Critical Evaluation of the ISCCP Simulator Using Ground-Based Remote Sensing Data

GERALD G. MACE, STEPHANIE HOUSER, AND SALLY BENSON

Department of Atmospheric Sciences, University of Utah, Salt Lake City, Utah

STEPHEN A. KLEIN

Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California

QILONG MIN

Atmospheric Sciences Research Center, University at Albany, State University of New York, Albany, New York

(Manuscript received 12 November 2009, in final form 3 September 2010)

ABSTRACT

Given the known shortcomings in representing clouds in global climate models (GCMs), comparisons with observations are critical. The International Satellite Cloud Climatology Project (ISCCP) diagnostic products provide global descriptions of cloud-top pressure and column optical depth that extend over multiple decades. Given the characteristics of the ISCCP product, the model output must be converted into what the ISCCP algorithm would diagnose from an atmospheric column with similar physical characteristics. This study evaluates one component of this so-called ISCCP simulator by comparing ISCCP results with simulated ISCCP diagnostics that are derived from data collected at the Atmospheric Radiation Measurement Program (ARM) Southern Great Plains (SGP) Climate Research Facility. It is shown that if a model were to simulate the cloud radiative profile with the same accuracy as can be derived from the ARM data, the likelihood of that occurrence being classified with similar cloud-top pressure and optical depth as ISCCP would range from 30% to 70% depending on optical depth. The ISCCP simulator improved the agreement of cloud-top pressure between ground-based remote sensors and satellite observations, and we find only minor discrepancies because of the parameterization of cloud-top pressure in the ISCCP simulator. The differences seem to be primarily due to discrepancies between satellite and ground-based sensors in the visible optical depth. The source of the optical depth bias appears to be due to subpixel cloud field variability in the retrieval of optical depths from satellite sensors. These comparisons suggest that caution should be applied to comparisons between models and ISCCP observations until the differences in visible optical depths are fully understood. The simultaneous use of ground-based and satellite retrievals in the evaluation of model clouds is encouraged.

1. Introduction

Clouds play an important role in the earth's climate system through their modification of the earth's radiative energy and hydrologic cycles. Not only do clouds act to modify the energy and water cycles, but they are themselves sensitive to changes in the climate state. Among the primary feedback processes in the earth's climate system [water vapor, surface albedo, and lapse rate feedbacks (Soden and Held 2006)], uncertainties in the representation of cloud feedbacks in global climate models (GCM) have been consistently identified as the primary source of uncertainty in prediction of anthropogenic climate change (Dufresne and Bony 2008).

GCMs in the recent Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (Solomon et al. 2007) have resolutions that are spatially and temporally much coarser than the spatial and temporal scales important to the evolution of cloud systems. Therefore, the impact of clouds systems (i.e., the radiative and hydrologic forcing) must be represented statistically through parameterizations of the dominant physical processes that result in the forcing (Randall et al. 2003). This task is difficult given the large variety of clouds ranging from deep convection to thin cirrus and the different processes involved. Many

Corresponding author address: Gerald G. Mace, Department of Meteorology, University of Utah, 135 South 1460 East, Room 819 (WBB), Salt Lake City, UT 84112-0110.
E-mail: jay.mace@utah.edu

DOI: 10.1175/2010JCLI3517.1

studies have shown that shortcomings in the prediction of present-day cloud forcing and cloud occurrence represent a major component of the cloud uncertainty associated with cloud feedbacks in future climates (e.g., Dufresne and Bony 2008; Williams and Tselioudis 2007; Williams and Webb 2009). A path forward to improved prediction of cloud feedbacks lies in the improved representation of clouds in the present climate state. Comparisons between model output and observations is, therefore, quite important.

The International Satellite Cloud Climatology Project (ISCCP) was initiated in the early 1980s with a goal of addressing the cloud feedback problem (Schiffer and Rossow 1983). This level of foresight is clearly a credit to the developers of ISCCP because, more than a quarter-century later, ISCCP remains a flagship description of the cloudy atmosphere. By analyzing visible and infrared radiances produced by geostationary and polar-orbiting meteorological satellites and applying assumptions regarding the layering of clouds in the atmosphere, their thermodynamic phases, and their properties, ISCCP describes a cloudy satellite pixel with the column visible optical depth (τ) and cloud-top pressure (P) of the highest cloud layer in the column. Hereafter, we refer to the ISCCP cloud-top pressure as P_{ISCCP} and the ISCCP visible optical depth as τ_{ISCCP}.

It would seem that the long-term global climatology of ISCCP addresses the needs of the GCM community. However, before comparing statistics derived from ISCCP with statistics derived from GCM output, the GCM-simulated atmospheric state must be interpreted with a set of equivalent assumptions as are used in calculating P_{ISCCP} and τ_{ISCCP} from the observed satellite data. This bridge between models and observations, known as the ISCCP simulator (Klein and Jakob 1999; Webb et al. 2001), has been and continues to be an important tool in model development, intercomparison (e.g., Zhang et al. 2005, hereafter Z05), and validation (e.g., Williams and Webb 2009).

There are two components to the ISCCP simulator. Since a GCM represents clouds within a finite spatial grid that is often much coarser than the satellite measurements, it is necessary to downscale the model output to a spatial scale that is more similar to that of the satellite measurement. This statistical downscaling technique, known as the Subgrid Cloud Overlap Profile Sampler (SCOPS), is based upon that reported in Klein and Jakob (1999). The other component, and the one we address here, is the representation of cloud-top pressure and visible optical depth from the model in a manner that is similar to what ISCCP would produce from satellite measurements. This component of the ISCCP simulator is known as the ISCCP Clouds and Radiances Using SCOPS (ICARUS).

The ISCCP simulator has become an important tool for evaluating the skill of GCMs to simulate the cloudy atmosphere. The Z05 study used output from 10 atmospheric general circulation models. The authors categorized the simulated clouds using what have become the standard nine ISCCP cloud types and compared them to the ISCCP climatology and to results from a similar algorithm known as the layer bispectral threshold method [LBTM; Minnis et al. 1995, referred to in Z05 as the Clouds and the Earth's Radiant Energy System (CERES) results]. Z05 shows that ISCCP and LBTM diagnose 30%–40% more midlevel clouds than produced by the GCMs, and that about half of the models underestimated the occurrence of low-topped clouds. Z05 also grouped the nine types into subgroups to better describe systematic model biases. The first subgroup consisted of the mid- and low-level clouds with optically thin ($\tau < 3.6$) and optically intermediate ($3.6 < \tau < 23$) thicknesses. Z05 found that the models simulated only about half of the clouds in this subgroup compared to ISCCP and LBTM. Another grouping of the model results combined all the optically thick ($\tau > 23$) clouds at all three cloud-top pressure intervals. The majority of the models significantly overestimated the occurrence frequency of this subgroup by more than a factor of 2 when compared to ISCCP and LBTM diagnostics.

While the ISCCP simulator has proven to be an important tool, the ISCCP simulator has not undergone a thorough validation with measurements. In an initial examination of the ISCCP simulator, Mace et al. (2006, hereafter M06) used cloud properties derived from ground-based remote sensors at the Atmospheric Radiation Measurement Program (ARM) Southern Great Plains site as input to the ICARUS algorithm and then compared the resulting cloud-top pressure and optical depth (hereafter P_{sim} and τ_{sim}) to P_{ISCCP} and τ_{ISCCP}. Comparisons were also made to LBTM-derived cloud-top pressures and visible optical depths (hereafter P_{LBTM} and τ_{LBTM}). Using data from the year 2000, the $P_{\mathrm{sim}} - \tau_{\mathrm{sim}}$ statistics compare much better to ISCCP than simply comparing the unaltered P and τ derived from the ground-based ARM data (hereafter P_{obs} and τ_{obs}) to P_{ISCCP} and τ_{ISCCP}. However, the statistics of P_{sim} and τ_{sim} when compared to P_{ISCCP} and τ_{ISCCP} were in some ways similar to the differences found between GCMs and ISCCP in Z05, suggesting that the ISCCP simulator should be examined more thoroughly. Such an examination is conducted here.

Our hypothesis is that if observed cloud property and thermodynamic profiles are provided as input to the ISCCP simulator, then the simulator will produce P_{sim} and τ_{sim} similar to P_{ISCCP} and τ_{ISCCP}. Our goal is not to evaluate the validity of ISCCP. Our goal is to evaluate the degree to which ICARUS simulates ISCCP when given an

observed physical distribution of cloud occurrence and cloud properties.

2. Data and technique

The simulation of ISCCP with ICARUS is a two-step process. The 10.5-μm radiance or brightness temperature of the clear and cloudy atmosphere is parameterized using a vertical profile of cloud properties and thermodynamics using a simple radiative transfer model similar to that reported in Klein and Jakob (1999). Then, $P_{\rm sim}$ and $\tau_{\rm sim}$ are derived using ISCCP-like assumptions (Rossow et al. 1996). To validate the first step in this process (the ICARUS parameterization of the IR radiances), we calculate clear and cloudy top-of-atmosphere (TOA) radiances using the more complete Moderate Spectral Resolution Atmospheric Transmittance (MODTRAN) model (Berk et al. 1989). We then applied the second component of the ICARUS algorithm to these MODTRAN radiances to calculate $P_{\rm MODT}$ and $\tau_{\rm MODT}$ reported on below.

To calculate $P_{\rm sim}$ from the parameterized IR radiance, the temperature at cloud top is calculated from the IR brightness temperatures and column visible optical depth assuming, as with ISCCP that only a single layer of cloud exists in the vertical column. Then, $P_{\rm sim}$ is set equal to the lowest pressure (highest altitude) in the troposphere for which the temperature of the input sounding matches the derived cloud-top temperature. Finally, $\tau_{\rm sim}$ is set equal to $\tau_{\rm obs}$ in all cases except for optically very thin clouds for which the single-layer cloud retrieval fails. In this case, a nominal value of optical depth is assigned following ISCCP documentation (Rossow et al. 1996). So, except for profiles with $\tau_{\rm obs} < 0.5$, $\tau_{\rm sim} = \tau_{\rm obs}$.

The primary goal of ICARUS is to calculate a value for $P_{\rm sim}$ that the ISCCP algorithm would derive from an atmospheric column with similar physical properties as that of the simulation. $P_{\rm ISCCP}$ can differ substantially from $P_{\rm obs}$, particularly where multiple cloud layers exist in the column and the highest cloud is transmissive to thermal IR radiation. In such situations, $P_{\rm sim}$ is higher (at lower altitudes) than $P_{\rm obs}$ and typically results in $P_{\rm sim}$ at middle levels of the atmosphere when the true cloud-top pressure is at high levels. $P_{\rm ISCCP}$ can also differ substantially from $P_{\rm obs}$ when a cloud layer is located beneath a strong temperature inversion. When this occurs, the $P_{\rm sim}$ is lower (at a higher altitude) than the $P_{\rm obs}$ and typically results in $P_{\rm sim}$ at middle levels of the atmosphere when $P_{\rm obs}$ is at low levels.

The area of focus for this study is the ARM SGP site in Oklahoma (Ackerman and Stokes 2003). Ground-based zenith-pointing cloud radar and lidar data have been collected continuously at that location since 1997. The cloud microphysical and radiative property profiles are derived using a combination of vertically pointing radar reflectivity, Doppler velocity, lidar-derived cloud boundaries, and liquid water paths derived from microwave radiometer measurements (M06). Using the derived cloud property profiles and observed thermodynamic profiles, $P_{\rm sim}$ and $\tau_{\rm sim}$ are calculated using the ICARUS component of the ISCCP simulator. The derived cloud microphysical and radiative property profiles have been validated against aircraft in situ data, surface radiometric fluxes, and TOA radiometric fluxes (M06; Mace and Benson 2008). Additionally, the M06 column optical depths compared favorably with optical depths derived from Multifilter Rotating Shadowband Radiometer (MFRSR) measurements using a technique described by Min and Harrison (1996). We also use the Min and Harrison optical depths (hereafter $\tau_{\rm MFRSR}$) below as an additional comparison dataset. It is important to note that the M06 methodology used to derive cloud properties from ground-based data does not use radiometric fluxes in either the solar or IR spectra as input. The common element between the M06 and Min and Harrison (1996) methods is that both use liquid water paths derived from the microwave radiometer at the SGP site.

As a reminder, our hypothesis is that if accurate observed cloud property and thermodynamic profiles are provided as input to the ISCCP simulator, then the simulator will produce $P_{\rm sim}$ and $\tau_{\rm sim}$ similar to $P_{\rm ISCCP}$ and $\tau_{\rm ISCCP}$. There are at least two significant challenges in testing our hypothesis. First, we assume that the cloud properties input to ICARUS represent a realistic version of the actual cloud properties for a given 5-min period. Because we use active remote sensing observations and soundings, the vertical locations of the cloud layers and the thermodynamics in the vertical column are reasonably certain. The vertical distribution of cloud properties is less certain. However, radiative closure studies at the TOA and surface suggest that the cloud radiative properties have minimal bias (M06; Mace and Benson 2008). Thus, we assume that, while any given profile will have significant uncertainty, statistics derived from many profiles will allow meaningful comparisons to emerge from the noise.

The second challenge in testing our hypothesis is that the ISCCP measurements are derived from spatially distributed radiances collected instantaneously, while the ARM data are collected as a function of time at a single point. Clearly, situations that have highly variable cloud fields in either space or time are not reasonable candidates for comparison. Therefore, we implement a strict set of criteria that a particular case must satisfy. We define a case to be the union of an interval in time during which the ARM data are averaged centered on the ISCCP observation time

TABLE 1. Sampling permutations of ISCCP and ARM data for examining the variability and covariance of the temporal and spatial averages.

ARM sampling versions	ISCCP sampling versions
Nearest 5-min average to nominal ISCCP measurement time	Nearest pixel to the SGP ARM central facility
The 30-, 60-, 90-, 120-min averages centered on the ISCCP measurement time	The 100- and 250-km-square domain centered on the ARM SGP (5 and 12 pixels, respectively)
Random sample of three of the six 5-min average profiles that comprise a 30-min average	Random sample of three of the five pixels that comprise the 100-km domain
Random sample of 6 of the 12 five-minute average profiles that comprise a 60-min average	Random sample of 6 of the 12 pixels in the 250-km domain

with a set of ISCCP retrievals that are averaged from within a geographic rectangular domain centered on the SGP site. To test the validity of the sampling statistics, we use variable time and space intervals as described below. For a case to be used in the comparison, that case had to have met all of the following criteria:

1) All ISCCP pixels within a 100-km averaging domain reported the presence of cloud.
2) The standard deviation of P_{ISCCP} in a 100-km domain must have been less than 100 mb.
3) All ARM 5-min profiles during a 1-h averaging period had to have contained cloud at some level.
4) All τ_{obs} during a 1-h averaging period were limited to values between 1 and 100.

We use the reported daylight P_{ISCCP} and τ_{ISCCP} from the ISCCP D series dataset from 1997 to 2002. These data are reported at 3-h intervals and sampled every 30 km from the native geostationary satellite data. We average the ISCCP data within 100- and 250-km domains centered on the ARM SGP central facility as well as use the single ISCCP pixel nearest to the SGP site to create versions of $\overline{P}_{\mathrm{ISCCP}}$ and $\overline{\tau}_{\mathrm{ISCCP}}$. These are compared with similar quantities derived from the ground-based data that have been averaged during 30-, 60-, 90-, and 120-min periods centered on the ISCCP nominal measurement time as well as the single 5-min averaged profile nearest to the ISCCP measurement time to create several versions of $\overline{P}_{\mathrm{obs}}$ and $\overline{\tau}_{\mathrm{obs}}$ and $\overline{P}_{\mathrm{sim}}$ and $\overline{\tau}_{\mathrm{sim}}$. The LBTM retrievals are used at the time nearest to the center of the averaging interval and for the single 0.3° spatial average [18 visible Geostationary Operational Environmental Satellite (GOES) pixels] nearest to the SGP central facility. In all of these various permutations, we use the set of events that pass the homogeneity criteria at the 100-km and 1-h averaging intervals. In other words, we do not define a new set of cases for each permutation but use the same set of cases in all comparisons.

Comparing ground-based and satellite measurements always raises questions of sampling uncertainty—especially when conducted between quantities derived from cloud fields that tend to be highly variable in both space and

TABLE 2. Correlation matrix of $\overline{P}_{\mathrm{sim}}$ and $\overline{P}_{\mathrm{ISCCP}}$ (values below the X diagonal) and $\overline{\tau}_{\mathrm{sim}}$ and $\overline{\tau}_{\mathrm{ISCCP}}$ (values above the X diagonal) for the indicated spatial and temporal averaging intervals. The quantities in parentheses show the correlation of the base-10 logarithms of the optical depths.

	ISCCP nearest pixel	ISCCP 100 km	ISCCP 250 km	ICARUS closest	ICARUS 30 min	ICARUS 60 min	ICARUS 90 min	ICARUS 120 min
ISCCP 50 km	X	0.952 (0.966)	0.873 (0.892)	0.392 (0.521)	0.439 (0.605)	0.476 (0.631)	0.443 (0.623)	0.448 (0.633)
ISCCP 100 km	0.993	X	0.919 (0.924)	0.378 (0.512)	0.433 (0.602)	0.477 (0.635)	0.441 (0.621)	0.447 (0.634)
ISCCP 250 km	0.951	0.960	X	0.376 (0.545)	0.421 (0.628)	0.459 (0.649)	0.433 (0.648)	0.447 (0.663)
ICARUS closest in time	0.706	0.705	0.685	X	0.917 (0.859)	0.819 (0.773)	0.878 (0.824)	0.833 (0.781)
ICARUS 30 min	0.780	0.784	0.795	0.708	X	0.902 (0.907)	0.931 (0.891)	0.882 (0.850)
ICARUS 60 min	0.792	0.795	0.802	0.711	0.952	X	0.909 (0.933)	0.892 (0.946)
ICARUS 90 min	0.794	0.796	0.812	0.710	0.977	0.975	X	0.972 (0.976)
ICARUS 120 min	0.792	0.794	0.812	0.704	0.967	0.973		X

time. While the criteria listed above that qualifies an event for comparison is stringent and ensures that only overcast and rather homogenous events are used in compiling statistics, we considered various renditions of temporal and spatial averaging (Table 1) to quantify the variability in the temporal statistics, the spatial statistics, and the reasonableness of comparing the spatially and temporally averaged quantities. In addition to the five temporal and three spatial averages listed above, we add to them a random sampling of the 30- and 60-min ARM data and a random sampling of the 100- and 250-km ISCCP domains.

In Table 2, we consider the degree to which these various sampling permutations covary by examining the correlation coefficients of $\overline{P}_{\text{sim}}$ and $\overline{P}_{\text{ISCCP}}$ in the bottom diagonal of the matrix and also $\overline{\tau}_{\text{sim}}$ and $\overline{\tau}_{\text{ISCCP}}$ in the top diagonal. We make the assumption that the most reasonable comparison between spatial and temporal statistics should be for the combination of temporal and spatial averaging that presents the strongest correlation between space and time in cloud-top pressure and optical depth. Overall, we find only marginal differences in the correlation coefficients for the various combinations of temporal and spatial averaging, suggesting that our initial screening of the events successfully captured fairly homogenous cloud fields. We find that the correlation coefficients tend to rise as the temporal averaging times of the ARM data increase. For the optical depth, the improvement seems to reach a maximum at the 1-h averaging period. For the spatial averaging, we find that ISCCP has the strongest correlation with the ARM data for the 250-km domain for cloud-top pressure, while for optical depth a stronger correlation is found for the 100-km averaging domain. Since the difference in the cloud-top pressure correlations is only slight and our emphasis later will be on optical depth differences, in the following discussion we use the 100-km spatial averages of ISCCP with the 60-min temporal averages of the ARM data. However, we also present the standard deviations of the various sampling permutations where appropriate as a measure of the sensitivity of the results to the sampling choice.

To build further confidence in the temporal–spatial comparison, we consider whether the magnitude of the differences in either optical depth or cloud-top pressure between the spatial and temporal averaging is a function of the variability of either quantity in space or time. If, for instance, we find that the differences between $\overline{\tau}_{\text{sim}}$ and $\overline{\tau}_{\text{ISCCP}}$ are correlated to the magnitude of the case-by-case spatial standard deviation of τ_{ISCCP}, then any systematic differences we find may be as much due to statistical offsets in the temporal and spatial averages as to real differences in the algorithms. Figure 1 illustrates one of these relationships. We find only minimal correlation (<0.15) in the magnitude of the differences of optical depth and cloud-top pressure and the variability of these quantities in either space or time, lending further confidence to our comparison of temporal and spatial statistics of these carefully selected cases. We continue to address this topic as we proceed.

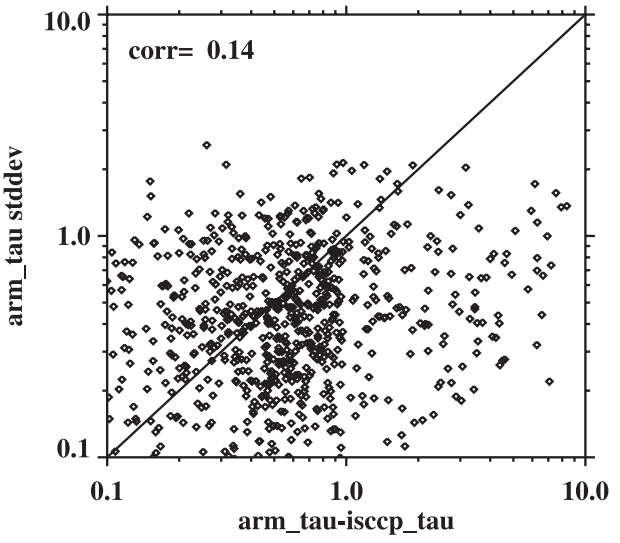

FIG. 1. The normalized standard deviation during the 1-h averaging period of ARM optical depth compared to the absolute value of the fractional difference between 1-h averaged ARM optical depth and the 100-km averaged ISCCP optical depth.

3. Results

In Fig. 2 we compare various renditions of $\overline{P}$, and in Fig. 3 the $\overline{\tau}$ quantities are compared. Regression statistics for $\overline{P}$ and $\overline{\tau}$ are listed in Tables 3 and 4, respectively. In the comparison of ISCCP with LBTM, a lack of any significant bias suggests that the two satellite algorithms tend to produce reasonably similar results, while the scatter in the comparisons likely arises from differences in the algorithms and from comparing the spatially averaged ISCCP to the 0.3° LBTM product nearest to the SGP central facility. For both ISCCP and LBTM, the improvement relative to observations in the comparison of $\overline{P}_{\text{sim}}$ is evident. The $\overline{P}_{\text{obs}}$ when compared to the satellite products shows two clusters of points in the lower and upper troposphere, with fewer points in the middle troposphere recorded by the active remote sensors. ICARUS correctly moves some fraction of those points into the middle troposphere as expected. Interestingly, while the normal deviation is slightly larger, the linear correlation coefficient of $\overline{P}_{\text{sim}}$ with $\overline{P}_{\text{ISCCP}}$ and $\overline{P}_{\text{sim}}$ with $\overline{P}_{\text{LBTM}}$ is nearly identical to that found comparing $\overline{P}_{\text{LBTM}}$ with $\overline{P}_{\text{ISCCP}}$. This suggests that the alterations of cloud-top pressure performed by ICARUS are performing as well as could be expected.

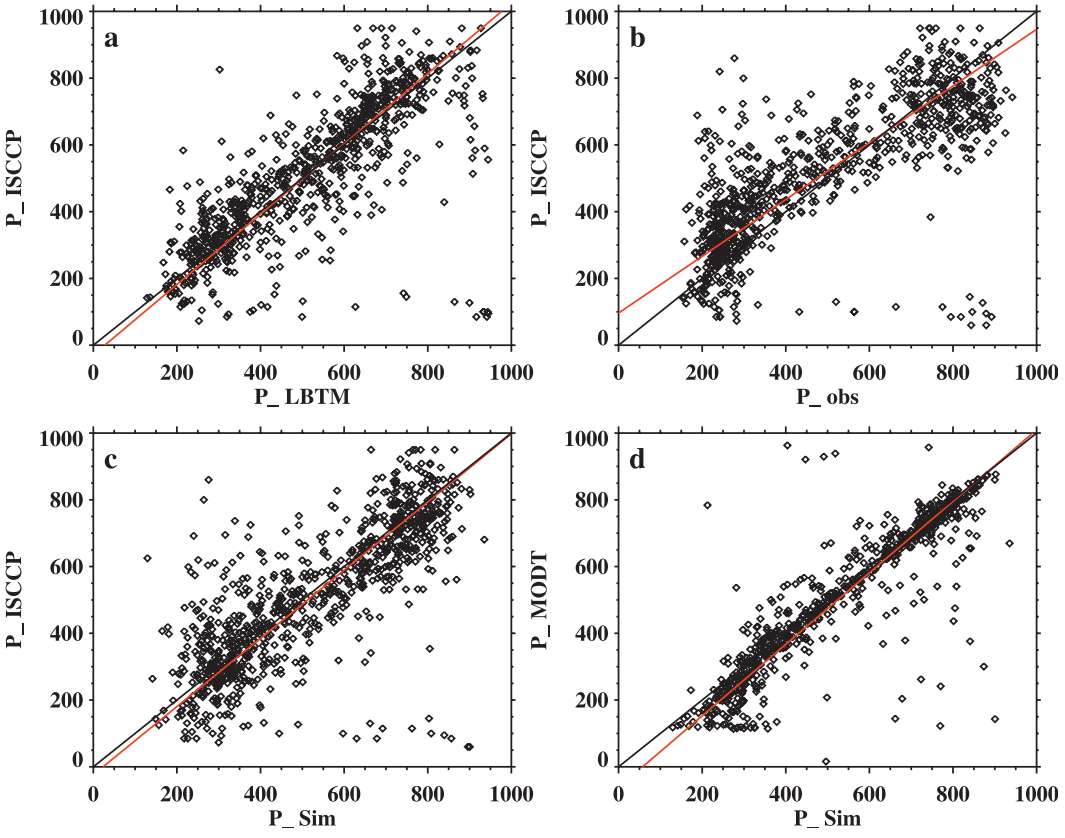

FIG. 2. Comparisons of cloud-top pressure (mb) between (a) $\overline{P}_{\mathrm{ISCCP}}$ and $\overline{P}_{\mathrm{LBTM}}$, (b) $\overline{P}_{\mathrm{ISCCP}}$ and $\overline{P}_{\mathrm{obs}}$, (c) $\overline{P}_{\mathrm{ISCCP}}$ and $\overline{P}_{\mathrm{sim}}$ and (d) $\overline{P}_{\mathrm{MODT}}$ and $\overline{P}_{\mathrm{sim}}$. The red line in each plot is a linear regression, and the black line is 1:1.

We compare the various renditions of $\overline{\tau}$ in Fig. 3. As in Fig. 2, the comparison between LBTM and ISCCP shows minimal bias. The comparison between $\overline{\tau}_{\mathrm{MFRSR}}$ and $\overline{\tau}_{\mathrm{obs}}$ show good agreement also with a slight positive bias in $\overline{\tau}_{\mathrm{obs}}$ that seems to be associated with higher optical depth events. Comparing the ground-based techniques to the satellite techniques, however, reveals a bias with the satellite retrievals of $\overline{\tau}$ on average 10% lower than the ground-based quantities.

The $\overline{P} - \overline{\tau}$ histograms that are derived from the approximately 1000 cases that pass our variability criteria are shown in Fig. 4, where we simplify the 42 ISCCP bins used by Rossow and Schiffer to the nine ISCCP cloud types used by many others, including Z05. As before, we

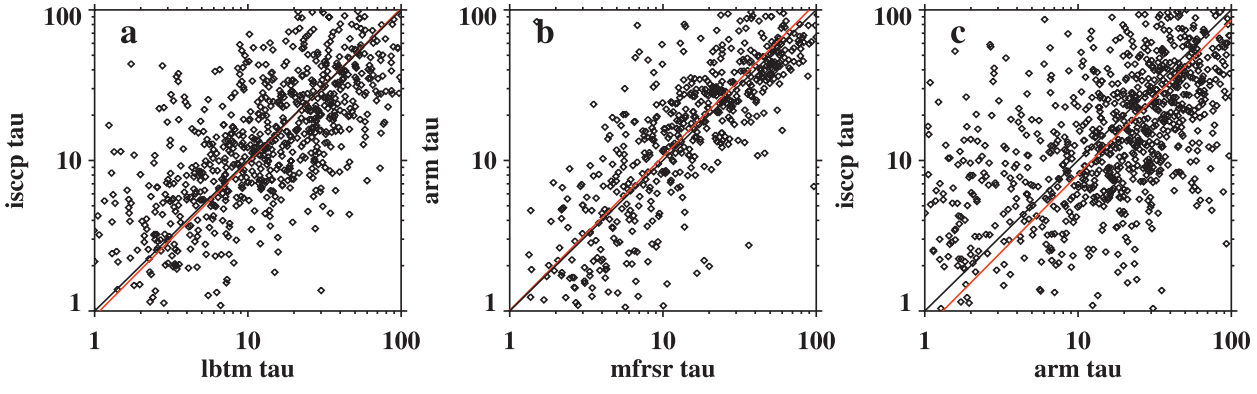

FIG. 3. Comparisons of total optical depth: (a) $\overline{\tau}_{\mathrm{ISCCP}} - \overline{\tau}_{\mathrm{LBTM}}$, (b) $\overline{\tau}_{\mathrm{ARM}} - \overline{\tau}_{\mathrm{MFRSR}}$, and (c) $\overline{\tau}_{\mathrm{ISCCP}} - \overline{\tau}_{\mathrm{ARM}}$. The red line in each plot is a linear regression, and the black line is 1:1. Recall that $\overline{\tau}_{\mathrm{ARM}} \cong \overline{\tau}_{\mathrm{sim}}$.

TABLE 3. Statistics of the cloud-top pressure comparisons seen in Fig. 2. All quantities are shown in millibars, except for the number of events.

Comparison	Number	Bias	Linear correlation	Linear slope	Normal deviation
$P_{ISCCP} - P_{LBTM}$	1000	−0.57	0.80	1.06	19.06
$P_{ISCCP} - P_{obs}$	1042	21.94	0.78	0.88	50.17
$P_{LBTM} - P_{obs}$	919	28.6	0.81	0.84	46.74
$P_{ISCCB} - P_{sim}$	1042	−16.90	0.80	1.05	26.57
$P_{LBTM} - P_{sim}$	919	−11.01	0.81	1.00	25.48
$P_{ISCCP} - P_{MODT}$	900	−1.27	0.75	1.00	17.29
$P_{LBTM} - P_{MODT}$	809	4.31	0.74	0.94	24.18
$P_{MODT} - P_{sim}$	900	−16.96	0.89	1.06	20.01

TABLE 4. Statistics of the optical depth comparisons seen in Fig. 3.

Comparison	Number	Bias	Linear correlation	Linear slope	Normal deviation
$\tau_{ISCCP} - \tau_{LBTM}$	789	−0.01	0.67	1.02	0.07
$\tau_{obs} - \tau_{MFRSR}$	555	0.03	0.79	1.02	0.08
$\tau_{ISCCP} - \tau_{obs}$	891	−0.09	0.59	1.03	0.10
$\tau_{LBTM} - \tau_{obs}$	789	−0.10	0.68	0.98	0.06
$\tau_{LBTM} - \tau_{MFRSR}$	492	−0.09	0.75	1.00	0.04
$\tau_{ISCCP} - \tau_{MFRSR}$	555	−0.05	0.64	1.06	0.06

find that ICARUS brings the ground-based observations into closer agreement with ISCCP. The most substantial changes made by ICARUS appear to be in the high- and middle-cloud categories, where ICARUS correctly moves cloud-top pressures into the middle troposphere. Similarly, large changes can be seen in the comparisons with LBTM. However, in comparing the $\overline{P}_{sim} - \overline{\tau}_{sim}$ statistics with the satellites in other categories, we find interesting differences. Specifically, the $\overline{P}_{sim} - \overline{\tau}_{sim}$ histograms show that the frequency of optically thick clouds are greater than the satellite algorithms, while the frequency of optically intermediate clouds are less than the satellite algorithms in approximately equal proportions. A difference that is common between the two satellite algorithms and the surface is that both LBTM and ISCCP diagnose less than half of the optically thickest lower tropospheric cloud type (hereafter stratus) compared to the surface results, while just the opposite is found in the optically intermediate middle troposphere cloud type (hereafter altostratus). The ISCCP and ICARUS results seem to agree in their frequency of optically thick high clouds (hereafter deep clouds). However, LBTM reports substantially fewer of these deep layers but diagnoses proportionally more of the optically thick midlevel clouds, commonly referred to as nimbostratus.

Examination of the panels in Fig. 4 is instructive. However, one must be cautious not to place too much stock in the quantitative agreement in Fig. 4, because there is potential for compensating errors that adjust the counts in a particular category that depends on factors unrelated to the agreement between the ground-based and satellite algorithms in that category. To illustrate this point, we list in Table 5 the fraction of cases that agreement is found between ISCCP or LBTM and ICARUS. Parts (c) and (f) show the fraction of the number of cases in part (a) for which ARM and ISCCP or LBTM agree for a particular type without application of the ICARUS algorithm. Parts (d) and (g) illustrate the effect of the ICARUS cloud-top pressure corrections. Part (b) illustrates the agreement among the two satellite algorithms. We find the agreement to range from approximately one-half to three-quarter of cases in most categories with the exception of the stratus, cumulus, and altocumulus classes. The small number of cases of altocumulus and cumulus make the interpretation of these results problematic. It is clear, however, that LBTM and ISCCP diagnose stratus clouds differently. Forty percent of the time that ISCCP diagnoses stratus, LBTM diagnoses nimbostratus, suggesting that under these circumstances the interpretation of cloud-top pressure is the issue.

We find that when ISCCP or LBTM diagnoses a high cloud, the ICARUS algorithm has little effect and actually acts to reduce the agreement in the cirrostratus and deep categories. This can be understood by considering that the role of ICARUS is to move the cloud-top pressure downward in altitude to higher cloud-top pressure values from its physical location to match the pressure of the column radiating temperature. ICARUS would not simulate the cloud-top pressure to be at lower pressures than it already is physically determined to be. The decrease in agreement in the cirrostratus and the deep categories are due to the presence of thin cirrus layers overlying thicker layers where ICARUS adjusts the cloud-top pressure downward so that the event is counted in the adjacent cloud-top pressure bin. While we also find that ICARUS has little influence on the optically thick stratus and stratocumulus agreement statistics, ICARUS does seem to successfully improve the altostratus and nimbostratus agreement, perhaps due to the upward shift in altitude for cloud layers under inversions.

The real question is why the overall percentage agreements in parts (c) and (f) of Table 5 are so small. One could argue, perhaps, that we should not expect the ground-based ICARUS results to agree any better than the two satellite algorithms agree. However, even with that criterion, we find in most cases that the agreement between ICARUS and the satellite results is smaller. On the other hand, the agreement with ICARUS applied to ARM observations significantly improves the agreement in the midlevel bins, while slightly decreasing the agreement in the deep cloud category. We are reasonably certain that the vertical distribution of cloud occurrence in the ARM

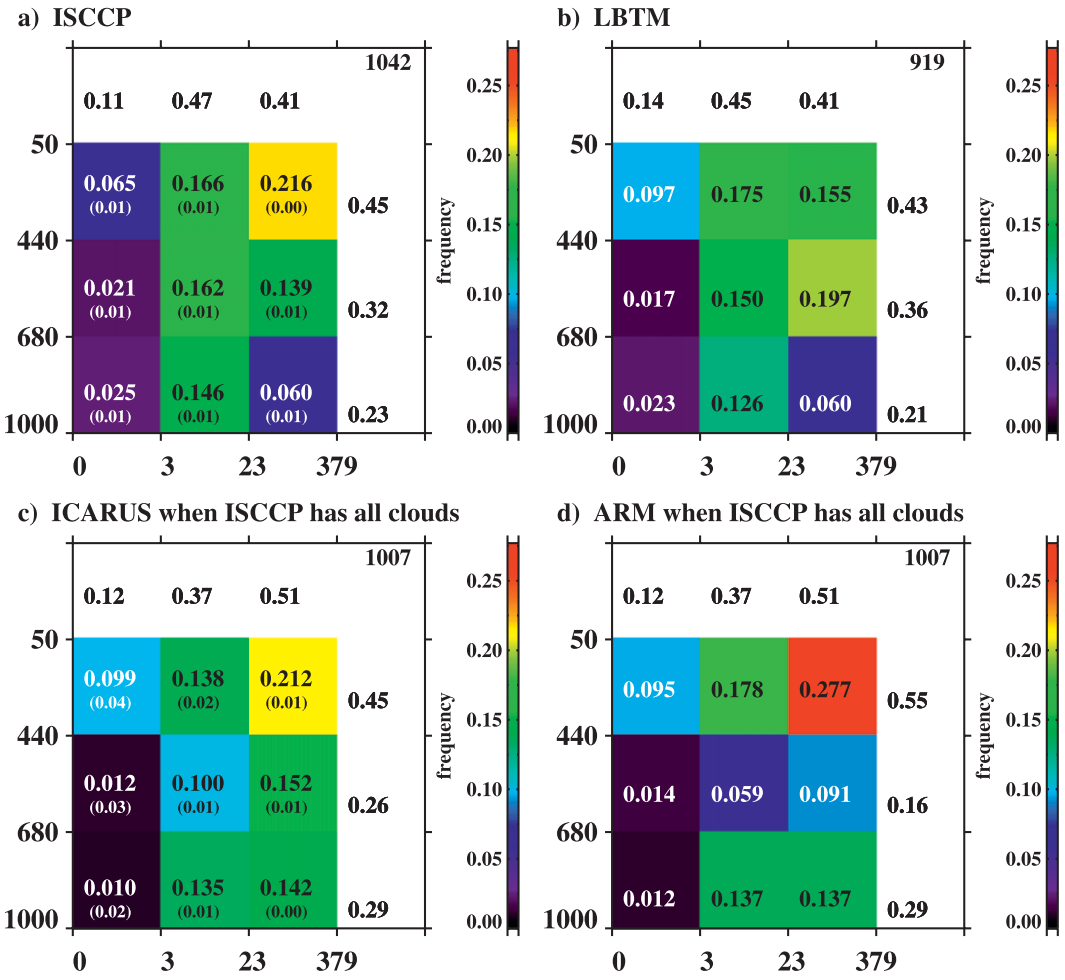

FIG. 4. The $\overline{P} - \overline{\tau}$ histograms for the 9 ISCCP cloud type classes with the numerical fraction of the total number of cases (listed in the upper-right corner of each plot). Coverage is between 1997 and 2002, and the events meet the criteria listed in section 2. The fractions in the right-most column are a summation of the fractional occurrence in each optical depth class. The fractions across the top are summations of the fractions in each cloud-top pressure class. (a) ISCCP, (b) LBTM, (c) ICARUS applied to ARM events, and (d) ARM events before application of ICARUS. In (a) and (c), the numbers in parentheses are the standard deviations of the fractions in each box of the sampling permutations listed in Table 1.

data is as correct as it could be—given a continuously operating millimeter radar and microwave radiometer and other ancillary data used as input to the algorithms. We have established by comparing with MFRSR above and elsewhere (M06) that the retrieved ARM radiative property profile is largely unbiased. We have also established that the ICARUS radiative parameterization is in reasonable agreement with similar quantities calculated from a more complicated radiative model, and that the differences in the $\overline{P}_{\mathrm{MODTRAN}} - \overline{\tau}_{\mathrm{MFRSR}}$ with $\overline{P}_{\mathrm{sim}} - \overline{\tau}_{\mathrm{sim}}$ are much smaller than the differences in any of the ground-based results with either of the satellite results.

To help shed light on this issue and to examine more closely the differences between the algorithms, we consider the $\overline{P} - \overline{\tau}$ statistics from ISCCP and LBTM when the ICARUS algorithm diagnoses clouds in each of the nine categories (Figs. 5 and 6). We present similar distributions in Fig. 7 for $\overline{P}_{\mathrm{MODTRAN}} - \overline{\tau}_{\mathrm{MFRSR}}$ to understand the uncertainty in $\overline{P}$ due to the ICARUS radiative parameterization. The depiction of the statistics in Figs. 5–7 is the converse approach taken in Table 5 and more closely represents the methodology of a model evaluation where the model output would be converted into an ISCCP equivalent using ICARUS. In other words, if a model were to predict the cloud occurrence and properties to within a similar uncertainty as the ground-based results, Figs. 5 and 6 would show the level of agreement that could be expected with ISCCP and LBTM.

Beginning with the high clouds and moving from the optically thickest to thinnest, we find that when ICARUS

TABLE 5. Evaluation of the agreement statistics when (a)–(d) ISCCP and (e)–(g) LBTM diagnose a particular cloud type. (a) Number of ISCCP cases; (b) the percentage of the ISCCP cases where $\overline{P}_{\mathrm{LBTM}} - \overline{\tau}_{\mathrm{LBTM}}$ are in the same class as ISCCP; (c) as in (b), but $\overline{P}_{\mathrm{obs}} - \overline{\tau}_{\mathrm{obs}}$; and (d) as in, (b) but $\overline{P}_{\mathrm{sim}} - \overline{\tau}_{\mathrm{sim}}$. The percentages in parentheses in part (d) show $\overline{P}_{\mathrm{MODT}} - \overline{\tau}_{\mathrm{sim}}$. (e) As in (a), but LBTM. (f) As in (c), but LBTM. (g) As in (d), but LBTM.

(a) ISCCP number of cases	$\overline{\tau}_{\mathrm{ISCCP}} < 3.6$	$3.6 < \overline{\tau}_{\mathrm{ISCCP}} < 23$	$\overline{\tau}_{\mathrm{ISCCP}} > 23$
$\overline{P}_{\mathrm{ISCCP}} < 440$	68	173	225
$680 < \overline{P}_{\mathrm{ISCCP}} < 440$	22	169	145
$\overline{P}_{\mathrm{ISCCP}} > 680$	26	152	62
(b) LBTM % agree	$\overline{\tau}_{\mathrm{ISCCP}} < 3.6$	$3.6 < \overline{\tau}_{\mathrm{ISCCP}} < 23$	$\overline{\tau}_{\mathrm{ISCCP}} > 23$
$\overline{P}_{\mathrm{ISCCP}} < 440$	63	53	57
$680 < \overline{P}_{\mathrm{ISCCP}} < 440$	16	47	73
$\overline{P}_{\mathrm{ISCCP}} > 680$	24	49	36
(c) ARM % agree	$\overline{\tau}_{\mathrm{ISCCP}} < 3.6$	$3.6 < \overline{\tau}_{\mathrm{ISCCP}} < 23$	$\overline{\tau}_{\mathrm{ISCCP}} > 23$
$\overline{P}_{\mathrm{ISCCP}} < 440$	53	47	79
$680 < \overline{P}_{ISCCP} < 440$	14	18	37
$\overline{P}_{\mathrm{ISCCP}} > 680$	8	47	66
(d) ICARUS % agree	$\overline{\tau}_{\mathrm{ISCCP}} < 3.6$	$3.6 < \overline{\tau}_{\mathrm{ISCCP}} < 23$	$\overline{\tau}_{\mathrm{ISCCP}} > 23$
$\overline{P}_{\mathrm{ISCCP}} < 440$	53 (57)	43 (47)	69 (65)
$680 < \overline{P}_{ISCCP} < 440$	9 (0)	27 (26)	57 (52)
$\overline{P}_{\mathrm{ISCCP}} > 680$	4 (4)	47 (44)	69 (72)
(e) LBTM number of cases	$\overline{\tau}_{\mathrm{LBTM}} < 3.6$	$3.6 < \overline{\tau}_{\mathrm{LBTM}} < 23$	$\overline{\tau}_{\mathrm{LBTM}} > 23$
$\overline{P}_{\mathrm{LBTM}} < 440$	89	161	116
$680 < \overline{P}_{\mathrm{LBTM}} < 440$	16	138	181
$\overline{P}_{\mathrm{LBTM}} > 680$	21	116	55
(f) ARM % agree	$\overline{\tau}_{\mathrm{LBTM}} < 3.6$	$3.6 < \overline{\tau}_{\mathrm{LBTM}} < 23$	$\overline{\tau}_{\mathrm{LBTM}} > 23$
$\overline{P}_{\mathrm{LBTM}} < 440$	50	37	87
$680 < \overline{P}_{\mathrm{LBTM}} < 440$	0	21	30
$\overline{P}_{\mathrm{LBTM}} > 680$	10	53	75
(g) ICARUS % agree	$\overline{\tau}_{\mathrm{LBTM}} < 3.6$	$3.6 < \overline{\tau}_{\mathrm{LBTM}} < 23$	$\overline{\tau}_{\mathrm{LBTM}} > 23$
$\overline{P}_{\mathrm{LBTM}} < 440$	50	34	77
$680 < \overline{P}_{\mathrm{LBTM}} < 440$	0	34	45
$\overline{P}_{\mathrm{LBTM}} > 680$	0	51	75

simulates a deep cloud, 62% of the time this diagnosis will agree with ISCCP (we refer to this as the hit rate, i.e., $H^{\mathrm{Deep}}_{\mathrm{ISCCP}} = 0.62$). Of the 38% of the time ICARUS places a cloud in the deep category in disagreement with ISCCP (we will refer to this as the miss rate, i.e., $M^{\mathrm{Deep}}_{\mathrm{ISCCP}} = 0.38$), ISCCP will report the majority of those in the next lower optical depth category, i.e., cirrostratus. A similar pattern is found with LBTM, except that $H^{\mathrm{Deep}}_{\mathrm{LBTM}} = 0.49$. For the cirrostratus category, $H^{\mathrm{Cirrostratus}}_{\mathrm{ISCCP}} = 0.53$ and $H^{\mathrm{Cirrostratus}}_{\mathrm{LBTM}} = 0.43$. ISCCP and LBTM diagnose a larger $\overline{P}$ in about the same fraction of cases; although, ISCCP places more events into a larger $\overline{\tau}$ bin, while LBTM places more into a smaller $\overline{\tau}$ bin. The quantity $H^{\mathrm{Cirrus}}_{\mathrm{ISCCP}}$ is significantly smaller than $H^{\mathrm{Cirrus}}_{\mathrm{LBTM}}$, with both algorithms placing the misses at larger $\overline{\tau}$ than at larger $\overline{P}$. For the high-cloud category, we find that the MODTRAN results show that the ICARUS parameterization places the clouds correctly more than 90% of the time. So, the misses in the high-cloud category are primarily due to differences in the interpretation of $\overline{\tau}$.

In the middle $\overline{P}$ classes, we see significant differences in skill from high to low $\overline{\tau}$ with the two satellite algorithms showing very similar hit rates. For the thicker classes at middle levels, the hit rate is on the order of 0.5 with the majority of misses being placed at smaller $\overline{\tau}$ for the nimbostratus category and at larger $\overline{\tau}$ for the altostratus. While there are only 10–12 cases for the altocumulus category, the miss rate seems quite high—much higher than for optically thin cirrus. One-half of the misses (80% for LBTM) are being diagnosed to have larger $\overline{\tau}$, while ISCCP diagnoses many of these events as cirrus. The comparison between MODTRAN and ICARUS shows that more uncertainty exists in the ICARUS parameterization in the middle levels with about 15% of the occurrences being placed in the wrong $\overline{P}$ category for nimbostratus and

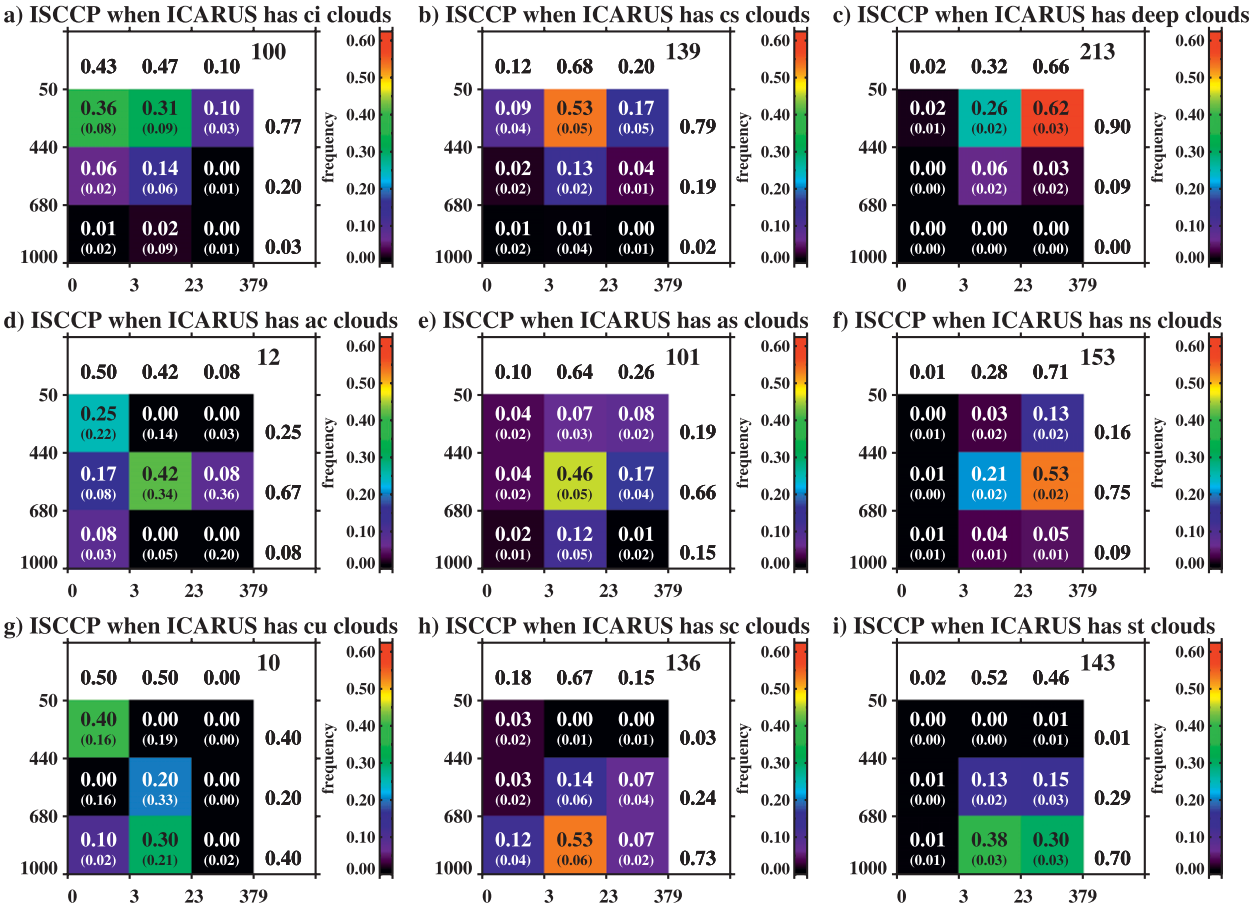

FIG. 5. The distribution of $\overline{P}_{\mathrm{ISCCP}} - \overline{\tau}_{\mathrm{ISCCP}}$ when $\overline{P}_{\mathrm{sim}} - \overline{\tau}_{\mathrm{sim}}$ is diagnosed in each of the nine cloud class bins. Each histogram is as described in Fig. 4: (a) $\overline{P}_{\mathrm{ISCCP}} - \overline{\tau}_{\mathrm{ISCCP}}$ when $\overline{P}_{\mathrm{sim}} - \overline{\tau}_{\mathrm{sim}}$ is diagnosed as cirrus ($\overline{P}_{\mathrm{sim}} < 440$ hPa and $\overline{\tau}_{\mathrm{sim}} < 3.6$); (b) $\overline{P}_{\mathrm{ISCCP}} - \overline{\tau}_{\mathrm{ISCCP}}$ when $\overline{P}_{\mathrm{sim}} - \overline{\tau}_{\mathrm{sim}}$ is diagnosed as cirrostratus ($\overline{P}_{\mathrm{sim}} < 440$ hPa and $3.6 > \overline{\tau}_{\mathrm{sim}} < 23$); (c) $\overline{P}_{\mathrm{ISCCP}} - \overline{\tau}_{\mathrm{ISCCP}}$ when $\overline{P}_{\mathrm{sim}} - \overline{\tau}_{\mathrm{sim}}$ is diagnosed as deep clouds ($\overline{P}_{\mathrm{sim}} < 440$ hPa and $\overline{\tau}_{\mathrm{sim}} > 23$; (d) $\overline{P}_{\mathrm{ISCCP}} - \overline{\tau}_{\mathrm{ISCCP}}$ when $\overline{P}_{\mathrm{sim}} - \overline{\tau}_{\mathrm{sim}}$ is diagnosed as altocumulus (680 hPa $< \overline{P}_{\mathrm{sim}} > 440$ hPa and $\overline{\tau}_{\mathrm{sim}} < 3.6$); (e) $\overline{P}_{\mathrm{ISCCP}} - \overline{\tau}_{\mathrm{ISCCP}}$ when $\overline{P}_{\mathrm{sim}} - \overline{\tau}_{\mathrm{sim}}$ is diagnosed as altostratus (680 hPa $< \overline{P}_{\mathrm{sim}} > 440$ hPa and $3.6 > \overline{\tau}_{\mathrm{sim}} < 23$); (f) $\overline{P}_{\mathrm{ISCCP}} - \overline{\tau}_{\mathrm{ISCCP}}$ when $\overline{P}_{\mathrm{sim}} - \overline{\tau}_{\mathrm{sim}}$ is diagnosed as nimbostratus (680 hPa $< \overline{P}_{\mathrm{sim}} > 440$ hPa and $\overline{\tau}_{\mathrm{sim}} > 23$); (g) $\overline{P}_{\mathrm{ISCCP}} - \overline{\tau}_{\mathrm{ISCCP}}$ when $\overline{P}_{\mathrm{sim}} - \overline{\tau}_{\mathrm{sim}}$ is diagnosed as cumulus ($\overline{P}_{\mathrm{sim}} > 680$ hPa and $\overline{\tau}_{\mathrm{sim}} < 3.6$); (h) $\overline{P}_{\mathrm{ISCCP}} - \overline{\tau}_{\mathrm{ISCCP}}$ when $\overline{P}_{\mathrm{sim}} - \overline{\tau}_{\mathrm{sim}}$ is diagnosed as stratocumulus ($\overline{P}_{\mathrm{sim}} > 680$ hPa and $3.6 > \overline{\tau}_{\mathrm{sim}} < 23$); and (i) $\overline{P}_{\mathrm{ISCCP}} - \overline{\tau}_{\mathrm{ISCCP}}$ when $\overline{P}_{\mathrm{sim}} - \overline{\tau}_{\mathrm{sim}}$ is diagnosed as stratus ($\overline{P}_{\mathrm{sim}} > 680$ hPa and $\overline{\tau}_{\mathrm{sim}} > 23$).

altostratus, while this uncertainty rises to 40% being placed in the cirrus category for the altocumulus class.

It is surprising that the agreement is not better between the $\overline{P}_{\mathrm{sim}} - \overline{\tau}_{\mathrm{sim}}$ events and the satellite products for the lower tropospheric clouds in the largest $\overline{P}$ classes. We find that $\overline{P}_{\mathrm{sim}} - \overline{\tau}_{\mathrm{sim}}$ has about the same hit rate, with the ISCCP and LBTM in the stratus and stratocumulus classes, respectively, being on the order of 30% for stratus and 50% for stratocumulus. One would expect that $H^{\mathrm{stratus}} > H^{\mathrm{stratocumulus}}$, given that the larger $\overline{\tau}$ would allow for a more accurate determination of $\overline{P}$, although uncertainties associated with surface inversions have been shown to cause these classes of cloud to be erroneously placed in the middle levels by the satellite algorithms. We note that the version of ICARUS that we are using mimics this satellite error by selecting under inversion conditions the lowest pressure (highest altitude) level in the sounding with matching cloud-top temperature (versions of the ISCCP simulator before version 4.0 did not have this feature). In the cumulus category, ISCCP diagnoses a higher $\overline{\tau}$ about half the time (100% for LBTM) and a smaller $\overline{P}$ about 60% of the time (40% of the time for LBTM). The MODTRAN results show that the accuracy of ICARUS to correctly diagnose $\overline{P}$ decreases from 90% in the stratus category to 56% in the cumulus category.

The sources of the discrepancies we illustrate in Figs. 5 and 6 likely arise from a combination of issues. In addition to uncertainties in the derived column radiative properties, the discrepancies noted above could arise from errors in the parameterization of $\overline{P}$ in ICARUS. To test this possibility, we bypassed the ICARUS radiative parameterization of $\overline{P}$ using the MODTRAN radiative

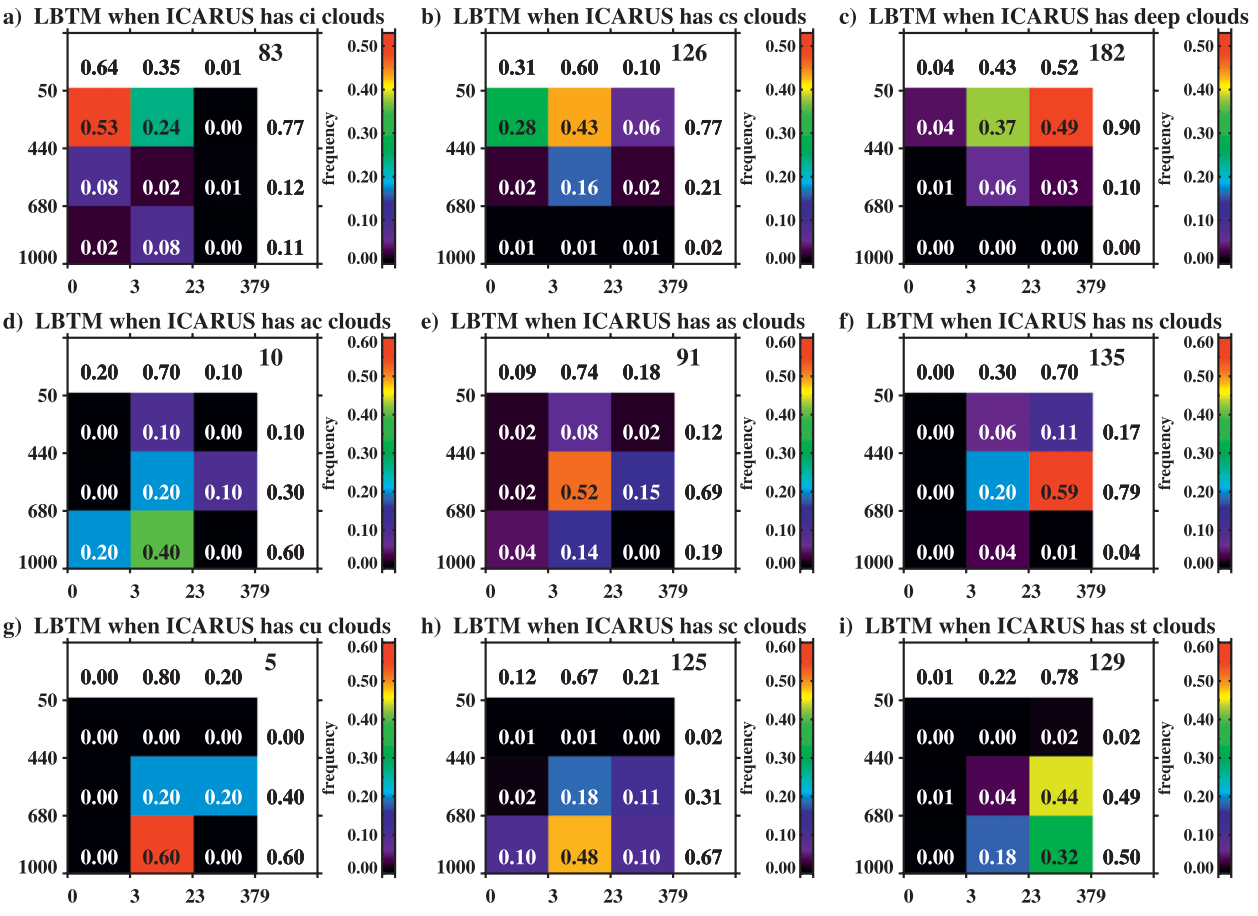

FIG. 6. As in Fig. 5, but for LBTM.

model as discussed above. We found that ICARUS tends to accurately parameterize estimates of $\overline{P}$ in high clouds and in stratus more than 90% of the time. The uncertainty in the parameterized $\overline{P}$ increases as the optical depth of the condensate decreases for middle and low clouds where the errors are on the order of 15% for the nimbostratus, altostratus, and stratocumulus cloud classes. The errors seemed to be larger for altocumulus and cumulus, although the number of events in these optically thin categories is small because of our method for selecting candidate cases.

Comparing the hits and misses in Figs. 5–7, it seems clear that differences in $\overline{\tau}$ are the dominant source of discrepancy between the ground-based and satellite-derived results. The two satellite algorithms tend to have similar hit and miss statistics, although LBTM does seem to have significantly better agreement in the cirrus and stratus categories. However, we do find a large fraction of the optically thin cases being placed by the satellite algorithms into higher optical depth categories. Conversely, a large fraction of the optically thick cases are being diagnosed by the satellite algorithms to occur in the optically intermediate categories.

Similar discrepancies have been reported several times in the literature. Min and Harrison (1996) and Barker et al. (1998) find, as we do, that ISCCP and LBTM optical depths are lower than optical depths derived from ground-based data. While there are numerous sources of uncertainty, optical depth retrievals from satellite radiances are particularly sensitive to assumptions regarding particle phase and single scattering properties as well as instrument calibration (Pincus et al. 1995). In the thicker cloud categories, uncertainties in satellite optical depth retrievals are further magnified because of the asymptotic relationship between reflectance and optical depth (Min and Harrison 1996) where small differences in reflectance equate to very large differences in optical depth as the optical depth becomes large. This uncertainty likely contributes much of the scatter in our comparisons.

However, the cause of the bias remains to be determined. Bias in visible optical depth retrievals from satellite radiances are known to occur because of the horizontal transport of photons when the scale of the satellite retrieval is less than a radiative smoothing scale that

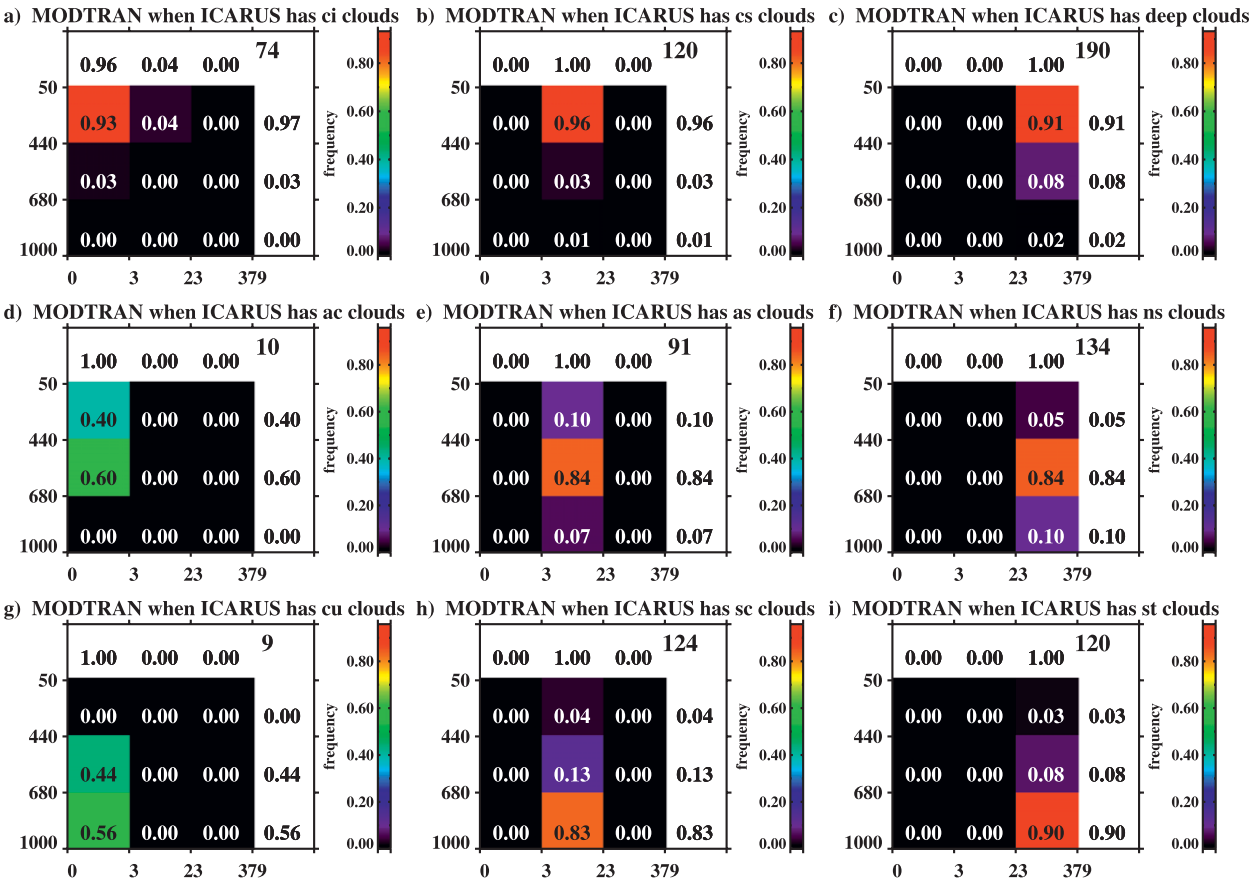

FIG. 7. As in Fig. 4, but for $\overline{P}_{\mathrm{MODT}} - \overline{\tau}_{\mathrm{sim}}$.

depends on cloud geometry (Davis et al. 1997). We have evaluated that source of error and found that the scales of the satellite retrievals averaged over several pixels are significantly larger than the radiative smoothing scale in most circumstances, suggesting that this source of error is not significant.

Another source of optical depth bias is caused by the subpixel variability of optical depth. A satellite radiometer measures pixel-mean radiance and, from this quantity, derives an optical depth that equates to an approximation of the logarithmic mean of the optical depth within the pixel. Because $\exp[\overline{\ln(\tau)}] \leq \overline{\tau}$, the bias is always negative, except when the cloud field is perfectly uniform. Therefore, the exact relationship in any given instance between pixel-mean radiance and the desired pixel-mean optical depth depends on the variability of the cloud field within the pixel (Cahalan et al. 1994). Presently in ICARUS, each subcolumn generated in the downscaling technique is treated as homogeneous and of sufficient size, such that satellites would have no bias in retrieving the true subcolumn optical depth. Thus, there is no facility for adjusting the model-predicted optical depth to account for any biases that might arise due to cloud field variability, because there is no way to know in coarse-resolution models the magnitude of cloud field variability at scales smaller than a satellite pixel. Essentially, it is assumed that $\exp[\overline{\ln(\tau)}] = \overline{\tau}$. Kato and Marshak (2009) most recently evaluate this source of error and show that it is generally small in clouds of moderate optical depths, such as marine stratocumulus. However, many of the cases we consider in this study have optical depths many times larger than those considered by Kato and Marshak. Table 6 shows that the average intra-event normalized standard deviation in optical depth derived from 15 min of MFRSR 20-s resolution retrievals centered on the ISCCP measurement times ranges from a minimum of 25% to nearly 40% in several of the optically thicker cloud categories. The effect of this variability on the optical depth retrieved from the mean radiance and the actual mean optical depth is shown in Fig. 8, where we assume a gamma distribution of optical depths with mean indicated along the abscissa and normalized standard deviations of 0.1, 0.25, 0.5, and 1.0 shown in the curves that extend increasingly to the right of the 1:1 line. Based on the statistics in Table 6 and the

TABLE 6. Mean intra-event normalized standard deviation of optical depth averaged within the 9 $\overline{P} - \overline{\tau}$ bins derived from 15-min averages of MFRSR 20-s optical depth retrievals for 3 cloud-top pressure P categories.

P (mb)	$\tau < 3.6$	$3.6 < \tau < 23$	$\tau > 23$
$P < 440$	0.26	0.35	0.30
$680 < P < 440$	0.34	0.27	0.32
$P > 680$	0.35	0.34	0.39

results in Fig. 8, it would seem that significant bias in comparisons of satellite-derived optical depths derived from pixel-mean radiances and optical depths that approximate true spatial means are likely in real-world situations. For instance, since Table 6 shows that a typical value of the intra-event standard deviation is approximately 30%, we show in Fig. 9 that the bias that would be expected from optical depths derived from pixel-mean radiances to illustrate that these biases become significant at larger optical depths. Such biases should be considered a potential source of uncertainty in comparing ISCCP statistics with model results until a means of adjusting the model optical depths to approximate the bias in the ISCCP simulator can be developed.

One could imagine a methodology to simulate the ISCCP optical depths given some assumed variance of optical depth within model grid boxes. Indeed, a preliminary attempt to adjust ground-based optical depths by accounting for subsatellite pixel variability essentially eliminated the bias between ARM and ISCCP optical depths (Fig. 3c) and significantly increased the agreement between ISCCP and ICARUS for the stratocumulus and deep clouds (note shown). Developing such a methodology will be the focus of the next phase of this work.

4. Summary and conclusions

The ISCCP simulator has gained wide use across the community, although it has not been well validated. The ISCCP simulator is designed to convert cloud property and thermodynamics profiles simulated by models into cloud-top pressure (P) and visible optical depth (τ) that would be diagnosed by ISCCP. This conversion from model output to satellite-like quantities enables global comparison of cloud properties that span several decades. Such comparisons of recent climate are critical to understanding and improving cloud feedbacks in GCMs (Williams and Tselioudis 2007; Williams and Webb 2009).

We find that the ICARUS portion of the ISCCP simulator does indeed facilitate comparisons between observed and simulated cloud-top pressures by adjusting some portion of the simulated high-topped and low-topped clouds into the middle troposphere (Fig. 2; Table 5). However, in comparing $\overline{P} - \overline{\tau}$ statistics from a carefully screened set of cases designed to minimize differences in sampling between satellite and ground-based measurements, the following discrepancies were found:

1) The ground-based observations converted to ISCCP-like quantities show significantly fewer (23%) mid-level clouds than found by ISCCP.
2) The ground-based observations converted to ISCCP-like quantities show significantly more (25%) optically thick cloud than reported by ISCCP.
3) The ground-based observations converted to ISCCP-like quantities show significantly fewer (27%) optically intermediate clouds than diagnosed by ISCCP.

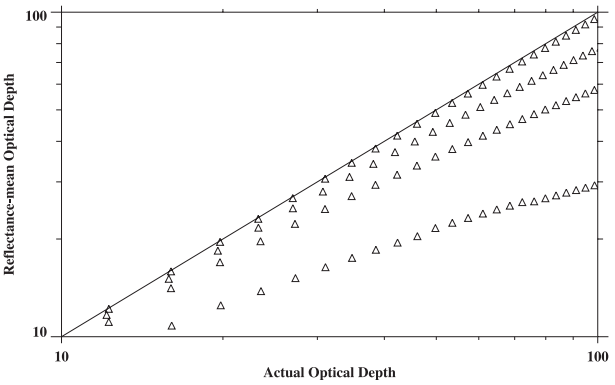

FIG. 8. The relationship between the actual spatially averaged optical depth (abscissa) and the optical depth derived from a spatially averaged mean reflectance, assuming that the optical depth is gamma distributed with the mean value that is the true spatial mean and differing values of optical depth standard deviation. The 1:1 line is shown as a solid line, and curves extending increasingly to the right of the 1:1 line represent normalized standard deviations of 0.1, 0.25, 0.5, and 1.0.

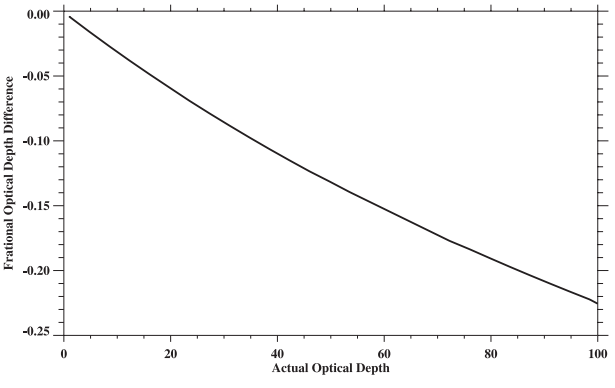

FIG. 9. The fractional bias in optical depth retrieved from pixel-mean radiances with an assumed 30% subpixel optical depth variability.

4) The discrepancies seem to be concentrated in the optically thick low-cloud category, where nearly a factor of 2.5 more clouds are found in the observations converted to ISCCP-like quantities than in ISCCP and in the optically intermediate middle-cloud categories.

We note that these discrepancies are nearly identical to several of the main findings reported by Z05 in comparing GCM statistics with ISCCP—albeit of lesser magnitude. For instance, Fig. 6b of Z05 shows the discrepancy in midlevel clouds (point 1 above) while Figs. 8a and 8b of Z05 show the discrepancies in optically thick and optically intermediate clouds (points 2 and 3 above). Figure 8e and 8i of Z05 are similar to our point 4 above.

Z05 interpreted these discrepancies (and others) to be due to deficiencies in the models. However, here we find several very similar discrepancies with ground-based measurements when passed through the same satellite simulator algorithm, suggesting that there may be unaccounted for issues in the comparison of ISCCP cloud statistics and model output that use the ISCCP simulator. This calls into question at least the severity of several of the main conclusions in Z05 and other studies that evaluate the fidelity of models by comparing them to ISCCP via the ISCCP simulator. For example, Williams and Tselioudis (2007) and Williams and Webb (2009) define cloud regimes based on $\overline{P} - \overline{\tau}$ statistics. They show that when models simulate the stratocumulus regime they tend to create clouds that are too optically thick when compared to ISCCP (points 2 and 4 above).

A more careful evaluation of the discrepancies (Fig. 5) show that if a model were to predict the actual occurrence of clouds with the same accuracy as a cloud radar and then the model made reasonable diagnostic interpretations of the column radiative properties, then agreement with satellite-derived $\overline{P} - \overline{\tau}$ after applying ICARUS would be successful in only approximately one-half to two-third of cases depending on the cloud type. Here, success is defined by ICARUS placing the simulated cloudy column into the same $\overline{P} - \overline{\tau}$ bin as ISCCP of the nine bins typically used for such comparisons.

The convolution of uncertainties in simulating $\overline{P}$ and $\overline{\tau}$ when model output is passed through the ISCCP simulator contrive to cause uncertainty and potential bias when comparing $\overline{P} - \overline{\tau}$ statistics from models to similar statistics derived from ISCCP data. While some uncertainty exists in the parameterization of $\overline{P}$ in ICARUS, the principal problem appears to be due to unaccounted for bias in the ISCCP $\overline{\tau}$ that may be because of subpixel variability in the cloud field that is not accounted for in the simulation of ISCCP in ICARUS. Based on these and earlier findings, we recommend that a systematic study of potential errors in visible optical depth be undertaken for ISCCP, LBTM, and the ground-based techniques so that corrections can be made as appropriate and/or the ISCCP simulator can be modified to account for any potential biases in $\overline{\tau}$ that do exist. Finally, we conclude that comparisons of optical depth made between ISCCP and similar algorithms with GCM results whether or not they have been modified to simulate ISCCP with the ISCCP simulator should be viewed with caution until these discrepancies are understood and accounted for, if necessary, in the ISCCP simulator. Thus GCM-ISCCP comparisons should be viewed with caution until these discrepancies are understood and accounted for, if necessary, in the ISCCP simulator. Where available, it would be prudent to use cloud retrievals from ground-based sensors in addition to those from satellites in the evaluation of model-simulated cloud properties.

Acknowledgments. The authors wish to gratefully acknowledge several helpful discussions with A. Marshak, A. Davis, and S. Kato. Primary funding for this work was supplied by the Environmental Science Division of the U.S. Department of Energy (Grant DE-FG0398ER62571). S. A. Klein is supported by the Office of Biological and Environmental Research in the Environmental Sciences Division of the U.S. Department of Energy as part of the Atmospheric Radiation Measurement Program. The contribution of S. A. Klein to this work is performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Q. Min is supported by ARM Grant DE-FG02-03ER63531. Data were obtained from the Atmospheric Radiation Measurements Program sponsored by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research, Environmental Science Division. An allocation of computer time from the Center for High Performance Computing at the University of Utah is acknowledged.

REFERENCES

Ackerman, T. P., and G. M. Stokes, 2003: The Atmospheric Radiation Measurement Program. *Phys. Today,* **56,** 38–45.

Barker, H. W., T. J. Curtis, E. Leontieva, and K. Stamnes, 1998: Optical depth of overcast cloud across Canada: Estimates based on surface pyranometer and satellite measurements. *J. Climate,* **11,** 2908–2994.

Berk, A., L. S. Bernstein, and D. C. Robertson, 1989: MODTRAN: A moderate resolution model for LOWTRAN 7. Spectral Sciences, Inc. Final Rep., Geophysics Laboratory Rep. GL-TR-89-0122, 44 pp.

Cahalan, R. F., W. Ridgway, W. J. Wiscombe, T. L. Bell, and J. B. Snider, 1994: The albedo of fractal stratocumulus clouds. *J. Atmos. Sci.,* **51,** 2434–2455.

Davis, A., A. Marshak, R. Cahalan, and W. Wiscombe, 1997: The Landsat scale break in stratocumulus as a three-dimensional

radiative transfer effect: Implications for cloud remote sensing. *J. Atmos. Sci.,* **54,** 241–260.

Dufresne, J.-L., and S. Bony, 2008: An assessment of the primary sources of spread of global warming estimates from coupled ocean–atmosphere models. *J. Climate,* **21,** 5135–5144.

Kato, S., and A. Marshak, 2009: Solar zenith and viewing geometry-dependent errors in satellite retrieved cloud optical thickness: Marine stratocumulus. *J. Geophys. Res.,* **114,** D01202, doi:10.1029/2008JD010579.

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

Mace, G. G., and S. Benson, 2008: The vertical structure of cloud occurrence and radiative forcing at the SGP ARM site as revealed by 8 years of continuous data. *J. Climate,* **21,** 2591–2610.

——, and Coauthors, 2006: Cloud radiative forcing at the Atmospheric Radiation Measurement Program Climate Research Facility: 1. Technique, validation, and comparison to satellite-derived diagnostic quantities. *J. Geophys. Res.,* **111,** D11S90, doi:10.1029/2005JD005921.

Min, Q., and L. C. Harrison, 1996: Cloud properties derived from surface MFRSR measurements and comparison with GOES results at the ARM SGP site. *Geophys. Res. Lett.,* **23,** 1641–1644.

Minnis, P., W. L. Smith Jr., D. P. Garber, J. K. Ayers, and D. R. Doelling, 1995: Cloud properties derived from GOES-7 for the spring 1994 ARM intensive observing period using version 1.0.0 of the ARM Satellite Data Analysis Program. NASA Reference Publ. 1366, 59 pp.

Pincus, R., M. Szczodrak, J. Gu, and P. Austin, 1995: Uncertainty in cloud optical depth estimates from satellite radiance measurements. *J. Climate,* **8,** 1453–1462.

Randall, D., M. Khairoutdinov, A. Arakawa, and W. Grabowski, 2003: Breaking the cloud parameterization deadlock. *Bull. Amer. Meteor. Soc.,* **11,** 1547–1564.

Rossow, W. B., A. W. Walker, D. E. Beuschel, and M. D. Roiter, 1996: International Satellite Cloud Climatology Project (ISCCP) documentation of new cloud datasets. WMO Rep. WMO/TD 737, 115 pp.

Schiffer, R. A., and W. B. Rossow, 1983: The International Satellite Cloud Climatology Project (ISCCP): The first project of the World Climate Research Programme. *Bull. Amer. Meteor. Soc.,* **64,** 779–784.

Soden, B. J., and I. M. Held, 2006: An assessment of climate feedbacks in coupled ocean–atmosphere models. *J. Climate,* **19,** 3354–3360.

Solomon, S., D. Qin, M. Manning, M. Marquis, K. Averyt, M. M. B. Tignor, H. L. Miller Jr., and Z. Chen, Eds., 2007: *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, 996 pp.

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. *Climate Dyn.,* **17,** 905–922.

Williams, K. D., and G. Tselioudis, 2007: GCM intercomparison of global cloud regimes: Present-day evaluation and climate change response. *J. Climate Dyn.,* **29,** 231–250.

——, and M. J. Webb, 2009: A quantitative performance assessment of cloud regimes in climate models. *Climate Dyn.,* **33,** 141–157, doi:10.1007/s00382-008-0443-1.

Zhang, M. H., and Coauthors, 2005: Comparing clouds and their seasonal variations in 10 atmospheric general circulation models with satellite measurements. *J. Geophys. Res.,* **110,** D15S02, doi:10.1029/2004jd005021.