A single-column model ensemble approach applied to the TWP-ICE experiment

L. Davies,[1] C. Jakob,[2] K. Cheung,[3] A. Del Genio,[4] A. Hill,[5] T. Hume,[6] R. J. Keane,[7] T. Komori,[8] V. E. Larson,[9] Y. Lin,[10,11] X. Liu,[12] B. J. Nielsen,[9] J. Petch,[5] R. S. Plant,[13] M. S. Singh,[14] X. Shi,[12] X. Song,[15] W. Wang,[16] M. A. Whitall,[13] A. Wolf,[17] S. Xie,[18] and G. Zhang[15]

Received 25 October 2012; revised 23 April 2013; accepted 24 April 2013; published 24 June 2013.

[1] Single-column models (SCM) are useful test beds for investigating the parameterization schemes of numerical weather prediction and climate models. The usefulness of SCM simulations are limited, however, by the accuracy of the best estimate large-scale observations prescribed. Errors estimating the observations will result in uncertainty in modeled simulations. One method to address the modeled uncertainty is to simulate an ensemble where the ensemble members span observational uncertainty. This study first derives an ensemble of large-scale data for the Tropical Warm Pool International Cloud Experiment (TWP-ICE) based on an estimate of a possible source of error in the best estimate product. These data are then used to carry out simulations with 11 SCM and two cloud-resolving models (CRM). Best estimate simulations are also performed. All models show that moisture-related variables are close to observations and there are limited differences between the best estimate and ensemble mean values. The models, however, show different sensitivities to changes in the forcing particularly when weakly forced. The ensemble simulations highlight important differences in the surface evaporation term of the moisture budget between the SCM and CRM. Differences are also apparent between the models in the ensemble mean vertical structure of cloud variables, while for each model, cloud properties are relatively insensitive to forcing. The ensemble is further used to investigate cloud variables and precipitation and identifies differences between CRM and SCM particularly for relationships involving ice. This study highlights the additional analysis that can be performed using ensemble simulations and hence enables a more complete model investigation compared to using the more traditional single best estimate simulation only.

Citation: Davies, L., et al. (2013), A single-column model ensemble approach applied to the TWP-ICE experiment, *J. Geophys. Res. Atmos.*, *118*, 6544–6563, doi:10.1002/jgrd.50450.

1. Introduction

[2] The Tropical Warm Pool International Cloud Experiment (TWP-ICE) took place around Darwin from 20 January to 13 February 2006 [*May et al.*, 2008]. The data collected during the experiment provides an opportunity to investigate several different states of tropical convection. The

[1]School of Mathematics, Monash University, Melbourne, Victoria, Australia.
[2]ARC Centre of Excellence for Climate System Science, Monash University, Melbourne, Victoria, Australia.
[3]Bureau of Meteorology, Melbourne, Victoria, Australia.
[4]NASA Goddard Institute for Space Studies, New York, New York, USA.
[5]Met Office, Exeter, UK.
[6]Centre for Australian Weather and Climate Research, Bureau of Meteorology, Melbourne, Victoria, Australia.
[7]Meteorologisches Institut, Ludwig-Maximilians-Universitat Munchen, Munich, Germany.
[8]Japan Meteorological Agency, Tokyo, Japan.
[9]University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, USA.
[10]University Corporation for Atmospheric Research, Boulder, Colorado, USA.
[11]NOAA Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, USA.
[12]Pacific Northwestern National Laboratory, Richland, Washington, USA.
[13]Department of Meteorology, University of Reading, Reading, UK.
[14]Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.
[15]University of California, San Diego, California, USA.
[16]IMSG@NOAA National Centers for Environmental Prediction, College Park, Maryland, USA.
[17]Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York, USA.
[18]Lawrence Livermore National Laboratory, Livermore, California, USA.

Corresponding author: L. Davies, School of Earth Sciences, The University of Melbourne, Melbourne 3010, Victoria, Australia. (laura.davies@unimelb.edu.au)

©2013. American Geophysical Union. All Rights Reserved.
2169-897X/13/10.1002/jgrd.50450

experiment collected sufficient information to derive both the large-scale heat and momentum and moisture budgets [*Xie et al.*, 2004] as well as detailed information on the state of the smaller scale convection and associated clouds. Such data sets are commonly used in the modeling community to carry out process-oriented studies in particular applying cloud-resolving models (CRM) and single-column models (SCM). One of the primary motivations for TWP-ICE was to enable the improvement of global climate models (GCM), which are known to be deficient in the representation of cloud and rainfall particularly associated with tropical convection. The international research community has conducted a suite of multimodel studies for TWP-ICE. A hierarchy of experiments enables the investigation of model errors as discussed in J. Petch et al. (Evaluation of intercomparisons of four different types of model simulating TWP-ICE, submitted to *Quarterly Journal of the Royal Meteorological Society*, 2012) and includes GCM [*Lin et al.*, 2012] and Limited Area Models [*Zhu et al.*, 2012] forced with European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis as well as a CRM study [*Fridlind et al.*, 2012] performing simulations driven by a single "best estimate" large-scale budget data set [*Xie et al.*, 2010]. This paper reports on the SCM component of the overall modeling strategy. One innovation applied here will be the use of an ensemble of SCM simulations to elucidate uncertainties in the estimation of model errors and to explore model sensitivities to changes in the data set driving the model simulations.

[3] The investigation of model shortcomings through SCMs is a well-used method in the model development research community. Model development studies, which include a SCM component, have been instigated by the Global Energy and Water-Cycle Experiment (GEWEX) Cloud System Study (GCSS) [*Randall et al.*, 2003] in conjunction with the U.S. Department of Energy Atmospheric System Research (ASR) to investigate a wide range of test cases including deep convection over the tropical ocean using data from the Tropical Ocean Global Atmosphere (TOGA) Coupled Ocean-Atmosphere Response Experiment (COARE) [*Webster and Lukas*, 1992] intensive observation period [e.g., *Woolnough et al.*, 2010; *Bechtold et al.*, 2000] and convection over land exploiting extensive observations [e.g., *Grabowski et al.*, 2006; *Xie et al.*, 2005, 2002; *Ghan et al.*, 2000]. Investigation of the specific problem of the diurnal cycle was conducted by the European Cloud Systems (EUROCS) project and discussed in *Guichard et al.* [2004]. These studies focussed on a limited number of model simulations forced by a single data set, from hereon referred to as the "best estimate" forcing. While best estimates of the large-scale atmosphere are usually derived to depict the most probable state of the large-scale atmosphere, they do contain errors of usually unknown magnitude. These errors complicate the interpretation of the results of SCM simulations, as the discrepancies between the model-simulated fields and observations may be attributed to two sources, from prescribing an incorrect large-scale state or due to errors in model processes. By using a single-model realization of the large-scale state, it is impossible to separate these two error sources.

[4] Ensemble techniques are commonly used in numerical weather prediction (NWP) and climate models to investigate model sensitivities and to determine uncertainty. These ensembles may include perturbed initial conditions or varying model parameters within a limited range. Multimodel ensembles have also been used to provide an estimate of the range of simulations. A limited number of studies also derived ensemble techniques for use in SCM studies. *Hack and Pedretti* [2000] added random perturbations to the initial conditions of their ensemble simulations and found considerable variations in simulated fields. Similar results were found when modifying the prescribed vertical motion field in a similar manner. Given the bifurcations discussed in *Hack and Pedretti* [2000], *Hume and Jakob* [2005, 2007] and *Ball and Plant* [2008] determined that an ensemble technique was appropriate for SCM. *Hume and Jakob* [2005] found that after about 18 h of simulation, results were increasingly sensitive to the prescribed forcing rather than differences in the initial conditions. For this reason, this TWP-ICE study uses an ensemble of large-scale forcing.

[5] The goal of this study is to apply an ensemble SCM technique to the TWP-ICE experiment and to highlight additional opportunities for model evaluation that such a technique may provide. The technique is applied to a wide range of SCMs as well as a small number of CRMs, enabling the investigation of a range of model behaviors. The results from the ensemble simulation will be compared to those of single "best estimate" simulations. It will be shown that a particularly interesting aspect of the use of the ensemble technique in this context is the possibility to study model sensitivities with changing forcing data set. It is shown that the different models exhibit distinctly different ensemble behavior that is not apparent when comparing simulations with a single forcing data set. Section 2 summarizes the experimental design including the methodology used in the derivation of the ensemble large-scale forcing, the case specification, and a description of the models. The main results of the study are discussed in section 3 followed by a summary and the main conclusions in section 4.

2. Experimental Design

[6] The experiments conducted here use both a best estimate forcing and an ensemble of forcing data sets. The best estimate data set used is that derived by *Xie et al.* [2010] and is identical to that used in *Fridlind et al.* [2012]. Using an ensemble approach enables a better understanding of model accuracy and model sensitivity to be calculated. As this study includes a number of different models, these characteristics are determined dependent on model. This section will detail the design of the study including the ensemble forcing design, the case specification, and the participating models.

2.1. Ensemble Design

[7] A number of techniques exist to derive budgets from observational data collected in field campaigns. Here, the variational analysis technique of *Zhang and Lin* [1997] is used in the analysis of TWP-ICE observations. This technique provides an estimate of area-averaged atmospheric and surface conditions using a combination of surface observations, vertical profiles of the atmosphere, satellite observations, and numerical model data. The variational analysis process minimizes a cost function for the heat, moisture, and

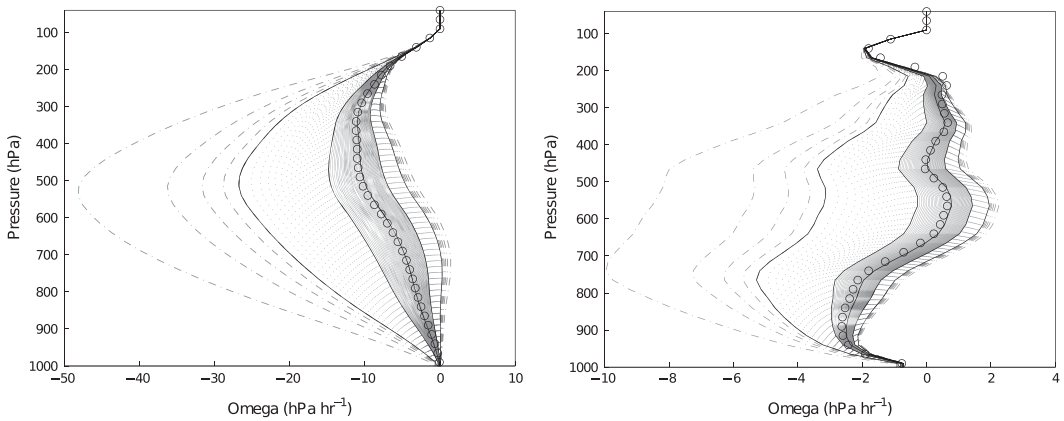

Figure 1. Time-averaged vertical profiles of omega over active (left) and suppressed (right) monsoon for all ensemble members. Broken and light-colored lines show all ensemble members with key ensemble members (5th, 25th, 50th, 75th, and 95th percentiles) as black continuous lines. The best estimate forcing, used here and in the CRM intercomparison, is shown by small circles. Note the different *x* axis.

momentum budgets using constraints of top of atmosphere and surface energy and moisture.

[8] One of the constraints used in the variational analysis method is the domain-average surface rainfall. In the case of the TWP-ICE experiment, this domain-mean surface rainfall is derived from radar data. Compared to the use of rain gauge observations, this improves the spatial representativeness of the estimate, but this comes at the expense of accuracy of the local rainfall estimates as radar measurements need to be converted to rainfall. It has been shown [*Zhang et al.*, 2001] that the surface rainfall has a large effect on the derived forcing data set; for example, the analyzed vertical velocity is very sensitive to rainfall. Furthermore, the derivation of surface rainfall from radar data is also highly complex and liable to large errors [*Joss and Waldvogel*, 1990]. These errors will have a large effect on the derived forcing data set.

[9] One method to address uncertainty in large-scale forcing data is to derive an ensemble of forcing data. Only a short summary of the method to derive such an ensemble is given here, with more details provided in the Appendix. The method is principally based on estimates of errors in the rainfall estimates that are a key input to the variational budget analysis. A comparison of radar-derived and rain gauge data is carried out to provide an estimate of the error in the radar estimates of domain-average rainfall. From these error estimates, 100 equally likely alternative domain-mean rainfall time series are calculated. The 100 rainfall time series are then used as inputs to the variational analysis to derive 100 alternative versions of the large-scale state using the same variational technique as is used to derive the best estimate large-scale state. These 100 large-scale states constitute the forcing ensemble used in this study.

[10] When deriving the large-scale state using these alternate rainfall time series, all other observations have the same values as the best estimate, for example, temperature, moisture, and horizontal wind fields. Given that the boundary values of temperature and moisture are identical between all realizations, the horizontal advection terms of temperature and moisture differ very little. The variational analysis process generally equates larger values of rainfall with increased low-level convergence and upper level divergence and therefore generally larger values of vertical velocity. The structure of the derived vertical velocity, however, is also dependent on other budget terms so that vertical velocity does not monotonically increase with rainfall.

[11] Figure 1 shows the vertical velocity profile averaged over both the active and suppressed monsoon for each ensemble member as well as key percentiles of the ensemble. Stronger vertical motion is derived from time series with larger rainfall. In the active monsoon, there is always strong upward vertical motion, although the ensemble members with weaker rainfall have weaker vertical motion. During the suppressed monsoon, the ensemble members with strong rainfall have upward vertical motion at all levels. The ensemble members with weaker rainfall have upward motion at lower levels (below 650 hPa) but downward motion above. In addition to the ensemble members, Figure 1 shows the standard best estimate for vertical velocity. As is evident, the best estimate results are close, but not identical, to the 50th percentile of the ensemble forcing. While there is a large spread in omega, it is worth noting that 50% of the ensemble members lie in the limited range between the 25th and 75th percentile lines. While each ensemble member is equally likely, most cluster around the 50th ensemble member and the best estimate, and the most extreme omega values are rare. These differences in omega imply changes in low-level convergence and upper level divergence through the continuity equation and will have an effect on convection. These 100 large-scale "forcing" data sets are then used as input to the model simulations discussed below.

2.2. Case Description

[12] The TWP-ICE experiment experienced a range of atmospheric conditions. At the start of the experiment, the region experienced monsoon conditions. Between 23 and 24 January, a strong mesoscale convective system (MCS) passed through the domain followed by relatively suppressed conditions. There were then clear conditions from 3 to 5 February with little rain followed by monsoon break conditions to the end of the field campaign. Full details of the meteorological conditions can be found in *May et al.* [2008]. In this study, the focus is on the active period defined as 20 00Z–25 12Z Jan and the suppressed period defined as

Table 1. Models Contributing to SCM Study[a]

Model	Full Name	Modeler	Reference
UM-GR	Unified Model-Gregory and Rowntree	M. Whitall/R. Plant	*Davies et al.* [2005]
UM-PC	Unified Model-Plant/Craig	R. Keane/R. Plant	*Davies et al.* [2005]
SCAMS	Single-Column Community Atmospheric Model	X. Liu/X. Shi	*Collins et al.* [2006]
SCAML	Single-Column Community Atmospheric Model	X. Liu/X. Shi	*Wang et al.* [2009]
SCAMR	Single-Column Community Atmospheric Model	X. Song/G. Zhang	*Collins et al.* [2006]
NCEPG	NCEP GFS Model	W. Wang	*EMC* [2003]
GFDL2	GFDL-AM2 Model	Y. Lin	*GAMDT* [2004]
GISS	GISS Model	A. Wolf/A. DelGenio	*Schmidt et al.* [2006]
CLUBB	Cloud Layers Unified by Binormals model	B. Nielsen/V. Larson	*Golaz et al.* [2002]
JMA1	Japan Meteorological Agency	T. Komori	*JMA* [2007]; *Nakagawa* [2009]
JMA2	Japan Meteorological Agency	T. Komori	*JMA* [2007]; *Nakagawa* [2009]
2-D LEM	UK Met Office Large Eddy Model	A. Hill	*Gray et al.* [2001]
3-D SAM	System for Atmospheric Modeling	L. Davies	*Khairoutdinov and Randall* [2003]

[a]Includes the acronym used in this paper, the full model name, contributing author(s), and the main reference for the model. Further model details are given in the text and the references therein. Note that there are two cloud-resolving models as part of this study.

28 00Z Jan–2 12Z Feb. The conditions during the clear and break periods are dominated by a strong diurnal cycle, which is driven by the land-sea contrast in the experiment domain. As SCMs cannot usually represent such contrasts in a single grid box, the later part of the experiment is excluded from the simulations presented here.

[13] In order to investigate the performance of the ensemble technique proposed here in different meteorological conditions, the study applies two sets of large-scale forcing data. The first is a best estimate simulation forced using the standard data set [*Xie et al.*, 2010]. These simulations can be directly compared to the CRM results [*Fridlind et al.*, 2012], and the best estimate simulations also form the basis of discussion in J. Petch et al. (submitted manuscript, 2012). In this study, the best estimate simulations will be used to form a SCM multimodel ensemble. In addition to the best estimate simulations, all models were run using the 100-member ensemble of forcing data derived above. It was found that some models showed numerical instabilities for the strongest forcing data sets (i.e., those derived from the largest rainfall) when using their standard time-stepping. As a consequence, the 10 strongest forcing data sets and, for reasons of symmetry, the 10 weakest ones are excluded from further analysis, reducing the ensemble size to 80.

[14] The aim when defining the model specification is to impinge as little as possible on the inherent characteristics of the individual models, and modelers are encouraged to use their preferred configurations; however, the following requirements are made for all simulations:

[15] 1. The TWP-ICE domain has mixed surface types making the choice of surface type unclear. All simulations assume an ocean surface consistent with *Fridlind et al.* [2012]. Fixed time-invariant SST = 29°C is used. Interactive surface fluxes are required to be calculated in the boundary layer scheme.

[16] 2. Simulations are initialized with observed temperature and moisture profiles at 0300Z 19 January 2006.

[17] 3. An observed ozone profile is used where possible, but the McClatchey ozone profile [*McClatchey et al.*, 1972] is used above the maximum height of observations (40 mbar).

[18] 4. Full interactive radiation is used with a diurnal cycle for a domain centered on the Atmospheric Radiation Measurement (ARM) site (12.425°S, 130.891°E).

[19] 5. Mean horizontal winds are relaxed to observed profiles with a 2 h time scale. There is no nudging of the temperature and moisture fields which are left free to respond to the forcing.

[20] 6. Horizontal advective tendencies for temperature and moisture are prescribed, but the vertical terms are calculated by the models. Sensitivity studies showed a warm temperature bias above the tropopause when prescribing a total forcing as the model cannot freely evolve vertical advection associated with this warming and reduce the temperature bias. This method differs from *Fridlind et al.* [2012] where temperature and moisture were nudged towards observed profiles to avoid such temperature biases.

2.3. Models

[21] In this section, a brief description of all models used in this study will be given. Table 1 gives a summary of the models with further details given below. The study also includes two CRM which also simulate the ensemble. The CRM provide an important reference for the SCM and link to the CRM study [*Fridlind et al.*, 2012].

[22] The UK Met Office SCM [*Davies et al.*, 2005] contains parameterizations for radiation [*Edwards and Slingo*, 1996], layer-cloud microphysics [*Wilson and Forbes*, 2004; *Wilson and Ballard*, 1999], boundary layer processes [*Lock et al.*, 2000], and convection; see also *Martin et al.* [2006]. Results were submitted for both the default UM convection scheme [*Gregory and Rowntree*, 1990; *Martin et al.*, 2006; *Derbyshire et al.*, 2011] (UM-GR) and the *Plant and Craig* [2008] stochastic spectral mass-flux scheme (UM-PC). In the default scheme, convection is triggered by instability of surface parcels at the lifting condensation level (LCL); a CAPE closure is used for deep convection, and the closure for shallow convection is based on *Grant* [2001]. In the Plant-Craig scheme, convection is triggered by constructing potential updraft source layers and evaluating their buoyancy at the LCL; a CAPE closure is used. The stochastic variability of the Plant-Craig scheme depends upon the column size—an area of $(50\ \mathrm{km})^2$ was used here.

[23] The single-column model of the NCAR CAM3 (SCAM) contains the radiation scheme as described in *Collins et al.* [2006]. The treatment of cloud condensation and microphysics in CAM3 [*Boville et al.*, 2006] is based on *Rasch and Kristjánsson* [1998] as updated by

Zhang et al. [2003] with separate prognostic equations for the liquid and ice-phase condensate. The boundary layer scheme is based on *Holtslag and Boville* [1993] and *Boville et al.* [2006]. CAM3 includes the convection scheme of *Zhang and McFarlane* [1995] with CAPE closure. CAM3-Liu (SCAML) [*Wang et al.*, 2009] differs from SCAM with modification for cloud microphysics by introducing a double-moment cloud microphysics [*Liu et al.*, 2007], explicit treatment of ice nucleation [*Liu and Penner*, 2005], and water vapor deposition on ice crystals and Bergeron-Findeisen process in pure ice and mixed-phase clouds. SCAMR differs most fundamentally from SCAMS as the deep convection parameterization is replaced by the revised *Zhang and McFarlane* [1995] scheme proposed by *Zhang* [2002]. The new convection scheme uses CAPE changes due to large-scale forcing (e.g., large-scale advection, radiative cooling) in the free troposphere, instead of CAPE itself, for closure.

[24] In the NCEP GFS model, the longwave radiation scheme follows *Fels and Schwarzkopf* [1975] and *Schwarzkopf and Fels* [1991]. The shortwave radiation formulation uses multiband techniques [*Slingo*, 1989; *Chou et al.*, 1998; *Kiehl et al.*, 1998]. The cloud condensate is prognosed from a single-moment microphysics scheme [*Zhao and Carr*, 1997]. The boundary layer parameterization uses a nonlocal scheme [*Hong and Pan*, 1996]. Penetrative convection scheme [*Pan and Wu*, 1995] is simplified from *Arakawa and Schubert* [1974], with a quasi-equilibrium assumption as a closure. Convection is trigged when a cloud work function exceeds a threshold. Shallow convection is parameterized as an extension of the vertical diffusion scheme [*Tiedtke*, 1983].

[25] The GFDL AM2 uses the shortwave radiation algorithm of *Freidenreich and Ramaswamy* [1999], and the longwave radiation follows *Schwarzkopf and Ramaswamy* [1999]. It uses *Slingo* [1989] and *Held et al.* [1993] for liquid cloud radiative properties and *Fu and Liou* [1993] for ice clouds. The microphysics scheme uses *Rotstayn* [1979] with cloud fraction prognosed following *Tiedtke* [1993]. The microphysics used for convective clouds is rather crude with prescribed precipitation efficiencies for shallow and deep convections. Boundary layer scheme follows *Lock et al.* [2000]. GFDL uses the relaxed Arakawa-Schubert scheme [*Moorthi and Suarez*, 1992] with a CAPE closure for shallow and deep convection.

[26] The GISS SCM used here is a developmental update of the *Schmidt et al.* [2006] model. Radiation uses explicit multiple scattering calculations and the k-distribution approach to absorption. Large-scale clouds are based on the prognostic cloud water parameterization of *Del Genio et al.* [1996], including all relevant microphysical processes, detrainment, and cloud top entrainment. Convective microphysics follows *Del Genio et al.* [2005], which interactively partitions condensate into precipitating, detrained, and vertically advected components. The boundary layer uses dry conserved variables and includes local (diffusive) and counter-gradient flux terms. Moist convection is parameterized using a mass-flux scheme with convection triggered when a lifted parcel becomes buoyant. The mass flux is that required to produce neutral buoyancy at cloud base, with updraft speeds and entrainment rates based on *Gregory* [2001].

[27] The CLUBB model, in these TWP-ICE simulations, is used in conjunction with the BUGSrad radiative transfer scheme [*Stephens et al.*, 2001] and a single microphysics scheme [*Morrison et al.*, 2009] for all clouds. Although in the prior literature CLUBB was tested only for boundary layer cloud cases [*Golaz et al.*, 2002; *Larson and Golaz*, 2005; *Larson et al.*, 2012], here CLUBB is used to simulate both deep and shallow clouds with a single, unified equation set. Unlike *Larson et al.* [2012], here CLUBB is run as a single-column model and handles all cloud types without the use of a cloud-resolving model or any other host model. CLUBB prognoses various higher-order moments and achieves closure by use of a single multivariate subgrid PDF of velocity, moisture, and temperature. CLUBB has no explicit convective trigger; rather, the turbulence and thermodynamic variability generated in shallow convection are intended to evolve into deep convection when and where the large-scale forcings are appropriate.

[28] The single-column model JMA1 contains the parameterizations of the default Global Spectral Model [*JMA*, 2007; *Nakagawa*, 2009]. The radiation scheme has two-stream with delta-Eddington approximation for shortwave and table look-up and k-distribution methods for longwave. Cloud condensation and microphysics are based on *Smith* [1990] and *Sundqvist et al.* [1989]. The boundary layer scheme is the level 2 closure scheme of *Mellor and Yamada* [1974]. The convection scheme is a multiplume type with cloud work function closure based on *Arakawa and Schubert* [1974], two types (for shallow and deep convection) of prognostic equations of the upward mass-flux [*Randall and Pan*, 1993] and triggering functions [*Xie and Zhang*, 2000]. JMA2 is the same as JMA1, except for using modified convection and cloud schemes (T. Komori and K. Yoshimoto, Evaluation from a perspective of spin-down problem: Moistening effect of convective parameterization, submitted to *CAS/JSC WGNE Research Activities in Atmospheric and Oceanic Modeling*, 2012).

[29] There are two CRM in the study which are briefly described here. The UKMO Large Eddy simulation model (LEM) uses the shortwave and longwave radiation scheme of *Edwards and Slingo* [1996]. The LEM employs a three-phase microphysics scheme, which is described in *Gray et al.* [2001], and the microphysical configuration is the same as the UKMO-2A setup described in *Fridlind et al.* [2012]. The subgrid mixing scheme is a modified first-order Smagorinksky-Lilly scheme, which is described in *MacVean and Mason* [1990].

[30] The model used in the System for Atmospheric Modeling (SAM) is described by *Khairoutdinov and Randall* [2003] and uses the BUGSrad radiation scheme described in *Stephens et al.* [2001]. Single-moment microphysics were used as outlined in *Khairoutdinov and Randall* [2003]. The subgrid mixing scheme is a 1.5-order closure model [*Khairoutdinov and Kogan*, 1999]. The SAM model simulates nine ensemble members equally spaced in the range 10–90.

3. Results

3.1. Simulations of Humidity and Precipitation

[31] This section gives an overview of both the temporal evolution and the vertical structure of the simulation

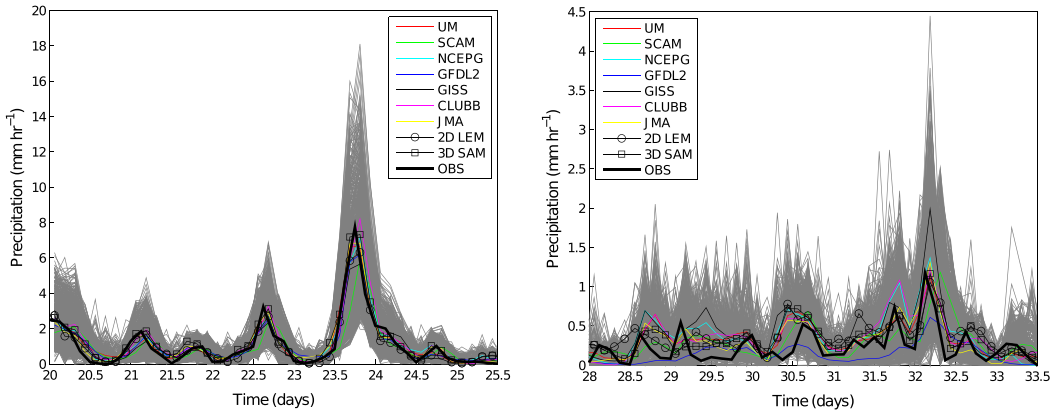

Figure 2. Time series of precipitation for the active period (left) and suppressed period (right) for each model type. Colored lines show the average for each model type (e.g., all UM SCM, all SCAM, and all JMA models are averaged together) and gray lines the 80-member ensemble for all models. The best estimate observed precipitation is given in the heavy black line. Note the different *y* axis.

of several moisture-related variables in the various models. Particular focus is given to comparing moisture-related variables as large errors can arise in models potentially due to the dependence of moisture on error-prone parameterizations. The convective component of total surface precipitation is discussed to highlight the different roles of model parameterization between the active and suppressed periods. Model accuracy will be discussed by comparison to observations for each model. The ensemble is then used to investigate model sensitivity in terms in the sources and sinks in the moisture budget. The best estimate is contrasted with the ensemble mean to directly determine how using an average of many simulations might affect results compared to a single simulation.

3.1.1. Overall Simulation Behavior

[32] Figure 2 shows time series of surface precipitation for the active and suppressed periods. Model ensemble means are shown as colored lines with individual ensemble members from all model simulations overlaid in gray. In this figure, all UM-type, SCAM-type, and JMA-type models are averaged together as they are very similar. Observations are shown as a heavy black line. This plot allows broad interpretation of the characteristics of each model while capturing the spread of the ensemble. Figure 2 shows that all models have a similar precipitation during the active period with moderate precipitation before the passage of the MCS on 23–24 January. All models have similar heavy rain associated with the MCS. The ensemble is spread around this mean

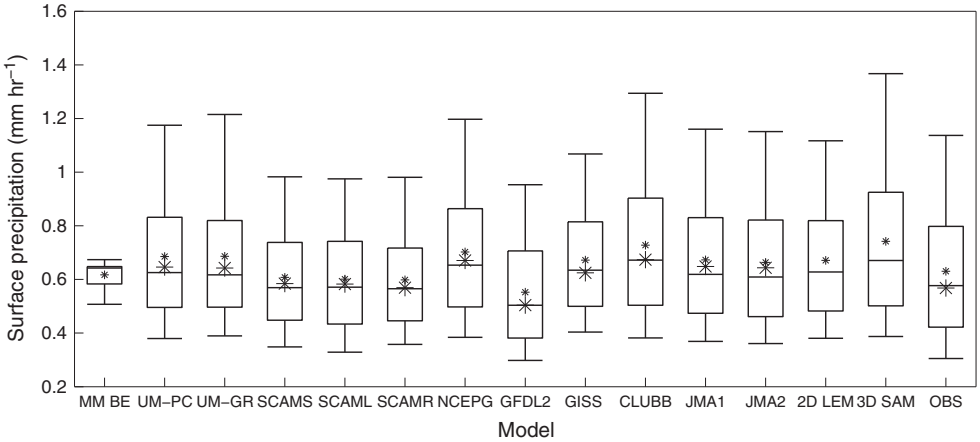

Figure 3. Mean precipitation averaged over the active and suppressed periods. The left box is the multimodel ensemble constructed from the best estimate simulations (MM BE) averaged over the period for each model. There are nine individual ensemble SCM with 80 members and two ensemble CRMs, a 2-D Met Office LEM simulation with 80 members, and a 3-D SAM model with 9 members. The far right has the ensemble of observations. The box represents the 25th, 50th, and 75th percentiles with the 5th and 95th percentiles being shown by the horizontal bars. The ensemble mean data are shown by the small asterisk. The best estimate is shown for the SCM data and observations as large asterisks. The ensemble is averaged for each ensemble member over all times.

with the largest spread occurring during the MCS. Modeled precipitation is close to observations which may be anticipated, as in strongly forced conditions precipitation will be predominantly driven by forcing in all ensemble members [*Xie et al*., 2005; *Xu et al*., 2002; *Woolnough et al*., 2010].

[33] Period-mean precipitation during the suppressed period is lower than during the active period. It is evident that the relative differences in the ensemble mean time evolution between models as well as the differences from observations are larger than those in the active period. This might be expected as the forcing is weaker and as a consequence has less of an influence on the model solutions. In weakly forced conditions, it is expected that the details of the parameterizations in the various models exert a stronger influence, which explains the larger differences in the suppressed period. The ensemble spread is rather uniform and does not increase substantially with rainfall, which remains light throughout the period. The CRM behave similarly to the SCM. In the active period, solutions from the two model types track each other closely, again highlighting that precipitation is constrained by the forcing in that period. Just as for the SCM, the differences between CRMs as well as to observations increase (in a relative sense) in the suppressed period. The CRM results in the active period strongly resemble the results of the larger CRM comparison [*Fridlind et al*., 2012], indicating that the CRMs shown here provide a representative sample for this family of models.

[34] Figure 3 provides a comparison of the multimodel best estimate ensemble and individual model ensembles for the time-mean surface precipitation averaged over the active and suppressed periods for all simulations used in this study. Each model is included as a box-whisker plot constructed from the time-averaged precipitation for each ensemble member. Observations are also included. It can clearly be seen that the ensemble SCM and CRM encompass a wide range of surface precipitation values. The models capture the spread seen in the observations. This is due to strong coupling between the forcing, which is primarily through vertical velocity, and rainfall.

[35] The multimodel ensemble has a limited spread of surface precipitation as all models are simulating the same forcing. Figure 3 provides a useful check that the multimodel ensemble has limited spread compared to the SCM and CRM simulations. This result supports findings of *Hume and Jakob* [2005] that largest spread in an SCM ensemble will be found by varying the forcing (the boundary conditions). Figure 3 also shows the ensemble mean (small asterisk) and best estimate mean (large asterisk) precipitation for the observations and models. For most models, the magnitude of the best estimate observed precipitation is very close to the 50th percentile (median) precipitation with the ensemble mean larger. This is due to the ensemble having a distribution which is skewed towards high values of precipitation leading to larger means than medians.

[36] Figure 4 shows time-height cross sections of the observed, SCM-mean and CRM-mean modeled relative humidity. Relative humidity provides a useful perspective on the model simulations, since unlike precipitation, which is primarily driven by forcing, relative humidity is less constrained by the forcing and more affected by model physics [*Emanuel and Zivkovic-Rothman*, 1999]. Given the model setup (section 2.2), the models have freedom to develop

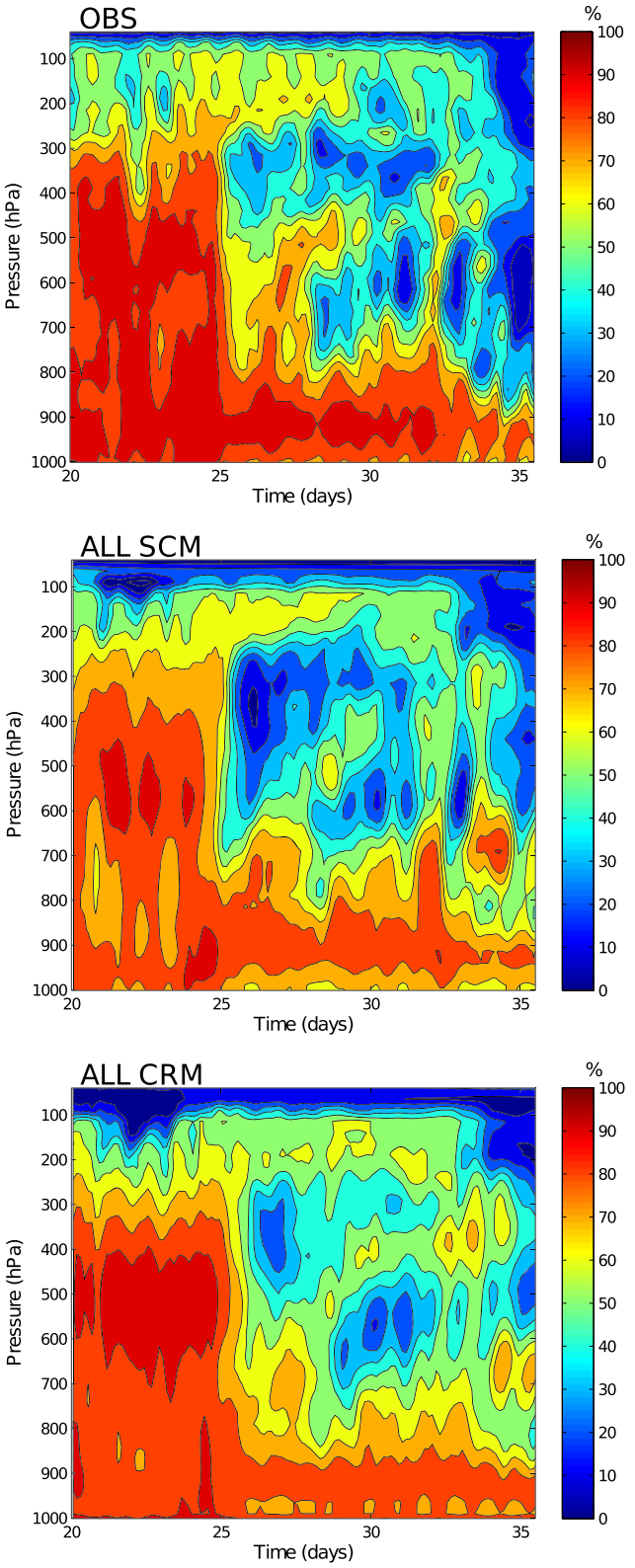

Figure 4. Time-pressure relative humidity (with respect to water) for the active and suppressed periods for observations and SCM and CRM simulations. The SCM and CRM data are averaged over all models and all available ensemble members in the range 10–90.

certain moisture source/sink terms such as moisture convergence and surface evaporation. The ensemble sensitivity to these terms will be addressed in section 3.1.3. Relative humidity with respect to water has been calculated using Teten's formula [*Lowe*, 1977, equation 6] for each individual simulation from values of temperature, water vapor, and pressure to ensure consistency across models. The modeled data shown in Figure 4 is an average over all models and all ensemble members used. Detailed investigation shows that relative humidity differences are primarily caused by differences in moisture as temperature varies little across the simulations and is close to observations.

[37] Observations show that the atmosphere has high relative humidity through a deep layer during the active period, but the models generally underestimate humidity particularly at low levels. During the suppressed period, observations show lower humidity above 800 hPa but large values in the boundary layer. All models capture the reduction in relative humidity caused by drying on the transition to the suppressed period above 700 hPa, although the SCM overestimate the reduction in humidity. Both SCM and CRM persist with low values of humidity in the boundary layer compared to observations.

[38] While Figure 4 shows the evolution of the mean state, the ensemble simulations also allow investigation of model sensitivity. Figure 5 shows time series of 500 mbar relative humidity for all ensemble members for each model compared to their best estimate simulations, ensemble mean, and observations. Relative humidity at 500 mbar is chosen, as accurate representation of moisture in midlevels is important if models are to correctly represent cloud. All models have high 500 mbar relative humidity during the active period consistent with the observations, but most SCM tend to have lower relative humidity than the CRM. The JMA and GISS models have particularly low relative humidity which is about 10% and 15% lower than the observations, respectively. The CLUBB and NCEP models have slightly larger relative humidity compared to the observations. All models have very limited spread during the active period.

[39] Observations show that during the transition to the suppressed period, humidity reduces to around 60% after the passage of the MCS. Relative humidity increases slightly before it reduces again from 70% to 30% between days 27 and 31 (27–31 January). There is a big difference between the responses of the SCM and CRM during this period. The CRM capture the transition to the suppressed period reasonably well with relative humidity 10% too low but its temporal evolution well captured. SCM generally reduce relative humidity too much in the transition period with mean values after the transition ranging from 40% (UM) to 10% (JMA). An exception to this is the CLUBB model which does not excessively reduce relative humidity during the transition and is then too moist during the suppressed period.

[40] The CRM show limited spread during the active period and the passage of the MCS. The spread in both model types is largest during the suppressed period. The SCM show larger but limited spread in the active period and in the transition associated with the MCS. Just like the CRM, they show increased spread during the suppressed period. This suggests a hypothesis that the simulation of midlevel relative humidity may be more sensitive to changes in the forcing when the forcing is weak. Furthermore, this sensitivity results in nonlinearity between the ensemble members which is particularly apparent during the suppressed period. For example, around 30 January, the SCAMS model shows that ensemble members with weaker (stronger) forcing have the lowest (highest) relative humidity despite the forcing not being the weakest (strongest) forcing.

[41] Figure 5 shows that in general the ensemble mean and best estimate simulation results follow each other quite closely so that their differences from observations are similar. On some limited occasions, the ensemble mean is closer to the observations than the best estimate, for example, UM-PC during both the active and suppressed periods and CLUBB and SCAMS during the suppressed period. To further investigate the ensemble mean to best estimate behavior, Figure 6 shows profiles of the difference between the best estimate and the ensemble mean relative humidity for all SCM for the active period. Figure 6 shows that when averaged over this period most models have similar best estimate and ensemble mean relative humidity. However, there are some important exceptions. For example, the UM-PC has larger ensemble mean relative humidity than its best estimate throughout the depth of the troposphere. This larger relative humidity in the ensemble mean represents an improvement in the model simulations by bringing the values closer to observations. As UM-PC is the only SCM to include a stochastic parameterization, this result highlights that ensemble simulations are necessary when using models with stochastic components. The usefulness of the ensemble approach will be investigated further below.

[42] When comparing the ensemble simulations with observations (Figure 5), it is possible, for some models and periods, to determine whether the errors are due to the prescribed forcing or are models errors. Given that the observed forcing spans the range of possible observations, none of the JMA ensemble members closely approximate the observed relative humidity during the active period. Therefore, this model clearly has limitations correctly simulating relative humidity during this period. For many models (including SCAM, NCEP, GISS, and JMA), the ensemble shows that the transition to the suppressed period is likely to be attributable to model error rather than errors in the forcing. The GISS model also consistently underestimates relative humidity during the suppressed period.

3.1.2. Precipitation Partitioning

[43] An interesting question in the simulation of tropical convection is how the various SCMs partition the precipitation between convection and the resolved scale motion. Furthermore, given the construction of the ensemble used here, it is possible to study how this partitioning changes with forcing strength and meteorological situation. Figure 7 shows the time average convective precipitation fraction (CPF), defined as the ratio of convective precipitation to total precipitation at the surface, against total precipitation for both the active and suppressed periods. Each SCM is shown by a color with different symbols used for the different models. Each point represents a single ensemble member averaged over the period of interest. An increase in total precipitation (x axis) indicates an increase in forcing strength. The multimodel best estimate ensemble is shown as large asterisks.

[44] Generally, there is a wide spread in the magnitude of CPF between the models ranging from 0.2 to 0.9 in the

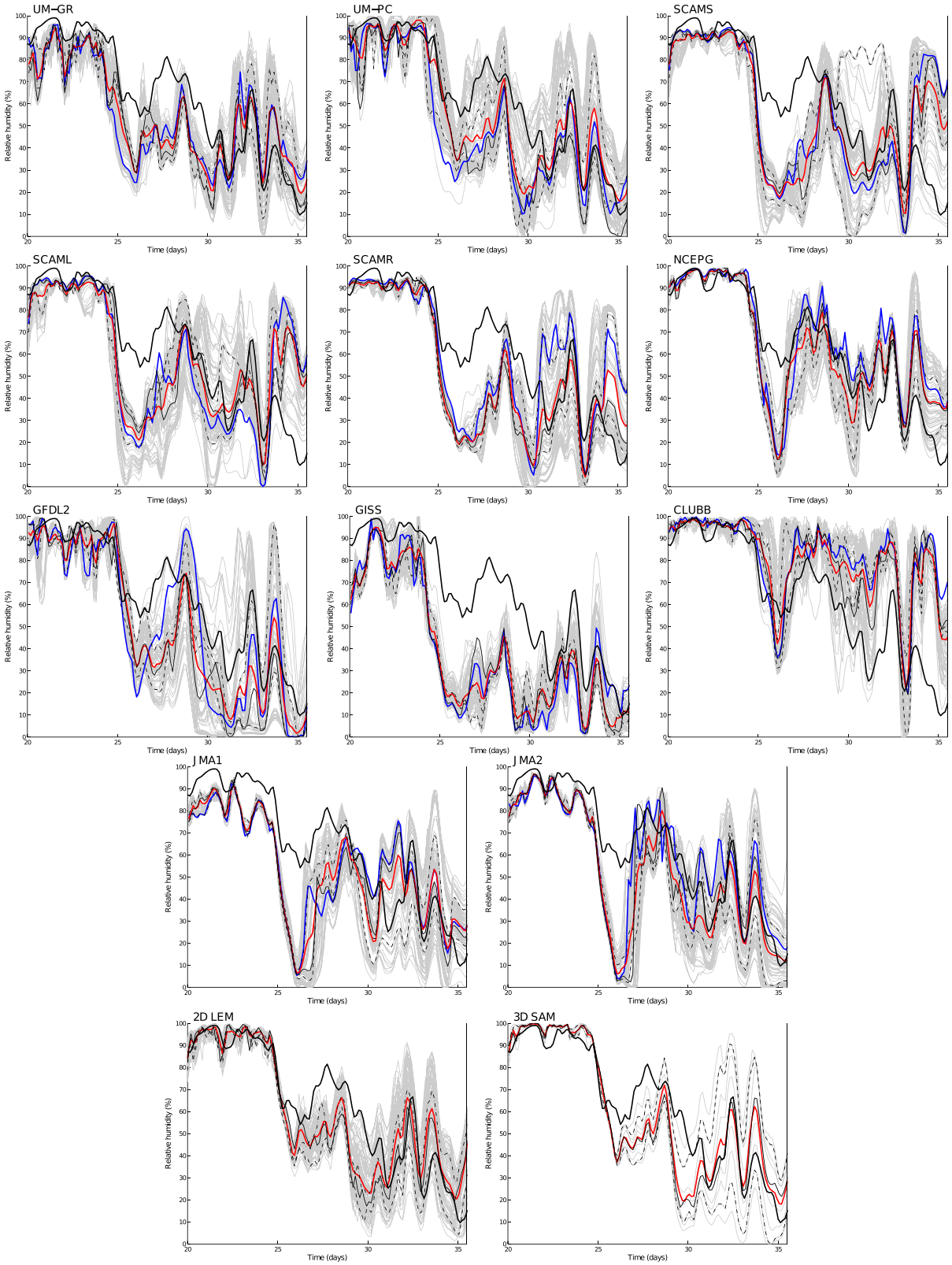

Figure 5. Time series of RH at 500 mbar for all SCM and CRM. Blue lines show best estimate simulations, red lines ensemble mean simulations, and the black line is observations. Gray lines show all ensemble members in the range 10–90. Key ensemble members, the 25th, 50th, and 75th percentiles of the 80-member large-scale "forcing," are highlighted as thin black lines which are dash-dotted, solid, and dashed, respectively. Note that the CRM do not report best estimate data.

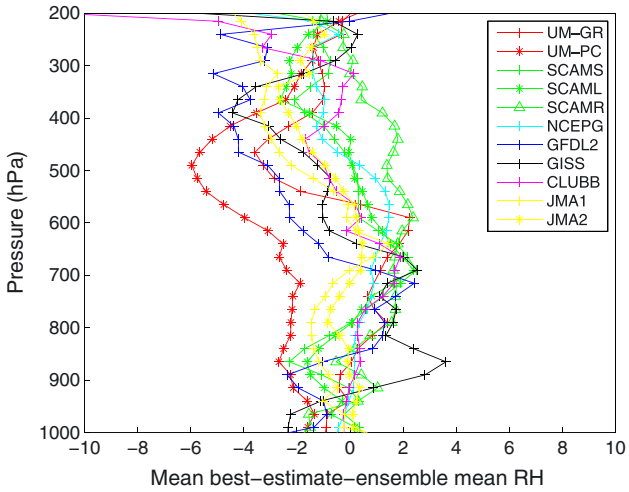

Figure 6. Mean period-averaged difference between the best estimate and ensemble mean relative humidity for the active period for each SCM.

active period and 0.5 to 1 in the suppressed period. In the active period, the models also show a very diverse behavior with forcing strength, with some showing an increase in CPF (e.g., GISS, UM-GR), some showing a near-constant CPF (e.g., NCEP, SCAM), and some showing a decrease (e.g., UM-PC). The GFDL2 model shows a somewhat erratic behavior. Models of the same type show different behavior depending on the parameterization scheme used (e.g., UM-PC versus UM-GR).

[45] In the suppressed period, all SCMs have a CPF of greater than 50%. There is a tendency in almost all models for the CPF to increase with increasing forcing although there is much scatter in the relationship. There are two groups of models, with either very high or relatively low CPF. There is some consistency between the periods, with the GISS and UM-PC models showing the lowest CPF in both.

[46] The rather wide spread in model behavior is likely indicative of large differences in the assumptions made in the different convection treatments on how to partition rainfall between convection and the larger scales. As this will likely have an impact on the vertical distribution of heating and moistening, an important issue for future work is to provide observational constraints for the relationships shown here.

3.1.3. Ensemble Moisture Budget Characteristics

[47] The ensemble provides an opportunity to investigate the interplay between modeled moisture and the moisture budget terms. In particular, this study permits a comparison between how the models control their moisture budgets. Given that the models are forced by prescribing horizontal advection terms and vertical velocity, they independently develop moisture budget terms such as vertical advection terms and moisture convergence in addition to the moisture contributions from parametrized processes such as convection and surface evaporation. This is an important difference between this study and previous intercomparisons [e.g., *Woolnough et al.*, 2010; *Guichard et al.*, 2004] where the total moisture forcing was prescribed. Furthermore, given that this study also includes both best estimate and ensemble simulations, comparison can be made about the additional model characteristics exposed using an ensemble compared to a single best estimate simulation.

[48] Figure 8 shows time average precipitable water against various terms in the moisture budget for the active period for all models and ensembles in this study. Very similar results are obtained for the suppressed period (not shown). Figure 8a shows that during the active period, the SCMs tend to divide into models in which lower precipitable water is associated with larger precipitation (GISS and SCAM), models where precipitable water is higher for larger values of precipitation (UM and CLUBB), and those models, including CRM, where precipitation is independent of precipitable water. The GFDL model is somewhat an exception as its relationship shows significant scatter.

[49] The largest term in the moisture budget is the moisture convergence term which is shown in Figure 8b. In all models, the moisture convergence term shows a similar magnitude and characteristics to precipitation which is not surprising as it is the largest source of moisture for the grid box exceeding surface evaporation by an order of magnitude (see below). Furthermore, Figure 8b shows that the moisture

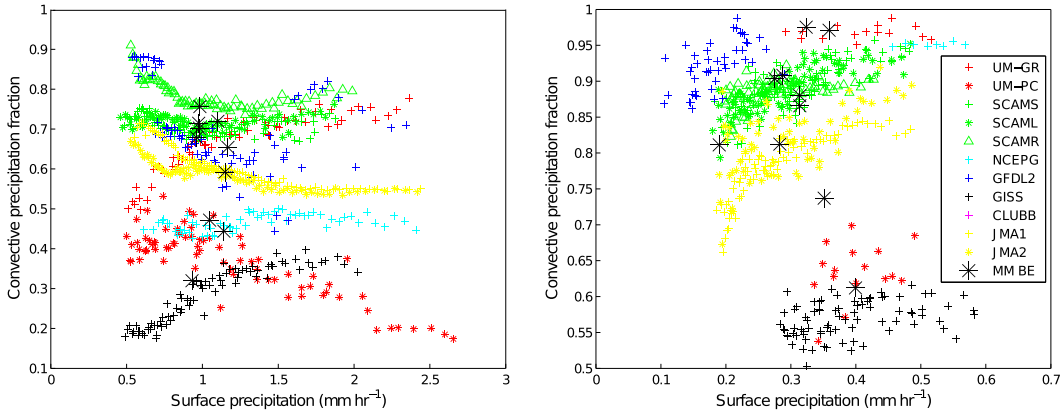

Figure 7. Time-averaged scatter plots of surface precipitation against convective precipitation (shown as a fraction of the total surface precipitation) over the active (left) and suppressed (right) periods for ensemble members 10–90. Each model type is represented by a color and each model of a given type by a symbol. The multimodel best estimate ensemble is represented by a large asterisk. The CLUBB model and CRM do not submit partitioned precipitation data.

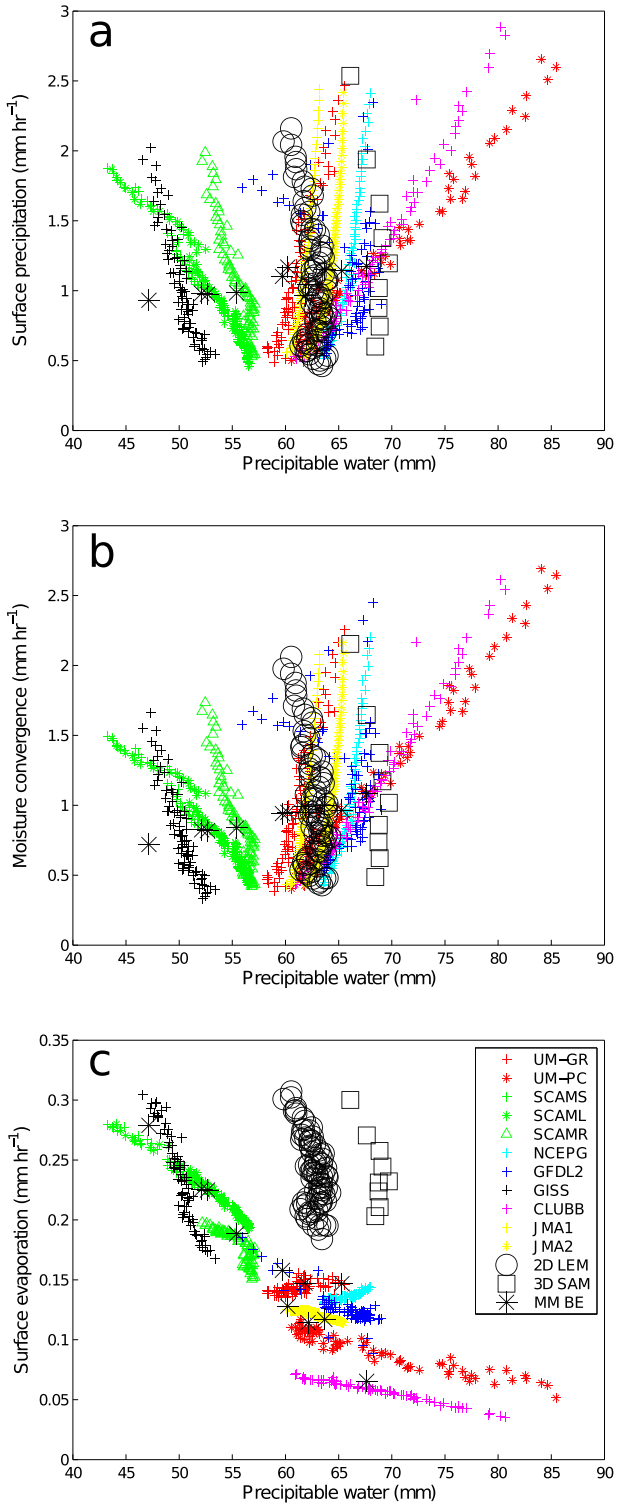

Figure 8. Time-averaged scatter plots of PW against (a) precipitation, (b) moisture convergence, and (c) surface evaporation over the active period for ensemble members 10–90. Each model type is represented by a color and each model of a given type by a symbol. The CRM are represented by large open symbols and the multimodel best estimate ensemble by a large asterisk.

convergence acts as feedback mechanism where SCM with larger values of precipitable water enhance moisture supply and produce more precipitation. Other models, despite the strong forcing, have lower precipitable water and lower moisture convergence. Petch et al. (submitted manuscript, 2012) discusses a likely reason by investigating the method used to force the SCM compared to the method used to force the CRM as used in *Fridlind et al.* [2012]. It was found that given a positive moisture bias, convergence (which occurs during the active period) increases that positive bias, and similarly convergence enhances a negative moisture bias. Models forced by prescribing the total moisture forcing, as used in *Fridlind et al.* [2012], do not develop these biases. The ensemble results shown in Figure 8b support the findings of Petch et al. (submitted manuscript, 2012). This model response to bias is not, however, apparent when only the best estimate simulations are considered. GISS and SCAM both have a drier atmosphere during the active period compared to the observations and other SCM which result in reduced precipitation compared to those SCM with a moister atmosphere.

[50] Another important term in the moisture budget is surface evaporation. Figure 8c shows this term for each model and ensemble member as before. Note that the surface evaporation term is an order of magnitude smaller than the moisture convergence contribution. It is evident that there is a fundamentally different relationship between forcing strength and evaporation in the SCMs and the CRMs indicating differences in the physical mechanisms at work in these two classes of models. All SCMs approximate a quasi-linear relationship of evaporation to precipitable water, albeit of varying strength, with larger surface evaporation at lower values of precipitable water and lower surface evaporation when precipitable water is high. This is consistent with the formulation of the SCMs as, given that low level winds and SST are prescribed in all models, evaporation can only change in response to atmospheric moisture. The CRMs on the other hand show a very different response to changes in the forcing. Here, the values of evaporation are independent of precipitable water. This indicates the importance of small-scale wind variability in driving surface evaporation. In the SCMs, this variability is not resolved. Unless it is parametrized, SCM surface fluxes are determined by the mean wind alone. In the CRMs, this wind variability is resolved and hence will enhance the surface fluxes. From the results, it is evident that the SCMs do not deal effectively with the subgrid variability. This result highlights the usefulness of the ensemble approach as this "error" in the SCMs would not have been evident from a set of single best estimate simulations.

[51] By using an ensemble approach, several interesting conclusions about model performance as well as simulation setup could be drawn. Given that strong precipitation in the models (and in nature) is strongly linked to moisture convergence, this exposes some interesting model behavior. By design of the simulations, moisture convergence is calculated by the models. Consequently, those models that develop a dry bias cannot develop large moisture convergence and do not produce as much precipitation, with the opposite effect occurring in models with a moist bias. The SCMs require a drier atmosphere to develop stronger surface evaporation. In contrast, the CRMs develop

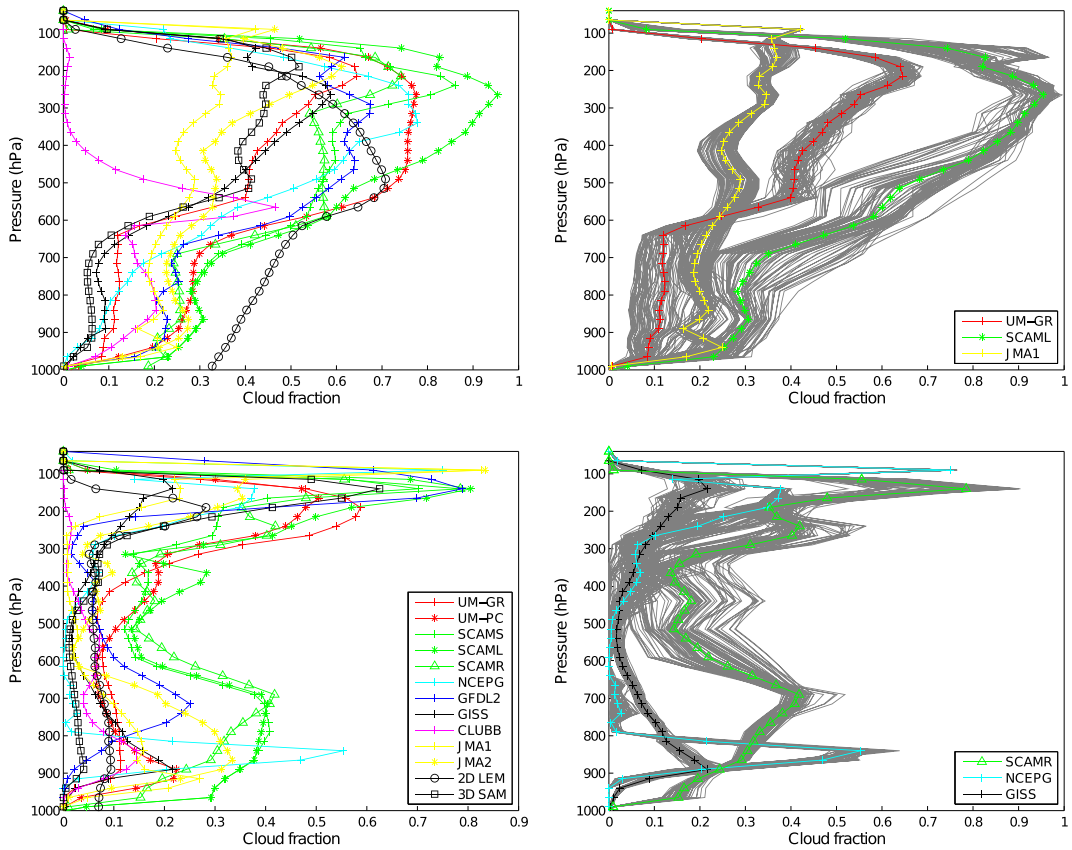

Figure 9. Mean period-averaged cloud fraction (left) for the active period (top) and suppressed period (bottom) for each model. For a limited number of models, the right-hand panels show the period mean together with ensemble members. Colored lines show the average for each model and gray lines ensemble members in the range 10–90. (Note that the CLUBB model does not include ice in cloud fraction and the LEM includes rain in cloud fraction.)

evaporation changes independent of atmospheric moisture likely due to the development of subgrid scale wind variability not present in the SCMs.

3.2. Clouds

[52] This section investigates the simulation of cloud-related variables in the CRMs and SCMs. Initially, the vertical structure of liquid water and ice clouds are discussed in both the active and suppressed monsoon. Following on from this, once again use of the ensemble will be made to investigate relationships between cloud-related variables as the forcing strength changes. This will expose several interesting characteristics of the various model parameterizations.

3.2.1. Profiles of Cloud Properties

[53] Figure 9 shows vertical profiles of the ensemble mean model cloud fraction for all models during the active (top left) and suppressed (bottom left) period as well as selected examples of the full ensemble from three models for the active (top right) and suppressed (bottom right) periods. Cloud fractions generally reflect the meteorological conditions shown in Figure 4 with cloud throughout the troposphere during the more moist, active period and two cloud layers during the suppressed period which are low cloud between 950-750 hPa, and high ice cloud above 200 hPa.

[54] During the active period, there are large differences in CRM cloud fraction of around 30% at all levels, and the SCMs mostly fall within the range of the CRMs. This can largely be explained by the definition of cloud fraction, which in the LEM includes both cloud and precipitating hydrometeors, while in the SAM model it only includes cloud water and ice. All SCMs have cloud fraction less than 30% below 600 hPa and more cloud (with the exception of JMA) above. There are large differences between the models with slightly better agreement in lower levels than in the upper troposphere.

[55] The differences in cloud fraction in the SCMs are also large in the suppressed period. One noticeable feature of the selected full ensembles (right panels) is that the difference of individual ensemble members from their mean tends to be smaller than the differences between models. This indicates that the differences in the simulated cloud structures are dominated by the structural properties of the models, not by the forcing data set, and shows that model representation of cloud is liable to error independent of the meteorological conditions. Best estimate simulations are therefore likely sufficient to expose model differences in this variable. This is investigated in Figure 10, which shows the differences of profiles of cloud cover between the ensemble mean and the best estimate simulation. As for relative humidity, most models show only small differences although with notable

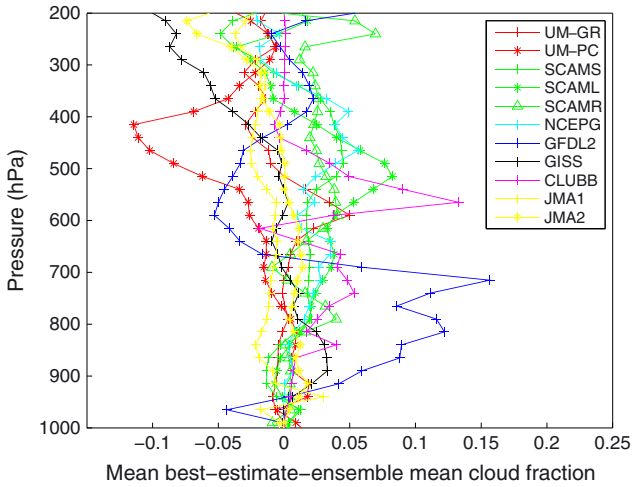

Figure 10. Mean period-averaged difference between the best estimate and ensemble mean cloud fraction for the active period for each SCM.

exceptions, the UM-PC around 400 hPa and GFDL below 700 hPa.

[56] Figure 11 shows profiles of ice water content in all models for the active period. Again, the ensemble means for all models are shown in the left panel, while selected full ensembles are shown in the right panel. The suppressed period is omitted from this Figure as the ice cloud during this period is not linked to local convection and is not well simulated. There are large differences between ice water content in both the CRMs and the SCMs during the active period which will impact on the model radiation budgets. Modeled ice water content differs in terms of both magnitude and vertical structure. Differences in the structural properties can again be noted in modeled ice water content with each SCM clustering around its ensemble mean. Difficulty in representing ice microphysics has been noted in all other TWP-ICE intercomparison studies and has been unanimously suggested as a focus for future model development.

[57] This section has shown that there are substantial differences in the vertical structure of parameterized cloud variables which may be attributable to systematic differences in the representation of clouds between the models. Structure in the cloud variables is clearly identifiable using the ensemble in both the active and suppressed periods. These persistent structures show that the models are not sensitive to changes in the forcing and that for most models best estimate simulations are likely sufficient to expose the mean model behavior in both periods. It is clear from the large differences between them that the CRMs only provide a limited estimate of the truth, especially during the suppressed period, as their representations of clouds are limited themselves [*Fridlind et al*., 2012].

3.2.2. Ensemble Cloud Characteristics

[58] While the previous section showed that it is likely that the mean cloud properties of each model can be exposed by a single best estimate simulation, the full ensemble results provide a useful tool to investigate how relationships between variables might change within each model as the forcing varies across ensemble members. Representing the correct relationships between variables is a greater challenge for models than representing means, but it is also a necessary condition for applying the models over a wide range of conditions, such as a full GCM. This subsection will investigate how the ensemble developed here can be used to investigate relationships between different variables. Each ensemble member, experiencing different forcing data, can be considered as a separate test case, albeit spaced in controlled manner from all other ensemble members.

[59] Figure 12 shows the mean liquid water path (LWP) as a function of the mean surface precipitation averaged over the active (left) and suppressed (right) periods for all models. Each symbol represents an individual ensemble member. While there are generally different relationships between the two periods (note the change in scale between periods in the Figure), the CRMs show that relationship between LWP and precipitation is linear (with a gradient of approximately 250 kg m^{-3} h in both the active and suppressed period). The CRMs agree very well during the suppressed period but differ at the larger precipitation rates during the active period.

[60] Most, but not all, SCMs also produce a linear relationship between LWP and surface precipitation. Notable exceptions are the GFDL, SCAM, and JMA models. The relationships in the SCMs differ somewhat between the

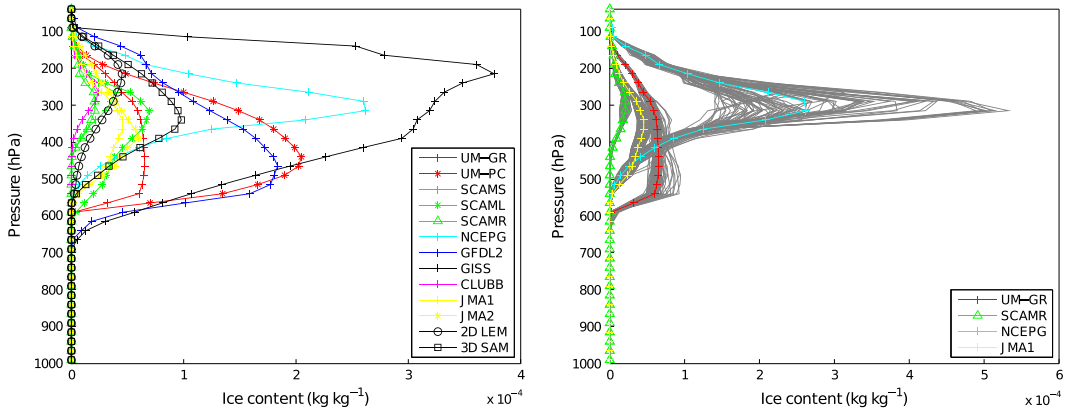

Figure 11. Mean period-averaged ice water content for the active period (left) for each model type. For a limited number of models, the right-hand panels show the period mean together with ensemble members. Colored lines show the average for each model and gray lines ensemble members in the range 10–90.

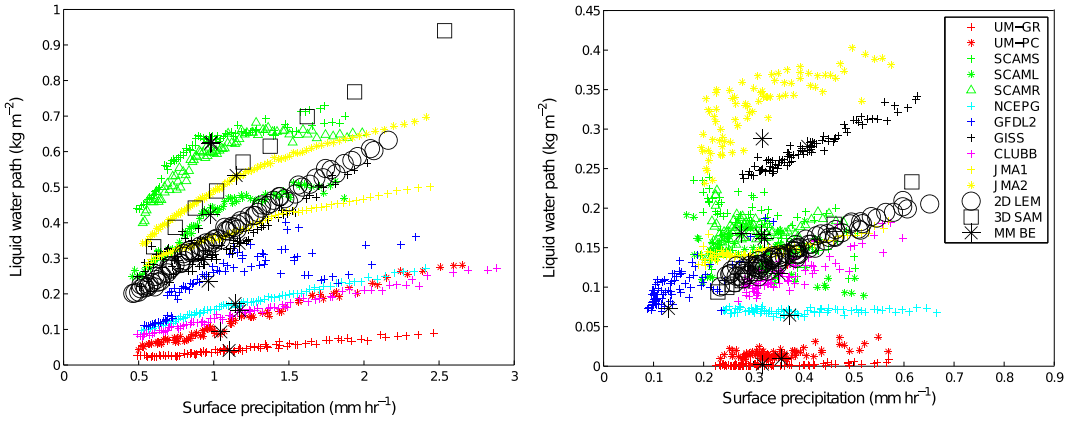

Figure 12. Time-averaged scatter plots of surface precipitation against liquid water path over the active (left) and suppressed (right) periods for ensemble members 10–90. Each model type is represented by a color and each model of a given type by a symbol. The CRM are represented by large open symbols and the multimodel best estimate ensemble by a large asterisk.

active and suppressed periods with a tendency for models to have tighter and more linear relationships during the active period. In the suppressed period when precipitation is small, both the UM and NCEP models tend to have precipitation independent of LWP, which itself is at an almost constant value. The GFDL and SCAM models tend to display significant scatter in LWP with only a weak relationship to precipitation. In fact, only the CLUBB, GISS, and JMA models increase LWP with precipitation as the CRMs suggest during the suppressed period. A linear relationship was observed between LWP and precipitation in *Fridlind et al.* [2012].

[61] The CRMs tend to lie in the middle of the SCM distribution, suggesting that the SCM ensemble mean may approximate the correct values of LWP, although individual models may differ quite considerably from the CRMs. The UM and NCEP models are biased low at all times, whereas the GISS and one of the JMA models have a LWP that is too large during the suppressed period. Unlike for cloud fraction before, the best estimate simulations do not always fall close to the center of the ensemble (note the large asterisks for GFDL and one JMA model to their associated ensemble during the suppressed period). The ensemble results also expose interesting nonlinearities in some of the models. For instance, there is a discontinuity in LWP in the GFDL around 0.15 kg m^{-2} during the active period. This possibly relates to the discontinuity in the convective precipitation fraction in Figure 7. While magnitude differences are apparent in the multimodel ensemble, the relationships between LWP and precipitation are only found in the full ensemble showing a potential usefulness of an ensemble technique when identifying model behavior.

[62] Figure 13a shows the relationship between IWP and precipitation during the active period. It can be seen that similar to LWP, IWP generally has a linear relationship with precipitation. Unlike the relationship of precipitation with LWP, the one with IWP is not consistent between the CRMs. There are very different magnitudes of IWP in the CRMs, and the slope of the relationship to precipitation varies strongly as well. Large differences in the simulation of

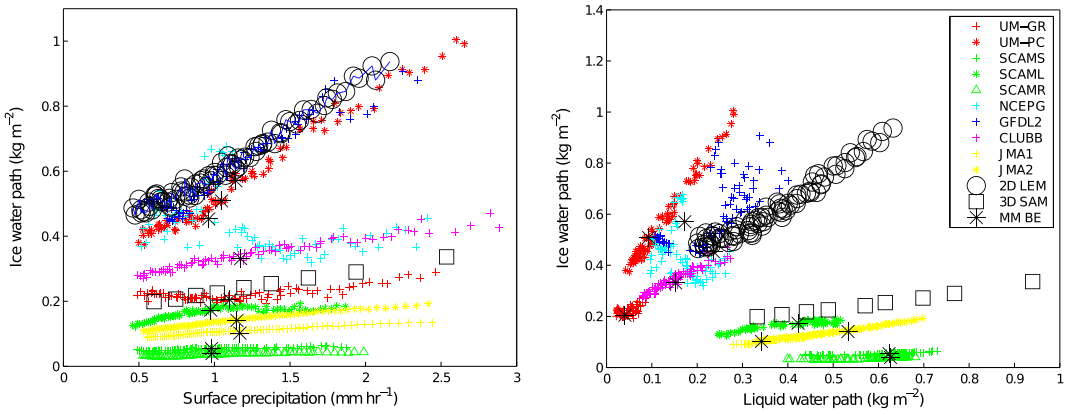

Figure 13. Time-averaged scatter plots of surface precipitation against ice water path (left) and liquid water path against ice water path (right) over the active period for ensemble members 10–90. Each model type is represented by a color and each model of a given type by a symbol. The CRM are represented by large open symbols and the multimodel best estimate ensemble by a large asterisk. Note: GISS model is not shown in these plots. GISS IWP exceeds all other models by a factor of 2.

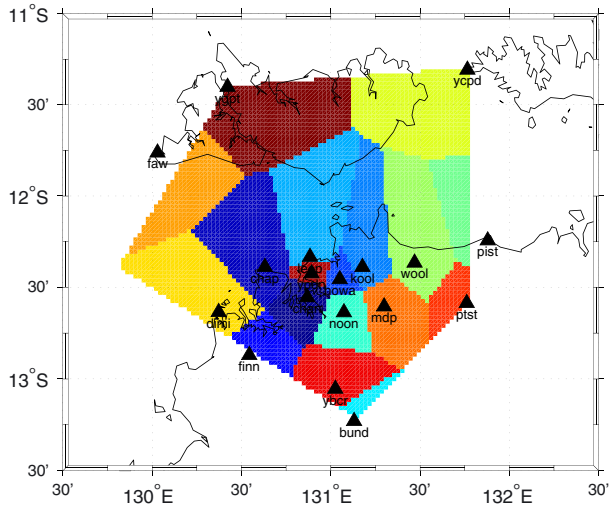

Figure 14. Location of rain gauges used in this study in and around the pentagon-shaped TWP-ICE domain. Comparison of radar-derived rainfall data to gauge rainfall data is conducted at all stations. Representative results are shown for Batchelor station (YBCR, 131.0252W, 13.0545S) and Charles Point (CHAP, 130.6309W, 12.389S). The colored area around each gauge shows the region of the TWP-ICE domain area closest to that gauge.

IWP in CRMs have been identified in other studies [*Fridlind et al.*, 2012]. The existence of those discrepancies makes it difficult to use the CRM results in assessing the SCM behavior. The LEM has approximately double IWP compared to SAM with a gradient of 300 kg m^{-3} h (LEM) compared to 50 kg m^{-3} h in SAM. Most SCM have gradients around this range, although in the NCEP model IWP is relatively insensitive to forcing.

[63] The ensemble enables the comparison not only between models but also of different versions of the same model. For example, SCAMS and SCAMR have very similar IWP, whereas SCAML, using a different microphysics scheme, has twice the IWP of the other SCAM models. There is a more marked difference between the two versions of the UM. UM-PC follows closely the gradient and approximate magnitude of the LEM (which is the UK Met Office's CRM) and which was used in the formulation of the *Plant and Craig* [2008] stochastic convection parameterization scheme. The UM-GR, on the other hand, is close to the SAM CRM which shows that there is complex interplay between the parameterization schemes. The UM SCM only differ in their convection parameterization, but this has a large effect on the IWP produced. In general, there is a split between models that follow the strong slope of the LEM and those closer to the weaker slope of the SAM. It is not possible, however, to attribute the relationship between precipitation and IWP simply based on the model microphysics scheme.

[64] Figure 13b shows the relationship between LWP and IWP and shows different aspects of the relationships between the variables in the models. There is a clear spilt between some models that have larger ranges in LWP (e.g., SCAM and JMA models) and others that have larger ranges in IWP (e.g., UM-PC and CLUBB). *Fridlind et al.* [2012] found that 2-D CRMs have a weaker relationship than 3-D CRMs between IWP and LWP, which is contrary to Figure 13a. However, the 2-D version of the LEM used here was not part of the *Fridlind et al.* [2012] study, and furthermore, the SAM here used a single-moment microphysics scheme, whereas the SAM in *Fridlind et al.* [2012] used a double-moment scheme [*Morrison et al.*, 2009] so a direct comparison is not possible.

[65] Interestingly, considering only the multimodel ensemble (Figure 13b) shows a different relationship between LWP and IWP compared to the relationship shown in the individual ensemble simulations. The ensemble within each model suggests increasing IWP with LWP, whereas the multimodel ensemble would suggest a tendency for IWP to increase with decreasing LWP. This shows the differences and potential limitations of using a multimodel

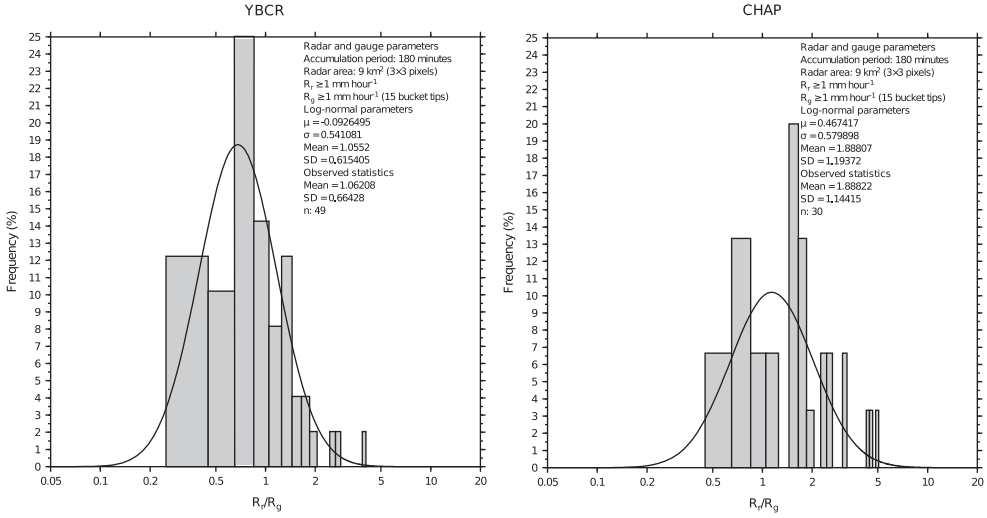

Figure 15. Distributions of radar-derived rainfall normalized by rain gauge rainfall for two rain gauges for TWP-ICE. A log-normal fit is shown in the solid line. Statistics of the observed data and the fit data are given in the top right corner of each panel.

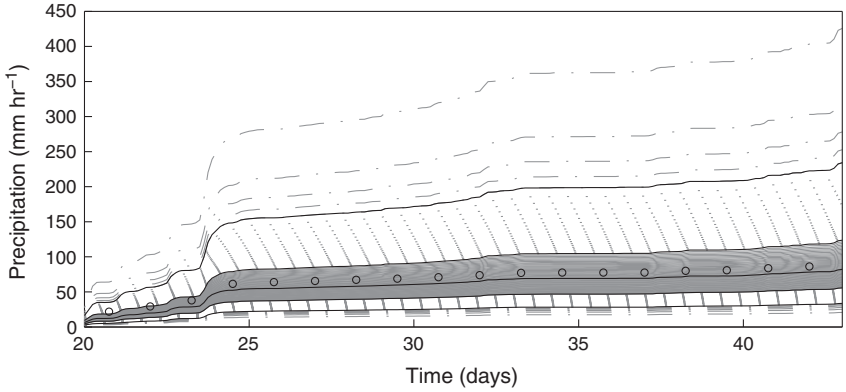

Figure 16. Ensemble cumulative rainfall time series for TWP-ICE derived from error estimates in radar-derived rainfall. Broken and light-colored lines show all ensemble members with key ensemble members (5th, 25th, 50th, 75th, and 95th percentiles) as black continuous lines. The best estimate forcing, used here and in the CRM intercomparison, is shown by small circles.

ensemble. Using a multimodel ensemble would suggest the reverse characteristic relationship between variables to that suggested by CRM and SCM each simulating their own ensemble.

4. Summary and Discussion

[66] This study presents an ensemble of SCM and CRM simulations for the TWP-ICE period. The first purpose of the study was to derive an ensemble of model forcings based on observational uncertainty. This data set was then applied to a variety of models to assess what new information about model behavior and model error might be gleaned from an ensemble approach that could not be attained by a single realization commonly used in CRM and SCM studies. It was found that the overall model behavior in terms of the time evolution of thermodynamic variables or the time-averaged vertical structure of those variables generally changes little between the ensemble mean and a single "best estimate" simulation. However, there were some notable exceptions to that finding. In some model simulations, like those with the UM-PC, ensemble means deviate from the best estimate simulations throughout the troposphere. Given that the ensemble mean forcing is close to that of the best estimate, this indicates nonlinearities in the simulation behavior possibly due to the stochastic component of the model. The ensemble also shows that models have greater sensitivity when weakly forced, and therefore, an ensemble is necessary. Perhaps the main value the ensemble adds to single simulations is the possibility to investigate the changes in model behavior with changes in forcing. This has proved invaluable in highlighting several aspects of model behavior in this study, namely, (i) a distinctly different behavior in the SCMs from that in the CRMs in achieving changes in surface evaporation; (ii) the sensitivity to the particular forcing method applied, (iii) a wide spread in the convective precipitation fraction in models and its sensitivity to forcing strength, and (iv) distinctly different model behavior in the relationships between cloud variables and precipitation.

[67] Examining the terms of the moisture budget using the ensemble enabled interesting conclusions about model behavior for two important terms; the surface evaporation and the moisture convergence. A clear distinction exists between the CRMs and the SCMs. In the CRMs, evaporation increases for constant atmospheric moisture, whereas the SCMs can only increase evaporation by drying the atmosphere. This suggests a role of subgrid variability likely brought about by cold pools in the CRMs that is not parameterized in SCMs. A representation of cold pool dynamics in SCMs would allow surface evaporation to occur in a moist atmosphere. Studying the moisture convergence term as a function of forcing strength revealed an interesting feedback between model error and the particular forcing approach chosen here. As the models are forced with horizontal moisture advection and vertical motion profiles (and hence profiles of mass convergence and divergence), they develop their own vertical moisture advection and moisture convergence terms. In models that develop a moist/dry bias, this bias is reinforced by an increase/decrease of the moisture convergence into the region. This behavior limitation can easily be deduced using the ensemble approach, while it would go largely unnoticed in single simulations with a number of models.

[68] The ensemble was also shown to be useful in investigating cloud variables and their relationships. Ensemble vertical profiles generally highlight structural differences between different models in that all ensemble members of a particular model tend to lie closer to its mean than to that of other models, even with large variations in the forcing. Consistent with the results in the accompanying modeling studies for TWP-ICE [*Lin et al*., 2012; *Fridlind et al*., 2012; *Zhu et al*., 2012], large differences are found in the models' simulation of cloud ice, highlighting this area once again as one warranting further study. The ensemble is used to identify relationships between liquid water, cloud ice, and precipitation. CRM simulations, while varying in magnitude, show clear linear relationships between those variables. This behavior is not reproduced in all SCMs, some of which show strongly nonlinear behavior or even jumps. The ensemble also reveals that the ice water path to liquid water path relationships are very different between models, with one group of models showing a very strong increase of IWP with LWP, while in others IWP is almost independent of LWP. This conclusion applies to both CRMs and SCMs. Using the multimodel best estimate ensemble only,

the important relationship of increasing ice water path with liquid water path in individual models is reversed.

[69] This study shows that the introduction of an ensemble to a modeling study provides more information than might be gathered by simulating only simple best estimate forcing. While the method does not replace the standard best estimate approach to single-column modeling, it complements it by (i) providing an easy framework to study model sensitivities and (ii) increasing confidence in detecting model behavior that is likely due to model, rather than forcing, limitations. Future SCM studies should therefore consider adding ensemble simulations in addition to, rather than instead of, the more conventional best estimate method. Despite the additional information provided by the ensemble, it remains difficult to conclusively link model behavior in an SCM to parameterization assumptions, highlighting the need to embed studies like the one presented here into a larger framework of model evaluation.

Appendix A: Derivation of the Large-Scale Forcing Ensemble

[70] An important part of this study is the use of an ensemble of large-scale forcing data sets. The motivation for doing so is to assess the inherent uncertainty in deriving a single best estimate of the large-scale atmosphere from observations and in its subsequent application to drive model simulations. This appendix describes the construction of the ensemble used in this study, which is based on two steps: (i) estimate errors in the estimate of area-mean rainfall and construct alternative rainfall scenarios and (ii) apply a constrained variational analysis to each of the rainfall scenarios derived in the first step to yield the final ensemble of large-scale atmospheric states.

A1. Deriving an Ensemble of Rainfall Estimates

[71] The main source area-mean rainfall information in this and other TWP-ICE studies [e.g., *Xie et al.*, 2010] are rainfall estimates from a C-band polarimetric radar located near Darwin [*Keenan et al.*, 1998]. The algorithm used to estimate rainfall from radar variables is that of *Bringi and Chandrasekar* [2001]. While the radar provides excellent spatial coverage to estimate area means, deriving rain rates from radar variables will lead to errors in the rainfall estimates. A first step in the ensemble construction is to estimate these errors. To do so, we use rain gauge observations around Darwin and apply a method very similar to that of *Jordan et al.* [2003].

[72] Radar rain rates vary in space and time, and radar errors may vary considerably based on location and timing of rain events. The array of rain gauge data shown in Figure 14 is used as a reference for the radar-derived rainfall data. A grid of 3×3 radar pixels (approximately 9 km^2) are averaged and compared to rain gauge measurements over an accumulated period of 180 min where both rain rates are greater than 1 mm. By performing this analysis at many locations over the TWP-ICE domain, it is anticipated that the differing sources of error may be better accounted for.

[73] Examples of the ratio of radar-derived rainfall data to rain gauge rainfall data are shown in Figure 15 for two rain gauges. Assuming that rain gauge data may be a better estimate of rainfall than radar-derived data, ratios close to 1 suggest small errors in the radar data, with smaller standard deviations showing the clustering of the errors. The statistics in Figure 15 for the observed data show differences in the mean values and standard deviations at the two locations, suggesting that indeed errors have different spatial patterns. As the data tend to cluster about 1, the two observed data sets predominantly agree on the magnitude of rainfall, although the long tails of the error distribution show that on occasions large errors can be identified.

[74] A log-normal distribution is fitted to the errors shown in Figure 15. The log-normal distribution parameters are estimated and used to construct an ensemble of rain rates at each radar pixel as follows. The distribution of radar to gauge rainfall ratios is divided into 100 percentiles. Then the ratio for each percentile is used to multiply the radar rain values, providing 100 rainfall values (one for each percentile) at each radar pixel. For each radar pixel, the error distribution derived at the nearest rain gauge is used. Figure 14 shows the areas (colored) for which error characteristics are assumed constant in space based on the nearest rain gauge behavior.

[75] Having derived rainfall error estimates at each radar pixel, which is expressed as 100 values of rainfall from the lowest to the highest, the next task is to estimate the error in the area-mean rainfall. This requires assumptions about the spatial correlation of the individual pixel errors. As our goal is to span the widest range of possibilities, we will assume the worst case scenario of maximum correlation. In other words, we assume that whenever the largest possible error occurs at 1 pixel, the largest error in the same direction occurs at all radar pixels. This is an extremely simple assumption and will maximize the possible error in the area-mean rainfall, consistent with our goal to maximize ensemble spread. Using this assumption, 100 values of area-mean rainfall are derived by simply averaging the pixel-rainfall rates within each percentile, i.e., the first percentile of the area-mean rainfall distribution is simply the average of all first-percentile values at each pixel and so on stepping through all percentiles. Figure 16 shows the 100 cumulative rainfall time series in this way for TWP-ICE. For comparison, the figure includes the best estimate rainfall time series as derived by *Xie et al.* [2010], which falls close to the 50th percentile as might be anticipated from the method the distribution was constructed. While the error estimates allow for a large range of possible rainfall values, 50% of the distribution falls between the 25th and 75th percentiles of the distribution which has a limited range of rainfall.

A2. Deriving the Large-Scale Atmospheric State

[76] Each of the 100 rainfall scenarios derived above is used separately in the variational analysis algorithm of *Zhang et al.* [2001] (all other observations, such as thermodynamic variables, horizontal winds, and radiation terms, are unchanged and are the same for each scenario) to produce 100 separate forcings that are all equally possible given the uncertainty in area-mean rainfall. The higher (lower) percentile corresponds to stronger (weaker) surface precipitation and generally stronger (weaker) vertical motion. The characteristics of the vertical motion for the active and suppressed periods are discussed in the main text.

[77] Investigations were made into whether the additional variational analysis inputs should be modified in order to

be more physically consistent. For example, an estimate of rainfall error has been used to derive alternative rainfall time series, but increased rainfall may, in the simplest terms, also be associated with more deep cloud and therefore reduced top-of-the-atmosphere longwave radiation, which is also an input to the variational analysis. Sensitivity studies where the radiation was varied in conjunction with rainfall had little impact on the resulting large-scale atmosphere. This supports *Zhang et al.* [2001], who suggested that rainfall provided the largest contribution term in the variational analysis.

[78] The 100 large-scale data sets so derived are used to provide forcing data for SCM and CRM as described in the main text.

[79] **Acknowledgments.** Davies and Jakob are supported by the Office of Science (BER), U.S. Department of Energy, under grant DE-SC0002731. Many of the other coauthors also participated through support from the U.S. Department of Energy Atmospheric System Research Program. V. Larson and B. Nielsen are grateful for financial support from the United States Department of Energy (grants DE-SC0006927 and DE-SC0008668) and the National Science Foundation (grant AGS-0968640). Support for X. Liu was provided by the U.S. Department of Energy (DOE), Office of Science, Atmospheric System Research (ASR) program. The Pacific Northwest National Laboratory is operated for DOE by Battelle Memorial Institute under contract DE-AC06-76RLO 1830. Dr. Weiguo Wang is partly supported by the National Natural Science Foundation of China under Grant No. 41075039. The contributions of S. Xie to this work were performed under the auspices of the U.S. Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research by Lawrence Livermore National Laboratory under contract No. DE-AC52-07NA27344 and supported by the Atmospheric Radiation Measurement Program of the Office of Science at the DOE.

References

Arakawa, A., and W. Schubert (1974), Interaction of a cumulus cloud ensemble with the large-scale environment, Part 1, *J. Atmos. Sci.*, *31*, 674–701, doi:10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2.

Ball, M. A., and R. S. Plant (2008), Comparison of stochastic parameterization approaches in a single-column model, *Phil. Trans. Roy. Soc.*, *366*, 2605–2623, doi:10.1098/rsta.2011.0377.

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. Meteorolog. Soc.*, *126*(564), 865–888, doi:10.1002/qj.49712656404.

Boville, B. A., P. J. Rasch, J. J. Hack, and J. R. McCaa (2006), Representation of clouds and precipitation processes in the Community Atmosphere Model version 3 (CAM3), *J. Climate*, *19*(11), 2184–2198, doi:10.1175/JCLI3749.1.

Bringi, V. N., and V. Chandrasekar (2001), *Polarimetric Doppler Weather Radar: Principles and Applications*, 636 pp., Cambridge University Press, The Edinburgh Building, Cambridge, CB2 2RU, UK.

Chou, M., M. J. Suarez, C. H. Ho, M. M. H. Yan, and K. T. Lee (1998), Parameterizations for cloud overlapping and shortwave single scattering properties for use in general circulation and cloud ensemble models, *J. Climate*, *11*, 202–214, doi:10.1175/1520-0442(1998)011<0202:PFCOAS>2.0.CO;2.

Collins, W. D., P. J. Rasch, B. A. Boville, J. J. Hack, J. R. McCaa, D. L. W. Williamson, B. P. Briegleb, C. M. Bitz, S.-J. Lin, and M. Zhang (2006), The formulation and atmospheric simulation of the Community Atmosphere Model version 3 (CAM3), *J. Climate*, *19*, 2144–2161, doi:10.1175/JCLI3760.1.

Davies, T., M. J. P. Cullen, A. J. Malcolm, M. H. Mawson, A. Staniforth, A. A. White, and N. Wood (2005), A new dynamical core for the Met Office's global and regional modelling of the atmosphere, *Q. J. R. Meteorol. Soc.*, *131*, 1759–1782, doi:10.1256/qj.04.101.

Del Genio, A., M.-S. Yao, W. Kovari, and K.-W. Lo (1996), A prognostic cloud water parameterization for global climate models, *J. Climate*, *9*, 270–304, doi:10.1175/1520-0442(1996)009<0270:APCWPF>2.0.CO;2.

Del Genio, A., W. Kovari, M.-S. Yao, and J. Jonas (2005), Cumulus microphysics and climate sensitivity, *J. Climate*, *18*, 2376–2387, doi:10.1175/JCLI3413.1.

Derbyshire, S. H., A. V. Maidens, S. F. Milton, R. A. Stratton, and M. R. Willett (2011), Adaptive detrainment in a convective parametrization, *Q. J. R. Meteorolog. Soc.*, *137*(660), 1856–1871, doi:10.1002/qj.875.

Edwards, J., and A. Slingo (1996), Studies with a flexible new radiation code. Part I: Choosing a configuration for a large-scale model, *Q. J. R. Meteorol. Soc.*, *122*, 689–719, doi:10.1002/qj.49712253107.

Emanuel, K., and M. Zivkovic-Rothman (1999), Development and evaluation of a convection scheme for use in climate models, *J. Atmos. Sci.*, *56*(11), 1766–1782, doi:10.1175/1520-0469(1999)056<1766:DAEOAC>2.0.CO;2.

EMC, (2003), The GFS atmospheric model, *Technical report*, National Oceanic and Atmospheric Administration, US Department of Commerce.

Fels, S., and M. Schwarzkopf (1975), The simplified exchange approximation: A new method for radiative transfer calculations, *J. Atmos. Sci.*, *37*, 2265–2297, doi:10.1175/1520-0469(1975)032<1475:TSEAAN>2.0.CO;2.

Freidenreich, S. M., and V. Ramaswamy (1999), A new multiple-band solar radiative parameterization for general circulation models, *J. Geophys. Res.*, *104*(D24), 31389–31409, doi:10.1029/1999JD900456.

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.

Fu, Q., and K. N. Liou (1993), Parameterization of the radiative properties of cirrus clouds, *J. Atmos. Sci.*, *50*, 2008–2025, doi:10.1175/1520-0469(1993)050<2008:POTRPO>2.0.CO;2.

GAMDT (2004), The new GFDL global atmosphere and land model AM2-LM2: Evaluation with prescribed SST simulations, *J. Climate*, *17*, 4641–4673, doi:10.1175/JCLI-3223.1.

Ghan, S., et al. (2000), A comparison of single column model simulations of summertime midlatitude continental convection, *J. Geophys. Res.*, *105*(D2), 2091–2124, doi:10.1029/1999JD900971.

Golaz, J.-C., V. Larson, and W. Cotton (2002), A PDF-based model for boundary layer clouds. Part I: Method and model description, *J. Atmos. Sci.*, *59*, 3540–3551, doi:10.1175/1520-0469(2002)059<3540:APBMFB>2.0.CO;2.

Grabowski, W. W., et al. (2006), Daytime convective development over land: A model intercomparison based on LBA observations, *Q. J. Roy. Meteor. Soc.*, *132*, 317–344, doi:10.1256/qj.04.147.

Grant, A. (2001), Cloud-base fluxes in the cumulus-capped boundary layer, *Q. J. Roy. Meteor. Soc.*, *127*(572), 407–421, doi:10.1002/qj.49712757209.

Gray, M. E. B., J. Petch, S. H. Derbyshire, A. R. Brown, A. P. Lock, H. A. Swann, and P. R. A. Brown (2001), Version 2.3 of the Met. Office Large Eddy Model: Part II. Scientific documentation. Met O (APR) Turbulence and Diffusion Note. No. 276.

Gregory, D. (2001), Estimation of entrainment rate in simple models of convective clouds, *Q. J. R. Meteorol. Soc.*, *127*, 53–72, doi:10.1002/qj.49712757104.

Gregory, D., and P. R. Rowntree (1990), A mass flux convection scheme with representation of cloud ensemble characteristics and stability dependent closure, *Mon. Wea. Rev.*, *118*, 1483–1506, doi:10.1175/1520-0493(1990)118<1483:AMFCSW>2.0.CO;2.

Guichard, F., et al. (2004), Modelling the diurnal cycle of deep precipitating convection over land with cloud-resolving models and single-column models, *Q. J. Roy. Meteor. Soc.*, *130*, 3139–3171, doi:10.1256/qj.03.145.

Hack, J. J., and J. A. Pedretti (2000), Assessment of solution uncertainties in single-column modelling frameworks, *J. Climate*, *13*(10), 352–365, doi:10.1175/1520-0442(2000)013<0352:AOSUIS>2.0.CO;2.

Held, I. M., R. S. Hemler, and V. Ramasway (1993), Radiative-convective equilibrium with explicit two-dimensional moist convection, *J. Atmos. Sci.*, *50*, 3909–3927, doi:10.1175/1520-0469(1993)050<3909:RCEWET>2.0.CO;2.

Holtslag, A. A. M., and B. A. Boville (1993), Local versus nonlocal boundary-layer diffusion in a global climate model, *J. Climate*, *6*(10), 1825–1842, doi:10.1175/1520-0442(1993)006<1825:LVNBLD>2.0.CO;2.

Hong, S., and H. Pan (1996), Nonlocal boundary layer vertical diffusion in a medium-range forecast model, *Mon. Wea. Rev.*, *124*(10), 2322–2339, doi:10.1175/1520-0493(1996)124<2322:NBLVDI>2.0.CO;2.

Hume, T., and C. Jakob (2005), Ensemble Single Column Modelling (ESCM) in the tropical western Pacific: Forcing datasets and uncertainty analysis, *J. Geophys. Res.*, *110*, D13109, doi:10.1029/2004JD005704.

Hume, T., and C. Jakob (2007), Ensemble Single Column Model (ESCM) validation in the tropical western Pacific, *J. Geophys. Res.*, *112*, D10206, doi:10.1029/2006JD008018.

JMA (2007), Outline of the Operational Forecast and Analysis System of the Japan Meteorological Agency. http://www.jma.go.jp/jma/jma-eng/jma-center/nwp/outline-nwp/index.htm.

Jordan, P., A. Seed, and P. Weinmann (2003), A stochastic model of radar measurement errors in rainfall accumulations at catchment

scale, *J. Hydro.*, *4*, 841–855, doi:10.1175/1525-7541(2003)004<0841:ASMORM>2.0.CO;2.

Joss, J., and A. Waldvogel (1990), *Precipitation Measurement and Hydrology. Radar in Meteorology*, chapter 29a, pp. 577–606, Academic Press, Boston, MA, USA.

Keenan, T. D., K. Glasson, F. Cummings, T. S. Bird, J. Keller, and J. Lutz (1998), The BMRC/NCAR C-band polarimetric (CPOL) radar system, *J. Atmos. Oceanic Technol.*, *15*, 871–886, doi:10.1175/1520-0426(1998)015<0871:TBNCBP>2.0.CO;2.

Khairoutdinov, M., and D. Randall (2003), Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties, and sensitivities, *J. Atmos. Sci.*, *60*, 607–625, doi:10.1175/1520-0469(2003)060<0607:CRMOTA>2.0.CO;2.

Khairoutdinov, M. F., and Y. L. Kogan (1999), A large eddy simulation model with explicit microphysics: Validation against aircraft observations of a stratocumulus-topped boundary layer, *J. Atmos. Sci.*, *56*(13), 2115–2131, doi:10.1175/1520-0469(1999)056<2115:ALESMW>2.0.CO;2.

Kiehl, J., J. J. Hack, G. B. Bonan, B. A. Boville, D. L. Williamson, and P. J. Rasch (1998), The national center for atmospheric research community climate model: CCM3*, *J. Climate*, *11*(6), 1131–1149, doi:10.1175/1520-0442(1998)011<1131:TNCFAR>2.0.CO;2.

Larson, V. E., and J.-C. Golaz (2005), Using probability density functions to derive consistent closure relationships among higher-order moments, *Mon. Wea. Rev.*, *133*, 1023–1042, doi:10.1175/MWR2902.1.

Larson, V. E., D. P. Schanen, M. Wang, M. Ovchinnikov, and S. Ghan (2012), PDF parameterization of boundary layer clouds in models with horizontal grid spacings from 2 to 16 km, *Mon. Wea. Rev.*, *140*, 285–306, doi:10.1175/MWR-D-10-05059.1.

Lin, Y., et al. (2012), TWP-ICE global atmospheric model intercomparison: Convection responsiveness and resolution impact, *J. Geophys. Res.*, *117*, D09111, doi:10.1029/2011JD017018.

Liu, X., and J. Penner (2005), Ice nucleation parameterization for global models, *Meteorol. Z.*, *14*(4), 499–514, doi:10.1127/0941-2948/2005/0059.

Liu, X., J. E. Penner, S. J. Ghan, and M. Wang (2007), Inclusion of ice microphysics in the NCAR community atmospheric model version 3 (CAM3), *J. Climate*, *20*, 4526–4547, doi:10.1175/JCLI4264.1.

Lock, A. P., A. R. Brown, M. R. Bush, G. M. Martin, and R. N. B. Smith (2000), A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests, *Mon. Wea. Rev.*, *128*, 3187–3199, doi:10.1175/1520-0493(2000)128<3187:ANBLMS>2.0.CO;2.

Lowe, P. (1977), An approximating polynomial for the computation of saturation vapor pressure, *J. Appl. Meteorol.*, *16*, 100–102, doi:10.1175/1520-0450(1977)016<0100:AAPFTC>2.0.CO;2.

MacVean, M., and P. Mason (1990), Cloud-top entrainment instability through small-scale mixing and its parameterization in numerical models, *J. Atmos. Sci.*, *47*(8), 1012–1030, doi:10.1175/1520-0469(1990)047<1012:CTEITS>2.0.CO;2.

Martin, G. M., M. A. Ringer, V. D. Pope, A. Jones, C. Dearden, and T. J. Hinton (2006), The physical properties of the atmosphere in the new Hadley Centre Global Environmental Model (HadGEM1). Part I: Model description and global climatology, *J. Climate*, *19*(7), 1274–1301, doi:10.1175/JCLI3636.1.

May, P. T., J. H. Mather, G. Vaughan, C. Jakob, G. M. McFarquhar, K. N. Bower, and G. G. Mace (2008), The tropical warm pool international cloud experiment, *Bull. Amer. Meteor. Soc.*, *89*, 629–645, doi:10.1175/BAMS-89-5-629.

McClatchey, R. A., R. W. Fenn, J. Selby, F. E. Volz, and J. S. Garing, (1972), The Tropical Warm Pool International Cloud Experiment, *Technical Report 411*, Air Force Cambridge Research Laboratory Environmental Research Paper.

Mellor, G. L., and T. Yamada (1974), A hierarchy of turbulence closure models for planetary boundary layers, *J. Atmos. Sci.*, *31*, 1791–1806, doi:10.1175/1520-0469(1974)031<1791:AHOTCM>2.0.CO;2.

Moorthi, S., and M. J. Suarez (1992), Relaxed Arakawa-Schubert: A parameterization of moist convection for general circulation models, *Mon. Wea. Rev.*, *120*, 978–1002, doi:10.1175/1520-0493(1992)120<0978:RASAPO>2.0.CO;2.

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. Wea. Rev.*, *137*(3), 991–1007, doi:10.1175/2008MWR2556.1.

Nakagawa, M., (2009), Outline of the high resolution global model at the Japan meteorological agency, *RSMC Tokyo-Typhoon Center Technical Review, **11**(1-13)*, Available from http://www.jma.go.jp/jma/jma--eng/jma--center/rsmc--hp--pub--eg/techrev/text11--1.pdf.

Pan, H., and W. Wu (1995), Implementing a mass flux convection parameterization package for the NMC medium-range forecast model. *NMC Office Note,* No. 409, 40 pp. [Available from NCEP, 5200 Auth Road, Washington, DC 20233].

Plant, R. S., and G. C. Craig (2008), A stochastic parameterization for deep convection based on equilibrium statistics, *J. Atmos. Sci.*, *65*, 87–105, doi:10.1175/2007JAS2263.1.

Randall, D., and D. M. Pan (1993), Implementation of the Arakawa-Schubert cumulus parameterization with a prognostic closure, *Meteorol. Monogr.*, *46*, 137–144.

Randall, D., et al. (2003), Confronting models with data: The GEWEX cloud systems study, *Bull. Amer. Meteor. Soc.*, *84*, 455–469, doi:10.1175/BAMS-84-4-455.

Rasch, P. J., and J. E. Kristjánsson (1998), A comparison of the CCM3 model climate using diagnosed and predicted condensate parameterizations, *J. Climate*, *11*(7), 1587–1614, doi:10.1175/1520-0442(1998)011<1587:ACOTCM>2.0.CO;2.

Rotstayn, L. D. (1979), A physically based scheme for the treatment of stratiform clouds and precipitation in large-scale models. I: Description and evaluation of the microphysical processes, *Q. J. Roy. Meteor. Soc.*, *123*, 1227–1282, doi:10.1002/qj.49712354106.

Schmidt, G. A., et al. (2006), Present day atmospheric simulations using GISS Model E: Comparison to in-situ, satellite and reanalysis data, *J. Climate*, *19*, 153–192, doi:10.1175/JCLI3612.1.

Schwarzkopf, M. D., and S. B. Fels (1991), The simplified exchange method revisited: An accurate, rapid method for computation of infrared cooling rates and fluxes, *J. Geophys. Res.*, *96*(D5), 9075–9096, doi:10.1029/89JD01598.

Schwarzkopf, M. D., and V. Ramaswamy (1999), Radiative effects of CH_4, N_2O, halocarbons and the foreign-broadened H_2O continuum: A GCM experiment, *J. Geophys. Res.*, *104*(D8), 9467–9488, doi:10.1029/1999JD900003.

Slingo, A. (1989), A GCM parameterization for the shortwave radiative properties of water clouds, *J. Atmos. Sci.*, *46*, 1419–1427, doi:10.1175/1520-0469(1989)046<1419:AGPFTS>2.0.CO;2.

Smith, R. N. B. (1990), A scheme for predicting layer clouds and their water content in a general circulation model, *Q. J. R. Meteorol. Soc.*, *116*, 435–460, doi:10.1002/qj.49711649210.

Stephens, G. L., P. M. Gabriel, and P. T. Partain (2001), Parameterization of atmospheric radiative transfer. Part I: Validity of simple models, *J. Atmos. Sci.*, *58*, 3391–3409, doi:10.1175/1520-0469(2001)058<3391:POARTP>2.0.CO;2.

Sundqvist, H., E. Berge, and J. E. Kristjansson (1989), Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model, *Mon. Wea. Rev.*, *117*, 1641–1657, doi:10.1175/1520-0493(1989)117<1641:CACPSW>2.0.CO;2.

Tiedtke, M. (1983), The sensitivity of the time-mean large-scale flow to cumulus convection in the ECMWF model, in *Workshop on Convection in Large-scale Numerical Models*, vol. *28*, Shinfield Park, Reading, pp. 297–316.

Tiedtke, M. (1993), Representation of clouds in large-scale models, *Mon. Wea. Rev.*, *121*(11), 3040–3061, doi:10.1175/1520-0493(1993)121<3040:ROCILS>2.0.CO;2.

Wang, W., X. Liu, S. Xie, J. Boyle, and S. A. McFarlane (2009), 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.

Webster, P., and R. Lukas (1992), TOGA-COARE: The Coupled Ocean-Atmosphere Response Experiment, *Bull. Amer. Meteor. Soc*, *73*, 1377–1416, doi:10.1175/1520-0477(1992)073<1377:TCTCOR>2.0.CO;2.

Wilson, D., and R. Forbes, (2004), Unified model documentation paper 26: The large-scale precipitation parametrization scheme, *Technical Report 26*, Met Office R&D.

Wilson, D. R., and S. P. Ballard (1999), A microphysically based precipitation scheme for the UK Meteorological Office Unified Model, *Q. J. R. Meteorol. Soc.*, *125*, 1607–1636, doi:10.1002/qj.49712555707.

Woolnough, S. J., P. Blossey, K. M. Xu, P. Bechtold, J. C. T. Hosomi, S. Iacobellis, Y. Luo, J. Petch, R. Wong, and S. Xie (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*(647), 333–353, doi:10.1002/qj.568.

Xie, S., et al. (2002), Intercomparison and evaluation of cumulus parametrizations under summertime midlatitude continental conditions, *Quart. J. Roy. Meteor. Soc.*, *128*, 1095–1136, doi:10.1256/003590002320373229.

Xie, S., R. T. Cederwall, and M. Zhang (2004), Developing long-term single-column model/cloud system—resolving model forcing data using numerical weather prediction products constrained by surface and

top of the atmosphere observations, *J. Geophys. Res.*, *109*, D01104, doi:10.1029/2003JD004045.

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. Climate*, *23*, 57–79, doi:10.1175/2009JCLI3071.1.

Xie, S., and M. Zhang (2000), Impact of the convection triggering function on single-column model simulations, *J. Geophys. Res.*, *105*(D11), 14983–14996, doi:10.1029/2000JD900170.

Xu, K. M., et al. (2002), An intercomparison of cloud-resolving models with the atmospheric radiation measurement summer 1997 intensive observation period data, *Quart. J. Roy. Meteor. Soc.*, *580*, 593–624, doi:10.1256/003590002321042117.

Zhang, G., and N. A. McFarlane (1995), Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian climate centre general circulation model, *Atmos. Ocean*, *33*, 407–446, doi:10.1080/07055900.1995.9649539.

Zhang, G. J. (2002), Convective quasi-equilibrium in midlatitude continental environment and its effect on convective parameterization, *J. Geophys. Res.*, *107*(D14), 4220, doi:10.1029/2001JD001005.

Zhang, M., W. Lin, C. S. Bretherton, J. J. Hack, and P. J. Rasch (2003), A modified formulation of fractional stratiform condensation rate in the NCAR Community Atmospheric Model (CAM2), *J. Geophys. Res.*, *108*(D1), 4035, doi:10.1029/2002JD002523.

Zhang, M. H., and J. L. Lin (1997), Constrained variational analysis of sounding data based on column integrated budgets of mass, heat, moisture, and momentum: Approach and application to ARM measurements, *J. Atmos. Sci.*, *54*, 1503–1524, doi:10.1175/1520-0469(1997)054<1503:CVAOSD>2.0.CO;2.

Zhang, M. H., J. L. Lin, R. T. Cederwall, J. J. Yio, and S. C. Xie (2001), Objective analysis of ARM IOP data: Method and sensitivity, *Mon. Wea. Rev.*, *129*, 295–311, doi:10.1175/1520-0493(2001)129<0295:OAOAID>2.0.CO;2.

Zhao, Q. Y., and F. H. Carr (1997), A prognostic cloud scheme for operational NWP models, *Mon. Wea. Rev.*, *125*, 1931–1953, doi:10.1175/1520-0493(1997)125<1931:APCSFO>2.0.CO;2.

Zhu, P., J. Dudhia, P. Field, K. Wapler, A. Fridlind, A. Varble, E. Zipser, J. Petch, M. Chen, and Z. Zhu (2012), A limited area model (LAM) intercomparison study of a TWP-ICE active monsoon mesoscale convective event, *J. Geophys. Res.*, *117*, D11208, doi:10.1029/2011JD016447.