

An observationally based evaluation of cloud ice water in CMIP3 and CMIP5 GCMs and contemporary reanalyses using contemporary satellite data

J.-L. F. Li,¹ D. E. Waliser,¹ W.-T. Chen,¹ B. Guan,¹ T. Kubar,¹ G. Stephens,¹ H.-Y. Ma,² M. Deng,³ L. Donner,⁴ C. Seman,⁴ and L. Horowitz⁴

Received 16 February 2012; revised 11 June 2012; accepted 11 June 2012; published 21 August 2012.

[1] We perform an observationally based evaluation of the cloud ice water content (CIWC) and path (CIWP) of present-day GCMs, notably 20th century CMIP5 simulations, and compare these results to CMIP3 and two recent reanalyses. We use three different CloudSat + CALIPSO ice water products and two methods to remove the contribution from the convective core ice mass and/or precipitating cloud hydrometeors with variable sizes and falling speeds so that a robust observational estimate can be obtained for model evaluations. The results show that for annual mean CIWP, there are factors of 2–10 in the differences between observations and models for a majority of the GCMs and for a number of regions. However, there are a number of CMIP5 models, including CNRM-CM5, MRI, CCSM4 and CanESM2, as well as the UCLA CGCM, that perform well compared to our past evaluations. Systematic biases in CIWC vertical structure occur below the mid-troposphere where the models overestimate CIWC, with this bias arising mostly from the extratropics. The tropics are marked by model differences in the level of maximum CIWC (~ 250 – 550 hPa). Based on a number of metrics, the ensemble behavior of CMIP5 has improved considerably relative to CMIP3, although neither the CMIP5 ensemble mean nor any individual model performs particularly well, and there are still a number of models that exhibit very large biases despite the availability of relevant observations. The implications of these results on model representations of the Earth radiation balance are discussed, along with caveats and uncertainties associated with the observational estimates, model and observation representations of the precipitating and cloudy ice components, relevant physical processes and parameterizations.

Citation: Li, J.-L. F., et al. (2012), An observationally based evaluation of cloud ice water in CMIP3 and CMIP5 GCMs and contemporary reanalyses using contemporary satellite data, *J. Geophys. Res.*, 117, D16105, doi:10.1029/2012JD017640.

1. Introduction

[2] Representing clouds and cloud climate feedback in global climate models (GCMs) remains a pressing challenge to reduce and quantify uncertainties associated with climate change projections. Until recently, useful (global) constraints for developing and evaluating clouds in GCMs were derived from radiation budget observations from Earth Radiation Budget Experiment/Clouds and the Earth's Radiant Energy

System (ERBE and CERES) [Wielicki et al., 1996], cloud cover observations from International Satellite Cloud Climatology Project (ISCCP) and related products [Han et al., 1999; Rossow and Schiffer, 1999], and other indirect observations such as precipitation observations [Xie and Arkin, 1997; Adler et al., 2003]. A key element of obtaining an accurate top of the atmosphere (TOA) radiation budget is the representation of clouds, which for GCMs and earth radiation budget considerations can be roughly broken down into realistic cloud cover, cloud water mass, cloud optical thickness, and cloud particle sizes. In general, the observed TOA radiative fluxes (e.g., CERES, ERBE) are used to constrain GCMs by selecting the coefficients in convective cloud, stratiform cloud fraction parameterizations. In addition, in a coupled atmosphere-ocean GCM, it is important to consider radiative effects/budgets emphasis on implied ocean heat transports and the global and regional patterns of TOA OLR and absorbed shortwave, with some emphasis on shortwave cloud forcing and longwave cloud forcing. In most cases, models such as in Coupled Model Intercomparison Project

¹Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA.

²Lawrence Livermore National Laboratory, Livermore, California, USA.

³Atmospheric Science, University of Wyoming, Laramie, Wyoming, USA.

⁴NOAA Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey, USA.

Corresponding author: J.-L. F. Li, Jet Propulsion Laboratory, California Institute of Technology, MS 233-306K, 4800 Oak Grove Dr., Pasadena, CA 91109, USA. (julin.f.li@jpl.nasa.gov)

©2012. American Geophysical Union. All Rights Reserved.
0148-0227/12/2012JD017640

Phase 3 (CMIP3) tend to underestimate cloud fraction when compared which is compensated by optically thick clouds. While ISCCP and other products have provided some information and guidance on cloud cover, there has been significant flexibility in tuning the latter two quantities due to the lack of observations for cloud water mass and particle size. This is especially the case for vertical profiles of information for cloud/hydrometeor particle sizes and water mass – leaving too many degrees of freedom unconstrained. The gap in available observations for cloud water mass and/or their lack of use in constraining the models was clearly evident from the wide disparity in the cloud ice and liquid water path (CIWP and CLWP) values exhibited in the CMIP3 GCMs [Li et al., 2008; Waliser et al., 2009].

[3] Since the period when the phase 3 of the Coupled Model Intercomparison Project (CMIP3) class of GCMs were developed, upper tropospheric ice values from the Microwave Limb Sounder (MLS) [Wu et al., 2008, 2009] and tropospheric ice values from CloudSat [Stephens et al., 2008] – both with some vertical-resolving capabilities – have become available to aid model development and provide a means for the global evaluation of the models [e.g., Li et al., 2007; Waliser et al., 2009; Chen et al., 2011; Gettelman et al., 2010]. While these observations are an altogether new resource, albeit with uncertainties [e.g., Hong, 2007; Heymsfield et al., 2008; Wu et al., 2008; Waliser et al., 2009], for improving GCM representation of clouds and cloud climate feedback, considerable care and caution have to be taken in order to make judicious comparisons between the GCM representations of (typically only the) clouds and the satellite observations that are an inherent combination of the clouds and falling hydrometeors (e.g., rain or snow). Such considerations include taking steps to make a sensible comparison for *cloud* ice water content (CIWC) or CIWP [e.g., Waliser et al., 2009; Chen et al., 2011; Ma et al., 2012a] or taking the steps needed for a viable comparison in terms of reflectivity or radiance, which typically means using satellite “simulators” [e.g., Klein and Jakob, 1999; Webb et al., 2001; Delanoë and Hogan, 2008, 2010; Delanoë et al., 2011; Bodas-Salcedo et al., 2011].

[4] In this study, we take the first approach mentioned above, and perform the evaluation in terms of the model representations of CIWC/CIWP. We utilize the experience we have gained from a number of studies performed in recent years on cloud ice and liquid [e.g., Li et al., 2005, 2007, 2008; Waliser et al., 2009; Chen et al., 2011; Ma et al., 2012a]. This includes developing a measure of observational uncertainty (discussed in section 2), and applying an illustrative and quantitative set of evaluation diagnostics. A prominent goal of the study is to examine how the fidelity of the models may have changed between CMIP3 and CMIP5. Moreover, we attempt to discriminate CMIP5 models that achieve a threshold capability of model fidelity via the Taylor diagram [Taylor, 2001] construct and the observational uncertainties just mentioned. In addition, as reanalyses data for some quantities such as the basic atmospheric state variables have become nearly synonymous, in some contexts, with “observations,” we also incorporate two recent reanalysis data products in our evaluation to provide some assessment of this questionable perception – particularly for quantities such as CIWP that are not strongly constrained by observations.

[5] In the following section, we describe the observational resources we use for this study, including the way the different retrievals and other methodologies are combined to form a robust observational estimate with some quantitative information on uncertainty. In section 3, we briefly describe the models and reanalyses data sets utilized in this evaluation study. In section 4, we illustrate and discuss the results of our model evaluation. Section 5 summarizes and draws conclusions.

2. Observed Estimates of IWC and IWP

[6] The A-Train constellation of satellites, which includes CloudSat and Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) flying only 15 s apart, provides a global view of the vertical structure of clouds, including cloud condensate such as IWC. CloudSat provides vertical profiles of radar reflectivity measured by a 94 GHz cloud profiling radar (CPR) with a minimum sensitivity of ~ -30 dBZ. The profiles extend between the surface and 30 km altitude with a vertical resolution of 240 m and have a footprint of about 2.5 km along track and 1.4 km across track. The CALIPSO lidar measures parallel and perpendicular backscattered laser energy at 532 nm and total backscattering at 1064 nm at altitude-dependent vertical resolutions and footprints (75 m vertically with about 0.3 km along track footprint above 8.2 km and 30 m vertically with about 1.0 km along track footprint below 8.2 km). To date, a series of retrieval algorithms either using CloudSat radar or CALIPSO lidar or both provide global retrievals of IWC, effective radius (Re), and the extinction coefficient from the thinnest cirrus (seen only by the lidar) to the thickest ice cloud [Austin and Stephens, 2001; Hogan, 2006; Delanoë and Hogan, 2008, 2010; Mace et al., 2009; Young and Vaughan, 2009; Sassen et al., 2009; Stein et al., 2011; M. Deng et al., Evaluation of several A-Train ice cloud retrieval products with in situ measurements collected during the SPartICus campaign, submitted to *Journal of Applied Meteorology and Climatology*, 2012].

[7] In this study, IWC and IWP products retrieved from three different algorithms are used to help account for observational uncertainty. They are:

[8] 1. 2B-CWC-RO4 [Austin et al., 2009]. The CloudSat Science Team radar-only product provides estimates of IWC and Re using measured radar reflectivity from CloudSat 2B-GEOPROF to constrain the retrieved IWC. The retrieved IWC profiles are obtained by assuming constant ice particle density with a spherical shape and a lognormal particle size distribution (PSD). An a priori PSD is specified based on its temperature dependencies obtained from European Centre for Medium-Range Weather Forecasts (ECMWF) operational analyses. The cloud water contents for both liquid and ice phases are retrieved for all heights using separate assumptions, then a composite profile is created by using the retrieved ice properties at temperatures colder than -20°C , the retrieved liquid water content at temperatures warmer than 0°C , and a linear combination of the two in intermediate temperatures. This reduces the total IWC as the temperature approaches 0°C . The sensitivity and uncertainty of this retrieval algorithm are discussed in Austin et al. [2009]. The time period of this data set is from January 2007 to December 2010. The vertical and horizontal resolutions are the same as

the CloudSat instrument discussed above. Our previous CMIP3 model evaluations [Waliser et al., 2009] were based on an earlier version of this product.

[9] 2. *DARDAR*. DARDAR (raDAR/liDAR) [Hogan, 2006; Delanoë and Hogan, 2008, 2010] is a synergistic ice cloud retrieval product derived from the combination of the CloudSat radar and CALIPSO lidar using a variational method for retrieving profiles of the extinction coefficient, IWC and Re of the ice cloud. DARDAR assumes a “unified” PSD given by Field et al. [2005]. The mass-size and area-size relations of non-spherical particles are considered using in situ measurements [Brown and Francis, 1995; Francis et al., 1998; Delanoë et al., 2011; Stein et al., 2011]. For DARDAR, it used CALIPSO backscatter and temperature to find supercooled water in the 0°C to –40°C range, while the depolarization is too noisy to use at the CALIPSO resolution [Delanoë and Hogan, 2010]. The time period of this data set is from July 2006 to June 2009.

[10] 3. *2C-ICE* (Deng et al., submitted manuscript, 2012). Similar to DARDAR, the CloudSat 2C-ICE cloud product is a synergistic ice cloud retrieval derived from the combination of the CloudSat radar and CALIPSO lidar using a variational method for retrieving profiles of the extinction coefficient, IWC, and Re in ice clouds. The CALIPSO attenuated backscattering coefficients are collocated to the CloudSat vertical and horizontal resolutions. The ice cloud microphysical model assumes a first-order Gamma particle size distribution of idealized non-spherical ice crystals [Yang et al., 2000]. The Mie scattering of radar reflectivity is calculated in a forward model look-up table according to a discrete dipole approximation calculation [Hong, 2007]. The 2C-ICE cloud identification is provided by CloudSat CLDCLASS-Lidar product, which takes advantage of CALIPSO lidar backscatter (sensitive to water clouds), lidar depolarization (sensitive to non-spherical ice particle) and CloudSat radar (sensitive to large ice particles). For the mixed-phase clouds, 2C-ICE treats it as ice clouds using radar only to perform the retrieval. For mixed-phase clouds in deep convection, where the lidar is usually attenuated, they retrieve clouds above –6°C mainly relying on CloudSat radar only [Deng et al., 2010]. Readers desiring a more in-depth description of the 2C-ICE algorithm should refer to Deng et al. [2010] for details. The time period of this data set is from January 2007 to December 2008.

[11] There are several differences between 2C-ICE and DARDAR. First, DARDAR is retrieved using the CALIPSO vertical resolution (60 m) instead of the CloudSat vertical resolution as in 2C-ICE. Second, the multiple scattering in the lidar signal is accounted for with a fast multiple-scattering code [Hogan, 2006] instead of assuming a constant multiple scattering factor as in 2C-ICE. Third, the lidar backscatter to extinction ratio is retrieved rather than assumed to be a constant as in 2C-ICE. Fourth, no parameterizations of radar or lidar signals are used for the lidar-only or radar-only regions of the ice cloud profile. Empirical relationships are heavily relied on for those regions in the DARDAR algorithm. Fifth, the DARDAR product assumes a “unified” PSD given by Field et al. [2005]. The mass-size and area-size relation of non-spherical particles is considered using relationships derived from in situ measurements [Francis et al., 1998; Brown and Francis, 1995], while 2C-ICE used a first order Gamma PSD of idealized ice particles in Yang et al. [2000]

with consistent size, area, mass, and scattering property relations (Deng et al., submitted manuscript, 2012).

[12] There are two important aspects to keep in mind regarding model and observation compatibility. First, because of the use of the CALIPSO in the latter two products described above, they will have more sensitivity to thin/cirrus ice clouds. However, thin ice clouds make very little contribution to the total ice mass and water content of clouds in a time-mean sense. Second, and more importantly, all three products, to first order, represent total tropospheric ice including “floating” ice and the precipitating cloud hydrometeors with variable size and falling speed, as the measurements are sensitive to a wide range of particle sizes, including small (quasi-suspended/cloud) particles and large (falling/precipitating) particles. The latter, including those particles associated with convective clouds, are generally not included as prognostic variables in most current GCMs [e.g., Li et al., 2008; Waliser et al., 2009]. It is generally assumed that convective core areas in a GCM grid box are small for a GCM with a grid box size that is commonly larger than one hundred km² and above. Thus, its contribution to total water content is not very large. Even they are prognostically determined, the relative contribution does not change. However, the grid box resolution in most current state-of-the-arts GCMs is much higher, with grid box size smaller than 100 km² to tens of km², so that their IWC contribution from the convective core should be considered. Thus, for a meaningful model-observation comparison between the satellite-estimated and model-simulated IWC, an estimate of the convective/precipitating ice mass needs to be removed from the satellite-derived IWC/IWP values.

[13] In this study, we use two independent approaches to distinguish ice mass associated with clouds from ice mass associated with precipitation and convection. The two approaches include:

[14] 1. *FLAG method* [Li et al., 2008; Waliser et al., 2009]. We exclude all the retrievals in any profile that are flagged as precipitating at the surface and exclude any retrieval within the profile whose cloud type is classified as “deep convection” or “cumulus” (from CloudSat 2B-CLDCLASS data). By excluding these portions of the ice mass, we obtain an estimate of the cloud-only portion of the IWP/IWC (hereafter, referred to as CIWP/CIWC). This methodology of estimating CIWP/CIWC was used in our previous CMIP3 model-data comparisons [e.g., Li et al., 2008; Waliser et al. 2009] and for model cloud parameterizations improvements in ECMWF Integrated Forecast System (IFS) (R. M. Forbes, personal communication, 2011), National Aeronautics and Space Administration/The Goddard Earth Observing System Model, Version 5 (NASA/GEOS5) GCM, NASA Goddard finite-volume multiscale modeling framework (fvMMF) GCM [Chen et al., 2011], National Center for Atmospheric Research, Community Atmosphere Model, version 5 (NCAR CAM5) [Gettelman et al., 2010; Song et al., 2012], Geophysical Fluid Dynamics Laboratory, Coupled Model, version 3(GFDL CM3) [Donner et al., 2011], and University of California at Los Angeles (UCLA) GCM [Ma et al., 2012, also Sensitivity of global tropical climate to land surface processes: Mean state and interannual variability, submitted to *Journal of Climate*, 2012].

[15] 2. *PSD method* [Chen et al., 2011]. We use the ice Particle Size Distribution (PSD) parameters associated with

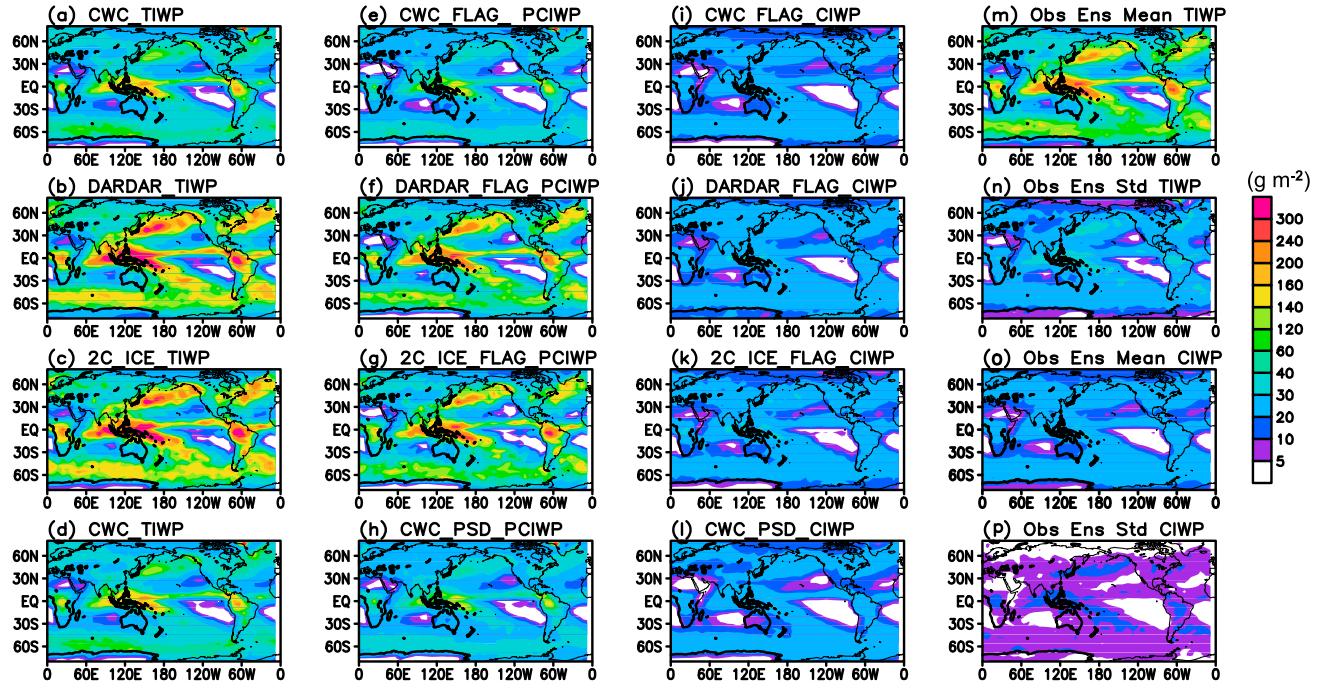


Figure 1. Annual mean maps of total ice water path (TIWP = cloud + precipitating) estimated from (a) CWC; (b) DARDAR; (c) 2C_ICE; and (d) CWC (repeated from Figure 1a) satellite products. CWC is derived from CloudSat radar only, while 2C_ICE and DARDAR are derived from both the CloudSat radar and CALIPSO lidar. See section 2 for details, including periods for time-mean averages. (e–h) The same as Figures 1a–1d, but for estimates of the precipitating ice water path (PIWP). Figures 1e–1g calculate this estimate based on surface precipitation and/or convective cloud flags and Figure 1h utilizes a method based on the particle size distribution (PSD) and a $150 \mu\text{m}$ cutoff. Further details and references for these methods are given in section 2. (i–l) The same as Figures 1e–1h, except for estimates of the cloud ice water path (CIWP). (m and n) Ensemble mean and standard deviation of Figures 1a–1d; (o and p) ensemble mean and standard deviation of Figures 1i–1l. Figures 1m and 1o are observed estimates of TIWP and CIWP used in this study. Unit is g m^{-2} .

each CloudSat retrieval to separate the total IWC into mass with particle sizes smaller and larger than a selected particle size threshold. Based on the analysis in *Chen et al. [2011]* and references therein, the size separation of cloud ice and precipitating ice on a global mean basis likely falls between $100 \mu\text{m}$ and $200 \mu\text{m}$ in diameter. A threshold of $150 \mu\text{m}$ is chosen for the present study, and the integrated mass of particles with diameter smaller than this size is considered representative of the CIWC/CIWP. In this case, such estimates are based on a quantitative, microphysical characterization (i.e., PSD) regardless of the presence of surface precipitation or cloud type, and thus the vertical distributions of cloud ice versus precipitating ice mass can be derived from each CloudSat profile. The CIWC derived by this method has been shown to agree well with estimates based on the FLAG method, and have been applied to evaluate the atmospheric ice in the ECMWF IFS and NASA Goddard fvMMF GCM [*Chen et al., 2011*].

[16] It should be underscored that with present satellite/retrieval technology, it is not possible to absolutely separate floating/cloudy forms of ice from falling/precipitating forms, yet models often try to make this distinction. Specific retrievals of this sort will require co-located vertical velocity information, such as from a Doppler radar capability, and/or multiple frequency radar to better characterize particle size.

[17] To account for observational uncertainty in our study, we produce four different estimates of “cloud” ice (i.e., CIWP/CIWC) from the three retrieval products and two precipitation/convection filtering methods described above. These include the FLAG method applied to all three of the retrieval products as well as the PSD method applied to the 2B-CWC-RO4 product. We use the ensemble mean of these four estimates as the “observed” or “reference” values, hereafter referred to as such, and the spread among the four estimates as a measure of observational uncertainty.

[18] Figure 1 shows long-term annual mean maps of IWP quantities associated with our observational estimates. The four columns represent estimates of total ice (TIWP: Figures 1a–1d), precipitating and convective ice (PCIWP: Figures 1e–1h), cloud ice (CIWP: Figures 1i–1l) and ensemble information (Figures 1m–1p). Overall, it is evident that the cloud ice (Figures 1i–1l) represents a smaller contribution to the total ice mass (Figures 1a–1d) than the precipitating/convective contribution (Figures 1e–1h), ranging from 10 to 30% depending on the product and location. It accounts for a smaller contribution in the two radar + lidar products (Figures 1b, 1c, 1f, 1g, 1j, 1k, 1n, and 1o) and in the tropics and storm track regions in all products. In general, the CIWP estimates typically agree relatively well, typically within a factor of two, and most of the

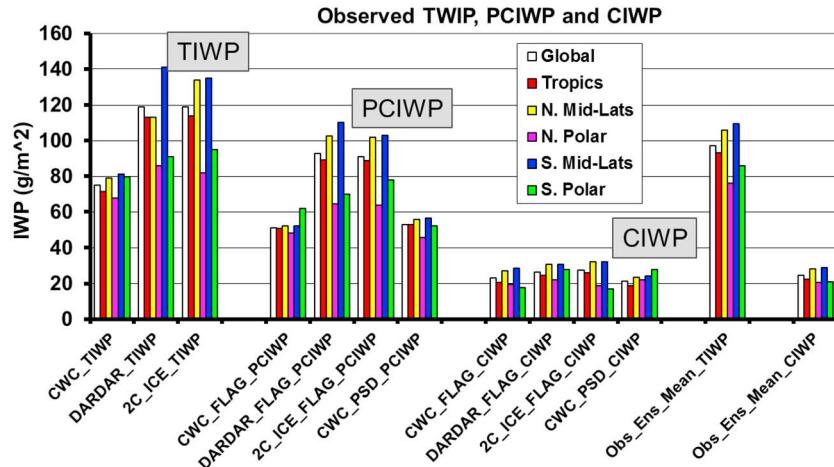


Figure 2. Bar charts of regional averages (see legend) of IWP (g m^{-2}), with the same data sets and methods as described in Figure 1. Here, “Global” is defined as 80°N – 80°S , “Tropics” is defined as 30°N – 30°S , “N. Mid-Lats” is defined as 30°N – 60°N , “S. Mid-Lats” defined as 30°S – 60°S , “N. Polar” is defined as 60°N – 80°N , and “S. Polar” is defined as 60°S – 80°S .

differences can be explained by the different microphysical assumptions [Delanoë *et al.*, 2011; Deng *et al.*, submitted manuscript, 2012]. The difference in mass-size, area-size relations and the cloud occurrence identified criteria between CloudSat 2C-ICE and DARDAR contributes to some subtle differences between those two data sets (Deng *et al.*, submitted manuscript, 2012). Figures 1m and 1n show the ensemble mean and standard deviation of the three observational estimates of TIWP (Figures 1a–1c), while the same is shown in Figures 1o and 1p for the four estimates of CIWP. It is the latter two maps, which are based on the four individual estimates in Figures 1i–1l, that represent the observational basis for CIWP and the GCM evaluations in this study.

[19] Figure 2 shows a synthesis of sorts of Figure 1, with bar charts of regional averages of TIWP, PCIWP and CIWP, which demonstrates that irrespective of the retrieval method, cloud/precipitation-filtering method, or region, there is generally good agreement among the different estimates of CIWP. That said, there are some latitudinal bands where the four estimates of CIWP can differ up to 30% or so (e.g., N. Midlatitudes, S. Polar) and thus it is prudent that this uncertainty be accounted for in the model evaluations.

[20] Figure 3 is similar to Figure 1, but displays the data as zonally averaged annual mean IWC as a function of height, rather than vertically integrated IWP. The general commonalities and differences between the products and different filtering methods are the same as described above for CIWP and Figure 1. Apart from that, the overall vertical structure of IWC in each exhibits three local maxima; one is in the tropics near 300 hPa and is associated with deep convection, and the other two are in the northern and southern midlatitudes at approximately 600 hPa and correspond to the storm tracks. In general, these maxima tend to be lower in the two radar + lidar products (2nd and 3rd row) and higher in the radar-only products (1st and 4th row). Moreover, there is a tendency for the maxima to be higher in the CIWC profiles (3rd column) compared to the PCIWC profiles (2nd column), particularly in the midlatitudes,

which has some intuitive merit as the larger, precipitating particles should be biased in the cloud(s) present at lower altitudes. As with Figure 1, the four estimates of CIWC (3rd column) and their ensemble information (Figures 3o and 3p) are used to evaluate the model/reanalysis representations of CIWC. The information in Figures 1m, 1n and 3m and 3n are used to compare with two GCMs (i.e., GFDL AM3, CM3) examined in this study that provide outputs of TIWC/CIWP.

[21] Apart from the uncertainty of the retrieval method, an additional uncertainty to consider in light of making model-observation comparisons concerns the differences in the spatial and temporal sampling between the observations and the GCMs, such as those in the CMIP archives. The former are based on the suborbital footprint of the CloudSat and CALIPSO sensors as they fly in the formation in the polar-orbiting A-train constellation, while the latter are well- and regularly sampled means across the diurnal cycle and the globe. To estimate the impact of this sampling, we utilize the 3-hourly IWC at a number of levels from Modern Era Retrospective-analysis for Research and Applications (MERRA) and sample them along the A-Train orbit to examine the sampling bias. The unsampled MERRA values represent the true state of the geophysical variable and the sampled data represent estimates of the time-space subset of the variable observable from the A-Train (i.e., like the actual observed values are). From these two data sets, two monthly IWC climatologies are then created and compared in the Taylor diagram shown in Figure 4 (red dots). The results show that the correlations between the two data sets are greater than 0.9 at all vertical levels examined. Standard deviation ratios fall between 1 and 1.2 for all vertical levels indicating that the sampled exhibit slightly more variability – which in this case would be reflective of sampling noise. This comparison suggests that the two climatologies closely resemble each other, particularly when considering the differences that are exhibited between the various GCMs examined in this study and their difference with the observed estimates that are shown in the next section. As a result of the above test –

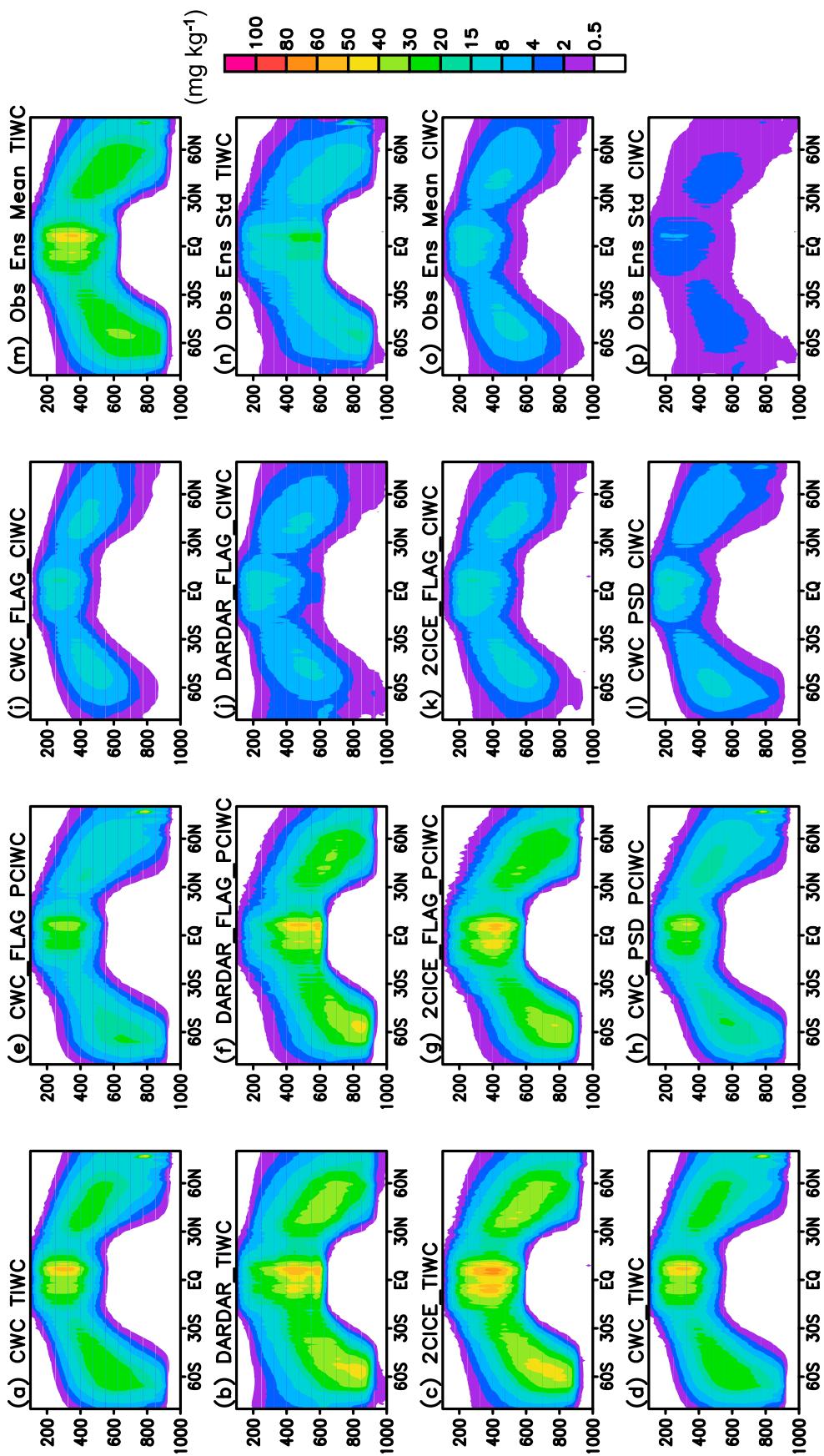


Figure 3. Zonal mean averages of ice water content (IWC) for each of the data sets and methods shown in Figure 1. Unit is mg kg^{-1} .

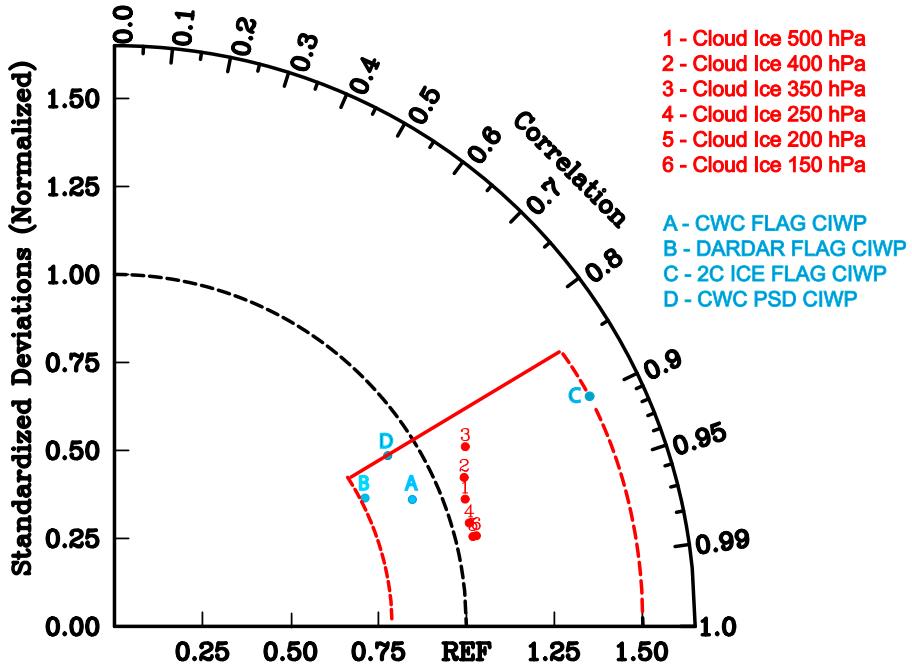


Figure 4. Taylor diagram illustrating two different components of observational uncertainty for the data sets and methodology used in this study (see section 2). Red dots illustrate the impact of subsampling of the climatologies associated with the A-Train and CloudSat/CALIPSO sensors. Here the reference is the complete, diurnally resolved (every 3 h) MERRA reanalysis values of CIWC. The dots represent comparisons of this well-sampled climatology against the same MERRA data set that has been re-sampled data along the A-Train ground track. Blue dots illustrate the uncertainty associated with four different observational estimates of CIWC (see section 2) including CWC_FLAG_CIWP, DARDAR_FLAG_CIWP and CWC_PSD_CIWP. In this case the reference is ensemble mean of these four observed estimates. The red bounding box represents the outer bounds of these various uncertainty estimates, and thus an overall observational uncertainty for this study. Specifically, the values used for observational uncertainty are a lower bound of 0.85 for correlation (red solid line) and a range of 0.8 to 1.5 for standard deviation ratios (red dashed curves). These Taylor metric values are used to discriminate models that exhibit relatively good model performance (see Figures 7 and 12 and Figure 13).

which does assume some fidelity by MERRA in representing the diurnal cycle, it is plausible to compare the observed, satellite-sampled, ice water estimates to those from the GCMs without the need to sample the GCMs along the A-Train satellite track. Also shown in Figure 4 are the four observed CIWP estimates (blue dots) shown in Figures 11i–11 plotted against their ensemble mean. Taken together, these sampling and retrieval uncertainties provide some quantitative information on the overall uncertainty in our observational estimate. This uncertainty is shown as the red outline in Figure 4, which is roughly delineated by a lower bound of 0.85 for correlation and a range of 0.8 to 1.5 for the standard deviation ratio. The high standard deviation of 2C-ICE in Figure 4 could be caused by the less data volume compared to other observation data set. In Taylor plots illustrated below, we will use these criteria as observational uncertainty and utilize it as a guideline for good model performance.

3. Modeled Values of IWC and IWP

[22] Using the observations described in the previous section, we evaluate CIWP/CIWC in reanalysis data sets, including ECMWF (ERA-Interim) [Dee et al., 2011] and NASA MERRA analyses, coupled atmosphere-ocean GCMs (CGCMs) from the CMIP3 (for CIWP only), CGCMs from

CMIP5, and two additional state-of-the-art GCMs, including the UCLA CGCM (Ma et al., submitted manuscript, 2012) and the NASA GEOS5 GCM. The CMIP3 simulations are the same as those described in Waliser et al. [2009] – although excluding the two UKMO models which we have since learned provided output on CIWP that was inconsistent with the experimental protocol's specifications [cf. Li et al., 2011]. Note that CMIP3 model output did not include CIWC. The CMIP5 simulations are listed in Table 1a. Table 1b is an outline of cloud microphysics parameterizations used in the selected CMIP5 models, as well as UCLA CGCM, GEOS5, MERRA and ERA-Interim model. Unlike that in a single column model intercomparison [e.g., Ghan et al., 2000; Xie et al., 2002, 2005; Xu et al., 2002, 2005; Klein et al., 2009; Morrison et al., 2009] which is conducted with the same large-scale conditions and forcing for each model, an undertaking to elicit the reason for the different model behaviors is more plausible within the scope of a single paper. For CMIP5 models, however, in this study, the performance of simulated cloud properties arises from a fully coupled system (land, ocean, atmosphere, etc.) and the behavior is not likely to be simply explained by any single component/scheme of, but rather by details of the model's specific schemes and the coupling among schemes related

Table 1a. Institution, Model Resolution and Label for the CMIP5 GCMs Examined in This Study

Model Label	Model Resolution	Institution/Model Name
BCC-CSM1	64 × 128 × 26	Beijing Climate Center, China/ BCC-CSM1-1
BCC	64 × 128 × 26	Beijing Climate Center, China/ BCC-CSM1-1_esm
CanESM2	64 × 128 × 35	Canadian Centre for Climate Modeling and Analysis, Canada/CanESM2
CCSM4	288 × 192 × 26	National Center for Atmospheric Research, USA/CCSM4
CNRM-CM5	128 × 256 × 17	Centre National de Recherches Meteorologiques, France/CNRM-CM5
CSIRO	96 × 192 × 18	Australian Commonwealth Scientific and Industrial Research Organization, Australia/CSIRO-Mk3-6-0
GISS-E2-H	90 × 144 × 29	NASA / Goddard Institute for Space Studies, USA/GISS-E2-H
GISS-E2-R	90 × 144 × 29	NASA / Goddard Institute for Space Studies, USA/GISS-E2-R
HadGEM2-ES	145 × 192 × 38	Hadley Centre for Climate Prediction and Research / Met Office, UK/ HadGEM2-ES
Inmcm4	120 × 180 × 21	Institute for Numerical Mathematics, Russian/ Inmcm4
Inmcm4ESM	120 × 180 × 21	Institute for Numerical Mathematics, Russian/ Inmcm4_esm
IPSL	96 × 96 × 39	Institute Pierre Simon Laplace, France/ IPSL-CM5A-LR
MIROC	64 × 128 × 80	University of Tokyo, NIES, and JAMSTEC, Japan/ MIROC-ESM-CHEM
MIROC4h	320 × 640 × 56	University of Tokyo, NIES, and JAMSTEC, Japan/ MIROC4h
MIROC5	128 × 256 × 40	University of Tokyo, NIES, and JAMSTEC, Japan/ MIROC5
MRI-CGCM3	160 × 320 × 35	Meteorological Research Institute, Japan/ MRI-CGCM3
NorESM	96 × 144 × 26	Norwegian Climate Centre, Norway/ NorESM1-M

with a particular process such as cloud and turbulence for boundary layer clouds, and cloud and convection for deeper clouds as well as the interactions with SSTs. However, material on attempting to explain the behavior of some of the best/worst performing models is discussed in the Summary. This includes for example discussion regarding the specifics associated with GFDL-CM3, GISS, MRI, UCLA CGCM and CAM5. In some cases, these are updates relative to the CMIP5 and thus provide possible explanations for good/poor model behavior and thus insight that modeling groups may find useful in improving their model physics. The specific experimental scenario is the historical 20th century simulation, which used observed 20th century greenhouse gas, ozone, aerosol and solar forcing. The time period used for the long-term mean is 1970–2005, and if a model provided an ensemble of simulations, only one of them was chosen for this evaluation. Unlike all other models examined in this study, which do not include ice mass from convective-type clouds in their CIWC, the two GFDL models include grid means over shallow cumulus, deep cumulus cells, and convective mesoscale clouds, weighted by their respective area fractions. In the GFDL-CM3, precipitating ice, however, that has fallen out of large-scale stratiform clouds and into clear areas is not included. Thus, the GFDL models should be considered somewhat carefully with respect to the others as they are including cloud mass from clouds whose contribution have been typically ignored, and their IWC/IWP fields would be more commensurate with TIWC/TIWP. In the CSIRO, diagnostic falling precipitation is considered while the convective-type clouds of cloud hydrometeors are not included. Thus, the CSIRO model somewhat should be considered between the cloud only and total ice water content/path. For both the GCM and observational data sets, all fields have been re-gridded to 40 levels (with a constant pressure interval of 25 hPa) and mapped onto common 8° × 4° longitude by latitude grids.

4. Results

[23] Figures 5a–5z show the long-term annual mean spatial distributions of simulated values of CIWP from fifteen

CMIP5 CGCMs (see Table 1a and Figures 5a–5o), the multimodel ensemble mean from the fifteen CMIP5 models (Figure 5p), and GEOS5 (Figure 5s), UCLA CGCM (Figure 5t), and two analyses including ECMWF-Interim (Figure 5u) and MERRA (Figure 5v) as well as the ensemble mean (Figure 5y) and standard deviation (Figure 5z) of the four observed estimates of CIWP discussed above. Overall, the multimodel mean CMIP5 CIWP values are spatially similar to observations but nonetheless are biased high. Individually, most models tend to qualitatively capture the global and regional CIWP patterns. This includes the relatively high values of CIWP in the ITCZ, warm pool, and storm tracks from the subtropics to high latitudes, and over convectively active continental areas over central Africa and South America. Note that the relative magnitudes between tropical and midlatitude values can be quite different across models; this will be more evident when discussing Figure 8 below. About three of the CMIP5 models do a good job at representing both the observed patterns and magnitudes of CIWP (i.e., CNRM-CM5, CanESM2, MRI). A number of models, however, significantly (~factor of 2) underestimate tropical CIWP (i.e., NorESM, BCC, BCC-CSM1, CCSM4) and two severely (~factor of 10) underestimate CIWP (i.e., Inmcm4, Inmcm4ESM). The two GISS GCMs greatly overestimate (~factor of 5) tropical CIWP. The IPSL, CSIRO, MIROC5, MIROC4h and the two GISS GCMs moderately overestimate CIWP in the extra-tropics. For the non-CMIP5 GCMs, the GEOS5 AGCM significantly underestimates (~factor of 3) CIWP in the storm tracks while the UCLA CGCM does remarkably well over most of the globe. The two reanalyses, ECMWF and MERRA, show relatively good CIWP patterns and magnitudes, with MERRA being biased a bit low in midlatitudes which is not surprising given the base model (GEOS5) exhibits such a strong negative bias. While the above model-observation differences are still substantial in many regards, it is worth noting that the ensemble of CMIP5 CIWP values examined here, appear to exhibit improvement compared to the ensemble CMIP3 models evaluated in our previous study (Appendix A Figure A1); this will be discussed and

Table 1b. Outline of Cloud Microphysics Parameterizations Used in the Selected CMIP5 Models, as Well as UCLA CGCM, GEOS5, MERRA and ERA-Interim Model

Models	Prognostic Cloud Variables	Bulk Single Moment or Double Moment	Cloud Fraction (PDF Based or Non-PDF Based)		References
			Diagnostic,	Non-PDF based, 0 or 1 Prognostic cloud fraction; PDF-based	
UCLA CGCM	Mixing ratio of cloud liquid and ice	Bulk single moment	Köhler [1999]; Ma et al. [2012, also submitted manuscript, 2012]; Tiedtke [1993]; Dee et al. [2011]		
ERA-Interim (CY31rl)	Mixing ratio of cloud condensate	Bulk single moment; mixing ratio of cloud condensate with temperature dependent partitioning (The bounds are adjustable constants with current settings ice at $T = -15^{\circ}\text{C}$ and liquid at $T = 0^{\circ}\text{C}$).			
GEOS5	Single mixing ratio of cloud condensate	Bulk single moment; mixing ratio of cloud condensate with temperature dependent partitioning; “anvil” cloud, originates in detaining convection. “large-scale cloud,” originates in a probability distribution function (PDF) based condensation calculation.			Rienecker et al. [2008]
MERRA	Single mixing ratio of cloud condensate	Bulk single moment; mixing ratio of cloud condensate with temperature dependent partitioning; “anvil” cloud, originates in detaining convection. “large-scale cloud,” originates in a probability distribution function (PDF) based condensation calculation.			Rienecker et al. [2008]
BCC-CSM1	Single mixing ratio of total water	Bulk single moment; mixing ratio of cloud condensate with temperature dependent partitioning (The bounds are adjustable constants with current settings ice at $T = -40^{\circ}\text{C}$ and liquid at $T = -10^{\circ}\text{C}$).			Collins et al. [2004]; Rasch and Kristjánsson [1998]; Zhang et al. [2003]; Sundqvist [1988]
BCC	Single mixing ratio of total water	Bulk single moment; mixing ratio of cloud condensate with temperature dependent partitioning (The bounds are adjustable constants with current settings ice at $T = -40^{\circ}\text{C}$ and liquid at $T = -10^{\circ}\text{C}$).			Collins et al. [2004]; Rasch and Kristjánsson [1998]; Zhang et al. [2003]; Sundqvist [1988]; Wu et al. [2010]
CanESM2	Mixing ratio of cloud liquid and ice	Bulk single moment			Lohmann and Roeckner [1996]; Charbouau and Bechtold [2005]
CCSM4	Single mixing ratio of total water	Bulk single moment; mixing ratio of cloud condensate with temperature dependent partitioning (The bounds are adjustable constants with current settings ice at $T = -40^{\circ}\text{C}$ and liquid at $T = -10^{\circ}\text{C}$).			Collins et al. [2004]; Rasch and Kristjánsson [1998]; Zhang et al. [2003]; Boville et al. [2006]
CNRM-CM5	Diagnostic cloud liquid water, PDF based	Statistical formulation			Bougeault [1981, 1982]; Mellor [1977]; Ricard and Royer [1993] Del Genio et al. [1996]
GISS-E2-H	Single mixing ratio of total water Diagnostic falling snow	Bulk single moment; mixing ratio of cloud condensate with temperature dependent partitioning (The bounds are adjustable constants with current settings ice at $T = -35^{\circ}\text{C}$ and liquid at $T = -4^{\circ}\text{C}$ over ocean; $T = -35^{\circ}\text{C}$ and liquid at $T = -10^{\circ}\text{C}$ over land).			Diagnostic, non-PDF based

Table 1b. (continued)

Models	Prognostic Cloud Variables	Bulk Single Moment or Double Moment	Cloud Fraction (PDF Based or Non-PDF Based)	References
GISS-E2-R	Single mixing ratio of total water Diagnostic falling snow	Bulk single moment; mixing ratio of cloud condensate (The bounds are adjustable constants with temperature dependent partitioning with current settings ice T = -35°C and liquid at T = -4°C over ocean; T = -35°C and liquid at T = -10°C over land).	Diagnostic, non-PDF based	<i>Del Genio et al. [1996]</i>
HadGEM2-ES	cloud ice water content (both cloud ice and snow); Mixing ratio of cloud liquid and ice	Bulk single moment; different phases determined by particle size information and temperature Bulk single moment; Large scale condensation in the case of relative humidity exceeds 1.	Diagnostic, non-PDF based	<i>Wilson and Ballard [1999]; Volodin et al. [2010]</i>
Inmcm4	Mixing ratio of cloud liquid and ice	Bulk single moment; Large scale condensation in the case of relative humidity exceeds 1.	Diagnostic, non-PDF based	<i>Volodin et al. [2010]</i>
Inmcm4ESM	Single mixing ratio of total water	Bulk single moment; mixing ratio of cloud condensate with temperature dependent partitioning (The bounds are adjustable constants with current settings ice at T = -15°C and liquid at T = 0°C). Bulk single moment; different phases determined by temperature	Diagnostic PDF based	<i>Bony and Emmanuel [2001]</i>
IPSL-CM5A-LR		Bulk single moment; different phases determined by temperature	Diagnostic PDF scheme with minor change for calculating anvil cloud	<i>Ogura et al. [2008]; Le Treut and Li [1991]; Hourdin et al. [2006]; Ogura et al. [2008]; Le Treut and Li [1991]</i>
MIROC(MIROC-ESM; MIROC-ESM-CHEM)	Mixing ratio of cloud liquid and ice	Bulk single moment; CCN is prognostic	Diagnostic PDF scheme with minor change for calculating anvil cloud	<i>Wilson and Ballard [1999]; Watanabe et al. [2009]; Tiedtke [1993]; Yukimoto et al. [2012]; Rasch and Kristjansson [1998]; Zhang et al. [2003]; Boville et al. [2006]</i>
MIROC4h	Mixing ratio of cloud liquid and ice	Double moment scheme.	Diagnostic PDF based	
MIROC5	Mixing ratio of cloud liquid and ice	Bulk single moment; mixing ratio of cloud condensate with temperature dependent partitioning (The bounds are adjustable constants with current settings ice at T = -40°C and liquid at T = -10°C).	Diagnostic, non-PDF based	
MRI-CGCM3	Mixing ratio of cloud liquid and ice	Bulk single moment	Prognostic cloud fraction; PDF based; The Tiedtke (1993) parameterization includes among its assumptions a simple PDF for humidity in clear air (uniform distribution).	<i>Tiedtke [1993]; Röistäy [1997]; Dommer [1993]</i>
NorESM1	Single mixing ratio of total water		The Dommer (1993) cumulus parameterization generates a discrete PDF (defined by its ensemble of cumulus elements), each member of which has a joint distribution of thermodynamic and microphysics properties and vertical velocities.	
GFDL-CM3	Mixing ratio of cloud liquid and ice			

Table 1b. (continued)

Models	Prognostic Cloud Variables	Bulk Single Moment or Double Moment	Cloud Fraction (PDF Based or Non-PDF Based)	References
CSIRO-Mk3.6.0	Mixing ratio of cloud liquid and ice; Diagnostic falling snow	Bulk single moment; ice crystal number concentration is diagnosed; mixing ratio of cloud condensate with temperature dependent partitioning (The bounds are adjustable constants with current setting ice at $T = -40^{\circ}\text{C}$);	Diagnostic, non-PDF based	Rotstayn [1997] Rotstayn <i>et al.</i> [2000]

quantified further below. The two GFDL models (Figures 5q and 5r) that simulate and provide output on TIWC each exhibit fairly good TIWP in the tropical ITCZ, warm pool and convectively active continental regions, but significantly underestimate TIWP in the extra-tropics storm track regions compared to ensemble mean TIWC shown in Figure 5w.

[24] Figure 5aa shows a bar chart illustrating CMIP5 CGCM global and regional CIWP averages for midlatitudes (30° N/S – 60° N/S, tropical (30° N 30° S) and polar latitudinal bands (60° – 80° N/S). Included are CMIP5 CGCMs, multimodel CMIP5 and CMIP3 CGCM ensemble means, GEOS5, UCLA AGCM, UCLA CGCM, ECMWF-Interim and MERRA, and also observed ensemble means. It is evident, and expected, that the model disagreement with observations is larger for regional values. For example, the GISS, IPSL, MIROC4h, MIROC5 and CSIRO models have extra-tropical IWP values that are larger than tropical values by often factors of two or more. The two Inmcm4 and GEOS5 models exhibit CIWP underestimates for all latitudinal bands. Interestingly, while the two GISS CGCMs grossly overestimate the CIWP in the tropics for CMIP5, with moderate overestimates for the extratropics, the opposite was the case for CMIP3 as shown in Figure A1. With the exception of the tropical values, the multimodel mean CIWP values from CMIP5, for the different latitudinal bands shown, are closer to observations than those from CMIP3. This demonstrates a quantitative improvement of CIWP simulations from CMIP3 to CMIP5; this will be further demonstrated below. The two GFDL models simulate TIWP well in tropical latitudinal bands but underestimate TIWP by 50% or more in the extra-tropical storm tracks and polar latitudinal bands.

[25] To summarize the multimodel performance of CMIP3 and CMIP5 in representing the time-mean pattern of CIWP, Figures 6a–6d illustrate the multimodel mean biases against the observed estimate and the root mean square error (RMSE) (Figure 6e) calculated across the models for each of the ensembles. In CMIP3, the high latitudes were biased high, while in CMIP5 the bias is more uniform. In the case of the latter, only the subtropics and deserts exhibit a low bias due, in part, to the relatively small cloud ice values in these regions. While the bias figure emphasizes the sign and magnitude of systematic biases across the two model archives, the RMSE figure emphasizes systematic errors irrespective of the sign, with larger errors emphasized as well because of the squaring operation. Again, high latitudes were more systematically incorrect in CMIP3 while low-latitudes were more problematic in CMIP5. However, we would like to point out the impact of the more extreme outliers on these figures. Examination of Figure 5 indicates that the GISS model is an extreme outlier, and biased very high, relative to the ensemble of CMIP5 models examined here. In addition, because there are two different versions of it analyzed here that do not differ much in regards to their CIWP representation, the effect of this model's extreme bias is doubled. Similarly, it is the case for the two INMCM GCMs but with a large negative bias of CIWP. Figures 6e and 6g show the impact on the spatial patterns of mean bias and RMSE if the two GISS models are removed from the ensemble, while the two right panels show the impact if the two GISS and the two INMCM models are removed. In this case, the bias of the CMIP5 ensemble analyzed here

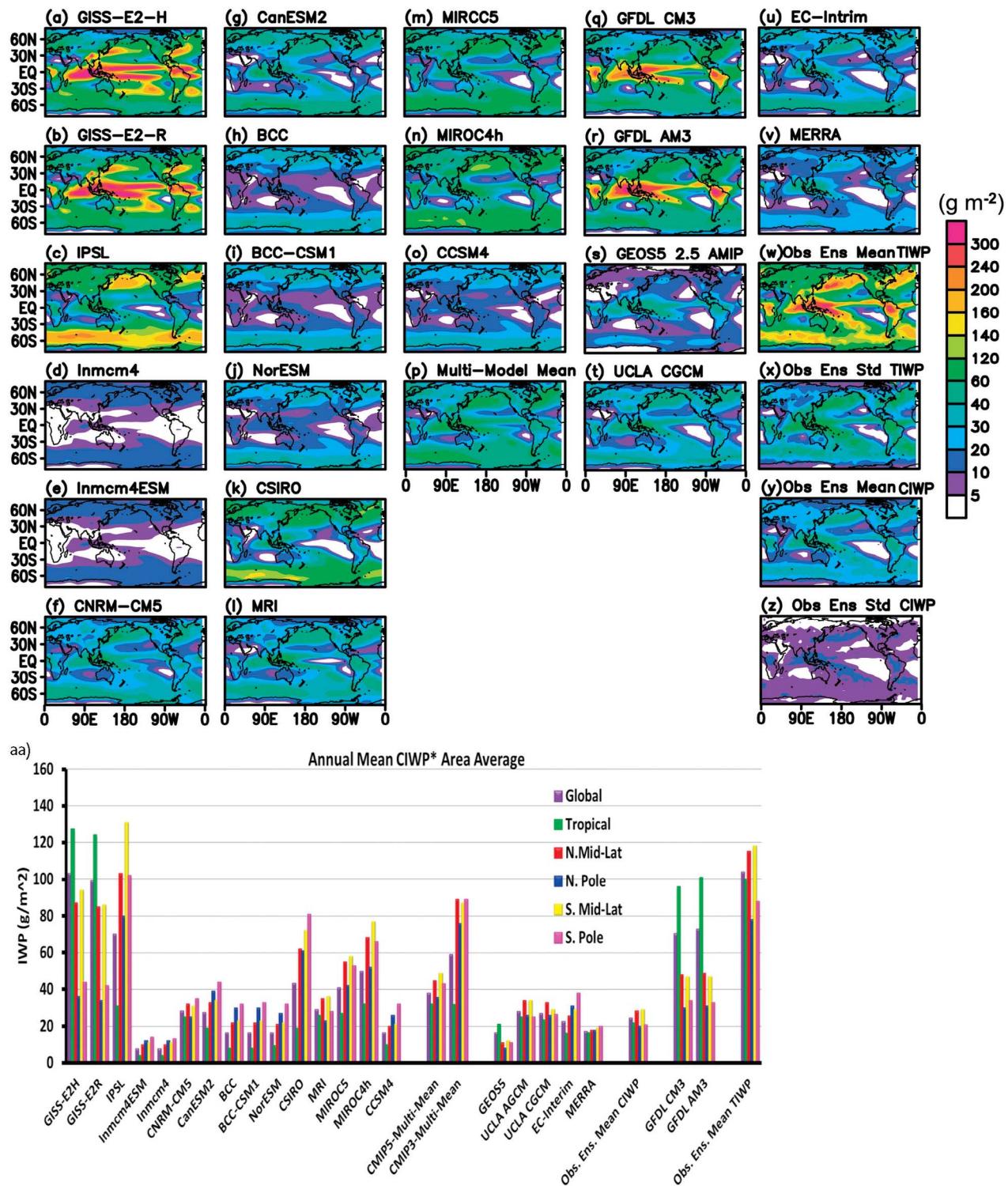


Figure 5. (a–o) Annual mean maps of ice water path (IWP; g m^{-2}) from the 20th century GCM simulations contributed to the CMIP5, with (p) their multimodel mean, (q) GFDL CM3 TIWP, (r) GFDL AM3 TIWP, (s) NASA GEOS5 2.5 AMIP, (t) UCLA CGCM, (u) ECMWF-Interim, (v) MERRA reanalyses as well as (w) ensemble-mean and (x) standard deviation of three TIWP observed estimates, (y) ensemble-mean and observed estimate (z) standard deviation of four CIWP observed estimates (see section 2). (aa) Regionally averaged, annual mean values of CIWP (g m^{-2}) for CMIP3, CMIP5, MERRA, ECMWF-Interim, NASA-GEOS5 and the observed estimate. Note that all values shown in Figure 1aa are CIWP except for the three groups on the far right which are TIWP, i.e., GFDL CM3, GFDL AM3 and the observed estimate of TIWP.

changes from a large, somewhat uniform (except the subtropics), positive value ($\sim 20\text{--}60 \text{ g m}^{-2}$), to a somewhat smaller positive bias in the extra-tropics with an offsetting negative bias in the tropics. In addition, the RMSE gets reduced substantially across the tropics and to a lesser extent in high latitudes. This change in the pattern of the overall bias and RMSE is substantial. What is left resembles a more canonical pattern associated with the observed ITCZ and storm tracks, possibly indicating a more straightforward bias/error to try and reconcile and improve (i.e., the cloud patterns are roughly realistic, but tropical clouds are too infrequent or have too little ice with the opposite being the case for storm track clouds).

[26] To further quantify and synthesize the comparative information discussed above, we can use a Taylor diagram [Taylor, 2001] to summarize both the degree of agreement in overall CIWP spatial pattern correlations along with the absolute sizes of the biases among the CMIP5 CGCMs, including their multimodel mean, two analyses, three other GCMs and four observed CWIP estimates. As with Figure 4 the ensemble mean of the latter is used as the reference data set and their spread to help quantify observational uncertainty. The Taylor diagram relates three statistical measures of model fidelity: the “centered” root mean square error, the spatial correlation, and the spatial standard deviations. These statistics are calculated for the long-term time mean and over the global domain (area-weighted). The reference data set is plotted along the *x* axis at the value 1.0.

[27] Figure 7 shows Taylor diagrams for CMIP3 (Figure 7a) and CMIP5 (Figure 7b), as well as select information from both as bar charts (Figure 7c). The observed estimates are plotted in blue, the CMIP GCMs in red, their ensemble means in green and the reanalyses and non-CMIP GCMs are in black. The red rectangular-like region illustrates a measure of observational uncertainty developed and shown in conjunction with Figure 4. Not surprisingly, the four individual observed estimates, reanalyses and AGCM simulations (i.e., specified SST; GEOS5 version2.5) perform as a group considerably better than the CMIP coupled GCMs. The former all tend to have correlations at around 0.9 or better and standard deviation ratios of between about 0.8 and 1.5. For the CMIP values (red), most of them have correlations between about 0.4 and 0.7 with standard deviation ratios well above 1, with some well above 3 and even up to 5 (see Figure 7c). The CMIP3 and CMIP5 multimodel means do not exhibit the best overall performance relative to the individual models due to the few strong outliers in the ensembles. Noteworthy, however, is that the CMIP3 and CMIP5 multimodel means (green) have correlations of 0.54 and 0.76, and standard deviation ratios of about 3.1 and 1.4, respectively, indicating a rather considerable performance improvement from CMIP3 to CMIP5 for representing CWIP. While this progress is encouraging, keep in mind also that all models shown still exhibit a very poor correlation against the reference data set, with values less than 0.8 and none of the CMIP GCMs fall within the (red “box”) range of observational uncertainty.

[28] In regards to specific models, the best performing CMIP5 model by this metric is CNRM-CM5 with correlation and standard deviation ratios of about 0.8 and about 1.0, with MRI and CanESM2 the next best performers. As mentioned above, GEOS2.5 performs well relative to the

others in this group but it has the advantage in this case of being an AGCM-only run, and thus uses specified SSTs, while all other models examined here are fully coupled. Noteworthy in this regard, is the very good performance of the UCLA CGCM (non-CMIP5), with metric values nearly identical to the best performing CMIP5 GCM. The two GISS models have standard deviation ratios that are off the scale of the Taylor plot, with values of about six. Other poorly represented CIWP fields as measured by this metric are exhibited by the CMIP5 Inmcm4, Inmcm4ESM, IPSL and CSIRO models. Along the lines of the discussion above regarding the impact of the biggest outliers, we note that even if the two GISS models and the two INMCM models are removed, the multimodel ensemble still does not quite mimic the observations within the estimated range of observational uncertainty (see caption). While expected, it is still encouraging to see the relatively good performance of the ECMWF-Interim and MERRA CIWP (e.g., correlations ~ 0.8 and standard deviations $\sim 0.8\text{--}1.0$), as neither does any direct assimilation of cloud ice observations. In summary, while there is progress in the overall performance of CMIP5 relative to CMIP3, and there is at least one model that would be considered quite good (i.e., CNRM-CM5) although not within the observational uncertainty used here, these disparities demonstrate a need for continued work in refining the model CIWP representations.

[29] Next, we examine the fidelity of the models’ CIWC vertical structure. A comparison is given in Figure 8 shows the CIWC zonal and annual mean values from thirteen CMIP5 CGCMs (Note that the CNRM-CM5 CGCM CIWC is not available from the CMIP5 data portal at the time), GEOS5 AGCM (Figure 8q) and UCLA CGCM (Figure 8p), as well as ECMWF-Interim (Figure 8r) and MERRA (Figure 8s). These models provide output specifically on cloud ice. The two GFDL GCMs, on the other hand, are shown in Figures 8n and 8o and provide output for TIWC. Overall, there are significant disparities among the CMIP5 CGCMs against the observed ensemble mean (Figure 8v) with overall discrepancies ranging from multiplicative factors of about 0.25 of the observations (i.e., Inmcm4) to factors of 10 (i.e., GISS GCMs). Moreover, the general character of their vertical distributions with respect to pressure levels is considerably different. For example, the IPSL exhibits significant overestimates of CIWC over the storm track regions. About five of the CMIP5 models do a fair job at representing the vertical structure and magnitude of IWC (i.e., CanESM2 (Figure 8f), BCC-ESM1 (Figure 8g), NorESM1 (Figure 8h), MIROC5 (Figure 8k), MRI (Figure 8l) and CCSM4 (Figure 8m)). The rest of the models (CSIRO ((Figure 8i), MIROC4h ((Figure 8j), ECMWF-Interim ((Figure 8o)) generally tend to qualitatively capture the patterns but overestimate CIWC over midlatitudes and below 700 hPa. The GEOS5 model, on the other hand, tends to slightly overestimate CIWC in the tropics but significantly underestimates CIWC in the mid to high latitudes by about a factor of 2 to 3. The analyses from MERRA as well as the simulation from the UCLA CGCM show realistic CIWC vertically with values close to the observed ensemble mean albeit not extending as close to the surface when compared to the observed estimate. However, it is reasonable to exercise caution when considering the robustness of the observed values in these lower tropospheric regions, or anywhere

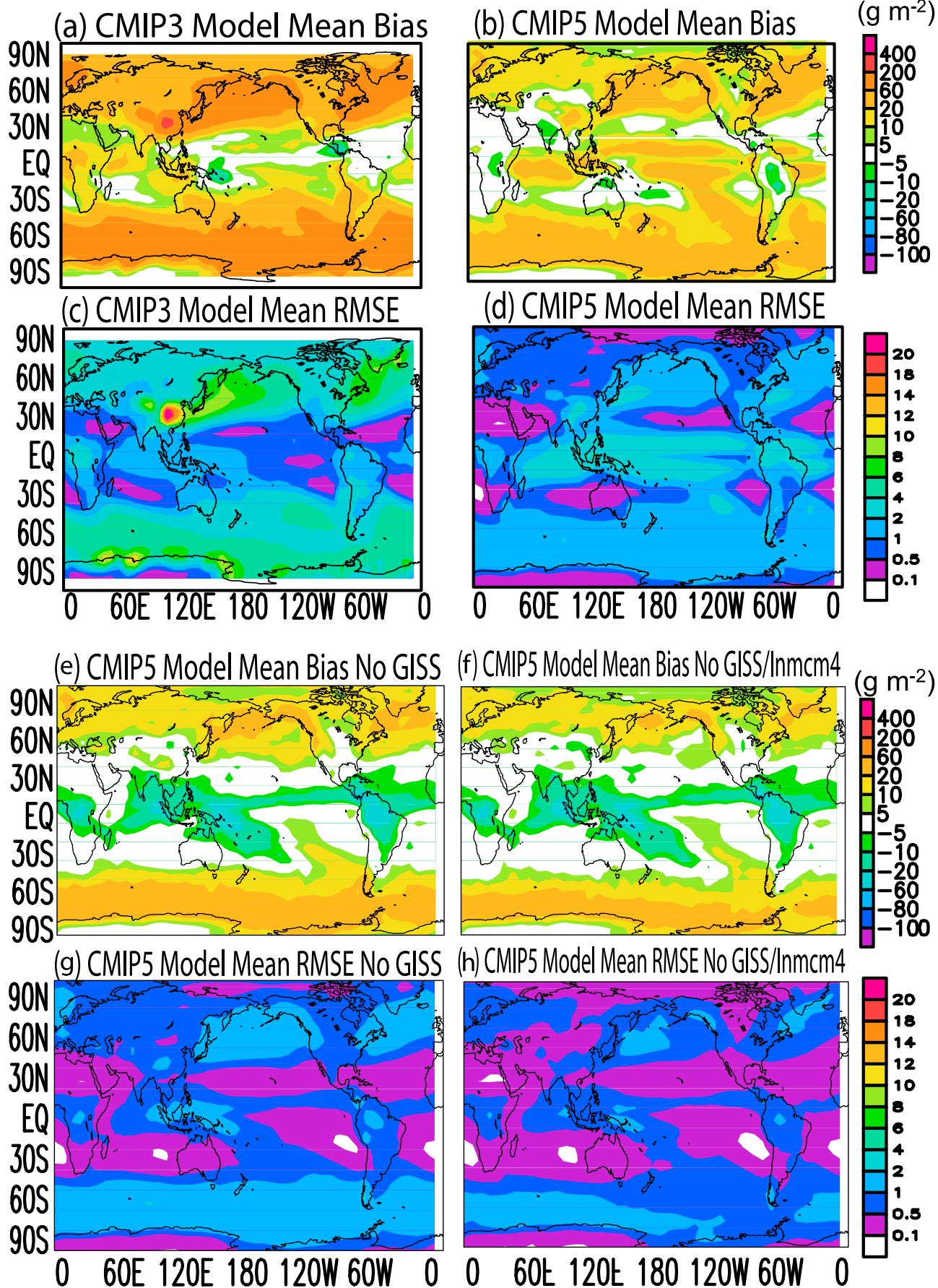


Figure 6

below the freezing level as there are artificial limitations applied to the retrievals that involve separating ice from liquid contributions. Compared to the observed TIWC (Figure 8s), the two GFDL models all capture the ITCZ in tropical regions pretty well but significantly underestimate TWIC in the extra-tropical storm track and Polar Regions. A realistic ITCZ is found in the GFDL uncoupled AGCM (Figure 8o) while a more notable double ITCZs is evident in the GFDL CGCM ((Figure 8n). Also, based on the caution above, the higher values in mid tropospheric tropics in these models relative to the observed value may not be an error. To summarize, apart from gross qualitative agreement with observations among many CMIP5 models, it is still apparent from Figures 4, 5, 6 and 8 that significant disparities exist not only horizontally but also in the vertical structure.

[30] Figure 9 summarizes some of the basic features of Figure 8 by showing the global (80°N - 80°S) mean vertical profiles of the models against observed ensemble mean CIWC (thick black). Again, the two GISS models and IPSL significantly overestimate CIWC while the two Inmcm4 models significantly underestimate CIWC at all levels. The best simulated CIWC vertical profile is from CanESM2 followed by the MRI, CCSM4, NorESM and BCC-ESM models (although keep in mind that CNRM-CM5 CGCM is not available here). Note that CanESM2 and MRI also performed very well in terms of horizontal structure. Both MERRA and GEOS5 overestimate CIWC in the upper troposphere, but values sharply decrease below 350 hPa. By decomposing Figure 9 into various regional averages (a subset of those used in Figure 5) shown in Figure 10, we find that the bias mainly comes from mid- and high-latitudes from both hemispheres, especially over the southern high-latitudes. Besides Inmcm4 and GEOS5, which are biased high relative to the observed estimate throughout the troposphere, the other models are biased high below 500 hPa, and increasingly so in the lowest levels. However, it is wise, as mentioned above, to exercise some caution regarding the observed estimates of cloud-ice at low levels. Simulation fidelity in the midlatitudes is similar to that in high-latitudes, although three models (BCC, CCSM4, NorESM and CanESM) compare quite favorably with observations in this region. Over the convectively active regions, the models generally are close to capturing the correct CIWC peak, but CIWC values vary greatly from model to model. Similar to Figure 10, Figure 11 is for GFDL CM3z global and regionally averaged vertical TIWC profiles in the tropics, northern mid- and high- latitudes (southern high-latitudes are similar). Compared to the ensemble mean TIWC (black profile), we find that the GFDL model (green) significantly underestimates TIWC throughout the troposphere in the mid- and high- latitudes. Over the convectively active regions, the model overestimates TIWC below 450 hPa but underestimates above 450 hPa. The IWC profiles spaghetti diagrams exhibited in Figure 9 is similar to that shown in Xu *et al.* [2002] and Xie *et al.* [2002]. They might be related to

cloud microphysics or convective parameterization used in each model.

[31] Finally, in order to determine if there are systematic biases across the models in the vertical structure of their CIWC fields, we examine the models', and their multimodel mean, bias and RMSE at each level including 700, 600, 500, 400, 300, 200, 150, and 100 hPa against the observed CIWC values. A Taylor diagram representing each pressure layer of the annual mean CIWC for the CMIP5 multimodel mean (red), MERRA (blue), and ECMWF-Interim (green) are shown in Figure 12. The information on Figure 12 is also given in Figure 13, which provides the Taylor diagram correlations (Figure 13a) and standard deviation ratios (Figure 13b) for the CMIP5 multimodel mean of CIWC at each pressure level for the CMIP5. In addition, Figure 13 also includes the values for each individual model in the CMIP5 ensemble analyzed here, along with the values for the UCLA and GEOS5 GCMs and the MERRA and ECMWF-Interim analyses, and correlation and standard deviation values for CIWP as well (left most column). For Figure 13, we utilize the combined sampling error estimates and observation data set sensitivity discussed in section 2 and shown in Figure 4 to classify reasonable model performance (pink boxes) relative to observed estimates for these Taylor diagram metrics. In this case, the threshold values for reasonable performance are 0.85 or better for correlation and 0.8 to 1.5 for standard deviation ratios.

[32] Considering Figure 12 and Figure 13, it is evident that CIWC is better simulated by most of the models in the lower, mid-tropospheric layers (i.e., correlation values are greater than 0.85 for 700, 600 and 500 hPa) and to some extent around 200 hPa. Interesting is that the models that tend to do well in terms of their pattern in the lower, mid-troposphere tend to more poorly simulated in the upper troposphere, and vice versa (see Figures 13a and 13b). All the models exhibit relatively poor scores at the 400 hPa level. From Figure 13b, it is evident that models do not uniformly overestimate or underestimate the magnitude of ice through the column. In some cases, the range of the standard deviation ratio for a given GCM across the different levels can easily be up to a factor of five (e.g., CSIRO, IPSL, BCC-ESM, GISS). In regards to CIWP, and as shown also in Figure 7c, a number of CMIP5 models exhibit good performance (i.e., shaded box) for the standard deviation ratio, while none of these models exhibit good performance for the correlation value – as indicated by the observational uncertainty level used here. Evident again here is the improvement in CMIP5 compared to CMIP3 in regards to both measures.

5. Summary and Discussion

[33] The objective of this study is to evaluate the representation of atmospheric ice in present-day GCMs, with an emphasis on: (1) the cloud portion of the ice (i.e., CIWC/CIWP) which is the quantity typically simulated and output by

Figure 6. (a) Multimodel mean bias (mg m^{-2}) for the CMIP3 GCMs shown in Figure A1, (b) multimodel mean bias for the CMIP5 GCMs in Table 1a, (c) same as Figure 6a, except for root mean square error (RMSE), (d) same as Figure 6b, except for RMSE. In all cases, the observed estimate is shown in Figure 10. (e and f) Same as Figures 6a and 6b, except excluding the two GISS GCMs. (g and h) Same as Figures 6c and 6d, except excluding the two GISS and two INMCM GCMs.

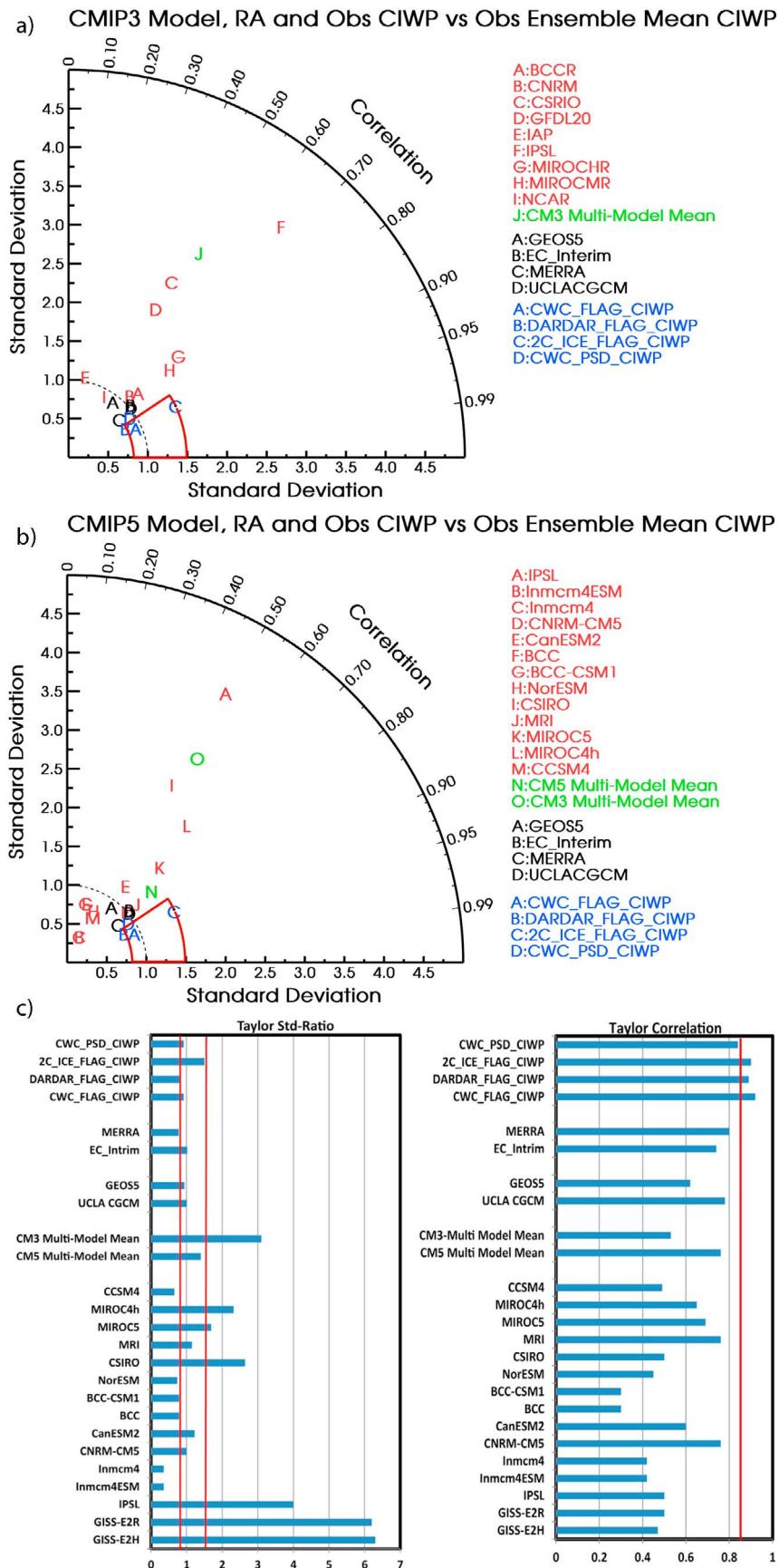


Figure 7

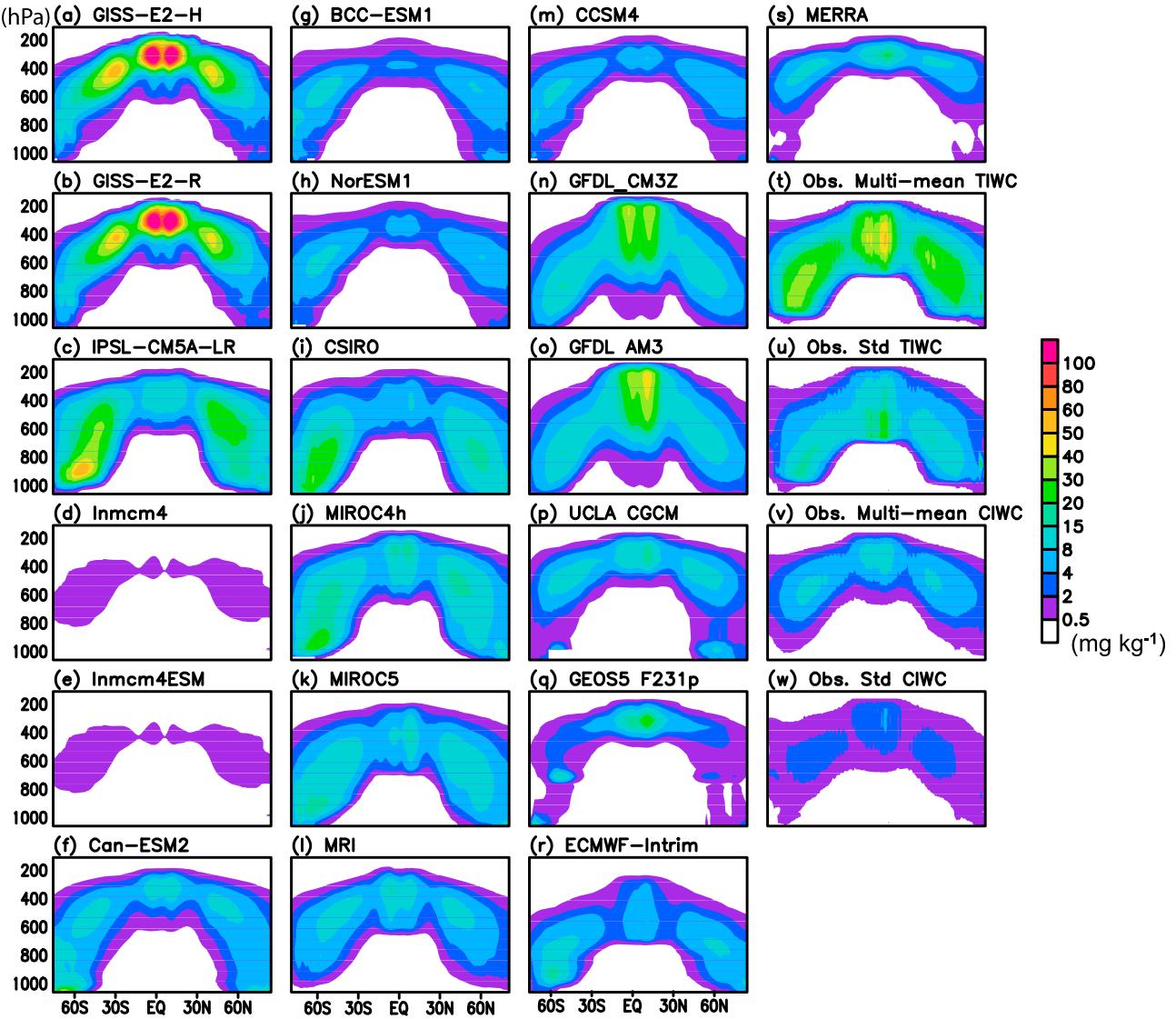


Figure 8. Zonally averaged annual mean (a–m) CIWC (mg kg^{-1}) from the 20th century GCM simulations contributed to the CMIP5, TIWC from GFDL (n) CM3 and (o) AM3, CIWC from (p) UCLA CGCM, (q) NASA GEOS5, (r) ECMWF-Interim reanalysis, (s) MERRA reanalyses as well as the (t) ensemble-mean and (u) standard deviation of three TIWP observed estimates, and the (v) ensemble-mean and (w) observed estimate standard deviation of four CIWP observed estimates (see section 2). Note that the CNRM-CM5 CGCM CIWC is not available from the CMIP5 data portal at the time.

GCMs and (2) model simulations of present-day climatology of CIWC/CIWP for CMIP5 and their comparison to CMIP3. Observational reference values and their uncertainty were addressed by using four different estimates of CIWC/CIWP

that accounted for different approaches to the retrieval and to the methods of filtering out the contribution to the ice mass in the retrievals due to large-particle/precipitating components which is a contribution to the mass that is typically not

Figure 7. (a) Taylor diagram for GCM annual mean (80°N – 80°S) cloud ice water path (CIWP; g m^{-2}), where the reference value is the observed estimate shown in Figure 1o. Modeled values include the CMIP3 GCMs (red) shown in Figure A1 and their multimodel mean (green), as well as the GEOS5 GCM, UCLA CGCM, and the MERRA and EC Interim analyses (black). Models with standard ratios that exceed 5 (GISS-E2-R and GISS-E2-H) are not shown in the Taylor diagram. Red box is defined in Figure 4 and discussed in section 2. (b) Same as Figure 7a, except for the CMIP5 GCMs instead of CMIP3, although the CMIP3 multimodel mean value is also replicated here. Note that the correlation and the standard deviation ratio of the CMIP5 multimodel mean when excluding the two GISS GCMs are 0.614 and 1.236, respectively. When excluding the two GISS and two Inmcm4 models, they are 0.645 and 1.09, respectively. For reference, when including all models, the values are 0.765 and 1.4, respectively. (c) Same as Figure 7b, except with the correlation and standard deviation ratios plotted in the form of a bar chart. The vertical lines represent the values of these quantities associated with observational uncertainty, taken from Figure 4, and which provide quantitative levels for indicating good model-observation agreement.

represented in GCMs as a prognostic or column-resolved quantity (see section 2). In addition, observational uncertainty also included considerations of the effects from the temporal sampling of the observations, i.e., the sun-synchronous, polar-orbiting A-Train satellites. The models evaluated (see section 3) included 15 simulations of present-day climate available to date in CMIP5 and two other GCMs of interest (GEOS-5 AGCM, and UCLA CGCM). The evaluation also included two modern reanalyses (MERRA and ECMWF-Interim). In addition, we included comparative evaluations of fourteen of the GCMs from CMIP3 that we analyzed in our previous study [Waliser et al., 2009], allowing for an assessment of whether GCM fidelity in regards to CIWP/CIWC has improved from CMIP3 to CMIP5.

[34] Overall, there is a fairly wide disparity in the fidelity of CIWP representations in the models examined. Even for the annual mean maps considered, there are easily factors of 2, and nearly up to 10, for the differences between observations and modeled values for most of the GCMs for a number of regions (Figures 5–7). That being said, there are about 3 models in the CMIP5 ensemble examined here that perform rather well in regards to the Taylor diagram metrics (i.e., standard deviation ratio, pattern correlation) for CIWP; these include the CNRM-CM5, MRI and CanESM2

[35] There is another five that perform particularly poorly, GISS-E2-H, GISS-E2-R, IPSL, INMCM4, and INMCM4-ESM, with the former (latter) three (two) being biased very significantly high (low) in terms of overall CIWP magnitude. The remaining seven exhibit performance in between (BCC, BCC-CSM1, NorESM, CSIRO, MIROC5, MIROC4h and CCSM4). As expected the two reanalyses examined performed relatively well compared to the group as a whole due to their use of observed SSTs and the incorporation of a wide array of constraining observations; a result that is still notable though since they don't assimilate cloud ice observations and thus rely on (parameterized) model physics to represent this quantity. However, even with the assimilation of many other/related quantities and the benefit of observed SSTs, neither MERRA's nor ECMWF-Interim's performance was within the uncertainty of the observations for both the standard deviation ratio and pattern correlation. The GEOS5 AGCM also performed well but it also has the benefit from specified, observed SSTs. Remarkable is the performance of the UCLA CGCM that was one of the best performing CGCMs of all those examined, along with the three identified above. Considering even these results alone and the remaining disparities between the observations and modeled values of CIWP, it is evident that while the models may be providing roughly the correct radiative energy budget, many are accomplishing it by means of unrealistic cloud characteristics of cloud ice mass at a minimum, which in turn likely indicates unrealistic cloud particle sizes and cloud cover.

[36] Examination of the vertical structure of CIWC in terms of global, zonal and large-region averages (e.g., high, mid and tropical latitudes) indicates similar findings in terms of overall performance across the models and re-analyses examined here. Setting aside the differences in magnitude discussed above, most of the systematic error in the global-mean vertical profile of CIWC occurs below the mid-troposphere where the models tend to overestimate CIWC compared to the observed estimate. When considered in more

detail, it is evident that this bias arises mostly from the mid and high latitudes (Figures 8–10). Note that some caution is warranted here in regards to the observations as discriminating the boundary between ice and liquid and quantifying the contributions from each in mixed-phase clouds is very difficult using present-day observing capabilities. Apart from the differences in magnitude, the models also display considerable differences in the pressure level of maximum CIWC, ranging from about 250 to 550 hPa. Finally, using the observational uncertainty to discriminate good from poorly performing models in terms of CIWC and CIWP (Figure 13) also indicated that in terms of pattern correlations, the bulk of the CMIP5 models tend to perform best in the mid-to-lower troposphere (\sim 500–700 hPa) with a few also performing well at about 200 hPa (Figure 13a).

[37] A specific objective of this study was to compare the overall performance between CMIP3 and CMIP5. Based on a number of diagnostics (e.g., Figures 5a–5z versus Figure A1, Figure 6, Figure 7) there has been significant and quantitative improvement in the representation of CIWP between CMIP3 and CMIP5 possibly due to the use of two-moment cloud microphysics schemes in more GCMs. This is clearly demonstrated by the reduction (by about 50% or more) in the multimodel mean bias and RMSE of the annual mean maps of CIWP between CMIP3 and CMIP5 (Figures 6a–6d and Figures 6e–6h). Similarly, when viewed on a Taylor plot, the multimodel mean values for the annual mean maps of CIWP are considerably better in CMIP5 than CMIP3 with the distance between the multimodel ensemble value and the reference point reduced by about a factor of 3, although the CMIP5 ensemble value (even with outliers removed – see more below) did not yet quite lie in the range of the observational uncertainty discussed above.

[38] While there are a number of models that perform quite well (e.g., CNRM-CM5, Can-ESM, and MRI-CGCM3), there are still a number that exhibit very significant biases in their mean spatial structure of CIWP (and CIWC). This includes differences up to factors of \sim 5 or more in a number of regions (e.g., Figure 5; GISS-E2-H, GISS-E2-R, INMCM4 and INMCM4ESM). The impact of these outliers on the multimodel mean bias and RMSE is substantial. For example, when excluding these outliers, not only is the overall magnitude of the bias and RMSE significantly reduced, the pattern of the bias changes markedly. With all models, the bias is generally positive over most regions where high clouds occur, and there is a notable double ITCZ imprint (Figures 6a–6d). When these outliers are removed, the bias is positive (negative) in the extra-tropics (tropics) and the pattern echoes a more canonical pattern of the ITCZ and storm tracks (Figures 6e–6h). Note that the substantial positive bias in CIWP/CIWC in the tropics in the two GISS models arises due to a compensation for a relative negative bias in clouds and CIWP/CIWC in the extra-tropics (Figures 5a–5z and A. Del Genio, GISS, personal communication, 2012) in conjunction with the need to have a globally averaged radiation budget consistent with observations.

[39] Given that there have been viable observed estimates of CIWP/CIWC for about 4–5 years from CloudSat, about 2–3 years from combined CloudSat + CALIPSO retrievals, and even longer for CIWC values from MLS [e.g., Wu et al., 2008, 2009; Li et al., 2005], yet there are still GCMs

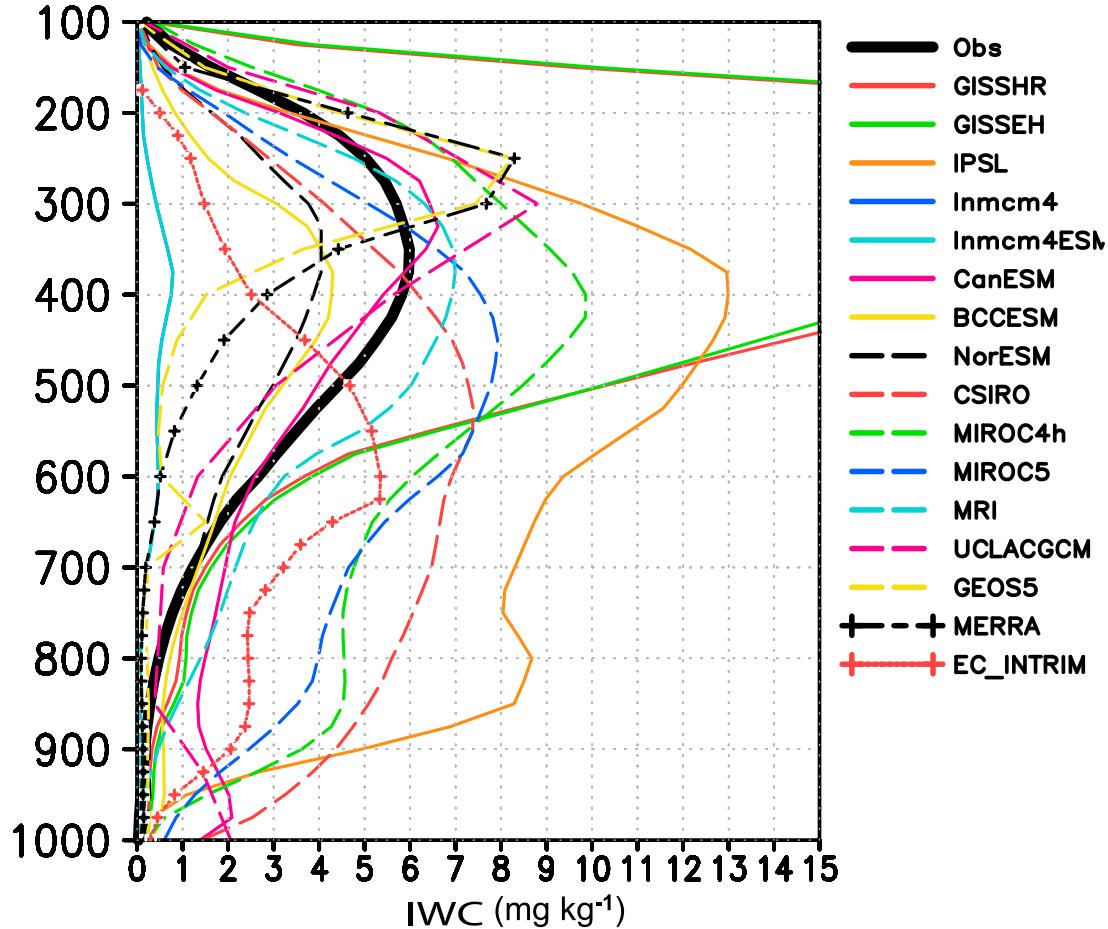


Figure 9. Globally averaged (80°N – 80°S), annual mean, vertical CIWC profiles (mg kg^{-1}) from the CMIP5 GCMs, UCLA CGCM, GEOS5 AGCM, EC Interim and MERRA reanalyses, as well as the observed estimate (thick black profile). Note that the CNRM-CM5 CGCM CIWC is not available from the CMIP5 data portal at the time.

exhibiting such large biases, indicates challenges to utilizing the observations by all the modeling groups. In addition, the deficiencies of model simulations on temperature, humidity, dynamics, radiation can also contribute to the IWC/TWP large bias to the reference values. These challenges likely include the painstaking work of model development and improvement, the relatively few people/resources available to perform such work within a modeling institute and group, the reduction in the above mentioned time periods of data availability due to the need to have “final” model versions ready for CMIP5 at least a year ago, and the non-trivial effort of obtaining and getting acquainted with the data [e.g., Waliser *et al.*, 2009; Gleckler *et al.*, 2011]. As cloud feedback will undoubtedly still represent a key uncertainty in the next IPCC assessment report, it is essential that these observations be utilized to their full extent to provide the key constraints they offer (see section 1). The instantiation of the Climate Metrics Panel sponsored by the Working Group on Coupled Modeling (WGCM) and Working Group on Numerical Experimentation, (WGNE) along with the development of the “obs4MIPs” activity [Gleckler *et al.*, 2011] are designed to facilitate the use of observations to provide multiple constraints and points of evaluation for climate

model performance. Thus, it can be expected that there will be fewer and fewer extreme outliers in model performance as these activities and resources become fully implemented.

[40] Our study also included evaluating the performance of two GFDL GCMs (AM3 and CM3) that stand apart from the others as they attempt to also model and include in their output atmospheric ice water associated with convective clouds and precipitation. Thus, the comparison in these cases is on TIWC/TIWP, and the unfiltered versions of the satellite retrievals can be utilized (e.g., Figure 5). For these two GCMs, the comparisons of TIWP show that these models perform very well in the tropics, but underestimate the TIWP by about a factor of $\sim 2\text{--}4$ in the extra-tropics. Examining the TIWC profiles shows modest pattern biases that come from the upper troposphere in the tropics (e.g., enhanced double ITCZ) with rather uniform underestimates of TIWC throughout the troposphere in the extra-tropics. The abilities of the GFDL models to represent and output both cloudy and precipitating ice profiles, combined with the observational capabilities to roughly distinguish these two types of ice mass, provides an additional means for constraints on the model physics.

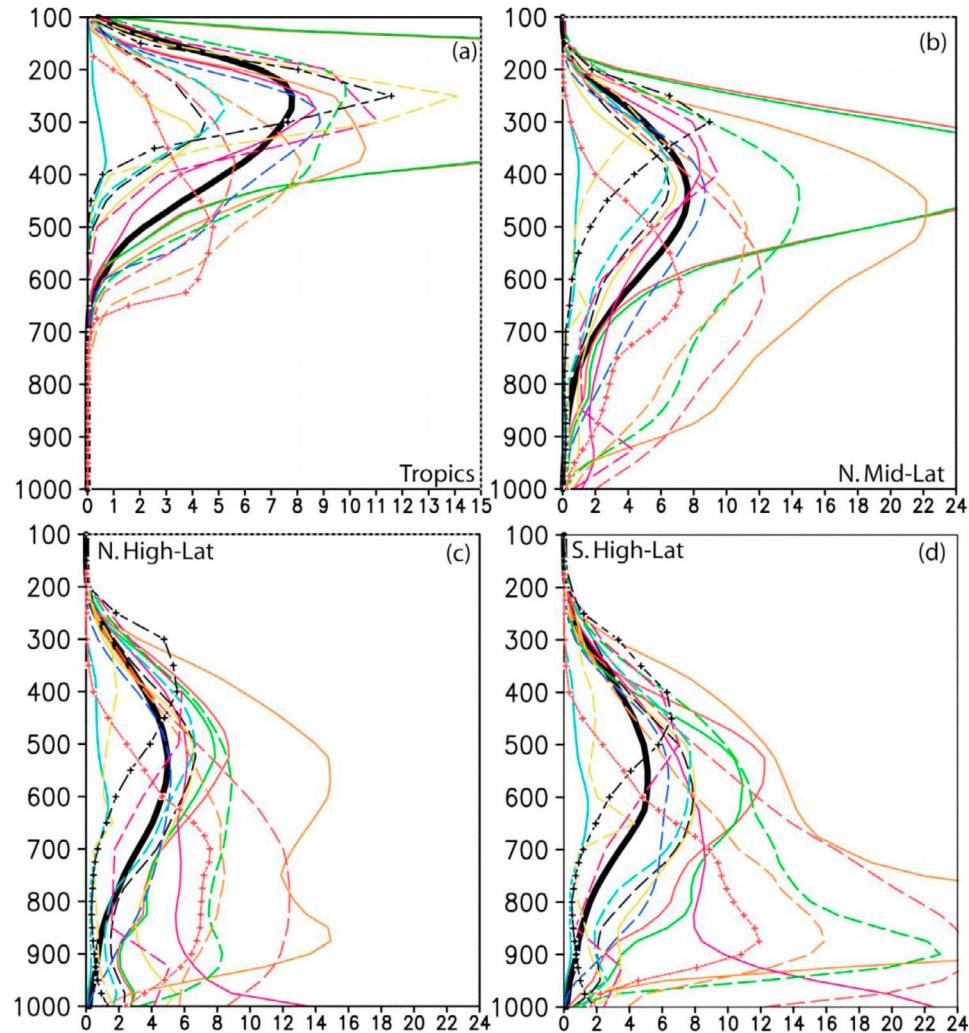


Figure 10. The same as Figure 9 but for (a) Tropics (30°N – 30°S), (b) N. Mid-Lat (30°N – 60°N), (c) N. High-Lat (60°N – 80°N) and (d) S. High-Lat (60°S – 80°S).

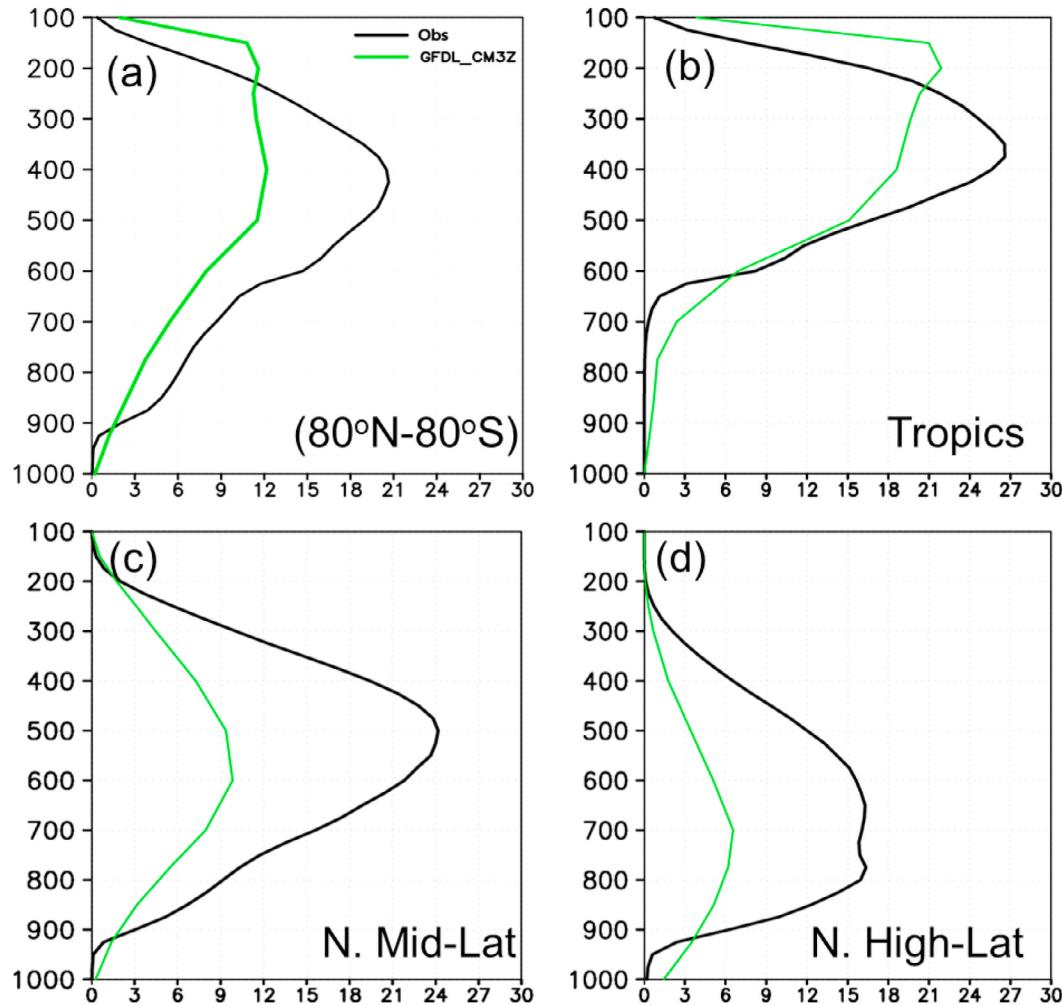


Figure 11. Same as Figure 10 but for TIWC (mg kg^{-1}) instead of CIWC, and only for GFDL CM3 (green). In this case, the latitudinal bands are: (a) 80°N – 80°S , (b) Tropics (30°N – 30°S), (c) N. Mid-Lat (30°N – 60°N), and (d) N. High-Lat (60°N – 80°N).

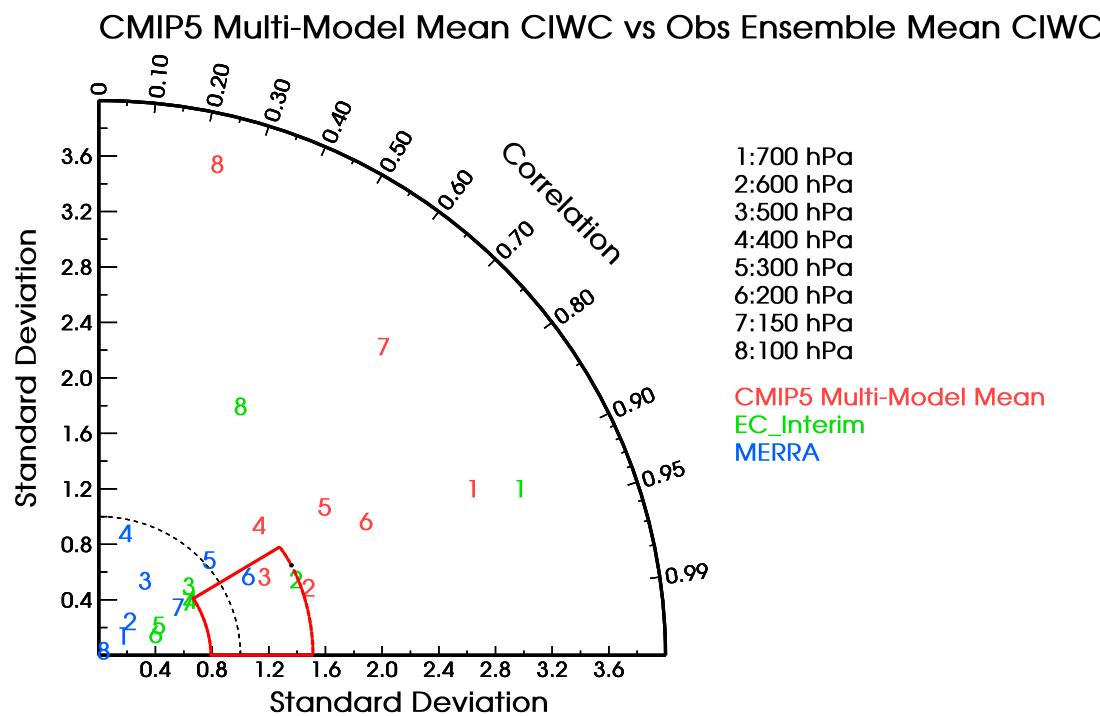


Figure 12. Taylor diagram for CMIP5 annual multimodel mean as well as GEOS5 GCM, UCLA CGCM, MERRA and EC Interim at each pressure level, area average (80°N – 80°S) cloud ice water content (CIWC; mg kg^{-1}) against CloudSat observed estimate ensemble mean of CIWC described in section 2.

a)

STD-Ratio	CMIP3 CIWP	CMIP5 CIWP	700 hPa	600 hPa	500 hPa	400 hPa	300 hPa	200 hPa	150 hPa
Multi-Model Mean	3.1	1.4(*1.23)(**1.09)	2.1	1.48	1.5	1.59	1.82	1.76	1.73
GISSEH	16.9	6.3	1.65	1.8	3.5	5	8.9	8	9.7
GISSER	16.8	6.2	1.66	1.8	3.4	5.4	8.3	8.6	9.3
IPSL	4	4	6.7	3.7	3.3	2.7	1.4	0.8	0.48
Inmcm4	N/A	0.36	0.2	0.21	0.112	0.1	0.06	0.02	0.03
Inmcm4ESM	N/A	0.36	0.2	0.21	0.12	0.1	0.06	0.02	0.03
Can_ESM	N/A	1.22	1.7	1.1	1.03	1.06	1.14	0.12	0.45
BCC-ESM	N/A	0.79	1.36	1.04	1.1	0.76	0.35	0.28	0.15
NorESM	N/A	0.74	1.32	1.02	1.08	0.8	0.47	0.363	0.27
CSIRO	2.6	2.64	4.9	2.6	1.66	1.2	0.7	1.2	0.24
MIROC4h	N/A	2.32	3.7	2.3	1.9	1.8	1.2	1.27	1.3
MIROC5	N/A	1.69	3.1	1.74	1.45	1.4	0.86	0.45	0.23
MRI-CGCM3	N/A	1.15	1.65	1.18	1.1	1.3	1.23	0.66	0.42
CNRM-CM5	1.1	0.76	N/A						
CCSM4	0.9	0.49	N/A						
GEOSS		0.96	0.3	0.42	0.2	0.28	1.47	1.17	0.843
EC-Intrrim		1.02	2.3	1.1	1	0.73	0.43	0.17	1.19
UCLA CGCM		1.01	0.81	0.73	0.94	1.1	1.2	1.33	1.51
MERRA		0.78	0.22	0.33	0.63	0.9	1.04	1.2	0.66

b)

STD-Ratio	CMIP3 CIWP	CMIP5 CIWP	700 hPa	600 hPa	500 hPa	400 hPa	300 hPa	200 hPa	150 hPa
Multi-Model Mean	3.1	1.4(*1.23)(**1.09)	2	1.48	1.45	1.51	1.82	1.76	1.61
GISSEH	16.9	6.3	1.65	1.8	3.5	5	8.9	8	9.7
GISSER	16.8	6.2	1.66	1.8	3.4	5.4	8.3	8.6	9.3
IPSL	4	4	6.7	3.7	3.3	2.7	1.4	0.8	0.48
Inmcm4	N/A	0.36	0.2	0.21	0.112	0.1	0.06	0.02	0.03
Inmcm4ESM	N/A	0.36	0.2	0.21	0.12	0.1	0.06	0.02	0.03
Can_ESM	N/A	1.22	1.7	1.1	1.03	1.06	1.14	0.12	0.45
BCC-ESM	N/A	0.79	1.36	1.04	1.1	0.76	0.35	0.28	0.15
NorESM	N/A	0.74	1.32	1.02	1.08	0.8	0.47	0.363	0.27
CSIRO	2.6	2.64	4.9	2.6	1.66	1.2	0.7	1.2	0.24
MIROC4h	N/A	2.32	3.7	2.3	1.9	1.8	1.2	1.27	1.3
MIROC5	N/A	1.69	3.1	1.74	1.45	1.4	0.86	0.45	0.23
MRI-CGCM3	N/A	1.15	1.65	1.18	1.1	1.3	1.23	0.66	0.42
CNRM-CM5	1.1	0.76	N/A						
CCSM4	0.9	0.49	1.2	0.9	1.02	0.83	0.47	0.37	0.29
GEOSS		0.96	0.3	0.42	0.2	0.28	1.47	1.17	0.843
EC-Intrrim		1.02	2.3	1.1	1	0.73	0.43	0.17	1.19
UCLA CGCM		1.01	0.81	0.73	0.94	1.1	1.2	1.33	1.51
MERRA		0.78	0.22	0.33	0.63	0.9	1.04	1.2	0.66

Figure 13. (a) Pattern correlation of long-term annual mean CIWP and CIWC values between the models and the observed values. Shading indicates values exceeding 0.85, which is the lower bound of the observational uncertainty used in this study (see Figure 4 and associated discussion in section 2). Single asterisks denote: excluding the two GISS GCMs. Double asterisks denote: excluding the two GISS and two Inmcm4 GCMs. (b) Same as Figure 13a except for standard deviation ratio. Shading indicates values between 0.8 and 1.5, which is the lower bound of the observational uncertainty used in this study (see Figure 4 and associated discussion in section 2). Single asterisks denote excluding the two GISS GCMs. Double asterisks denote excluding the two GISS and two Inmcm4 GCMs.

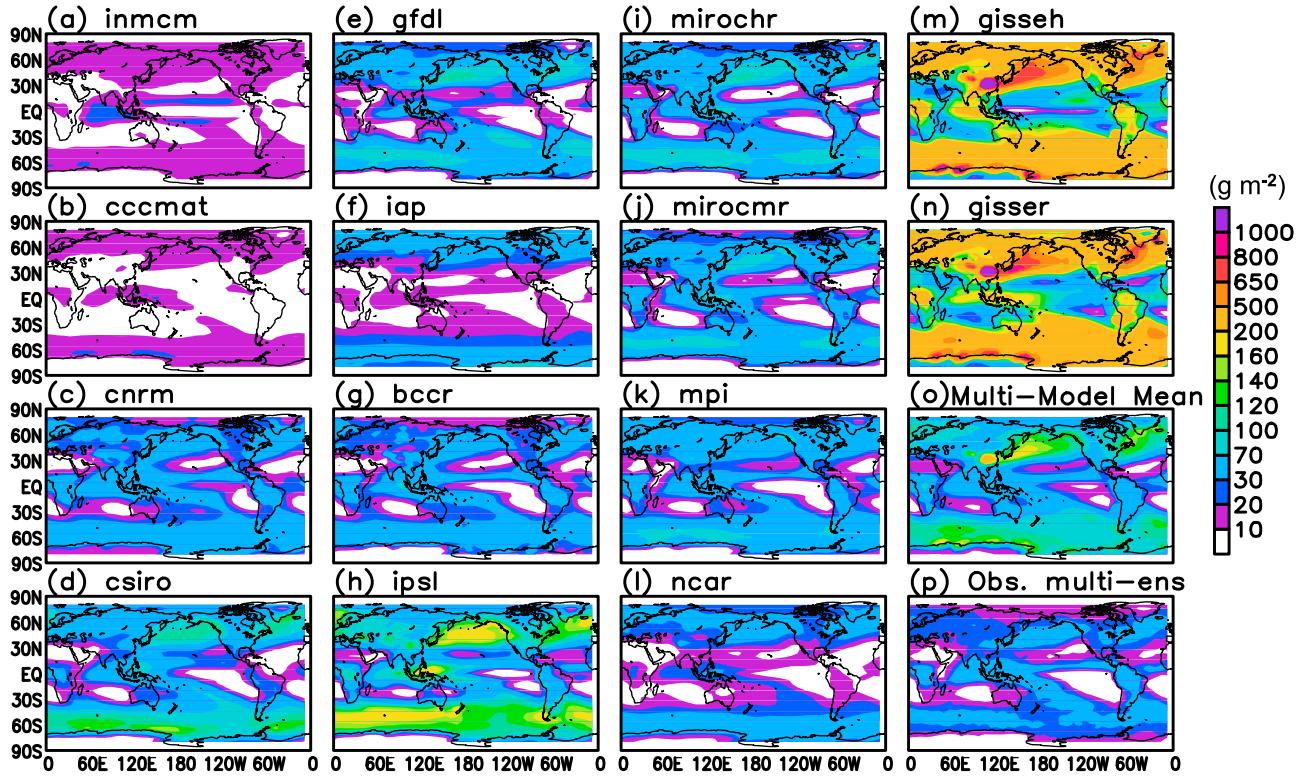


Figure A1. (a–n) The same as Figure 5, except for a number of CMIP3 GCMs included in Waliser et al. [2009], (o) their multimodel mean, and (p) the observed estimate of CIWP. Note the change in color bar between Figures 5 and A1.

[41] While it was beyond the scope of this study to probe the causes of the model-to-model differences and model-to-observation biases, highlighting a few recent developments is instructive to help keep in mind the complexities associated with modeling atmospheric ice. An issue is that, the inconsistency, sensitivity and uncertainties of temperature thresholds used for the temperature-dependent mixed-phase assumptions for cloud ice and cloud liquid water in observations and models can also contribute to the biases [Cheng et al., 2012; Ma et al., 2012]. For example, Note that, MRI is one of the CMIP5 models that perform quite well, is the only model using a double moment cloud scheme with two separate prognostic equations for cloud liquid and ice. The model treats atmospheric transport (including cumulus convection) of the aerosols and detailed cloud-aerosol effects [Yukimoto et al., 2012].

[42] Recent progress in this regard includes a study by Gettelman et al. [2010] who reported improvements from the incorporation of a process-based ice cloud representation that allowed supersaturation with respect to ice and an ice cloud cover consistent with the treatment of ice microphysics. The sensitivities and improvements found in association with ice supersaturation were also reported by Tompkins et al. [2007] for a recent version/update of the ECMWF Integrated Forecast System. In another development, Ma et al. [2012a] indicates that the notable fidelity of simulated CIWC/CIWP in the UCLA GCM stems from better representation of deep convection and the associated detrainment, as well as cloud macro- and microphysics schemes. Song et al. [2012] incorporated two-moment microphysical

scheme to convective areas of grid cells, helping to rectify the negative CIWP bias evident in CCSM4.

Appendix A

[43] Figures A1a–A1n are the same as Figure 5, except for a number of CMIP3 GCMs included in Waliser et al. [2009], their multimodel mean (Figure A1o), and the observed estimate of CIWP (Figure A1p). Note the change in color bar between Figures 5 and A1.

[44] **Acknowledgments.** We would like to thank editor and three reviewers for their very comprehensive review and constructive comments/suggestion. We thank Anthony Del Genio/NASA GISS, Ken Lo /NASA GISS, Voldoir/Aurore/CNRM, Masahiro Watanabe and Shingo Watanabe /MIROC, Leon Rotstayn/CSIRO, Knut von Salzen/CCCma, Gary Strand/NCAR, Alf Kirkevag/NCC, Seiji Yukimoto/MRI, Jean-Louis Dufresne/IPSL, Tongwen Wu/BCC-CMA and many other colleagues from climate modeling centers for providing model information. Thanks also go to Gregory Huey with data and Matthew Lebsack for reading and providing comments on the manuscript. The contributions by D.E.W. and J.L.L. to this study were carried out on behalf of the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The contribution of Hsi-Yen Ma to this work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344.

References

- Adler, R. F., et al. (2003), The version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present), *J. Hydrometeorol.*, 4, 1147–1167, doi:10.1175/1525-7541(2003)004<1147:TVGCP>2.0.CO;2.

- Austin, R., and G. L. Stephens (2001), Retrieval of stratus cloud microphysical parameters using millimeter-wave radar and visible optical depth in preparation for CloudSat: 1. Algorithm formulation, *J. Geophys. Res.*, **106**, 28,233–28,242, doi:10.1029/2000JD000293.
- Austin, R. T., A. J. Heymsfield, and G. L. Stephens (2009), Retrieval of ice cloud microphysical parameters using the CloudSat millimeter-wave radar and temperature, *J. Geophys. Res.*, **114**, D00A23, doi:10.1029/2008JD010049.
- Bodas-Salcedo, A., et al. (2011), COSP: Satellite simulation software for model assessment, *Bull. Am. Meteorol. Soc.*, **92**, 1023–1043, doi:10.1175/2011BAMS2856.1.
- Bony, S., and K. A. Emmanuel (2001), A parameterization of the cloudiness associated with cumulus convection: Evaluation using TOGA COARE data, *J. Atmos. Sci.*, **58**, 3158–3183.
- Bougeault, P. (1981), Modeling the trade-wind cumulus boundary layer. Part I: Testing the ensemble cloud relations against numerical data, *J. Atmos. Sci.*, **38**, 2414–2428, doi:10.1175/1520-0469(1981)038<2414:MTTWC>2.0.CO;2.
- Bougeault, P. (1982), Cloud ensemble relations based on the gamma probability distribution for the higher-order models of the planetary boundary layer, *J. Atmos. Sci.*, **39**, 2691–2700, doi:10.1175/1520-0469(1982)039<2691:CERBOT>2.0.CO;2.
- Boville, B. A., P. J. Rasch, J. J. Hack, and J. R. McCaa (2006), Representation of clouds and precipitation processes in the Community Atmosphere Model Version 3(CAM3), *J. Clim.*, **19**, 2184–2198, doi:10.1175/JCLI3749.1.
- Brown, P. R. A., and P. N. Francis (1995), Improved measurements of the ice water content in cirrus using a total-water probe, *J. Atmos. Oceanic Technol.*, **12**, 410–414, doi:10.1175/1520-0426(1995)012<0410:IMO-TIW>2.0.CO;2.
- Chaboureau, J.-P., and P. Bechtold (2005), Statistical representation of clouds in a regional model and the impact on the diurnal cycle of convection during tropical convection, cirrus and nitrogen oxides (TROCCINOX), *J. Geophys. Res.*, **110**, D17103, doi:10.1029/2004JD005645.
- Chen, W.-T., C. P. Woods, J.-L. F. Li, D. E. Waliser, J.-D. Chern, W.-K. Tao, J. H. Jiang, and A. M. Tompkins (2011), Partitioning CloudSat ice water content for comparison with upper-tropospheric ice in global atmospheric models, *J. Geophys. Res.*, **116**, D19206, doi:10.1029/2010JD015179.
- Cheng, A., K.-M. Xu, Y. X. Hu, and S. Kato (2012), Impact of a cloud thermodynamic phase parameterization based on CALIPSO observations on climate simulation, *J. Geophys. Res.*, **117**, D09103, doi:10.1029/2011JD017263.
- Collins, W. D., et al. (2004), Description of the NCAR Community Atmosphere Model (CAM 3.0), *Tech Rep. NCAR/TN-464+STR*, 210 pp., Natl. Cent. for Atmos. Res., Boulder, Colo.
- Dee, D. P., et al. (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, **137**, 553–597, doi:10.1002/qj.828.
- Delanoë, J., and R. J. Hogan (2008), A variational scheme for retrieving ice cloud properties from combined radar, lidar, and infrared radiometer, *J. Geophys. Res.*, **113**, D07204, doi:10.1029/2007JD009000.
- Delanoë, J., and R. J. Hogan (2010), Combined CloudSat-CALIPSO-MODIS retrievals of the properties of ice clouds, *J. Geophys. Res.*, **115**, D00H29, doi:10.1029/2009JD012346.
- Delanoë, J., R. J. Hogan, R. M. Forbes, A. Bodas-Salcedo, and T. H. M. Stein (2011), Evaluation of ice cloud representation in the ECMWF and UK Met Office models using CloudSat and CALIPSO data, *Q. J. R. Meteorol. Soc., Part B*, **137**, 2064–2078, doi:10.1002/qj.882.
- Del Genio, A. D., M.-S. Yao, W. Kovari, and K. K.-W. Lo (1996), A prognostic cloud water parameterization for global climate models, *J. Clim.*, **9**(2), 270–304, doi:10.1175/1520-0442(1996)009<0270:APCWP>2.0.CO;2.
- Deng, M., G. G. Mace, Z. Wang, and H. Okamoto (2010), Tropical Composition, Cloud and Climate Coupling Experiment validation for cirrus cloud profiling retrieval using CloudSat radar and CALIPSO lidar, *J. Geophys. Res.*, **115**, D00J15, doi:10.1029/2009JD013104.
- Donner, L. J. (1993), A cumulus parameterization including mass fluxes, vertical momentum dynamics, and mesoscale effects, *J. Atmos. Sci.*, **50**, 889–906.
- Donner, L. J., et al. (2011), The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component of the GFDL coupled model CM3, *J. Clim.*, **24**, 3484–3519, doi:10.1175/2011JCLI3955.1.
- Field, P. R., R. J. Hogan, P. R. A. Brown, A. J. Illingworth, T. W. Choularton, and R. J. Cotton (2005), Parameterization of ice particle size distributions for mid-latitude stratiform cloud, *Q. J. R. Meteorol. Soc.*, **131**, 1997–2017, doi:10.1256/qj.04.134.
- Francis, P. N., P. Hignett, and A. Macke (1998), The retrieval of cirrus cloud properties from aircraft multi-spectral reflectance measurements during EUCREX'93, *Q. J. R. Meteorol. Soc.*, **124**, 1273–1291, doi:10.1002/qj.49712454812.
- Gettelman, A., X. Liu, S. J. Ghan, H. Morrison, S. Park, A. J. Conley, S. A. Klein, J. Boyle, D. L. Mitchell, and J.-L. F. Li (2010), Global simulations of ice nucleation and ice supersaturation with an improved cloud scheme in the Community Atmosphere Model, *J. Geophys. Res.*, **115**, D18216, doi:10.1029/2009JD013797.
- Ghan, S., et al. (2000), A comparison of single column model simulations of summertime midlatitude continental convection, *J. Geophys. Res.*, **105**, 2091–2124, doi:10.1029/1999JD900971.
- Gleckler, P., R. Ferraro, and D. E. Waliser (2011), Meeting summary: Improving use of satellite data in evaluating climate models, *Eos Trans. AGU*, **92**(20), 172, doi:10.1029/2011EO200005.
- Han, Q. Y., W. B. Rossow, J. Chou, K. S. Kuo, and R. M. Welch (1999), The effects of aspect ratio and surface roughness on satellite retrievals of ice-cloud properties, *J. Quant. Spectrosc. Radiat. Transfer*, **63**(2–6), 559–583.
- Heymsfield, A. J., et al. (2008), Testing IWC retrieval methods using radar and ancillary measurements with in situ data, *J. Appl. Meteorol. Climatol.*, **47**(1), 135–163, doi:10.1175/2007JAMC1606.1.
- Hogan, R. J. (2006), Fast approximate calculation of multiply scattered lidar returns, *Appl. Opt.*, **45**, 5984–5992, doi:10.1364/AO.45.005984.
- Hong, G. (2007), Radar backscattering properties of nonspherical ice crystals at 94 GHz, *J. Geophys. Res.*, **112**, D22203, doi:10.1029/2007JD008839.
- Hourdin, F., et al. (2006), The LMDZ4 general circulation model: Climate performance and sensitivity to parametrized physics with emphasis on tropical convection, *Clim. Dyn.*, **19**(15), 3445–3482, doi:10.1007/s00382-006-0158-0.
- Klein, S. A., and C. Jakob (1999), Validation and sensitivities of frontal clouds simulated by the ECMWF model, *Mon. Weather Rev.*, **127**(10), 2514–2531, doi:10.1175/1520-0493(1999)127<2514:VASOFC>2.0.CO;2.
- Klein, S. A., et al. (2009), Intercomparison of model simulations of mixed phase clouds observed during the ARM Mixed-Phase Arctic Cloud Experiment, Part I: Single-layered cloud, *Q. J. R. Meteorol. Soc.*, **135**, 979–1002, doi:10.1002/qj.416.
- Köhler, M. (1999), Explicit prediction of ice clouds in general circulation models, PhD dissertation, Dep. of Atmos. Sci., Univ. of Calif., Los Angeles.
- 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, J.-L. F., et al. (2005), Comparisons of EOS MLS cloud ice measurements with ECMWF analyses and GCM simulations: Initial results, *Geophys. Res. Lett.*, **32**, L18710, doi:10.1029/2005GL023788.
- Li, J.-L. F., J. H. Jiang, D. E. Waliser, and A. M. Tompkins (2007), Assessing consistency between EOS MLS and ECMWF analyzed and forecast estimates of cloud ice, *Geophys. Res. Lett.*, **34**, L08701, doi:10.1029/2006GL029022.
- Li, J.-L. F., D. Waliser, C. Woods, J. Teixeira, J. Bacmeister, J. Chern, B.-W. Shen, A. Tompkins, W.-K. Tao, and M. Köhler (2008), Comparisons of satellites liquid water estimates to ECMWF and GMAO analyses, 20th century IPCC AR4 climate simulations, and GCM simulations, *Geophys. Res. Lett.*, **35**, L19710, doi:10.1029/2008GL035427.
- Li, J.-L. F., D. E. Waliser, and J. H. Jiang (2011), Correction to “Comparisons of satellites liquid water estimates to ECMWF and GMAO analyses, 20th century IPCC AR4 climate simulations, and GCM simulations,” *Geophys. Res. Lett.*, **38**, L24807, doi:10.1029/2011GL049956.
- Lohmann, U., and E. Roeckner (1996), Design and performance of a new cloud microphysics scheme developed for the ECHAM general circulation model, *Clim. Dyn.*, **12**, 557–572, doi:10.1007/BF00207939.
- Ma, H.-Y., M. Köhler, J.-L. F. Li, J. D. Farrara, C. R. Mechoso, R. M. Forbes, and D. E. Waliser (2012), Evaluation of an ice cloud parameterization based on a dynamical-microphysical lifetime concept using CloudSat observations and the ERA-Interim reanalysis, *J. Geophys. Res.*, **117**, D05210, doi:10.1029/2011JD016275.
- Mace, G. G., Q. Zhang, M. Vaughan, R. Marchand, G. Stephens, C. Trepte, and D. Winker (2009), A description of hydrometeor layer occurrence statistics derived from the first year of merged CloudSat and CALIPSO data, *J. Geophys. Res.*, **114**, D00A26, doi:10.1029/2007JD009755.
- Mellor, G. L. (1977), The Gaussian cloud model relations, *J. Atmos. Sci.*, **34**, 356–358.
- Morrison, H., et al. (2009), Intercomparison of model simulations of mixed phase clouds observed during the ARM Mixed-Phase Arctic Cloud Experiment, part II: Multi-layered cloud, *Q. J. R. Meteorol. Soc.*, **135**, 1003–1019, doi:10.1002/qj.415.
- Ogura, T., S. Emori, M. J. Webb, Y. Tushima, T. Yokohata, A. Abe-Ouchi, and M. Kimoto (2008), Towards understanding cloud response in

- atmospheric GCMs: The use of tendency diagnostics, *J. Meteorol. Soc. Jpn.*, 86, 69–79, doi:10.2151/jmsj.86.69.
- 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, doi:10.1175/1520-0442(1998)011<1587:ACOTCM>2.0.CO;2.
- Ricard, J. L., and J. F. Royer (1993), A statistical cloud scheme for use in an AGCM, *Ann. Geophys.*, 11, 1095–1115.
- Rienecker, M. M., et al. (2008), The GEOS-5 Data Assimilation System - Documentation of versions 5.0.1, 5.1.0, and 5.2.0, *NASA Tech. Memo.*, NASA/TM-2008-104606, vol. 27, 118 pp.
- Rossow, W. B., and R. A. Schiffer (1999), Advances in understanding clouds from ISCCP, *Bull. Am. Meteorol. Soc.*, 80(11), 2261–2287, doi:10.1175/1520-0477(1999)080<2261:AIUCFI>2.0.CO;2.
- 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, *Q. J. R. Meteorol. Soc.*, 123, 1227–1282.
- Rotstayn, L. D., B. F. Ryan, and J. Katzfey (2000), A scheme for calculation of the liquid fraction in mixed-phase clouds in large-scale models, *Mon. Weather Rev.*, 128, 1070–1088, doi:10.1175/1520-0493(2000)128<1070:ASFCOT>2.0.CO;2.
- Sassen, K., Z. Wang, and D. Liu (2009), Cirrus clouds and deep convection in the tropics: Insights from CALIPSO and CloudSat, *J. Geophys. Res.*, 114, D00H06, doi:10.1029/2009JD011916.
- Song, X., G. J. Zhang, and J.-L. F. Li (2012), Evaluation of microphysics parameterization for convective clouds in the NCAR Community Atmosphere Model CAM5, *J. Clim.*, in press.
- Stein, T. H. M., J. Delanoë, and R. J. Hogan (2011), A comparison among four different retrieval methods for ice-cloud properties using data from CloudSat, CALIPSO, and MODIS, *J. Appl. Meteorol. Climatol.*, 50, 1952–1969, doi:10.1175/2011JAMC2646.1.
- Stephens, G. L., et al. (2008), CloudSat mission: Performance and early science after the first year of operation, *J. Geophys. Res.*, 113, D00A18, doi:10.1029/2008JD009982. [printed 114(D8), 2009]
- Sundqvist, H. (1988), Parameterization of condensation and associated clouds in models for weather prediction and general circulation simulation, in *Physically Based Modeling and Simulation of Climate and Climate Change*, vol. 1, edited by M. E. Schlesinger, pp. 433–461, Kluwer Acad., Norwell, Mass.
- Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, 106, 7183–7192, doi:10.1029/2000JD900719.
- Tiedtke, M. (1993), Representation of clouds in large-scale models, *Mon. Weather Rev.*, 121, 3040–3061, doi:10.1175/1520-0493(1993)121<3040:ROCLIS>2.0.CO;2.
- Tompkins, A. M., K. Gierens, and G. Radel (2007), Ice supersaturation in the ECMWF integrated forecast system, *Q. J. R. Meteorol. Soc.*, 133, 53–63, doi:10.1002/qj.14.
- Volodin, E. M., N. A. Dianskii, and A. V. Gusev (2010), Simulating present day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations, *Izv. Russ. Acad. Sci. Atmos. Oceanic Phys., Engl. Transl.*, 46(4), 414–431, doi:10.1134/S000143381004002X.
- Waliser, D. E., et al. (2009), Cloud ice: A climate model challenge with signs and expectations of progress, *J. Geophys. Res.*, 114, D00A21, doi:10.1029/2008JD010015.
- Watanabe, M., S. Emori, M. Satoh, and H. Miura (2009), A PDF based hybrid prognostic cloud scheme for general circulation models, *Clim. Dyn.*, 33, 795–816, doi:10.1007/s00382-008-0489-0.
- 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(12), 905–922, doi:10.1007/s003820100157.
- Wielicki, B. A., B. R. Barkstrom, E. F. Harrison, R. B. Lee, 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, doi:10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2.
- Wilson, D. R., and S. P. Ballard (1999), A microphysically based precipitation scheme for the UK Meteorological Office unified model, *Q. J. R. Meteorol. Soc.*, 125, 1607–1636, doi:10.1002/qj.49712555707.
- Wu, D. L., J. H. Jiang, W. G. Read, R. T. Austin, C. P. Davis, A. Lambert, G. L. Stephens, D. G. Vane, and J. W. Waters (2008), Validation of the Aura MLS cloud ice water content measurements, *J. Geophys. Res.*, 113, D15S10, doi:10.1029/2007JD008931.
- Wu, D. L., et al. (2009), Comparisons of global cloud ice from MLS, CloudSat, and correlative data sets, *J. Geophys. Res.*, 114, D00A24, doi:10.1029/2008JD009946.
- Wu, T., R. Yu, F. Zhang, Z. Wang, M. Dong, L. Wang, X. Jin, D. Chen, and L. Li (2010), The Beijing Climate Center Atmospheric General Circulation Model: Description and its performance for the present-day climate, *Clim. Dyn.*, 34, 123–147, doi:10.1007/s00382-008-0487-2.
- Xie, P., and P. A. Arkin (1997), Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates and numerical model outputs, *Bull. Am. Meteorol. Soc.*, 78, 2539–2558, doi:10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2.
- 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, doi:10.1256/003590002320373229.
- Xie, S., et al. (2005), Simulations of midlatitude frontal clouds by SCMs and CRMs during the ARM March 2000 Cloud IOP, *J. Geophys. Res.*, 110, D15S03, doi:10.1029/2004JD005119.
- Xu, K.-M., et al. (2002), An intercomparison of cloud-resolving models with the ARM summer 1997 IOP data, *Q. J. R. Meteorol. Soc.*, 128, 593–624, doi:10.1256/003590002321042117.
- Xu, K.-M., et al. (2005), Modeling springtime shallow frontal clouds with cloud-resolving and single-column models, *J. Geophys. Res.*, 110, D15S04, doi:10.1029/2004JD005153.
- Yang, P., K. Liou, K. Wyser, and D. Mitchell (2000), Parameterization of the scattering and absorption 25 properties of individual ice crystals, *J. Geophys. Res.*, 105, 4699–4718, doi:10.1029/1999JD900755.
- Young, S. A., and M. A. Vaughan (2009), The retrieval of profiles of particulate extinction from Cloud-Aerosol Lidar Infrared Pathfinder Satellite Observations (CALIPSO) data: Algorithm description, *J. Atmos. Oceanic Technol.*, 26, 1105–1119, doi:10.1175/2008JTECHA1221.1.
- Yukimoto, S., et al. (2012), Meteorological Research Institute-Earth System Model Version 1(MRI-ESM1)—Model description, *Tech. Rep. Meteorol. Res. Inst.* 64, 96 pp., Meteorol. Res. Inst., Tsukuba, Japan. [Available at http://www.mri-jma.go.jp/Publish/Technical/DATA/VOL_64/tec_rep_mri_64.pdf.]
- Zhang, M., W. Lin, C. S. Bretherton, J. J. Hack, and P. J. Rasch (2003), A modified formulation of fractional stratiform condensation rate in the NCAR Community Atmospheric Model (CAM2), *J. Geophys. Res.*, 108(D1), 4035, doi:10.1029/2002JD002523.