A comparison of TWP-ICE observational data with cloud-resolving model results

A. M. Fridlind,[1] A. S. Ackerman,[1] J.-P. Chaboureau,[2] J. Fan,[3] W. W. Grabowski,[4] A. A. Hill,[5] T. R. Jones,[6] M. M. Khaiyer,[7] G. Liu,[8] P. Minnis,[9] H. Morrison,[4] L. Nguyen,[9] S. Park,[10,11] J. C. Petch,[5] J.-P. Pinty,[2] C. Schumacher,[12] B. J. Shipway,[5] A. C. Varble,[13] X. Wu,[8] S. Xie,[14] and M. Zhang[15]

Received 21 July 2011; revised 7 December 2011; accepted 18 December 2011; published 13 March 2012.

[1] Observations made during the TWP-ICE campaign are used to drive and evaluate thirteen cloud-resolving model simulations with periodic lateral boundary conditions. The simulations employ 2D and 3D dynamics, one- and two-moment microphysics, several variations on large-scale forcing, and the use of observationally derived aerosol properties to prognose droplet numbers. When domain means are averaged over a 6-day active monsoon period, all simulations reproduce observed surface precipitation rate but not its structural distribution. Simulated fractional areas covered by convective and stratiform rain are uncorrelated with one another, and are both variably overpredicted by up to a factor of ~2. Stratiform area fractions are strongly anticorrelated with outgoing longwave radiation (OLR) but are negligibly correlated with ice water path (IWP), indicating that ice spatial distribution controls OLR more than mean IWP. Overpredictions of OLR tend to be accompanied by underpredictions of reflected shortwave radiation (RSR). When there are two simulations differing only in microphysics scheme or large-scale forcing, the one with smaller stratiform area tends to exhibit greater OLR and lesser RSR by similar amounts. After ~10 days, simulations reach a suppressed monsoon period with a wide range of mean precipitable water vapor, attributable in part to varying overprediction of cloud-modulated radiative flux divergence compared with observationally derived values. Differences across the simulation ensemble arise from multiple sources, including dynamics, microphysics, and radiation treatments. Close agreement of spatial and temporal averages with observations may not be expected, but the wide spreads of predicted stratiform fraction and anticorrelated OLR indicate a need for more rigorous observation-based evaluation of the underlying micro- and macrophysical properties of convective and stratiform structures.

Citation: Fridlind, A. M., et al. (2012), A comparison of TWP-ICE observational data with cloud-resolving model results, *J. Geophys. Res.*, *117*, D05204, doi:10.1029/2011JD016595.

1. Introduction

[2] The Tropical Warm Pool–International Cloud Experiment (TWP-ICE) took place over and around Darwin, Australia, from 20 January through 13 February 2006. According to *May et al.* [2008], TWP-ICE is "the first field program in the tropics that attempted to describe the evolution of tropical convection, including the large-scale heat,

[1]NASA Goddard Institute for Space Studies, New York, New York, USA.
[2]Laboratoire d'Aerologie, University of Toulouse/CNRS, Toulouse, France.
[3]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, Washington, USA.
[4]Mesoscale and Microscale Meteorology, National Center for Atmospheric Research, Boulder, Colorado, USA.
[5]Met Office, Exeter, UK.
[6]Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado, USA.
[7]Science Systems, and Applications, Inc., Hampton, Virginia, USA.
[8]Department of Earth, Ocean and Atmospheric Science, Florida State University, Tallahassee, Florida, USA.
[9]NASA Langley Research Center, Hampton, Virginia, USA.
[10]Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa, USA.
[11]Now at Department of Earth and Atmospheric Sciences, University of Nebraska, Lincoln, Nebraska, USA.
[12]Department of Atmospheric Sciences, Texas A&M University, College Station, Texas, USA.
[13]Department of Atmospheric Sciences, University of Utah, Salt Lake City, Utah, USA.
[14]Lawrence Livermore National Laboratory, Livermore, California, USA.
[15]Institute for Planetary and Terrestrial Atmospheres, State University of New York at Stony Brook, Stony Brook, New York, USA.

Copyright 2012 by the American Geophysical Union.
0148-0227/12/2011JD016595

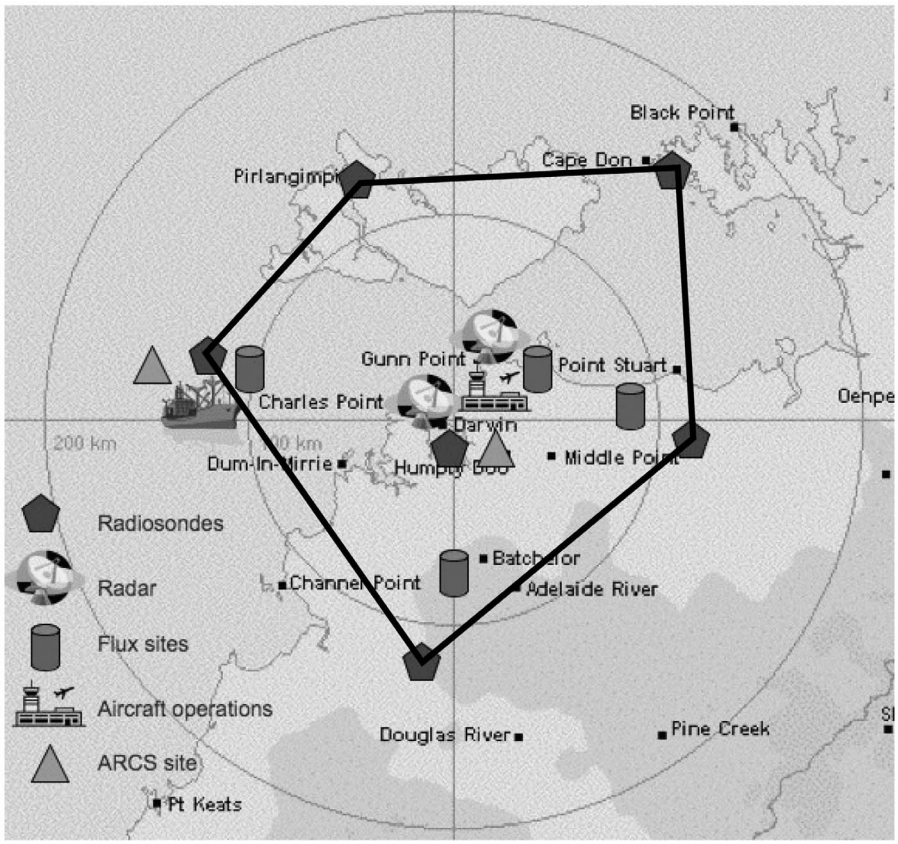

Figure 1. The TWP-ICE observational domain, with sounding array locations enclosing a pentagonal area of roughly 31,000 km^2. Latitude and longitude of each radiosonde site listed in Table 1.

moisture, and momentum budgets at 3-hourly time resolution, while at the same time obtaining detailed observations of cloud properties and the impact of the clouds on the environment." The experiment specifically focused on the properties of convectively generated cirrus, aiming to document their relationship to environmental conditions. The experimental domain (Figure 1) was centered on a highly instrumented site operated by the US Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) program and a C-band polarimetric (C-POL) weather radar operated by the Australian Bureau of Meteorology, surrounded by a 3-hourly sounding array (Table 1) and surface energy budget sites. TWP-ICE was also coordinated with the Aerosol and Chemical Transport in tropIcal conVEction (ACTIVE) program, funded by the UK Natural Environment Research Council, which gathered extensive in situ measurements of environmental aerosol properties [*Vaughan et al.*, 2008]. The data gathered during TWP-ICE and ACTIVE are now archived by ARM and the British Atmospheric Data Centre, respectively.

[3] Improving the climate projection skill of general circulation models (GCMs), a principal motivation for TWP-ICE, has been hindered by inadequate representation of cloud properties and their relationship to environmental conditions [e.g., *Randall et al.*, 2007]. Since cloud properties vary on short temporal and spatial scales that are not well resolved in GCMs, approaches to improve GCM cloud representation have commonly included direct or indirect use of cloud-resolving models (CRMs) [e.g., *Randall et al.*, 2003a]. With the explicit goal of using CRMs to guide the improvement of climate models, under the auspices of the World Meteorological Organization, a primary activity of the Global Energy and Water-Cycle Experiment (GEWEX) Cloud Systems Study (GCSS) has been the organization of intercomparison studies in which modeling groups worldwide are invited to participate [*Randall et al.*, 2003b]. The first GCSS modeling study of deep convective processes, based on the Tropical Oceans Global Atmosphere Coupled Ocean–Atmosphere Research Experiment (TOGA-COARE), included both CRMs and single-column models (SCMs) [*Moncrieff et al.*, 1997; *Wu et al.*, 1998, 1999; *Wu and Moncrieff*, 2001a; *Redelsperger et al.*, 2000; *Bechtold et al.*, 2000]. SCM and CRM simulations of midlatitude continental convection were next compared based on observations at the ARM Southern Great Plains (SGP) site [*Ghan et al.*, 2000; *Xu et al.*, 2002; *Xie et al.*, 2002, 2005]. Later GCSS convection studies focused on the transitions

Table 1. Sounding Array Sites Defining the TWP-ICE Pentagonal Domain[a]

Site Name	Latitude	Longitude
Mount Bundy	−13.2287	131.1355
Ship	−12.4	129.8
Garden Point	−11.4089	130.4167
Cape Don	−11.3081	131.7651
Point Stuart	−12.5858	131.7609

[a]See Figure 1.

from shallow to deep convection over tropical land [*Grabowski et al.*, 2006] and from suppressed to deep convection over tropical ocean, and included analysis of global atmospheric models, in addition to CRMs and SCMs [*Petch et al.*, 2007; *Willett et al.*, 2008; *Woolnough et al.*, 2010]. A number of related studies have used a similar approach to investigate the sensitivity of a single CRM or SCM over a much wider parameter space than can generally be accommodated in a multimodel study [e.g., *Grabowski et al.*, 1996, 1998; *Wu et al.*, 1998; *Grabowski et al.*, 1999; *Wu et al.*, 1999; *Wu and Moncrieff*, 2001a, 2001b].

[4] Here we present the results of a CRM study based on data gathered during the TWP-ICE and ACTIVE programs. The specification for CRM initialization and forcing [*Fridlind et al.*, 2010] was developed jointly through the DOE ARM, GCSS, and Stratospheric Processes And their Role in Climate (SPARC) programs. SPARC participation was motivated by the goal of understanding the influence of tropical deep convection on water vapor concentrations and convective transport through the tropical tropopause, and led to the derivation and adoption of a large-scale forcing data set at 10-mb vertical resolution in order to improve representation of near-tropopause thermodynamic conditions. A unique aspect of this case is the availability of an idealized aerosol number size distribution profile, composed of three lognormal modes with fixed geometric mean radius and standard deviation and number concentrations that vary with altitude (see section 2), derived from measurements as described by *Fridlind et al.* [2010]. To our knowledge this is the first CRM comparison study to provide a vertically varying profile of aerosol size distribution properties; the inclusion of modal information extends upon the simpler specification of total number concentration profile provided by *Barth et al.* [2007]. Whereas the work presented here focuses only on 2D and 3D CRM simulations with fully periodic boundary conditions (an approach used most commonly in combination with SCM simulations), three complementary studies based on TWP-ICE data have been simultaneously conducted using SCMs (L. Davies, manuscript in preparation, 2012), limited-area models (LAMs) with open boundary conditions and nested grids (P. Zhu et al., A limited area model (LAM) intercomparison study of a TWP-ICE active monsoon mesoscale convective event, manuscript submitted to *Journal of Geophysical Research*, 2012), and GCMs operated in short-term forecast mode (Y. Lin et al., TWP-ICE global atmospheric model intercomparison: convection responsiveness and resolution impact, manuscript submitted to *Journal of Geophysical Research*, 2012). A summary and comparison of all four studies will focus on common and contrasting results as well as methodological issues (J. Petch, manuscript in preparation, 2012).

[5] TWP-ICE data have already been widely used in other modeling studies. Among those focused primarily on dynamics and precipitation, several analyzed CRM dynamical behaviors under TWP-ICE conditions to inform GCM parameterization development [*Wu et al.*, 2009; *Del Genio and Wu*, 2010; *Wang and Liu*, 2009]. Others directly evaluated GCM parameterizations with respect to closure assumptions [*Zhang*, 2009], the effect of two-moment microphysics on simulated versus observed stratiform precipitation [*Song and Zhang*, 2011], and simulated versus observed relative humidity among other factors [*Franklin et al.*, 2012]. *Wapler et al.* [2010] concluded that judiciously formulated LAM simulations could reasonably reproduce observed precipitation rate statistics. Among studies focused more on ice properties, *Wang et al.* [2009b] found substantial discrepancies between simulated and observed ice cloud properties in all the CRM simulations they considered. *Wang et al.* [2009a] found that SCM radiative fluxes are sensitive to the representation of ice properties that are not directly constrained by ground and satellite measurements. Other studies reported on the sensitivity of simulated aerosol, microphysical, dynamical, and radiative processes to changes in aerosol specification and ice nucleation assumptions [*Fan et al.*, 2010a, 2010b; *Morrison and Grabowski*, 2011; *Zeng et al.*, 2011]. In a companion study using 3D simulations from this study, *Varble et al.* [2011] have also examined the characteristics of precipitating cloud structures in greater detail.

[6] Here we first briefly describe the specification for CRM initialization and forcing (section 2), the CRMs used (section 3), and the observational data sets used to evaluate the simulations (section 4). Most aspects of the specification are based on methodologies developed for earlier GCSS cases, and like all prior GCSS studies cited above, this work compares observed and simulated thermodynamic variables, hydrometeor paths, and precipitation rates. Each prior study also addressed specific focus areas, such as the treatment of boundary conditions and large-scale forcing terms [e.g., *Ghan et al.*, 2000] or the effect of adding basic model features such as the ice phase or a third spatial dimension [e.g., *Redelsperger et al.*, 2000]. In this paper we focus on the following questions: (1) do simulations and observations agree within experimental uncertainties and (2) how robust is the methodology used here for producing realistic simulations? In the course of addressing these questions in section 5, simulations are also compared with one another; here we intentionally limit our focus to quantities that are observationally constrained, but note that companion studies include additional model variables [e.g., *Varble et al.*, 2011] and compare them with other model results (e.g., Zhu et al., submitted manuscript, 2012).

2. Case Description

[7] Over the month-long TWP-ICE campaign, Darwin experienced active monsoon conditions only during the first week, culminating with the passage of a large mesoscale convective system (MCS) directly through the center of the observational domain on 23–24 January, followed by suppressed monsoon conditions through 3 February, and monsoon break conditions thereafter [*May et al.*, 2008]. This study focuses only on the active and suppressed periods. Although the TWP-ICE experimental domain contains both land and ocean, the low-lying land areas become saturated during monsoon periods, behaving in a manner that is maritime in nature. To allow CRM representation of relatively slowly developing and advecting monsoon features (such as cold pools) over the TWP-ICE region in a framework that remains as simple as possible, the following idealized marine conditions are specified (additional details provided by *Fridlind et al.* [2010]): (1) model domain footprint representative of the TWP-ICE observation domain (circa 176 × 176 km), (2) sea surface temperature fixed at 29°C

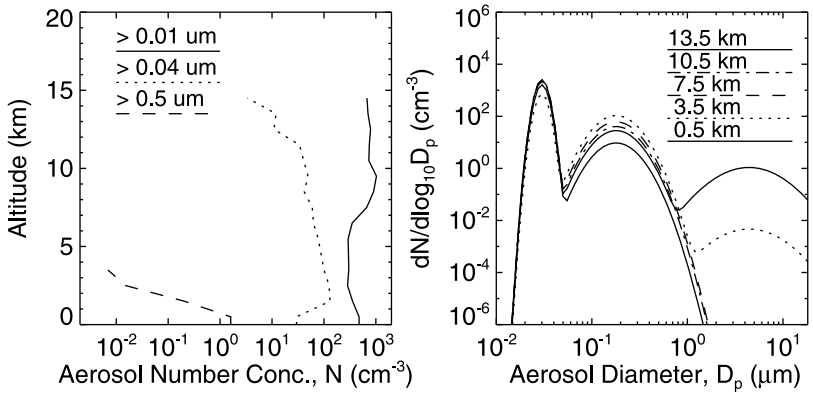

Figure 2. (left) Mean profiles of aerosol number concentration in three size cuts and (right) derived tri-modal size distributions as a function of elevation based on ACTIVE in situ measurements as described by *Fridlind et al.* [2010] (see section 2).

(diagnosed surface fluxes), (3) fully periodic horizontal boundary conditions, (4) surface albedo fixed at 0.07 in all shortwave bands, (5) diurnally varying insolation with domain centered on the Darwin ARM site (12.425°S, 130.891°E), (6) run time of 16 days (0Z 18 January to 0Z 3 February 2006), (7) nudging of mean horizontal winds above 500 m to observed profiles with a two-hour timescale, (8) large-scale advective forcing of potential temperature and water vapor (vertical and horizontal, derived from observations) and condensate (vertical only, calculated using large-scale wind derived from observations) [*Xie et al.*, 2010], applied at full strength below 15 km, linearly decreasing above to zero strength at 16 km, and (9) nudging of mean water vapor and potential temperature to observed profiles with a six-hour timescale, adopted at full strength above 16 km and linearly decreasing below to zero strength at 15 km (baseline) or adopted at full strength above 1 km and linearly decreasing below to zero strength at 0.5 km (optional sensitivity test).

[8] Nudging of water vapor and potential temperature in the upper troposphere was found to be necessary to keep simulated environmental conditions realistic aloft, consistent with an understanding that large-scale forcings are poorly constrained by measurements above about 15 km [cf. *Petch et al.*, 2007; *Morrison and Grabowski*, 2011]. An optional sensitivity test with nudging extended down to the lower troposphere was included because drift of simulated conditions from observations at lower elevations was found to influence the strength and depth of convection and the area covered by stratiform precipitation. Horizontally uniform application of all nudging terms preserves variations from the mean.

[9] Input files are archived as described in Appendix A. These include an idealized profile of aerosol size distribution properties that was derived from observations during the active period (Figure 2) as described by *Fridlind et al.* [2010].

[10] Allowing 36 hours of model spin-up, analysis here and in companion studies is focused on several time periods after 12Z on 19 January (day of year range in parentheses): (1) 6 days of active monsoon conditions (19.5–25.5), (2) 6 days of suppressed monsoon conditions (27.5–33.5), and (3) three shorter periods of intense precipitation during the active monsoon (19.5–20.625, 22.125–23.125, and 23.125–24.5), referred to hereafter as events A, B, and C.

[11] We emphasize that this case specification assumes an all-ocean surface over the coastal TWP-ICE observational domain, an assumption that is being examined in companion studies that include land surfaces. Comparison of CRM results with LAM simulations spanning event C have shown little impact of land surface on initiation and maintenance of convection during the active period (Zhu et al., submitted manuscript, 2012), but impacts could well be greater during the suppressed period.

3. Simulations

[12] Simulations include ten combinations of dynamics and microphysics. Six dynamics models were used: the Distributed Hydrodynamic-Aerosol-Radiation Model Application (DHARMA) [*Stevens et al.*, 2002; *Ackerman et al.*, 2000], the Eulerian semi-Lagrangian model (EULAG) [*Smolarkiewicz and Margolin*, 1997], the Iowa State University 2D Cloud Resolving Model (ISUCRM) [*Wu et al.*, 2008], the Meso-NH Atmospheric Simulation System (MESONH) [*Lafore et al.*, 1998], the System for Atmospheric Modeling (SAM) [*Khairoutdinov and Randall*, 2003], and the UK Met Office Large Eddy Model (UKMO) [*Shutts and Gray*, 1994; *Petch and Gray*, 2001]. All model dynamics are based on anelastic equations, but with varying treatments of subgrid-scale turbulence, surface fluxes, radiative transfer, advection, and time stepping. General model features and optional setup parameters are summarized in Table 2. Three groups submitted the optional sensitivity test (with nudging added throughout the free troposphere to offset accumulation of errors; see section 2); the sensitivity test simulations are identified with an "s" (DHARMA-1s, EULAG-2s, and SAM-2Ms).

[13] Simulated convective cloud properties are expected to be sensitive to the microphysics scheme [e.g., *Wang et al.*, 2009b; *Fan et al.*, 2010b], which can be classified in terms of the number of prognostic variables used for condensed water. The one-moment schemes (DHARMA-1 and MESONH-1) prognose only 4–5 hydrometeor mixing ratios, whereas the two-moment schemes additionally prognose 1–5 number concentrations (see Table 2). EULAG-2 uses a

Table 2. Model Parameters

Model	Dimension	Domain[a] (km)	$\triangle x$ (m)	$\triangle z$[b] (m)	Microphysical Reference[c]	Prognostic Microphysical Variables[d]	Ice Nucleation Mechanisms[e]	Sensitivity Test[f]
DHARMA-1	3	176	900	100–250	G99	$q_c\ q_r\ q_s\ q_g$	I H	yes
DHARMA-2M	3	176	900	100–250	M09	$q_c\ q_r\ q_i\ q_s\ q_g\ N_c\ N_r\ N_i\ N_s\ N_g$	D C I H M S	
EULAG-2	2	200	1000	100–300	M08	$q_c\ q_r\ q_i\ N_c\ N_r\ N_i\ RMF_i$	D C I H M S	yes
ISUCRM-2	2	600	3000	100–1000	K76	$q_c\ q_r\ q_s\ q_g\ N_s\ N_g$	D	
MESONH-1	3	192	1000	100–250	P98	$q_c\ q_r\ q_i\ q_s\ q_g$	D H	
MESONH-2	3	192	1000	100–250	P02	$q_c\ q_r\ q_i\ q_s\ q_g\ N_c\ N_r\ N_i$	D C H A M	
SAM-2M	3	192	1000	100–400	M09	$q_c\ q_r\ q_i\ q_s\ q_g\ N_c\ N_r\ N_i\ N_s\ N_g$	D C I H M S	yes
UKMO-2A	3	176	900	225–500	B06	$q_c\ q_r\ q_i\ q_s\ q_g\ N_i$	D C H M	
UKMO-2B	3	176	900	225–500	B06	$q_c\ q_r\ q_i\ q_s\ q_g\ N_i\ N_s\ N_g$	D C I H M S	
UKMO-2M	3	176	900	225–500	M09	$q_c\ q_r\ q_i\ q_s\ q_g\ N_r\ N_i\ N_s\ N_g$	D C I H M S	

[a]Domain footprint is square for 3D models.
[b]Range of model layer depths between surface and typical tropopause elevation of 17 km.
[c]Microphysics references: G99, *Grabowski* [1999]; M09, *Morrison et al.* [2009]; M08, *Morrison and Grabowski* [2008b] (see also section 3); K76, *Koenig and Murray* [1976]; P98, *Pinty and Jabouille* [1998]; P02, *Pinty* [2002]; and B06, *Brown and Heymsfield* [2006].
[d]Prognostic microphysical variables: respective mixing ratios and number concentrations of cloud water (q_c, N_c), rain (q_r, N_r), ice (q_i, N_i), snow (q_s, N_s), and graupel (q_g, N_g; treated as hail in SAM-2M). In EULAG-2, cloud ice includes all ice types and is characterized by rimed mass fraction (RMF_i).
[e]Ice crystal formation mechanisms (see section 3): D, deposition and condensation nucleation (may be used for diagnostic N_c); C, contact nucleation; I, immersion nucleation; H, homogeneous freezing of cloud or raindrops; A, aerosol freezing (homogeneous); M, Hallett-Mossop ice multiplication; S, snow breakup.
[f]Optional sensitivity test submitted (see section 2, item 9).

single size distribution for all ice that is further characterized by a prognostic rimed mass fraction [*Morrison and Grabowski*, 2007, 2008a, 2008b]. Here each simulation moniker includes either a "1" to indicate one-moment (no number concentrations prognosed) or a "2" to indicate two-moment (at least one number concentration prognosed). Simulations identified with "2M" use versions of the two-moment *Morrison et al.* [2009] scheme. Other analyses of these simulations may use differing naming conventions [e.g., *Varble et al.*, 2011].

[14] Of the four schemes that prognose cloud droplet number concentration (N_c), DHARMA-2M and SAM-2M used the vertically varying trimodal aerosol profile provided. In DHARMA-2M, the aerosol in each mode were advected, consumed by hydrometeor collision-coalescence, and nudged on a domain-mean basis to their initial profiles with a six-hour timescale, whereas in SAM-2M the number concentrations were fixed. In EULAG-2, the three modes were populated with vertically uniform number concentrations of 295, 95, and 0.4 cm^{-3}, respectively. In MESONH-2, an activation spectrum was fitted using the diameter and standard deviation of the middle specified mode and the number concentration of 130 cm^{-3} specified at 1500 m in altitude. N_c was fixed at 240 cm^{-3} in UKMO-2A and UKMO-2B and at 100 cm^{-3} in the remaining schemes.

[15] While it is beyond the scope of this paper to provide an exhaustive comparison of the microphysical processes, the mechanisms of primary and secondary ice nucleation are listed in Table 2 owing to their expected importance in simulation results [e.g., *Fan et al.*, 2010b]. Most simulations include a single diagnostic equation for the number concentration of heterogeneous ice nuclei that form ice crystals directly from the vapor phase in the deposition or condensation modes when the air is supersaturated with respect to ice or water, expressed as an exponential function of either supercooling only (ISUCRM-2 uses equation (13) from *Koenig and Murray* [1976] with A_{06} = 464 and A_{07} = 12 in SI units; SAM-2M and UKMO-2M follow the *Thompson et al.* [2004] implementation of *Cooper et al.* [1986] except with an upper limit of 500 L^{-1}) or supersaturation only (DHARMA-2M, E ULAG-2, MESONH-2, UKMO-2A, and UKMO-2B use equation (2.4) from *Meyers et al.* [1992] with $a = -0.639$ and $b = 0.1296$; EULAG-2 sets a limit of 100 L^{-1} on the resulting number concentration). Most schemes independently diagnose ice nuclei active in the contact mode as an exponential function of supercooling only following *Meyers et al.* [1992] (DHARMA-2M, EULAG-2, MESONH-2, SAM-2M and UKMO-2M use their equation (2.6) with $a = -2.80$ and $b = 0.262$) or *Young* [1974] (UKMO-2A and UKMO-2B use his equation (12) with N_{a0} = 2000 m^{-3}). Most models that include immersion freezing assume a stochastic treatment that is an exponential function of supercooling only, following *Bigg* [1953]; the only exception is DHARMA-1, which diagnoses a number concentration of heterogeneous immersion nuclei as an exponential function of temperature only [*Grabowski*, 1999, equation (A.20)], following *Fletcher* [1962]. Most models include near-instantaneous homogeneous freezing of activated cloud droplets and/or raindrops at temperatures colder than roughly −40°C [e.g., *Pinty and Jabouille*, 1998]. Freezing of unactivated aerosol at colder temperatures and higher supersaturations is included only in MESONH-2, following *Kärcher and Lohmann* [2002]. Secondary nucleation processes are Hallett-Mossop rime splintering and snow breakup (one or both are included in most simulations; see Table 2). SAM-2M and UKMO-2M also set an upper limit of 10 cm^{-3} on cloud ice number concentration.

[16] Submitted model results are archived for public use as described in Appendix A. Archived results for 3D models include Rayleigh-scattering radar reflectivities that have been independently calculated using uniform assumptions [*Varble et al.*, 2011], which are used here. Numerous non-standard diagnostics were requested for comparison with specific observational data streams [see *Fridlind et al.*, 2010], which resulted in most participants running computationally intensive simulations more than once to increase compliance with the full specification. Since a principal objective of this study is to compare a variety of simulations with the observations (rather than with one another), three unique simulations that do not hew precisely to the full

specification have been included here. In ISUCRM-2, which is valuable as one of only two 2D models, nudging of tropospheric water vapor and potential temperature aloft is neglected, which has little impact on overall convective fluxes and precipitation by specification design (aside we note that the meridional wind direction is also reversed in this simulation, to which results are invariant in this modeling framework). In MESONH-1 and MESONH-2, which include unique graupel schemes [cf. *Varble et al.*, 2011], meridional winds are not consistent with the specification. All diagnostics were made optional owing to the long list requested [see *Fridlind et al.*, 2010]; if a diagnostic is not available for a given simulation, then that simulation is omitted from evaluation against measurements in the following without further comment.

4. Observations

[17] In this paper, emphasis is placed on domain-wide observational data sets rather than individual point and profile measurements. To bridge the spatial scale mismatch between CRMs and data derived from scanning radar, satellite imaging, or global analysis models, CRMs reported some diagnostics at a coarsened horizontal resolution, ranging from 2.5 km (scanning radar-scale) to 55 km (global analysis-scale). By contrast, it is not possible to directly manipulate CRM fields to bridge the mismatch between the model grids and very high spatial resolution point and column measurements. Taking precipitation rate as an example, averaging times of 5–15 min have been found to be optimal in point comparisons of time-integrated rain gauge measurements with instantaneous precipitation radar measurements at 2-km horizontal resolution [*Habib and Krajewski*, 2002]. Even if it were in principle possible to find an optimal averaging time for intercomparison of each point or column data source with each CRM horizontal grid spacing (0.9–3 km from Table 2), statistical results would still be challenging to robustly use for constraining models, as evidenced by the difficulties encountered when comparing precipitation radars (roughly comparable to CRM grid cell size) and rain gauges [e.g., *Nikolopoulos et al.*, 2008]. Past work has furthermore indicated that average collocated radar-gauge precipitation measurements should not be expected to agree to better than about 10% until 20 or so convective events are sampled [*Habib and Krajewski*, 2002, and references therein], far more than sampled here. Additional point and column measurements will be considered in future work.

[18] Original data were downloaded from the ARM online archive unless otherwise indicated. Processed values have been archived for public use in a CF-compliant format (see Appendix A).

4.1. C-Band Polarimetric (C-POL) Radar

[19] Data obtained from the 5.5–cm-wavelength scanning C-POL radar at Darwin [*Keenan et al.*, 1998] are gridded reflectivities at 0.5-km vertical resolution and retrieved precipitation rate at an elevation of 2.5 km and 1-km vertical resolution. All radar data are reported at 2.5-km horizontal resolution and 10-min frequency throughout the TWP-ICE domain (bounded by sites listed in Table 1). Recalibrated radar data were provided by Peter May. Uncertainty in retrieved precipitation rate is estimated to be 25% at rain rates above 10 mm h^{-1} and 100% at the lowest reported rain rates. Uncertainty in rain rate averaged domain-wide over the suppressed period is 25% but a bit higher during the active period (33%); we use 25% as a representative value for both periods since the difference during the active period does not impact conclusions. Uncertainties of the occurrence frequencies over a range of rates (e.g., 2–20 mm h^{-1}) were found by recalculating the frequencies in each range with uncertainties added or subtracted; the uncertainty range is then obtained from the minimum and maximum frequencies found in each rain rate category.

[20] We identify the fractional area covered by convective and stratiform rain over the TWP-ICE domain using C-POL reflectivity as described in Appendix B. Uncertainty in the fractional areas is estimated by applying the same uncertainty algorithm to reflectivity fields with the grid cell uncertainty of approximately 1 dBZ added or subtracted. Resulting relative uncertainty in the convective and stratiform area fractions is within 20% and 5%, respectively, during both active and suppressed periods.

[21] Latent heating rate profiles retrieved over the TWP-ICE domain were provided by Courtney Schumacher [*Schumacher et al.*, 2004] based on a separate processing of C-POL raw radar data, including gridding at 2-km rather than 2.5-km resolution. For comparison with simulations here, latent heating rate profiles are normalized by the ratio of surface precipitation rate in the large-scale forcing data set to vertically integrated latent heating rate, which is computed using time-dependent thermodynamic profiles also obtained from the large-scale forcing data set. Latent heating rate profiles are compared with model results on a qualitative basis.

4.2. Visible Infrared Shortwave-Infrared Split-Window Technique (VISST) Retrievals

[22] Broadband top-of-atmosphere (TOA) outgoing longwave radiation (OLR), shortwave albedo, and ice water path (IWP) were derived from radiances measured by the imager on the geostationary satellite MTSAT-1R. The OLR and shortwave albedo were derived from the 10.8-μm and 0.73-μm radiances, respectively, following the approach of *Minnis and Smith* [1998] with modifications similar to those described by *Khaiyer et al.* [2010]. The relevant MTSAT-1R channels were calibrated against the corresponding spectral channels on the Terra MODerate-resolution Imaging Spectroradiometer. The cloud properties were derived using the methods of *Minnis et al.* [2008, 2011a] to detect cloudy pixels and retrieve cloud properties such as phase, effective particle size, and optical depth. IWP was computed from the product of the last two parameters for ice-cloud pixels, and therefore may include contributions from liquid underlying an ice layer. Analysis of the TWP-ICE data set is summarized by *Minnis et al.* [2006]. All of these MTSAT-derived data streams are referred to hereafter as VISST for brevity. Values are reported at a 15–60-min frequency and 4-km resolution over 5–17°S and 125–136°E. In each swath, pixels in the TWP-ICE domain are identified and relevant statistics calculated. Domain-mean relative uncertainties are estimated as +9/−4% in OLR and +7/−15% in reflected shortwave radiation (RSR) and shortwave albedo, based

on comparisons with Terra (CERES Edition 3) after *Khaiyer et al.* [2010].

[23] VISST IWP data are limited to daytime because the retrieval requires visible reflectance to estimate optical depth. Since the maximum retrievable optical depth is 128 for this data set, IWP will be underestimated when optical depth is higher, consistent with a comparison of annual mean VISST and CloudSat retrievals finding close agreement except in regions of tropical convection [*Waliser et al.*, 2009]. There, the mean VISST IWP values were ~25% less than their CloudSat counterparts, suggesting a negative bias of roughly one-third. For thin cirrus, the instantaneous VISST IWP retrievals are typically within 40% of surface-based radar-radiometer retrievals [*Minnis et al.*, 2011b]. VISST IWP retrievals are considered here qualitatively, and the unknown contribution of liquid hydrometeors to retrieved IWP is neglected when comparing with simulations. Daytime is defined conservatively as any time when instantaneous TOA downwelling flux exceeds 200 W m^{-2}. Since simulated TOA downwelling solar fluxes are not identical, the UKMO-2A 10-min TOA shortwave downwelling flux time series, with mean diurnal values that precisely match those in the 3-h large-scale forcing data set, is used as a benchmark to define daytime temporally for all comparisons.

4.3. Total Sky Imager (TSI) Retrievals

[24] The TSI provides time series of hemispheric sky images during daylight hours and retrievals of fractional opaque and thin cloud cover at 30-s frequency when the solar elevation is greater than 10°. Uncertainty in opaque cloud retrievals depends upon cloud aspect ratio [*Kassianov et al.*, 2005] and is not used here. Owing also to the difficulty of reconciling model-based and differing VISST and TSI instrument-based definitions of clear and cloudy conditions, we therefore consider all cloud cover results qualitatively rather than quantitatively.

4.4. Surface Sensible and Latent Heat Fluxes

[25] An eddy covariance system mounted on a short tower over Darwin Harbor provided surface sensible and latent heat flux measurements at 30-min resolution. Gap-filled data are used, wherein gaps shorter than two hours are filled using interpolation and longer gaps are filled using a neural network algorithm [*Beringer et al.*, 2007]. Here we consider the harbor measurements qualitatively since they were made at a single location. We note that the domain-wide fluxes in the large-scale forcing data set are an area-weighted average that includes eddy covariance measurements at land sites and profile-based calculations at sea (see section 4.7).

4.5. Microwave Radiometer (MWR) Retrievals

[26] Liquid water path is retrieved from MWR measurements at Darwin and on the ship [*Turner et al.*, 2007]. Data are reported at 20 to 35-s resolution when the measurements are not contaminated by surface precipitation. Indeterminate and missing fields are removed, small negative values set to zero, and the mean of retrieved values at both stations taken over 10-min intervals. The degree to which an average of two stations is representative of the domain mean is unknown; the representativeness of point measurements likely depends upon the variable being measured and the meteorological conditions encountered [e.g., *Barnett et al.*, 1998; *Habib and Krajewski*, 2002]. LWP is therefore considered qualitatively in this study.

4.6. European Center for Medium-Range Weather Forecasts (ECMWF) Global Analyses

[27] ECMWF supplied results of the Operational Analysis and Forecasting System in three grid cells representative of the TWP-ICE domain. Surface precipitation rate considered here is an average over 1-h time periods (ending at reporting time) and roughly 55-km resolution. Statistics used here are mean and maximum of surface precipitation rate across the three grid cells provided. They are compared with the domain-wide mean and maximum of instantaneous values obtained at 10-min frequency from C-POL measurements and simulations.

4.7. Large-Scale Forcing Data Set

[28] The variational analysis used to derive the domain-mean large-scale forcing data set at 10-mb vertical resolution and 3-h temporal resolution (centered in time) is based on inputs that include surface heat and radiative fluxes and C-POL, VISST, MWR, and ECMWF products listed above. As described by *Xie et al.* [2010], environmental profiles in the large-scale forcing data set are an integration of available soundings with analysis products. Data set components used here are domain-mean surface precipitation rate, environmental profiles, surface latent and sensible heat fluxes, and TOA and surface radiative fluxes.

4.8. The 3D Ice Water Content (3D-IWC) Retrieval

[29] The 3D-IWC retrieval employs a Bayesian algorithm to retrieve IWC and IWP from MWR, cloud radar, and sounding measurements at Darwin, and high-frequency microwave data collected on NOAA satellites from the Advanced Microwave Sounding Unit - B (AMSU-B) [*Seo and Liu*, 2005, 2006]. The retrieval algorithm includes a range of ice types and the input observations are variably sensitive to all ice types (see discussion in section 5.2). Retrievals are reported at the temporal resolution of available satellite overpasses at ~16–25 km spatial resolution within 10° latitude and longitude of Darwin. An uncertainty is provided for each reported value as described by *Seo and Liu* [2006]. For each available swath that spans the TWP-ICE domain, measurements and uncertainty are averaged over pixels identified within the domain. Averaging uncertainty of point measurements in this manner is equivalent to assuming that all errors are perfectly correlated. Statistics calculated are thus domain-mean IWC profiles and IWP. Mean uncertainty in IWP, which is strongly dependent upon amount of ice present [*Seo and Liu*, 2006], is just under 20% during the active period, roughly 600% during the more dormant suppressed period, and just under 40% when averaging over the full reported simulation period (19.5–34). Ice water content profiles are considered qualitatively.

[30] In order to account for the temporal sparseness of retrievals dependent upon polar orbiting satellites, simulations are sampled at the frequency of available 3D-IWC retrievals whenever quantitative comparisons are made between retrievals and simulations. This introduces a degree of inconsistency when plotting observed and simulated IWP versus, for example, observed and simulated OLR, in terms

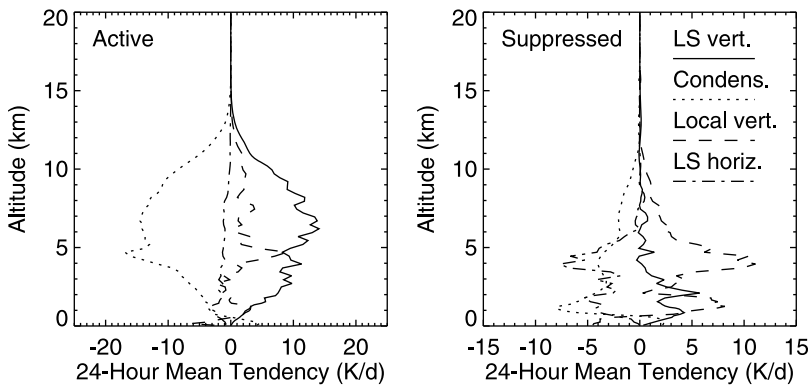

Figure 3. Leading terms in the water vapor budget profiles (expressed in terms of latent heat) over (left) the active monsoon period and (right) the suppressed monsoon period from the DHARMA-2M baseline simulation example: net flux convergence from large-scale (LS) vertical advection, net condensation (including deposition and sublimation), local vertical mixing (including resolved and subgrid-scale), and LS horizontal advection.

of temporal sampling. However, we find that the temporal sampling is important to the quantitative comparison of observations and simulations of IWP (and therefore it should be done to properly consider whether simulations are within the uncertainty of retrievals) but it does not qualitatively impact the arrangement of ensemble members in correlation plots considered here (thus retrievals are qualitatively representative despite relatively low sampling frequency).

5. Results

5.1. Precipitation Features

[31] The surface precipitation rate over the TWP-ICE domain obtained from C-POL measurements (see section 4.1) is an input to the variational analysis (see section 4.7) and a principal indicator of the strong large-scale ascent that is dominant during the active period (e.g., Figure 3). It is therefore not surprising that models reproduce the temporal evolution of domain-mean surface precipitation under the strongly forced active conditions (Figure 4a), consistent with similar past studies [e.g., *Xu et al.*, 2002; *Xie et al.*, 2005; *Woolnough et al.*, 2010]. Figure 5 shows that every simulation reproduces mean surface precipitation rate within the uncertainty of retrievals during the active period, and in a manner that is closely correlated with total liquid water path (LWP, defined throughout this study as cloud plus rainwater). Aside, we note that Figure 5 is the first of several figures in which one domain-mean quantity is plotted against another. These plots illustrate the degree to which two quantities are related across the ensemble of simulations. If observations are available, they also concisely illustrate whether simulated quantities fall within the range of observational uncertainty and whether deviations from observations are correlated. In all such figures, results for the active and suppressed period are plotted separately owing to commonly differing patterns. During the weakly forced suppressed period in Figure 5, for instance, all simulations overestimate surface precipitation except SAM-2Ms, unlike during the active period. However, even when the domain mean of surface precipitation is robustly reproduced under strongly forced conditions, we find the following evidence that the underlying structural features of simulated precipitation fields differ substantially across models, associated with large differences in radiative flux terms.

[32] Because models can report precipitation rate at the 2.5-km altitude and 2.5-km resolution of the C-POL retrievals, it is possible to closely compare precipitation rate statistics. Before doing so we note that domain-mean precipitation rates are 10–20% greater at 2.5-km than at the surface in all reporting models except EULAG (comparing Figures 4a and 4b). Although mean C-POL reflectivity is actually lower at 2.5 km than at 0.5 km during this period (not shown), consistent with past measurements in tropical maritime convection [*Houze et al.*, 2004], it is unknown whether the actual precipitation rate is lower or higher at the surface in this case. Any such actual difference can be viewed as a source of uncertainty in the large-scale forcing data set, as discussed further below. With respect to the frequencies of precipitation rate at 2.5-km resolution and 2.5-km altitude (Figure 6), the most apparent feature is that all reporting models overestimate occurrences in the light 0.2–2 mm h^{-1} range during both active and suppressed periods. Compared with DHARMA-1, this tendency is notably reduced in the DHARMA-1s sensitivity test simulation that is nudged toward domain-mean thermodynamic profiles throughout the free troposphere, suggesting that baseline simulations produce rain that is too widespread owing in part to deviations of simulated mean temperature and water vapor profiles from those observed. EULAG-2s shows a similar but weaker trend during the active period. In a more detailed analysis of precipitation statistics in 3D simulations from this study, *Varble et al.* [2011] found that total precipitating area matches observations to within 1–2% in DHARMA-1s and SAM-2Ms sensitivity test simulations during the active monsoon period, whereas baseline simulations overestimate precipitating area by 35–65%.

[33] Thus the baseline simulations systematically produce light rain that is far more widespread than observed, and this particular error is ameliorated to some degree with 6-h nudging of the domain-mean tropospheric thermodynamic profiles. It is possible that such overly widespread light precipitation could arise or persist from horizontally uniform domain-wide application of large-scale forcing terms, which are relatively strong in the TWP-ICE data set [cf. *Hagos*,

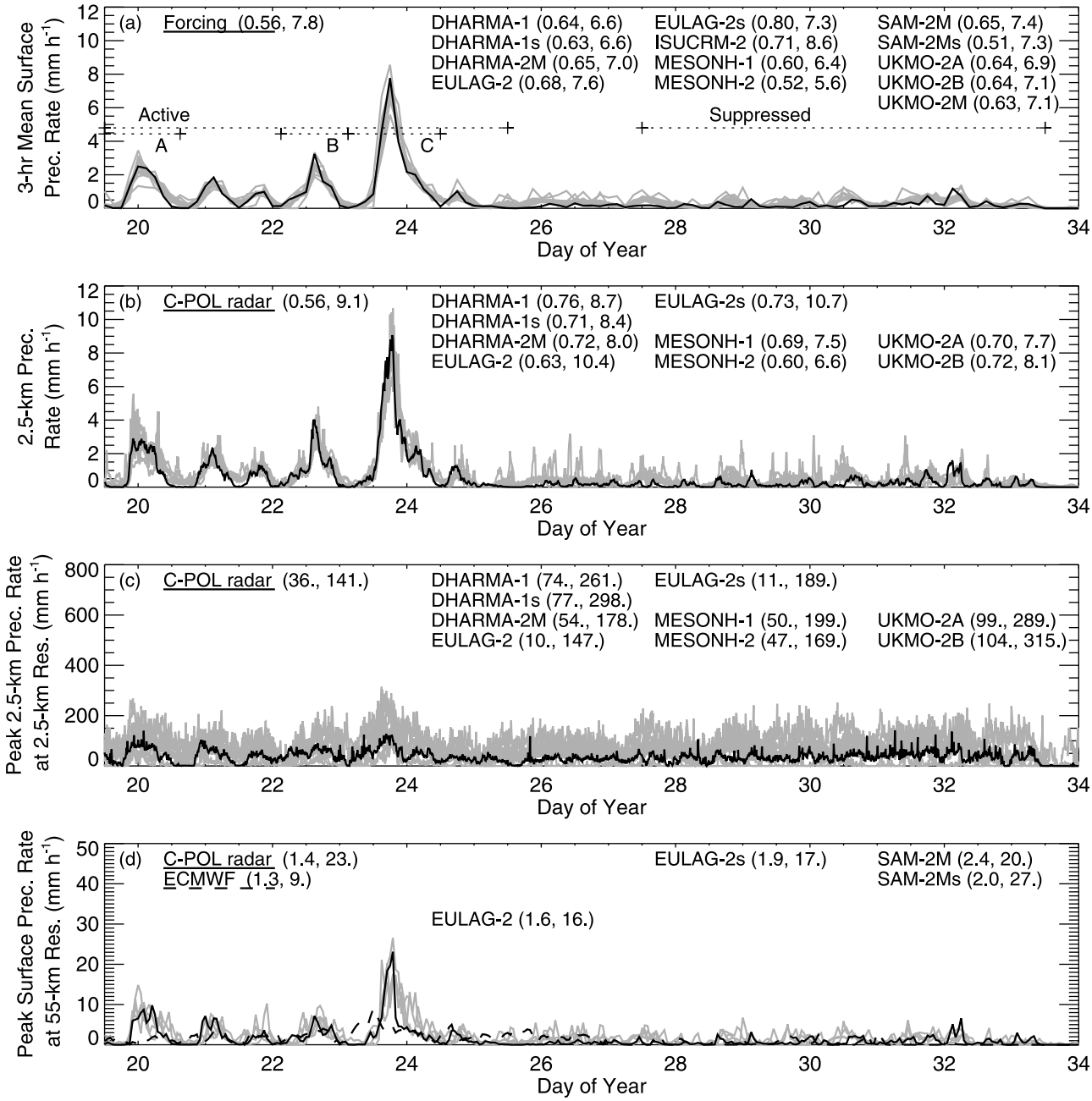

Figure 4. Precipitation rates: (a) simulated 3-h domain mean at the surface compared with large-scale forcing data derived from 2.5-km C-POL retrievals, (b) simulated 10-min domain mean at 2.5-km elevation compared with C-POL retrievals, (c) simulated 10-min domain maximum at 2.5-km elevation and 2.5-km horizontal resolution compared with C-POL retrievals, and (d) simulated 1-h domain maximum at the surface at 55-km horizontal resolution compared with ECMWF analyses and C-POL. Listed in parentheses are the mean and maximum of plotted values.

2010]. Surface rainfall frequencies were by contrast found to be in reasonable agreement with C-POL observations in a study of the TWP-ICE time period using a LAM with open boundary conditions [*Wapler et al.*, 2010]. Comparison of these CRM results with the relative frequencies of light and heavy rain in the associated LAM intercomparison study should shed light on the effects of differing boundary conditions and large-scale forcing approach. Those used in this study are more similar to a cloud-resolving convection parameterization approach in GCMs [e.g., *Grabowski*, 2001], whereas those used in the LAM study are more similar to a global cloud-resolving model.

[34] Compared with the excessive frequencies of lighter rain rates, the frequencies of rain rates >2 mm h^{-1} are generally reproduced better by the simulations (see Figure 6). However, the mean of simulated maximum precipitation rates (based on the 10-min sampling of models at identical 2.5-km horizontal resolution and elevation as retrievals)

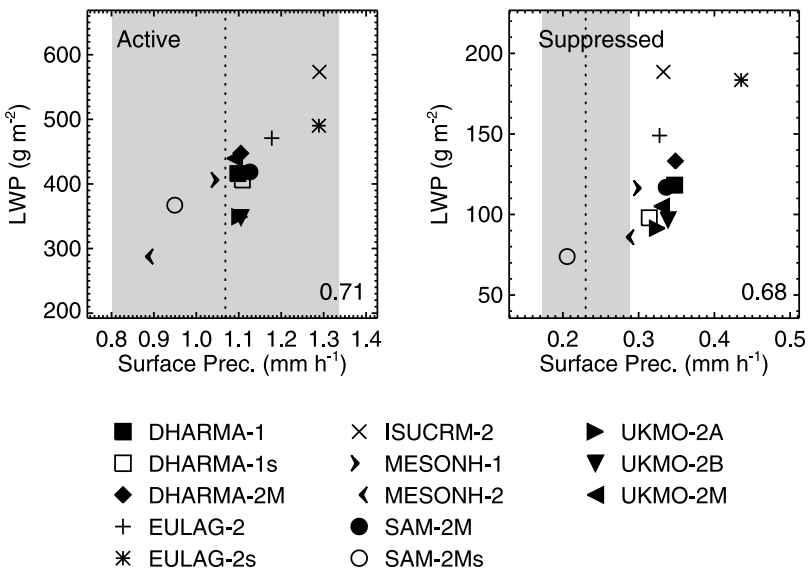

Figure 5. Simulated surface precipitation rate versus liquid water path (LWP, defined as cloud plus rainwater). Domain means averaged over (left) active period (19.5–25.5) and (right) suppressed period (27.5–33.5), respectively (see Figure 4). Degree of correlation is given as the Spearman rank coefficient. Precipitation rate from C-POL retrievals (dotted lines) shown with estimated uncertainty range (shading). Domain-wide LWP not observed (see section 5.2).

varies over a factor of ten range from roughly 10 mm h^{-1} in the 2D EULAG simulations to roughly 50–100 mm h^{-1} in reporting 3D simulations, as compared with roughly 40 mm h^{-1} retrieved from C-POL (Figure 4c). As the high-intensity tail of the frequency distribution, domain-wide maximum rain rate is appealing mostly because it is easily sampled in models and observations. That typical simulated peak rain rates increase with dimensionality is consistent with past findings that updraft strength and vertical mass fluxes increase with dimensionality [e.g., *Phillips and Donner*, 2006; *Petch et al.*, 2008; *Zeng et al.*, 2008], although the smaller sample size in 2D simulations is expected to reduce the likelihood of generating typical domain-wide maxima. Sample sizes are nearly identical in 3D simulations and observations, but peak rain rates in 3D simulations are systematically higher than those retrieved, which could be associated with overly broad and intense updrafts at the ~1-km horizontal resolution of most simulations [e.g., *Bryan and Morrison*, 2012]. We note that although pixel-level C-POL retrievals never exceed about 140 mm h^{-1}, maximum 10-min-mean tipping bucket gauge and disdrometer measurements do exceed 140 mm h^{-1} during events A, B and C, and at other times (not shown). Since convective core regions are the most important sources of rainfall and are locations where precipitation efficiency influences the production of longer-lived convective outflow aloft, systematic discrepancies between retrievals and simulations warrant further study.

[35] The structure of precipitation fields in the 3D simulations can also be compared with C-POL measurements by applying a textural algorithm to objectively identify the domain fraction covered by convective and deep stratiform rain in simulated and observed radar reflectivity fields (see Appendix B and examples in Figure 7). Results indicate that convective area fraction is commonly overpredicted by a factor of two or more (Figure 8a), indicating that regions with strong updrafts are systematically too large or too frequent or both. Two baseline simulations lie within the 15% relative uncertainty of observed convective area during the active period, and only one sensitivity test does during the suppressed period (Figure 9). The maximum convective area fraction observed, ~30% during event C, is larger than typical maxima of ~20% observed over larger domains at Darwin and elsewhere [e.g., *Frederick and Schumacher*, 2008; *Holder et al.*, 2008]. Although differences in area identification algorithms and source data resolution do significantly impact area calculations [e.g., *Steiner et al.*, 1995; *Yuter et al.*, 2005], an exaggerated maximum in this case could be attributable at least partly to the MCS of event C covering an area substantially larger than the TWP-ICE domain. The maximum convective area fraction is never as strongly overestimated as the time average, possibly reflecting physical limits on convective fraction under given environmental conditions [cf. *Holder et al.*, 2008].

[36] Whereas the mean convective area fractions are consistently overpredicted, the stratiform area fractions range from being underpredicted to overpredicted with relatively wider ranges (Figure 8b), indicating that stratiform outflows and evolution are more sensitive to model differences. The narrow range of uncertainty in stratiform area (Figure 9) is likely substantially smaller than the uncertainty in the large-scale forcings driving the simulations (e.g., ~25% uncertainty in surface precipitation rate). Compared to their respective baseline simulations, stratiform area is substantially reduced in the SAM-2Ms sensitivity test but not in DHARMA-1s, suggesting a microphysics-dependent sensitivity to tropospheric moisture and temperature. Aside, we note that stratiform area fraction can be underpredicted and light precipitation rate frequencies simultaneously overpredicted when the light rain is originating from shallow

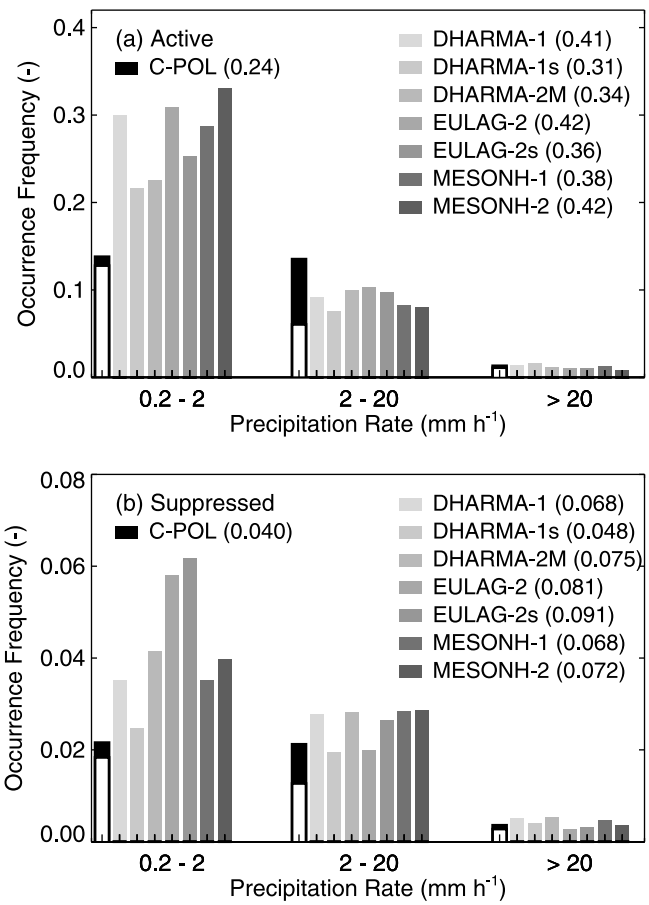

Figure 6. Precipitation rate statistics simulated and observed at 2.5-km resolution and 2.5-km elevation. Shaded section of first bar in each range indicates minimum and maximum frequency considering uncertainties in C-POL retrievals (see section 4.1). In parentheses is summed mean occurrence frequency of all rates > 0.2 mm h^{-1}.

clouds rather than deep stratiform clouds (see Appendix B). Within a given dynamics model (e.g., DHARMA, UKMO), two-moment microphysics schemes based on *Morrison et al.* [2009] produce similar or greater stratiform areas than one-moment schemes, consistent with past results [*Morrison et al.*, 2009; *Luo et al.*, 2010; *Bryan and Morrison*, 2012]. But baseline simulations with versions of the same two-moment scheme also differ substantially (e.g., stratiform fraction is notably larger in DHARMA-2M than in UKMO-2M), suggesting that dynamics also plays a role. Prognosing droplet number concentration evidently does not produce a strongly distinguishing effect in Figure 8 (N_c prognosed using differing approaches in DHARMA-2M, MESONH-2, and SAM-2M), but could be associated with larger stratiform fractions in DHARMA-2M and SAM-2M than in UKMO-2M.

[37] Across the ensemble during the active period, it is notable that stratiform area fractions ranging over ~25–60% are poorly correlated with either convective area fractions or with ice water path (IWP, defined throughout this study as the sum of all ice-phase hydrometeors, including cloud ice, snow, and graupel; see Table 2). However, the stratiform area fraction is strongly correlated with a 60 W m^{-2} range of predicted outgoing longwave radiation (OLR, Figure 9), indicating that the spatial distribution of ice controls simulated OLR more than domain-mean IWP. It is also notable that convective area predictions tend to be better during the onset of convective events than during the decay (e.g., during event C over 23–24 January in Figure 8a), which may be at least in part attributable to the fact that mature cells can pass out of the observational domain whereas periodic boundary conditions require their decay to be completed within the modeling domain. During the suppressed period, stratiform area fractions are less than 15% in all simulations (variably underestimated and overestimated); they are more poorly correlated with OLR than during the active period owing in part to a wider range of high cloud fraction (see section 5.2).

[38] The observed mean ratio of 0.86 for stratiform to stratiform plus convective area found in this study is higher than the active, suppressed, and experiment wide values of 0.75–0.79 during TWP-ICE over the full C-POL domain reported by *Frederick and Schumacher* [2008, Table 2] and at the upper limit of the 0.66–0.86 range reported for various tropical regions [*Holder et al.*, 2008]. The observed ratio could be higher owing to (1) poor Eulerian sampling of a small number of Lagrangian events passing through the geographically limited TWP-ICE observational domain and (2) differences in observational data characteristics such as horizontal grid resolution. Aside, we note that the additional requirement on stratiform area in our algorithm of a minimum reflectivity above the melting level would tend to reduce rather than increase the ratio, all else being equal (see Appendix B). The simulations exhibit stratiform to stratiform plus convective ratios of 0.73 (DHARMA-1s and MESONH-1) to 0.89 (DHARMA-2M), roughly spanning the observational range over tropical regions, and thus none appear to be strong outliers by this simple metric.

[39] Although it is beyond the scope of this paper to examine the details of convection organization, simulations with more linear convection features appear to produce larger stratiform fractions (compare Figures 7 and 8). More specifically, simulations with little organization exhibit the least stratiform area (DHARMA-1 and MESONH simulations), whereas those with the most linear squall lines tend to exhibit the greatest stratiform area (e.g., DHARMA-2M, SAM, and UKMO simulations; observed conditions appear more similar to this latter class). Such simulation tendencies are roughly consistent with observations of greater stratiform rainfall being associated with linear organization in tropical systems [*Rickenbach and Rutledge*, 1998], although in this case model physics is responsible rather than environmental conditions (as evidenced by the difference between DHARMA-1s and SAM-2Ms, despite both being nudged to observed conditions throughout the troposphere). Convective area does not appear to be closely associated with degree of linear organization (e.g., both MESONH and UKMO simulations span a wide range of convective area fractions in Figure 9). Thus it appears that the simulated convection organization mode could be more closely associated with stratiform rain generation than absolute convective area. The simulated degree of linear organization can in turn be substantially modified by microphysics (e.g., DHARMA-1 versus DHARMA-2M), consistent with past modeling results [e.g., *Lynn et al.*, 2005].

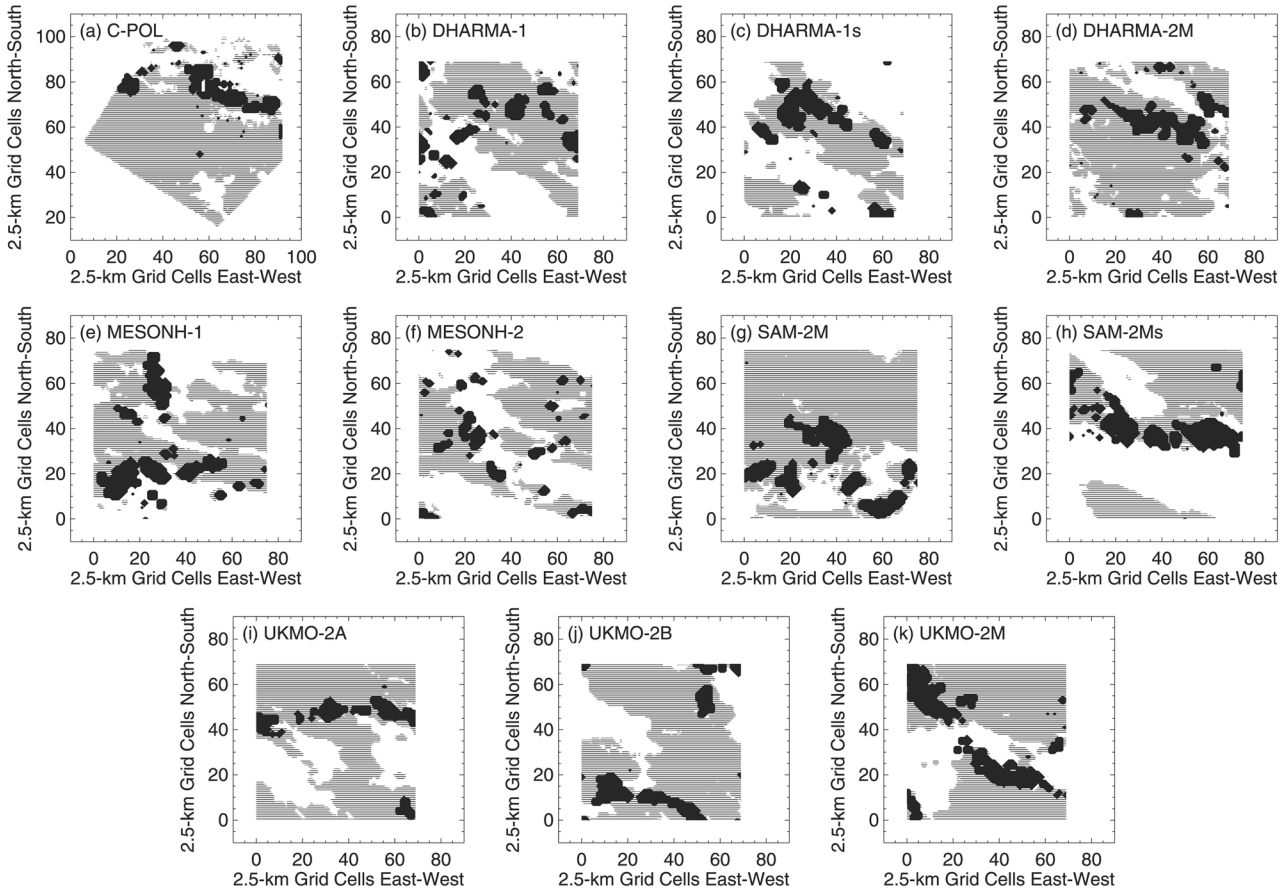

Figure 7. Convective (black) and stratiform (shaded) area fractions (see Appendix B) identified at day 20.125 from (a) C-POL measurements over the TWP-ICE domain and (b-k) simulations.

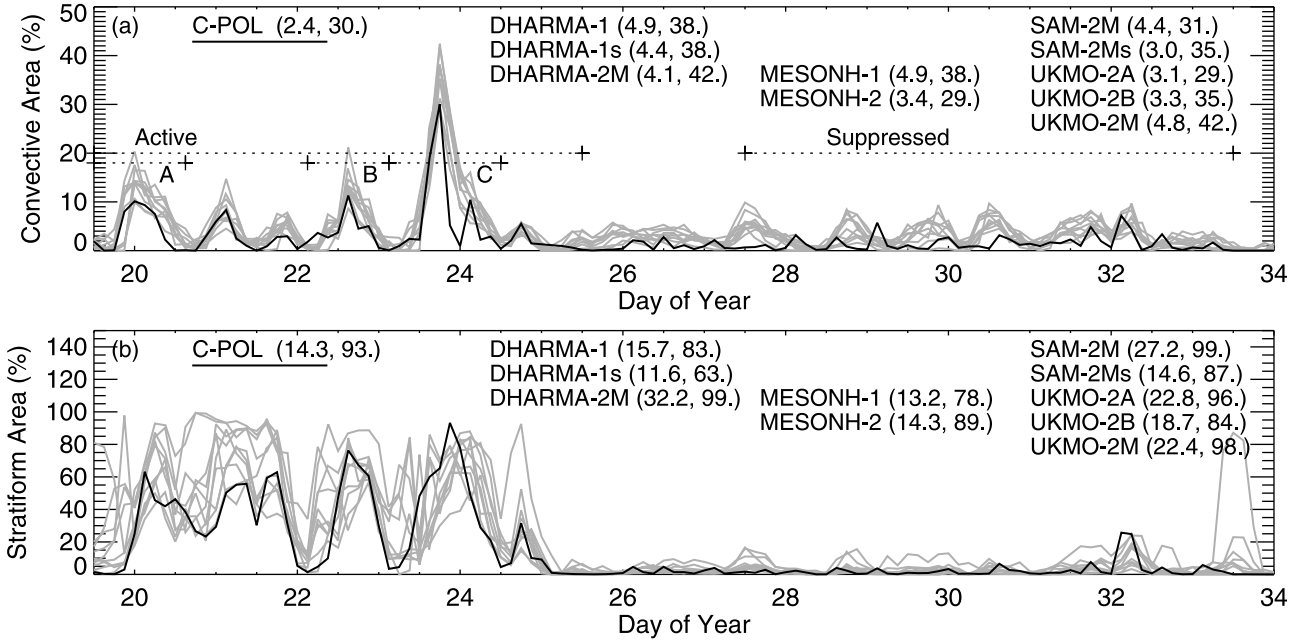

Figure 8. Simulated and observed (a) convective and (b) stratiform area fractions at 2.5-km elevation and 2.5-km resolution. Listed in parentheses are the mean and maximum of plotted values.

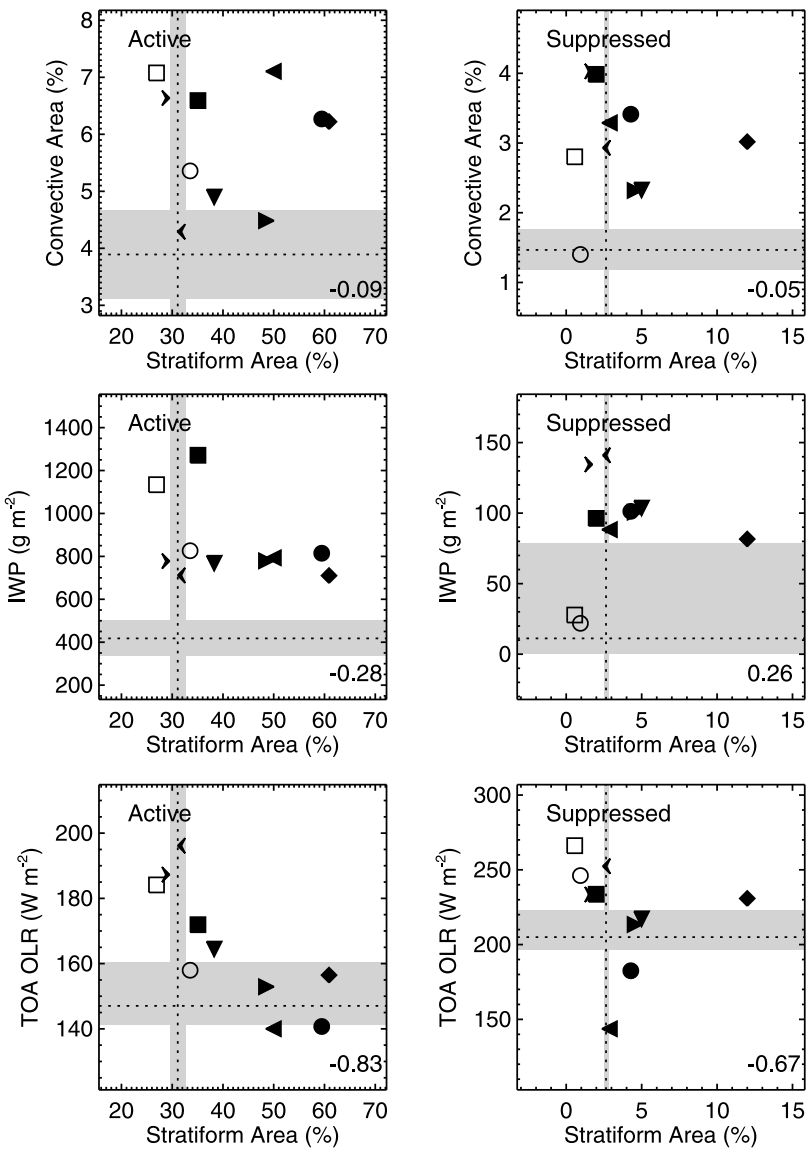

Figure 9. Stratiform area versus (top) convective area, (middle) ice water path (IWP), and (bottom) outgoing longwave radiation (OLR). Averaging times, symbols, and Spearman rank coefficients as in Figure 5. Domain means of convective and stratiform area fractions, OLR, and IWP from C-POL, VISST, and 3D-IWC retrievals (dotted lines) shown with estimated uncertainty ranges (shading).

[40] Finally, owing to the use of ECMWF analyses to drive the TWP-ICE LAM intercomparison, we briefly consider the relationship among ECMWF, observed, and simulated maximum precipitation rates at comparable horizontal scales. Taking 55-km resolution as roughly that of the ECMWF analyses, we find that the maximum of peak surface precipitation rates in the local ECMWF fields is about one-third of that retrieved from C-POL (Figure 4d). During each major event (A, B, and C), maximum intensity in ECMWF fields is lower than observed by an amount that far exceeds the C-POL observational uncertainty of 25%. Reduced maximum intensity in ECMWF fields is compensated by more frequent mid-range precipitation rates, as evidenced by less than 10% difference between the mean of peak intensities at 55-km resolution from ECMWF and C-POL fields. This pattern of overly frequent rainfall events with overall maximum intensity substantially lower than observed is consistent with extensive recent comparisons of ECMWF and other global models with CloudSat data [*Stephens et al.*, 2010]. The reporting CRM simulations, on the other hand, predict maximum values of 55-km-resolution precipitation rates that are greater than or within experimental uncertainty of that derived from C-POL. The reporting 3D simulations also sustain such rates more commonly than observed, as evidenced by mean peak 55-km intensities that are roughly 50% too high. Alongside variably high 2.5-km peak intensities and excessive convective area (Figures 4c and 8a), this provides additional evidence that convective precipitation structures in 3D simulations tend to be too intense or extensive.

5.2. Condensate, Latent Heating, and Cloud Cover

[41] We next consider observational constraints on the domain-wide column and profiles of condensate, latent

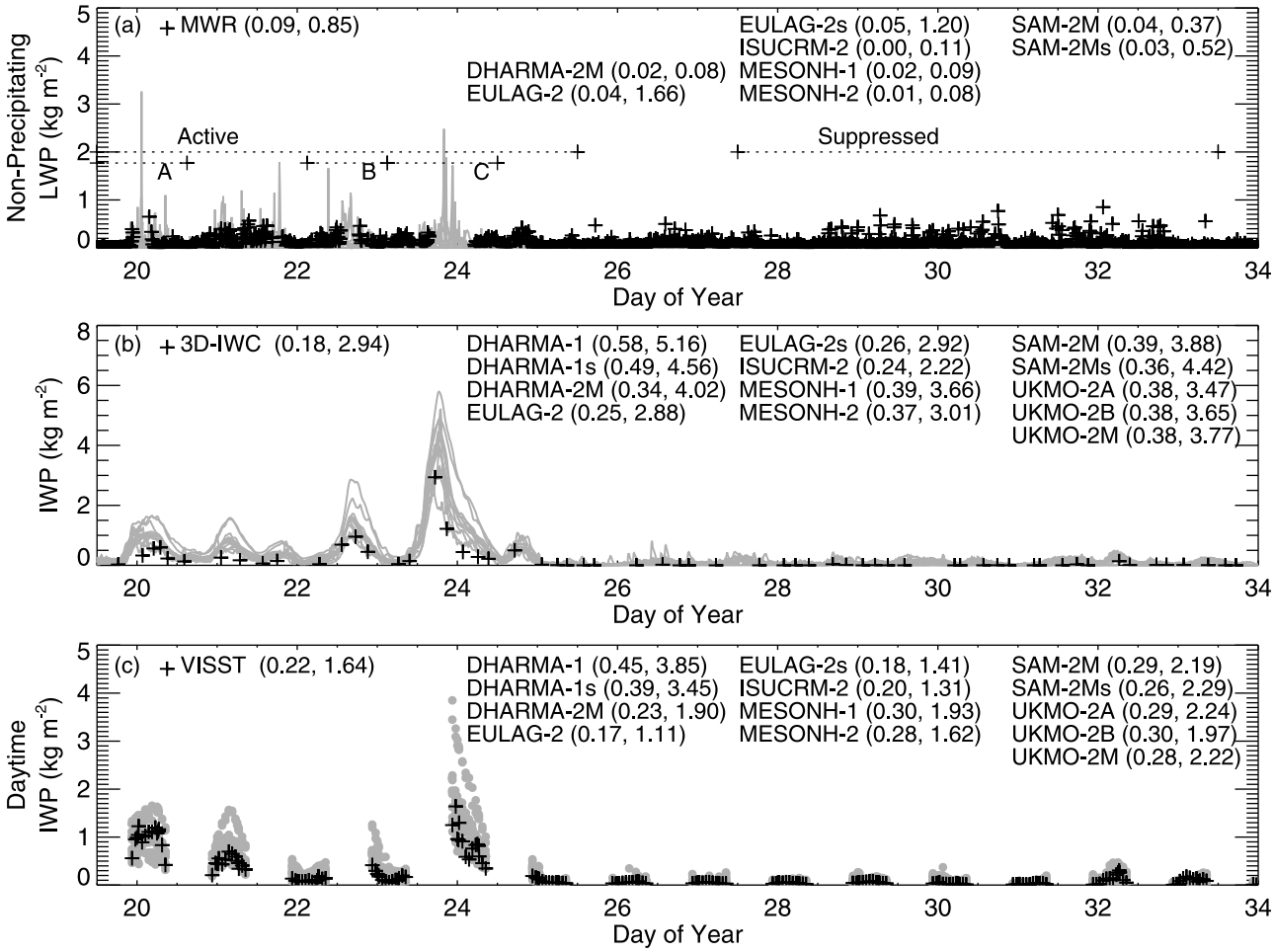

Figure 10. (a) Simulated non-precipitating LWP compared with MWR retrievals (see text), (b) IWP compared with 3D-IWC retrievals, and (c) daytime IWP compared with VISST retrievals. Listed in parentheses are the mean and maximum of plotted values. Simulation IWC statistics listed in Figure 10b are calculated after subsampling at the observational frequency (see section 4.8).

heating, and cloud cover. First, we find that observations are unfortunately too sparse to provide a robust constraint on LWP. MWR retrievals are available only at Darwin and on the ship, and are available only when surface precipitation is less than ~0.02 mm h^{-1} (D. Turner, personal communication, 2008). In simulations that reported non-precipitating LWP (domain mean of contributions from columns where surface precipitation <0.02 mm h^{-1}, Figure 10a), it constitutes ~5–25% of total LWP (see Figure 5). By comparison, cloud water alone accounts for ~25–50% of total LWP (not shown; no observational analog). The domain-mean non-precipitating LWP is therefore a relatively small fraction of total LWP and is more variable across simulations than domain-mean cloud or rainwater path or their sum, which could be attributable to differences in the simulated frequency of light precipitation. Satellite microwave-based retrievals of LWP are expected to be strongly influenced by assumptions regarding cloud and rainwater partitioning in this region of high average rainfall [*O'Dell et al.*, 2008], and are beyond the scope of this work to assess.

[42] Although LWP remains thus unconstrained, simulated IWP can be robustly compared with 3D-IWC retrievals, which are based on a synthesis of polar-orbiting satellite and ground-based measurements. The CRMs reproduce the temporal evolution of domain-mean IWP from both 3D-IWC and VISST retrievals quite well (Figures 10b and 10c). All simulated values below 0.2 kg m^{-2} lie within the uncertainty of VISST retrievals during daytime, and over the active period only DHARMA-1 and DHARMA-1s overestimate VISST daytime mean IWP by more than one-third, which is the estimated amount that it may be biased low owing to maximum retrievable optical depth (see section 4.2). However, simulated IWP is systematically greater than 3D-IWC retrievals, and only in the 2D baseline simulations (EULAG-2, ISUCRM-2) is domain-mean IWP over days 19.5–34 just within the associated uncertainty of 40%. Across all baseline simulations, the ratio of IWP to LWP is also notably higher in 3D than in 2D (Figure 11) because LWP is higher and IWP lower in all 2D versus all 3D baseline simulations. Aside, we note that over the (high-IWP) active period, when the 3D-IWC retrieval uncertainty is relatively lower, no simulations actually lie within the associated uncertainty range, whereas roughly half of simulations lie within the far larger uncertainty under the (low-IWP) conditions of the suppressed period. With respect to the vertical distribution of IWP, the CRMs locate most ice

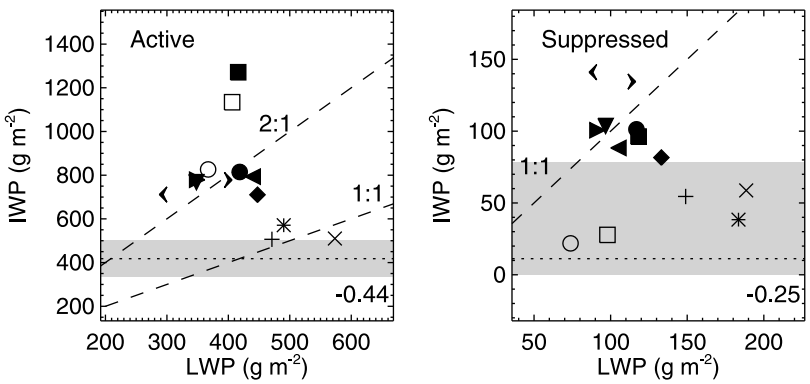

Figure 11. IWP versus LWP (including cloud water and rain). Averaging times, symbols, and Spearman rank coefficients as in Figure 5. Domain means of IWP derived from 3D-IWC retrievals (dotted lines) shown with estimated uncertainty ranges (shading). Simulation IWP is subsampled at the observational frequency (see section 4.8).

mass in the ~5–13-km altitude range during the active period, qualitatively consistent with 3D-IWC retrievals (Figure 12a). But the CRM results tend to exceed 3D-IWC retrievals by up to a factor of two at those elevations, commonly by more above 13 km. During the suppressed period, the CRMs predict up to an order of magnitude more IWC than retrieved in the 5–13-km altitude range and many show a secondary peak above 13 km that does not appear in the retrievals (Figure 12b).

[43] One conceivable explanation for the systematic difference between simulated IWP and 3D-IWC retrievals is a possible lack of sensitivity of those retrievals to dense ice contributions from convective core regions, which could arise since the input vertically pointing millimeter cloud radar data are interpreted using the properties of cloud ice and snow [*Seo and Liu*, 2006]. Both millimeter radar and satellite microwave radiometer also lack sensitivity to thin ice clouds, which could lead to underestimates in IWP particularly during the suppressed period. Robust methods of comparing models with measurements using retrievals of varying sensitivity to cloud, snow, and dense ice contributions are not yet generally in hand even when ice classes and properties are well-defined as in CRMs [e.g., *Waliser et al.*, 2009]. A detailed analysis of both retrieval inputs and model outputs would be required to quantitatively assess whether this explanation can account for the systematic differences seen between retrievals and particular simulations in this case. Further work to establish the robust use of microwave-based remote-sensing measurements to constrain CRM and LAM simulations should have a high priority given the sensitivity of results to poorly constrained ice microphysical processes [e.g., *Wu et al.*, 2009; *Fan et al.*, 2010b; *Morrison and Milbrandt*, 2011] and the considerable potential of such measurements to constrain simulation results [e.g., *Matsui et al.*, 2009; *Waliser et al.*, 2009].

[44] Offset a bit lower than simulated and retrieved IWC peaks, simulated latent heating rates peak at ~4–12 km, consistent with retrievals from C-POL measurements (see section 4.1 and Figure 13). Almost all simulations agree remarkably well with retrievals above ~8 km during the active period. Since the vertical integral of latent heating rate is nearly equivalent to the reported surface precipitation rate in all plotted simulations, the larger heating rates above 10 km in SAM simulations appear to be reliably reported features, as discussed further below. In the SAM-2M baseline simulation, the divergence from latent heating rate retrievals above 12 km can be traced to event C alone, whereas in the SAM-2Ms sensitivity test, latent heating rate diverges from retrievals in all three convective events A–C, consistent with greater updraft speeds and vertical mass

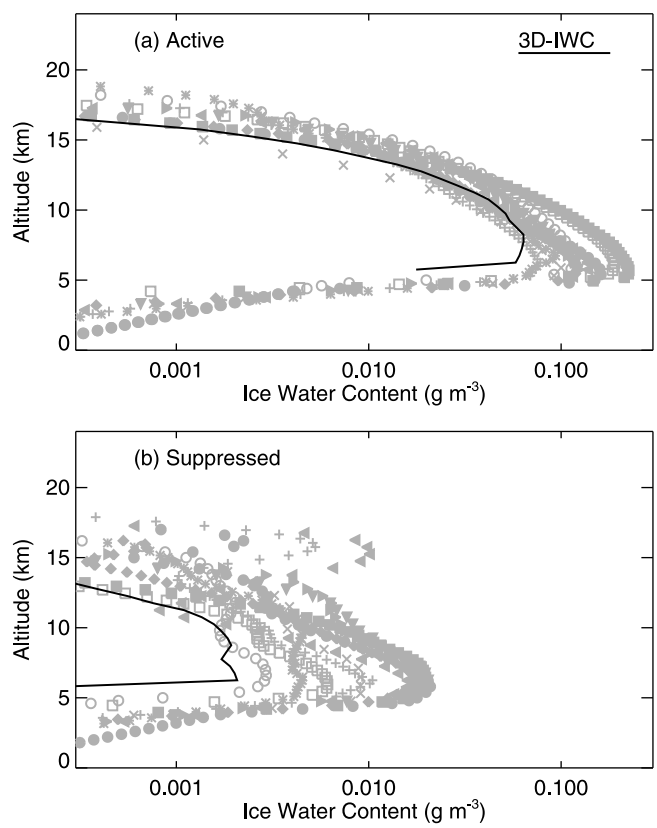

Figure 12. Domain-mean ice water content profiles (including cloud ice, snow, and graupel) averaged over (a) active period (19.5–25.5) and (b) suppressed period (27.5–33.5) compared with 3D-IWC retrievals. Simulation IWC is subsampled at the observational frequency (see section 4.8). Symbols as in Figure 5.

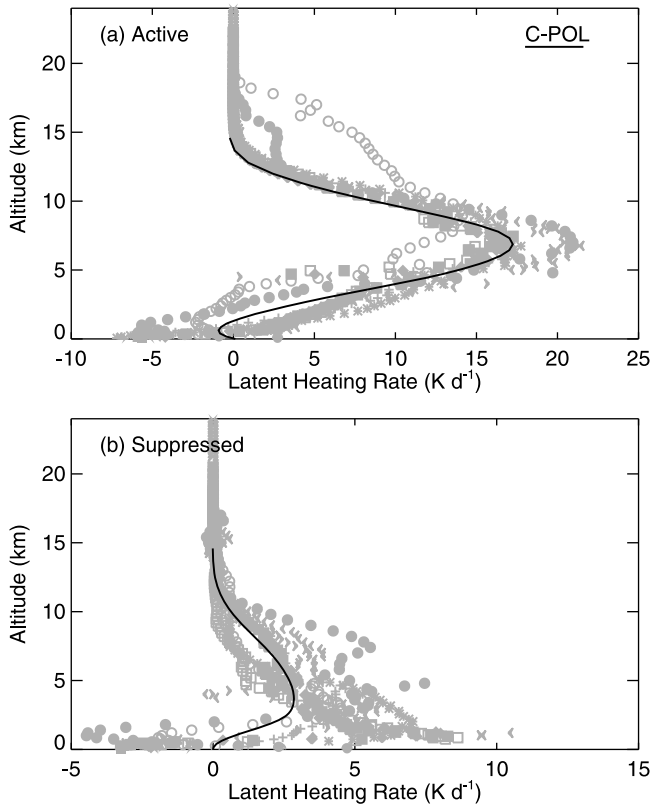

Figure 13. Latent heating rate simulated during (a) active and (b) suppressed periods compared with retrievals from C-POL (normalized as described in section 4.1). Symbols as in Figure 5.

fluxes in sensitivity test simulations (not shown). At the melting level (~5 km), some simulations exhibit a sharp localized reduction in latent heating rate (all DHARMA and MESONH simulations) whereas the others do not. Maximum latent heating rates also appear to fall into two groups during the active period: those that exceed 20 K d^{-1} (MESONH and SAM simulations) and those that do not (all others). Most peak rates in both groups fall within the minimum expected retrieval uncertainty of roughly 25%. In contrast to the relative consistency of simulated and observed latent heating rate profiles during the active period, simulations deviate variously from retrievals during the suppressed period, consistent in part with differences between observed and simulated surface precipitation rates (see Figure 5).

[45] Finally, we note that although cloud cover is available from both ground-based and satellite-based measurements, it is difficult to use retrievals as a quantitative constraint on simulations for two reasons. First, conditions are often continuously overcast in both retrievals and simulations (Figure 14), as during the active period, thus providing little signal. And second, when both VISST and TSI retrievals indicate that clear-sky regions are present, as during the suppressed period, a robust quantitative comparison of say minimum cloud cover obtained by VISST (0.02), TSI (0.1), and DHARMA-1 (0.16) cannot be made owing to fundamental differences in the definition of cloudiness. Since the model definition of cloud cover (a grid cell mixing ratio of ice plus cloud water $>10^{-6}$kg kg^{-1}) and the two measurement-based definitions of cloud cover are not easily reconciled quantitatively, this is a problem that probably requires forward-simulation approaches that are beyond the scope of this work [e.g., *Henderson and Pincus*, 2009].

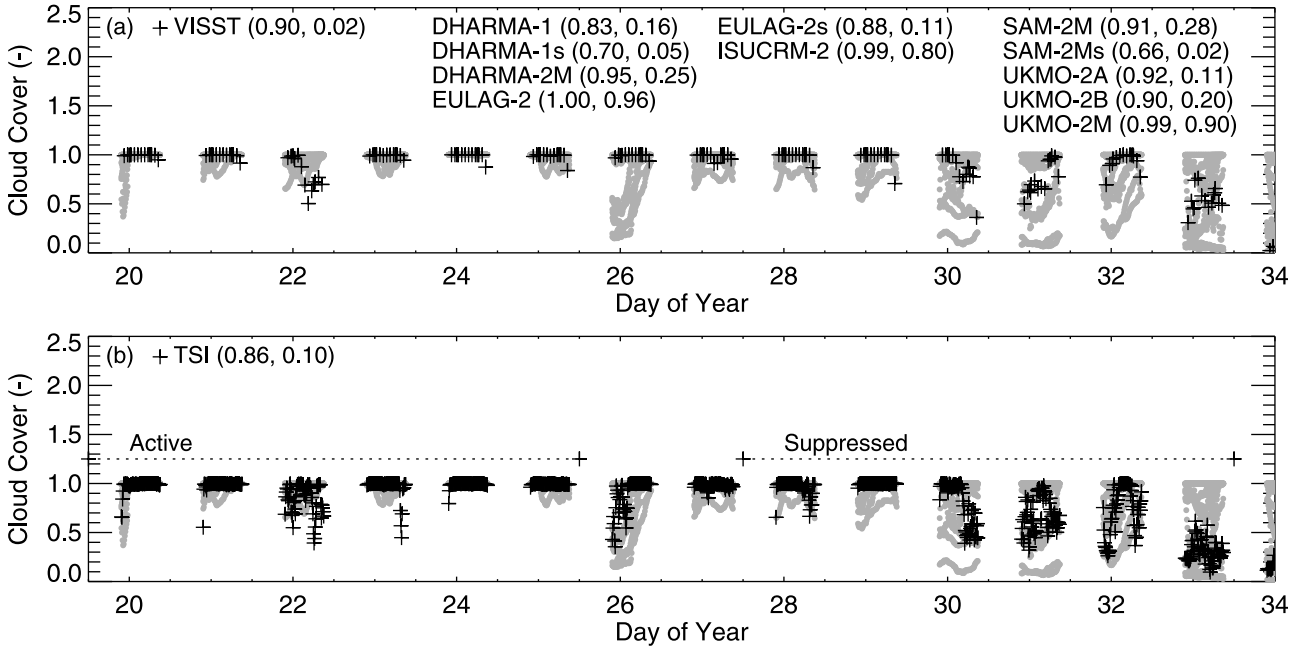

Figure 14. Simulated cloud cover (defined as domain fraction overlain by grid cells with combined liquid and ice condensate excluding rain in excess of 10^{-6} kg kg^{-1}) compared with (a) the fraction of domain grid cells identified as cloudy in VISST retrievals or (b) the opaque cloud cover derived from TSI measurements. Listed in parentheses are the mean and minimum of plotted values.

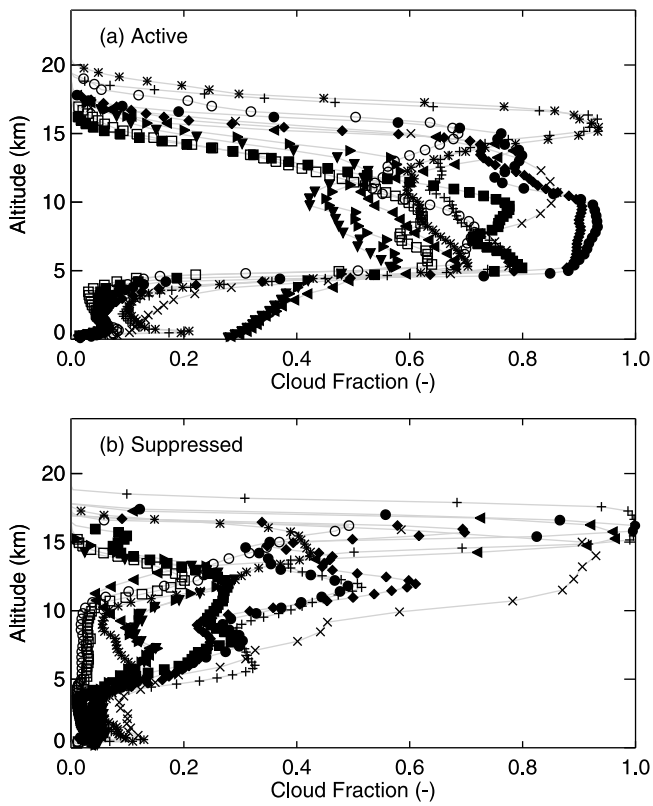

Figure 15. Cloud fraction profiles simulated during (a) active and (b) suppressed periods, where cloudy grid cells are defined as those containing combined liquid and ice condensate excluding rain in excess of 10^{-6} kg kg^{-1}. Symbols as in Figure 5.

[46] Nonetheless, it can be seen that minimum cloud cover varies over 2–96% across the model ensemble compared with 2–10% across the VISST and TSI retrievals. In the profile of cloud fraction, model differences can be traced primarily to extent of ice cloud fraction above the freezing level (Figure 15). Above 15 km during the suppressed period, overcast conditions persist in all baseline simulations that include ice nucleation directly from the vapor phase in the deposition mode, whether based on supercooling or supersaturation (see section 3), with the apparent exception of UKMO-2A and UKMO-2B (see Table 2). The coverage of the persistent cloud layer aloft peaks ~15 km, where large-scale forcings are linearly diminished owing to uncertainties in derived values (see section 2 and discussion by *Fridlind et al.* [2010]), but extending nudging to lower elevations in the sensitivity tests also substantially reduces cloud fraction aloft if it is present in the baseline simulation (e.g., EULAG-2s and SAM-2Ms), indicating a role for convective outflow. Water vapor mixing ratios in excess of the large-scale forcing data set conditions above 15 km tend to occur in SAM and UKMO-2M simulations but not in others, indicating that they are not a determining factor (not shown; results near the tropopause will be compared with radiosonde and in situ water vapor measurements in future work). During the active phase, EULAG is the only model that sustains a high cloud layer. However, the other models sustaining high cloud aloft during the suppressed period also produce the highest cloud tops during the active phase. Overall, differences in high cloud occurrence clearly vary with microphysics scheme (e.g., across DHARMA and UKMO simulations) and are likely attributable in part specifically to treatments of ice nucleation, consistent with past findings [e.g., *Fan et al.*, 2010b].

5.3. Precipitable Water Vapor, Moist Static Energy, and Radiative Fluxes

[47] We have noted already that model overprediction of OLR during the active period is closely correlated with the properties of deep stratiform clouds in the 3D simulations (see Figure 9). *Varble et al.* [2011] have demonstrated that simulated TOA 10.8-μm brightness temperature features are influenced by variable high-level anvil ice outside of precipitating stratiform areas, so this correlation is not the result of singular and direct causation. However, all else being equal, it is expected that sustained overprediction of OLR could lead to excessive radiative cooling of the troposphere during the course of simulations. Given near-saturated conditions, such cooling could increase condensation and subsequent precipitation, thereby reducing precipitable water vapor (PWV). In fact, PWV does fall to levels often persistently more than 5 kg m^{-2} lower than observed in most baseline simulations (Figure 16a). The troposphere has also cooled more than observed in most baseline simulations, as evidenced by biases in mass-weighted dry static energy averaged over 0–17 km (Figure 16b). In contrast to baseline simulations, which are free-running below 15 km, sensitivity test simulations are nudged throughout the troposphere, guaranteeing little deviation from observed PWV and dry static energy.

[48] Because of variations in thermodynamic evolution that are apparent in Figure 16, the free-running baseline simulations enter the suppressed period with a broad range of mean conditions. We therefore focus the remainder of this section on the active period, when simulated PWV values are well correlated with their rates of decline (Figure 17). We also focus on quantities that impact mass-weighted mean tropospheric moist static energy (MSE, Figure 16c), which is negligibly different from a frozen moist static energy in this case [e.g., *Blossey et al.*, 2007]. Since MSE is conserved during hydrostatic adiabatic condensation and evaporation processes, differences in the simulated distribution of total water between vapor and condensate (including precipitation) do not directly modify it [e.g., *Bretherton et al.*, 2006]. In our modeling framework, tropospheric MSE changes between two points in time therefore can be attributed to the accumulated sum of surface turbulent heat fluxes, tropospheric radiative flux divergence, and large-scale forcing and nudging terms [cf. *Blossey et al.*, 2007].

[49] Figure 18 shows total surface turbulent heat fluxes (latent plus sensible) and mean radiative flux convergence over the full atmospheric column (not precisely equal to a tropospheric mean, but shown here for the sake of comparison with observations). Taking DHARMA-1 as an example, surface heat fluxes are only ~5 W m^{-2} lower than in the large-scale forcing data set during the active period, whereas radiative flux convergence is ~45 W m^{-2} lower. The sum of these terms accounts for the MSE drift in DHARMA-1 over the active period (based on closure of

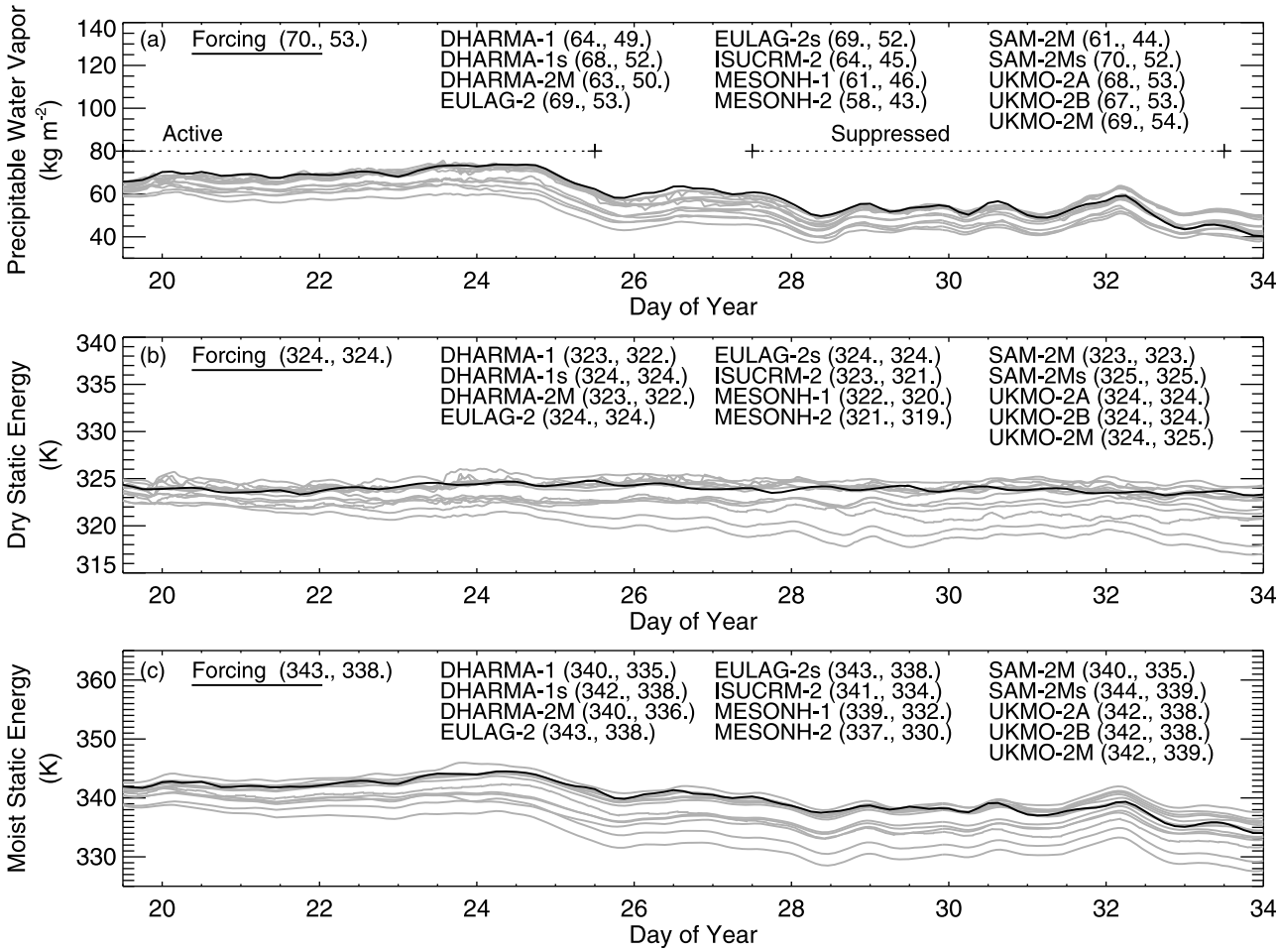

Figure 16. Simulated (a) precipitable water vapor (PWV) and mass-weighted (b) dry and (c) moist static energy averaged over 0–17 km. Static energies are normalized by the specific heat of dry air at constant pressure. Also shown are values derived from the large-scale forcing data set profiles, which differ from simulations at day 19.5 since the first 36 h of simulations were disregarded as spin-up (see section 2). Listed in parentheses are the means during active and suppressed periods.

the tropospheric MSE budget; budget terms not available for most models). Since DHARMA-1s is a sensitivity test simulation, in which tropospheric conditions are nudged toward observed conditions, MSE drift remains minimal despite lower surface fluxes and an even more negative radiative convergence that would otherwise amplify MSE drift. In DHARMA-2M as compared with DHARMA-1, mean surface heat fluxes are little changed. Substantially

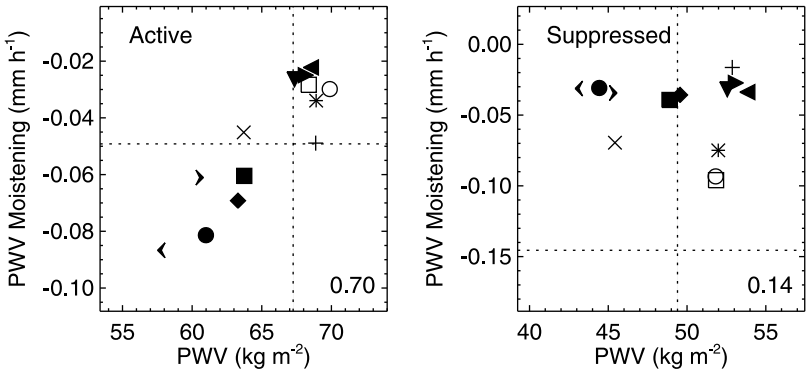

Figure 17. Precipitable water vapor (PWV) versus rate of change in PWV. Averaging times, symbols, and Spearman rank coefficients as in Figure 5. Domain means in the large-scale forcing data set (dotted lines).

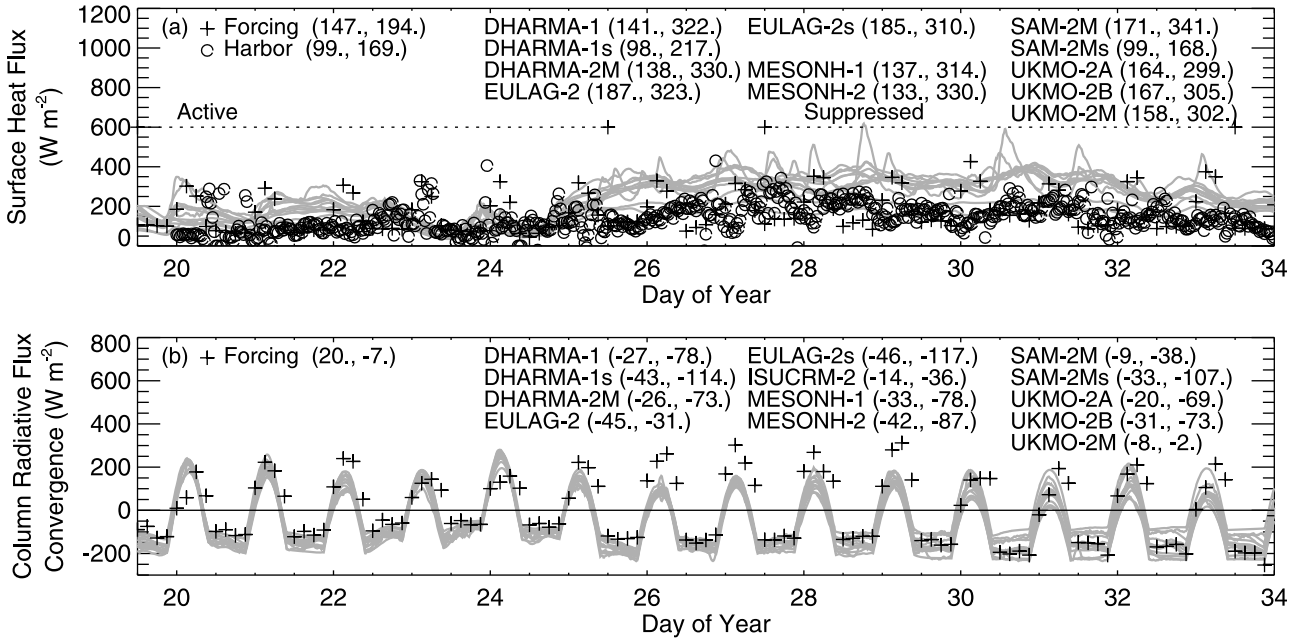

Figure 18. (a) Surface turbulent heat fluxes (latent plus sensible) and (b) radiative flux convergence between the surface and TOA. Also shown are values in the large-scale forcing data set based on domain-wide observational data streams [*Xie et al.*, 2010] and gap-filled point measurements at the Darwin Harbor surface flux site (see section 4.4). Listed in parentheses are the means during active and suppressed periods.

greater stratiform area and high-cloud fractions in DHARMA-2M have little impact on radiative flux convergence relative to DHARMA-1 because reduced longwave cooling (Figure 19b) is offset by reduced solar heating (Figure 20b). Thus MSE drift in DHARMA-1 and DHARMA-2M can be attributed primarily to cloud-modulated radiative flux divergence that is opposite in sign from and larger than that in the large-scale forcing data set.

[50] Considering radiative flux divergence across the ensemble, the most notable feature is that every simulation is

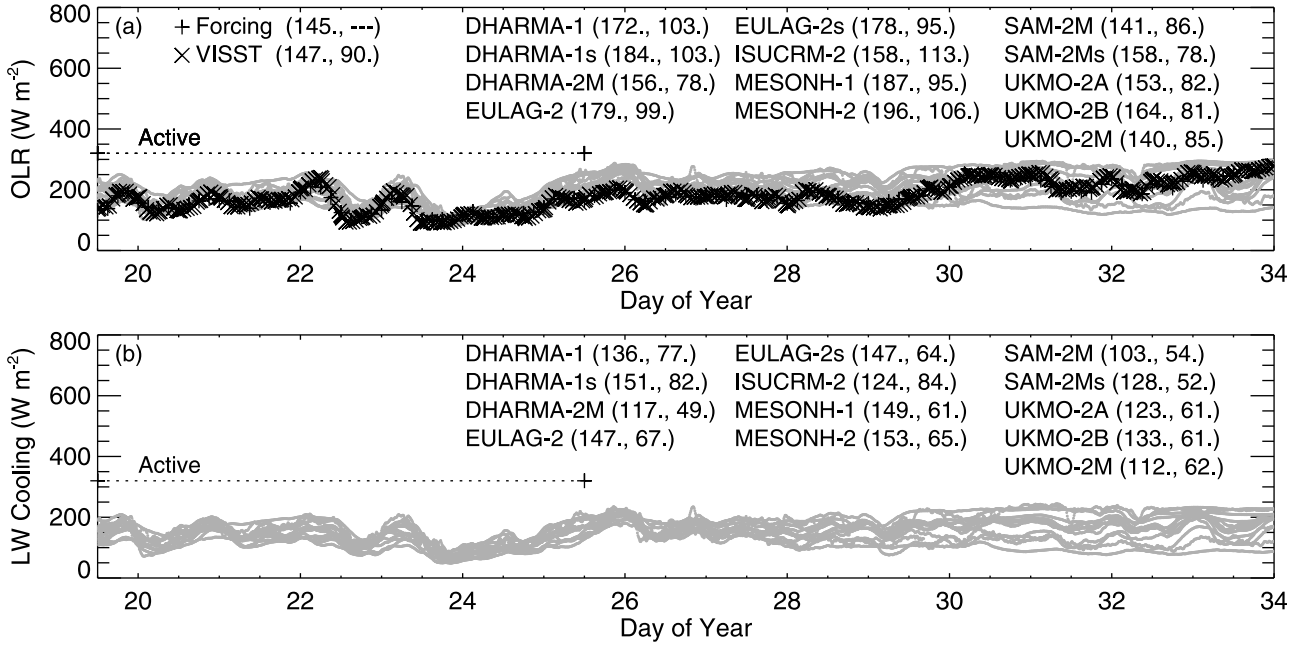

Figure 19. Simulated top-of-atmosphere (TOA) (a) outgoing longwave radiation (OLR) and (b) column longwave (LW) cooling. Column LW cooling not available from the large-scale forcing data set owing to necessary omission of net-only ship-based measurements from available upwelling and downwelling LW fluxes. Listed in parentheses are the mean and minimum during the active period.

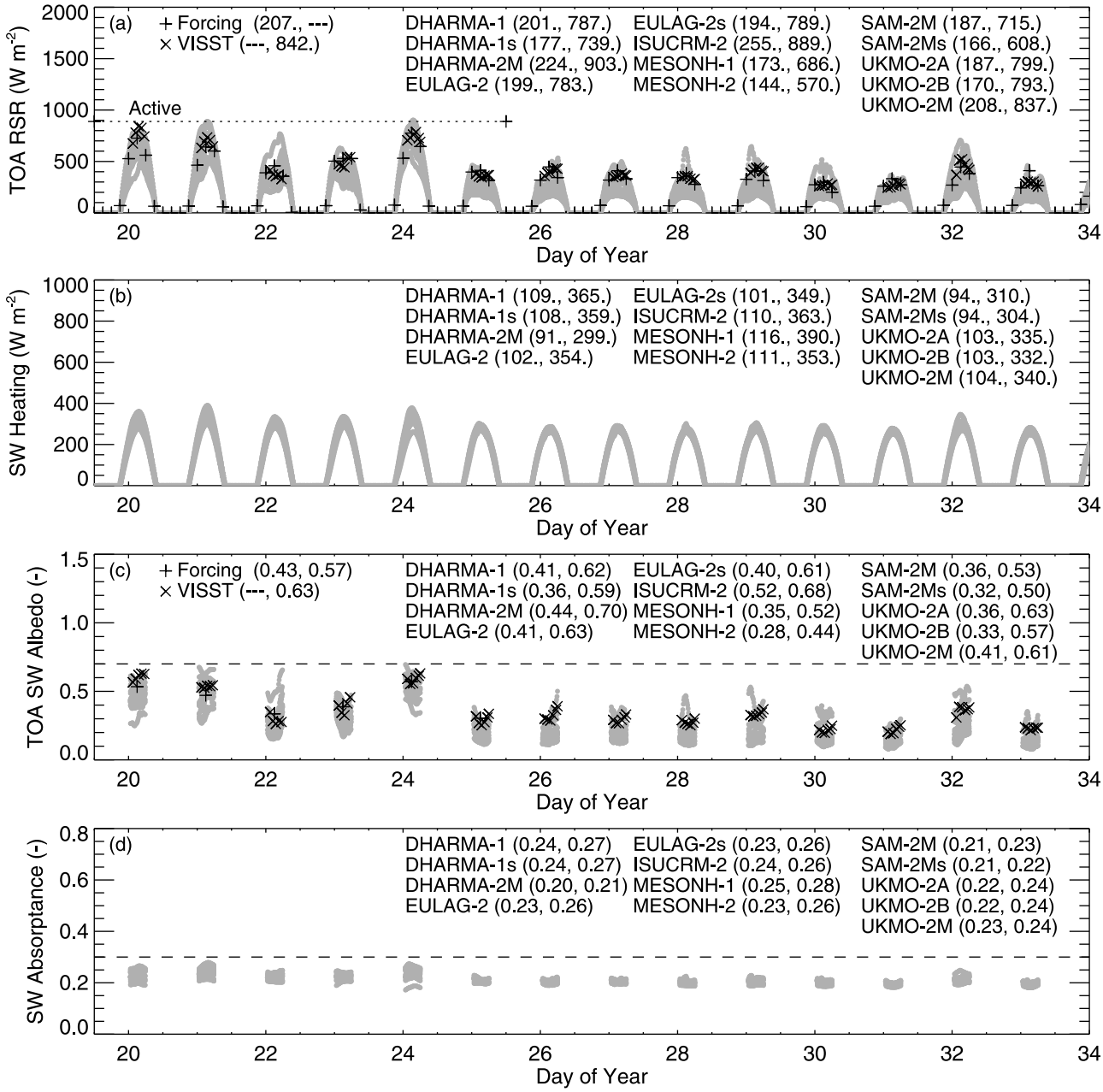

Figure 20. Simulated (a) top-of-atmosphere (TOA) reflected shortwave radiation (RSR), (b) column shortwave heating, (c) TOA shortwave (SW) broadband albedo, and (d) column SW absorptance. Albedo and absorptance shown only for low solar zenith angles, as defined by instantaneous TOA downwelling SW flux > 1100 W m^{-2}; dashed lines indicate observationally derived high-optical-depth limits of about 0.7 and 0.3, respectively [*Dong et al.*, 2008]. Column SW heating not available from the large-scale forcing data set owing to necessary omission of net-only ship-based measurements from available upwelling and downwelling SW fluxes. Listed in parentheses are the mean and maximum during the active period.

biased high by at least 25 W m^{-2}. It is also notable that in the UKMO-2M simulation in particular, TOA shortwave and longwave upwelling fluxes both closely match the forcing data set (Figures 19a and Figure 20a). Thus, differences between simulated and forcing radiative fluxes occur only at the surface in some cases. Simulated surface shortwave and longwave fluxes are biased to an unknown degree by the idealized treatment of the portion of land surfaces within the domain as oceanic. Surface downwelling shortwave fluxes also exhibit high spatiotemporal variability and were necessarily obtained from a sparse network of observing stations. It seems possible that uncertainty in surface fluxes over the TWP-ICE domain could be large enough to reconcile the weak column cooling found in some simulations with the column warming in the forcing data set; biases of 25–30 W m^{-2} in radiative divergence

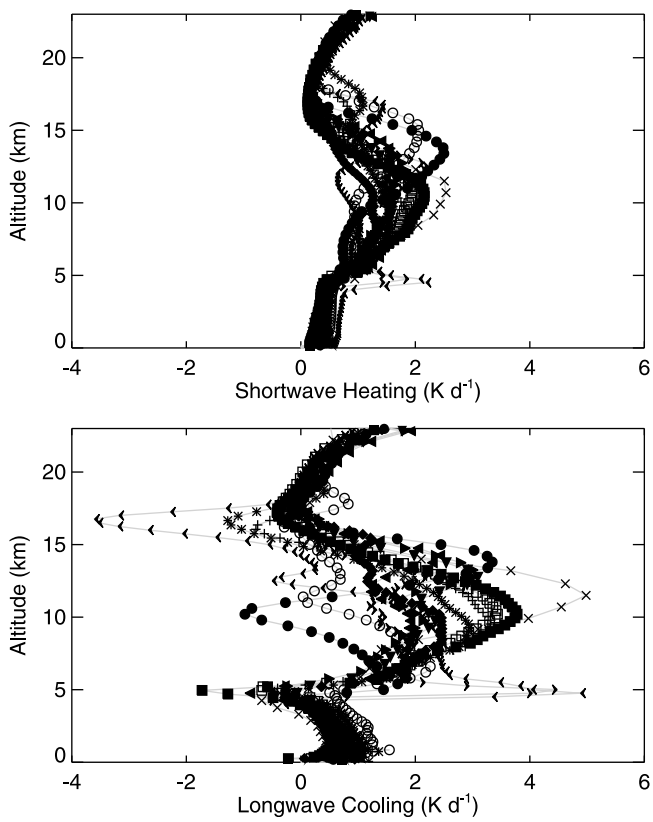

Figure 21. (top) Shortwave heating and (bottom) longwave cooling profiles simulated during the active period. Symbols as in Figure 5.

constitute roughly 15% of shortwave downwelling flux during the active period. Such uncertainties in surface radiative fluxes hinder the accuracy of global radiative budgets [e.g., *Trenberth et al.*, 2009]. In this modeling scenario, radiative divergence uncertainty is also associated with uncertainty in MSE changes over simulation durations of days to weeks.

[51] Although observational uncertainty in the domain mean of net surface radiative fluxes could be too large to strongly constrain simulations, observational analysis of solar albedo and column absorptance at low solar zenith angles has indicated maximum respective values of about 0.7 and 0.3 in the optically thick limit [*Dong et al.*, 2008]. Although maximum domain-wide albedos reach 0.7 at high solar zenith angles in one baseline simulation (Figure 20c), none exceed that value, and no domain-wide absorptances exceed 0.3 (Figure 20d). SAM simulations produce the greatest solar heating in a layer at ~13–16 km (Figure 21), perhaps associated with processes that led to the greatest latent heating at similar elevations (see Figure 13). Such excursions in the shortwave heating profile are generally associated with excursions in longwave cooling, as also seen in MESONH simulations near the melting level. Differences in shortwave and longwave heating rate profiles across the ensemble can be attributed in part to differences in the vertical distribution of hydrometeors (see Figure 15); the treatment of hydrometeor radiative properties also probably plays a role that deserves further scrutiny.

[52] Considering surface turbulent heat fluxes during the active period, values are highest in EULAG-2 and EULAG-2s, resulting in little MSE drift in EULAG-2 despite relatively low radiative convergence. Surface heat fluxes are by contrast lowest in the 3D sensitivity tests (DHARMA-1s and SAM-2Ms), in which they nearly equal the Darwin Harbor data from the active period (see Figure 18a), consistent with idealized marine conditions. That the 2D sensitivity test EULAG-2s fluxes are dissimilar indicates a possible role of dimensionality. Latent heat flux dominates sensible heat flux in simulations and observations, and simulation differences are related to near-surface relative humidities (not shown). For instance, the 3D sensitivity tests with lowest heat fluxes also exhibit the highest mean near-surface relative humidities (90%). However, near-surface relative humidities are lower in MESONH than in EULAG, indicating that some combination of dimensionality, near-surface winds (see section 3), and flux parameterization may determine results. The role of microphysics appears comparatively weak (e.g., across DHARMA, MESONH, and UKMO simulations). Overall, whether or not surface heat fluxes play a determining role in the local initiation and maintenance of deep convective systems during the active period (see section 2), they can modulate tropospheric MSE and PWV evolution over CRM integration times as short as several days in the modeling framework used here.

[53] Returning to the general question of how large differences in deep stratiform cloud properties could influence tropospheric heating and cooling rates, and considering only TOA, where observational constraints are strongest, we lastly note that most models overpredict OLR during the active period. However, if lines of offsetting changes in TOA OLR and RSR are drawn through the limits of observational uncertainty in OLR and RSR (see Figure 22), most simulations fall between these lines during both active and suppressed periods. The tendency of cloud-associated shifts in TOA OLR and RSR to balance in the tropics has long been noted [e.g., *Kiehl*, 1994], and a similar pattern is evident across the ensemble of simulations. Given any two simulations that differ only in microphysics scheme or large-scale forcing during the active period (e.g., across DHARMA, SAM or UKMO simulations), greater stratiform area fraction tends to be consistently associated with a decrease in OLR and an increase in RSR that are similar in magnitude.

[54] Identifying which factors determine the relatively stable baseline level of TOA RSR versus OLR in each model, from which departures appear to result in roughly equal offsets of longwave and shortwave fluxes, is beyond the scope of this study owing to a lack of sufficient diagnostics (additional diagnostics are suggested below). It is unknown whether such model differences could be attributable to dynamically or microphysically modulated relative distributions of primarily high or low clouds, for instance, or also their radiative treatment. In 3D models, compensating offsets in OLR and RSR can be associated only in part with deep stratiform area fraction, as evidenced by its weaker correlation with broadband albedo than with OLR (Figure 22 versus Figure 9). TOA albedo may also be influenced by variations in effective radius of liquid-phase condensate (Figure 22), which contributes most optical depth to simulations (not shown). 2D TOA radiative flux fields, which

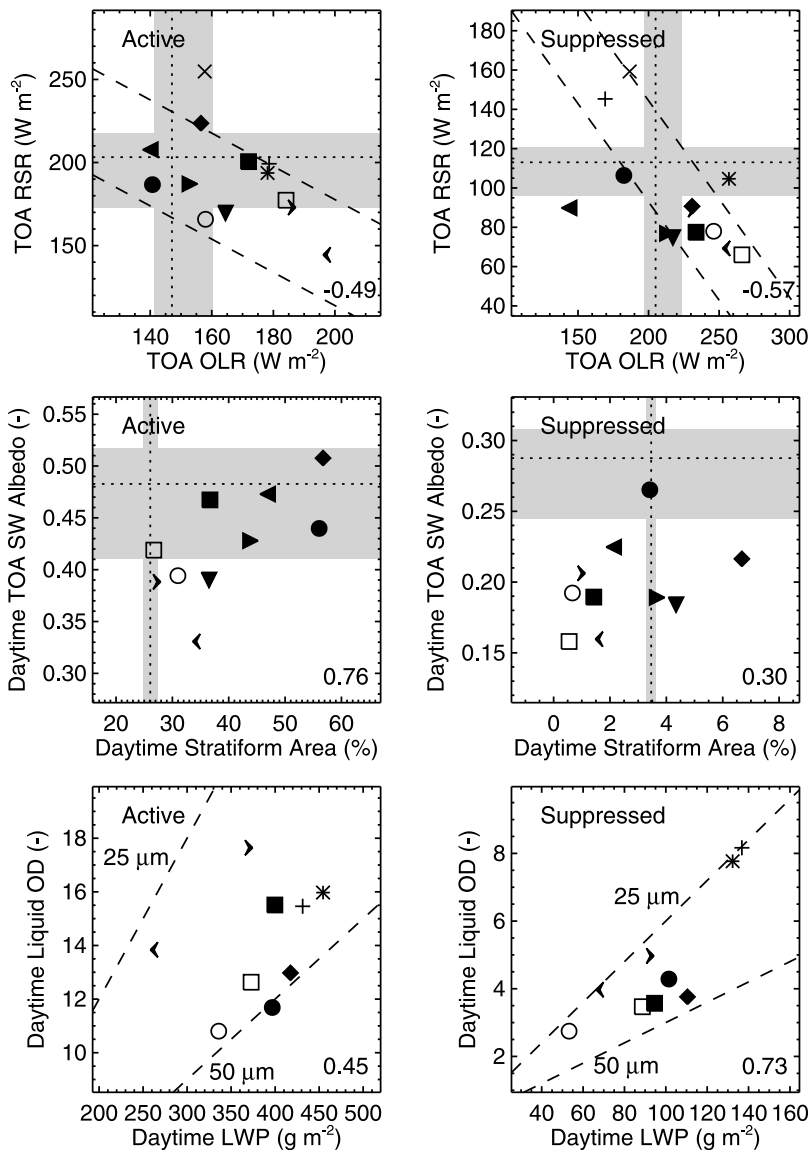

Figure 22. (top) Top-of-atmosphere reflected shortwave radiation (TOA RSR) versus TOA outgoing longwave radiation (OLR; dashed lines are 1:1 drawn through the intersection of observational values plus and minus their uncertainties), (middle) daytime SW albedo versus stratiform area fraction, and (bottom) daytime LWP versus liquid optical depth (OD; dashed lines indicate domain-mean effective radius, defined as $1.5 \cdot \mathrm{LWP}/(\rho_w \cdot \mathrm{OD})$, of 25 μm and 50 μm). Averaging times, symbols, and Spearman rank coefficients as in Figure 5. Domain means of RSR and OLR from large-scale forcing data set (derived from VISST retrievals), daytime SW albedo directly from VISST retrievals, and stratiform area fraction from analysis of C-POL data shown with dotted lines and uncertainty ranges (shading).

were not requested output here, could be used to more fully diagnose differences between simulations and observations, as demonstrated by *Varble et al.* [2011].

6. Summary and Conclusions

[55] Observations of consecutive active and suppressed monsoon periods around Darwin, Australia from 18 January to 3 February 2006 during the TWP-ICE campaign are used to drive and evaluate multiple cloud-resolving models (CRMs) with periodic boundary conditions, most using a horizontal grid resolution of ~1 km. Baseline simulations represent an idealized marine case study. Sensitivity test simulations include nudging domain-mean water vapor and potential temperature to observations throughout the troposphere. Since baseline simulations enter the suppressed monsoon period with a wide range of mean conditions, results from the earlier active monsoon period provide a better indication of ensemble performance. Agreement of the 13 simulations with domain-wide observational data sets over the active period is summarized in Table 3, and associated conclusions can be summarized as follows.

[56] 1. All 13 simulations reliably reproduce domain-mean precipitation rates (Table 3, Precipitation column), consistent

Table 3. Simulated Domain-Mean Quantities Averaged Over the Active Monsoon Period (19.5–25.5) That Are Within the Range of Observational Data Plus and Minus Uncertainties[a]

Simulation	Precipitation[b]	Area[c]	Convective[d]	Stratiform[d]	IWP[e]	OLR[f]	RSR[g]
DHARMA-1	yes	+	+	+	+	+	yes
DHARMA-1s	yes	+	+	–	+	+	yes
DHARMA-2M	yes	+	+	+	+	yes	+
EULAG-2	yes	+			+	+	yes
EULAG-2s	yes	+			+	+	yes
ISUCRM-2	yes				+	yes	+
MESONH-1	yes	+	+	–	+	+	yes
MESONH-2	yes	+	yes	yes	+	+	–
SAM-2M	yes		+	+	+	–	yes
SAM-2Ms	yes		+	+	+	yes	–
UKMO-2A	yes		yes	+	+	yes	yes
UKMO-2B	yes		+	+	+	+	–
UKMO-2M	yes		+	+	+	–	yes

[a]Within the range: yes; higher than that range: +; lower than that range: –; not diagnosed or reported: blank.
[b]Mean surface precipitation rate versus C-POL with uncertainty of 25% (see section 4.1, Figure 5).
[c]Total occurrence frequency of precipitation rates exceeding 0.2 mm h^{-1} at 2.5-km elevation and 2.5-km resolution versus C-POL range of 0.21–0.28 (see section 4.1, Figure 6a).
[d]Fractional area of convective and stratiform rain in 3D models versus C-POL (using the algorithm described in Appendix B) with uncertainties of 20% and 5%, respectively (see section 4.1, Figure 9).
[e]Ice water path versus 3D-IWC with uncertainty of 20% (see section 4.8, Figure 11).
[f]TOA outgoing longwave radiation (OLR) versus VISST (same as large-scale forcing data set) with uncertainty of +9/–4% (see section 4.2, Figure 22).
[g]TOA reflected shortwave radiation (RSR) versus the large-scale forcing data set (based on VISST) with uncertainty of +7/–15% (see section 4.2, Figure 22).

with similar past studies in which models are constrained by strong large-scale forcing terms [e.g., *Xu et al.*, 2002; *Xie et al.*, 2005; *Woolnough et al.*, 2010]. However, simulations deviate systematically from observations with respect to the underlying precipitation rate distributions (Table 3, Area, Convective and Stratiform columns). The area covered by rain rates greater than 0.2 mm h^{-1} is overestimated (Table 3, Area column), a tendency that is reduced in sensitivity test simulations (Figure 6a). Thus excessive precipitating area appears at least partly attributable to drift of mean tropospheric conditions from those observed. Several lines of evidence also indicate that the strongest rain is locally too intense or widespread in the 3D simulations (see 2 below), consistent with results from a similar past study of tropical convection around Kwajalein Island [*Blossey et al.*, 2007].

[57] 2. In the ten 3D simulations, areas covered by convective and deep stratiform rain are diagnosed from simulated radar reflectivity using a textural algorithm (Appendix B, Figure 7). Simulated convective area fractions are systematically larger than observed by C-POL, by up to a factor of ~2 (Figure 9), and are within measurement uncertainty in only two cases (Table 3, Convective column). Simulated stratiform area fractions by contrast can be too large by up to a factor of two or too small (Figure 9), consistent with past findings of variable stratiform rain across CRMs [*Xie et al.*, 2002], and only one lies within the narrow measurement uncertainty (Table 3, Stratiform column). The largest stratiform area fractions are simulated with *Morrison et al.* [2009] two-moment microphysics, consistent with the expected importance of microphysics scheme [*Morrison et al.*, 2009; *Li et al.*, 2009; *Luo et al.*, 2010; *Bryan and Morrison*, 2012]. However, the slower-than-observed decay of convective area from simulated peaks seen here (Figure 8) is probably caused at least partly by using periodic boundary conditions and neglecting large-scale advective divergence of condensate. The usefulness of this modeling approach for directly constraining stratiform areal coverage may therefore be limited; a companion study using limited-area models demonstrates an approach that could be more suitable (see section 1).

[58] 3. Simulated ice water path (IWP) in all 13 simulations is systematically higher than domain-mean 3D-IWC retrievals (Table 3, IWP column), commonly by a factor of two, with only 2D simulations nearly within estimated uncertainty (Figure 11). However, CRMs reproduce the retrieved temporal evolution and vertical distribution of IWP and retrieved latent heating rate profiles qualitatively well (Figures 10b, 12a, and 13a). 3D-IWC retrievals may underestimate dense ice contributions to IWP (see section 5.2), but systematic overestimation of IWP by simulations cannot be ruled out. Identifying the sources of discrepancy between simulated and retrieved IWP should be a priority owing to the wide variability of ice distribution documented across models generally, its importance to radiative fluxes, and the potential of satellite microwave data for providing strong constraints [e.g., *Waliser et al.*, 2009; *Wu et al.*, 2009; *Su et al.*, 2011].

[59] 4. At the top-of-atmosphere (TOA), where radiative fluxes are most strongly constrained by observations, simulated outgoing longwave radiation (OLR) is usually higher than observed (Figure 22), and is within experimental uncertainty in only four cases (Table 3, OLR column). TOA reflected shortwave radiation (RSR) is usually lower than observed (Figure 22), although within the estimated uncertainty range (Table 3, RSR column). Persistently high OLR (Figure 19a) and low RSR (Figure 20a) are strikingly similar to those reported by *Blossey et al.* [2007] (cf. their Figure 6). In 3D models the OLR is strongly anticorrelated with stratiform area fraction (Figure 9). That OLR is by contrast negligibly correlated with IWP indicates that the spatial distribution of ice is more important than its area-averaged path. Furthermore, in multiple simulations from a given dynamics model, increasing stratiform coverage is

associated with roughly equal changes in OLR and RSR (Figure 22). However, the absolute ratio of RSR to OLR varies as a function of OLR (e.g., DHARMA > UKMO > SAM) for reasons that cannot be adequately assessed from reported output (see section 5.3).

[60] 5. The UKMO-2A simulation agrees best overall with domain-wide data streams (Table 3), but systematic differences between observations and all simulations including UKMO-2A (e.g., in timing of convective area in Figure 8) suggest that close agreement may not be expected using this modeling framework. Errors associated with neglecting advective divergence of condensate, such as the cirrus inflow following event C during TWP-ICE [*Cohen*, 2008; *May et al.*, 2008], could be significant especially in the upper troposphere [*Grabowski et al.*, 1996; *Petch and Dudhia*, 1998]. However, across the ensemble of simulations, the wide spread of stratiform fraction and its close correlation with OLR indicate a need for more rigorous assessment of the structural, microphysical, and radiative properties of convective, stratiform, and anvil clouds. Analysis of such structural features could have been usefully extended if requested results had included (1) 2D TOA broadband longwave, shortwave, and relevant narrow-band radiative fluxes, (2) associated 3D radiative fluxes for attribution to underlying hydrometeor distributions, (3) 3D precipitation rates for division of rainfall into convective and stratiform components, and (4) 2D surface radiative and turbulent heat fluxes.

[61] Although it is beyond the scope of this study to quantify uncertainties in column radiative flux divergence and surface turbulent heat fluxes associated with sparse surface station measurements over ocean and land, and treatment of the observational domain as entirely marine in these idealized simulations introduces bias, several additional conclusions can nonetheless be drawn. First, simulated values of domain-mean column solar absorption are below an observationally determined maximum for the optically thick limit (see section 5.3). Second, deviations of predicted column radiative flux divergence and surface turbulent heat fluxes from the large-scale forcing data set are commonly large enough to drive substantial drifts in tropospheric water vapor, temperature, and moist static energy over only a few days. Nudging the domain-mean profiles to large-scale conditions as done in sensitivity tests here is one means of preventing drift. An exhaustive closure approach [e.g., *Fridlind and Jacobson*, 2003] would be to use an ensemble of large-scale forcings that includes uncertainties of inputs to the variational analysis [e.g., *Hume and Jakob*, 2007]. A companion TWP-ICE SCM intercomparison study considers such uncertainty in the surface precipitation rate input (see section 1).

[62] The variability of simulation drifts from observed conditions limits the value of comparing simulations with either observations or with one another during the suppressed period; simulated conditions entering the suppressed period are too broad to attribute differences to model physics going forward. It is nonetheless notable that predicted surface precipitation rates exceed observations by more than experimental uncertainty in all cases except one (Figure 5), suggesting that models may form or sustain convection too easily during the suppressed period. Convective area fractions correspondingly exceed observations, but stratiform area fractions are again variably overpredicted and underpredicted and uncorrelated with either convective area fractions or IWP (Figure 9). Highly variable latent heating rates in the lower troposphere (Figure 13) and ice water contents aloft (Figure 12) point to a general divergence of simulation results during the suppressed period.

[63] Differences across the simulation ensemble arise noticeably from multiple sources, including treatments of dynamics, microphysics, and radiation. Compared with 2D models, the 3D models tend to produce higher peak rain rates and IWP/LWP ratios (Figures 4c and 11), consistent with past studies finding stronger updrafts and larger vertical mass fluxes that impact ice formation processes and increase anvil ice mass in 3D [*Redelsperger et al.*, 2000; *Petch and Gray*, 2001; *Phillips and Donner*, 2006; *Petch et al.*, 2008; *Zeng et al.*, 2008]. Ensemble spread can be attributed also to microphysics schemes (see 2–4 above), consistent with past CRM studies [e.g., *Wu et al.*, 1999; *Grabowski et al.*, 1999; *Redelsperger et al.*, 2000; *Xu et al.*, 2002; *Xie et al.*, 2005; *Morrison et al.*, 2009; *Fan et al.*, 2010b; *Luo et al.*, 2010; *Bryan and Morrison*, 2012]. Prognosing droplet number concentration, an important advance for incorporating aerosol effects on cloud properties, did not produce systematically distinguishing effects across the ensemble here, for instance in IWP, radiative fluxes, or domain-mean water-drop effective radius (e.g., DHARMA-2M, EULAG, MESONH-2, and SAM-2M in Figure 22), likely owing in part to differences in aerosol and activation treatments (see section 3).

[64] Returning to our first originating question, do simulations and observations agree within experimental uncertainties? The short answer is that even when simulations reproduce domain-mean surface precipitation, cloud structural and radiative properties are often not reproduced (see Table 3). Resulting deviations of column radiative flux divergence from observationally derived values, if not compensated by errors in surface turbulent heat fluxes, lead to substantial drift of predicted tropospheric conditions from those observed. In addition, systematic differences between simulated and observed timing of convective structures suggest that periodic boundary conditions could be an important source of discrepancy. Applying large-scale horizontal advective tendencies to condensate would not likely resolve timing errors that differ for convective and stratiform features. The modeling approach used here therefore does not appear sufficiently robust to reproduce the observed precipitating cloud structures and their radiative effects in this case. Regardless, the wide spread of predicted stratiform area fraction and closely correlated radiative impacts indicate a need for more rigorous observation-based evaluation of the ability of models to reproduce the fundamental micro- and macrophysical properties of convective cloud structures. The factors specifically controlling simulated convective and stratiform properties deserve further study owing to their effects on tropical dynamics, large-scale circulation and precipitation patterns, and climate sensitivity [e.g., *Donner et al.*, 2001; *Schumacher and Houze*, 2003; *Song and Zhang*, 2011; *Houze*, 2004; *Fu and Wang*, 2009].

Appendix A: Data Archive

[65] Unless otherwise indicated, all measurements used here were downloaded from the ARM data archive (http://

www.arm.gov). The modeling case study specification, input files, simulation results, and processed observational data sets produced for this study are also stored there (http://www.arm.gov/campaigns/twp2006twp-ice/). Archived simulation results include scalars and profiles at 10-minute frequency, and 3D model fields at 3-h frequency. All model fields and processed data are archived in compliance with the netCDF Climate and Forecast (CF) metadata convention (http://cf-pcmdi.llnl.gov) version 1.3, insuring standardized metadata, variable names, and units, enhancing ease of use. Archived material is intended to allow the case to be run independently and results compared with those shown here, or users can alternatively download simulation results and treat them as an ensemble without running the case.

Appendix B: Convective and Stratiform Area

[66] We identify regions of convective and stratiform rain in observed and simulated radar reflectivity fields based on the textural algorithm described by *Steiner et al.* [1995], with the added requirement that reflectivity be at least 0 dBZ at 6 km elevation in stratiform columns to avoid inclusion of isolated shallow convection as stratiform outflow. Adding such an echo requirement aloft (whether 0 or 5 dBZ at 6 or 8 km) exposed strong correlations of identified stratiform area with both OLR and RSR across the simulation ensemble that were otherwise absent. Thus echo aloft provides an efficient proxy for stratiform structural properties that are closely associated with radiative fluxes. To treat observations and models identically, we linearly interpolate model fields of equivalent reflectivity (Z_e) to obtain a single slice at an elevation of 3 km and degrade model resolution to 2.5-km horizontal resolution in a manner that conserves Z_e. Reflectivities weaker than 0 dBZ are set to missing values in both model and C-POL fields. We then apply the three-step algorithm to identify convective pixels as described by *Steiner et al.* [1995] (their section 2c), where background intensity considered in steps two and three is averaged over values of dBZ $\geq$ 0 ($Z_e \geq 1$). Since the optimal stepwise algorithm coefficients reported by *Steiner et al.* [1995] were developed using the same instrument, location, and meteorological conditions, we consider them adequate for our purpose of comparing observed and simulated structures despite the reduction in resolution from 2 km (used in their study) to 2.5 km (the resolution at which C-POL data are archived). A more detailed analysis of convective and stratiform rain structures in 3D simulations is made by *Varble et al.* [2011], where the algorithm for identifying convective and stratiform regions differs slightly, including exclusion of all columns with 2.5-km reflectivity <5 dBZ and no echo requirement aloft in stratiform columns.

[67] **Acknowledgments.** This research was supported by the NASA Radiation Sciences Program and by the DOE Office of Science, Office of Biological and Environmental Research, through Contracts DE-AI02-06ER64173, DE-AI02-08ER64547, and DE-FG03-02ER63337 (Fridlind and Ackerman), DE-FG02-08ER64574 (Grabowski and Morrison), DE-AI02-07ER64546 (Minnis), and DE-FG02-08ER64559 (Wu), and the DOE Atmospheric System Research Program (Fan). Computational support was provided by the DOE National Energy Research Scientific Computing Center and the NASA Advanced Supercomputing Division. We thank the TWP-ICE and ACTIVE field campaign teams led by Peter May and Geraint Vaughan. TWP-ICE data were obtained from the ARM Program archive, sponsored by the DOE Office of Science, Office of Biological and Environmental Research, Environmental Science Division. ECMWF analyses were provided to the ARM data archive under a site license agreement. Sally McFarlane is thanked for help in Rayleigh reflectivity calculations for SAM simulations. We thank Ed Zipser and Chris Bretherton for helpful discussions. We thank Steven Krueger and an anonymous reviewer for detailed corrections and comments.

References

Ackerman, A. S., O. B. Toon, D. E. Stevens, A. J. Heymsfield, V. Ramanathan, and E. J. Welton (2000), Reduction of tropical cloudiness by soot, *Science*, *288*(5468), 1042–1047, doi:10.1126/science.288.5468.1042.

Barnett, T. P., J. Ritchie, J. Foat, and G. Stokes (1998), On the space-time scales of the surface solar radiation field, *J. Clim.*, *11*(1), 88–96.

Barth, M. C., et al. (2007), Cloud-scale model intercomparison of chemical constituent transport in deep convection, *Atmos. Chem. Phys.*, *7*, 4709–4731.

Bechtold, P., et al. (2000), A GCSS model intercomparison for a tropical squall line observed during TOGA-COARE. II: Intercomparison of single-column models and a cloud-resolving model, *Q. J. R. Meteorol. Soc.*, *126*(564), 865–888.

Beringer, J., L. B. Hutley, N. J. Tapper, and L. A. Cernusak (2007), Savanna fires and their impact on net ecosystem productivity in North Australia, *Global Change Biol.*, *13*(5), 990–1004, doi:10.1111/j.1365-2486.2007.01334.x.

Bigg, E. K. (1953), The supercooling of water, *Proc. Phys. Soc. London, Sect. B*, *66*, 688–694.

Blossey, P. N., C. S. Bretherton, J. Cetrone, and M. Kharoutdinov (2007), Cloud-resolving model simulations of KWAJEX: Model sensitivities and comparisons with satellite and radar observations, *J. Atmos. Sci.*, *64*(5), 1488–1508.

Bretherton, C. S., P. N. Blossey, and M. E. Peters (2006), Interpretation of simple and cloud-resolving simulations of moist convection–radiation interaction with a mock-Walker circulation, *Theor. Comput. Fluid Dyn.*, *20*(5), 421–442.

Brown, P. R. A., and A. J. Heymsfield (2006), The microphysical properties of tropical convective anvil cirrus: A comparison of models and observations, *Q. J. R. Meteorol. Soc.*, *127*(575), 1535–1550.

Bryan, G. H., and H. Morrison (2012), Sensitivity of a simulated squall line to horizontal resolution and paramerization of microphysics, *Mon. Weather Rev.*, *140*, 202–225, doi:10.1175/MWR-D-11-00046.1.

Cohen, E. A. (2008), Observations of cirrus cloud evolution during the Tropical Warm Pool International Cloud Experiment, Master's thesis, Univ. of Utah, Salt Lake City.

Cooper, W. A. (1986), Ice initiation in natural clouds, in *Precipitation Enhancement—A Scientific Challenge*, *Meteorol. Monogr.*, vol. 43, pp. 29–32, Am. Meteorol. Soc., Boston, Mass.

Del Genio, A. D., and J. Wu (2010), The role of entrainment in the diurnal cycle of continental convection, *J. Clim.*, *23*(10), 2722–2738, doi:10.1175/2009JCLI3340.1.

Dong, X., et al. (2008), Using observations of deep convective systems to constrain atmospheric column absorption of solar radiation in the optically thick limit, *J. Geophys. Res.*, *113*, D10206, doi:10.1029/2007JD009769.

Donner, L. J., C. J. Seman, R. S. Hemler, and S. Fan (2001), A cumulus parameterization including mass fluxes, convective vertical velocities, and mesoscale effects: Thermodynamic and hydrological aspects in a general circulation model, *J. Clim.*, *14*, 3444–3463.

Fan, J., J. M. Comstock, and M. Ovchinnikov (2010a), The cloud condensation nuclei and ice nuclei effects on tropical anvil characteristics and water vapor of the tropical tropopause layer, *Environ. Res. Lett.*, *5*, 044005, doi:10.1088/1748-9326/5/4/044005.

Fan, J., J. M. Comstock, M. Ovchinnikov, S. A. McFarlane, G. McFarquhar, and G. Allen (2010b), Tropical anvil characteristics and water vapor of the tropical tropopause layer: Impact of heterogeneous and homogeneous freezing parameterizations, *J. Geophys. Res.*, *115*, D12201, doi:10.1029/2009JD012696.

Fletcher, N. H. (1962), *The Physics of Rainclouds*, 390 pp., Cambridge Univ. Press, New York.

Franklin, C. N., C. Jakob, M. Dix, A. Protat, and G. Roff (2012), Assessing the performance of a prognostic and a diagnostic cloud scheme using single column model simulations of twp-ice, *Q. J. R. Meteorol. Soc.*, doi:10.1002/qj.954, in press.

Frederick, K., and C. Schumacher (2008), Anvil characteristics as seen by C-POL during the Tropical Warm Pool International Cloud Experiment (TWP-ICE), *Mon. Weather Rev.*, *136*(1), 206–222.

Fridlind, A. M., and M. Z. Jacobson (2003), Point and column aerosol radiative closure during ACE 1: Effects of particle shape and size, *J. Geophys. Res.*, *108*(D3), 4094, doi:10.1029/2001JD001553.

Fridlind, A. M., A. S. Ackerman, J. Petch, P. R. Field, A. Hill, G. G. McFarquhar, S. Xie, and M. Zhang (2010), ARM/GCSS/SPARC TWP-

ICE CRM intercomparison study, *NASA Tech. Memo., NASA TM-2010-215858*, 24 pp.
Fu, X., and B. Wang (2009), Critical roles of the stratiform rainfall in sustaining the Madden-Julian Oscillation: GCM experiments, *J. Clim.*, *22*, 3939–3959.
Ghan, S. J., et al. (2000), An intercomparison of single column model simulations of summertime midlatitude continental convection, *J. Geophys. Res.*, *105*, 2091–2124.
Grabowski, W. W. (1999), A parameterization of cloud microphysics for long-term cloud-resolving modeling of tropical convection, *Atmos. Res.*, *52*(1–2), 17–41.
Grabowski, W. (2001), Coupling cloud processes with the large-scale dynamics using the Cloud-Resolving Convection Parameterization (CRCP), *J. Atmos. Sci.*, *58*(9), 978–997.
Grabowski, W. W., X. Wu, and M. W. Moncrieff (1996), Cloud-resolving modeling of tropical cloud systems during Phase III of GATE. Part I: Two-dimensional experiments, *J. Atmos. Sci.*, *53*(24), 3684–3709.
Grabowski, W. W., X. Wu, M. W. Moncrieff, and W. D. Hall (1998), Cloud-resolving modeling of cloud systems during Phase III of GATE. Part II: Effects of resolution and the third spatial dimension, *J. Atmos. Sci.*, *55*(21), 3264–3282.
Grabowski, W., X. Wu, and M. Moncrieff (1999), Cloud resolving modeling of tropical cloud systems during Phase III of GATE. Part III: Effects of cloud microphysics, *J. Atmos. Sci.*, *56*(14), 2384–2402.
Grabowski, W. W., et al. (2006), Daytime convective development over land: A model intercomparison based on LBA observations, *Q. J. R. Meteorol. Soc.*, *132*(615), 317–344.
Habib, E., and W. F. Krajewski (2002), Uncertainty analysis of the TRMM ground-validation radar-rainfall products: Application to the TEFLUN-B field campaign, *J. Appl. Meteorol.*, *41*, 558–572.
Hagos, S. (2010), Building blocks of tropical diabatic heating, *J. Atmos. Sci.*, *67*(7), 2341–2354, doi:10.1175/2010JAS3252.1.
Henderson, P. W., and R. Pincus (2009), Multiyear evaluations of a cloud model using ARM data, *J. Atmos. Sci.*, *66*(9), 2925–2936.
Holder, C. T., S. E. Yuter, A. H. Sobel, and A. R. Aiyyer (2008), The mesoscale characteristics of tropical oceanic precipitation during Kelvin and mixed Rossby–gravity wave events, *Mon. Weather Rev.*, *136*(9), 3446–3464, doi:10.1175/2008MWR2350.1.
Houze, R. A., Jr. (2004), Mesoscale convective systems, *Rev. Geophys.*, *42*, RG4003, doi:10.1029/2004RG000150.
Houze, R. A., Jr., S. Brodzik, C. Schumacher, S. E. Yuter, and C. R. Williams (2004), Uncertainties in oceanic radar rain maps at Kwajalein and implications for satellite validation, *J. Appl. Meteorol.*, *43*(8), 1114–1132.
Hume, T., and C. Jakob (2007), Ensemble single column model validation in the tropical western Pacific, *J. Geophys. Res.*, *112*, D10206, doi:10.1029/2006JD008018.
Kärcher, B., and U. Lohmann (2002), A parameterization of cirrus cloud formation: Homogeneous freezing of supercooled aerosols, *J. Geophys. Res.*, *107*(D2), 4010, doi:10.1029/2001JD000470.
Kassianov, E., C. N. Long, and M. Ovtchinnikov (2005), Cloud sky cover versus cloud fraction: Whole-sky simulations and observations, *J. Appl. Meteorol.*, *44*(1), 86–98.
Keenan, T. D., K. Glasson, F. Cummings, T. S. Bird, J. Keeler, and J. Lutz (1998), The BMRC/NCAR C-Band polarimetric (C-POL) radar system, *J. Atmos. Oceanic Technol.*, *15*, 871–886.
Khairoutdinov, M. F., and D. A. Randall (2003), Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties, and sensitivities, *J. Atmos. Sci.*, *60*(4), 607–625.
Khaiyer, M. M., M. L. Nordeen, R. Palikonda, D. A. Rutan, Y. Yi, P. Minnis, and D. R. Doelling (2010), Improved TOA broadband shortwave and longwave fluxes derived from satellites over the Tropical Western Pacific, paper presented at 13th Conference on Atmospheric Radiation, Am. Meteorol. Soc., Boston, Mass.
Kiehl, J. T. (1994), On the observed near cancellation between longwave and shortwave cloud forcing in tropical regions, *J. Clim.*, *7*(4), 559–565.
Koenig, L. R., and F. W. Murray (1976), Ice-bearing cumulus cloud evolution: Numerical simulation and general comparison against observations, *J. Appl. Meteorol.*, *15*(7), 747–762.
Lafore, J. P., et al. (1998), The Meso-NH atmospheric simulation system. Part I: Adiabatic formulation and control simulations, *Ann. Geophys.*, *16*(1), 90–109.
Li, X., W. Tao, A. P. Khain, J. Simpson, and D. E. Johnson (2009), Sensitivity of a cloud-resolving model to bulk and explicit bin microphysical schemes. Part I: Comparisons, *J. Atmos. Sci.*, *66*, 3–21.
Luo, Y., Y. Wang, H. Wang, Y. Zheng, and H. Morrison (2010), Modeling convective-stratiform precipitation processes on a Mei-Yu front with the Weather Research and Forecasting Model: Comparison with observations and sensitivity to cloud microphysics parameterizations, *J. Geophys. Res.*, *115*, D18117, doi:10.1029/2010JD013873.
Lynn, B. H., A. P. Khain, J. Dudhia, D. Rosenfeld, A. Pokrovsky, and A. Seifert (2005), Spectral (bin) microphysics coupled with a mesoscale model (MM5). Part II: Simulation of a CaPE rain event with a squall line, *Mon. Weather Rev.*, *133*(1), 59–71.
Matsui, T., X. Zeng, W.-K. Tao, H. Masunaga, W. S. Olson, and S. Lang (2009), Evaluation of long-term cloud-resolving model simulations using satellite radiance observations and multifrequency satellite simulators, *J. Atmos. Oceanic Technol.*, *26*(7), 1261–1274, doi:10.1175/2008JTECHA1168.1.
May, P. T., J. H. Mather, G. Vaughan, and C. Jakob (2008), Characterizing oceanic convective cloud systems—The Tropical Warm Pool International Cloud Experiment, *Bull. Am. Meteorol. Soc.*, *154*, 153–155, doi:10.1175/BAMS-89-2-153.
Meyers, M. P., P. J. DeMott, and W. R. Cotton (1992), New primary ice-nucleation parameterizations in an explicit cloud model, *J. Appl. Meteorol.*, *31*(7), 708–721.
Minnis, P., and W. L. Smith, Jr. (1998), Cloud and radiative fields derived from GOES-8 during SUCCESS and the ARM-UAV spring 1996 flight series, *Geophys. Res. Lett.*, *25*, 1113–1116.
Minnis, P., et al. (2006), Large-scale cloud properties and radiative fluxes over Darwin during TWP-ICE, paper presented at 16th ARM Science Team Meeting, U.S. Dep. of Energy, Albuquerque, N. M.
Minnis, P., et al. (2008), Cloud detection in non-polar regions for CERES using TRMM VIRS and Terra and Aqua MODIS data, *IEEE Trans. Geosci. Remote Sens.*, *46*, 3857–3884.
Minnis, P., et al. (2011a), CERES Edition 2 cloud property retrievals using TRMM VIRS and Terra and Aqua MODIS data–Part I: Algorithms, *IEEE Trans. Geosci. Remote Sens.*, *11*, 4374–4400, doi:10.1109/TGRS.2011.2144601.
Minnis, P., et al. (2011b), CERES Edition 2 cloud property retrievals using TRMM VIRS and Terra and Aqua MODIS data–Part II: Examples of average results and comparisons with other data, *IEEE Trans. Geosci. Remote Sens.*, *11*, 4401–4430, doi:10.1109/TGRS.2011.2144602.
Moncrieff, M. W., S. K. Krueger, D. Gregory, J. Redelsperger, and W. Tao (1997), GEWEX Cloud System Study (GCSS) working group 4: Precipitating convective cloud systems, *Bull. Am. Meteorol. Soc.*, *78*(5), 831–845.
Morrison, H., and W. W. Grabowski (2007), Comparison of bulk and bin warm-rain microphysics models using a kinematic framework, *J. Atmos. Sci.*, *64*(8), 2839–2861.
Morrison, H., and W. W. Grabowski (2008a), A novel approach for representing ice microphysics in models: Description and tests using a kinematic framework, *J. Atmos. Sci.*, *65*(5), 1528–1548.
Morrison, H., and W. W. Grabowski (2008b), Modeling supersaturation and subgrid-scale mixing with two-moment bulk warm microphysics, *J. Atmos. Sci.*, *65*(3), 792–812.
Morrison, H., and W. W. Grabowski (2011), Cloud-system resolving model simulations of aerosol indirect effects on tropical deep convection and its thermodynamic environment, *Atmos. Chem. Phys. Discuss.*, *11*, 15,573–15,629.
Morrison, H., and J. Milbrandt (2011), Comparison of two-moment bulk microphysics schemes in idealized supercell thunderstorm simulations, *Mon. Weather Rev.*, *139*(4), 1103–1130, doi:10.1175/2010MWR3433.1.
Morrison, H., G. Thompson, and V. Tatarskii (2009), Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one-and two-moment schemes, *Mon. Weather Rev.*, *137*(3), 991–1007. doi:10.1175/2008MWR2556.1.
Nikolopoulos, E. I., A. Kruger, W. F. Krajewski, C. R. Williams, and K. S. Gage (2008), Comparative rainfall data analysis from two vertically pointing radars, an optical disdrometer, and a rain gauge, *Nonlinear Processes Geophys.*, *15*, 987–997.
O'Dell, C. W., F. J. Wentz, and R. Bennartz (2008), Cloud liquid water path from satellite-based passive microwave observations: A new climatology over the global oceans, *J. Clim.*, *21*(8), 1721–1739.
Petch, J. C., and J. Dudhia (1998), The importance of the horizontal advection of hydrometeors in a single-column model, *J. Clim.*, *11*(9), 2437–2452, doi:http://dx.doi.org/10.1175/1520-0442(1998)011<2437:TIOTHA>2.0.CO;2.
Petch, J. C., and M. E. B. Gray (2001), Sensitivity studies using a cloud-resolving model simulation of the tropical west Pacific, *Q. J. R. Meteorol. Soc.*, *127*(577), 2287–2306.
Petch, J. C., M. Willett, R. Y. Wong, and S. J. Woolnough (2007), Modelling suppressed and active convection. Comparing a numerical weather prediction, cloud-resolving and single-column model, *Q. J. R. Meteorol. Soc.*, *133*(626), 1087–1100.
Petch, J. C., P. N. Blossey, and C. S. Bretherton (2008), Differences in the lower troposphere in two-and three-dimensional cloud-resolving model simulations of deep convection, *Q. J. R. Meteorol. Soc.*, *134*(636), 1941–1946.

Phillips, V. T. J., and L. J. Donner (2006), Cloud microphysics, radiation and vertical velocities in two-and three-dimensional simulations of deep convection, *Q. J. R. Meteorol. Soc.*, *132*(621C), 3011–3033.

Pinty, J. (2002), A quasi 2-moment microphysical scheme for mixed-phase clouds at mesoscale with sensitivity to aerosols (CCN and IN), paper presented at 11th Conference on Cloud Physics, Am. Meteorol. Soc., Boston, Mass.

Pinty, J., and P. Jabouille (1998), A mixed-phase cloud parameterization for use in a mesoscale non-hydrostatic model: Simulations of a squall line and of orographic precipitation, paper presented at Conference on Cloud Physics, Am. Meteorol. Soc., Boston, Mass.

Randall, D., M. Khairoutdinov, A. Arakawa, and W. Grabowski (2003a), Breaking the cloud parameterization deadlock, *Bull. Am. Meteorol. Soc.*, *84*(11), 1547–1564, doi:10.1175/BAMS-84-11-1547.

Randall, D., et al. (2003b), Confronting models with data: The GEWEX Cloud Systems Study, *Bull. Am. Meteorol. Soc.*, *84*(4), 455–469, doi:10.1175/BAMS-84-4-455.

Randall, D. A., et al. (2007), Climate models and their evaluation, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 589–662, Cambridge Univ. Press, Cambridge, U. K.

Redelsperger, J. L., et al. (2000), A GCSS model intercomparison for a tropical squall line observed during TOGA-COARE. Part I: CRM results, *Q. J. R. Meteorol. Soc.*, *126*, 823–863.

Rickenbach, T. M., and S. A. Rutledge (1998), Convection in TOGA COARE: Horizontal scale, morphology, and rainfall production, *J. Atmos. Sci.*, *55*(17), 2715–2729.

Schumacher, C., and R. A. Houze Jr. (2003), Stratiform rain in the tropics as seen by the TRMM precipitation radar, *J. Clim.*, *16*(11), 1739–1756, doi:10.1175/1520-0442(2003)016<1739:SRITTA>2.0.CO;2.

Schumacher, C., R. A. Houze Jr., and I. Kraucunas (2004), The tropical dynamical response to latent heating estimates derived from the TRMM Precipitation Radar, *J. Atmos. Sci.*, *61*, 1341–1358.

Seo, E. K., and G. Liu (2005), Retrievals of cloud ice water path by combining ground cloud radar and satellite high-frequency microwave measurements near the ARM SGP site, *J. Geophys. Res.*, *110*, D14203, doi:10.1029/2004JD005727.

Seo, E. K., and G. Liu (2006), Determination of 3D cloud ice water contents by combining multiple data sources from satellite, ground radar, and a numerical model, *J. Appl. Meteorol. Climatol.*, *45*(11), 1494–1504.

Shutts, G. J., and M. E. B. Gray (1994), A numerical modelling study of the geostrophic adjustment process following deep convection, *Q. J. R. Meteorol. Soc.*, *120*, 1145–1178.

Smolarkiewicz, P. K., and L. G. Margolin (1997), On forward-in-time differencing for fluids: An Eulerian/semi-Lagrangian non-hydrostatic model for stratified flows, in *Numerical Methods in Atmospheric and Oceanic Modelling: The André J. Robert Memorial Volume*, edited by C. A. Lin, R. Laprise, and H. Ritchie, pp. 127–152, NRC Res. Press, Ottawa, Ont., Canada.

Song, X., and G. J. Zhang (2011), Microphysics parameterization for convective clouds in a global climate model: Description and single-column model tests, *J. Geophys. Res.*, *116*, D02201, doi:10.1029/2010JD014833.

Steiner, M., R. A. Houze, Jr., and S. E. Yuter (1995), Climatological characterization of three-dimensional storm structure from operational radar and rain gauge data, *J. Appl. Meteorol.*, *34*(9), 1978–2007.

Stephens, G. L., T. L'Ecuyer, R. Forbes, A. Gettlemen, J. C. Golaz, A. Bodas-Salcedo, K. Suzuki, P. Gabriel, and J. Haynes (2010), Dreary state of precipitation in global models, *J. Geophys. Res.*, *115*, D24211, doi:10.1029/2010JD014532.

Stevens, D. E., A. S. Ackerman, and C. S. Bretherton (2002), Effects of domain size and numerical resolution on the simulation of shallow cumulus convection, *J. Atmos. Sci.*, *59*, 3285–3301.

Su, H., J. H. Jiang, J. Teixeira, A. Gettelman, X. Huang, G. Stephens, D. Vane, and V. S. Perun (2011), Comparison of regime-sorted tropical cloud profiles observed by CloudSat with GEOS5 analyses and two general circulation model simulations, *J. Geophys. Res.*, *116*, D09104, doi:10.1029/2010JD014971.

Thompson, G., R. M. Rasmussen, and K. Manning (2004), Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis, *Mon. Weather Rev.*, *132*, 519–542.

Trenberth, K. E., J. T. Fasullo, and J. Kiehl (2009), Earth's global energy budget, *Bull. Am. Meteorol. Soc.*, *90*(3), 311–323, doi:10.1175/2008BAMS2634.1.

Turner, D. D., S. A. Clough, J. C. Liljegren, E. E. Clothiaux, K. Cady-Pereira, and K. L. Gaustad (2007), Retrieving liquid water path and precipitable water vapor from Atmospheric Radiation Measurement (ARM) microwave radiometers, *IEEE Trans. Geosci. Remote Sens.*, *45*, 3680–3690, doi:10.1109/TGRS.2007.903703.

Varble, A., A. Fridlind, E. Zipser, A. Ackerman, J.-P. Chaboureau, J. Fan, A. Hill, S. McFarlane, J.-P. Pinty, and B. Shipway (2011), Evaluation of cloud-resolving model intercomparison simulations using TWP-ICE observations: Precipitation and cloud structure, *J. Geophys. Res.*, *116*, D12206, doi:10.1029/2010JD015180.

Vaughan, G., C. Schiller, A. R. MacKenzie, K. N. Bower, T. Peter, H. Schlager, N. R. P. Harris, and P. T. May (2008), SCOUT-O3/ACTIVE: High-altitude aircraft measurements around deep tropical convection, *Bull. Am. Meteorol. Soc.*, *89*, 647–662, doi:10.1175/BAMS-89-5-647.

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.

Wang, W., and X. Liu (2009), Evaluating deep updraft formulation in NCAR CAM3 with high-resolution WRF simulations during ARM TWP-ICE, *Geophys. Res. Lett.*, *36*, L04701, doi:10.1029/2008GL036692.

Wang, W., X. Liu, S. Xie, J. Boyle, and S. A. McFarlane (2009a), Testing ice microphysics parameterizations in the NCAR Community Atmospheric Model Version 3 using Tropical Warm Pool–International Cloud Experiment data, *J. Geophys. Res.*, *114*, D14107, doi:10.1029/2008JD011220.

Wang, Y., C. N. Long, L. R. Leung, J. Dudhia, S. A. McFarlane, J. H. Mather, S. J. Ghan, and X. Liu (2009b), Evaluating regional cloud-permitting simulations of the WRF model for the Tropical Warm Pool International Cloud Experiment (TWP-ICE), Darwin, 2006, *J. Geophys. Res.*, *114*, D21203, doi:10.1029/2009JD012729.

Wapler, K., T. P. Lane, P. T. May, C. Jakob, M. J. Manton, and S. T. Siems (2010), Cloud system resolving model simulations of tropical cloud systems observed during the Tropical Warm Pool-International Cloud Experiment, *Mon. Weather Rev.*, *138*, 55–73, doi:10.1175/2009MWR2993.1.

Willett, M. R., P. Bechtold, D. L. Williamson, J. C. Petch, S. F. Milton, and S. J. Woolnough (2008), Modelling suppressed and active convection: Comparisons between three global atmospheric models, *Q. J. R. Meteorol. Soc.*, *134*(636), 1881–1896.

Woolnough, S. J., et al. (2010), Modelling convective processes during the suppressed phase of a Madden-Julian oscillation: Comparing single-column models with cloud-resolving models, *Q. J. R. Meteorol. Soc.*, *136*, 333–353.

Wu, J., A. D. Del Genio, M. S. Yao, and A. B. Wolf (2009), WRF and GISS SCM simulations of convective updraft properties during TWP-ICE, *J. Geophys. Res.*, *114*, D04206, doi:10.1029/2008JD010851.

Wu, X., and M. Moncrieff (2001a), Long-term behavior of cloud systems in TOGA COARE and their interactions with radiative and surface processes. Part III: Effects on the energy budget and SST, *J. Atmos. Sci.*, *58*(9), 1155–1168.

Wu, X., and M. W. Moncrieff (2001b), Sensitivity of single-column model solutions to convective parameterizations and initial conditions, *J. Clim.*, *14*(12), 2563–2582.

Wu, X., W. Grabowski, and M. Moncrieff (1998), Long-term behavior of cloud systems in TOGA COARE and their interactions with radiative and surface processes. Part I: Two-dimensional modeling study, *J. Atmos. Sci.*, *55*(17), 2693–2714.

Wu, X. Q., W. D. Hall, W. Grabowski, M. Moncrieff, W. Collins, and J. Kiehl (1999), Long-term behavior of cloud systems in TOGA COARE and their interactions with radiative and surface processes. Part II: Effects of ice microphysics on cloud-radiation interaction, *J. Atmos. Sci.*, *56*(18), 3177–3195.

Wu, X., S. Park, and Q. Min (2008), Seasonal variation of cloud systems over ARM SGP, *J. Atmos. Sci.*, *65*(7), 2107–2129.

Xie, S., et al. (2002), Intercomparison and evaluation of cumulus parametrizations under summertime midlatitude continental conditions, *Q. J. R. Meteorol. Soc.*, *128*(582), 1095–1135.

Xie, S., et al. (2005), Simulations of midlatitude frontal clouds by single-column and cloud-resolving models during the Atmospheric Radiation Measurement March 2000 cloud intensive operational period, *J. Geophys. Res.*, *110*, D15S03, doi:10.1029/2004JD005119.

Xie, S., T. Hume, C. Jakob, S. A. Klein, R. B. McCoy, and M. Zhang (2010), Observed large-scale structures and diabatic heating and drying profiles during TWP-ICE, *J. Clim.*, *23*, 57–79.

Xu, K. M., et al. (2002), An intercomparison of cloud-resolving models with the Atmospheric Radiation Measurement summer 1997 Intensive Observation Period data, *Q. J. R. Meteorol. Soc.*, *128*(580), 593–624.

Young, K. C. (1974), A numerical simulation of wintertime, orographic precipitation: Part I. Description of model microphysics and numerical techniques, *J. Atmos. Sci.*, *31*(7), 1735–1748.

Yuter, S. E., R. A. Houze, Jr., E. A. Smith, T. T. Wilheit, and E. Zipser (2005), Physical characterization of tropical oceanic convection observed in KWAJEX, *J. Appl. Meteorol.*, *44*(4), 385–415.

Zeng, X., W. Tao, S. Lang, A. Y. Hou, and M. Zhang (2008), On the sensitivity of atmospheric ensembles to cloud microphysics in long-term cloud-resolving model simulations, *J. Meteorol. Soc. Jpn.*, *86A*, 45–65.

Zeng, X., W. Tao, T. Matsui, S. Xie, S. Lang, M. Zhang, D. O. Starr, and X. Li (2011), Estimating the ice crystal enhancement factor in the tropics, *J. Atmos. Sci.*, *68*, 1424–1434.

Zhang, G. J. (2009), Effects of entrainment on convective available potential energy and closure assumptions in convection parameterization, *J. Geophys. Res.*, *114*, D07109, doi:10.1029/2008JD010976.

A. S. Ackerman and A. M. Fridlind, NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA. (ann.fridlind@nasa.gov)

J.-P. Chaboureau and J.-P. Pinty, Laboratoire d'Aerologie, University of Toulouse/CNRS, 14, Avenue Belin, F-31400 Toulouse, France.

J. Fan, Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, PO Box 999, MSIN k9-24, Richland, WA 99352, USA.

W. W. Grabowski and H. Morrison, Mesoscale and Microscale Meteorology, National Center for Atmospheric Research, PO Box 3000, Boulder, CO 80307-3000, USA.

A. A. Hill, J. C. Petch, and B. J. Shipway, Met Office, Fitzroy Road, Exeter EX1 3PB, UK.

T. R. Jones, Department of Atmospheric Science, Colorado State University, 1371 Campus Delivery, Fort Collins, CO 80523-1371, USA.

M. M. Khaiyer, Science Systems, and Applications, Inc., 1 Enterprise Pkwy., Hampton, VA 23666, USA.

G. Liu and X. Wu, Department of Earth, Ocean and Atmospheric Science, Florida State University, 404 Love Bldg., Tallahassee, FL 32306, USA.

P. Minnis and L. Nguyen, NASA Langley Research Center, MS 420, 21 Langley Blvd., Hampton, VA 23681-0001, USA.

S. Park, Department of Earth and Atmospheric Sciences, University of Nebraska, Lincoln, NE 66588, USA.

C. Schumacher, Department of Atmospheric Sciences, Texas A&M University, 3150 TAMU, College Station, TX 77840-3150, USA.

A. C. Varble, Department of Atmospheric Sciences, University of Utah, 135 S 1460 E, Rm. 819 (WBB), Salt Lake City, UT 84112-0110, USA.

S. Xie, Lawrence Livermore National Laboratory, PO Box 808, Livermore, CA 94550, USA.

M. Zhang, Institute for Planetary and Terrestrial Atmospheres, State University of New York at Stony Brook, Stony Brook, NY 11794-5000, USA.