models (1750–2015). *Geoscientific Model Development*, *10*, 3329–3357. https://doi.org/10.5194/gmd-10-3329-2017

Waliser, D., Lau, K., & Kim, J.-H. (1999). The influence of coupled sea surface temperatures on the Madden-Julian oscillation: A model perturbation experiment. *Journal of the Atmospheric Sciences*, *56*, 333–358.

Walsh, K., Camargo, S., Vecchi, G. A., Daloz, A. S., Elsner, J., Emanuel, K., et al. (2015). Hurricanes and climate: The U.S. CLIVAR Working Group on Hurricanes. *Bulletin of the American Meteorological Society*, *96*, 997–1017.

Walsh, K., Fiorino, M., Landsea, C., & McInnes, K. (2007). Objectively determined resolution-dependent threshold criteria for the detection of tropical cyclones in climate models and reanalyses. *Journal of Climate*, *20*, 2307–2314.

Weller, R., Wagenbach, D., Legrand, M., Elsasser, C., Tian-Kunze, X., & Konig-Langlo, G. (2011). Continuous 25-yr aerosol records at coastal Antarctica—I: Inter-annual variability of ionic compounds and links to climate indices. *Tellus, Series B*, *63*, 901–919. https://doi.org/10.1111/j.1600-0889.2011.00542.x

Wheeler, M., & Kiladis, G. N. (1999). Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber-frequency domain. *Journal of the Atmospheric Sciences*, *56*, 374–399.

Wittenberg, A. (2009). Are historical records sufficient to constrain ENSO simulations? *Geophysical Research Letters*, *36*, L12702. https://doi.org/10.1029/2009GL038710

Wittenberg, A., Rosati, A., Lau, N.-C., & Ploshay, J. J. (2006). GFDL's CM2 global coupled climate models. Part III: Tropical Pacific climate and ENSO. *Journal of Climate*, *19*, 698–722.

Xiang, B., Zhao, M., & Held, I. M. (2017). Predicting the severity of spurious "double ITCZ" problem in CMIP5 coupled models from AMIP simulations. *Geophysical Research Letters*, *44*, 1520–1527. https://doi.org/10.1002/2016GL071992

Zhao, M. (2014). An investigation of the connections among convection, clouds, and climate sensitivity in a global climate model. *Journal of Climate*, *27*, 1845–1862.

Zhao, M., Golaz, J.-C., Held, I., Ramaswamy, V., Lin, S.-J., Ming, Y., et al. (2016). Uncertainty in model climate sensitivity traced to representations of cumulus precipitation microphysics. *Journal of Climate*, *29*, 543–560.

Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., et al. (2018). The GFDL global atmosphere and land model AM4.0/LM4.0: 2. Model description, sensitivity studies, and tuning strategies. *Journal of Advances in Modeling Earth Systems*. https://doi.org/10.1002/2017MS001209, in press.

Zhao, M., & Held, I. M. (2010). An analysis of the effect of global warming on the intensity of Atlantic hurricanes using a GCM with statistical refinement. *Journal of Climate*, *23*, 6382–6393.

Zhao, M., Held, I. M., Lin, S.-J., & Vecchi, G. A. (2009). Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50 km resolution GCM. *Journal of Climate*, *22*, 6653–6678.