

## RESEARCH ARTICLE

10.1002/2017MS001209

Companion to Zhao et al. [2018],  
<https://doi.org/10.1002/2017MS001208>.

## Key Points:

- Part 2 of the AM4.0/LM4.0 paper provides documentation of key changes in individual components from previous GFDL models
- Some sensitivities to choices of model formulation and parameter values are presented with emphasis on convection and tuning of clouds
- We describe the extent to which the effect on aerosol forcing and Cess sensitivity has been taken into account during model development

## Supporting Information:

- Supporting Information S1
- Figures S1–S9

## Correspondence to:

M. Zhao,  
[Ming.Zhao@noaa.gov](mailto:Ming.Zhao@noaa.gov)

## Citation:

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*, 10, 735–769. <https://doi.org/10.1002/2017MS001209>

Received 18 OCT 2017

Accepted 15 FEB 2018

Accepted article online 19 FEB 2018

Published online 15 MAR 2018

© 2018. The Authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

This article has been contributed to by US Government employees and their work is in the public domain in the USA.

## The GFDL Global Atmosphere and Land Model AM4.0/LM4.0: 2. Model Description, Sensitivity Studies, and Tuning Strategies

M. Zhao<sup>1</sup> , J.-C. Golaz<sup>1,2</sup> , I. M. Held<sup>1</sup> , H. Guo<sup>3</sup>, V. Balaji<sup>4</sup> , R. Benson<sup>1</sup>, J.-H. Chen<sup>1</sup> , X. Chen<sup>4</sup> , L. J. Donner<sup>1</sup> , J. P. Dunne<sup>1</sup> , K. Dunne<sup>5</sup>, J. Durachta<sup>1</sup>, S.-M. Fan<sup>1</sup> , S. M. Freidenreich<sup>1</sup> , S. T. Garner<sup>1</sup>, P. Ginoux<sup>1</sup> , L. M. Harris<sup>1</sup> , L. W. Horowitz<sup>1</sup> , J. P. Krasting<sup>1</sup> , A. R. Langenhorst<sup>1,8</sup>, Z. Liang<sup>1</sup>, P. Lin<sup>1</sup> , S.-J. Lin<sup>1</sup>, S. L. Malyshev<sup>1</sup> , E. Mason<sup>6</sup>, P. C. D. Milly<sup>1,5</sup>, Y. Ming<sup>1</sup> , V. Naik<sup>1</sup> , F. Paulot<sup>4</sup> , D. Paynter<sup>1</sup> , P. Philipps<sup>1</sup>, A. Radhakrishnan<sup>6</sup>, V. Ramaswamy<sup>1</sup> , T. Robinson<sup>6</sup> , D. Schwarzkopf<sup>1</sup> , C. J. Segan<sup>1</sup>, E. Shevliakova<sup>1</sup>, Z. Shen<sup>4</sup> , H. Shin<sup>4</sup> , L. G. Silvers<sup>4</sup> , J. R. Wilson<sup>7</sup>, M. Winton<sup>1</sup> , A. T. Wittenberg<sup>1</sup> , B. Wyman<sup>1</sup>, and B. Xiang<sup>3</sup> 
<sup>1</sup>GFDL/NOAA, Princeton, NJ, USA, <sup>2</sup>Lawrence Livermore National Laboratory, Livermore, CA, USA, <sup>3</sup>UCAR/GFDL, Princeton, NJ, USA, <sup>4</sup>Princeton University/GFDL, Princeton, NJ, USA, <sup>5</sup>U.S. Geological Survey, Princeton, NJ, USA, <sup>6</sup>Engility Corporation/GFDL, NJ, USA, <sup>7</sup>NASA Ames Research Center, Moffett Field, CA, USA, <sup>8</sup>Deceased 31 March 2016

**Abstract** In Part 2 of this two-part paper, documentation is provided of key aspects 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). The quality of the simulation in AMIP (Atmospheric Model Intercomparison Project) mode has been provided in Part 1. Part 2 provides documentation of key components and some sensitivities to choices of model formulation and values of parameters, highlighting the convection parameterization and orographic gravity wave drag. The approach taken to tune the model's clouds to observations is a particular focal point. Care is taken to describe the extent to which aerosol effective forcing and Cess sensitivity have been tuned through the model development process, both of which are relevant to the ability of the model to simulate the evolution of temperatures over the last century when coupled to an ocean model.

## 1. Introduction

A description of the simulation, in AMIP mode, provided by the AM4.0/LM4.0 atmosphere/land model has been provided in Zhao et al. (2018) (below referred to as Part 1). This model has roughly 100 km horizontal resolution, a relatively low top with 33 levels, and a "light" chemistry that allows simulation of aerosols from emissions but prescribes ozone and other oxidants. The land model assumes static vegetation. Models with more stratospheric resolution and with more comprehensive stratospheric and tropospheric chemistry will be described elsewhere. Aspects of the simulation that have been discussed in Part 1 include top-of-atmosphere (TOA) radiative fluxes, mean precipitation, surface pressure, surface land temperatures, zonal-mean temperatures and zonal winds, extratropical transients, the Madden-Julian Oscillation (MJO), tropical storms, stratospheric variability, responses of precipitation and surface fluxes to sea surface temperature (SST) anomalies associated with the El Niño-Southern Oscillation (ENSO), the aerosols simulation, and the resulting Cess sensitivity and aerosol radiative forcing. The description of the simulation in Part 1 is necessarily incomplete but includes some impressive aspects of the simulation (e.g., TOA fluxes and surface responses to ENSO), some persistent problems (e.g., the "Philippines hotspot" in precipitation and the diurnal cycle of precipitation over land), and some aspects that are understandably deficient due to low horizontal or stratospheric resolution in this version (tropical storms and stratospheric influences on tropospheric winds, respectively).

The goal in Part 2 is to explain the key choices made in the development of the model, describe some sensitivities to model formulation and parameter values, and discuss the optimization strategy. The model formulation is first described in a series of sections, focusing on changes from previous GFDL models, AM2.1 (GFDL-GAMDT, 2004), AM3 (Donner et al., 2011), and HiRAM (Zhao et al., 2009). The topics covered will be the atmospheric dynamical core, the radiation code and radiative forcing, the shallow and deep convection

schemes, the orographic and nonorographic gravity wave drag formulation, the light chemistry and aerosol modules, the clouds and aerosol-cloud interaction scheme, the PBL and surface flux formulation, the land model, and the computational efficiency and workflow (sections 2–11). The sensitivity of the simulation to some of the choices within these model components is discussed within each section as appropriate, with particular attention to the choices in the convection, cloud, and orographic gravity wave drag schemes. We make a special effort to describe our optimization strategy for the simulation of clouds and TOA fluxes in section 12. Finally, we try to make clear the extent to which we have tuned toward values of RFP and Cess sensitivity in sections 13 and 14.

For studies of model sensitivity to changes in various parameterizations, we use simulations prescribed with no interannual variability of SSTs or sea-ice extent. The control we use for this purpose (referred to here as C0 and in Part 1 as 2010RAD) uses radiative forcing agents from 2010 and the climatological SSTs and sea-ice averaged for the 1981–2014 period based on the CMIP6 specification as described in Part 1. In general, C0 produces model mean climatology as well as the subseasonal to seasonal variability very similar to its corresponding AMIP simulations shown in Part 1. The quality of a model's simulation of present-day climatology and variability forced by this climatological SST, sea-ice concentration, and fixed radiative forcings is also well correlated to the quality of its corresponding AMIP simulation, thus this kind of climatological run has played an important role during the AM4 development and tuning processes.

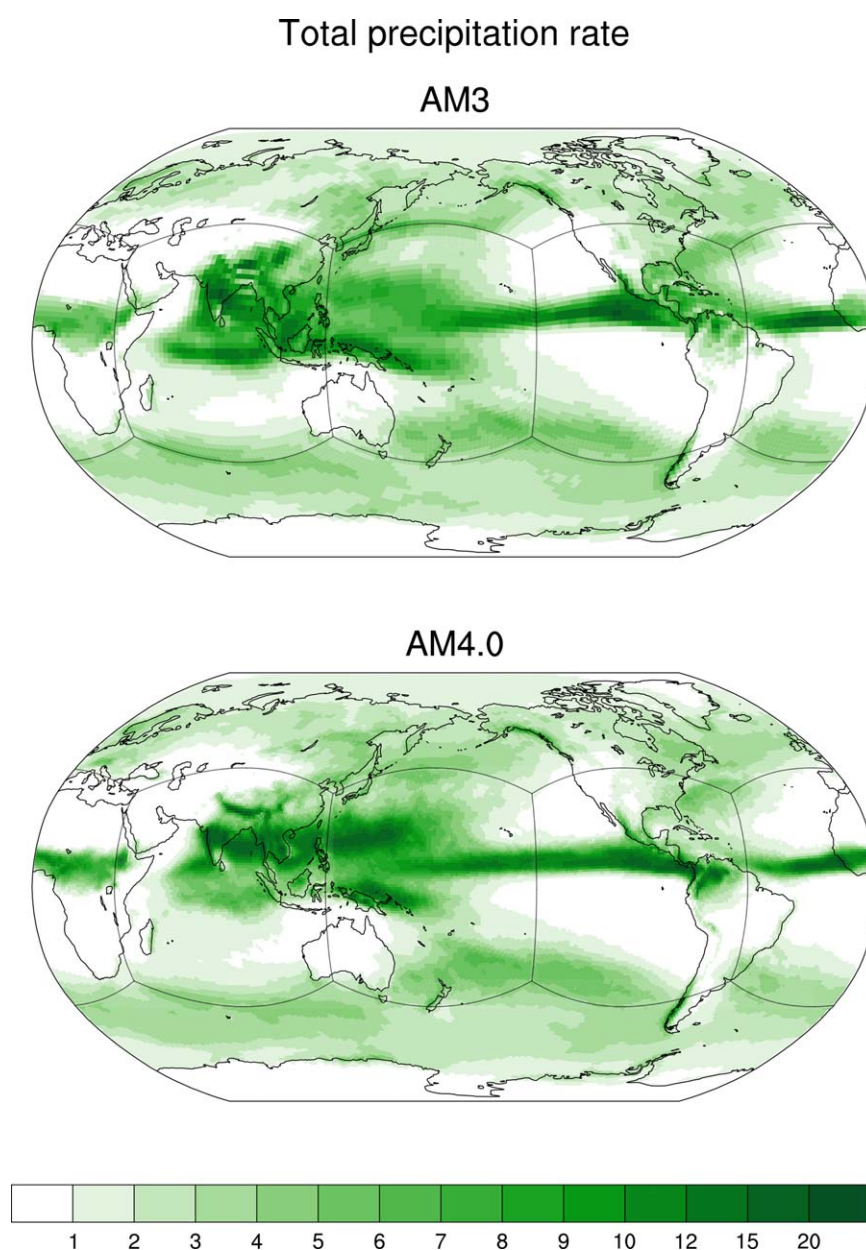
## 2. Dynamical Core

AM4.0 uses the most recent version of the GFDL Finite-Volume Cubed-Sphere dynamical core (FV<sup>3</sup>; Harris & Lin, 2013; Putman & Lin, 2007) developed for weather and climate applications at GFDL, NASA Goddard Space Flight Center, and at the National Centers for Environmental Prediction. FV<sup>3</sup> solves the hydrostatic primitive equations (with a runtime option to instead solve the nonhydrostatic fully compressible Euler equations) using the finite-volume solution method described by Lin and Rood (1997) based on the multidimensional monotone transport scheme of Lin and Rood (1996), with the finite-volume pressure gradient force of Lin (1997) and Lagrangian vertical coordinate described by Lin (2004). The dynamical core is broadly configured the same as in HiRAM and AM3 (AM2.1 utilizes an older version of the finite-volume code on a lat-lon grid), although many minor refinements, bug fixes, and performance improvements have been implemented.

One significant change in the configuration of the dynamics in AM4.0 is the use of a sixth-order divergence damping (described in Zhao et al., 2012) rather than the more diffusive second-order damping used in AM3. The divergence damping can have strong effects on the TC frequency in the model, with stronger damping resulting in more TCs, as described in Zhao et al. (2012). We have not considered the overall frequency of TCs as a critical target for this 100 km resolution model, and we have not tuned the divergence damping strength to optimize TC numbers, which are, in any case, sensitive to the criteria used to count TCs. This issue is of more relevance for the tuning of higher resolution versions of the model.

Also, the previous means of applying diffusion in the model-top sponge layer by the use of first-order upwind fluxes has been deprecated, and is replaced by a second-order horizontal hyperdiffusion. This allows finer control of the strength of the diffusive sponge-layer damping. In AM4.0, the sponge is only applied at the top two model levels, so is primarily meant to control noise rather than to prevent reflection of planetary waves. The model simulation in the stratosphere is sensitive to this change in the sponge but the convective gravity wave drag was adjusted to compensate for changes in the sponge to minimize changes in mean winds. In higher top versions of the model, we find that the sponge has less effect on the circulation and there is less need to adjust the convective gravity wave drag when the sponge is altered.

A concern when using the cubed-sphere grid is imprinting of the grid on the model's climatology. Figure 1 shows an example of the grid imprinting in AM3 in a plot of the climatological mean precipitation distribution in August, the season in which this distortion is most evident. The effect on precipitation of the northern edge of the cube face that passes through the North Pacific is especially visible. It is the large-scale rather than parameterized convective precipitation that suffers from this imprinting. This is the worst case of distortion that we have found in any single month in this field. In this figure, we plot the precipitation on the model's native cubed-sphere grid because interpolation to a lat-lon grid obscures this distortion. We also show the same result for AM4.0 in which the imprinting is hardly visible in any months that we have



**Figure 1.** August climatological mean precipitation (unit:  $\text{mm d}^{-1}$ ) plotted in (top) AM3 and (bottom) AM4.0's native cubed-sphere grid.

visualized. We have not had the opportunity to study the cause of this difference, but we presume that it is primarily due to the factor of two reduction in grid size in AM4.0.

The treatment of dissipation of kinetic energy and associated energy conservation is described in supporting information S1. The overall structure of the interaction between the dynamics and physics is unchanged from previous GFDL models. A distinguishing feature of this structure is the separation of the atmospheric code into two parts, which we refer to Atmosphere-down and Atmosphere-up, with the state of the land model updated between these two steps. This structure allows a fully implicit treatment of the diffusive fluxes in the atmosphere and land without any need for iteration, while retaining a fully modular separation between atmosphere and land models, as described in Balaji (2012).

There are several distinct time steps in the model; from shortest to longest, these are the gravity wave and the Lagrangian dynamics step (both 150 s), the vertical remapping and the horizontal advection step

(1,800 s), the physics and the land step (also 1,800 s), and the radiation steps, which differ for the shortwave (1 h) and longwave (3 h) computations. The choice of different radiation time steps for short and long waves is discussed in section 3.

### 3. Radiation

Both the longwave and shortwave components of the GFDL radiation code have been extensively benchmarked and updated for AM4. This represents the first significant update to the GFDL radiation code since AM2 (GFDL-GAMDT, 2004), as the updates for AM3 (Donner et al., 2011) were generally minor. The longwave and shortwave codes can be considered as independent and will be dealt with separately in the following. These updates and their resulting impact upon climate will be detailed in an upcoming paper. Here we summarize the key changes.

The longwave code continues to be based upon the simplified exchange approximation (SEA; Fels & Schwarzkopf, 1975; Schwarzkopf & Fels, 1991; Schwarzkopf & Ramaswamy, 1999). For the AM4 version, we have retained the SEA approximation for H<sub>2</sub>O line calculations but have updated the spectral information used for other species. Random band coefficients for H<sub>2</sub>O, O<sub>3</sub>, CO<sub>2</sub>, N<sub>2</sub>O, CH<sub>4</sub>, and halocarbons were all updated from HITRAN 2000 (Rothman et al., 2003) to HITRAN 2012 (Rothman et al., 2013), while the water vapor continuum was updated from CKD 2.1 (Clough et al., 1989) to MT-CKD 2.5 (Mlawer et al., 2012). The code also allows for and provides accurate calculation of radiative transfer at much higher quantities of CO<sub>2</sub>, CH<sub>4</sub>, and N<sub>2</sub>O than previously. CO<sub>2</sub> amounts may be as high as 10,000 ppmv; CH<sub>4</sub> (N<sub>2</sub>O) amounts may range up to 6,000 (800) ppbv. The interpolation algorithm for these species has also been modified to increase code stability for the higher species amounts.

The most significant update to the longwave radiation code is the inclusion of the CO<sub>2</sub> bands in the 990–1,200 cm<sup>−1</sup> range (commonly called the 10 μm band). While these weaker CO<sub>2</sub> lines have a small impact on forcing due to present-day CO<sub>2</sub>, they become increasingly significant for higher amounts of CO<sub>2</sub>. Accordingly, our 2010 effective radiative forcing (ERF) due to CO<sub>2</sub> (Pincus et al., 2016) increases by 10%. The ERF from quadrupling CO<sub>2</sub> increases from 7.74 to 8.46 W m<sup>−2</sup>. This update also makes the GFDL radiation code estimate of CO<sub>2</sub> instantaneous clear-sky radiative forcing agree much better with clear-sky forcing estimates of a benchmark line-by-line radiation code (Paynter & Ramaswamy, 2012). For example in the CIRC1 (Oreopoulos et al., 2012), benchmark atmosphere (which represents a typical midlatitude summer case) the instantaneous forcing at the tropopause due to a quadrupling of CO<sub>2</sub> was previously 9% too small; now it is within 1% of the line-by-line calculation.

The impact of all the longwave updates on the control climatological simulation (C0) is modest. For example, the updates only alter the clear-sky OLR by 0.40 W m<sup>−2</sup> and clear-sky downwelling surface radiation by 1.9 W m<sup>−2</sup>. Nearly all of these changes are due to the water vapor continuum changing from CKD 2.1 to MTCKD 2.5. When benchmarked against a line-by-line radiation code these updates do not lead to a significant improvement, with both old and new code underestimating OLR by 2% in CIRC2 (a typical tropical profile) and 1% for CIRC4 (a typical Arctic summer case). We have identified the major causes of these underestimates, which are related to the treatment of O<sub>3</sub> and H<sub>2</sub>O and also to the neglect of bands above 2,200 cm<sup>−1</sup>. An update to fix these issues is in preparation for future releases of AM4.

The shortwave code is an update of the 18 band formulation utilized in AM2 and AM3 simulations (Freidenreich & Ramaswamy, 2005). We have updated the H<sub>2</sub>O, CO<sub>2</sub>, and O<sub>2</sub> formulations, and added parameterizations to account for shortwave absorption due to the CH<sub>4</sub> and N<sub>2</sub>O, based on line-by-line reference calculations utilizing a more recent version (2012) of the HITRAN catalog. In addition, we have incorporated the effects of the shortwave water self continuum (BPS 2.0), updated the foreign continuum from CKD 2.1 to BPS 1.1 (Paynter & Ramaswamy, 2012, 2014) and added the effects of the O<sub>2</sub> and N<sub>2</sub> continua. Two specific improvements are (1) including additional exponential-sum-fit terms in the parameterization of H<sub>2</sub>O to now account for stratospheric absorption and improve tropospheric absorption and (2) utilizing the exponential-sum-fit technique to more accurately parameterize the CO<sub>2</sub> absorption. The impact of these updates has been to considerably increase the shortwave absorption by the atmosphere in C0. The globally average shortwave absorption by the atmosphere is now 76.2 W m<sup>−2</sup> in AM4.0 AMIP simulation (1981–2014) compared to 74.3 W m<sup>−2</sup> in AM3 and AM2 AMIP simulation. This increase in atmospheric absorption helps to lower AM4.0 global mean precipitation, which, as discussed in Part 1, is biased high compared to



**Table 1**
*Sensitivity of AM4.0 Simulated TOA Radiative Fluxes (unit:  $W m^{-2}$ ) to the Model's SW Radiation Time Step  $\Delta t_{swrad}$* 

Exp	OLR	SWABS	NETRAD	LW CRE	SW CRE	Total CRE
As C0 except $\Delta t_{swrad} = 0.5$ h	238.52	240.57	2.05	23.68	−48.40	−24.72
C0 (AM4.0 CNTL) ( $\Delta t_{swrad} = 1$ h)	238.54	240.23	1.69	23.68	−48.54	−24.86
As C0 except $\Delta t_{swrad} = 2$ h	238.46	239.10	0.64	23.77	−49.10	−25.33
As C0 except $\Delta t_{swrad} = 3$ h	238.70	238.15	−0.55	23.57	−49.34	−25.77

*Note.* 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.

standard observational data sets, but the changes in this direction are modest compared to the size of the bias.

In the CIRC2 benchmark, the shortwave absorption by atmosphere is now within 1% of the line-by-line calculation, compared to a previous underestimate of 6% in the AM3 code. We see similar results in the benchmark CIRC4 atmosphere (a typical Arctic summer case), where a previous 6% underestimate is reduced to within 1% of the line-by-line calculation. Recent analysis comparing the shortwave code at every GCM grid point for a day also reveals a less than 1% bias. A list of the specific differences in radiation code between AM3 and AM4 is provided in supporting information Table S1.

The radiative fluxes in AM4.0 have some sensitivity to the radiative time step, as discussed in Balaji et al. (2016). Comparing simulations with a radiative time step of 3 h (the AM3 value) and 30 min (the physics time step at which all radiatively active species are updated), the difference in the net TOA flux is substantial,  $\sim 3 W m^{-2}$ . Most of this difference is in the shortwave and the major part of this shortwave difference is due to the cloud radiative effect. The shortwave computation in all GFDL models since AM2 includes a solar interpolator that renormalizes all shortwave fluxes each physics time step according to the evolving incident solar flux. But the atmospheric state (temperatures and all radiatively active species) enter the radiation computation only at each radiative time step. While we do not have a complete analysis of this sensitivity to time step, we believe that it may be primarily associated with the diurnal cycle of low clouds over the oceans. (The sign of the dependence on time step can be understood if it is dominated by the part of the day in which cloud cover is decaying, being burned away as the incident shortwave flux is increasing, resulting in an overestimate of the reflection.)

As illustrated in Table 1, simulations with a 1 h time step produce results much closer to the 1/2 h case. Because the longwave computation is roughly 50% of the total radiation workload when using equal shortwave and longwave time steps, we have compromised between accuracy and efficiency in AM4.0 by choosing a 1 h shortwave step, but retaining the 3 h step for the longwave.

#### 4. Moist Convection

A variety of convection schemes have been utilized in the GFDL AM2-AM3-HIRAM series of models. A Relaxed Arakawa-Schubert scheme (Moorthi & Suarez, 1992) is used in AM2. This is a unified scheme for both shallow and deep convection involving a spectrum of entraining plumes which detrains only at the plume top, with a mass flux closure determined by relaxation of the cloud work function in which the relaxation time and precipitation production efficiency are specified to vary across the plume spectrum. The entraining plume model which detrains only at plume top has long been argued as a poor representation of shallow cumulus clouds (e.g., Bretherton et al., 2004a; Emanuel, 1991; Raymond & Blyth, 1986; von Salzen & McFarlane, 2002; Zhao & Austin, 2003). The plumes in AM2 are tuned in a way that undilute plumes are permitted for shallow plumes with top below 500 hPa while they are prevented for deep plumes with top above 500 hPa following a procedure suggested by Tokioka et al. (1988). The model's TOA radiation could not plausibly be tuned to balance when undilute shallow plumes are also prevented from forming. Moreover, the critical cloud work function toward which the convective plumes are relaxed is set to zero for all plumes with top below 600 hPa. With this tuning, the shallow convection in AM2 tends to be dominated by the least entraining plumes and the simulated low cloud amount and condensate tend to peak at a higher level (600–800 hPa) than in observations and other models (Wyant et al., 2006).

In addition to the unrealistic aspects of the mixing behavior in its shallow convection, the treatment of precipitation in AM2 convection scheme is also crude. A plume precipitates only at its top with an arbitrarily specified efficiency parameter. There is extreme sensitivity of the model's TOA energy balance to the precipitation efficiency of the most weakly entraining plumes (Held et al., 2007). Furthermore, the AM2 convective plume ignores the convective inhibition below its neutral buoyancy level and it does not explicitly compute its buoyancy and vertical velocity. This makes it difficult to couple with an aerosol activation scheme for which the vertical velocity is a key.

In part due to these unrealistic aspects of the shallow convection in AM2, shallow and deep convection were separated in AM3, with shallow convection replaced by the University of Washington Shallow Cumulus scheme originally developed by Bretherton et al. (2004a). This is a continuously entraining-detraining bulk plume with the profiles of entrainment and detraining determined by a simple parcel buoyancy-sorting method. It also includes plume-top entrainment, taking air from layers between the neutral buoyancy level and the point at which the plume vertical motion reaches zero and mixing it into the air at neutral buoyancy level. This scheme has support from observations and large-eddy-simulations (LES) of shallow cumulus clouds (e.g., Siebesma et al., 2003; Zhao & Austin, 2005a, 2005b). The cloud base mass flux for this shallow plume is determined by convective inhibition and boundary layer turbulent velocities, similar to that proposed by Mapes (2000). The deep convection in AM3 is treated using a version of the Donner et al. (2001) scheme discretized on the model's native vertical grid, in which it assumes a spectrum of entraining plumes with a fixed fraction of the cloud base mass flux in each plume and the total cloud base mass flux is determined by relaxation of the convective available potential energy (CAPE). AM3 also includes an explicit treatment of cloudiness and precipitation from convective anvils.

The goal in HiRAM was to simplify the treatment of convection while trying to produce a high-quality climate simulation, with an eye toward convection schemes appropriate for higher resolution models. To this end, the explicit treatment of convective anvils was removed with the hope that these anvils would be produced by the resolved scale as horizontal resolution increased, and the shallow and deep convection were again unified using a single plume of the University of Washington Shallow Cumulus type to represent all convection. The overall strength of the lateral mixing between environment and plume was reduced to make it easier for the plume to penetrate deeply. Some local modifications to the plume model described by Bretherton et al. (2004a) have previously been documented in Zhao et al. (2009, Appendix A). This modified version has been used in GFDL HiRAM for the CMIP5 high resolution time-slice simulations and the intermodel comparison project for the US CLIVAR Hurricane Working Group (Walsh et al., 2015).

Despite its simplicity, HiRAM, and an early version of AM4 incorporating HiRAM's convective scheme, produce high-quality simulations of the tropical atmosphere, including tropical cyclone statistics in versions of the model with 50 km grid size or smaller, when the model is forced by observed SSTs. Some noteworthy exceptions are that, especially with grid size of 100 km or larger, the model produces too little rainfall over tropical continents, as noted in Zhao et al. (2009), and too weak MJO. Furthermore, when this HiRAM-like version of AM4 is coupled to an ocean model we find that it produces overly strong equatorial Pacific cold biases and too zonal a South Pacific convergence zone (SPCZ), which in turn negatively affect model simulations of ENSO. Our hypothesis is that these deficiencies are at least in part related to the inability of a single fractional lateral mixing rate  $\varepsilon$  in HiRAM's single bulk plume to simultaneously mimic shallow and deep convection (Zhao et al., 2009).

In AM4.0, we introduce an additional bulk plume so that the new scheme can contain two plumes at a given time, one for shallow and one for deep convection. We refer to the resulting new scheme as a double plume convection scheme. The assumptions concerning the determination of the entrainment and detraining rate through inhomogeneous mixing and a buoyancy-sorting algorithm and the representations of vertical velocity in each of the two plumes are identical to the original scheme described in Bretherton et al. (2004a) except that the fractional lateral mixing rates  $\varepsilon$  in the two plumes are parameterized differently. We use a simple constant value  $\varepsilon_s = 3 \text{ km}^{-1}$  for the shallow plume. The fractional lateral mixing rate of the deep plume  $\varepsilon_d$  is parameterized to be a linear function of the free troposphere column relative humidity  $\zeta = W/W_*$ , where  $W$  and  $W_*$  denote respectively the vertically integrated specific humidity and saturation specific humidity above the planetary boundary layer.

$$\varepsilon_d = \varepsilon_1 + \frac{\zeta - \zeta_0}{1 - \zeta_0} (\varepsilon_2 - \varepsilon_1), \zeta_0 \leq \zeta \leq 1 \quad (1)$$

where  $\zeta_0 = 0.4$ ,  $\varepsilon_1 = 0.9 \text{ km}^{-1}$ , and  $\varepsilon_2 = 0.1 \text{ km}^{-1}$ . While  $\varepsilon_s$ ,  $\varepsilon_1$ , and  $\varepsilon_2$  are tunable parameters, the values for shallow and deep plumes that we have chosen are broadly consistent with those used in typical GCM cumulus convection schemes. They are also consistent with a recent numerical study that produces a measure of cumulus entrainment rates (i.e., roughly  $2\text{--}3 \text{ km}^{-1}$  for shallow convection and  $0.5\text{--}1 \text{ km}^{-1}$  for deep convection) using a LES of cumulus clouds (Roms, 2011).

In this parameterization, deep convection may occur only when the ambient environment is sufficiently moist (i.e.,  $\zeta > \zeta_0$ ). Reducing the lateral mixing rate in a more humid environment attempts to very roughly account for the effect of convective organization on cumulus mixing because a more humid atmosphere is conducive for organized convection, resulting in less mixing of air with the properties of the large-scale environment (e.g., Bretherton et al., 2004b; Brown & Zhang, 1997; Holloway & Neelin, 2009; Mapes & Neale, 2011; Redelsperger et al., 2002; Sherwood, 1999; Zhang et al., 2003). While the nature of convective organization in modifying cumulus mixing remains a topic of active research, a more realistic representation of the convection organization would require a sophisticated model which accounts for not only the lateral mixing rate but also the cloud immediate environment (Becker et al., 2017, 2018). Thus, our current approach was deemed to be expedient in this development cycle. Similar approaches have also been used in the ECMWF model and other models for improving simulations of tropical variability such as the MJO, and previous studies suggest broad support by observations and the cloud resolving model (CRM) results (e.g., Bechtold et al., 2008; DelGenio, 2012; Derbyshire et al., 2004; de Rooy et al., 2013; Johnson, 1997; Raymond, 2000; Raymond & Zeng, 2000; Redelsperger et al., 2002; Sherwood, 1999; Tompkins, 2001).

For the deep plume, AM4.0 uses a cloud work function (CWFN) relaxation closure to determine the cloud base mass flux (e.g., Arakawa & Schubert, 1974; Moorthi & Suarez, 1992; Neale et al., 2008; Zhang & McFarlane, 1995). Specifically, we relax the CWFN defined as the vertical integral of plume buoyancy from cloud base to the level of neutral buoyancy, to a reference value  $\text{CWFN}_0 = 10 \text{ J kg}^{-1}$  with the relaxation time scale  $\tau_d$  set to be 8 h. The relaxation time and the small value to which the CWFN is relaxed are tuning parameters. While it is not ideal to use independent closures for shallow and deep convective mass fluxes, the decoupling of the deep convective mass flux from any measure of convective inhibition was deemed expedient, in part because the latter is poorly resolved over the ocean by the model's coarse vertical grid and can be noisy in time, and results in a model that is not easily tuned.

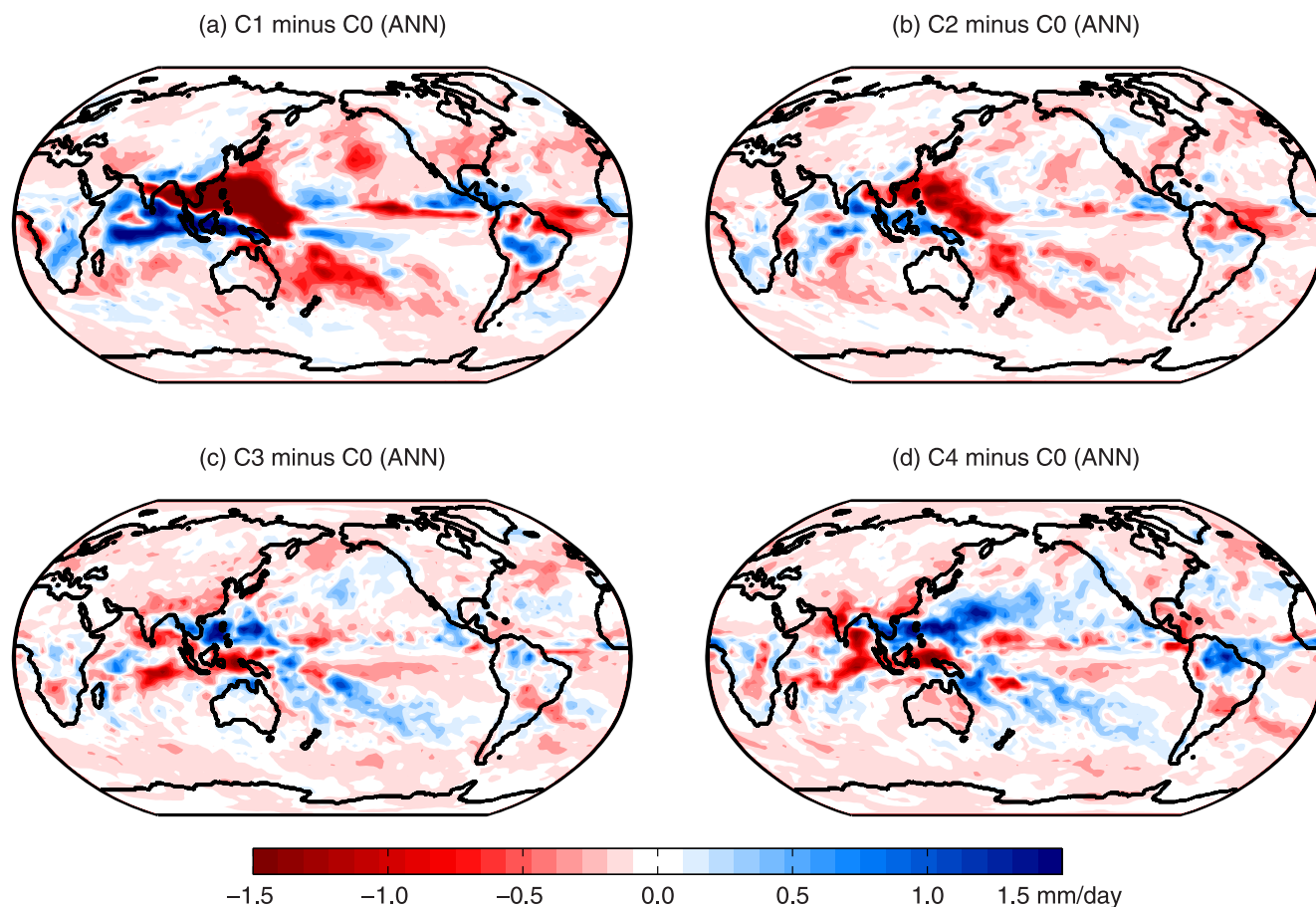
A description of the treatment of convective precipitation and its reevaporation and a parameterization of convective gustiness is provided in supporting information sections S3 and S4.

#### 4.1. Sensitivities to Convective Parameterization

We do not try to provide an exhaustive analysis of sensitivities to all aspects of this convection scheme. Instead, here and in supporting information S5, we illustrate several of the key sensitivities that played significant roles in the AM4 development process. For this purpose, we use 10 year climatological simulations similar to C0 except with perturbations of parameters in the convection scheme. Changes in the convection scheme generally result in significant change in the global TOA energy balance (see supporting information Table S2). We choose not to rebalance the TOA radiative fluxes here when we change a model parameter so as not to obscure the underlying sensitivity with other changes that would be required to create an acceptably balanced alternative model.

Two parameters that have large impact on AM4.0 simulations of the spatial distribution of precipitation are the strength of the lateral mixing rate in the deep plume and the strength of precipitation reevaporation. We discuss the former here and the latter in supporting information S5.3.

We modify the deep plume base lateral mixing rate  $\varepsilon_1$  from the AM4.0 default value  $0.9 \text{ km}^{-1}$  to 0.5, 0.7, 1.1, and  $1.3 \text{ km}^{-1}$  and refer to the four models as C1, C2, C3, and C4, respectively. Figure 2 shows the distribution of the difference in annual mean surface precipitation between each perturbation experiment and the control experiment. As discussed in Part 1, a particularly significant bias in AM4.0 is excessive precipitation in the "Philippines hotspot." As  $\varepsilon_1$  decreases, the most significant changes are, indeed, a reduction of precipitation in the Philippine hotspot and in the SPCZ with an increase of precipitation over the maritime continent and the equatorial Indian Ocean, as well as the Gulf of Mexico. This sensitivity is monotonic within

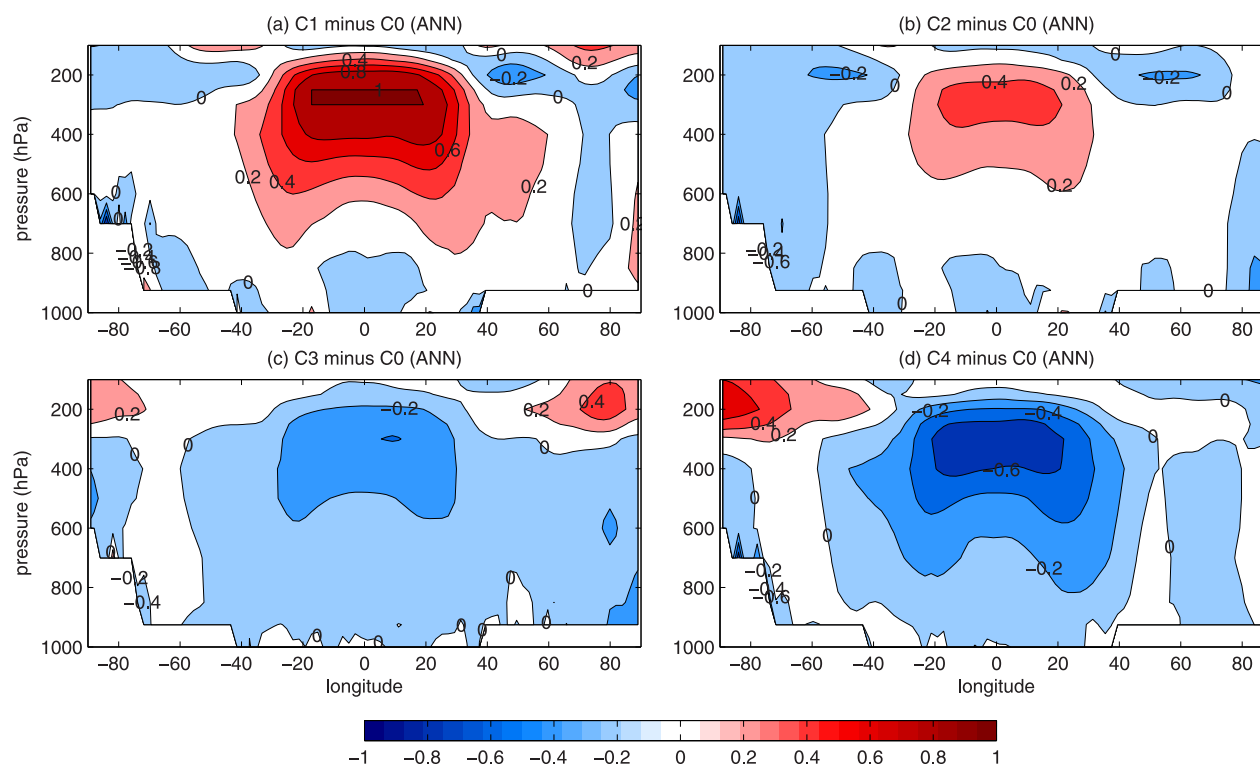


**Figure 2.** Geographical distribution of the difference in long-term annual mean surface precipitation (unit:  $\text{mm d}^{-1}$ ) for (a) C1 minus C0, (b) C2 minus C0, (c) C3 minus C0, and (d) C4 minus C0. C1–C4 denote experiments with increasing lateral mixing rate  $\varepsilon_1$  (i.e.,  $\varepsilon_1 = 0.5, 0.7, 1.1$ , and  $1.3 \text{ km}^{-1}$ ) while C0 is the control with  $\varepsilon_1 = 0.9 \text{ km}^{-1}$ . See text for a description of the experiments.

the range of the parameter explored, and overall a decrease in  $\varepsilon_1$  does decrease precipitation biases in these regions. However, the explanation for this precipitation sensitivity to  $\varepsilon_1$  is not straightforward due to strong interactions between the parameterized convection and the PBL, stratiform clouds and associated radiation, as well as resolved dynamics. For example, as  $\varepsilon_1$  decreases, parameterized deep plumes become more penetrative due to reduced lateral mixing, which in isolation would increase convective rainfall. While this is indeed true for the convective precipitation averaged over the entire tropics or the ITCZ, the spatial distribution of this increase is far from uniform because a change in parameterized convection modifies the vertical distribution of temperature and humidity, which feeds back to convection. We believe that a possible cause of the reduction of precipitation in the Philippine hotspot with decreasing  $\varepsilon_1$  is that more penetrative plumes over this region produce more drying in the PBL and lower troposphere and more warming in the upper troposphere which together increase the gross moist stability (e.g., Neelin & Held, 1987). Everything else being equal, this would reduce the climatological rainfall over this region. We hope to pursue analysis of this sensitivity elsewhere.

In Part 1, we show that AM4.0 has roughly  $1^\circ\text{C}$  cold temperature bias in the tropical upper troposphere (200–300 hPa). Figure 3 shows that this bias can be largely controlled by  $\varepsilon_1$ . This response is roughly linear in  $\varepsilon_1$ . The temperature response results directly from the changes in the deep plumes' lateral mixing, which determines both the convective updraft temperature and their frequency of occurrence, both of which affect the temperature. If the tropical upper tropospheric temperature cold bias was the only concern, we would choose a smaller  $\varepsilon_1$  for AM4.0 optimization. However, a smaller  $\varepsilon_1$  tends to degrade AM4.0 simulations of upper level winds. For example, the global root-mean-square error (RMSE) of 200 hPa zonal wind

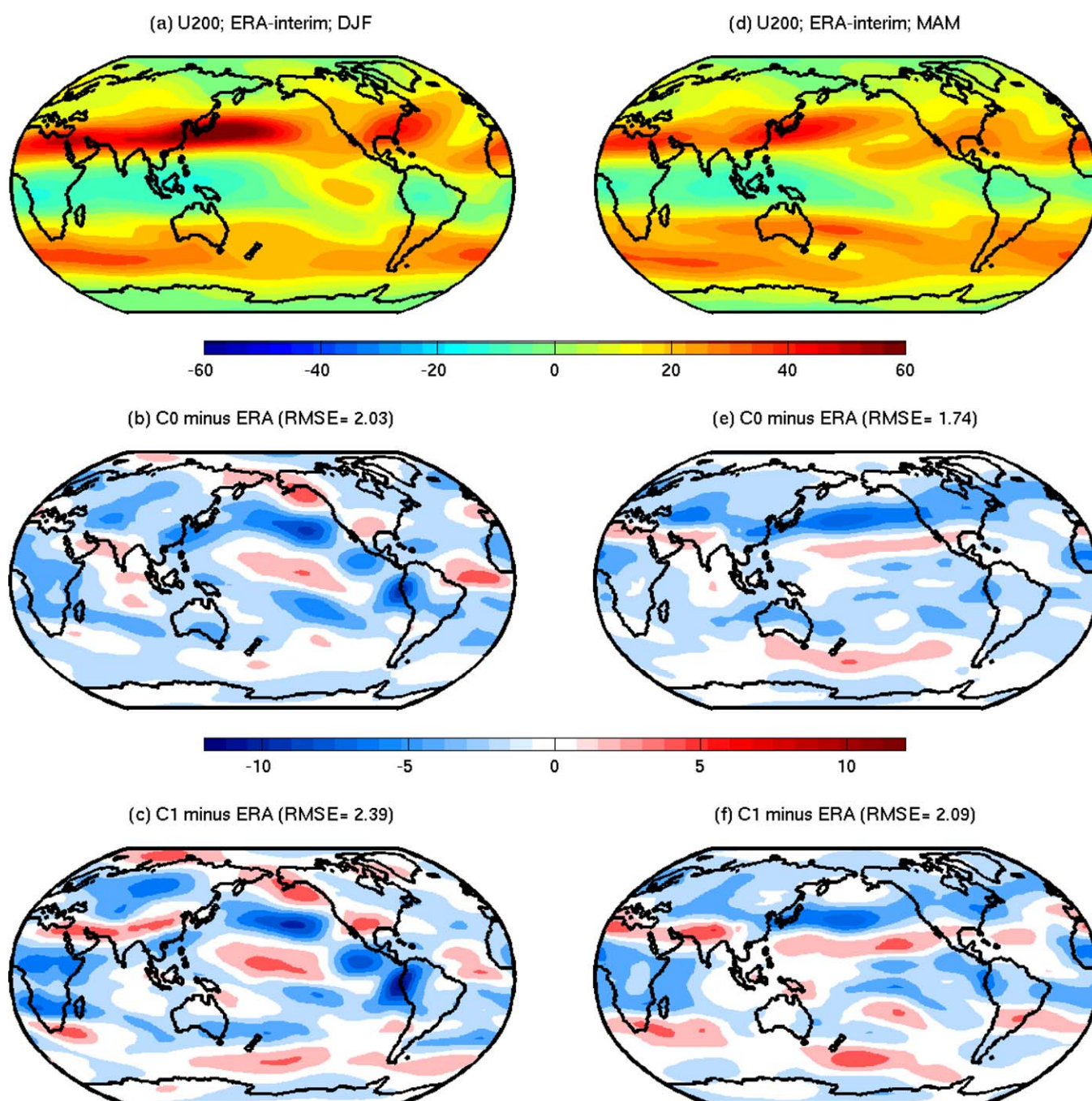




**Figure 3.** Annual and zonal-mean tropospheric temperature difference (unit: K) for (a) C1 minus C0, (b) C2 minus C0, (c) C3 minus C0, and (d) C4 minus C0. C1–C4 denote experiments with increasing lateral mixing rate  $\varepsilon_1$  (i.e.,  $\varepsilon_1 = 0.5, 0.7, 1.1$ , and  $1.3 \text{ km}^{-1}$ ) while C0 is the control with  $\varepsilon_1 = 0.9 \text{ km}^{-1}$ . See text for a description of the experiments.

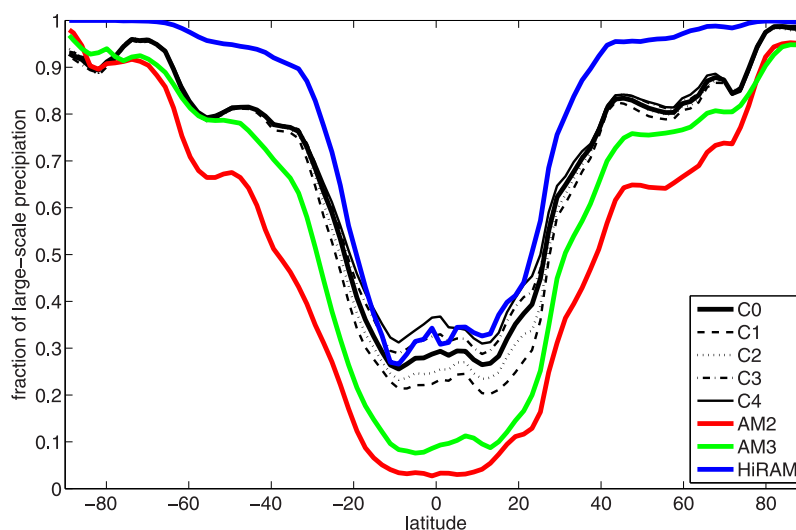
increases significantly from C0 to C1, especially in the boreal winter and spring ( $\sim 15\%$ , see Figure 4). This bias may result from the fact that warming the tropical upper troposphere in this model cools the extratropical lower stratosphere, increasing biases in horizontal temperature gradients. Some of the degradation of wind biases may be due to changes in the distribution of tropical convection and its induced change in extratropical circulation.

An even more significant problem associated with a smaller  $\varepsilon_1$  is that it causes a dramatic deterioration in AM4.0 simulations of tropical transient activity. In these perturbed  $\varepsilon_1$  simulations, the global and tropical mean precipitation hardly vary at all ( $< 1.7\%$ ) because these are controlled primarily by the atmospheric radiative cooling rate which does not change much in this fixed SST model. However, the partition of total precipitation between the parameterized convection and large-scale (stratiform) clouds in the tropics can vary substantially. As the parameterized convection is more inhibited with increasing  $\varepsilon_1$ , large-scale precipitation increases to balance the atmospheric radiative cooling. Figure 5 shows the latitudinal distribution of  $f_l$ , the fraction of precipitation that is large-scale, generated by C0, C1, C2, C3, and C4 along with that from AM2, AM3, and HiRAM simulations. There is a large difference in  $f_l$  among the models. In the tropics, AM4.0 produces roughly 30% large-scale precipitation (for comparison, ERA-Interim produces a similar fraction for the large-scale precipitation in the tropics). As the deep plume lateral mixing rate decreases from C0 to C1, there is roughly a 8% decrease in the large-scale precipitation in the tropics. The opposite is true when  $\varepsilon_1$  increases. This parameter has less impact in the extratropics where precipitation is already dominated by the large-scale component. In contrast to AM4.0, AM2 and AM3 produce much less large-scale precipitation in the tropics ( $\sim 5\%$  for AM2 and  $\sim 10\%$  for AM3) while HiRAM generates a similar amount of large-scale precipitation in the tropics and significantly more outside the tropics. These results are consistent with the fact that the lateral mixing rate of the deep convective plumes in AM2 and AM3 are significantly smaller than that in AM4.0 and HiRAM. In addition, the higher spatial resolution in AM4.0 and HiRAM also contribute to their increase of large-scale precipitation.



**Figure 4.** (a) DJF season 200 hPa zonal wind from ERA-Interim reanalysis averaged for 1980–2014 (unit:  $\text{m s}^{-1}$ ); (b) as in Figure 4a except for model bias from C0 ( $\epsilon_1 = 0.9 \text{ km}^{-1}$ ) simulation; (c) as in Figure 4b except for C1 ( $\epsilon_1 = 0.5 \text{ km}^{-1}$ ) simulation. The titles of Figures 4b and 4c show the global RMS errors. (d–f) As in Figures 4a–4c but for the MAM season.

While the GCM's explicit clouds, which generate the large-scale precipitation, are often referred to as stratiform clouds, they do not always manifest the characteristics of observed stratiform clouds in the tropics. Often, they are associated with deep ascent throughout the troposphere (Held et al., 2007). Many of these large-scale clouds and precipitation may be better considered as related to mesoscale convective systems and tropical storms, which the model struggles to resolve as well as it can, when the parameterized convection is significantly inhibited. The decaying phase of these large-scale systems do behave like stratiform clouds in that they produce significant radiative heating aloft and evaporative cooling in the lower



**Figure 5.** Latitudinal distribution of the fraction of total precipitation that results from the large-scale (or stratiform) cloud parameterization module generated in C0, C1, C2, C3, C4, AM2, AM3, and HiRAM. C1–C4 denote experiments with increasing lateral mixing rate  $\varepsilon_1$  (i.e.,  $\varepsilon_1 = 0.5, 0.7, 1.1$ , and  $1.3 \text{ km}^{-1}$ ) while C0 is the control ( $\varepsilon_1 = 0.9 \text{ km}^{-1}$ ).

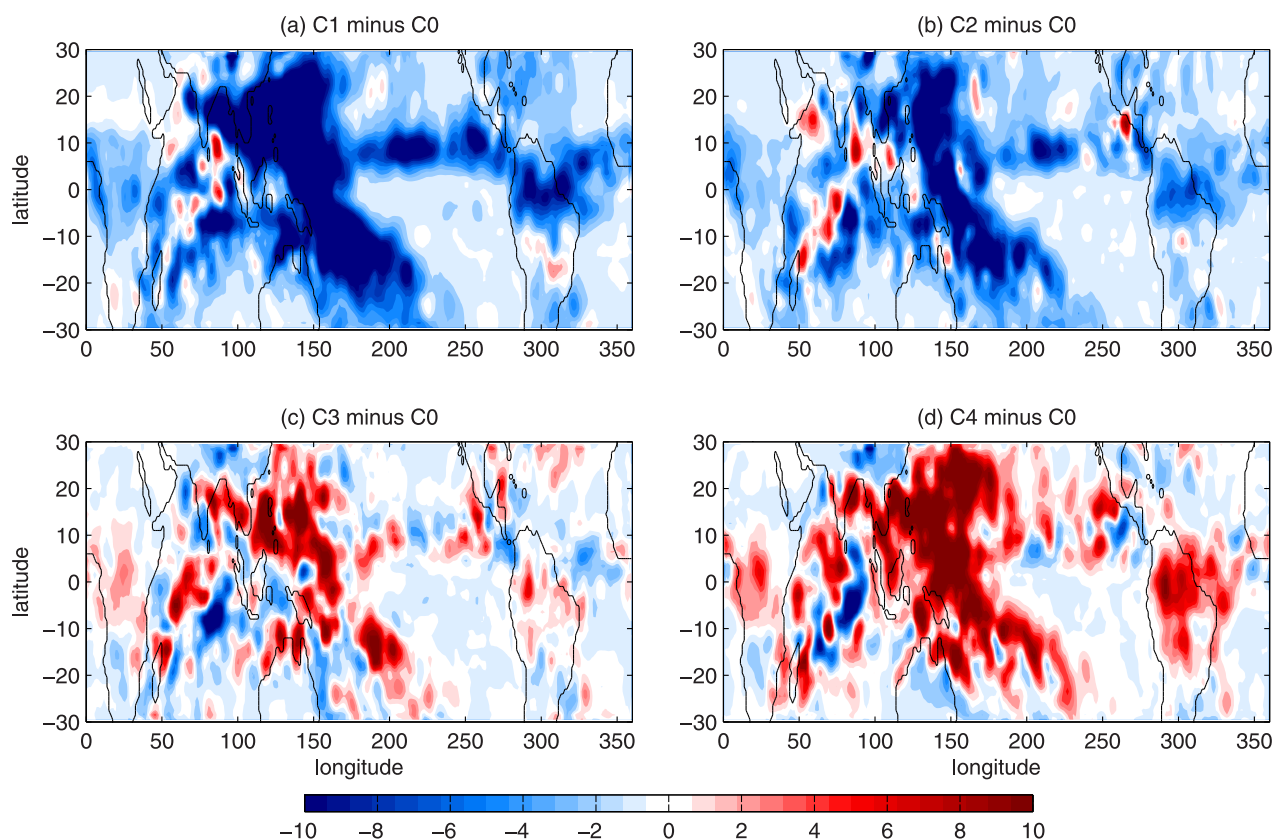
troposphere. The amount of these large-scale cloud systems, which are controlled here by the inhibition of the parameterized convection, can have large impacts to both the model's mean climate (e.g., Figures 2 and 3) and tropical transient activities.

Compared to C0, the reduction in large-scale precipitation in C1 and C2 is associated with a large decrease in tropical transient activity. Figure 6 shows the difference in 20–100 day band-pass filtered intraseasonal variance of precipitation between each perturbation experiment and C0. As the parameterized convection becomes incrementally less restricted through reduced lateral mixing, there is a large reduction ( $\sim 50\%$  from C0 to C1) in intraseasonal variance of precipitation. A comparison of these experiments using the familiar Wheeler-Kiladis wave number-frequency power spectrum plot is provided in supporting information Figure S1, which is consistent with the variance plot. In particular, as  $\varepsilon_1$  decreases (increases), there are decreased (increased) activities in the equatorial Kelvin waves, the MJO, as well as the westward moving disturbances associated with the tropical storms/depressions.

These results are broadly consistent with many previous GCM studies which suggest that an overly active parameterized convection tends to kill tropical transients in many models, resulting in too quiescent tropics (e.g., Crueger et al., 2013; Lin et al., 2006). On the other hand, when the parameterized deep convection is overly inhibited (or turned off in the extreme limit), the model-simulated mean state can deteriorate dramatically. In particular, it has often been reported that models with better MJO simulations tend to produce poor mean climate (e.g., Kim et al., 2012). Consistent with this literature, in optimizing AM4.0 convection, we have also compromised between the mean state (precipitation, TOA radiation, temperature, and circulation) and tropical transients (MJO and tropical storms). The compromise chosen has many realistic features, as portrayed in Part 1, but with, e.g., a weak tropical upper tropospheric temperature cold bias and a “Philippines hotspot” (unrealistic precipitation maximum).

Another parameter in the convection scheme which also strongly affects AM4.0 simulations of spatial distribution of precipitation is the strength of precipitation reevaporation. To demonstrate this, we modified a parameter that controls the fraction of precipitation falling outside the convective updrafts and conducted four sensitivity experiments which are referred to as C5, C6, C7, and C8, respectively. We find that as the convective precipitation reevaporation decreases, the precipitation in the Philippine Sea decreases substantially with an increase in precipitation over the eastern Pacific ITCZ. The shift of precipitation from the west to east in the Pacific helps to produce more eastern and less western Pacific TCs. However, it weakens the model's simulation of MJO. More importantly, coupled experiments suggested that this strengthening of the eastern Pacific ITCZ tends to produce a larger equatorial cold SST bias. We describe the experiments and discuss the results in more detail in supporting information S5.3.



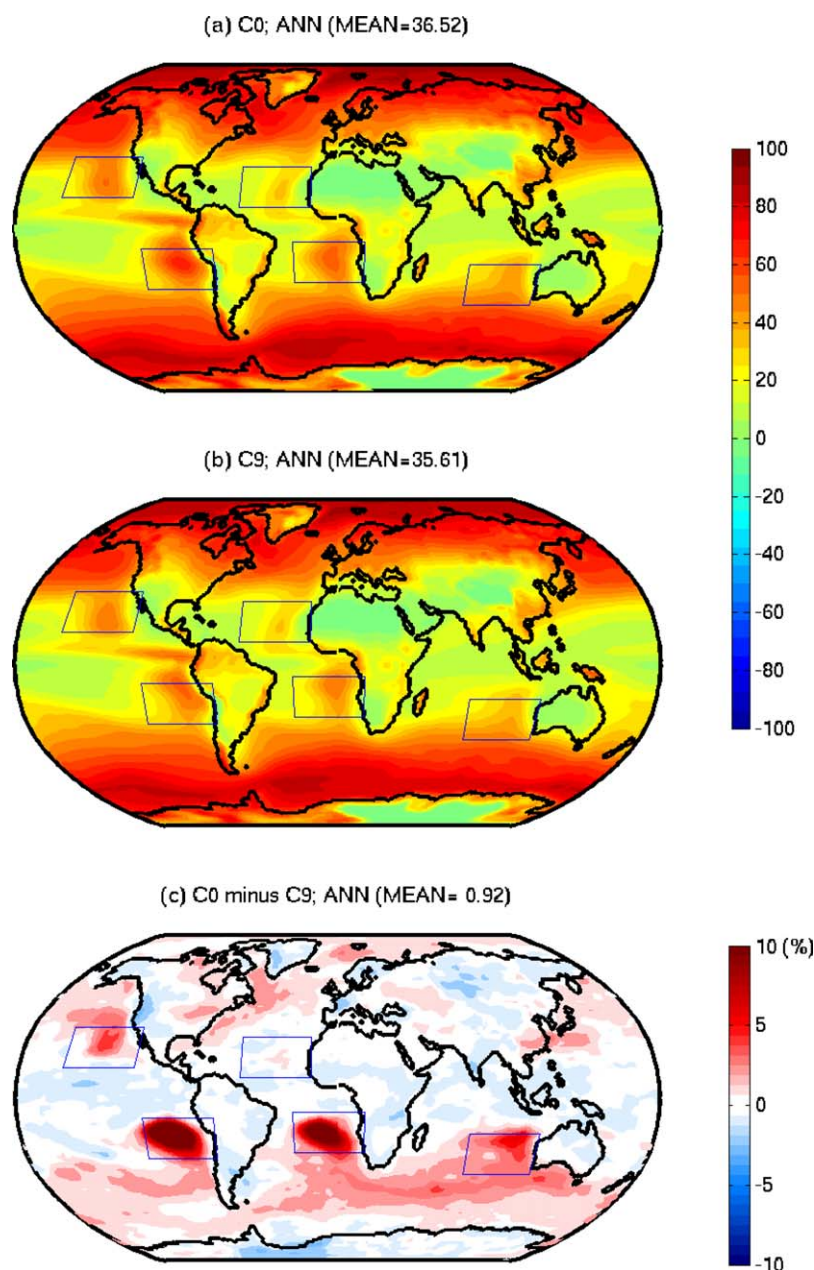


**Figure 6.** Difference in 20–100 day band-pass filtered variance of precipitation (unit:  $\text{mm}^2 \text{d}^{-2}$ ) for (a) C1 minus C0, (b) C2 minus C0, (c) C3 minus C0, and (d) C4 minus C0. C1–C4 denote experiments with increasing lateral mixing rate  $\varepsilon_1$  (i.e.,  $\varepsilon_1 = 0.5, 0.7, 1.1$ , and  $1.3 \text{ km}^{-1}$ ) while C0 is the control ( $\varepsilon_1 = 0.9 \text{ km}^{-1}$ ).

We also emphasize one modification to the shallow convective closure in AM4.0 and the model's sensitivity to this change. The starting point for AM4.0's parameterization of the shallow plume cloud base mass flux is the same as in the original University of Washington Shallow Cumulus scheme (Bretherton et al., 2004a), which is based on boundary layer TKE and convective inhibition (CIN; Mapes, 2000). However, we find the modeled shallow convection is occasionally unrealistically active (although weak) in regions with very large lower tropospheric static stability. This degrades the model's simulation of stratocumulus and stratus clouds. One plausible reason for this deficiency is that CIN is not well resolved due to the model's poor vertical and horizontal resolution, and, as a result, shallow convection is not properly shut down by the closure. It is also likely that our choice of shallow plume lateral mixing rate is too small in this regime. (The lateral mixing rate in shallow plumes is a constant value optimized for representing all shallow cumuli including regimes of both large-scale ascent and descent.)

To alleviate this deficiency in AM4.0, shallow convection is turned off where the lower tropospheric stability as measured by an estimated inversion strength (EIS) exceeds a critical value. The EIS was originally designed to provide an estimate of the strength of the PBL inversion given the temperatures at 700 hPa and at the surface. In particular, we use the EIS formulation derived in Wood and Bretherton (2006, see their equation (4)), which assumes a moist adiabat in the lower troposphere above the boundary layer. This EIS accounts for the general observation that the free-tropospheric temperature profile is often close to a moist adiabat and its lapse rate is strongly temperature dependent. Thus, it attempts to remove the impact of surface warming on lower tropospheric stability, which may be particularly important in a climate change scenario. We choose the critical EIS value to be 8K. The arbitrariness of this cutoff is of concern, and may become relevant if this model is used to simulate climates very different from Earth's. One can conceive of two atmospheric columns with the same EIS: one with smaller inversion strength but larger stability immediately above the inversion and another with a larger inversion and less stability immediately above the inversion. When the EIS is sufficiently large, AM4.0 cuts off shallow convection even if the model's inversion

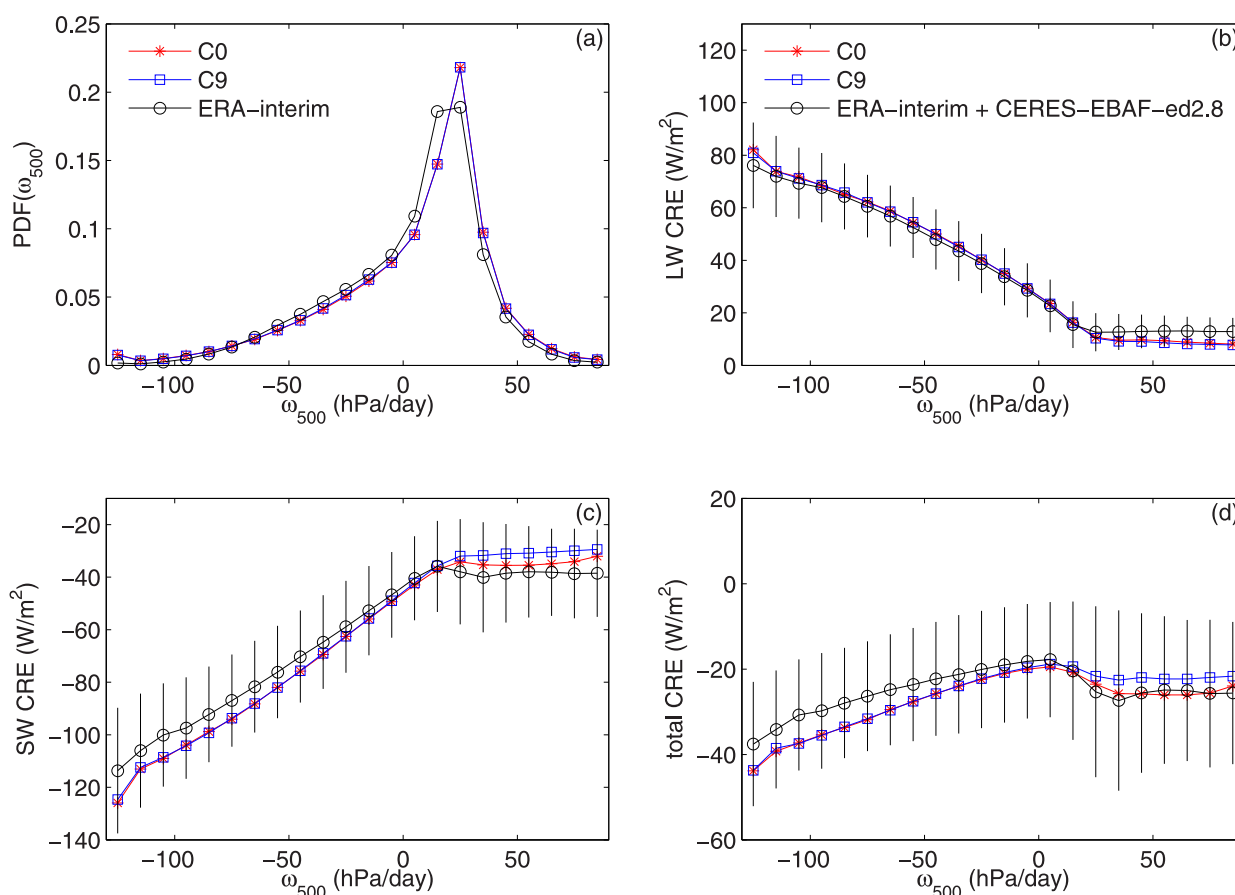




**Figure 7.** Long term climatological low cloud amount (below 680 hPa; unit: %) simulated by (a) C0 (control), (b) C9 (as C0 except without the EIS constraint), and (c) their difference: C0 minus C9. Boxes indicate observed regions of persistent stratocumulus decks.

is weak, thereby transferring control of the convection from the model's poorly represented inversion to a measure of the bulk stability of the lower troposphere. The clouds are not directly tied to the EIS; rather, the EIS affects the convection which, in turn affects the cloud simulation.

We show a comparison of AM4.0 with an alternative model that is identical to AM4.0 except with the EIS constraint on shallow convection removed. We have conducted a similar climatological run with this model and refer to it as C9. Figure 7 shows the model-simulated low cloud amount from C0, C9, and their difference. The EIS constraint tends to produce a modest increase in low cloud amount over the southern high-latitude ocean with especially large increase concentrated off the west coasts of the south America, the south Africa, and to a less extent the Australia and the northern America, where stratocumulus clouds



**Figure 8.** (a) Probability distribution of monthly mean 500 hPa vertical pressure velocity  $\omega_{500}$  (unit: hPa d<sup>-1</sup>) over the tropical (30°S–30°N) ocean simulated in C0, C9 and ERA-Interim. (b) The averaged LW CRE sorted in the large-scale dynamical regimes characterized by  $\omega_{500}$ . Black line shows the observational estimates based on the CERES-EBAF-ed2.8 LW CRE and the ERA-Interim  $\omega_{500}$ . The bars show one standard deviation of the LW CRE for each regime. (c) As in Figure 8b except for the SW CRE. (d) As in Figure 8b except for the total CRE.

typically occur. Globally, there is less than 1% increase of the low cloud amount. While it is not hard to tune the global low cloud amount, it has been proved challenging to improve the bias in underestimate of the low cloud amount over the stratocumulus region. One motivation for using this EIS constraint in AM4.0 is to improve its simulation of stratocumulus cloud, although significant underestimates of the low cloud cover and associated SW CRE remain near the coast. Other than the increase in stratocumulus and stratus clouds and associated SW CRE, we found little impacts of the EIS constraint on the model-simulated climate.

To further contrast the results from C0 and C9, and simultaneously to take a closer look at CRE biases in the tropics of AM4.0, we sort the simulated CREs over the tropical ocean into different large-scale dynamical regimes characterized by 500 hPa vertical pressure velocity,  $\omega_{500}$ , following Bony et al. (2004). Figure 8 shows that C0 and C9 produce very similar probability distribution of  $\omega_{500}$  as well as the LW CRE for all  $\omega_{500}$  regimes. Compared to the ERA-Interim reanalysis (Dee et al., 2011), both models tend to underestimate the frequency of weak ascent and descent and overestimate the frequency of strong descent. Both models slightly overestimate LW CRE over the ascent regimes and underestimate it in the modest and strong descent regimes ( $\omega_{500} > 30$  hPa d<sup>-1</sup>). The two models also produce a very similar overestimate of the SW CREs in all large-scale ascent regimes compared to the CERES-EBAF-ed2.8. The difference between the two models occurs nearly entirely in the SW CRE over the subsidence regimes, where C0 produces systematically more negative SW CRE and agrees better with the CERES-EBAF-ed2.8 data. This reduction in SW CRE bias further results in better agreement of total CRE over the subsidence regimes in C0 (Figure 8d). This is our primary justification for using the EIS constraint for shallow convection in AM4.0. We return to the effects of this EIS constraint on Cess sensitivity in section 13.

## 5. Subgrid Orographic Drag

AM4.0 uses the subgrid orographic drag scheme developed by Garner (2005), referred to here as G05. The scheme used in AM2.1 and AM3 is documented in Stern and Pierrehumbert (1988) and is referred to here as SP88. The new scheme differs in two important respects.

First, the base flux is built on an analytical expression for the tensor relationship between the near-surface wind and the total column drag. This handles anisotropy naturally and directly via the nonisotropic statistics of the topography. Specifically, the linear drag is given by

$$\vec{D} = \bar{\rho} \frac{\bar{N}}{N_0} \begin{pmatrix} \langle \partial_x \chi \partial_x h \rangle & \langle \partial_x \chi \partial_y h \rangle \\ \langle \partial_y \chi \partial_x h \rangle & \langle \partial_y \chi \partial_y h \rangle \end{pmatrix} \begin{pmatrix} \bar{u} \\ \bar{v} \end{pmatrix} \quad (2)$$

where  $(\bar{u}, \bar{v})$ ,  $\bar{\rho}$ , and  $\bar{N}$  are the resolved low-level velocity, density, and static stability (with  $N_0$  a reference value),  $\chi$  is the analytical disturbance velocity potential, and  $h(x, y)$  is the topographic height and the angle brackets denote an average over the grid cell. With  $\vec{x}$  denoting position on the sphere, the velocity potential is the integral

$$\chi(\vec{x}) = \frac{N_0}{2\pi} \iint \frac{h(\vec{x}')}{|\vec{x} - \vec{x}'|} dA \quad (3)$$

over spherical area (G05). The evaluation of this integral and the remaining work of generating the tensor elements in equation (2) can all be done off-line. This approach requires no approximate description of the characteristic orientation of the subgrid topography. Alternative approaches to anisotropic topography (e.g., Kim & Doyle, 2005; Scinocca & McFarlane, 2000) are more ad hoc.

Second, as also described in G05, the transition to nonlinearity (blocking) at the ground is accomplished by setting up an idealized distribution of mountains within the grid cell and comparing the corresponding Froude numbers, or nondimensional mountain heights, with a critical Froude number, rather than by using a single characteristic Froude number and an arbitrarily specified shape of the transition from laminar to blocked flow. The drag associated with an individual local maximum of  $h(x, y)$  is estimated according to dimensional analysis of an orographically adjusted flow (Pierrehumbert & Wyman, 1985). The degree of blocking is governed by an individual feature's Froude number and the critical Froude number. The resulting estimate is integrated analytically over the simplified distribution of heights and widths (see G05).

As in SP88, the vertical profile of forcing is derived from the wave saturation principle (Lindzen, 1981; Palmer et al., 1986), but with a modification described in G05 to incorporate the distribution of mountain heights, with the drag in the blocked flows being distributed over the lowest vertical half-wavelength of a stationary wave. Unlike the SP88 scheme, the blocked part of the drag is separate and additive in the G05 scheme.

Let  $D_p$  and  $D_b$  refer to the dimensional formulas for the propagating and blocked parts of the drag, normalized so that  $D_p$  matches the analytical drag in the limit of small mountains. We allow ourselves the freedom of introducing two parameters,  $a_p$  and  $a_b$ , and write

$$D = a_p D_p + a_b D_b \quad (4)$$

for the total drag  $D$ . While a departure of  $a_b$  from unity is not difficult to justify due to the approximate treatment of blocked flows, the drag coefficient  $a_p$  should be set to unity if we trust the linear theory for the propagating component. We mention two sources of uncertainty in the treatment of the latter.

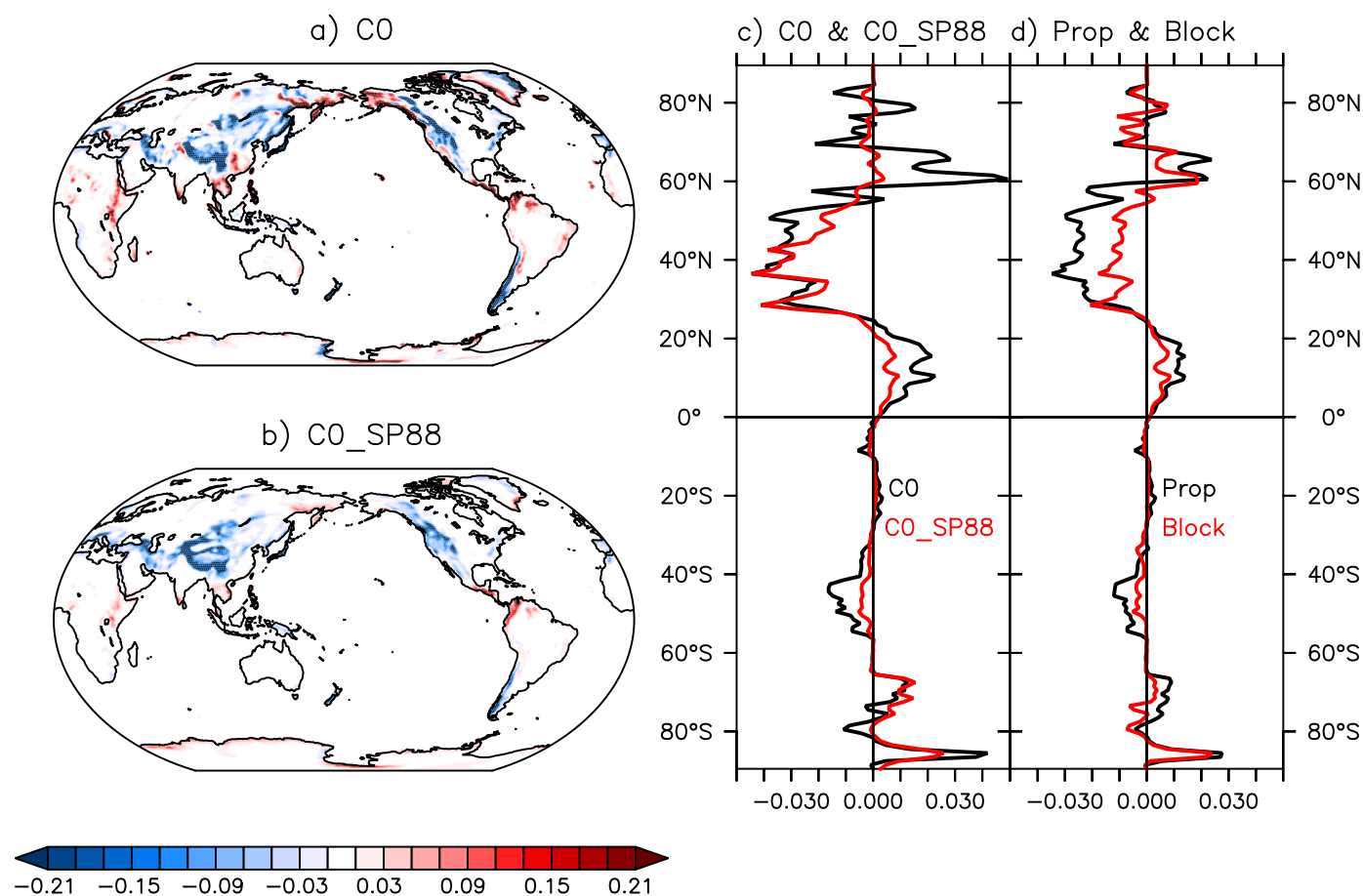
Drag schemes must ignore topographic scales that generate drag that the model resolves (to avoid double counting) and scales that are inconsistent with physical assumptions in the scheme. The common approach is to apply a high-pass filter to the raw topography before performing any other analysis. For AM4.0, we apply the spatial filtering simultaneously with the linear transformation (equation (3)) by multiplying the integrand by a kernel  $B(\vec{x}', \vec{x})$ . We use a Blackman kernel on the sphere that has nonzero weights only within a radius  $R$ . We set  $R$  to 100 km (the approximate model grid size) equatorward of  $20^\circ$  and  $100 \times \frac{\sin 20^\circ}{|\sin \theta|}$  km, i.e., inversely proportional to the Coriolis parameter, at higher latitudes  $\theta$ . In this way, the subgrid Rossby number based on a fixed velocity scale remains large and uniform across all latitudes (the theory is

invalid if the response to topography is affected by rotation). There is some arbitrariness in how to define this window function  $B$ .

The scheme also requires a specification of the level at which the resolved flow, density and static stability are defined in the computation for the base flux. The choice of these resolved variables is a source of some uncertainty. We generally use the first two layers above the planetary boundary layer (PBL) as defined by the model's PBL scheme. However, when the PBL depth collapses to zero, we use the 10 m wind provided by the boundary layer scheme, rather than the wind from the lowest model level. Because the 10 m wind is smaller, this reduces the drag in very stable regions, leading to certain improvements in the climatology, including a reduction of the cold bias in the Arctic 2 m temperature. Within the PBL, there is no parameterized orographic drag (SP88 excludes forcing from a certain mass fraction of the lower atmosphere that is usually deeper than the boundary layer).

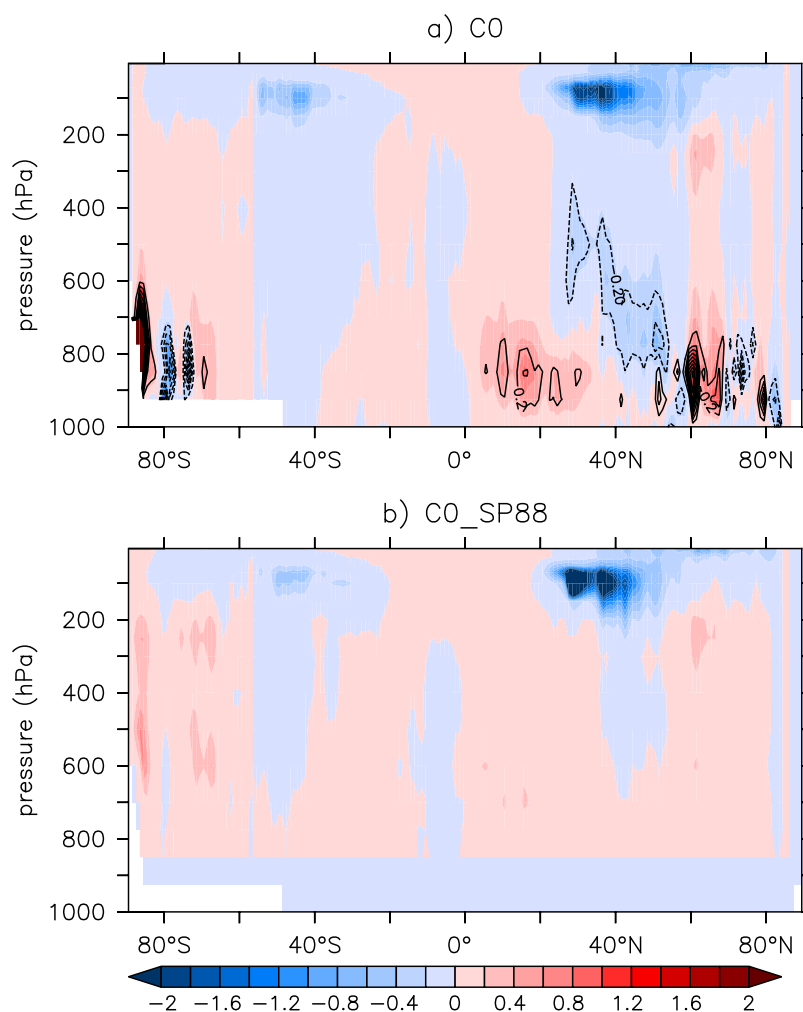
Given the uncertainties, we believe some level of tuning is justified. In AM4.0, we use  $a_p = 0.9$  and  $a_b = 3.0$ . The large value of  $a_b$  is justified by high resolution simulations of flow over the Rockies in which much of the drag parametrized here is directly simulated. We hope to describe these simulations elsewhere. Sensitivity of the AM4.0 simulation to  $a_p$  and  $a_b$  are described in supporting information S6.

We compare the results from AM4.0 incorporating G05 orographic drag with results from the same model but incorporating the scheme of SP88, as used in AM2.1 and AM3. The control simulation is C0 as described above. All results are averages over DJF (northern winter), the season in which the effects of subgrid topographic drag are largest, and using 30 years of the simulation in each model to help reduce the large natural variability in this season.



**Figure 9.** (a) Geographical distribution of the DJF seasonal mean zonal base flux (unit:  $\text{N m}^{-2}$ ) from the AM4.0 climatological experiment (C0). (b) As in Figure 9a except using the previous orographic drag parameterization (i.e., C0\_SP88). (c) The zonally averaged zonal base fluxes from C0 (black) and C0\_SP88 (red). (d) The decomposition of the total zonal base flux into the propagating (black) and blocking (red) components for C0.



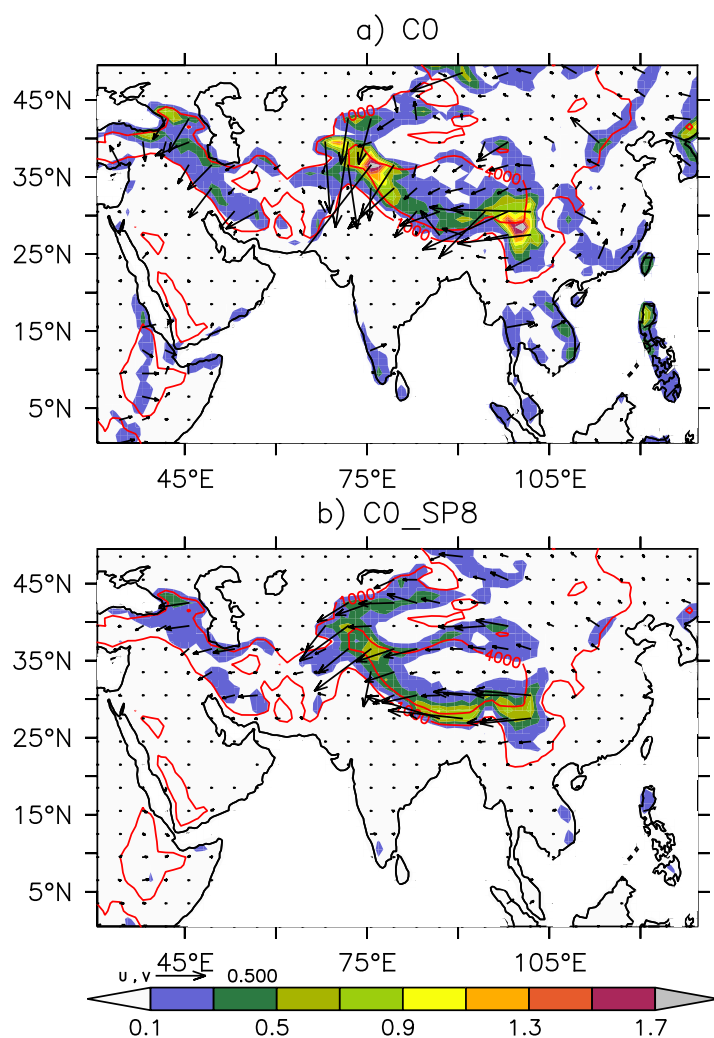


**Figure 10.** (a) Vertical cross section of the zonal-mean zonal wind tendencies due to the orographic drag parameterization from the AM4.0 climatological simulation (C0). Shading: total tendencies due to orographic drag parameterization. Contour: the component of the total tendencies due to blocking. Unit:  $\text{m s}^{-1} \text{d}^{-1}$ . (b) As in Figure 10a except for the results from C0\_SP88.

Figure 9 shows the geographical distribution of total zonal component of the base flux for the two schemes along with the zonal averages and the partition between blocked and propagating drag for the G05 scheme. The geographical distributions (Figures 9a and 9b) are generally similar, with the most noticeable differences appearing in Siberia, Alaska, and Greenland. As seen in the zonal averages (Figure 9c), G05 can generate twice as much drag as SP88 in the tropics and much of the middle latitudes. The increase is locally even greater at some higher northern latitudes. In the G05 scheme, the blocking part of the drag is generally about half as strong as the propagating drag (Figure 9d).

Figure 10 shows the zonal acceleration due to mountain drag in the two schemes. The blocking part of the drag is contoured for the G05 scheme (top figure). The base flux seen in Figure 9 is the vertical integral of these accelerations weighted by the density. Most of the acceleration below 500 hPa comes from the blocked component, which is maximized just above the PBL. By contrast, the SP88 scheme (bottom figure) deposits all of the drag according to the saturation hypothesis without regard to the degree of nonlinearity at the ground. The strongest accelerations occur in the stratosphere and are weaker for G05 in the northern hemisphere but stronger in the southern hemisphere compared to SP88.

To illustrate the implications of basing the drag on a linear analysis that automatically accounts for the anisotropy of the terrain, we show the total DJF base-flux vector in Figure 11 for a region centered on the Himalayas. The SP88 scheme (bottom figure) simply orients the base flux opposite the low-level wind. There



**Figure 11.** (a) Base-flux vector for the DJF season over the Himalayas from the AM4.0 climatological simulation (C0). (b) As in Figure 11a except for C0\_SP88. Shading shows the magnitude of the base flux (unit:  $\text{N m}^{-2}$ ). The 1,000 and 4,000 m topographic contours are shown in red.

any fundamental significance to these values. As evidence of this, we find that the values need to be modified to retain reasonable stratospheric winds when the treatment of the sponge layer at the model top is altered, a sensitivity that is not present in models with more stratospheric resolution.

## 7. “Light” Chemistry and Aerosols

AM4.0 simulates the mass distribution of five aerosol types including sulfate, dust, black carbon, organic carbon, and sea salt. Their size distribution is prescribed as lognormal for all but dust and sea salt, which are discretized into five size bins from 0.1 to 10  $\mu\text{m}$  radius. Their concentrations are calculated based on their emission (and precursor emissions), chemical production for sulfate and secondary organics, dry and wet (rainout and washout) deposition, transport by advection, and dry and wet convection. Modeling of these processes has been described in detail for AM3 by Donner et al. (2011) and Naik et al. (2013). A major difference in the treatment of aerosols with AM4.0 is in the simplified chemical production of sulfate. This simplification reduces the number of tracers by a factor of 3, which allows maintaining sustainable throughput ( $\sim 10$  simulation years per day at 1,152 cores) at high (50 km) spatial resolution.

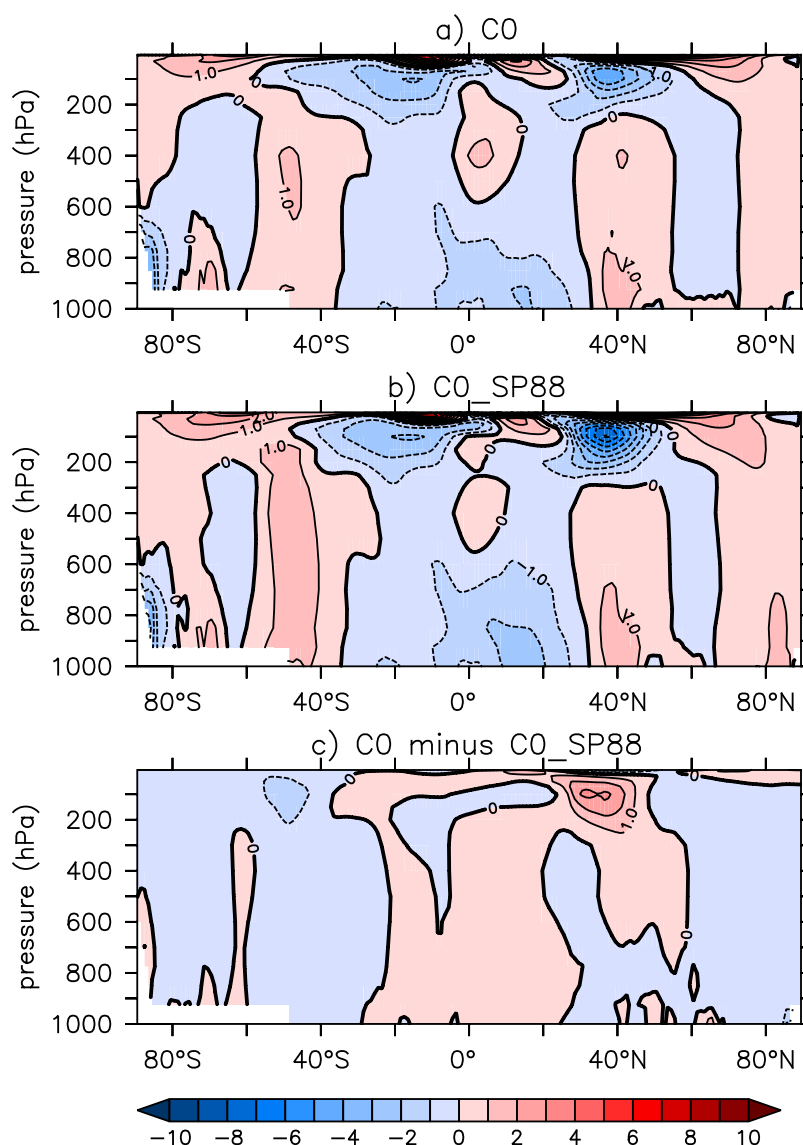
are differences in wind direction between the simulations, but most of the differences in drag direction are due to the incorporation of mountain anisotropy in the G05 scheme (top figure).

Figure 12 displays the biases relative to ERA-Interim reanalysis in the zonal-mean zonal wind when G05 and SP88 are used in otherwise identical models. G05 alleviates tropospheric westerly biases in the middle latitudes of the southern hemisphere and high latitudes of the northern hemisphere. It also mitigates the easterly bias at the top of the tropospheric jet (consistent with the reduced drag on the westerlies) and the westerly bias in the stratospheric jet. In response to increased eastward drag on the trade winds, there is also a modest improvement in the tropical lower tropospheric easterly bias.

There is also sensitivity of the Arctic surface air temperature to high-latitude orographic drag in AM4.0. The drag from Greenland, which dominates the Arctic forcing, is much larger in the G05 scheme than in the SP88 scheme. The latter produces warmer Arctic temperatures. We find that completely removing the drag over Greenland in simulations using G05 leads to an increase in the Arctic surface air temperature. The cold bias when Greenland is included may be related to the northward shift of the polar jet and a consequent reduction in northward energy transport but is not fully understood.

## 6. Nonorographic Gravity Wave Drag

Nonorographic gravity wave drag, of importance in the stratosphere, is unchanged in form from that in AM3 (AM2.1 does not have a convective gravity wave drag parameterization due to its lower model top). Assumptions about the wave spectrum and the deposition of the drag follow Alexander and Dunkerton (1999), but no attempt is made to relate the wave source to parameterized convection or aspects of the resolved flow. Instead, as in AM3 we simply specify the magnitude of the wave flux above 350 hPa, with three different values ( $S_t$ ,  $S_n$ , and  $S_s$ ) for the tropics, northern extratropics, and southern extratropics. These values are simply tuned to improve the stratospheric winds. In AM4.0, they are given the relative magnitude of 0.005, 0.007, and 0.0047 ( $\text{m}^2 \text{s}^{-2}$ ), respectively. Given the relatively low resolution in the model stratosphere, we do not pretend that there is



**Figure 12.** Bias (compared to ERA-Interim) in zonal-mean zonal wind from (a) C0 and (b) C0\_SP88. (c) Their difference.

As in AM3, optical properties of aerosols are precalculated using Mie theory assuming sphericity. The extinction efficiency, single scattering albedo, and asymmetry parameter are tabulated with dependency on wavelength, aerosol type, aerosol size (for dust and sea salt), and relative humidity. Sulfate is assumed to be internally mixed with black carbon for the calculation of optical properties, which are tabulated using their relative mass concentration.

As in AM3, dust emission is calculated interactively following the parameterization of Ginoux et al. (2001) with a threshold of wind erosion and global scaling factor of  $3.5 \text{ m s}^{-1}$  and  $0.2 \mu\text{g s}^2 \text{ m}^{-5}$ , respectively. The latter parameter is tuned to fit surface concentration of dust on islands measured by the University of Miami (Arimoto et al., 1995; Prospero et al., 1989). For sea salt emission, the particles larger than  $1.4 \mu\text{m}$  are parameterized following Monahan et al. (1986) while smaller particles are parameterized according to Martensson et al. (2003), with a 1.33 global scaling factor for both parameterizations. The treatment of emission of carbonaceous aerosols and sulfate precursors, including dimethylsulfide (DMS) and  $\text{SO}_2$ , is unchanged from AM3 (Donner et al., 2011; Naik et al., 2013), but the emissions inventory has been updated following the CMIP6 protocol. An additional direct emission of sulfate aerosols is considered by assuming that 1% of  $\text{SO}_2$  emissions are released as sulfate. DMS is emitted from phytoplankton in seawater and is parameterized as a

function of windspeed following Chin et al. (2002). Chemical production of sulfate is from oxidation of sulfur dioxide ( $\text{SO}_2$ ) in the gas phase by OH, and in the aqueous phase by  $\text{O}_3$  and  $\text{H}_2\text{O}_2$ . The rate of aqueous-phase oxidation of sulfate by  $\text{O}_3$  and  $\text{H}_2\text{O}_2$  depends on the solution pH (Seinfeld & Pandis, 1998). Here we assume a constant pH = 4.5 in cloud droplets. In AM4.0, there is no explicit treatment of ozone chemistry. Instead, the monthly concentrations of  $\text{O}_3$  and other oxidants (OH,  $\text{HO}_2$ , and  $\text{NO}_3$ ) and the photolysis rate of  $\text{H}_2\text{O}_2$  used by the sulfate chemistry are calculated from a 20 year climatology (1981–2000) simulated by AM3 (Naik et al., 2013), which was done using the MOZART comprehensive mechanism of ozone chemistry (Horowitz et al., 2003). Ozone for radiation is specified using the CMIP6 data set.

Following Koch et al. (1999), we include  $\text{H}_2\text{O}_2$ , dimethylsulfide (DMS), methanesulfonic acid (MSA),  $\text{SO}_2$ , and sulfate as prognostic species. Formation of  $\text{H}_2\text{O}_2$  occurs by reaction of  $\text{HO}_2$  with itself in two-body and three-body reactions.  $\text{H}_2\text{O}_2$  is destroyed photochemically and by reaction with OH. DMS is oxidized by OH during the day and  $\text{NO}_3$  at night. Due to the short lifetime of the radicals and their source from daylight photodissociation, a diurnal variation is imposed to their concentration. Null concentration is imposed at night for OH and  $\text{HO}_2$ , while during daylight OH is assumed constant and  $\text{HO}_2$  concentration follows the cosine of the solar zenith angle. The reverse is imposed for  $\text{NO}_3$ , with a constant concentration at night and zero concentration during daylight. Gas-phase reaction rates are prescribed following DeMore et al. (1997) and Atkinson et al. (1989). Aqueous-phase reactions rates are taken from Jacobson (2005).

Three model improvements are primarily responsible for the improved aerosol simulations in AM4 described in Part 1: (1) reduced wet removal by frozen precipitation produced by the Bergeron process (Liu et al., 2011; Paulot et al., 2017), (2) increased wet removal by convective precipitation (Paulot et al., 2016, 2017), and (3) more accurate representation of  $\text{SO}_2$  oxidation in clouds (Paulot et al., 2016, 2017). In particular, (1) and (3) increase the long-range transport of sulfate from midlatitude sources to the Arctic in winter and spring, while (2) reduces transport in summer and in the tropics. As these sensitivities are fully described in the cited papers, we do not discuss them further here.

## 8. Clouds and Aerosol-Cloud Interaction

The treatment of large-scale cloud macrophysics and microphysics in AM4.0 is essentially the same as that in AM2 and AM3. The aerosol microphysics is similar to AM3. The source and sink terms for cloud condensates (liquid and ice) are parameterized largely following Rotstajn (1997) and Rotstajn et al. (2000). AM3 and AM4.0 also include a prognostic scheme of cloud droplet number concentrations ( $N_d$ ; Ming et al., 2007) with droplet activation dependent on updraft velocity and aerosol properties (Ming et al., 2006). The sub-grid distribution of updraft velocity used for droplet activation is linked to the eddy diffusivity coefficient, with a specified lower bound on the standard deviation ( $\sigma_{w,\min}$ ; Golaz et al., 2011), which is set at  $0.7 \text{ m s}^{-1}$  in both AM3 and AM4. The autoconversion from cloud liquid to rain is related to  $N_d$  and a specified threshold value for droplet radius ( $r_{\text{thresh}}$ ), which is set at  $8.5 \mu\text{m}$  in AM4, only slightly larger than  $8.2 \mu\text{m}$  in AM3. As shown in Golaz et al. (2011), the magnitude of the aerosol indirect effects is sensitive to the choice of  $r_{\text{thresh}}$  in AM3, an issue that we will return to in section 13.

Droplet activation is assumed to occur both in newly formed clouds and in preexisting clouds in AM4 (see Golaz et al., 2011, equation 7). In contrast, AM3 considers only the former (see Ming et al., 2007, equation 7). This switch from partial to total activation tends to increase droplet number concentrations considerably. Another difference is that the size distribution of sulfate aerosol for computing activation no longer varies with the mass concentration of sulfate or that of primary aerosols in AM4. Sulfate is always assumed to follow distribution 1 in Donner et al. (2011) ( $N_1 : N_2 = 17:3$ ,  $D_{p1} = 0.01 \mu\text{m}$ ,  $\sigma_1 = 1.6$ ,  $D_{p2} = 0.15 \mu\text{m}$ , and  $\sigma_2 = 2$ , with  $N_i$  denoting number concentrations,  $D_{pi}$  median diameters,  $\sigma_i$  geometric standard deviations, and the subscripts  $i$  different lognormal modes). This change has an effect of lowering droplet number concentrations. The upper bound for supersaturation used in the activation parameterization is also increased to avoid sporadic nonconvergence of the iteration involved in the algorithm.

Similar to AM3, AM4 contains a prognostic variable for grid-box mean cloud liquid droplet number, which is transported by both resolved-scale flow and the parameterized convection and diffusion. Potential numerical issues related to advecting drop number and prognosticating droplet activation have been pointed out in Stevens et al. (1996). In AM4.0's convection scheme, the liquid cloud drop number in the convective updrafts is modified by dilution due to entrainment as well as new aerosol activation at the cloud base and



any layers where entrainment takes place (i.e., secondary activation). Given the updraft vertical velocity, temperature, and total aerosol mass at a pressure level, the activation of aerosols into droplets uses the parameterization of Ming et al. (2006). Above cloud base, only the entrained aerosol mass is used for secondary activation. The aerosol tracers are transported by the convective plumes as passive tracers following the plume's mixing, entrainment, and detrainment characteristics described in Bretherton et al. (2004a). A portion of the aerosols is removed by wet deposition due to convective precipitation. The removal scheme is as in AM3, although the efficiency of removal of tracers by convective precipitation is significantly greater than the unrealistically weak removal in AM3 (Fang et al., 2011; Paulot et al., 2016). The convective scavenging of tracers in AM3 was reduced by "limiters" implemented to enforce monotonicity and positive definiteness in the tracer mixing ratio profiles. The relaxation of the monotonicity constraint in AM4.0, as well as other changes in the convection scheme contribute to the stronger convective scavenging in AM4.0.

## 9. PBL and Surface Fluxes

In AM4.0, we retain the PBL scheme of Lock et al. (2000) that was utilized in both AM2.1 and AM3 (as described in GFDL-GAMDT, 2004). The only parameter in this scheme that we vary in the process of tuning is the strength of the turbulent mixing generated in a cloud-topped PBL by radiative cooling at the cloud top, as described in section 12.

In the development of AM4.0, we implemented the CLUBB (Cloud Layers Unified by Binormals) higher order turbulence closure (Golaz et al., 2002; Larson & Golaz, 2005) to unify the treatment of PBL, shallow convection and large-scale clouds (Guo et al., 2014) as well as deep convection (Guo et al., 2015). This approach showed interesting results, for example, better simulation of coastal stratocumulus, better capture of the transition of stratocumulus to cumulus, and better capture of liquid water path responses to aerosols (Guo et al., 2015). But it was computationally expensive. Compared to standard AM3, adding CLUBB increased overall expense by 70%. With the choice of a "light" chemistry package in AM4.0, the relative cost increase for CLUBB would have been around 140%, which was deemed to be prohibitive. We worked on formulating a simpler and more cost effective scheme retaining key concepts of CLUBB but were unable to configure this model successfully within the time frame of this development.

We have incorporated the neutral drag formulation over the oceans endorsed by the Coupled Ocean-Atmosphere Response Experiment version 3.5 (COARE3.5) consortium (Edson et al., 2013) into AM4.0. This change has little effect on our simulations in AMIP mode. Stability corrections in the Monin-Obukhov similarity theory are unchanged from those used in AM2 and AM3. The contribution of gusts in the wind speed used for the flux computation is estimated as before from a convective velocity scale based on boundary layer depth and surface buoyancy flux. For the stable boundary layer, the diffusivity is zero when the bulk Richardson number ( $Ri$ ) exceeds 1 as in AM3. AM2 uses the same formulation but with a much longer tail, the critical  $Ri$  is set to 10. A topographic roughness parameterization that ties the roughness length over land to the subgrid-scale topography is also identical to that in AM2.1/AM3.

## 10. The Land Model LM4.0

LM4.0, which uses the same grid as AM4.0, is based on the model LM3.1 described by Milly et al. (2014) and is summarized here. Land grid cells are tiled to represent different land types (soil, lake, or glacier) and, for soil tiles, differing histories of land use/land cover change. The reconstructions of land use/land cover change are from CMIP5 forcing (Hurtt et al., 2011). Runoff is transported to the ocean by a grid-scale river and lake network. Prognostic variables for soil, lake, and overlying (constant-density) snowpack, all vertically distributed, are liquid content, ice content, and temperature. The soil is treated as a dual-domain (Beven & Germann, 1982), saturated-unsaturated, soil-bedrock continuum, in which matrix flow is governed by Richard's equation (Hillel, 1980), and both infiltration and parameterized subgrid discharge to rivers may be facilitated by idealized macropores. Lakes are represented by one or more vertical columns, with density-driven stratification and overturning. Prognostic variables for each river reach are depth-averaged liquid mass, solid mass, and sensible heat content. Rivers flow according to the hydraulic geometry of Leopold and Maddock (1953). Prognostic variables for vegetation are mass of carbon in each of five plant compartments (leaves, fine roots, heartwood, sapwood, and labile), and liquid and solid water storage and temperature on leaves. Vegetation type ( $C_3$  or  $C_4$  grass, temperate deciduous tree, tropical tree, or cold evergreen

tree) is determined by plant biomass and climate (Shevliakova et al., 2009). Photosynthesis and stomatal conductance are determined mechanistically (Leuning, 1995).

Several changes were made from LM3.1 to LM4.0. Soil depth was decreased from 200 to 10 m (the LM3.0 value) in order to reduce spin-up times. Significant changes also were made to the soil-type characterization, to surface albedo, and to plant hydraulics and biogeography, as described below. LM3.1 had been developed with dynamic vegetation and without the land use/land cover change, while AM4.0/LM4.0 instead uses prescribed land use tile distribution and seasonal cycle of vegetation characteristics. The latter were derived from stand-alone LM4.0 simulations forced by Sheffield et al. (2006) surface-climate reanalysis.

The Harmonized World Soil Database Version 1.2 (Fischer et al., 2008) was used to define the global distribution of soil texture, in order to have an updated and higher-resolution soil representation, with a wider range of soil types.

The finite xylem resistance and increased root membrane permeability that were introduced in LM3.1, which had been chosen in stand-alone simulations with the Sheffield et al. (2006) forcing, were set back to the zero xylem resistance and lower root permeability of LM3.0 (Milly et al., 2014); liquid interception capacity per unit leaf area was changed from 0.1 mm back to the LM3.0 value of 0.02 mm. In an early prototype of AM4.0/LM4.0, the LM3.1 values of these parameters were judged to have inhibited transpiration sufficiently to suppress precipitation and realistic vegetation growth over the Amazon basin.

Values of several parameters affecting surface albedo were changed, largely to adjust for biases in near-surface air temperature. Leaf reflectances were set back to LM3.0 values, except that values in the near-infrared were decreased from LM3.0's 0.5 to 0.39 for cold evergreen trees and 0.45 for other trees. All leaf reflectances were higher than in LM3.1, having the direct effect of brightening and cooling snow-free vegetated surfaces. The isotropic weights in the RossThickLiSparseReciprocal bidirectional reflectance distribution function (Schaaf et al., 2002; similar in magnitude to snow albedo) for snowpack were made to vary linearly from their cold (LM3.1) values (0.92, 0.58 for visible and near-infrared, respectively) to lower warm values (0.77, 0.43) over a 10 K range below the freezing point. Additionally, the critical snowpack depth at which snow cover of bare ground reaches 50% was decreased from 5 to 1.25 cm. The direct effect these changes is to brighten and cool the surface when snow depth is low and to do the opposite when snow is deep and warm.

## 11. Computational Efficiency and Work-Flow

All models at GFDL are built upon a software infrastructure known as the Flexible Modeling System (FMS; Balaji, 2012) and its associated workflow system, the FMS Runtime Environment (FRE; Balaji & Langenhorst, 2012). FMS provides a highly efficient parallel programming environment using MPI and OpenMP constructs (as well as hardware-specific optimizations such as hyperthreading), and FRE enables remote execution, postprocessing, analysis, and monitoring of jobs. The computations were performed on the NOAA Cray supercomputer Gaea, in Oak Ridge, Tennessee, and all analysis and postprocessing on the GFDL analysis cluster Pan, in Princeton.

In measuring computational efficiency, one must take into account both the time to solution, measured in simulated years per day (SYPD), as well the best use of the allocation, measured in compute-hours per simulated year (CHSY). With perfect scaling, CHSY will not change as one increases parallelism. However, as scaling efficiency drops, CHSY will increase and it may not be the most economical use of one's resources to run at the maximum SYPD possible (see Balaji et al., 2017, for a full analysis of these metrics).

The AM4.0/LM4.0 configuration used in this paper achieves  $\sim 24$  SYPD at 864 processors, at a cost of 900 CHSY. While it can run on higher counts to achieve speeds as high as 77 SYPD on 6,912 PEs, this increases the cost to 2,200 CHSY.

As the current simulations are based on relatively short AMIP runs, we have chosen to run the  $\sim 24$  SYPD configuration for best use of the available compute resources, as this allows individual simulations to complete within 1–2 days. For the same reason, we have forgone further scaling optimizations such as those achieved by the “concurrent radiation” configuration of AM4 (Balaji et al., 2016). This again serves to increase SYPD but at additional cost in CHSY.

## 12. Tuning of TOA Radiative Flux and Cloud Radiative Effects

We calibrate AM4.0 TOA radiative fluxes (OLR and SW absorption) toward the observed values primarily through tuning the parameters related to cloud processes. These parameters affect distributions of high and low (including boundary layer) cloud amount ( $q_a$ ), cloud liquid ( $q_l$ ) and ice ( $q_i$ ) water content, cloud liquid droplet number ( $q_n$ ), as well as cloud overlap assumptions which together determine the net TOA LW and SW CRE. The cloud ice crystal number (or effective radius of ice particles), which is parameterized as a function of temperature in the model, is an additional important cloud optical property that affects CRE, but we did not explore it for tuning purposes in AM4.0.

We try to optimize the LW and SW CREs (both global means and spatial distribution) instead of cloud properties. This choice is a consequence of our feeling that the consistency with observed CREs is fundamental to simulation quality, especially when the model is coupled. The imprecision and lack of consistency between existing cloud data sets, especially when combined with potentially serious limitations in our cloud model formulation, have made it difficult to use water and ice condensate amounts, for example, in the model optimization process. So we choose in AM4.0 to focus on the initial step of optimizing CREs within the chosen model formulation, leaving process-level critiques of our simulations to guide future development strategies, accepting the likelihood of compensating errors that can arise with this strategy.

Large-scale or stratiform clouds are the dominant clouds in the model that interact with radiation; convective clouds are also seen by the model's radiation scheme but play a much less important role due to their small spatial extent. AM4.0 utilizes an explicit prognostic equation for each of the four cloud properties ( $q_a$ ,  $q_l$ ,  $q_i$ , and  $q_n$ ), with the cloud amount in particular following the scheme of Tiedtke (1993) as in AM2.1 and AM3. Therefore, it is useful to think of the cloud optimization through tuning of the source and sink terms, each of which may involve a variety of processes (e.g., subgrid-scale condensation/deposition, cloud and precipitation microphysics, and turbulent mixing) and have large uncertainty in parameter estimates. A key aspect of the explicit cloud model is its connections to the parameterized convection, which acts directly as a source term through convective detrainment but also indirectly as a sink through convective precipitation and associated drying. Because convective detrainment tends to dominate near cloud top while convective drying dominates near cloud base, the parameterized convection typically increases high cloud amount and reduces low cloud amount. But the strength of this interaction is sensitive to details of the parameterized convection (e.g., Vial et al., 2016), especially the representation of convective precipitation microphysics and cumulus mixing, which together determine the bulk convective precipitation efficiency (Zhao, 2014; Zhao et al., 2016).

In the tropics, high clouds in AM4.0 are primarily generated by convective detrainment and dissipated by the gravitational settling of ice particles in the representation of explicit clouds. In regions of significant large-scale precipitation (mostly the midlatitude storm track regions), high clouds are also formed directly by cooling due to large-scale ascent. To tune the amount of convective detrainment, we use parameters that control the precipitation efficiency in the convection scheme ( $e_l$ ,  $e_i$ , and  $q_{l0}$  in supporting information equation (S2)). An increase of the convective precipitation efficiency (especially  $e_i$ ) can be thought of as assuming that more condensate falls through the convective updrafts, leaving less water for detrainment and therefore reduces high clouds.

Independent from this convective source term, a parameter we used to tune the sink of the high (ice) clouds is the fall speed of ice particles  $V_f$ , which is parameterized as a function of in-cloud ice water content ( $V_f = c_1 3.29 (\rho q_i / q_a)^{0.16}$ ). This formulation with  $c_1 = 1$  was suggested by Heymsfield and Donner (1990) and was used in AM2.1. The tuning parameter  $c_1$  is set to  $c_1 = 1.5$  in AM3 and to 0.9 in AM4.0. With the parameters controlling both the source and sink terms, we can tune not only the global mean OLR toward the observed value but also its distribution between the tropics and extratropics, because convective detrainment impacts mostly the tropics while the ice fall speed affects the entire globe. While the representation of ice cloud microphysics in both the convection and the explicit cloud schemes is rudimentary in AM4.0, as in many current GCMs, with tuning they do appear to be capable of producing realistic simulations of high clouds and associated LW CRE and OLR in AM4.0 (see Figure 5 in Part 1).

Compared to high clouds, tuning of low clouds in AM4.0 is more complicated due to the variety of processes which directly affect the low cloud distribution or affect it indirectly through impact on humidity and temperature. Above the PBL, in regions of abundant shallow convection, AM4.0 low clouds are primarily

generated by convective detrainment (controlled by  $e_l$  in supporting information equation (S2)) and dissipated by cloud erosion (controlled by a cloud erosion time scale  $\tau_{eros}$ ), which roughly accounts for the effects of turbulent mixing of dry air into existing clouds (Tiedtke, 1993). In addition, because the convective plumes are allowed to penetrate beyond their neutral buoyancy level and mix dry air down to the convective layer this process tends to dry the cloud layer and diminish clouds. The strength of this cloud top penetrative mixing is parameterized to be proportional to the lateral mixing rate with a constant enhancement factor  $r_p$  (Bretherton et al., 2004a). An increase in  $r_p$  enhances the penetrative mixing and reduces low clouds. We find this cloud top penetrative mixing to be a powerful control on low cloud amount. In these shallow convection regions,  $e_l$ ,  $r_p$ , and  $\tau_{eros}$  are all used for tuning the amount and distribution of low clouds in AM4.0.

Within the PBL, moist convection and its associated downward flux of free-tropospheric air mass through either downdrafts or PBL entrainment typically reduces relative humidity and PBL cloudiness through ventilation of moisture (Vial et al., 2017). However, reevaporation of the rain associated with the convection can also moisten the PBL and reduce this drying. These effects depend again on the details of convective parameterization, including the treatment of cumulus lateral mixing and convective rain reevaporation. The parameters we used to tune the efficiency of convective rain reevaporation in AM4.0 are  $H_c$  and  $\alpha$  in supporting information equations (S3) and (S4). An increase in  $H_c$  and/or  $\alpha$  enhances convective rain reevaporation and associated moistening and therefore formation of low cloud. In contrast, cumulus lateral mixing affects low clouds primarily through impacts on convective depth and precipitation efficiency. A reduction of lateral mixing rate  $\varepsilon_s$ ,  $\varepsilon_d$  (see section 4) allows plumes to penetrate more deeply, producing more rain and drying in the lower troposphere. This drying propagates further into the PBL through PBL entrainment and leads to a strong reduction of low clouds both above and within the PBL.

Compared to other aspects of cloud parameters/processes, changes in cumulus lateral mixing rate and precipitation reevaporation also impact clouds indirectly through changes in circulation and precipitation, by their modification of the vertical distribution of heating profiles and through strong interactions with other processes (e.g., large-scale condensation, the PBL and surface fluxes). As a result, one must be cautious in using these parameters for tuning clouds. In particular, we do not use these parameters to fine-tune the global TOA fluxes.

In contrast, a parameter that is relatively benign in its effects on circulation and precipitation, and even the clouds themselves, is a parameter affecting the cloud overlap assumption used in taking into account partial cloudiness in the radiative transfer. We continue to use the Monte Carlo Independent Column Approximation (McICA; Pincus et al., 2003) as in AM3, which treats subgrid-scale cloud variability by creating subcolumns with consistent cloud properties and vertical structure (Pincus et al., 2005, 2006). As in AM3, we use the exponentially decaying overlap assumption which assumes a specified length scale controlling the decorrelation of clouds in the column. This coherence length has been used for tuning the TOA fluxes. In particular, we choose 2 km (a value broadly consistent with the CRM results in Pincus et al. (2005)) in AM4.0 (AM3 used 1 km while AM2.1 did not use the stochastically generated subcolumns for cloud radiation calculation).

AM4.0 continues to use Lock et al. (2000) for its boundary layer scheme, as in AM2 and AM3. In AM4.0, boundary layer clouds are also strongly affected by the PBL entrainment of free-tropospheric air, which controls the PBL depth and relative humidity in regions of little or no moist convection. In particular, a parameter used in tuning the AM4.0 boundary layer clouds is a scaling factor  $\beta_{rad}$  that controls the strength of PBL entrainment velocity due to radiative cooling (Lock et al., 2000). In AM4.0,  $\beta_{rad}$  is set to 0.23, the default value used in Lock et al. (2000).  $\beta_{rad}$  was previously tuned to 0.5 in AM2.1 and AM3. Everything else being equal, a reduction of  $\beta_{rad}$  reduces the PBL entrainment and increases PBL clouds especially over the eastern Pacific stratus and stratocumulus regions.

Among the numerous parameters which affect AM4.0 simulations of low clouds, one requiring special mention is a liquid cloud microphysics parameter in the explicit (stratiform or large-scale) cloud module, referred to as critical cloud drop radius ( $r_{thresh}$ ), above which cloud liquid is converted to rain. For a given cloud liquid water content and droplet number, an increase of  $r_{thresh}$  makes it harder for the liquid cloud to precipitate and therefore increases cloud liquid water content. This parameter has larger impact in the extratropics where large-scale stratiform clouds dominate. A key aspect of the AM4.0 sensitivity to this parameter is that



it not only affects the present-day liquid water path and associated SW CRE but also exerts strong control on the aerosol RFP similar to that seen in AM3 (Golaz et al., 2011, 2013). See discussions in section 13.

In addition to cloud amount/coverage, liquid and ice water content, the optical properties of clouds (e.g., droplet number concentration and effective radius) can also affect AM4.0 simulated CREs. In AM4.0, cloud liquid drops are generated by nucleation of aerosols and destroyed primarily by cloud dissipation processes associated with precipitation and turbulent mixing (erosion). The parameterized convection, which also activates aerosols and detrains liquid drops, serves as an additional source to drop number. But the convective source tends to be secondary compared with the aerosol activation in the explicit cloud scheme. The primary parameter we have used to tune the cloud droplet number is a specified limiting standard deviation of the subgrid-scale distribution of vertical velocity ( $\sigma_{w,\min}$ , see section 8 for details).

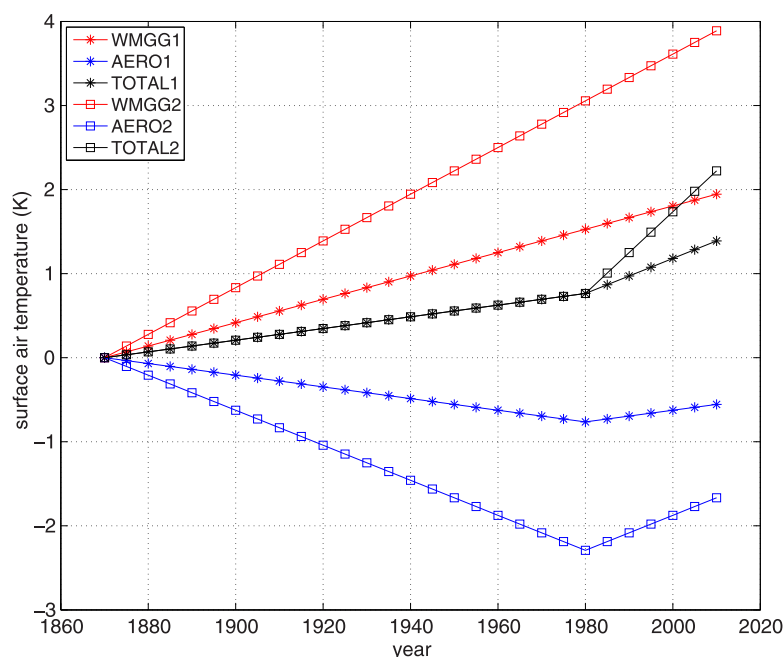
### 13. Cess Sensitivity and RFP

A goal of our model development is to create a model that simulates a climate evolution of the past century that is consistent with observations. This often involves model tuning either explicitly or implicitly. The climate model tuning process has been discussed in some detail from different perspectives in Mauritsen et al. (2012), Hourdin et al. (2017), and Schmidt et al. (2017). While it is conceptually useful to distinguish between model formulation and parameter estimation, the distinction is not always clear-cut in climate model tuning. In general, model parameters or parameterization may be tuned for two different considerations. One is to improve physical realism based on evaluation of specific processes (Randall & Wielicki, 1997). The other is to improve some overall emergent property of a simulation based on some hypothesis and experience. In some cases, these can be pursued without contradiction. But there are also aspects of our AM4.0 development that were admittedly driven primarily by the quality of the emerging climate simulation rather than process-level fidelity, effectively admitting that our process-level parameterizations are not fully adequate to this task. The characteristics of the tuning process are of special interest in the context of obtaining a plausible simulation of temperature trends over twentieth century. This issue has come to the forefront of modern model development with the advent of models of the indirect aerosol effect and the accompanying uncertainty in the strength of aerosol cooling, increasing the potential for creating unrealistic historical simulations without implicit or explicit tuning.

One often sees the argument that the twentieth century warming does not strongly constrain either climate sensitivity or the strength of aerosol cooling because similar overall warming can result from relatively low sensitivity to  $\text{CO}_2$  and weak aerosol cooling, or by high sensitivity and strong aerosol cooling. But this is true only to a limited extent, because of the likelihood of their having been a peak, or at least a plateau, in aerosol forcing in the 1980–1990s. As a result, in order to create the correct overall warming if climate sensitivity is high, one requires large enough aerosol forcing to cancel much of the warming prior to the 1980s, while after the aerosols peak the high sensitivity and reductions in aerosols combine to produce very rapid warming, as illustrated in the crude schematic in Figure 13. The result is a warming trajectory that can be difficult to reconcile with observations. See Golaz et al. (2013) for discussion of this type of behavior in the CM3 model. This is also a central aspect of one of the arguments used by Stevens (2015) to argue for a less negative lower bound in aerosol forcing.

The most relevant measure of climate sensitivity in this context is the transient climate response (TCR), the warming at the time of doubling when  $\text{CO}_2$  increases at a rate of 1%/yr. Most attempts at inferring the TCR from the historical record yield a value on the low end of typical GCM estimates. For example, Gillett et al. (2012) estimate 1.3–1.8K for TCR (as compared to the median of the CMIP5 models of 1.8K) when fitting the historical temperature evolution. The case is far from foolproof, given the uncertainty in the historical aerosol forcing, the presence of other forcing agents, and internal variability, but we have considered it desirable to avoid a large value of TCR ( $>2\text{K}$ ) to minimize the potential for the creation of an unrealistic warming evolution. To this end, and given that the CM2.1 coupled model has a TCR of 1.5–1.6K while the CM3 model has a TCR of 2.0–2.1K, we considered it desirable to create a model with Cess sensitivity closer to that of AM2.1 rather than AM3, assuming that the TCR of the resulting coupled model would follow the Cess sensitivity qualitatively.

Our first decision along these lines, early in the development process, was a choice of scheme for converting convective cloud condensate to precipitation (or equivalently the transfer of convective condensate to



**Figure 13.** A schematic diagram of two alternative models' simulation of twentieth century warming trend. Model 1 (stars) assumes low sensitivity to well-mixed greenhouse gases (WMGG) and low aerosol cooling effect (AERO); Model 2 (squares) assumes high sensitivity and higher aerosol cooling effect. Both models produce the same surface temperature trend before 1980, but diverge significantly after 1980 (black symbols: TOTAL = WMGG + AERO). Vertical coordinate is for illustrative purpose and is not meant to be quantitative.

the large scale) in the convection parameterization. From the work of Zhao et al. (2016), we knew that this choice does affect cloud feedback and Cess sensitivity in a prototype AM4 model. In particular, if we take the formulation similar to that used in AM2 rather than that used in the shallow plume of AM3, the resulting Cess sensitivity in the prototype model analyzed by Zhao et al. (2016) is reduced significantly. The AM2 version effectively produces condensate that is proportional to precipitation, while the scheme in AM3 shallow convection results in condensation per unit precipitation decreasing with increasing temperature, due to increased efficiency in generating precipitation, producing less cloud and more positive low and middle cloud feedback. Given the limitations of the plume model that underlies the convection scheme and the complexity of convective organization seen in nature, our preference for one or the other scheme on a process level was not strong, and we chose the AM2-like version for its lower Cess sensitivity. In large part because of this choice, AM4.0's Cess sensitivity ( $0.57\text{KW}^{-1}\text{m}^2$ ) is closer to that of AM2.1 than AM3, as described in Part 1.

The model's high-latitude cloud feedbacks may also be influenced by behavior of the cloud droplet number module. We have constructed a fixed drop number version of AM4.0 by setting the droplet number concentration to one value over land ( $150\text{cm}^{-3}$ ) and another over ocean ( $50\text{cm}^{-3}$ ), holding all other aspects of the model's parameterization unchanged. The resulting model has a lower Cess sensitivity than our standard AM4.0 ( $0.52\text{KW}^{-1}\text{m}^2$ ), in which, on average, the number of cloud drops decreases with warming. We do not fully understand at present the process within the model generating this result, nor are we confident that it is physical, but it does appear to play a modest role in raising our Cess sensitivity by  $\sim 10\%$ . While we were aware of this result in the development process, and tried model versions with different total droplet numbers to mitigate this effect (changing the assumptions concerning the subgrid distribution of vertical motion), we were not able to obtain values of Cess sensitivity close to this fixed droplet number result. (The sense of these small changes was to decrease TCR with increasing droplet numbers through an increase of high-latitude negative cloud feedback.) We did not pursue this further due to the smallness of the effects seen.

In spite of our attention to the Cess sensitivity, our first coupled simulations with prototypes of AM4.0 still generated large values of TCR, often greater than 2K and still substantially higher than CM2.1's value of 1.5–

1.6K. The assumption of proportionality between TCR and Cess sensitivity would have predicted a lower value, closer to 1.6–1.8K, taking into account the increased CO<sub>2</sub> forcing due to changes in the radiation module discussed in section 3. While analysis of this behavior is still underway, we believe that this difference is in large part due to the stronger polar amplification in coupled models incorporating AM4.0 than in CM2.1, which has unusually low polar amplification. The polar amplified warming combined with relatively weak negative low cloud feedback over the midlatitude and subpolar oceans evidently produces a larger ratio of TCR to Cess sensitivity in AM4.0-based coupled models.

Given these results, we looked for other options for reducing the TCR, but with modest results. One option that was adopted has already been described above, the EIS constraint on shallow convection. The model without this modification included has been referred to as C9 above in contrast to C0, the standard AM4.0. These two models have similar Cess sensitivities but significant difference in low cloud feedback over regions of large-scale subsidence (see supporting information Figures S5 and S6). In particular, both C0 and C9 produce a positive cloud feedback over the tropical large-scale subsidence region. However, the magnitude of this positive feedback is significantly smaller in C0 (supporting information Figures S6c and S6d). The importance of this change in coupled simulations has yet to be determined, but our incorporation of this change into AM4.0 can be thought of as partially motivated by an improvement in stratocumulus simulation and partly by the desire to nudge the model to lower sensitivity. It is worth pointing out the difficulty of documenting and evaluating the physical realism of all modeling decisions that effect how the system response to perturbations, since many are embedded deeply in legacy code inherited from previous development cycles. We do not claim to be providing a complete accounting of all of these decision trees here.

The difference in aerosol RFP between AM4.0 and AM3 has been presented in Part 1. In particular, the time variation of the aerosol RFP has been estimated using pairs of long AMIP simulations starting in 1870, running with time-varying SSTs, with and without anthropogenic aerosol emissions. Before 1980, the magnitude of AM4.0's aerosol RFP is uniformly smaller than AM3's by a factor of 0.72. Experimentation suggests that there is not one single cause for this reduction, but at least three significant contributors: (1) We see a decrease in RFP with increasing horizontal resolution in both models. Because AM4.0 has twice the horizontal resolution as AM3, we estimate that this difference results in roughly a 10% reduction in the magnitude of the aerosol RFP, consistent with the scaling estimate by Donner et al. (2016); (2) the changes in the aerosol model and convection scheme, specifically the change in aerosol wet deposition that result in an improved aerosol simulation in AM4, with reduced sulfate in heavily polluted regions and more sulfate and black carbon in northern high latitudes than AM3, result in another 10% reduction; and (3) the changes in aerosol activation and the resulting increase in droplet numbers in AM4, described in section 8 above, also result in roughly a 10% decrease in the magnitude of the aerosol RFP. The detailed mechanisms explaining these three contributions all require further study.

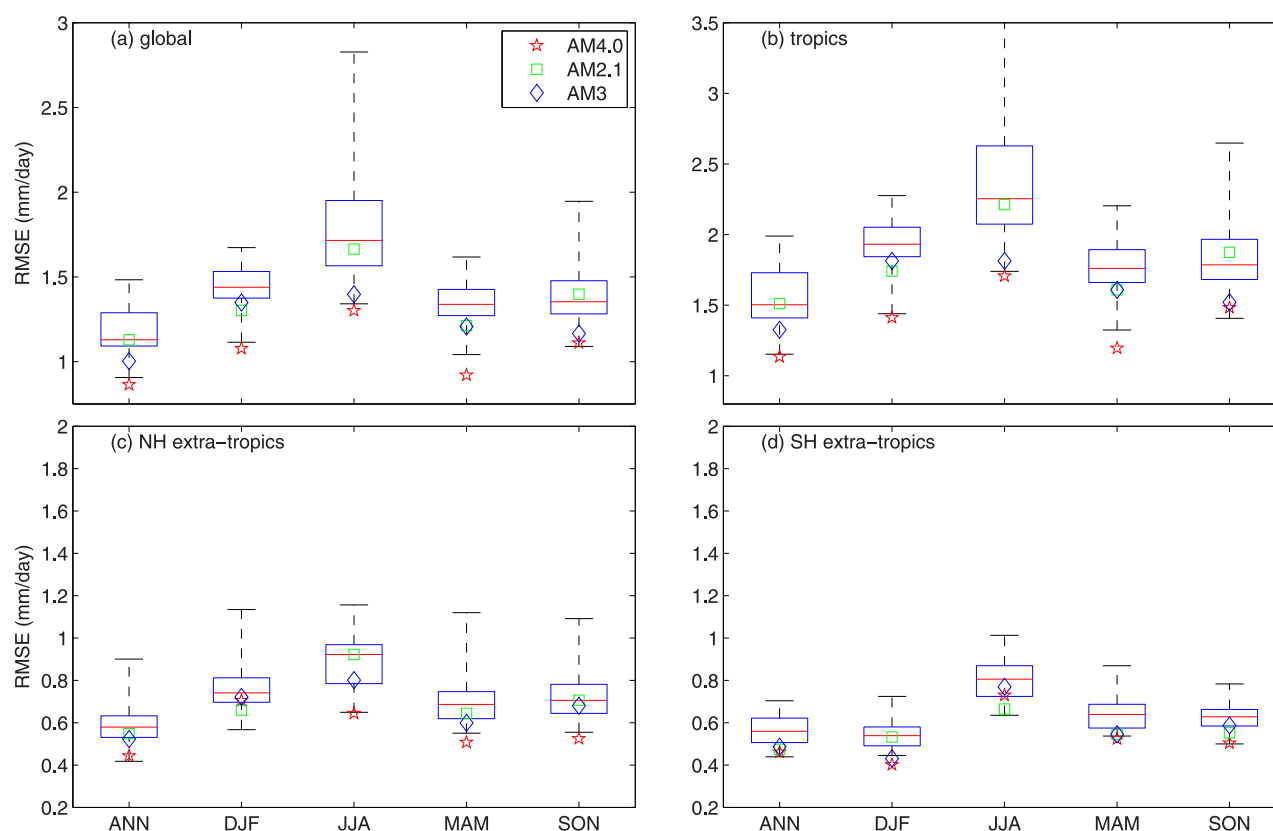
However, the long AMIP simulations also indicate that the aerosol RFP difference between AM4.0 and AM3 increases after 1980, with the ratio changing from 0.72 to 0.55. In particular, AM4.0 shows a substantial reduction in global aerosol RFP magnitude (and aerosol optical depth) after 1980 while AM3 shows a flatter evolution. This is not caused by the differences in emissions as running AM4.0 with CMIP5 emissions (which are used in AM3) yields a similar decrease. Further analysis shows that it is mostly due to the change in the spatial distribution of emissions and the models' difference in representing wet deposition. As sulfur emissions shift gradually from the NH midlatitudes (Atlantic sector) to the tropics Pacific sector after 1980, AM3 gives rise to an increase in sulfate burden due to its suppressed efficiency in convective wet removal. In contrast, AM4.0 produces no such increase as the wet removal efficiency is comparable between the midlatitudes and the tropics. Other modeling studies also suggest the effect of shifting emissions from the Atlantic to the Pacific sector and the strength of the effect appears to be related to differences in precipitation scavenging in the downwind regions of Asia versus North America (e.g., Fiedler et al., 2017). We leave a full investigation of the two models' difference in post-1980 aerosol trend for future work.

Golaz et al. (2013) document the sensitivity of aerosol RFP in AM3 to the parameter  $r_{thresh}$  controlling the value at which precipitation is assumed to begin in the large-scale cloud scheme, and AM4.0 has a similar sensitivity to this parameter. The model's second aerosol indirect effect (the response of cloud liquid water to aerosols) is responsible, an effect that most GCMs with aerosol-cloud interactions tend to overproduce compared to recent observational studies (e.g., Malavelle et al., 2017; Quaas et al., 2009). Recent LES studies of shallow cumulus clouds suggest a negative lifetime effect due to deepening of cloud layer in association

with a larger droplet concentration (e.g., Seifert et al., 2015; Stevens, 2017). However, this effect is not parameterized in our GCMs. We have not adjusted  $r_{thresh}$  in a significant way during AM4.0 development ( $r_{thresh} = 8, 8.2$ , and  $8.5$  in AM2.1, AM3, and AM4.0, respectively). While it has been argued that a larger value ( $>10 \mu\text{m}$ ) may be more consistent with satellite observations (Suzuki et al., 2015), it produces stronger aerosol cooling effect in AM4.0 which likely would negatively affect coupled model simulations of historical temperature trend in a way similar to CM3 (Golaz et al., 2013). Further, it also degrades model simulations of present-day clouds and TOA radiative fluxes by generating excessive amount of middle latitude clouds especially over the storm track regions. The parameterization of clouds and cloud microphysics in most current GCMs (including AM3 and AM4.0) is still very crude. In particular, the cloud scheme uses grid-box in-cloud mean drop size to determine precipitation, which does not account for subgrid variability of clouds and size distribution of hydrometeors. In reality, precipitation depends mostly on the size of the largest drops. This makes many observational estimates and comparisons with the parameters used in a cloud model far from straightforward. While further investigations of the model-observation discrepancies discussed in Suzuki et al. (2015) are clearly needed, future development of cloud parameterizations will focus not only on parameter estimates but more importantly on formulation refinement such as use of two-moment microphysics schemes.

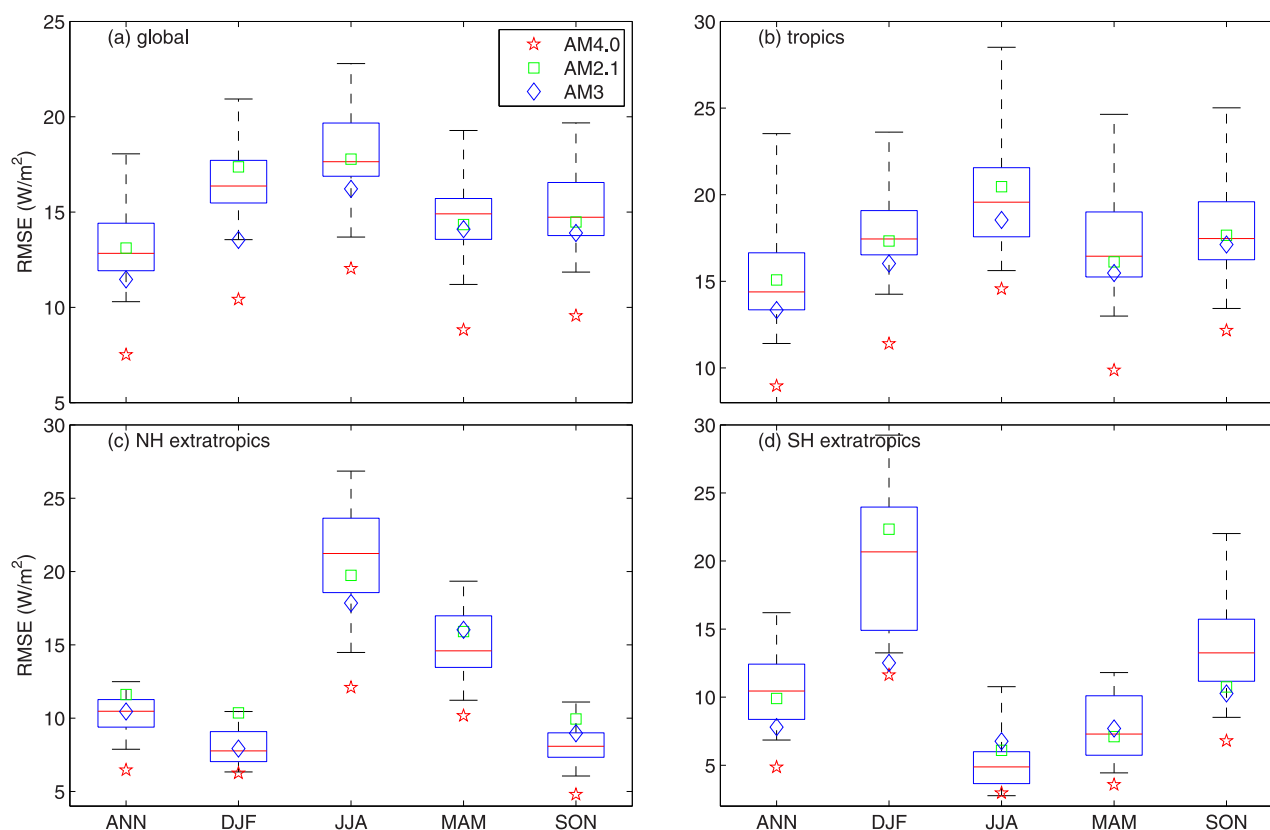
## 14. Summary and Discussions

The starting point for the development of AM4.0/LM4.0 model can be thought of as its dynamical core, its radiative transfer code, and its land model, none of which have changed in their fundamental formulation from previous models. The radiative transfer has been modified significantly to reduce its biases compared to line-by-line results using up-to-date spectroscopy, as described in section 3, resulting in a  $\sim 10\%$  increase



**Figure 14.** Comparison of the RMS errors of AM4.0 simulated long-term (1980–2014) climatological distribution of precipitation with the AMIP simulations from CMIP5 models using the box plots. Red lines are the median values while whiskers show the maximum and minimum values from the CMIP5 ensemble. The RMS errors are computed for both the annual mean and individual seasons and for (a) the entire globe, (b) the tropics, (c) the NH extratropics, and (d) the SH extratropics. The observational data are from GPCP.





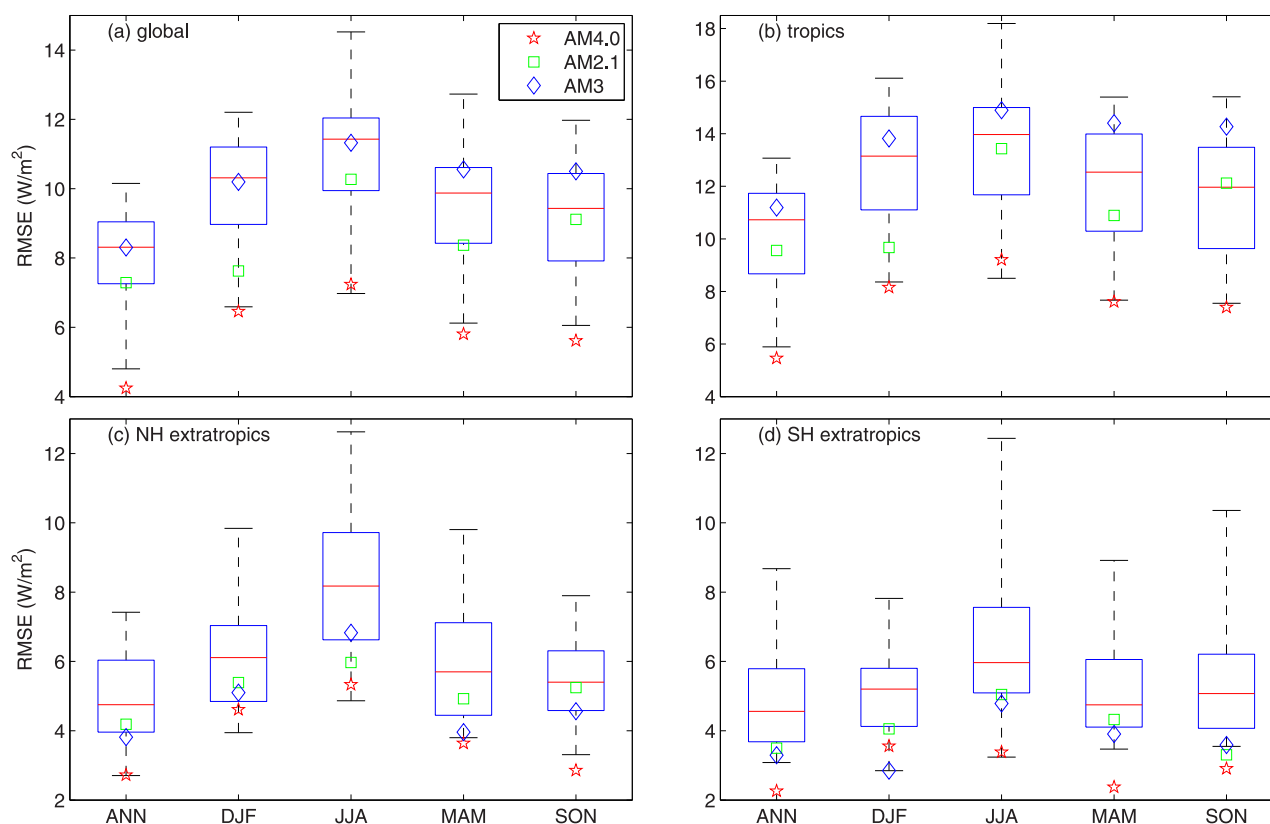
**Figure 15.** As in Figure 14 except for TOA net SW downward radiation. The observational data are from CERES.

in the radiative forcing at the TOA for doubling of  $\text{CO}_2$ . Recent development of the FV<sup>3</sup> dynamical core has focused primarily on its nonhydrostatic version, which is not relevant at the roughly 100 km resolution of this model. The land model LM4.0 is very similar to LM3.1, with some changes in parameter settings.

The deep convection scheme has been completely retooled as described in section 4 and is a key distinctive feature of this model. This double plume convection scheme has been developed in large part with an eye toward optimizing tropical precipitation and, especially, TOA fluxes. The result is a model that is closer to the observed TOA fluxes than any other AMIP model that has been entered into the CMIP database. To illustrate this, we include in Figures 14–16, box and whisker plots showing the biases in precipitation, TOA net SW and LW radiative fluxes in the CMIP5 ensemble compared to AM4.0.

The tuning leading to the cloud and precipitation simulations has been described in sections 12 and 13. Our confidence in the resulting model is increased by 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 described in Part 1. 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. Importantly, also documented in Part 1, the response to ENSO SSTs in precipitation, wind stresses, and surface heat fluxes agree well with observations and recent reanalyses. 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 as regards 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.

Another distinctive aspect of this model is the parameterization of subgrid topographic gravity wave stresses due to Garner (2005). We have tried to incorporate this scheme into both AM2 and AM3 in the past, attracted especially by its natural treatment of anisotropy. Our success in AM4.0 has resulted from a



**Figure 16.** As in Figure 14 except for TOA outgoing LW radiation.

better appreciation of the importance of some key parameters in the scheme as well as a greater willingness to tune some of these parameters to optimize the resulting circulation.

Also noteworthy are the improvements in the aerosol simulation, resulting from a combination of changes, including the removal of aerosols by convection and by precipitation from mixed phase clouds, and changes in the sulfate chemistry within cloud drops (see section 7).

Following earlier work by Mauritsen et al. (2012) and Hourdin et al. (2017) on climate model tuning processes, Schmidt et al. (2017) provides a convenient summary of tuning strategies at a number of major US climate modeling centers, allowing a comparison of the strategy used in AM4.0 development and that used by other climate modelers. The models used for decadal-to-centennial simulations in the set described by Schmidt et al. (2017) tune global mean top-of-atmosphere energy balance to some extent, as it is universally recognized that simulations leading to an accurate enough balance for long coupled simulations cannot otherwise be obtained. There is an interesting difference between groups that tune energy balance for PD (present-day) conditions and those that tune for PI (preindustrial) conditions. Our tuning has been entirely based on the standard AMIP period (1980–2014), tuning to estimates of PD energy balance using observed SSTs as boundary conditions. Some of the advantages of trying to tune instead to zero net balance in PI conditions are discussed in Schmidt et al. (2017). We feel it is important to optimize the SW and LW fluxes separately, and not just the net difference, for which we have no PI constraints.

As in AM4.0 development, models used for decadal-to-centennial simulations are often tuned, not simply for global mean TOA fluxes but for regional patterns of mean climate (e.g., shortwave and longwave cloud radiative effects, precipitation, sea level pressure, geopotential height, temperature, humidity, zonal wind, near-surface wind stress, implied ocean heat transports, and aerosol optical depth). The emphasis placed upon the various aspects of mean climate varies among modeling groups described in Schmidt et al. (2017). But it still remains rare to base a tuning strategy on an automated optimization process. In AM4.0 we have also refrained from trying to automate the tuning process. The development of AM4.0 to date has

isolated key parameters, and metrics sensitive to those parameters, helping to define the parameter space of most interest. We believe that we have prepared the way for future work on automating the final stages of the tuning process in this model, working with the key parameters identified by our work to date.

An area of difference in strategies across models is the extent to which decisions on model optimization are made in atmosphere/land AMIP simulations or in the fully coupled context. Our approach in AM4.0 development has been to develop primarily in AMIP mode, with an eye toward coupled simulations for reassurance that no serious deficiencies emerge. Aspects of the coupled model on which we focused attention were the MJO, ENSO, the strength of the Atlantic overturning circulation, the Arctic ice cover, the structure of drifts in SSTs, and ocean heat uptake, typically in simulations lasting 100 years or less. But these inspections of coupled simulations rarely provided systematic ways of modifying the atmospheric component. For example, coupled simulations of ENSO, which varied in quality through the development of the coupled model, did not always feed back directly on atmospheric development due in part to our limited ability to predict how coupled simulations would react to changes in the atmospheric model and partly to the fact that ENSO characteristics also depended on oceanic tuning in prototypes of our CM4 coupled model.

## 15. Conclusions

The tuning of the climate sensitivity and aerosol radiative forcing are of special interest when documenting a model development path, because they effectively determine the amplitude and shape of the anthropogenic warming over the past century. Other than preconceptions based on the simplest energy balance models as to the qualitative historical trajectory to expect from particular values of Cess sensitivity and aerosol RFP, we did not explicitly vary parameters to fit coupled model historical simulations. To maximize transparency, we list here the concerns and ways we tried to push Cess sensitivity and aerosol RFP in a particular direction.

1. At the start of AM4.0 development we made a choice of a microphysics scheme within the moist convection module, based on the analysis of Zhao et al. (2016) of some early AM4 prototypes. By choosing a scheme similar to that in AM2 rather than that in AM3 shallow convection, without a clear process-level rationale for doing so, we began the model development process with a relatively low Cess sensitivity close to AM2 rather than the higher value in AM3.
2. Early preliminary coupled simulations suggested that our TCR was still relatively high, 2K or even higher, more comparable to AM3 than AM2, despite a lower Cess sensitivity. This result, still not understood fully but presumably related to strong polar amplification, has encouraged us to continue thinking about alternatives for lowering the sensitivity.
3. Contributing to this concern were results from the model's newly updated radiation scheme, which increased the radiative forcing for doubling CO<sub>2</sub> by ~10%.
4. The EIS modification to the shallow convection scheme (see sections 4 and 13) reduces the positive low cloud feedback in tropical subsidence region. It also improves the simulation of stratocumulus decks. This modification was implemented not only to reduce the sensitivity but also because it improved simulations.
5. Golaz et al. (2013) describes how AM3's aerosol RFP can be manipulated by changing an autoconversion threshold parameter controlling the drop size required before precipitation occurs in the model's large-scale cloud scheme. Reducing AM3's value from 8.2 to 6  $\mu\text{m}$  was beneficial in that model for qualitatively fitting the evolution of global mean temperature. While larger values (~10  $\mu\text{m}$ ) were preferred observationally (Suzuki et al., 2015), precipitation in the real atmosphere depends not on the grid-box averaged cloud droplet size, but rather on the size of largest drops. Experimentation with a coupled version of a prototype of AM4.0 suggested that we had a similar sensitivity to that parameter but could get by with a value close to 8.2  $\mu\text{m}$  (we chose 8.5  $\mu\text{m}$ ). We then left this value unchanged as we developed AM4.0.)
6. The several changes in the aerosol model that happened to change the aerosol RFP from that in AM3 were accepted simply to improve the aerosol simulation or to simplify and improve the numerical stability of the activation scheme (section 8). They happened to decrease the magnitude of the aerosol RFP by about 30% in combination. The one change that we thought might have this effect, but that was also motivated by an improvement in the droplet numbers (i.e., allowing aerosol activation when cloud is

present rather than requiring an increasing cloud amount) in fact had very little effect on the model's RFP value.

The final model, AM4.0/LM4.0, has a Cess sensitivity of  $0.57 \text{ K W}^{-1} \text{ m}^2$ , but with continuing hints from coupled experimentation of a TCR close to 2K, and an aerosol RFP that peaks at about  $-1.1 \text{ W m}^{-2}$  around 1980 and decreases to  $-0.7 \text{ W m}^{-2}$  in 2014. While the final coupled model CM4 has not yet been frozen at the time our submission of this manuscript, the final AM4/LM4 truly parallel with the GFDL CMIP6 DECK submission will involve some changes in the land model. This involves mostly the use of dynamical vegetation and some changes in snow albedo over glacier.

## Acknowledgments

We provide the AM4.0/LM4.0 code and selected model output data from an AMIP simulation at <http://data1.gfdl.noaa.gov/nomads/forms/am4.0/>. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. We thank the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the IPCC Data Archive at the Lawrence Livermore National Laboratory/Department of Energy (LLNL/DOE) for collecting, archiving the CMIP5 data and providing the standard portrait plots for model comparison. Lawrence Livermore National Laboratory is operated by Lawrence Livermore National Security, LLC, for the U.S. Department of Energy, National Nuclear Security Administration under contract DE-AC52-07NA27344. M. Zhao, J.-C. Golaz, B. Xiang, and Y. Ming acknowledge partial support by NOAA's Climate Program Office (CPO) Climate Variability and Predictability (CVP) Program (GC14-252) through a CPO CVP funded proposal for understanding AM4/CM4 biases. We are grateful for helpful comments and suggestions from Nathaniel Johnson and Hiroyuki Murakami. We thank Catherine Raphael for assistance with particular figures and the many GFDL scientists and support staff who have not been explicitly listed as authors but supported this effort through their insight, work on previous model development efforts on which this effort is based, and work of GFDL and NOAA's software and hardware infrastructures.

## References

- Alexander, M. J., & Dunkerton, T. J. (1999). A spectral parameterization of mean-flow forcing due to breaking gravity waves. *Journal of the Atmospheric Sciences*, 56, 4167–4182.
- Arakawa, A., & Schubert, W. H. (1974). Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. *Journal of the Atmospheric Sciences*, 31, 674–701.
- Arimoto, R., Duce, R. A., Ray, B. J., Ellis, W. G., Cullen, J. D., & Merrill, J. T. (1995). Trace-elements in the atmosphere over the North-Atlantic. *Journal of Geophysical Research*, 100, 1199–1213.
- Atkinson, R., Baulch, D. L., Cox, R. A., Hampson, R. F., Jr., Kerr, J. A., & Troe, J. (1989). Evaluated kinetic and photochemical data for atmospheric chemistry: Supplement III. *International Journal of Chemical Kinetics*, 21, 115–150.
- Balaji, V. (2012). The Flexible Modeling System. In S. Valcke, R. Redler, & R. Budich (Eds.), *Earth system modelling—Volume 3, Springer briefs in earth system sciences* (pp. 33–41). Berlin, Germany: Springer.
- Balaji, V., Benson, R., Wyman, B., & Held, I. (2016). Coarse-grained component concurrency in Earth system modeling: Parallelizing atmospheric radiative transfer in the GFDL AM3 model using the Flexible Modeling System coupling framework. *Geoscientific Model Development*, 9(10), 3605–3616. <https://doi.org/10.5194/gmd-9-3605-2016>
- Balaji, V., & Langenhorst, A. (2012). ESM workflow. In R. Ford, G. Riley, R. Budich, & R. Redler (Eds.), *Earth system modelling—Volume 5, Springer briefs in earth system sciences* (pp. 5–13). Berlin, Germany: Springer.
- Balaji, V., Maisonnave, E., Zadeh, N., Lawrence, B. N., Biercamp, J., Fladrich, U., et al. (2017). CPMIP: Measurements of real computational performance of earth system models. *Geoscientific Model Development*, 10, 19–34. <https://doi.org/10.5194/gmd-10-19-2017>
- Bechtold, P., Kohler, M., Jung, T., Doblas-Reyes, F., Leutbecher, M., Rodwell, M., et al. (2008). Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*, 134, 1337–1351.
- Becker, T., Bretherton, C. S., Hohenegger, C., & Stevens, B. (2018). Estimating bulk entrainment with unaggregated and aggregated convection. *Geophysical Research Letters*, 45, 455–462. <https://doi.org/10.1002/2017GL076640>
- Becker, T., Stevens, B., & Hohenegger, C. (2017). Imprint of the convective parameterization and sea-surface temperature on large-scale convective self-aggregation. *Journal of Advances in Modeling Earth Systems*, 9, 1488–1505. <https://doi.org/10.1002/2016MS000865>
- Beven, K., & Germann, P. (1982). Macropores and water flow in soils. *Water Resources Research*, 18, 1311–1325.
- Bony, S., Dufresne, J.-L., Treut, H. L., Morcrette, J.-J., & Senior, C. (2004). On dynamic and thermodynamic components of cloud changes. *Climate Dynamics*, 22, 71–86.
- Bretherton, C. S., McCaig, J. R., & Grenier, H. (2004a). A new parameterization for shallow cumulus convection and its application to marine subtropical cloud-topped boundary layers. Part I: Description and 1-D results. *Monthly Weather Review*, 132, 864–882.
- Bretherton, C. S., Peters, M. E., & Back, L. E. (2004b). Relationships between water vapor path and precipitation over the tropical oceans. *Journal of Climate*, 17, 1517–1528.
- Brown, R. G., & Zhang, C. (1997). Variability of midtropospheric moisture and its effect on cloud-top height distribution during TOGA COARE. *Journal of the Atmospheric Sciences*, 54, 2760–2774.
- Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., et al. (2002). Tropospheric optical thickness from the GOCART model and comparisons with satellite and Sun photometer measurements. *Journal of the Atmospheric Sciences*, 59, 461–483.
- Clough, S. A., Kneizys, F. X., & Davies, R. W. (1989). Line shape and the water vapor continuum. *Atmospheric Research*, 23, 229–241.
- Crueger, T., Stevens, B., & Brokopf, R. (2013). The Madden Julian Oscillation in ECHAM6 and the introduction of an objective MJO metric. *Journal of Climate*, 26, 3241–3257.
- 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>
- DelGenio, A. (2012). Representing the sensitivity of convective cloud systems to tropospheric humidity in general circulation models. *Surveys in Geophysics*, 33, 637–656. <https://doi.org/10.1007/s10712-011-9148-9>
- DeMore, W., Sander, S., Golden, D., Hampson, R., Kurylo, M., Howard, C., et al. (1997). *Chemical kinetics and photochemical data for use in stratospheric modeling* (Publ. 97-4). Pasadena, CA: JPL.
- Derbyshire, S., Beau, I., Bechtold, P., Grandpeix, J.-Y., Piriou, J.-M., Redelsperger, J.-L., et al. (2004). Sensitivity of moist convection to environmental humidity. *Quarterly Journal of the Royal Meteorological Society*, 130, 3055–3079.
- de Rooy, W., Bechtold, P., Frohlich, K., Hohenegger, C., Jonker, H., Mironov, D., et al. (2013). Entrainment and detrainment in cumulus convection: An overview. *Quarterly Journal of the Royal Meteorological Society*, 139, 1–9.
- Donner, L., Seman, C., Hemler, R., & Fan, S.-M. (2001). A cumulus parameterization including mass fluxes, convective vertical velocities, and mesoscale effects. *Journal of Climate*, 14, 3444–3463.
- Donner, L., Seman, C., Hemler, R., & Fan, S.-M. (2016). Are atmospheric updrafts a key to unlocking climate forcing and sensitivity? *Atmospheric Chemistry and Physics*, 16, 12983–12992. <https://doi.org/10.5194/acp-16-12983-2016>
- 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.
- Edson, J. B., Jampana, V., Weller, R. A., Bigorre, S. P., Plueddemann, A. J., Fairall, C. W., et al. (2013). On the exchange of momentum over the open ocean. *Journal of Physical Oceanography*, 43, 1589–1610. <https://doi.org/10.1175/JPO-D-12-0173.1>



- Emanuel, K. A. (1991). A scheme for representing cumulus convection in large-scale models. *Journal of the Atmospheric Sciences*, 48, 2313–2335.
- Fang, Y., Fiore, A. M., Horowitz, L. W., Gnanadesikan, A., Held, I. M., Chen, G., et al. (2011). The impacts of changing transport and precipitation on pollutant distributions in a future climate. *Journal of Geophysical Research*, 116, D18303. <https://doi.org/10.1029/2011JD015642>
- Fels, S., & Schwarzkopf, D. M. (1975). The simplified exchange approximation: A new method for radiative transfer calculations. *Journal of the Atmospheric Sciences*, 32, 1475–1488.
- Fiedler, S., Stevens, B., & Mauritsen, T. (2017). On the sensitivity of anthropogenic aerosol forcing to model-internal variability and parameterizing a Twomey effect. *Journal of Advances in Modeling Earth Systems*, 9, 1325–1341. <https://doi.org/10.1002/2017MS000932>
- Fischer, G., Nachtergaele, F., Prieler, S., van Velthuisen, H., Verelst, L., & Wiberg, D. (2008). *Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008)*. Laxenburg, Austria: FAO.
- Freidenreich, S. M., & Ramaswamy, V. (2005). Refinement of the Geophysical Fluid Dynamics Laboratory solar benchmark computations and an improved parameterization for climate models. *Journal of Geophysical Research*, 110, D17105. <https://doi.org/10.1029/2004JD005471>
- Garner, S. (2005). A topographic drag closure built on an analytical base flux. *Journal of the Atmospheric Sciences*, 62, 2302–2315.
- GFDL-GAMDT. (2004). The new GFDL global atmosphere and land model AM2/LM2: Evaluation with prescribed SST simulations. *Journal of Climate*, 17, 4641–4673.
- Gillett, N. P., Arora, V. K., Flato, G. M., Scinocca, J. F., & vonSalzen, K. (2012). Improved constraints on 21st-century warming derived using 160 years of temperature observations. *Geophysical Research Letters*, 39, L01704. <https://doi.org/10.1029/2011GL050226>
- Ginoux, P., Chin, M., Tegen, I., Prospero, J. M., Holben, B., Dubovik, O., et al. (2001). Sources and distributions of dust aerosols simulated with the GOCART model. *Journal of Geophysical Research*, 106, 20255–20273. <https://doi.org/10.1029/2000JD000053>
- Golaz, J.-C., Horowitz, L. W., & Levy, H., II (2013). Cloud tuning in a coupled climate model: Impact on 20th century warming. *Geophysical Research Letters*, 40, 2246–2251. <https://doi.org/10.1002/grl.50232>
- Golaz, J.-C., Larson, V. E., & Cotton, W. R. (2002). A PDF-based model for boundary layer clouds Part I: Method and model description. *Journal of the Atmospheric Sciences*, 59, 3540–3551.
- Golaz, J.-C., Salzmann, M., Donner, L. J., Horowitz, L. W., Ming, Y., & Zhao, M. (2011). Sensitivity of the aerosol indirect effect to subgrid variability in the cloud parameterization of the GFDL atmosphere general circulation model AM3. *Journal of Climate*, 24, 3145–3160.
- Guo, H., Golaz, J.-C., Donner, L. J., Ginoux, P., & Hemler, R. S. (2014). Multi-variate probability density functions with dynamics in the GFDL atmospheric general circulation model: Global tests. *Journal of Climate*, 27, 2087–2108. <https://doi.org/10.1175/JCLI-D-13-00347.1>
- Guo, H., Golaz, J.-C., Donner, L. J., Wyman, B., Zhao, M., & Ginoux, P. (2015). CLUBB as a unified cloud parameterization: Opportunities and challenges. *Geophysical Research Letters*, 42, 4540–4547. <https://doi.org/10.1002/2015GL063672>
- Harris, L., & Lin, S.-J. (2013). A two-way nested global-regional dynamical core on the cubed-sphere grid. *Monthly Weather Review*, 141, 283–306. <https://doi.org/10.1175/MWR-D-11-00201.1>
- Held, I. M., Zhao, M., & Wyman, B. (2007). Dynamic radiative-convective equilibria using GCM column physics. *Journal of the Atmospheric Sciences*, 64, 228–238.
- Heymsfield, A. J., & Donner, L. J. (1990). A scheme for parameterizing ice-cloud water content in general circulation models. *Journal of the Atmospheric Sciences*, 47, 1865–1877.
- Hillel, D. (Ed.). (1980). *Fundamentals of soil physics* (413 pp.). Cambridge, MA: Academic.
- Holloway, C. E., & Neelin, J. (2009). Moisture vertical structure, column water vapor, and tropical deep convection. *Journal of the Atmospheric Sciences*, 66, 1665–1683.
- Horowitz, L. W., Walters, S., Mauzerall, D., Emmons, L., Rasch, P., Granier, C., et al. (2003). A global simulation of tropospheric ozone and related tracers: Description and evaluation of MOZART, version 2. *Journal of Geophysical Research*, 108(D24), 4784. <https://doi.org/10.1029/2002JD002853>
- Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., et al. (2017). The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, 98, 589–602. <https://doi.org/10.1175/BAMS-D-15-00135.1>
- Hurt, 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>
- Jacobson, M. Z. (Ed.). (2005). *Fundamentals of atmospheric modeling* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Johnson, R. (1997). Recent observations of deep convection: TOGA COARE. In *New insights and approaches to convective parametrization* (pp. 1–24). Reading, UK: ECMWF.
- Kim, D., Sobel, A., DelGenio, A., Chen, Y., Camargo, S., Yao, M.-S., et al. (2012). The tropical subseasonal variability simulated in the NASA GISS general circulation model. *Journal of Climate*, 25, 4641–4659.
- Kim, Y.-J., & Doyle, J. (2005). Extension of an orographic-drag parameterization scheme to incorporate orographic anisotropy and flow blocking. *Quarterly Journal of the Royal Meteorological Society*, 131, 1893–1921.
- Koch, D., Jacob, D., Tegen, I., Rind, D., & Chin, M. (1999). Tropospheric sulfur simulation and sulfate direct radiative forcing in the goddard institute for space studies general circulation model. *Journal of Geophysical Research*, 104, 23799–23822. <https://doi.org/10.1029/1999JD900248>
- Larson, V. E., & Golaz, J.-C. (2005). Using probability density functions to derive consistent closure relationships among higher-order moments. *Monthly Weather Review*, 133, 1023–1042.
- Leopold, L., & T. Maddock (Eds.). (1953). *Hydraulic geometry of stream channels and some physiographic implications* (Prof. Pap. 252, 55 pp.). Reston, VA: U.S. Geological Survey.
- Leuning, R. (1995). A critical appraisal of a combined stomatal-photosynthesis model for C-3 plants. *Plant, Cell & Environment*, 18, 339–355.
- Lin, J.-L., Kiladis, G., Mapes, B., Weickmann, K., Sperber, K., Lin, W., et al. (2006). Tropical intraseasonal variability in 14 IPCC AR4 climate model. Part I: Convective signals. *Journal of Climate*, 19, 2665–2690.
- Lin, S.-J. (1997). A finite-volume integration method for computing pressure gradient force in general vertical coordinates. *Quarterly Journal of the Royal Meteorological Society*, 123, 1749–1762.
- Lin, S.-J. (2004). A “vertically Lagrangian” finite-volume dynamical core for global models. *Monthly Weather Review*, 132, 2293–2307.
- Lin, S.-J., & Rood, R. (1997). An explicit flux-form semi-Lagrangian shallow-water model on the sphere. *Quarterly Journal of the Royal Meteorological Society*, 123, 2477–2498.
- Lin, S.-J., & Rood, R. B. (1996). Multidimensional flux-form semi-lagrangian transport schemes. *Monthly Weather Review*, 124, 2046–2070.
- Lindzen, R. S. (1981). Turbulence and stress owing to gravity wave and tidal breakdown. *Journal of Geophysical Research*, 86, 9707–9714.
- Liu, J., Fan, S., Horowitz, L. W., & Levy, H., II (2011). Evaluation of factors controlling long-range transport of black carbon to the Arctic. *Journal of Geophysical Research*, 116, D04307. <https://doi.org/10.1029/2010JD015145>

- Lock, A. P., Brown, A. R., Bush, M. R., Martin, G. M., & Smith, R. (2000). A new boundary layer mixing scheme. part I: Scheme description and single-column model tests. *Monthly Weather Review*, 128, 3187–3199.
- Malavelle, F. F., Haywood, J. M., Jones, A., Gettelman, A., Clarisse, L., Bauduin, S., et al. (2017). Strong constraints on aerosol cloud interactions from volcanic eruptions. *Nature*, 546, 485–491. <https://doi.org/10.1038/nature22974>
- Mapes, B. (2000). Convective inhibition, subgrid-scale triggering energy, and stratiform instability in a toy tropical wave model. *Journal of the Atmospheric Sciences*, 57, 1515–1535.
- Mapes, B., & Neale, R. (2011). Parameterizing convective organization to escape the entrainment dilemma. *Journal of Advances in Modeling Earth Systems*, 3, M06004. <https://doi.org/10.1029/2011MS000042>
- Martensson, E., Nilsson, E., de Leeuw, G., Cohen, H., & Hansson, H.-C. (2003). Laboratory simulations and parameterization of the primary marine aerosol production. *Journal of Geophysical Research*, 108(D9), 4297. <https://doi.org/10.1029/2002JD002633>
- Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., et al. (2012). Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems*, 4, M00A01. <https://doi.org/10.1029/2012MS000154>
- 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.
- Ming, Y., Ramaswamy, V., Donner, L. J., & Phillips, V. (2006). A robust parameterization of cloud droplet activation. *Journal of the Atmospheric Sciences*, 63, 1348–1356.
- Ming, Y., Ramaswamy, V., Donner, L. J., Phillips, V., Klein, S., Ginoux, P., et al. (2007). Modeling the interactions between aerosols and liquid water clouds with a self-consistent cloud scheme in a general circulation model. *Journal of the Atmospheric Sciences*, 64, 1189–1209. <https://doi.org/10.1175/JAS3874.1>
- Mlawer, E. J., Payne, V. H., Moncet, J.-L., Delamere, J. S., Alvarado, M. J., & Tobin, D. C. (2012). Development and recent evaluation of the MT\_CKD model of continuum absorption. *Philosophical Transactions of the Royal Society A*, 370, 2520–2556. <https://doi.org/10.1098/rsta.2011.0295>
- Monahan, E., Spiel, D., & Davidson, K. (1986). A model of marine aerosol generation via whitecaps and wave disruption. In E. Monahan (Ed.), *Oceanic whitecaps: And their role in air-sea exchange processes* (pp. 167–174). Dordrecht, Holland: D. Reidel.
- Moorthi, S., & Suarez, M. (1992). Relaxed Arakawa Schubert: A parameterization of moist convection for general circulation models. *Monthly Weather Review*, 120, 978–1002.
- 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: Atmospheres*, 118, 8086–8110. <https://doi.org/10.1002/jgrd.50608>
- Neale, R. B., Richter, J. H., & Jochum, M. (2008). The impact of convection on ENSO: From a delayed oscillator to a series of events. *Journal of Climate*, 21, 5904–5924.
- Neelin, J. D., & Held, I. M. (1987). Modeling tropical convergence based on the moist static energy budget. *Monthly Weather Review*, 115, 3–12.
- Oreopoulos, L., Mlawer, E., Delamere, E., Shippert, T., Cole, J., Fomin, B., et al. (2012). The continual intercomparison of radiation codes: Results from phase I. *Journal of Geophysical Research*, 117, D06118. <https://doi.org/10.1029/2011JD016821>
- Palmer, T., Shutts, G., & Swinbank, R. (1986). Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parametrization. *Quarterly Journal of the Royal Meteorological Society*, 112, 1001–1039.
- Paulot, F., Fan, S., & Horowitz, L. W. (2017). Contrasting seasonal responses of sulfate aerosols to declining SO<sub>2</sub> emissions in the Eastern U.S.: Implications for the efficacy of SO<sub>2</sub> emission controls. *Geophysical Research Letters*, 43, 455–464. <https://doi.org/10.1002/2016GL070695>
- Paulot, F., Ginoux, P., Cooke, W. F., Donner, L. J., Fan, S., Lin, M.-Y., et al. (2016). Sensitivity of nitrate aerosols to ammonia emissions and to nitrate chemistry: Implications for present and future nitrate optical depth. *Atmospheric Chemistry and Physics*, 16, 1459–1477.
- Paynter, D. J., & Ramaswamy, V. (2012). Variations in water vapor continuum radiative transfer with atmospheric conditions. *Journal of Geophysical Research*, 117, D16310. <https://doi.org/10.1029/2012JD017504>
- Paynter, D. J., & Ramaswamy, V. (2014). Investigating the impact of the shortwave water vapor continuum upon climate simulations using gfdl global models. *Journal of Geophysical Research: Atmospheres*, 119, 10720–10737. <https://doi.org/10.1002/2014JD021881>
- Pierrehumbert, R. T., & Wyman, B. (1985). Upstream effects of mesoscale mountains. *Journal of the Atmospheric Sciences*, 42, 977–1003.
- Pincus, R., Barker, H. W., & Morcrette, J.-J. (2003). A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields. *Journal of Geophysical Research*, 108(D13), 4376. <https://doi.org/10.1029/2002JD003322>
- 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., Hannay, C., Klein, S., Xu, K.-M., & Hemler, R. (2005). Overlap assumptions for assumed probability distribution function cloud schemes in large-scale models. *Journal of Geophysical Research*, 110, D15S09. <https://doi.org/10.1029/2004JD005100>
- Pincus, R., Hemler, R., & Klein, S. A. (2006). Using stochastically generated subcolumns to represent cloud structure in a large-scale model. *Monthly Weather Review*, 134, 3644–3656.
- Prospero, J. M., Uematsu, M., & Savoie, D. L. (1989). Mineral aerosol transport to the Pacific Ocean. In *Chemical oceanography* (chap. 58, pp. 187–218). New York, NY: Academic.
- Putman, W. M., & Lin, S.-J. (2007). Finite-volume transport on various cubed-sphere grid. *Journal of Computational Physics*, 227, 55–78.
- Quaas, J., Ming, Y., Menon, S., Takemura, T., Wang, M., Penner, J. E., et al. (2009). Aerosol indirect effects—General circulation model intercomparison and evaluation with satellite data. *Atmospheric Chemistry and Physics*, 9, 8697–8717.
- Randall, D. A., & Wielicki, B. A. (1997). Measurements, models, and hypotheses in the atmospheric sciences. *Bulletin of the American Meteorological Society*, 78, 399–406.
- Raymond, D. J. (2000). Thermodynamic control of tropical rainfall. *Quarterly Journal of the Royal Meteorological Society*, 126, 889–898.
- Raymond, D. J., & Blyth, A. M. (1986). A stochastic mixing model for nonprecipitating cumulus clouds. *Journal of the Atmospheric Sciences*, 43, 2708–2718.
- Raymond, D. J., & Zeng, X. (2000). Instability and large-scale circulations in a two-column model of the tropical troposphere. *Quarterly Journal of the Royal Meteorological Society*, 126, 3117–3135.
- Redelsperger, J.-L., Parsons, D. B., & Guichard, F. (2002). Recovery processes and factors limiting cloud-top height following the arrival of a dry intrusion observed during TOGA COARE. *Journal of the Atmospheric Sciences*, 59, 2438–2457.
- Roms, D. M. (2011). A direct measure of entrainment. *Journal of the Atmospheric Sciences*, 67, 1908–1927.
- Rothman, L. S., Barbe, A., Benner, D. C., Brown, L. R., Camy-Peyret, C., Carleer, M. R., et al. (2003). The HITRAN molecular spectroscopic database: Edition of 2000 including updates through 2001. *Journal of Quantitative Spectroscopy & Radiative Transfer*, 82, 5–44.
- Rothman, L. S., Gordon, I. E., Babikov, Y., Barbe, A., Benner, D. C., Bernath, P. F., et al. (2013). The HITRAN2012 molecular spectroscopic database. *Journal of Quantitative Spectroscopy & Radiative Transfer*, 130, 4–50.

- Rotstayn, L. D. (1997). A physically based scheme for the treatment of stratiform clouds and precipitation in large-scale models. I: Description and evaluation of the microphysical processes. *Quarterly Journal of the Royal Meteorological Society*, 123, 1227–1282.
- Rotstayn, L. D., Ryan, B. F., & Katzfey, J. (2000). A scheme for calculation of the liquid fraction in mixed-phase clouds in large-scale models. *Monthly Weather Review*, 128, 1070–1088.
- Schaaf, C., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., et al. (2002). First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sensing of Environment*, 83, 135–148.
- Schmidt, G. A., Bader, D., Donner, L., Elsaesser, G. S., Golaz, J.-C., Hanay, C., et al. (2017). Practice and philosophy of climate model tuning across six US modeling centers. *Geoscientific Model Development*, 10, 3207–3223. <https://doi.org/10.5194/gmd-10-3207-2017>
- Schwarzkopf, D. M., & Fels, S. (1991). The simplified exchange method revisited: An accurate, rapid method for computation of infrared cooling rates and fluxes. *Journal of Geophysical Research*, 96, 9075–9096.
- Schwarzkopf, D. M., & Ramaswamy, V. (1999). Radiative effects of CH<sub>4</sub>, N<sub>2</sub>O, halocarbons and the foreign-broadened H<sub>2</sub>O continuum: A GCM experiment. *Journal of Geophysical Research*, 104, 9467–9488.
- Scinocca, J., & McFarlane, N. (2000). The parametrization of drag induced by stratified flow over anisotropic orography. *Quarterly Journal of the Royal Meteorological Society*, 126, 2353–2393.
- Seifert, A., Heus, T., Pincus, R., & Stevens, B. (2015). Large-eddy simulation of the transient and near-equilibrium behavior of precipitating shallow convection. *Journal of Advances in Modeling Earth Systems*, 7, 1918–1937. <https://doi.org/10.1002/2015MS000489>
- Seinfeld, J. H., & Pandis, S. N. (1998). *Atmospheric chemistry and physics*. New York, NY: John Wiley.
- 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.
- Sherwood, S. C. (1999). Convective precursors and predictability in the tropical western pacific. *Monthly Weather Review*, 127, 2977–2991.
- Shevliakova, E., Pacala, S. W., Malyshev, S., Hurr, G. C., Milly, P. C. D., Caspersen, J. P., et al. (2009). Carbon cycling under 300 years of land use change: Importance of the secondary vegetation sink. *Global Biogeochemical Cycles*, 23, GB2022. <https://doi.org/10.1029/2007GB003176>
- Siebesma, A. P., Bretherton, C. S., Brown, A. R., Chlond, A., Cuszrt, J., Duynkerke, P. G., et al. (2003). An intercomparison study for cloud resolving models of shallow cumulus convection. *Journal of the Atmospheric Sciences*, 60, 1201–1219.
- Stern, W. F., & Pierrehumbert, R. T. (1988). The impact of an orographic gravity wave drag parameterization on extended range predictions with a GCM. In *Eighth Conference on Numerical Weather Prediction* (pp. 745–750). Baltimore, MD: American Meteorological Society.
- Stevens, B. (2015). Rethinking the lower bound on aerosol radiative forcing. *Journal of Climate*, 28, 4794–4819.
- Stevens, B. (2017). Clouds unfazed by haze. *Nature*, 546, 483–484.
- Stevens, B., Walko, R. L., & Cotton, W. R. (1996). The spurious production of cloud-edge supersaturations by eulerian models. *Monthly Weather Review*, 124, 1034–1041.
- Suzuki, K., Stephens, G. L., Bodas-Salcedo, A., Wang, M., Golaz, J.-C., Yokohata, T., et al. (2015). Evaluation of the warm rain formation process in global models with satellite observations. *Journal of the Atmospheric Sciences*, 72, 3996–4014. <https://doi.org/10.1175/JAS-D-14-0265.1>
- Tiedtke, M. (1993). Representation of clouds in large-scale models. *Monthly Weather Review*, 121, 3040–3061.
- Tokioka, T., Yamazaki, K., Kitoh, A., & Ose, T. (1988). The equatorial 30–60 day oscillation and the Arakawa-Schubert penetrative cumulus parameterization. *Journal of the Meteorological Society of Japan*, 66, 883–901.
- Tompkins, A. M. (2001). Organization of tropical convection in low vertical wind shears: The role of water vapor. *Journal of the Atmospheric Sciences*, 58, 529–545.
- Vial, J., Bony, S., Dufresne, J.-L., & Roehrig, R. (2016). Coupling between lower-tropospheric convective mixing and low-level clouds: Physical mechanisms and dependence on convection scheme. *Journal of Advances in Modeling Earth Systems*, 8, 1892–1911. <https://doi.org/10.1002/2016MS000740>
- Vial, J., Bony, S., Stevens, B., & Vogel, R. (2017). Mechanisms and model diversity of trade-wind shallow cumulus cloud feedbacks: A review. *Surveys in Geophysics*, 38, 1331–1353. <https://doi.org/10.1007/s10712-017-9418-2>
- von Salzen, K., & McFarlane, N. A. (2002). Parameterization of the bulk effects of lateral and cloud-top entrainment in transient shallow cumulus clouds. *Journal of the Atmospheric Sciences*, 59, 1405–1430.
- 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.
- Wood, R., & Bretherton, C. S. (2006). On the relationship between stratiform low cloud cover and lower-tropospheric stability. *Journal of Climate*, 19, 6425–6432.
- Wyant, M. C., Bretherton, C. S., Bacmeister, J. T., Kiehl, J. T., Held, I. M., Zhao, M., et al. (2006). A comparison of low-latitude cloud properties and their response to climate change in three AGCMs sorted into regimes using mid-tropospheric vertical velocity. *Climate Dynamics*, 27, 261–279.
- Zhang, C., Mapes, B. E., & Soden, B. J. (2003). Bimodality in tropical water vapour. *Quarterly Journal of the Royal Meteorological Society*, 129, 2847–2866.
- Zhang, G. J., & McFarlane, N. A. (1995). Sensitivity of climate simulations to the parameterization of cumulus convection in the canadian climate centre general circulation model. *Atmosphere-Ocean*, 33(3), 407–446.
- 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., & Austin, P. H. (2003). Episodic mixing and buoyancy-sorting representations of shallow convection: A diagnostic study. *Journal of the Atmospheric Sciences*, 60, 892–912.
- Zhao, M., & Austin, P. H. (2005a). Life cycle of numerically simulated shallow cumulus clouds. Part I: Transport. *Journal of the Atmospheric Sciences*, 62, 1269–1290.
- Zhao, M., & Austin, P. H. (2005b). Life cycle of numerically simulated shallow cumulus clouds. Part II: Mixing dynamics. *Journal of the Atmospheric Sciences*, 62, 1292–1310.
- 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: 1. Simulation characteristics with prescribed SSTs. *Journal of Advances in Modeling Earth Systems*. <https://doi.org/10.1002/2017MS001208>, in press.
- Zhao, M., Held, I. M., & Lin, S.-J. (2012). Some counterintuitive dependencies of tropical cyclone frequency on parameters in a GCM. *Journal of the Atmospheric Sciences*, 69, 2272–2283.
- 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.