# Comparing clouds and their seasonal variations in ten atmospheric general circulation models with satellite measurements

M. H. Zhang,[1] W. Y. Lin,[1] S. A. Klein,[2,3] J. T. Bacmeister,[4] S. Bony,[5] R. T. Cederwall,[6] A. D. Del Genio,[7] J. J. Hack,[8] N. G. Loeb,[9] U. Lohmann,[10,11] P. Minnis,[9] I. Musat,[5] R. Pincus,[12] P. Stier,[13] M. J. Suarez,[4] M. J. Webb,[14] J. B. Wu,[1] S. C. Xie,[6] M.-S. Yao,[7] and J. H. Zhang[10]

Received 14 May 2004; revised 18 October 2004; accepted 26 November 2004; published XX Month 2005.

[1] To assess the current status of climate models in simulating clouds, basic cloud climatologies from ten atmospheric general circulation models are compared with satellite measurements from the International Satellite Cloud Climatology Project (ISCCP) and the Clouds and Earth's Radiant Energy System (CERES) program. An ISCCP simulator is employed in all models to facilitate the comparison. Models simulated a four-fold difference in high-top clouds. There are also, however, large uncertainties in satellite high thin clouds to effectively constrain the models. The majority of models only simulated 30–40% of middle-top clouds in the ISCCP and CERES data sets. Half of the models underestimated low clouds, while none overestimated them at a statistically significant level. When stratified in the optical thickness ranges, the majority of the models simulated optically thick clouds more than twice the satellite observations. Most models, however, underestimated optically intermediate and thin clouds. Compensations of these clouds biases are used to explain the simulated longwave and shortwave cloud radiative forcing at the top of the atmosphere. Seasonal sensitivities of clouds are also analyzed to compare with observations. Models are shown to simulate seasonal variations better for high clouds than for low clouds. Latitudinal distribution of the seasonal variations correlate with satellite measurements at >0.9, 0.6–0.9, and −0.2–0.7 levels for high, middle, and low clouds, respectively. The seasonal sensitivities of cloud types are found to strongly depend on the basic cloud climatology in the models. Models that systematically underestimate middle clouds also underestimate seasonal variations, while those that overestimate optically thick clouds also overestimate their seasonal sensitivities. Possible causes of the systematic cloud biases in the models are discussed.

**Citation:** Zhang, M. H., et al. (2005), Comparing clouds and their seasonal variations in ten atmospheric general circulation models with satellite measurements, *J. Geophys. Res.*, *110*, XXXXXX, doi:10.1029/2004JD005021.



## 1. Introduction

[2] Clouds are intrinsically coupled with the chaotic moist atmospheric circulations. Aside from directly interacting with air motions through latent heating, clouds also produce a net energy loss or gain to the atmosphere-Earth system through their radiative effects. Variations of clouds thus have the potential to either amplify or reduce a climate change. It has been known that the sensitivity of a climate model strongly depends on its clouds [*Cess et al.*, 1990; *Senior and Mitchell*, 1993, 1996; *Le Treut and Li*, 1991; *Roeckner et al.*, 1987], and models simulate different cloud feedbacks [*Cubasch et al.*, 2001; *Bony et al.*, 2004]. In the last 10 years, several research programs have been orga-

[1]Institute For Terrestrial and Planetary Atmospheres, Stony Brook University, Stony Brook, New York, USA.
[2]NOAA Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey, USA.
[3]Now at Atmospehric Science Division, Lawrence Livermore National Laboratory, Livermore, California, USA.
[4]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.
[5]Laboratoire de Météorologie Dynamique, Institut Pierre-Simon Laplace, Centre Nationale de Recherche Scientifique, Paris, France.
[6]Atmospheric Science Division, Lawrence Livermore National Laboratory, Livermore, California, USA.
[7]NASA Goddard Institute for Space Studies, New York, New York, USA.
[8]National Center for Atmospheric Research, Boulder, Colorado, USA.
[9]NASA Langley Research Center, Hampton, Virginia, USA.
[10]Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada.
[11]Now at Institut für Atmosphäre und Klima, Eidgenössische Technische Hochschule, Zürich, Switzerland.
[12]NOAA Cooperative Institute for Research in Environmental Studies, Climate Diagnostics Center, Boulder, Colorado, USA.
[13]Max Planck Institute for Meteorology, Hamburg, Germany.
[14]Hadley Centre for Climate Prediction and Research, UK Met Office, Bracknell, UK.


0148-0227/05/2004JD005021$09.00

nized to narrow the range of uncertainty in cloud-climate feedbacks.

[3] The Cloud Parameterization and Modeling Working Group (CPM) within the Atmospheric Radiation Measurement (ARM) program organized a model intercomparison project to compare cloud climatologies from general circulations models with satellite measurements. The first purpose of the project is to assess the current status of climate models in simulating clouds so that future progress can be more objectively measured. The second purpose is to reveal serious deficiencies in these models so as to guide future measurement and single-column modeling/cloud system resolving modeling (SCM/CSRM) activities [*Randall et al.*, 1996; *Ghan et al.*, 2000; *Xu et al.*, 2002; *Xie et al.*, 2002]. *Weare and AMIP Modeling Groups* [1996] compared zonally averaged model clouds from the Atmospheric Model Intercomparison Projects (AMIP I) with surface and satellite measurements. *Weare* [2004] also evaluated clouds in six AMIP II models against observations. The AMIP I study showed that global means of model high cloudiness are about two to five times larger than satellite measurements, but low clouds are 10% to 20% less than satellite and surface observations. The AMIP II study showed that the models simulated moderately well cloud albedo, but not the cloud water path.

[4] The effort to more quantitatively characterize cloud errors in models has been hampered by the considerable amount of uncertainties of available cloud data. Each cloud data set, whether satellite based or surface based, has its specific viewing geometry that blocks some clouds. Past efforts have tried to empirically adjust model clouds into diagnostics that are comparable to observations. Assumptions made in these adjustments introduce additional uncertainties in the comparison. Another uncertainty is about the exact definition of cloudiness. Within a model, one can either use a threshold of hydrometeor concentration or an optical depth of the cloud condensate to define clouds. In observations, however, the threshold is dependent on the cloud detection algorithm, which is further related with satellite pixel sizes. In this regard, cloud radiative forcing (CRF) is a more objective measure. However, CRF only measures the accumulative effects of clouds and therefore it does not necessarily provide the physical insight on particular model biases.

[5] The present study of evaluating clouds in climate models is aided by simulating the results of the International Satellite Cloud Climatology Project (ISCCP; see *Rossow and Schiffer* [1999]) through an ISCCP simulator developed in *Klein and Jakob* [1999] and *Webb et al.* [2001]. The ISCCP simulator not only minimizes the difference of sampling geometry between the models and data, but also allows the comparison of model cloud types with measurements that are stratified into both altitude ranges and optical properties. This paper reports results from this intercomparison project. Several studies have used the same approach to evaluate model clouds and their associated cloud radiative forcing at the top of the atmosphere [*Webb et al.*, 2001; *Tselioudis and Jakob*, 2002; *Williams et al.*, 2003; *Lin and Zhang*, 2004]. This study employs a wide range of models with different physical parameterization components to analyze common model biases. Furthermore, seasonal sensitivities of clouds in the models are evaluated and their relationships with the basic climatology are examined. The paper is organized as follows. Section 2 gives a brief description of the data and the models used. Section 3 analyzes common model biases. Section 4 studies the seasonal sensitivity of cloud types in the models. The last section contains a summary and discussion of the results.

## 2. Data and Models

### 2.1. ISCCP and CERES Cloud Data

[6] The ISCCP D2 monthly cloud frequencies are used in the comparison. ISCCP combines infrared and visible radiances from geostationary satellites with the TIROS Operational Vertical Sounder (TOVS) atmospheric temperature and humidity as well as correlative surface ice/snow data to obtain cloud information. It first identifies whether an image pixel of size 4–10 km is cloudy or clear, and then retrieves the optical thickness. Optical depth is used to estimate cloud emissivity, which is then used to determine the cloud-top pressure. The cloudy pixels are sorted into different bins defined by ranges in cloud-top pressure (height) and cloud optical depth [*Rossow et al.*, 1996] sampled at a nominal resolution of 30 km and 3 hours. The frequencies of each cloud bin are accumulated to an equal area map with 280 km resolution. They are further averaged to monthly means. Three bins of cloud-top pressures (high: <440 mb, middle: 440–680 mb, and low: >680 mb) and three bins of cloud optical thickness (thin: <3.6, medium: 3.6–23, and thick >23) define nine ISCCP cloud types. They will be referred to as high thin or middle thick clouds, and so on.

[7] Four seasons of CERES clouds binned into the same ISCCP optical depth and cloud altitude ranges are also used in this study. This product is based on pixel-level radiance data in multiple channels from the *Terra* Moderate Resolution Imaging Spectrometer (MODIS) [*King et al.*, 1992]. The CERES cloud detection algorithm differs from ISCCP in that it uses four instead of two wavelengths to decide whether a given pixel is clear or cloudy [*Trepte et al.*, 1999]. Radiance measurements at these four wavelengths are also used to estimate cloud phase, optical depth, particle size, and temperature [*Minnis et al.*, 1995, 1998]. Cloud-top height is determined from cloud temperature using a lapse rate method [*Minnis et al.*,. 1992; *Garreaud et al.*, 2001] for low clouds over ocean areas and using the vertical temperature and water vapor profiles from the NASA Global Modeling and Assimilation (GMAO)'s Goddard Earth Observing System DAS product (GEOS-DAS V4.0.3) (see http://dao.gsfc.nasa.gov/subpages/atbd.html) for other regions. In addition, the CERES cloud analysis is performed using 1-km MODIS pixels sampled to a 4-km resolution instead of taking one 4–10 km pixel in each 30 km by 30 km box as in ISCCP. Regional cloud type frequencies derived for 1° × 1° latitude-longitude equal area regions are used to determine cloud type frequencies of occurrence over the entire 60°S–60°N domain.

[8] Caveats in both the ISCCP and CERES data include uncertainties in cloud detection, partial cloudy pixels, and emissivity correction of cloud top altitudes. These mostly affect optically thin clouds. Multilayer clouds with thin high clouds above low clouds pose a special difficulty for ISCCP

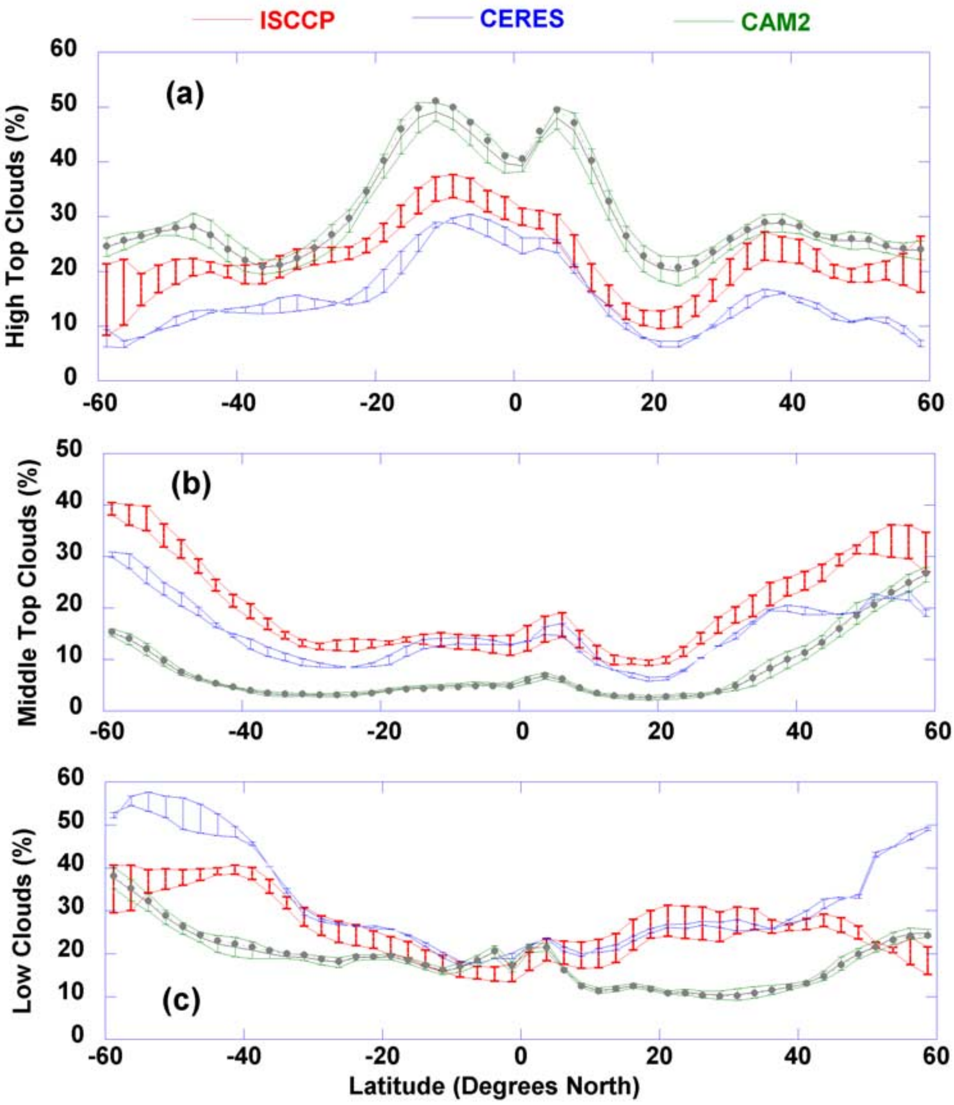


**Figure 1.** Interannual ranges of clouds DJF seasons from ISCCP (red), CERES (blue), and CAM2 (green) for: (a) high top clouds, (b) middle top clouds, and (c) low clouds. El Nino years have been excluded in the ISCCP data and CAM simulations. The CERES data refer to the two DJF seasons of 2001 and 2002. The difference between the black line and dashed line with solid circle for the CAM2 represent whole day sampling versus daytime sampling of calculating the ISCCP clouds in the simulator.

and CERES to correctly retrieve cloud tops since all these methods are based on single-layer clouds. This factor however is considered in the ISCCP simulator, which uses a simulated infrared brightness temperature to determine cloud pressure under a single-layer cloud assumption.

[9] The ISCCP D2 data are available from 1983 to 2001. The CERES cloud data are from the DJF seasons of 2001 and 2002, and JJA seasons of 2000 and 2001. Each DJF season refers to the months of January and February of the year and December of the previous year. The red lines in Figure 1a show the range of maximum and minimum high cloud frequencies in ISCCP from ten DJF seasons after excluding the El Nino years of 1983, 1987, 1990–1995, and 1998. The blue lines show the range of high cloud frequencies from the 2 CERES DJF seasons of 2001 and 2002. While the patterns of the latitudinal distributions are similar in the two data sets, high clouds are systematically less frequent in the CERES, especially at middle latitudes. Interannual variability of high cloud frequencies is generally smaller than the difference between the two satellite data sets. For the overlapping season of JJA in 2001 (not shown), the difference in the two data sets is similar to that in Figure 1a. This suggests that the main source of difference is from the cloud retrieving algorithms. This discrepancy is primarily due to differences in optically thin clouds. We have also examined high clouds from the High Resolution Infrared Radiometer Sounder (HIRS) [*Jin et al.*, 1996; *Wylie and Menzel*, 1999] and the adjusted surface-based observations [*Norris*, 1999], and found similar differences among all these data sets. While it is beyond the scope of this paper to explain the difference between the satellite data sets, we believe that the cloud detection algo-

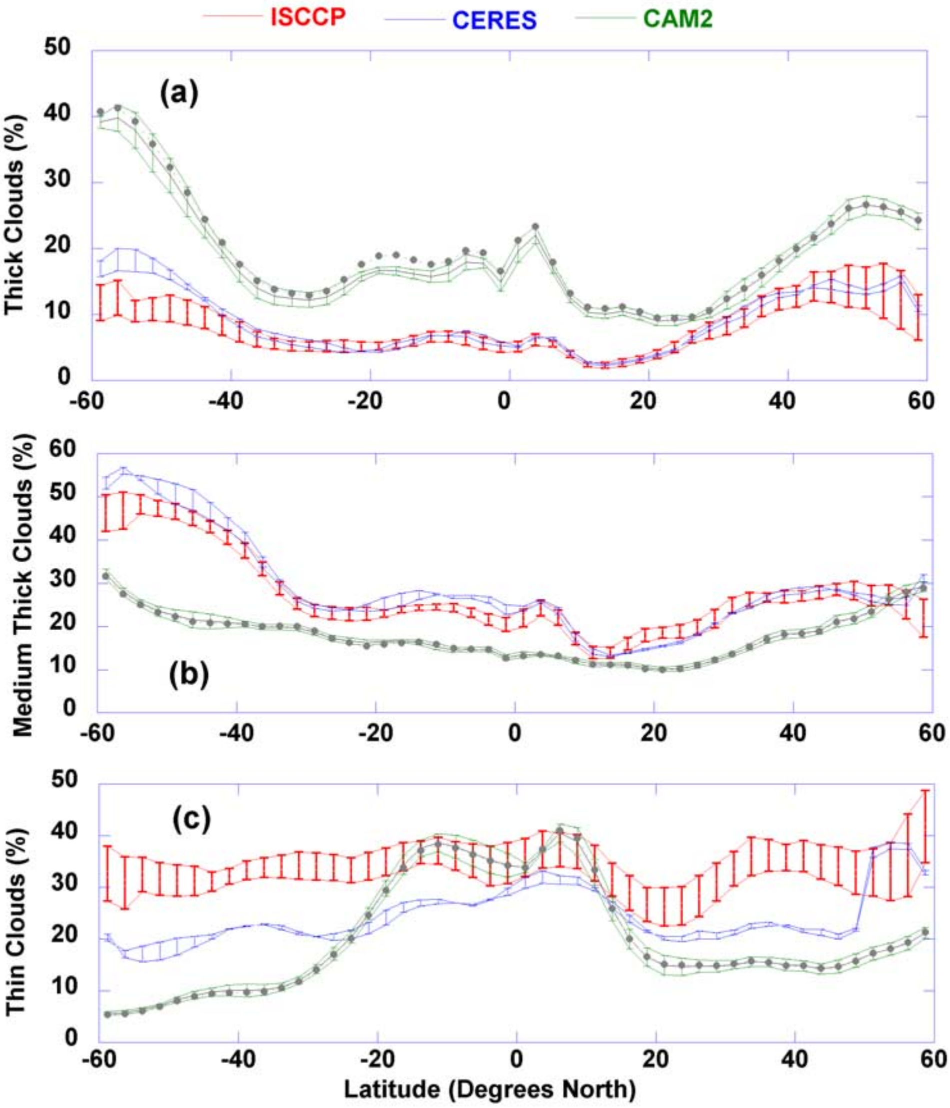


**Figure 2.** Same as Figure 1 except that clouds are stratified into ranges of optical thickness (a) for optical thick clouds, (b) for optically intermediate clouds, and (c) for optically thin clouds.

rithms, which are further related with the satellite pixel sizes, contributed significantly to these differences.

[10] Middle and low clouds from ISCCP and CERES agree better except at high latitudes (Figures 1b and 1c). Both data sets show increasing amount of middle and low clouds toward high latitudes, especially the southern high latitudes. This latitudinal variation is also seen in surface-based observations. It is therefore unlikely due to the masking of high clouds in the satellite product. Poleward of about 40° CERES reported fewer middle clouds, but more low clouds than ISCCP. These differences in middle and low clouds at high latitudes may be caused by several factors, including the specifications of radiative and micro-physical properties of liquid and ice particles, the temperature profiles, and the surface properties.

[11] When clouds are stratified against the optical thickness ranges, Figure 2 shows that optically thick and intermediate clouds in ISCCP and CERES agree well. The discrepancy between the two data sets is mainly in optically thin clouds (Figure 2c) with ISCCP reporting about 32% thin clouds versus 25% in CERES.

[12] Given the differences between the two data sets, in our comparison, if a model is statistically different from both data sets toward one direction, it is judged as biased. As an example, Figures 1 and 2 also included simulated clouds from the CAM2 as reported in the work of *Lin and Zhang* [2004]. For each cloud type, if the model result satisfies the following

$$C_m > \max(C_{ISCCP}, C_{CERES}) + \delta_{\max}$$

or

$$C_m < \min(C_{ISCCP}, C_{CERES}) - \delta_{\min}$$

where $C_m$, $C_{ISCCP}$, $C_{CERES}$ represent cloud frequencies from the model, ISCCP, and CERES, the model is judged to overestimate or underestimate clouds. $\delta_{\max}$ and $\delta_{\min}$ are the

95% confidence intervals based on a student *t* test. Standard deviations for the satellite data are from ten seasons of ISCCP data.

### 2.2. Models and Simulations

[13] Table 1 lists the participating models, along with their physical parameterization components related to clouds. Cloud fraction schemes can be categorized broadly into three groups: relative humidity-based schemes, statistical total water schemes, and prognostic cloud fraction schemes. In principle, the relative humidity cloud scheme can be also considered as a special case of the statistical scheme in which the probability distribution is implicitly assumed rather than diagnosed. All models used the maximum-random cloud overlapping assumptions in the vertical except for the GFDL model that will be described later.

[14] Two versions of the UKMO GCM are used. These are the HadAM3 and HadAM4. Multiple enhancements are made in HadAM4 relative to HadAM3, including the boundary layer scheme, precipitation scheme and the statistical cloud scheme. These are described in the work of *Webb et al.* [2001] and *Williams et al.* [2003]. Three versions of the NCAR CAM are used (CAM2, CAM2c, and CAM2x). CAM2c differs from CAM2 with only one change: the triggering condition of convection in the model. It is known that convection in the CAM2 is too frequent. An empirical convection triggering condition was imposed based on measurements of the large-scale atmospheric dynamics and SCM simulations at the ARM SGP site [*Xie and Zhang*, 2000]. This modification was implemented in the CAM2c by *Xie et al.* [2004] and was found to improve many aspects of the model climate. CAM2x is a developmental version of the CAM3. Main differences include the parameterization of clouds from shallow and deep convections, advection of condensates, and separation of cloud ice and cloud water in the CAM2x (W. D. Collins et al., The formulation and atmospheric simulation of the Community Atmosphere Model: CAM3, submitted to *J. Climate*, 2005, hereinafter referred to as Collins et al., submitted manuscript, 2005).

[15] All models used the ISCCP simulator described by *Klein and Jakob* [1999] and *Webb et al.* [2001] and updated at http://gcss-dime.giss.nasa.gov/simulator.html. Since model grids are much coarser than the ISCCP or CERES pixels, clouds at each model layer are first downscaled to cloudy and cloud-free subcolumns, along with liquid and ice water contents. The columns are then vertically aligned with overlapping assumptions consistent with the radiation routines in the models. Optical depth of each subcolumn is calculated based on the model radiation code. The optical depths are then used to calculate the emissivity, brightness temperature, and emissivity-adjusted cloud top pressure. The frequencies in different bins of cloud optical depths and cloud top pressures are used to compare with satellite measurements.

[16] The ISCCP simulator incorporates the satellite view of clouds and adjusts the physical cloud top to mimic satellite measurements. There are still issues in comparing its output with measurements. Chief among them is the cutoff value of the optical depth to define cloudiness. The satellite algorithms used a cutoff value of 0.02 as the lowest detectable limit. Because the satellites may not actually report clouds at this level, a cutoff optical depth value of 0.3 is used in the ISCCP simulator as a proxy (B. Rossow, personal communication, 2003). This choice is somewhat arbitrary and it affects the model cloud amount. To understand the sensitivity of ISCCP clouds with this cutoff value, Figure 3 shows the frequency differences of high thin clouds from the ISCCP simulator relative to a cutoff optical thickness value of 0.4 in the CAM2. High thin clouds are almost linear to the cutoff value in the optical thickness range of 0.1 to 0.4 in this model, with about 6% relative increase of cloud frequency per 0.1 optical thickness decrease. Below 0.1, the sensitivity is larger. The HIRS and ISCCP comparison by *Jin et al.* [1996] also implied large sensitivity of high thin clouds to the cutoff value of optical thickness in real observations. Figure 3 also includes the differences of middle and low thin clouds between the optical thickness of 0.01 and 0.4. There is very little sensitivity in these cloud types. We therefore have less confidence in optically thin clouds than optically thick to intermediate clouds in inferring model biases.

[17] Other issues about the ISCCP simulator include the possibility of subpixel clouds that the satellite sees as optically thin clouds, assumptions on the vertical overlapping of the subcolumns, and the distribution of condensates among the subcolumns. These are expected to be on the secondary order. Two versions of the GFDL model are used to highlight the impact of subgrid-scale cloud structure on the ISCCP simulator results. One version (GFDL0) uses the random overlap assumption, and a constant value of condensate in each layer for every subcolumn in the ISCCP simulator. The second version (GFDL) infers internal inhomogeneity in each layer by fitting a symmetric beta distribution of total water with fixed shape parameters to the model values of cloud fraction and mean condensate. Cloud overlapping varies smoothly between maximum and random as the distance between adjacent layer increases [*Hogan and Illingworth*, 2000; R. Pincus et al., Overlap assumptions for assumed-PDF cloud schemes in large-scale models, submitted to *Journal of Geophysical Research*, 2004]. Since this structure is used in both the ISCCP simulator and the radiation calculations [*Pincus et al.*, 2003], the model produces slightly different clouds in the two simulations. The second version is more physically based in terms of both its overlapping assumption and its subgrid-scale inhomogeneity.

[18] Figure 4 shows the cloud frequencies produced by the different implementations of the ISCCP simulator in these two simulations relative to ISCCP and CERES data, averaged equatorward of 60°, for the nine ISCCP cloud types in DJF. The second version simulated slightly more optically thin clouds but less optically thick clouds. This is likely due to the impact of the consideration of subgrid-scale inhomogeneity. This version also simulated less low clouds. This is consistent with the impact of the new overlapping scheme relative to the random assumption. None of these differences in the two calculations modifies the nature of the model biases relative to the two satellite measurements, and so the second version is used in the rest of this paper. For all other models, including those that employ statistical distribution of cloud liquid, the ISCCP simulator only assigns constant condensates to the subcolumns.

t1.1 **Table 1.** Description of Physical Parameterization Schemes in the Models

| t1.2 | | Stratiform Clouds | Convective Clouds | Convection | PBL | Cloud Microphysical and Precipitation | Resolution |
|---|---|---|---|---|---|---|---|
| t1.3 | CAM2 | RH-based; *Kiehl et al.* [1996] | diagnostic; *Rasch and Kristjánsson* [1998] | *Hack* [1994]; mass flux; *Zhang and McFarlane* [1995] | first-order nonlocal; *Holtslag and Boville* [1993] | *Rasch and Kristjánsson* [1998] | T42L26 |
| t1.4 | CAM2c | as above | as above | as above, modified with *Xie et al.* [2004] | as above | as above | as above |
| t1.5 | CAM2x | as above with modifications in the work of Collins et al. (submitted manuscript, 2005) | diagnostic amount; Collins et al. (submitted manuscript, 2005) | as above | as above | as above with modifications; Collins et al. (submitted manuscript, 2005) | as above |
| t1.6 | GFDL | prognostic; *Tiedtke* [1993]; *Geophysical Fluid Dynamics Laboratory Global Atmospheric Model Development Team* (*GFDL GAMDT*) [2004] | prognostic; *Tiedtke* [1993]; *GFDL GAMDT* [2004] | RAS; *Moorthi and Suarez* [1992] | cloud entrainments; *Lock et al.* [2000]; *GFDL GAMDT* [2004] | *Rotstayn* [1997]; *GFDL GAMDT* [2004] | 2.5 × 2.0 L24 |
| t1.7 | GISS | RH based, Sundqvist type; *Del Genio et al.* [2005] | diagnostic; *Del Genio et al.* [2005] | mass flux; *Del Genio and Yao* [1993] | second-order; *Cheng et al.* [2002] | *Del Genio et al.* [2005] | 4 × 5 L12 |
| t1.8 | GSFC | RH based, Sundqvist type; *Del Genio et al.* [1996] | diagnostic; *Del Genio et al.* [1996] | RAS; *Moorthi and Suarez* [1992] | 2.5 order; *Helfand and Labraga* [1988] | *Del Genio et al.* [1996]; *Sud and Walker* [1999] | 2.5 × 2 L40 |
| t1.9 | HadAM3 | statistical; *Smith* [1990] | diagnostic; *Gregory and Rowntree* [1990] | mass flux; *Gregory and Rowntree* [1990]; *Gregory and Allen* [1991] | first-order; *Smith* [1990]; *Pope et al.* [2000] | *Smith* [1990] | 3.75 × 2.5L19 |
| t1.10 | HadAM4 | as above with modifications of *Webb et al.* [2001], *Cusack et al.* [1999] | as above with modifications *Gregory* [1999] | mass flux; *Gregory and Rowntree* [1990]; *Gregory and Allen* [1991] | first-order with cloud entrainment; *Lock et al.* [2000]; *Martin et al.* [2000] | *Wilson and Ballard* [1999] | 3.75 × 2.5L30 |
| t1.11 | ECHAM5 | prognostic; *Tompkins*, [2002] | diagnostic; *Roeckner et al.* [1996] | mass flux; *Tiedtke* [1989]; *Nordeng* [1994] | first-order; *Brinkop and Roeckner* [1995] | *Lohmann and Roeckner* [1996] | T42L19 |
| t1.12 | LMD | statistical; *Le Treut and Li* [1991] | statistical; *Bony and Emanuel* [2001] | *Emanuel* [1991] | first-order *Li* [1999] | *Le Treut and Li* [1991] | 3.75 × 2.5L19 |

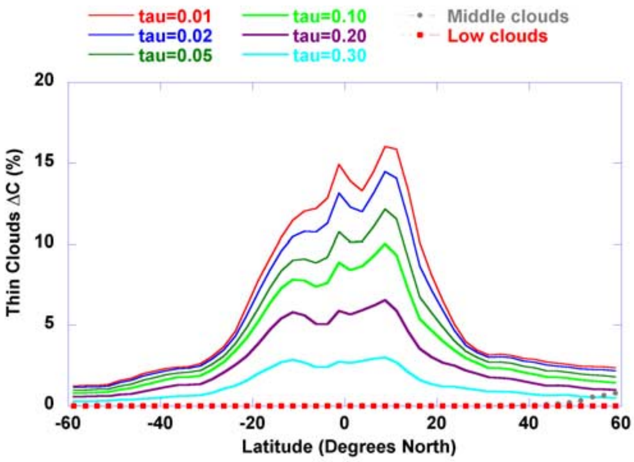


**Figure 3.** Frequency differences of high thin clouds from the ISCCP simulator with different optical cutoff values of optical thickness relative to the cutoff optical thickness value of 0.4 in the CAM2. For middle and low thin clouds, the differences are between the optical thickness of 0.01 and 0.4.

[19] Since the calculation of optical depth in the models does not explicitly involve sunlight, the ISCCP simulator can calculate cloud types during nighttime, while the ISCCP cloud data are for daytime only. In Figures 1 and 2, simulated clouds from the whole day and daytime only calculations in the CAM2 are shown as the black line and the dashed line with solid circles. While diurnal variations of clouds are seen, their magnitudes are much smaller than the signatures sought in this study. Clouds are from daytime-only simulations for most models.

[20] Some participating models carried out AMIP long simulations. Others only carried out one-year simulations. Ranges of interannual variabilities of model clouds from 10 DJF seasons in CAM2 are shown in Figures 1 and 2 as the green lines. The interannual variations in the model are smaller than those in ISCCP. When statistical tests are carried out, we use the ISCCP standard deviations as an upper bound estimates for the models. All model results are from 1-year integrations spun off from an AMIP type simulation with prescribed monthly sea surface temperature (SST). Most models used SST for the year 2000. The ISCCP cloud data are also from year 2000. The CERES data are averages for the two winter and summer seasons of 2001 and 2002, respectively.

## 3. Results

[21] Frequencies of high clouds from ISCCP, CERES and the models for the DJF season are shown in Figure 5a. Most models are able to simulate the observed cloud maximum in the tropics and minimum in the northern subtropics. The simulated frequencies however differ greatly among the models even though the same cutoff values of the optical thickness (0.3) are used in the ISCCP simulators. There is a four-fold difference in the simulated high clouds among the models, ranging from 11% in the GISS GCM to 44% in the GSFC when averaged from 60°S to 60°N. The ISCCP and CERES measurements are 23% and 15% respectively with an estimate of the interannual standard deviation of 1.5%. In the tropics from 30°N to 30°S, the GSFC model and the CAM series overestimated high clouds above the 95% confidence level; the GISS and HadAM3 underestimated high clouds. In middle latitudes poleward of 30°, GSFC, CAM2, CAM2c, and ECHAM5 overestimated high clouds, while GISS underestimated high clouds. High clouds from half of the models are within the range of the two observational data sets. This is due to the large spread in the two data sets. Along with uncertainties related to the cutoff value of optical thickness for high thin clouds, the accuracy of currently available satellite high clouds is therefore still not sufficient to effectively constrain the models.

[22] While half of the models simulated high clouds within the range of the two satellite data sets, the situation is worse for middle and low clouds. Figure 5b shows that most models substantially underestimated middle clouds. Nine out of the ten models significantly underestimated middle clouds in the tropics, and eight models did so in middle latitudes. The grand mean of middle clouds from all models is only one third of the satellite measurements in the tropics and one half in middle latitudes. The GFDL model is the only exception to simulate middle clouds in close agreement with the ISCCP and CERES data at all latitudes, while the GISS model did well at middle latitudes. Because of the satellite view, middle clouds may be affected by shielding from high clouds. We have examined the middle cloud distribution in the CAM2 directly from the model output without this shielding effect and they are still significantly less than the satellite observation.

[23] Figure 5c compares the low clouds with satellite measurements. In the tropics from 30°S to 30°N, eight out of the ten models underestimated low clouds at the 95% significant level. Poleward of 30°, half of the models underestimated low clouds at the statistically significant level. None of the models overestimated low clouds. The grand mean of model low clouds is about 70 and 80% of satellite observations in the tropics and middle latitudes respectively. This underestimation is probably a lower limit because there is less shielding of low clouds by middle clouds in the models. We have also examined the surface-based low clouds adjusted to satellite view, courtesy of Joel Norris. In low latitudes, the surface-based low clouds are more than the two satellite data sets. In middle latitudes, they are less, but are still more than those in half of the models.

[24] Compensations of model clouds at different heights are seen in some models. The GISS model simulated the smallest amount of high clouds, but it had the largest amount of low and middle clouds that are in close agreement with the satellite data. The GSFC model simulated the largest amount of high clouds, but it was among the models simulating the smallest amount of middle clouds. The HadAM3 and the LMD, on the other hand, were at the lower end of the spectrum of simulating clouds at all altitudes. Because of the use of the ISCCP simulator, a model with excessive high cloud cover tends to have fewer low clouds because of masking by high clouds. Examination of the geographical distribution of different cloud types however does not suggest this shielding effect to be the leading cause of these compensations in the models.

[25] In summary, models simulated a four-fold difference in high clouds. The available observations however also

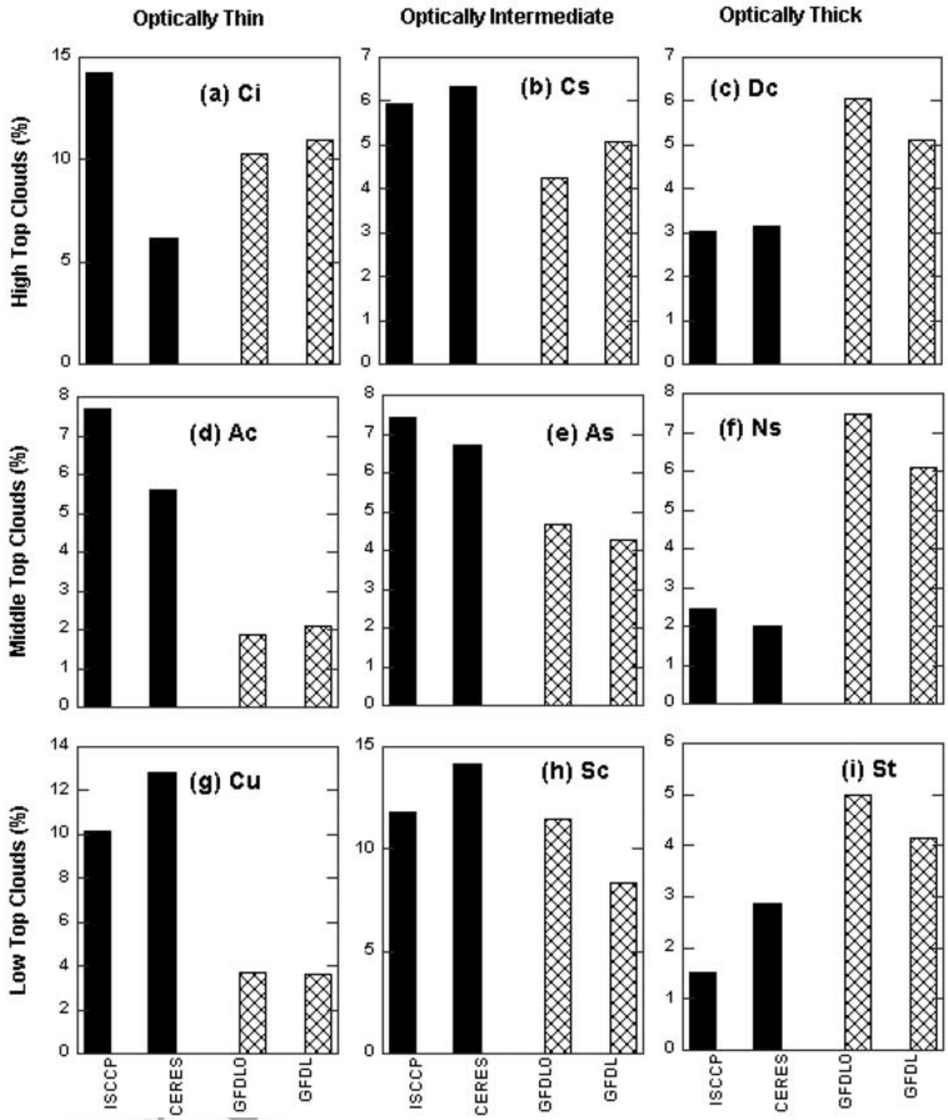


**Figure 4.** Cloud frequency averaged from 60°N to 60°S in the DJF season for the nine ISCCP cloud types in ISCCP, CERES, and two versions of the ISCCP simulator in the GFDL model. See text for details.

show large spread. Models are found to significantly underestimate middle and low clouds. Averaged over 60°S to 60°N, the models simulated a grand mean of 6.9% middle clouds, in contrast to 17.5% in ISCCP and 14.3% in CERES. This is only 40% of the smaller satellite value from CERES. The models simulated a grand mean of 20.2% low clouds, in contrast to 23.4% and 29.9% in ISCCP and CERES measurements.

[26] We next examine the collective impact of these clouds on the cloud radiative forcing (CRF) at the top of the atmosphere (TOA). Figure 6 shows the longwave, shortwave, and net CRFs at the TOA, averaged from 60°S to 60°N for the DJF season from the ISCCP FD product [*Zhang et al.*, 1995], ERBE, CERES, and all models. Standard deviations of interannual variation in the ISCCP FD, ERBE, the CAM2 for non-El Nino years, and the range of the two seasons in CERES, are also shown in Figure 6 as error bars. Mean values from ERBE and CERES are also drawn as dashed horizontal lines. The ERBE data are from the monthly S-4 product that combined measurements from ERBS, NOAA-9 and NOAA-10 [*Harrison et al.*, 1990]. The CERES data are from its ERBE-like Monthly Regional Averages (ES-9) product [*Wielicki et al.*, 1996]. Both data sets were acquired from the NASA Langley Distributed Active Archive Center. The magnitudes of cloud forcing in CERES are smaller than in ERBE. Similar to the cloud products, we do not know the exact reasons of this difference. We thus use the same statistical procedure to judge the model biases by using the interannual variations from the ERBE in the $t$ test. Only the GSFC model significantly overestimated longwave cloud forcing, which is consistent with the overestimation of high clouds in this model (Figure 6a). Seven models underestimated the longwave cloud forcing. This is consistent with the underestimation of middle and low clouds in the models. CAM2 and CAM2x simulated longwave cloud forcing that falls

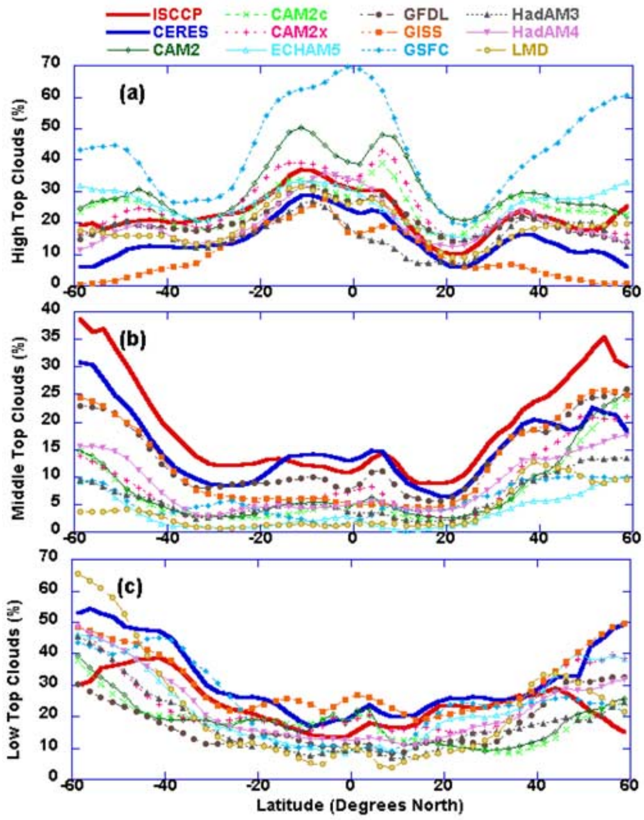


**Figure 5.** (a) High top clouds, (b) middle top clouds, and (c) low top clouds in the DJF from satellite measurements and from the models. ISCCP data are from year 2000, CERES data are from two seasons of 2001 and 2002. Model results are from one year simulations with most of them forced with prescribed monthly sea-surface temperature of year 2000.

within the range of the ERBE and CERES measurements since these models overestimated high clouds but with deficient middle and low clouds. The GISS GCM simulated the smallest longwave cloud forcing, consistent with underestimation of high clouds in this model. Overall, the underestimations of middle and low clouds in the models showed signatures in the longwave cloud forcing except for those that have overestimated high clouds.

[27] Since middle and low clouds should have a large impact on the shortwave CRF at the TOA, one would expect most models to also underestimate the magnitude of shortwave cloud forcing. This is, however, not seen in Figure 6b. Eight models simulated shortwave cloud forcing that is within the range of the two observational data sets. Two models (CAM2x and GISS) even overestimated the magnitude of shortwave cloud forcing. As a result, half of the models simulated the net cooling of clouds within the observational range, and the other half overestimated the net cooling effect (Figure 6c). This is consistent with *Potter and Cess* [2004] who demonstrated negative biases of the net CRF in many GCMs.

[28] The biases in the shortwave cloud forcing can only be explained by compensatory changes in cloud types. The ISCCP simulator allows us to compare model clouds with observation in optical thickness ranges. For optically thick clouds (Figure 7a), eight models are found to overestimate them at all latitudes. The GSFC and LMD models are the exceptions. The grand mean of all models in the tropics is 12.4%, more than double the satellite measurements of 4.7% in ISCCP and 5.1% in CERES. In middle latitudes, the simulated mean of optically thick cloud amount of 19.5% is also about twice of the observed values of 10% in ISCCP and 12% in CERES. The HadAM3 simulated the largest amount of optically thick clouds. This compensates for the small amount of middle and low clouds to explain its shortwave cloud forcing that was close to HadAM4 simulations.

[29] Figure 7b shows the comparison of optically intermediate clouds. The GSFC model is an outlier that simulated significantly more than observations. Except for the GISS GCM, all eight other models underestimated this type of clouds at the 95% statistically significance level. The grand average of simulated cloud amount in the tropics, including the GSFC values, is 16.2% in contrast to observations of 20.0% in ISCCP and 21.8% in CERES. In middle

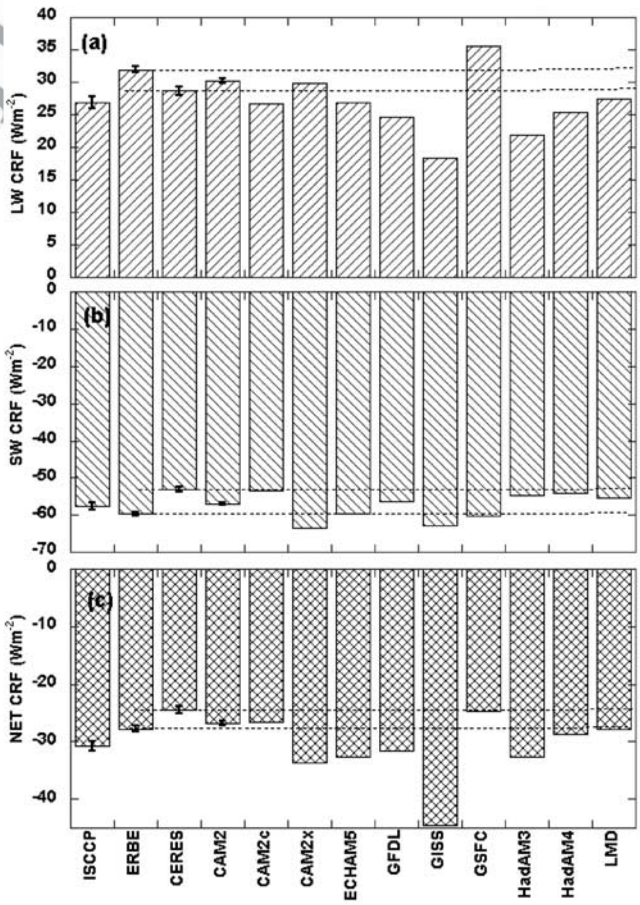


**Figure 6.** DJF cloud radiative forcing at the TOA averaged from 60°N to 60°S from measurements and from the models: (a) longwave CRF, (b) shortwave CRF, and (c) net CRF. The error bars for ISCCP FD, ERBE, CAM2 are their interannual standard deviations excluding the El Nino years. The error bar for CERES is from the two DJF seasons of 2001 and 2002. The ERBE and CERES values are also drawn as horizontal lines.

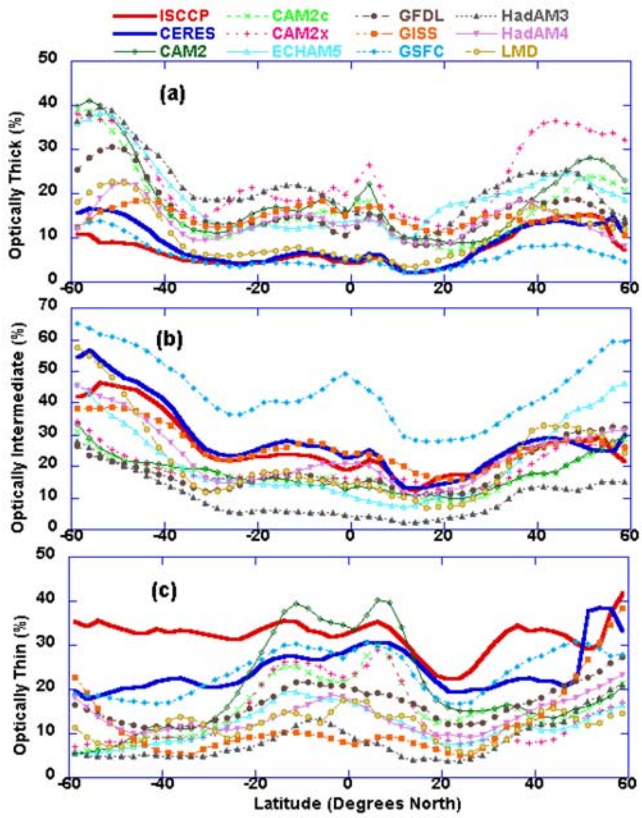


**Figure 7.** (a) Optically thick clouds, (b) optically intermediate clouds, and (c) optically thin clouds in the DJF from satellite measurements and from the models.

latitudes, all models except the GSFC and GISS models underestimated this cloud type, even though some models were able to simulate better in one of the hemispheres. The grand average of simulated amount in middle latitudes is 27.0% versus 32.2% in ISCCP and 34.7% in CERES. The HadAM3 simulated about one third of the satellite observations.

[30] Comparison of the optically thin clouds is shown in Figure 7c. Except for the CAM2 and the GSFC models, all other models significantly underestimated thin clouds. The models simulated about 60% of the smaller satellite values from CERES. These underestimations are mostly contributed by middle and low thin clouds, which are not very sensitive to the cutoff optical depth in the ISCCP simulators. The HadAM3 simulated one third of the CERES value.

[31] We have therefore seen systematic cloud biases in the models as follow: In the altitude ranges, models systematically underestimate middle and low clouds. In the optical thickness range, they overestimate optically thick clouds and underestimate optically thin and intermediate clouds. These biases can be more clearly seen in Figure 8 where the nine ISCCP cloud types are averaged from 60°N to 60°S. The same statistical procedure with error bars from ISCCP is used to judge the models against both satellite data sets.

[32] For high thin clouds (Figure 8a), the two satellite data sets show the largest differences, with CERES giving less than half of the ISCCP value. Eight models simulated high thin clouds within the two data sets. The two exceptions are the CAM2 that simulated excessive high thin clouds and the GISS GCM that simulated too few of them. The data sets therefore cannot effectively constrain the models. Moreover, as noted in the previous section, the cutoff value of the optical thickness used in the ISCCP simulator has the largest impact on this cloud type. It is difficult to draw firm conclusions from the comparison of this cloud type with observations.

[33] For all other cloud types, the spread between the two data sets is smaller. The impact of the optical thickness cutoff value is also small or absent. For high clouds with intermediate thickness, the GSFC model had significantly more clouds than those in the other models and in the two observations. The majority of models only slightly underestimated this cloud type, but the HadAM3 simulated one third of the observed values. For high thick clouds, eight of the ten models had significantly positive biases. The grand average of high thick clouds from the models is 6.1%, about twice the ISCCP value of 3.0% and the CERES value of 3.2%. The two exceptions are the GISS and the GSFC models that simulated the correct amount of this cloud type. The GISS model simulated small amount of high clouds in all optical ranges, while the GSFC simulated more than twice high clouds with intermediate optical depth.

[34] Middle clouds with thin and intermediate optical depths are both significantly underestimated in all models. The grand average of simulated middle thin clouds is only 15% of ISCCP value and 20% of the CERES value. The grand average of middle intermediate clouds is only 40% of the two satellite measurements. For middle thick clouds, seven models overestimated this cloud type, with the GFDL simulated three times more than observations. Three models underestimated this cloud type, with the GSFC model simulated about one quarter of the observed values.

[35] All models also underestimated low thin clouds, with the grand mean only about 30% of the satellite measurements. Because low thin clouds may be subpixel to satellite measurements, they are better combined with low intermediate clouds as a single cloud type to minimize the mismatch between observation and the ISCCP simulator results. The grand mean of this combined type in all models is only 55% of the ISCCP value of 21.8% and 65% of the CERES value of 27.1%. On the other hand, eight models significantly overestimated low top optically thick clouds, with the grand mean two to three times of the satellite measurements. The LMD model simulated this cloud type comparable to measurements, while the GSFC model is the only one that significantly underestimated this cloud type.

[36] To summarize the above results for systematic model biases, we can categorize the nine ISCCP cloud types into three groups. In one group, we combine the four ISCCP cloud types with middle and low tops, thin and intermediate optical depths. All models significantly underestimated clouds in this group. The grand mean of all models is about half of both the ISCCP and CERES measurements of 41% and 43% respectively. In the second group, we combine the three optical thick clouds of all altitudes. The majority of the models significantly overestimated this group of clouds. The grand mean of these cloud types in all models is 15.4%, while the two satellite values are 6.9% and 8.1% respectively. The only exception is the GSFC model that simulated less than half of the low and middle top optically thick

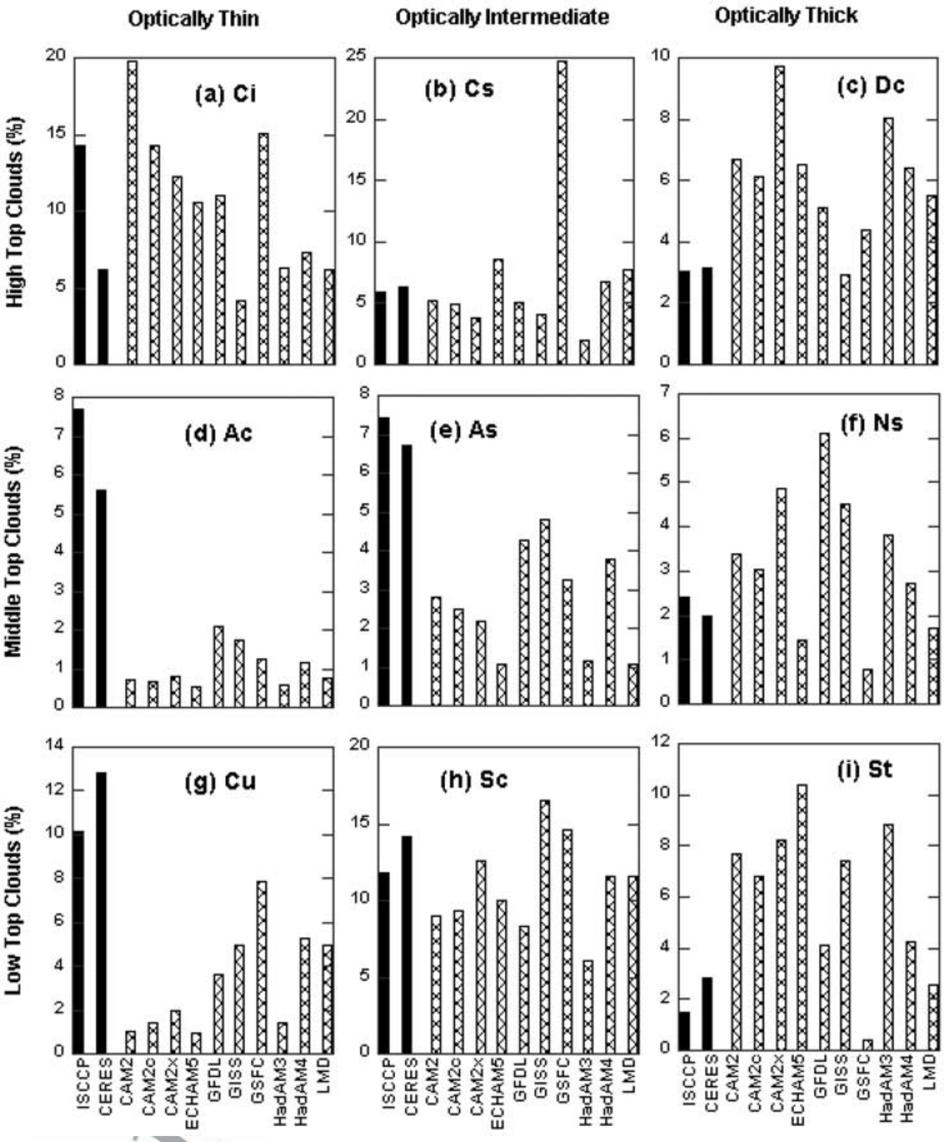


**Figure 8.** Cloud frequency averaged from 60°N to 60°S in the DJF season for the nine ISCCP cloud types in satellite measurements and in the models.

clouds. The systematic overestimation of optically thick clouds is more pronounced for high and low top thick clouds than for middle top thick clouds, suggesting possibly multiple sources of model errors. The third group of clouds consists of high thin clouds. The spread in the two satellite data sets is too large to constrain the models.

[37] It is difficult to completely rule out the possibility that problems in the observations and the ISCCP simulator caused the systematic differences between models and data. In particular, it has been argued that middle clouds in ISCCP may be exaggerated by the misrepresentation of cirrus over low clouds, which ISCCP simulator may not completely capture because of possible sampling differences. We believe however that the above results reveal true physical deficiencies of clouds in the models. As an example, Figures 9a and 9b show the visible and infrared cloud images, respectively, from GOES east associated with an Atlantic cyclone on 19 February 2004 at 1500 UTC. Even though the optical properties of these clouds are not available, the classic high-altitude comma-shaped frontal cloud band and the low shallow cumulus and stratocumulus after the cold front can be clearly identified from the images. Figure 9c shows the 15-hour forecast simulation of visible clouds as measured by TOA albedo from the CAM2, which was initialized with the NCEP operational analysis. It is seen that, consistent with the systematic models biases discussed above, the CAM2 missed a considerable amount of low and middle clouds behind the front. Figure 9d shows a forecast simulation of clouds from using the Weather Research and Forecasts (WRF) mesoscale model. Although the frontal cloud structure is improved in the mesoscale mode, it still missed the low clouds behind the cold front.

## 4. Seasonal Variations of Clouds

[38] Figure 10 shows the seasonal variations of clouds from DJF to JJA for high, middle and low clouds in ISCCP,

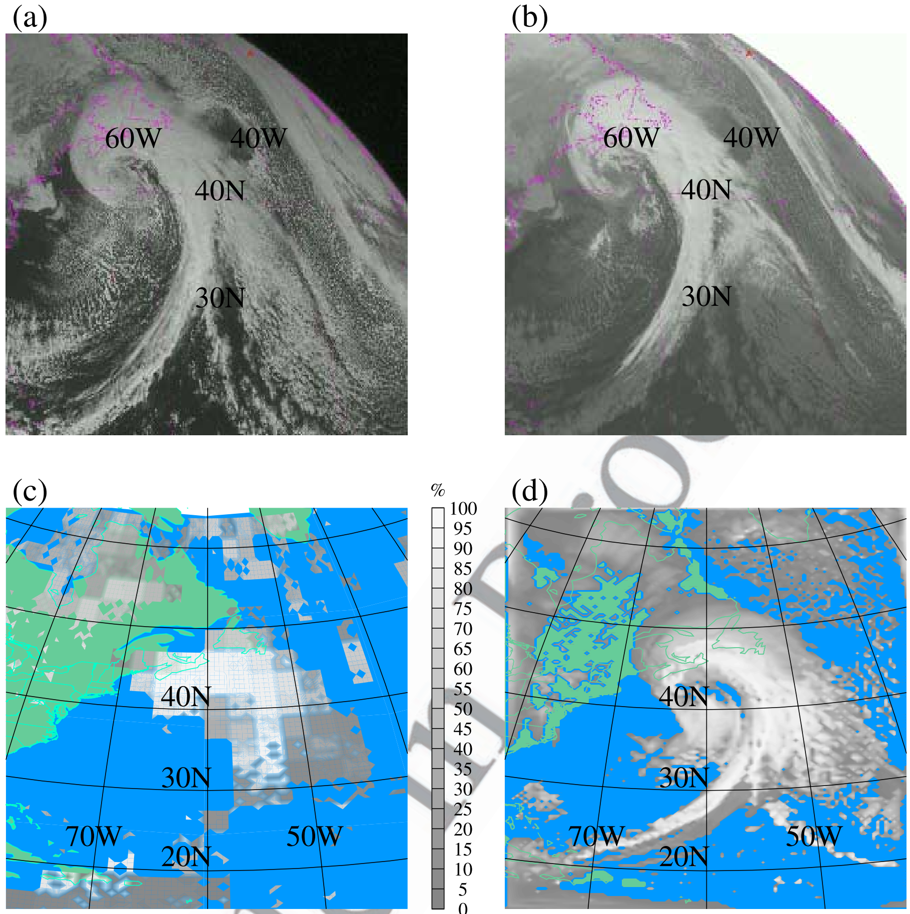


**Figure 9.** (a) Visible and (b) infrared cloud images from GOES east at 1500 UTC on 19 February 2004. Simulated TOA albedo from 15-hour forecasts with NCEP operational initial conditions in (c) CAM2 and (d) WRF.

CERES and the models. Seasonal variation is defined as the difference of cloud frequencies between the JJA and the DJF seasons. In observations, the dominant pattern of seasonal high cloud variation is the movement of the ITCZ. There is little seasonal variation of high clouds poleward of 30°. The seasonal variation of middle clouds has a similar ITCZ pattern in the tropics, but it shows summertime reductions at middle latitudes, indicating the impact of wintertime middle latitude storm track clouds. This seasonal variability is more pronounced in the Northern Hemisphere than in the Southern Hemisphere. North of 50°, the two satellite data sets diverge significantly from each other for middle clouds, with the CERES data indicating summertime increase. Top view middle clouds from HIRS also show summertime reduction that is more like ISCCP. For low clouds, the two satellite data sets show reduction of clouds in the summer. This could be partly related with shielding of low clouds by high and middle clouds that maximize in the summer. The two satellite data sets again diverge from each other north of 50°, and the ISCCP seasonal variation is more consistent with both surface-based and HIRS measurements at these latitudes (figures not shown). We therefore use the ISCCP seasonal variation as a benchmark when quantitative evaluations are carried out.

[39] Most models were able to simulate the seasonal variation of high clouds (Figure 10a). The correlations with ISCCP and CERES are all above 0.9. *Weare and AMIP Modeling Groups* [1996] showed that nearly all AMIP I models had tropical peaks in seasonal variability that were poleward of observations. Figure 10a shows that only the CAM2, CAM2x and the GSFC model still have this tendency. The GISS model simulated a tropics-like variation at middle latitudes. To facilitate the discussion, we define seasonal amplitude as the area weighted root mean square of the seasonal variation. The GISS GCM simulated the smallest amplitude, 81% of the ISCCP measurement. The GSFC model simulated the largest seasonal amplitude, 180% of the satellite data.

[40] For middle clouds, most models simulated summertime reduction in middle latitudes, but they tend to simulate little seasonal variability in the tropics (Figure 10b). The correlations of the seasonal changes with observations are in the range of 0.6 to 0.9. The ECHAM5 simulated the smallest seasonal variations, with amplitudes about 50% the ISCCP measurement, while the GISS model had the largest seasonal amplitude, about 140% of the ISCCP value.

[41] The ability of the models in simulating low clouds is poorer (Figure 10c). Correlations of simulated low cloud variations with ISCCP values are from −0.2 in the CAM2 to 0.7 in ECHAM5. The majority of the models had correlations from 0.2 to 0.5. The RMS differences between the models and ISCCP are as large as the observed seasonal

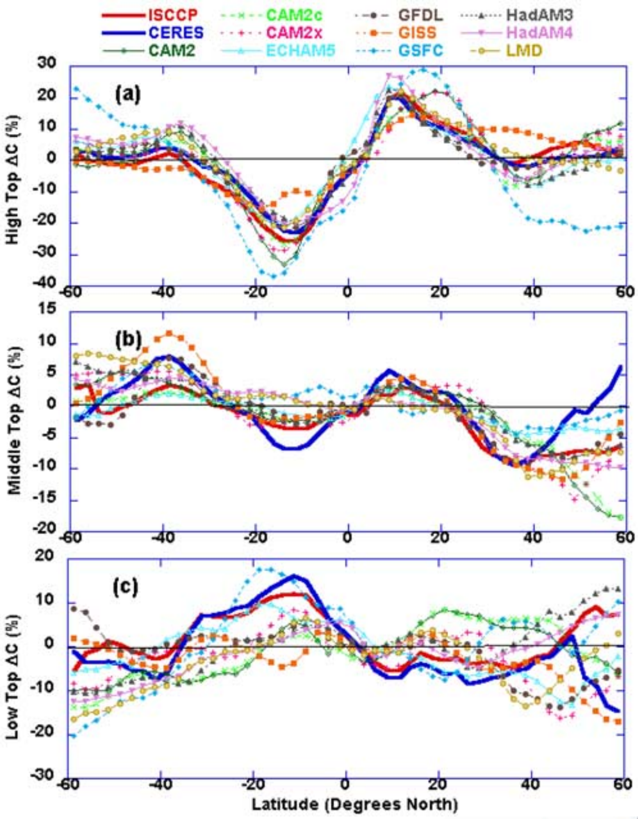


**Figure 10.** Seasonal variations (JJA minus DJF) of clouds from measurements and the models: (a) high top clouds, (b) middle top clouds, and (c) low top clouds.

amplitudes. These results are similar to what were found in the work of *Weare and AMIP Modeling Groups* [1996] about 10 years ago. While satellite-view low clouds are necessarily affected by middle and high clouds and may contain larger uncertainties than high and middle clouds, after examining the geographical distributions of seasonal variations of low clouds, we tend to conclude that the poorer quality of low cloud seasonal variation in the models is not mainly caused by the shielding effects of middle and high clouds.

[42] We next examine the individual ISCCP cloud types to search for the first-order controlling factors of the seasonal cloud variations in the models. Figures 11a and 11b show the seasonal variations of high clouds with intermediate and thick optical depth. The two satellite data sets agree well with each other. The seasonal amplitudes of high intermediate clouds in the models differ by several folds, with the HadAM3 showing the smallest amplitude of 30% of the ISCCP value, and the GSFC model showing the largest amplitude of four times the ISCCP measurement. Large differences are also seen for high thick clouds. The GSFC model had the smallest amplitude of 70% of the satellite data, while the two HadAM models showed the largest seasonal amplitudes of about twice the ISCCP and CERES values. Figures 11c and 11d relate the magnitudes of the seasonal amplitudes to the mean annual cloud frequencies for these two cloud types. The solid circle inside a square denotes the ISCCP data, and that inside a triangle represents the CERES data. Models that simulated large annual frequencies also had greater seasonal variation, and vice versa. The linear correlation between the seasonal amplitudes and the basic cloud amounts is 0.85 for high intermediate clouds and 0.7 for high thick clouds. For high thin clouds, the relationship (not shown) is similar to those shown in Figures 11c and 11d among the models.

[43] A different type of relationships between the seasonal variation and the basic distribution of clouds can be seen in Figures 12a and 12c for middle thin clouds. Figure 12a shows that the models produced very little seasonal changes in this cloud type in the tropics and subtropics. This can be explained by the basic annual cloud frequency in the models shown in Figure 12c. All models simulated very little middle thin clouds between 40°N and 40°S. Therefore there is also little seasonal change and intermodel variability. A contrasting case is shown in Figures 12b and 12d for low thick clouds. The seasonal variations of low thick clouds in many models are substantially larger than in the observations. With the exception of the GSFC and the LMD models, all models simulated more than twice the seasonal amplitudes of the ISCCP value. The ECHAM5 amplitude is six times the satellite data, followed by HadAM3 and CAM2x with amplitudes 5 and 4 times of the ISCCP data. This exaggerated variation in most models is clearly related to the basic cloud distributions shown in Figure 12d since most models substantially overestimated the mean frequency of this cloud type. The GSFC model had the smallest amount of low thick clouds, and thus the smallest seasonal variation. The ECHAM5 generated the greatest amount of this cloud type and had the largest amount of seasonal variation. The LMD model produced the best climatology of low thick clouds and it simulated the best seasonal cycle of this cloud type relative to ISCCP and CERES.

[44] The cloud biases illustrated in the previous section thus have direct relevance to the sensitivities of clouds in models. Even though our results do not invalidate cloud feedbacks and climate sensitivity results from the models, the fact that cloud feedback uncertainties have not been reduced appreciably in the last 15 years [*Cubasch et al.*, 2001; *Bony et al.*, 2004] suggests the need to improve the model clouds beyond what have been done.

## 5. Summary and Discussion

[45] We have used ISCCP simulators in ten GCMs to compare with satellite cloud analysis from ISCCP and CERES. We have shown a four-fold difference in high clouds among the models. There is also a large difference in high thin clouds between the satellite data sets and a large sensitivity of high thin clouds to the cutoff value of optical thickness. The available satellite data are therefore not accurate enough to constrain high thin clouds in most models. We have also shown that the majority of models only simulated thirty to 40% of the observed middle clouds. Some models only simulated less than a quarter of observed middle clouds. For low clouds, half of the models underestimated them while none overestimated them at the statistically significant level. The grand mean of low clouds from all models is about 70–80% of observations. When stratified in the optical thickness ranges, the majority of the models simulated optically thick clouds more than twice the

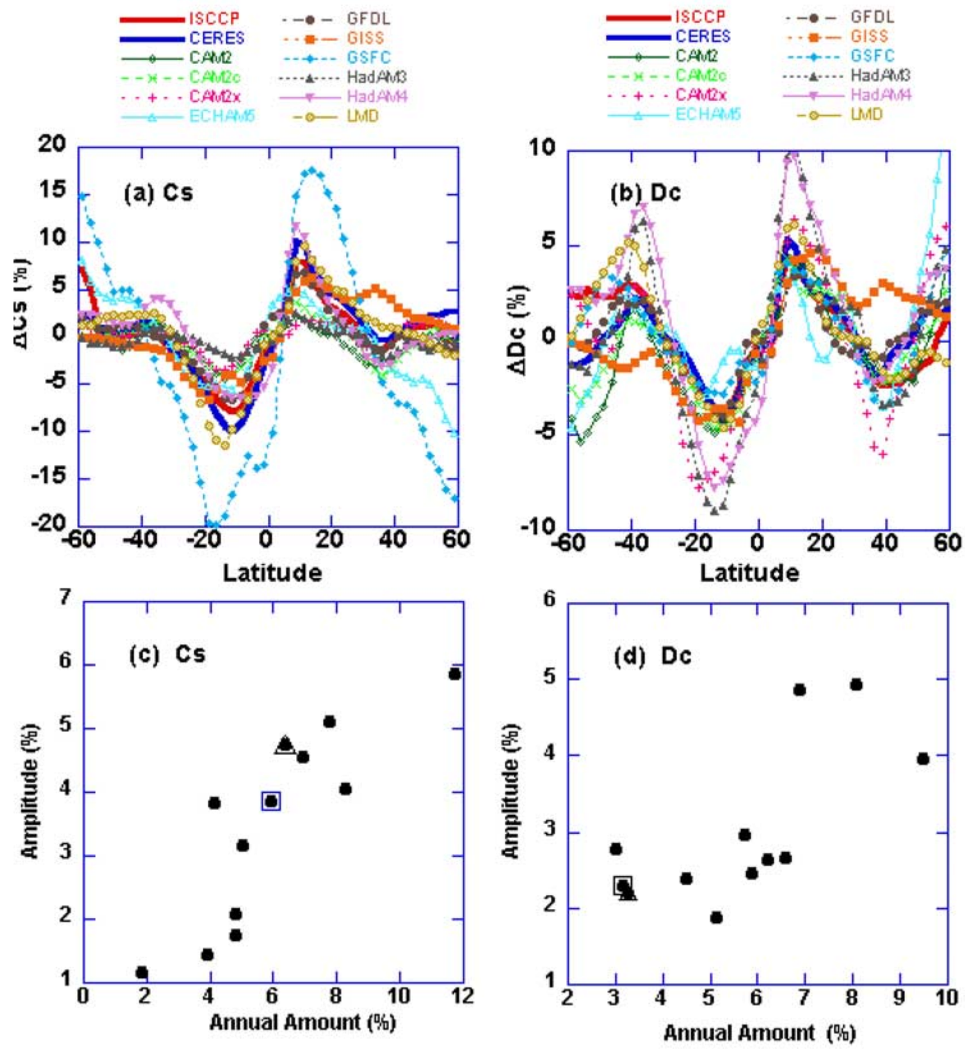


**Figure 11.** Seasonal variations of high top clouds with (a) intermediate and (b) thick optical depths. Relationships between the seasonal amplitudes averaged from 60°N to 60°S with the mean cloud amount for high-top clouds with (c) intermediate and (d) thick. To fit the GSFC model in Figure 11c that is represented by the data point in the upper right corner, the magnitudes of the amplitude and basic frequency are all scaled by half.

satellite observations. Most models however underestimated optically intermediate and optically thin clouds. The grand mean of all models simulated about 80% of optical intermediate clouds and 60% of optically thin clouds. The underestimations of middle and low clouds are related with the negative biases of TOA longwave cloud forcing in some models, while the overestimations of optically thick clouds explains the reasonable or excessive shortwave cloud forcing in the models. These results further quantify the model cloud biases reported in the work of *Weare and AMIP Modeling Groups* [1996]. They also explain the *Weare* [2004] result that models tend to simulate the albedo moderately well, but not the cloud water path.

[46] We have categorized the nine ISCCP cloud types into three groups to describe the systematic model biases. The first group consists of middle and low clouds with intermediate and thin optical thickness. Models underestimate this group of clouds. The grand mean of all model results is about half of both the ISCCP and CERES measurements of 41% and 43% respectively. The second group consists of the three optical thick clouds of all altitudes. The majority of the models significantly overestimated this group of clouds. The grand mean of these cloud types in all models is twice the two satellite measurements of 6.9% and 8.1% respectively. The third group consists of thin cirrus for which the models show a several fold difference but they cannot be accurately constrained by the available satellite data.

[47] We also presented seasonal sensitivities of clouds. Models are shown to simulate latitudinal distributions of seasonal variations that correlate with satellite measurements at >0.9, 0.6–0.9, and −0.2 to 0.7 respectively for high, middle and low clouds. For individual ISCCP cloud types, the differences of seasonal amplitudes among the models and satellite measurements can reach several hundred percent. The dominant factor that determines the seasonal amplitude of a particular cloud type in the models

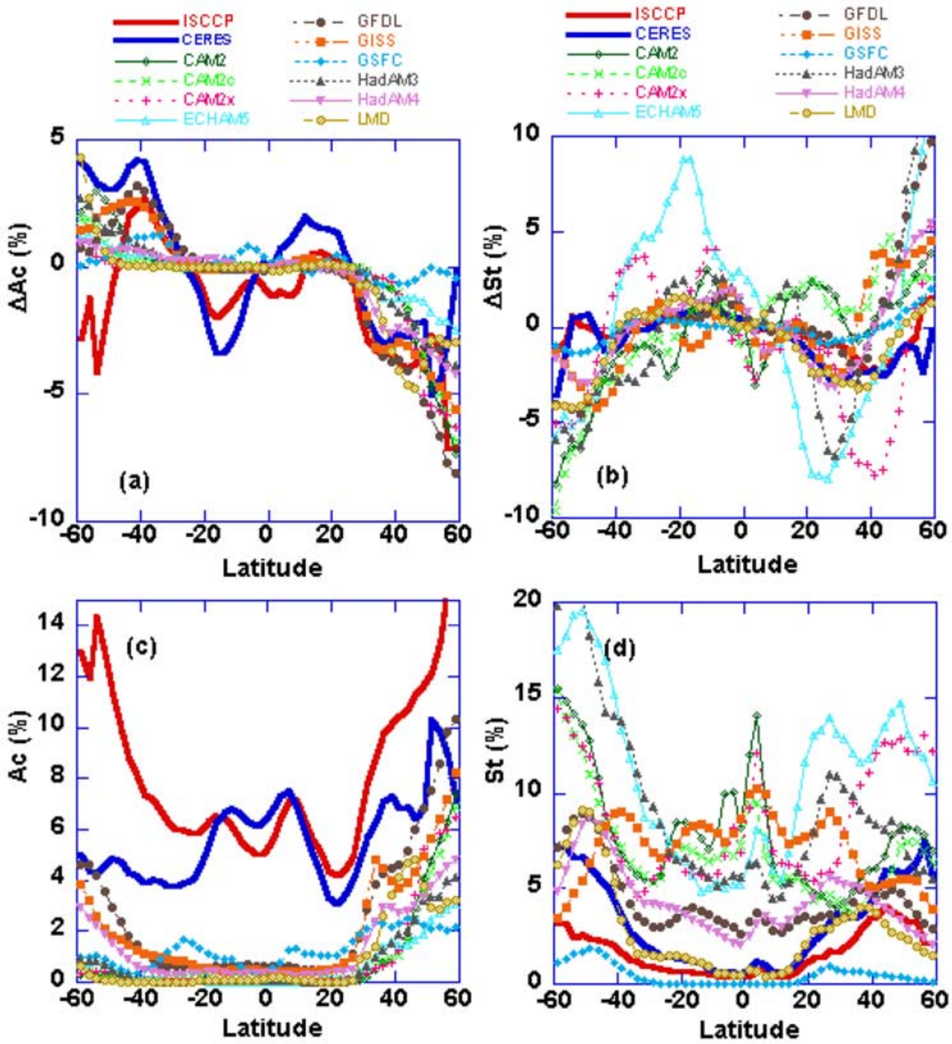


**Figure 12.** Seasonal variations of (a) middle top thin clouds and (b) low top thick clouds. Basic annual cloud distributions for (c) middle top thin clouds and (d) low-top thick clouds.

is the simulated magnitude of the basic cloud frequencies. Models that systematically underestimate middle clouds also underestimate seasonal variations, while models that overestimate optically thick clouds also overestimate the seasonal sensitivity of these clouds.

[48] Even though the systematic cloud biases are common to most models, some models simulate certain cloud types better than others. It is highly desirable if the positive and negative attributes of model clouds can be associated with specific physical parameterizations. As pointed out in the work of *Webb et al.* [2001], in assessing clouds in models, many model components can be as important as the cloud and precipitation schemes. Without carrying out controlled experiments by isolating individual physical parameterization components, it is difficult to pinpoint the source of the model differences. Nevertheless, certain relationships with physical parameterizations can be observed. Both the HadAM4 and the GFDL models used the *Lock et al.* [2000] PBL scheme that contained additional turbulent mixing due to cloud top entrainment, and these two models simulated relatively better low clouds in all optical thickness ranges. On the other hand, the overestimation of optically thick clouds is common to models that used very different cloud schemes. For example, CAM used a relative humidity-based cloud scheme, the HadAM used a statistical cloud scheme, while the GFDL model used a prognostic cloud scheme. They have similar biases in optically thick high top clouds, in both the tropics and middle latitudes.

[49] *Williams et al.* [2003] and *Lin and Zhang* [2004] showed that optically thick clouds occur with strong vertical ascent associated with either large-scale convective systems or middle latitude frontal systems in models. Cloud schemes in current GCMs are designed to account for partial cloud cover, but not for subgrid distribution of temperature and moisture tendencies as a result of the subgrid circulations. These subgrid structures are sometimes part of the large-scale systems, sometimes they are self organized and maintained within the grid box [*Katzfey and Ryan*, 2000; *Ryan et al.*, 2000]. When a grid box has only a fractional area of strong upward motion, the mean vertical motion for the GCM grid box is upward. In a relative humidity-based fractional cloud scheme, the fractional cloud cover merely allows water to condense before the whole grid is saturated, but it does not prevent the humidity in the clear-sky portion

of the grid from rising as a result of the mean vertical motion. In traditional statistical schemes, the grid-scale upward motion typically shifts the total water distribution upward and the saturation threshold downward, rather than changing the subgrid-scale distributions of total water. In the prognostic cloud amount scheme of *Tiedtke* [1993], stratiform cloud tendency was formulated based on the difference of the grid box mean mixing ratio, which is controlled by the mean vertical motion, and its saturation value. One possible cause of the overestimation of high thick clouds in most models may be therefore due to the discretization of advective tendencies for the grid boxes, which has not been adequately accounted for in current fractional cloud schemes. Diagnosing the subgrid-scale distribution of total water as a function of subgrid-scale processes (e.g., convective or boundary-layer processes) may improve that situation. For instance, the LMD model, whose cloud scheme is physically coupled to the convection scheme and uses both large-scale and subgrid-scale predictors to diagnose the total water distribution [*Bony and Emanuel*, 2001], does not overestimate the occurrence of optically thick clouds as much as most other models (Figures 7 and 8). Other possible causes of the model biases include vertical resolution and cloud microphysical properties. The overestimation of optically thick high clouds in the models can be also confirmed if daily TOA radiative fluxes are used to compare with model results. This has been reported in the work of *Norris and Weaver* [2001].

[50] The overestimation of low optically thick clouds in many models could be due to completely different reasons. Again, regardless of which type of cloud schemes is used, the mean relative humidity is a major control variable of clouds in the models. In the planetary boundary layer (PBL), the dynamic range of the relative humidity variation is small, which makes it a poor predictor of clouds. None of the models had a PBL scheme that directly predicts clouds. Instead, clouds are predicted by the model's stratiform or convective cloud scheme with its parameters modified by the PBL processes.

[51] With respect to the underestimation of low and middle optically thin clouds, it is well known that even mesoscale models cannot simulate clouds from shallow convections (see Figure 9 and *Bretherton et al.* [2004]). The simulation of cumulus and stratocumulus in the tropics has been a challenge for the modeling community for a long time. In middle latitudes, especially over the oceans, these same types of clouds frequently occur after a cold front due to the temperature contrast between warm water and cold air. Insufficient vertical resolution in the models is also a possible cause. In addition, observations also show that surface heterogeneity and topography can often generate shallow mesoscale and synoptic-scale circulation systems that cannot be adequately resolved by coarse resolution models.

[52] Much more needs to be done to fully understand the physical causes of model cloud biases presented here and to improve the models. Process oriented study, with observations of vertical cloud distributions and cloud optical properties, will be most useful to associate clouds with transient atmospheric dynamical circulations on different scales. Some of these challenges are discussed in several papers in this volume.

[53] **Acknowledgments.** We wish to thank Joel Norris for providing us with the adjusted surface-based cloud climatology that served as an independent check of our results. We also thank William Rossow for many discussions on the ISCCP data. Constructive comments from the three anonymous reviewers have helped us to significantly improve the original version of the paper. This research is supported by the DOE Atmospheric Radiation Measurement Program (ARM) to the Stony Brook University, the Geophysical Fluid Dynamics Laboratory, the Goddard Institute for Space Studies, the Lawrence Livermore National Laboratory, and the NASA Earth Science Enterprise through the CERES Project. Work at Dalhousie University was supported by MOC2 project funded by NSERC and CFCAS. Work and numerical simulations done at LMD/IPSL were supported by CNRS and IDRIS, respectively. Additional support is provided by the National Science Foundation to the Stony Brook University and the National Center for Atmospheric Research and by the NASA Radiation Science Program to the Goddard Institute for Space Studies.

## References

Bony, S., and K. A. Emanuel (2001), A parameterization of the cloudiness associated with cumulus convection: Evaluation using TOGA COARE data, *J. Atmos. Sci.*, *58*, 3158–3183.

Bony, S., J.-L. Dufresne, H. Le Treut, J.-J. Morcrette, and C. Senior (2004), On dynamic and thermodynamic components of cloud changes, *Clim. Dyn.*, *22*(2/3), 71–76.

Bretherton, C. S., J. R. McCaa, and H. Grenier (2004), A new parameterization for shallow cumulus convection and its application to marine subtropical cloud-topped boundary layers. Part I: Description and 1D results, *Mon. Weather Rev.*, *132*, 864–882.

Brinkop, S., and E. Roeckner (1995), Sensitivity of a general circulation model to parameterizations of cloud-turbulence interactions in the atmospheric boundary layer, *Tellus, Ser. A*, *47*, 197–220.

Cess, R. D., et al. (1990), Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models, *J. Geophys. Res.*, *95*, 16,601–16,615.

Cess, R. D., et al. (1997), Comparison of the seasonal change in cloud-radiative forcing from atmospheric general circulation models and satellite observations, *J. Geophys. Res.*, *102*, 16,593–16,604.

Cheng, Y., V. M. Canuto, and A. M. Howard (2002), An improved model for the turbulent PBL, *J. Atmos. Sci.*, *59*, 1550–1565.

Cubasch, U., et al. (2001), Projection of future climate change, in *Climate Change 2001: The Scientific Basis*, edited by J. T. Houghton et al., pp. 527–582, Cambridge Univ. Press, New York.

Cusack, S., J. M. Edwards, and R. Kershaw (1999), Estimating subgrid variance of saturation and its parameterization for use in a GCM cloud scheme, *Q. J. R. Meteorol. Soc.*, *125*, 3057–3076.

Del Genio, A. D., and M.-S. Yao (1993), Efficient cumulus parameterization for long-term climate studies: The GISS scheme, in *The Representation of Cumulus Convection in Numeric Models*, *Meteorol. Monogr. Ser.*, vol. 46, pp. 181–184, Am. Meteorol. Soc., Boston, Mass.

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

Del Genio, A. D., W. Kovari, M.-S. Yao, and J. Jonas (2005), Cumulus microphysics and climate sensitivity, *J. Clim.*, in press.

Emanuel, K. A. (1991), A scheme for representing cumulus convection in large-scale models, *J. Atmos. Sci.*, *48*, 2313–2335.

Garreaud, R. D., J. Rutllant, J. Quintana, J. Carrasco, and P. Minnis (2001), CIMAR-5: A snapshot of the lower troposphere over the subtropical southeast Pacific, *Bull. Am. Meteorol. Soc.*, *92*, 2193–2208.

Geophysical Fluid Dynamics Laboratory Global Atmospheric Model Development Team (GFDL GAMDT) (2004), The new GFDL global atmosphere and land model AM2/LM2: Evaluation with prescribed SST conditions, *J. Clim.*, *17*(24), 4641–4673.

Ghan, S. J., et al. (2000), An intercomparison of single column model simulations of summertime midlatitude continental convection, *J. Geophys. Res.*, *105*, 2091–2124.

Gregory, D., and S. Allen (1991), The effect of convective scale downdrafts upon NWP and climate simulations, in *9th Conference on Numerical Weather Prediction*, pp. 122–123, Am. Meteorol. Soc., Boston, Mass.

Gregory, D., and P. R. Rowntree (1990), A mass flux convection scheme with representation of cloud ensemble characteristics and stability dependent closure, *Mon. Weather Rev.*, *118*, 1483–1506.

Gregory, J. M. (1999), Representation of the radiative effect of convective anvils, *Hadley Cent. Tech. Note 7*, Hadley Cent. for Clim. Predict. and Res., Exeter, UK.

Hack, J. J. (1994), Parameterization of moist convection in the National Center for Atmospheric Research Community Climate Model (CCM2), *J. Geophys. Res.*, *99*, 5551–5568.

Harrison, E. F., P. Minnis, B. R. Barkstrom, V. Ramanathan, R. D. Cess, and G. G. Gibson (1990), Seasonal variation of cloud radiative forcing

derived from the Earth Radiation Budget Experiment, *J. Geophys. Res.*, *95*, 18,687–18,703.
Helfand, H. M., and J. C. Labraga (1988), Design of a non-singular level 2.5 second order closure model for the prediction of atmospheric turbulence, *J. Atmos. Sci.*, *45*, 113–132.
Hogan, R. J., and A. J. Illingworth (2000), Deriving cloud statistics from radar, *Q. J. R. Meteorol. Soc.*, *126*, 2903–2909.
Holtslag, A. A. M., and B. A. Boville (1993), Local versus nonlocal boundary-layer diffusion in a global climate model, *J. Clim.*, *6*, 1825–1842.
Jin, Y., W. B. Rossow, and D. P. Wylie (1996), Comparison of the climatologies of high-level clouds from HIRS and ISCCP, *J. Clim.*, *9*, 2850–2879.
Katzfey, J. J., and B. F. Ryan (2000), Midlatitude frontal clouds: GCM-scale modeling implications, *J. Clim.*, *13*, 2729–2745.
Kiehl, J. T., J. Hack, G. Bonan, B. Boville, B. Briegleb, D. Williamson, and P. Rasch (1996), Description of the NCAR Community Climate Model (CCM3), *Tech. Rep. NCAR/TN-420+STR*, 152 pp., Natl. Cent. for Atmos. Res., Boulder, Colo.
King, M. D., Y. J. Kaufman, W. P. Menzel, and D. Tanré (1992), Remote sensing of cloud, aerosol, and water vapor properties from the Moderate Resolution Imaging Spectrometer (MODIS), *IEEE Trans. Geosci. Remote Sens.*, *30*, 2–27.
Klein, S. A., and C. Jakob (1999), Validation and sensitivities of frontal clouds simulated by the ECMWF model, *Mon. Weather Rev.*, *127*, 2514–2531.
Le Treut, H., and Z. X. Li (1991), Sensitivity of an atmospheric general circulation model to prescribed SST changes: Feedback effects associated with the simulation of cloud optical properties, *Clim. Dyn.*, *5*, 175–187.
Li, Z. X. (1999), Ensemble atmospheric GCM simulation of climate interannual variability from 1979 to 1994, *J. Clim.*, *12*, 986–1001.
Lin, W. Y., and M. H. Zhang (2004), Evaluation of clouds and their radiative effects simulated by the NCAR Community Atmospheric Model CAM2 against satellite observations, *J. Clim.*, *17*, 3302–3318.
Lock, A. P., A. R. Brown, M. R. Bush, G. M. Martin, and R. N. B. Smith (2000), A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests, *Mon. Weather Rev.*, *128*, 3187–3199.
Lohmann, U., and E. Roeckner (1996), Design and performance of a new cloud microphysics scheme developed for the ECHAM4 general circulation model, *Clim. Dyn.*, *12*, 557–572.
Martin, G. M., M. R. Bush, A. R. Brown, A. P. Lock, and R. N. B. Smith (2000), A new boundary layer mixing scheme. Part II: Tests in climate and mesoscale models, *Mon. Weather Rev.*, *128*, 3200–3217.
Minnis, P., P. W. Heck, D. F. Young, C. W. Fairall, and J. B. Snider (1992), Stratocumulus cloud properties derived from simultaneous satellite and island-based instrumentation during FIRE, *J. Appl. Meteorol.*, *31*, 317–339.
Minnis, P., D. P. Kratz, J. A. Coakley Jr., M. D. King, D. Garber, P. Heck, S. Mayor, D. F. Young, and R. Arduini (1995), Cloud optical property retrieval (subsystem 4.3), in *Clouds and the Earth's Radiant Energy System (CERES) Algorithm Theoretical Basis Document*, vol. 3, *Cloud Analyses and Radiance Inversions (Subsystem 4)*, edited by CERES Science Team, pp. 135–176, NASA, Greenbelt, Md.
Minnis, P., D. P. Garber, D. F. Young, R. F. Arduini, and Y. Takano (1998), Parameterizations of reflectance and effective emittance for satellite remote sensing of cloud properties, *J. Atmos. Sci.*, *55*, 3313–3339.
Moorthi, S., and M. J. Suarez (1992), Relaxed Arakawa-Schubert: A parameterization of moist convection for general circulation models, *Mon. Weather Rev.*, *120*, 978–1002.
Nordeng, T. E. (1994), Extended versions of the convective parameterization scheme at ECMWF and their impact on the mean and transient activity of the model in the tropics, *Tech. Memo. 206*, Eur. Cent. for Med. Range Weather Forecast., Reading, UK.
Norris, J. R. (1999), On trends and possible artifacts in global ocean cloud cover between 1952 and 1995, *J. Clim.*, *12*, 1864–1870.
Norris, J. R., and C. P. Weaver (2001), Improved techniques for evaluating GCM cloudiness applied to the NCAR CCM3, *J. Clim.*, *14*, 2540–2550.
Pincus, R., H. W. Barker, and J. Morcrette (2003), A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields, *J. Geophys. Res.*, *108*(D13), 4376, doi:10.1029/2002JD003322.
Pope, V. D., M. Gallani, P. R. Rowntree, and R. A. Stratton (2000), The impact of new physical parameterizations in the Hadley Centre climate mode—HadAM3, *Clim. Dyn.*, *16*, 123–146.
Potter, G. L., and R. D. Cess (2004), Testing the impact of clouds on the radiation budgets of 19 atmospheric general circulation models, *J. Geophys. Res.*, *109*, D02106, doi:10.1029/2003JD004018.
Randall, D. A., K. M. Xu, R. J. C. Somerville, and S. Iacobellis (1996), Single-column models and cloud ensemble models as links between observations and climate models, *J. Clim.*, *9*, 1683–1697.
Rasch, P. J., and J. E. Kristjánsson (1998), A comparison of the CCM3 model climate using diagnosed and predicted condensate parameterizations, *J. Clim.*, *11*, 1587–1614.
Roeckner, E., et al. (1987), Cloud optical depth feedbacks and climate modeling, *Nature*, *329*, 138–140.
Roeckner, E., et al. (1996), The atmospheric general circulation model ECHAM-4: Model description and simulation of present-day climate, *Rep. 128*, Max-Planck-Inst. für Meteorol., Hamburg, Germany.
Rossow, W. B., and R. A. Schiffer (1999), Advances in understanding clouds from ISCCP, *Bull. Am. Meteorol. Soc.*, *80*, 2261–2288.
Rossow, W. B., A. W. Walker, D. Beuschel, and M. Roiter (1996), International Satellite Cloud Climatology Project (ISSCP) description of new cloud datasets, *Rep. WMO/TD 737*, 115 pp., World Clim. Res. Prog., World Meteorol. Org., Geneva.
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 microphysical processes, *Q. J. R. Meteorol. Soc.*, *123*, 1227–1282.
Ryan, B. F., et al. (2000), Simulations of a cold front by cloud-resolving, limited-area, and large-scale models, and a model evaluation using in situ and satellite observations, *Mon. Weather Rev.*, *128*, 3218–3235.
Senior, C. A., and J. F. B. Mitchell (1993), Carbon dioxide and climate: The impact of cloud parameterization, *J. Clim.*, *6*, 393–418.
Senior, C. A., and J. F. B. Mitchell (1996), Cloud feedbacks in the unified UKMO GCM, in *Climate Sensitivity to Radiative Perturbations, Physical Mechanism and Their Validation*, edited by H. Le Treut, pp. 191–202, Springer, New York.
Smith, R. N. B. (1990), A scheme for predicting layer clouds and their water content in a general circulation model, *Q. J. R. Meteorol. Soc.*, *116*, 435–460.
Sud, Y. C., and G. K. Walker (1999), Microphysics of Clouds with the Relaxed Arakawa-Schubert Scheme (McRAS). Part II: Implementation and performance in GEOS II GCM, *J. Atmos. Sci.*, *56*(18), 3221–3240.
Tiedtke, M. (1989), A comprehensive mass flux scheme for cumulus parameterization in large scale models, *Mon. Weather Rev.*, *117*, 1779–1800.
Tiedtke, M. (1993), Representation of clouds in large-scale models, *Mon. Weather Rev.*, *121*, 3040–3061.
Tompkins, A. (2002), A prognostic parameterization for the subgrid-scale variability of water vapor and clouds in large-scale models and its use to diagnose cloud cover, *J. Atmos. Sci.*, *59*, 1917–1942.
Trepte, Q., Y. Chen, S. Sun-Mack, P. Minnis, D. F. Young, B. A. Baum, and P. W. Heck (1999), Scene identification for the CERES cloud analysis subsystem, in *Proceedings of the AMS 10th Conference on Atmospheric Radiation*, pp. 169–172, Am. Meteorol. Soc., Boston, Mass.
Tselioudis, G., and C. Jakob (2002), Evaluation of midlatitude cloud properties in a weather and a climate model: Dependence on dynamic regime and spatial resolution, *J. Geophys. Res.*, *107*(D24), 4781, doi:10.1029/2002JD002259.
Weare, B. C. (2004), A comparison of AMIP II model cloud layer properties with ISCCP D2 estimates, *Clim. Dyn.*, *22*, 281–292.
Weare, B. C., and AMIP Modeling Groups (1996), Evaluation of the vertical structure of zonally averaged cloudiness and its variability in the Atmospheric Model Intercomparison Project, *J. Clim.*, *9*, 3419–3431.
Webb, M., C. Senior, S. Bony, and J. J. Morcrette (2001), Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models, *Clim. Dyn.*, *17*, 905–922.
Wielicki, B. A., B. R. Barkstrom, E. F. Harrison, R. B. Lee III, G. L. Smith, and J. E. Cooper (1996), Clouds and the Earth's Radiant Energy System (CERES): An Earth observing system experiment, *Bull. Am. Meteorol. Soc.*, *77*, 853–868.
Williams, K. D., M. A. Ringer, and C. A. Senior (2003), Evaluating the cloud response to climate change and current climate variability, *Clim. Dyn.*, *20*, 705–721.
Wilson, D. R., and S. P. Ballard (1999), A microphysical based precipitation scheme for the UK Meteorological Office numerical weather prediction model, *Q. J. R. Meteorol. Soc.*, *125*, 1607–1636.
Wylie, D. P., and W. P. Menzel (1999), Eight years of high cloud statistics using HIRS, *J. Clim.*, *12*, 170–184.
Xie, S. C., and M. H. Zhang (2000), Impact of the convection triggering function on single-column model simulations, *J. Geophys. Res.*, *105*, 14,983–14,996.
Xie, S. C., et al. (2002), Intercomparison and evaluation of cumulus parameterizations under summertime midlatitude continental conditions, *Q. J. R. Meteorol. Soc.*, *128*, 1095–1135.
Xie, S., M. Zhang, J. S. Boyle, R. T. Cederwall, G. L. Potter, and W. Lin (2004), Impact of a revised convective triggering mechanism on Community Atmosphere Model, Version 2, simulations: Results from short-range weather forecasts, *J. Geophys. Res.*, *109*, D14102, doi:10.1029/2004JD004692.

Xu, K. M., et al. (2002), An intercomparison of cloud-resolving models with the Atmospheric Radiation Measurement summer 1997 intensive observation period data, *Q. J. R. Meteorol. Soc.*, *128*, 593–624.

Zhang, G. J., and N. A. McFarlane (1995), Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centre general circulation model, *Atmos. Ocean*, *33*, 407–446.

Zhang, Y. C., W. B. Rossow, and A. A. Lacis (1995), Calculation of surface and top of atmosphere radiative fluxes from physical quantities based on ISCCP data sets: 1. Method and sensitivity, *J. Geophys. Res.*, *100*, 1149–1165.

J. T. Bacmeister and M. J. Suarez, Global Modeling and Analysis Office, NASA GSFC, Code 900.3, Greenbelt, MD 20771, USA.

S. Bony and I. Musat, LMD/IPSL, 4 Place Jussieu, Boite 99, F-75252 Paris Cedex 05, France.

R. T. Cederwall, S. A. Klein, and S. C. Xie, Atmospheric Science Division, Lawrence Livermore National Laboratory, L-103 PO Box 808, Livermore, CA 94551, USA.

A. D. Del Genio and M.-S. Yao, NASA GISS, 2880 Broadway, New York, NY 10025, USA.

J. J. Hack, NCAR, PO Box 3000, Boulder, CO 80307, USA.

W. Y. Lin, J. B. Wu, and M. H. Zhang, Institute for Terrestrial and Planetary Atmospheres, Marine Sciences Research Center, Stony Brook University, Stony Brook, NY 11794-5000, USA. (mzhang@notes.cc.sunysb.edu)

N. G. Loeb and P. Minnis, Atmospheric Science Division, MS 420, NASA Langley Research Center, Hampton, VA 23681-0001, USA.

U. Lohmann, Institut für Atmosphäre und Klima, ETH Hönggerberg, HPP L 1.2, Schafmattstrasse 30, CH-8093 Zürich, Switzerland.

R. Pincus, NOAA CIRES Climate Diagnostics Center, 325 S. Broadway, R/CDC1, Boulder, CO 80305, USA.

P. Stier, Max Planck Institute for Meteorology, D-20146 Hamburg, Germany.

M. J. Webb, Hadley Centre for Climate Prediction and Research, UK Met Office, FitzRoy Road, Exeter EX1 3PB, UK.

J. H. Zhang, Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia B3H 3J5, Canada.