Final report

Ensemble Single Column Modeling in the Tropics - Derivation of observed forcing data sets, estimation of observation uncertainty and application to parametrization improvements

DE-FG02-08ER64527

Christian Jakob, Peter May, Alan Seed, Timothy Hume, Laura Davies

The project was carried out in 2 distinct phases. In the first phase we established the general validity of using an ensemble approach to Single Column Modeling (SCM) using the Manus and Nauru sites. In the second phase we derived and applied an ensemble forcing derivation technique to observations. First we derived an ensemble forcing estimate for the TWP-ICE experiment and then by extended the method to provide three wet seasons of “continuous ensemble forcing” for the Darwin site. The main purpose of using ensemble techniques in SCM simulation is to be able to assess how much of the overall SCM simulation error is due to model errors and how much due to errors in the forcing. Below we briefly summarize the main results of the entire project. All relevant publications are appended to this report.

Step 1: Establish the ESCM technique

Using NWP re-analysis products we first derived forcing data sets for SCM models over the TWP sites of Manus and Nauru. We thoroughly evaluated the quality of the derived data sets against observations from radiosondes at the sites. We then applied the derived forcing data sets to two SCMs to identify the main source of uncertainties resulting from forcing data in SCM studies. We conclusively showed that it is mainly the boundary conditions, and not as previously thought the initial conditions, that drive uncertainties in SCMs. We further showed that the ensemble methodology does allow to separate model error from forcing errors. The results of this part of the study are summarized in Hume and Jakob (2005, attached)

Step 2: Use the ESCM technique in model evaluation

Having established the ESCM technique, we applied it to the evaluation of the BMRC SCM to demonstrate how ensemble techniques can be combined with regime-oriented model evaluation and probabilistic model evaluation techniques to identify model errors and to separate them from errors in the prescribed forcing. We also applied the technique to assess possible modifications to the parametrization of convection in the ECWMF model. We showed that the ESCM techniques is capable of clearly identifying the effect of the changes on the model results, an important characteristic in its application to judge model improvements. The result of this part of the study are summarized in Hume and Jakob (2007, attached).

Step 3: Derive best-estimate forcing data sets for TWP-ICE

The TWP-ICE experiment in Jan/Feb 2006 in Darwin spent much effort taking observations to derive forcing data sets for SCM studies. The data collected during the experiment provided the foundation for the main thrust of the research over the second phase of the project. As outlined above, the purpose of the research was to assess the uncertainties in the derivation of SCM forcing data and to provide an alternative

methodology of SCM simulations based on ensemble techniques that inherently include the uncertainties in the forcing data sets and hence allow us to estimate the uncertainties in the model results directly resulting from forcing errors. The first step to deriving an ensemble forcing for TWP-ICE was to derive the conventional best-estimate forcing data set for TWP-ICE. This was achieved in collaboration with the LLNL group and the result of this work are published in Xie et al. (2010, attached).

Step 4: Derive ensemble forcing data set for TWP-ICE

Having established the best-estimate data set, the next step was to derive an ensemble forcing data set for the TWP-ICE period. As a publication on this is still in preparation, we summarize the main ideas of doing so here. The basic approach relies on assessing the uncertainties in the observations that form the input to the constrained variational analysis technique that is used in ARM to derive model forcing data from observations and that was used above in determining the best-guess data set.

Figure 1: 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. The coloured area around each gauge shows the region of the TWP-ICE domain closest area closest to that gauge.

The main source area-mean rainfall information in this and other TWP-ICE studies are rainfall estimates from a C-band polarimetric radar located near Darwin. 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.

Figure 2: Distributions of radar-derived rainfall normalised 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 is given in the top right of each panel

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 1 is used as a reference for the radar-derived rainfall data. A grid of 3x3 radar pixels (approximately 9 km²) are averaged and compared to rain gauge over an accumulated period of 180min where both rain rates are greater than 1mm. 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.

Examples of the ratio of radar-derived rainfall data to rain gauge rainfall data are shown in Figure 2 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 2 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 tends to cluster about 1, the two observed datasets predominantly agree on the magnitude of rainfall although the long tails of the error distribution show large errors can be identified.

Radar errors have been shown to have a log-normal distribution so an estimate of the errors is found by fitting a log-normal curve to the observed errors shown in Figure 2. By design the statistics of the log-normal distribution are similar to those given for the observations. Therefore the log-normal distributions gives a good estimate of the errors found comparing observations of radar and rain gauge observed rainfall data.

The log-normal distribution parameters are 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 mean 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.

Having derived rainfall error estimates at each radar pixel 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 one pixel, the largest error in the same direction occurs at all radar pixels. This is an extremely simple assumption and will maximise the possible error in the area-mean rainfall, consistent with our goal to maximise ensemble spread.

Figure 3: Ensemble cumulative rainfall timeseries for TWP-ICE derived from error estimates in radar-derived rainfall. Broken and light-coloured lines show all ensemble members with key ensemble members (5th, 25th, 50th, 75th and 95th) as black continuous lines. The best estimate forcing, used here and in the CRM intercomparison, is shown by small circles.

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 3 shows the 100 cumulative rainfall timeseries in this way for TWP-ICE. For comparison, the figure includes the best-estimate rainfall timeseries 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. Whilst the error estimates allow for a large range of possible rainfall values 50% of the distribution falls between the 25th and 75th percentile of the distribution which has a limited range of rainfall. Each of these 100 scenarios are then processed separately using the variational analysis (all other observations are unchanged and are the same for each scenario) to produce 100 separate forcings that are all equally possible given the uncertainty in rainfall estimate. The higher (lower) the percentile corresponds to stronger (weaker) surface precipitation. Figure 4 shows the vertical velocity profile averaged over both the active and suppressed monsoon for each ensemble member as well as key percentiles. It can be seen the stronger rainfall (higher percentiles) is associated with stronger vertical motion. 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 lower percentiles, with weaker rainfall, have upward motion at lower levels (below 650hPa) but downward motion above. In addition to the ensemble members, Figure 4 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.

Figure 4: Time-averaged vertical profiles of omega over active (left) and suppressed (right) for all ensemble members. Broken and light-coloured lines show all ensemble members with key ensemble members (5th, 25th, 50th, 75th and 95th) 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.

Whilst there is a large spread in omega it is again worth noting that 50% of the ensemble members lie in the limited range between the 25-75th percentile lines. While each ensemble member is equally likely, most cluster around the 50th ensemble member and the best estimate and that the most extreme omega values are rare. These difference in omega imply changes in low-level convergence and upper-level divergence through the continuity equation and will have an effect on convection.

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 timeseries 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.

The 100 large-scale datasets may be used to provide forcing data for SCM and CRM and when simulated provide additional information, when compared to using a best estimate simulation alone, of model accuracy and spread.

Step 5: Lead a community SCM comparison study using the ensemble forcing

In the final step of this research we applied the 100 forcing data sets to a large number of single column model to study their ability to simulate the TWP-ICE cloud and convective conditions. We find that the largest spread in the model results exist in the suppressed period of TWP-ICE, while the active and break periods (both periods of deep convection) show smaller spread. This may be seen as somewhat surprising given that the active and break periods exhibit a much larger spread in the forcing itself. We interpret this result as showing that under strong forcing (active and break period) the model (and perhaps the real atmosphere) are strongly controlled by the forcing data itself. During the suppressed period on the other hand the model has a wider range of possible solutions and so small changes to the forcing allow large deviations in the model results. These early results have very recently been shown to also hold when the ensemble forcing is applied to a simple 2-dimensional cloud resolving model. The results of this intercomparison study are currently being prepared for publication.

Step 6: Deriving a long-term ensemble data set for Darwin

Another line of our research focused on comparing the observation-derived model forcing data sets with those directly derived from NWP systems we evaluated the ability of ECMWF NWP analyses to correctly represent the large scales in the tropics. Through this assessment we can explore the use of NWP data to generate long-term forcing data sets in particular by combining the NWP analysis with additional observations (e.g., rainfall). We found that there are significant differences between the observed and NWP data sets. The NWP analysis shows some difficulties in representing the observed distribution of vertical motion. When deriving vertical velocity using a “hybrid” approach where the radiosonde observations used in the variational analysis are replaced by data from the ECMWF NWP analysis, while the rainfall and other observed boundary conditions are used in deriving the forcing in a way identical to that used in the fully observed case. We can show that many of the differences between the ECMWF and observed forcing disappear. This indicates that the combination of NWP analyses with observations of the boundary conditions, most notably rainfall observations, can provide a very close approximation of the results based entirely on observations. While this has been highlighted before for the extratropical ARM SGP site, our work is the first to apply these techniques in the tropics. As the rainfall and other boundary condition data is available from radar and satellites for long period of times, this finding will enable the application of the forcing derivation techniques to long periods of time, thereby, for the first time, providing the community with long-term reliable estimates of the large-scale state of the tropical atmosphere in conjunction with detailed observations of the convective state. A publication on this part of our research is also in preparation.

Having established its feasibility long-term forcing data sets for the Darwin region using both the techniques described above, i.e. the use of NWP instead of soundings and the use of ensemble techniques were derived. At the end of our research we have derived 3 wet-seasons of ensemble forcing data (~1900 six-hourly samples).

Products:

We have delivered both an ensemble model forcing data sets for Naura and Manus and those for TWP-ICE and the long-term forcing data sets to the ARM PI archive, where they are available under <http://www.arm.gov/data/pi/10> and <http://www.arm.gov/data/pi/46>. We also delivered a model evaluation data set for Nauru and Manus spanning the same period as the forcing data. The latter is available under <http://www.arm.gov/data/pi/11>.

Ensemble single column modeling (ESCM) in the tropical western Pacific: Forcing data sets and uncertainty analysis

Timothy Hume and Christian Jakob

Bureau of Meteorology Research Centre, Melbourne, Victoria, Australia

Received 15 December 2004; revised 11 March 2005; accepted 14 April 2005; published 12 July 2005.

[1] Single column models (SCMs) are useful tools for the evaluation of parameterisations of radiative and moist processes used in general circulation models (GCMs). Most SCM studies to date have concentrated on regions where there is a sufficiently dense observational network to derive the required forcing data. This paper describes an ensemble single column modeling (ESCM) approach where the forcing data are derived from numerical weather prediction (NWP) analysis products. To highlight the benefits of the ESCM approach, four forcing data sets were derived for a two year period at the Tropical Western Pacific ARM (Atmospheric Radiation Measurement Program) sites at Manus Island and Nauru. In the first section of the study, the NWP derived forcing data are validated against a range of observations at the tropical sites. In the second section, the sensitivity of two different SCMs to uncertainties in the forcing data sets are analysed. It is shown that despite the inherent uncertainties in the NWP derived forcing data, an ESCM approach is able to identify errors in the SCM physics. This suggests the ESCM approach is useful for testing parameterisations in relatively observation sparse regions, such as the TWP.

Citation: Hume, T., and C. Jakob (2005), Ensemble single column modeling (ESCM) in the tropical western Pacific: Forcing data sets and uncertainty analysis, *J. Geophys. Res.*, 110, D13109, doi:10.1029/2004JD005704.

1. Introduction

[2] Radiative and moist processes play a critical role in the Earth's climate system, yet the parameterisations of these processes in current general circulation models (GCMs) still contain many uncertainties. Consequently, considerable effort has been put into developing frameworks to test GCM parameterisations. These range from testing parameterisations inside a full three dimensional model [e.g., *Grell*, 1993] through to cloud resolving models (CRMs) and Single Column Models (SCMs) [e.g., *Randall et al.*, 1996; *Bechtold et al.*, 2000; *Bergman and Sardeshmukh*, 2003]. Each of these methods have their own strengths and weaknesses. For example, a full GCM tests the parameterisation as it is intended to be used. However, the results from a full GCM can sometimes be difficult to interpret because of the many complex interactions and feedbacks between physical and dynamical processes. Additionally, GCMs require considerable computational resources, thereby limiting the number of tests which can be performed. For reasons such as these, SCMs have become popular for testing parameterisations outside the full GCM.

[3] As the name implies, a SCM is a model containing a single column of grid points. As such, SCMs are only capable of modeling processes in the vertical dimension.

Information about dynamical processes in the horizontal dimensions must be provided to the model as prescribed forcing data. There are several methods for specifying these data, as discussed by *Randall and Cripe* [1999]. One commonly used method is horizontal advective forcing. In this method, the horizontal advective tendencies of specific humidity, temperature and momentum, or equivalent variables, are prescribed. Additionally, it is necessary to prescribe the large scale vertical velocity and the pressure gradient or geostrophic wind field. This is sufficient information to tell the model what the effect of the large scale dynamical processes will be. Obviously the model also needs the initial temperature, humidity and wind profiles throughout the atmosphere. Typical data requirements for running a SCM using this method are summarised in Table 1. It is possible, and may be useful, to initialise SCMs with additional variables such as the cloud liquid and ice mixing ratios, and force the models using the advective tendencies of these variables. Unfortunately, these variables are not often observed. Therefore they are not listed in Table 1. A common approach to overcome this problem, which is adopted in this study, is to focus on SCM predictions greater than about 12 hours. This allows the model sufficient time to develop reasonable cloud representations.

[4] Most previous studies have derived the required forcing data from observations [e.g., *Randall and Cripe*, 1999]. An objective analysis method, such as the scheme described by *Zhang and Lin* [1997], is used to derive the advective tendencies and vertical velocity at the site where the SCM is

Table 1. Minimum Data Requirements for a Typical SCM Run Using Horizontal Advective Forcing

Initial Conditions	Forcing Data
Temperature sounding	Advective tendency of temperature
Specific humidity sounding	Advective tendency of specific humidity
U and V wind sounding	Advective tendency of momentum
	Vertical velocity
	Pressure gradient (or geostrophic winds)
	Surface characteristics (elevation, albedo etc)

being run. Unfortunately, there are few locations where there is a sufficiently dense observational network to derive the forcing data. When sufficient observations do exist, they are usually part of a short duration observational program, such as the Atmospheric Radiation Measurement (ARM [Ackerman and Stokes, 2003; Stokes and Schwartz, 1994]) program's Intensive Operational Periods (IOPs).

[5] Another method is to derive the forcing data from a CRM. For example, Bechtold *et al.* [2000] derived forcing data for a GEWEX (Global Energy and Water-cycle EXperiment) Cloud System Study (GCSS) SCM inter-comparison from a three dimensional CRM. This method is useful for situations when the available observations are insufficient to derive accurate forcing data. However, often CRM data are not available, or they are computationally prohibitive to generate. Therefore, several researchers have used standard Numerical Weather Prediction (NWP) products, constrained by surface and top of the atmosphere observations, to derive the SCM forcing data [e.g., Xie *et al.*, 2003, 2004].

[6] A variation on the previous approach, which is adopted in this paper, is to derive the required forcing data solely from NWP products. There are several advantages to this method. Firstly, forcing data can be derived anywhere where NWP analyses are available. Additionally, it is easy to derive forcing data for long continuous periods. Some authors [e.g., Emanuel and Živković-Rothman, 1999] argue that in order to test convective schemes, it is necessary to make fully prognostic model runs over periods of twenty or more days. This is very difficult to achieve with forcing data derived from observations, as most intensive observational campaigns are of limited duration.

[7] On the other hand, it is well known that NWP analyses contain considerable uncertainties, especially in relatively data sparse regions, such as the tropical Pacific. This is of concern for SCMs, which are sensitive to uncertainties in the initial conditions and external forcing data, as shown by Hack and Pedretti [2000]. By adding small random perturbations to the initial conditions, they found clear bifurcations in the SCM solutions. This led them to suggest that an ensemble approach was desirable when using SCMs. Based on the above considerations, it

seems that an ensemble single column modeling (ESCM) approach is also likely to be beneficial when using forcing data derived solely from NWP analyses.

[8] The first objective of this paper is to describe the derivation from NWP analyses, and validation against observations, of an ensemble of four forcing data sets at the ARM Tropical Western Pacific (TWP) sites of Manus Island and Nauru during 1999 and 2000. Since 1998, high quality, continuous observations have been available from these sites. Unfortunately, being isolated islands, the radiosonde network is insufficient to enable SCM forcing data to be derived from observations. Therefore, apart from some studies using TOGA-COARE (Tropical Ocean/Global Atmosphere Program: Coupled Ocean-Atmosphere Response Experiment) data [e.g., Bechtold *et al.*, 2000], there has been relatively little SCM research in this region. The preceding considerations make the TWP an ideal location to test SCM forcing data derived from NWP analyses.

[9] The second part of the paper investigates the use of the NWP derived forcing data in the European Centre for Medium Range Weather Forecasting (ECMWF) and Australian Bureau of Meteorology Research Centre (BMRC) SCMs. The ECMWF SCM is the single column version of the model used for the ERA-40 reanalysis [Uppala, 2002]. The BMRC SCM is described by Roff [2004]. The two questions which are addressed are: how sensitive are the SCMs to uncertainties in the initial conditions and forcing data?; and can an ESCM approach be used to distinguish errors caused by the physical parameterisations from errors caused by uncertainties in the initial conditions and forcing data?

2. Forcing Derivation for ESCM

[10] As mentioned above, SCM forcing data were derived for Manus Island and Nauru from four NWP analyses; namely ERA-40 reanalyses, ECMWF operational analyses, analyses from the Australian Bureau of Meteorology Global Analysis and Prediction System (GASP) and NCEP R-2 reanalyses. Descriptions of the four NWP sources can be found in Uppala [2002], Gregory *et al.* [2000], Seaman *et al.* [1995], and Kanamitsu *et al.* [2002], respectively. The main features of each model are also summarised in Table 2.

[11] Despite having comprehensive archives of the aforementioned NWP analyses, the forcing data requirements for SCMs are still rather onerous. For example, the advective tendencies of temperature, humidity and momentum are not archived, nor are the pressure gradients or geostrophic winds. These quantities obviously have to be derived from quantities which have been archived. Fortunately, there are some simplifications that can ease the process of deriving the forcing terms, yet not significantly affect the utility of the SCMs for the purposes of this study.

Table 2. NWP Analyses Used to Derive the SCM Forcing Data

	ERA-40	ECMWF	GASP	NCEP
Resolution	2.5° lat-lon.	0.5° lat-lon.	0.75° lat., 0.5° lon.	2.5° lat-lon.
Times, UTC	00:00, 06:00, 12:00, 18:00	00:00, 06:00, 12:00, 18:00	23:00, 05:00, 11:00, 17:00	00:00, 06:00, 12:00, 18:00
Levels	1000–1 hPa 23 pressure + surface	1000–10 hPa 15 pressure + surface	Surface–10 hPa 29 sigma	1000–10 hPa 15 pressure + surface

Figure 1. Derivation of SCM forcing data from NWP analyses. The locations of Manus Island (M: 2.058°S, 147.425°E) and Nauru (N: 0.521°S, 166.916°E) are also shown.

[12] Firstly, it is standard practice to avoid deriving advective tendencies of momentum by nudging the SCM with the two analysed (or observed) horizontal wind components, as described by *Randall and Cripe* [1999]. A consequence of doing this is that the SCM loses any predictive power for momentum. Considering that the main focus of this study is on cloud and radiative processes, this loss of predictive power is considered acceptable. Therefore, the SCM was nudged with the six hourly analysed horizontal wind components throughout this study.

[13] Another consequence of nudging the SCM with the horizontal wind components is that the pressure gradient or geostrophic wind term becomes less of an issue. Therefore, this term was set to zero. In any case, the ARM sites are close to the equator (Manus Island is at 2.058°S and Nauru is at 0.521°S), and consequently the coriolis term is small. The small coriolis term makes it difficult to accurately calculate the geostrophic wind, because numerical considerations can cause significant uncertainties.

[14] The process for deriving SCM forcing data is illustrated in Figure 1. Forcing data were derived for the entire 1999 and 2000 period. As alluded to earlier, this period was primarily chosen based on the availability of the NWP analyses and suitable validating observations. The temporal resolution of the forcing data is limited to six hours by the availability of the NWP analyses. Previous studies show that, in the absence of strong diurnal cycles in the forcing data, this is adequate for SCMs [e.g., *Emanuel and Živković-Rothman*, 1999; *Randall and Cripe*, 1999; *Xie et al.*, 2003].

[15] Firstly, each of the NWP analyses are regridded, if necessary, onto a 2.5° × 2.5° grid covering the tropical

Pacific (20°S–20°N and 110°E–80°W). 2.5° is the lowest resolution of the NWP analyses used in this study, it is also a typical resolution for current GCMs. When the NWP resolution is higher than 2.5° (e.g., the ECMWF operational analyses are on a 0.5° grid), it is reduced to 2.5° by averaging.

[16] Despite the simplifications described earlier, the archived NWP analyses still do not include all the fields required to run a SCM (refer to Table 1). Therefore, it is necessary to derive some terms from variables which are archived. The required derivations are performed after the data have been remapped to the 2.5 degree grid, so as to minimise the effect of different model resolutions. In particular, none of the analyses include the advective tendencies of temperature and specific humidity. These are derived by

$$A_T = -\mathbf{V} \cdot \nabla T \quad (1)$$

$$A_q = -\mathbf{V} \cdot \nabla q \quad (2)$$

where T is the temperature, q is the specific humidity, $\mathbf{V}$ is the horizontal wind velocity and A_T and A_q are respectively the horizontal advective tendencies of temperature and specific humidity.

[17] The gradients in equations (1) and (2) are calculated using a cubic spline method. Other numerical techniques for calculating the gradients, such as centred differences, were also tested. It was found that the differences between the advective tendencies calculated using different numerical techniques were less than the differences between tenden-

Table 3. Observations Used in This Study

Observation Type	Frequency	Notes
Radiosonde	12 hourly	Reduced frequency before May 99 at Nauru Reduced frequency before March 99 at Manus Relative humidity profiles corrected for dry bias Observations contaminated by rainfall removed.
MWR	Hourly	Observations of temperature, rain etc. Starting in June 1999.
Surface Met.	Hourly	Quantities such as outgoing long wave radiation, cloud top temperature, cloud cover and optical thickness derived from GMS-5 IR and VIS channels.
GMS-5	Hourly	

cies calculated using different NWP data sets. Furthermore, SCM runs were not significantly affected by the numerical technique used to derive the gradients.

[18] Finally, it is obvious that an SCM will only require forcing data from a single column of the large tropical Pacific domain described previously. For the purposes of this study, the columns of grid points closest to Manus Island and Nauru were extracted from each analysis.

3. Validation of Initial Conditions and Forcing Data

[19] From the description of the data derivation in the preceding section, it is clear that the initial condition and forcing data are likely to contain uncertainties which have a significant impact on the SCM runs. Therefore, this section will focus on quantifying these uncertainties.

[20] There are a number of possible sources of uncertainties in the initial condition and forcing data. These include uncertainties in the NWP analyses used to derive the forcing terms, and uncertainties resulting from the numerical techniques used in the derivation. Furthermore, it must be borne in mind that the NWP analyses are on a 2.5° grid, so there will be issues with how representative these data are at the point locations where the SCMs are run. It is worth noting that the mountainous terrain of Papua New Guinea is only two or three grid points away from Manus Island. This may affect, for example, the advective tendency derivations, where the values of the surrounding grid points are used to calculate the temperature and specific humidity gradients.

[21] Obviously, validation of the initial condition and forcing data is restricted by the availability of suitable observations. The observations used for validation are briefly described below. As discussed earlier, there are insufficient observations at the TWP ARM sites to derive all the forcing terms. A corollary of this is that direct validation of all the NWP derived forcing terms with observations is not possible. On the other hand, there are observations of the initial condition data required by the SCMs. It is therefore convenient to split the validation into two parts. Firstly, the initial condition terms are directly validated against observations. In the second part, indirect methods are used to validate some of the most important forcing terms, namely the vertical velocity and the horizontal advective tendencies of temperature and specific humidity.

3.1. Observations Used for Validation

[22] The observations used for validation are summarised in Table 3. All the surface and radiosonde observations were

made at the ARM sites [Mather *et al.*, 1998]. Quality control checks were applied to remove low quality data.

[23] The radiosondes used at the TWP ARM sites during 1999 and 2000 were Vaisala RS-80H models. These are known to produce a dry bias of as much as 10% [e.g., Revercomb *et al.*, 2003; Wang *et al.*, 2002]. To remove this bias, the water vapor profiles were calibrated against the microwave radiometers (MWR) at Manus Island and Nauru using the method described by Turner *et al.* [2003].

[24] In addition to the ARM data, various cloud and radiation products derived from GMS-5 (Geostationary Meteorological Satellite) infrared (IR) and visible (VIS) imagery were available, starting in June 1999. These data were produced using the method of Minnis *et al.* [1995] as reported by Nordeen *et al.* [2001].

[25] It is worth noting that validation of the NWP data will also be affected by uncertainties inherent in the observations. There are two components to these observational errors. Firstly, instrumental error is typically quite small, except in the case of humidity. Lesht [2004] reports uncertainties of about 0.2 m s^{-1} and 0.2 K for the TWP radiosonde observations of wind speed and temperature respectively. As described above, the radiosonde humidity data were corrected for a known dry bias using MWR data. This correction may introduce uncertainties associated with the MWR instrument. Nevertheless, it is hoped the errors in the humidity data are minimised, and will be less than the original dry bias of up to 10%.

[26] The second component of observational uncertainty arises because the soundings are essentially point measurements, whereas the NWP data are representative of a 2.5 degree grid box, and are unable to resolve the spatial variability which affects the point observations. Mapes *et al.* [2003] attempted to quantify the unresolved spatial variability for an array of radiosondes in the TWP. The standard deviations of unresolved variability for temperature, wind speed and relative humidity soundings made from two ships about 200 km apart were respectively about 0.5 K , $2\text{--}3 \text{ m s}^{-1}$ and $5\text{--}15\%$. Although Mapes *et al.* [2003] were concerned with an array of soundings, the unresolved variability they measured is probably similar to the uncertainty in comparing our grid point data with soundings.

3.2. Initial Condition Validation

3.2.1. Horizontal Wind Components

[27] The horizontal wind components are important not only as initial conditions for the SCMs, but also as nudging terms throughout the runs, as described earlier. Figure 2 shows the bias and RMSE for the west-east (U) and south-

Figure 2. Bias (dotted lines) and RMSE (dashed lines) of U (top panels) and V (bottom panels) forcing data at Manus Island (left panels) and Nauru (right panels). The solid lines show the mean bias for all the models. Validations were made using all available radiosondes wind observations during 1999 and 2000.

north (V) wind component profiles at Manus Island and Nauru. Most of the models are quite good at analysing the wind components, with the biases being close to 0 m s^{-1} throughout most of the troposphere, and RMSE scores between about 2 m s^{-1} and 5 m s^{-1} . The exception is the GASP model, which exhibits significantly larger RMSE scores in the middle and upper troposphere.

[28] A slight negative bias is seen in the wind speed near the tropopause (between about 200 hPa and 100 hPa). A similar bias was also observed by *Nagarajan and Aiyyer* [2004] in ECMWF analyses over the Tropical Indian Ocean. A number of “steps” can be seen in the RMSE scores, particularly for the U wind component at Manus Island, but also to a lesser extent in the other plots. The cause of these is not known.

3.2.2. Temperature

[29] Figure 3 shows the bias and root mean square error (RMSE) for the initial temperature profiles at Manus and Nauru, compared to radiosonde observations. The mean bias is quite small for all the temperature profiles, except between about 250 hPa and 100 hPa, where all the analyses exhibit a well documented warm bias [e.g., *Zhou et al.*,

2001; *Randel et al.*, 2000; *Nagarajan and Aiyyer*, 2004]. In the case of the ECMWF model, *Gregory et al.* [2000] suggest the warm bias is partly caused by factors related to the model’s representation of cloud ice in the upper troposphere.

3.2.3. Specific Humidity

[30] Figure 4 shows the bias and RMSE of the specific humidity initial conditions, compared with radiosonde observations which were corrected using the method described earlier. At Manus Island, the NCEP and GASP data have a dry bias throughout the troposphere. In contrast, the ERA-40 and ECMWF analyses do not exhibit much bias, except near the surface, where there is a slight dry bias. The biases at Nauru tend to be smaller than at Manus Island, except near the level of the trade inversion, where all the analysis data exhibit a moist bias.

[31] The net effect of the biases described above are summarised in Table 4, which shows the integrated water vapor total from the forcing data compared with MWR observations. Overall, the NCEP forcing data at Manus Island are more than 15% too dry, and have a very large RMSE compared to the other forcing data. Although the

Figure 3. Bias (dotted lines) and RMSE (dashed lines) of temperature data at Manus Island (left) and Nauru (right). The solid line is the mean bias for all the models. Validations were made using all available radiosonde temperature observations during 1999 and 2000.

bias for the NCEP forcing data at Nauru is not as large as at Manus, the RMSE remains quite large compared to the other forcing data. *Trenberth and Guillemot* [1998] found the NCEP R-1 reanalyses had a dry bias that sometimes exceeded 5 kg m^{-2} in the Indonesian and New Guinea region, which includes Manus Island. They attributed the relatively large uncertainties in water vapor fields at least partially to water vapor information from satellite systems such as SSM/I (Special Sensor Microwave Imager) and TOVS (TIROS Operational Vertical Sounder) not being utilised in the reanalysis. While the Tropical Western Pacific dry bias is slightly less in the R-2 reanalysis than in the R-1 reanalysis [Kanamitsu *et al.*, 2002], it seems likely that the uncertainties in the NCEP water vapor forcing data used here are at least partially caused by the analysis not using satellite data. In contrast to the NCEP reanalyses, all the

other analyses used here incorporate water vapor information from either TOVS or SSM/I, or both.

3.3. Forcing Data Validation

3.3.1. Vertical Velocity

[32] Vertical velocity is difficult to validate, because it is not directly observed, and can not be derived from observations made at the TWP sites. Figure 5 shows the average vertical velocity (in Pa s^{-1}) from each forcing data set at Manus Island and Nauru. Most data show average ascent at Manus Island and descent at Nauru. This is because, in the absence of an El Niño, Manus Island is located in the convectively active region of the TWP, while Nauru is located in the mostly suppressed central Pacific.

[33] There are significant differences between the forcing sets. For example, the GASP forcing set shows more ascent

Figure 4. Bias (dotted lines) and RMSE (dashed lines) of specific humidity data at Manus Island (left) and Nauru (right). The solid line is the mean bias for all the models. Validations were made using all available corrected radiosonde data (as described in the text) during 1999 and 2000.

Table 4. Bias and RMSE of the Total Column Water Vapor Derived From the Forcing Data, Compared With Observations by Microwave Radiometers at Manus Island and Nauru^a

Data Set	Manus Island			Nauru		
	Mean	Bias	RMSE	Mean	Bias	RMSE
ERA-40	51.1	−0.5	4.5	43.1	1.6	4.5
ECMWF	51.5	−0.1	4.2	43.4	1.9	4.9
GASP	47.5	−3.9	7.1	42.4	1.0	5.3
NCEP	43.1	−8.5	12.3	40.2	−1.3	9.3
Observed	51.6			41.5		

^aUnits are in kg m^{-2} .

than the other forcing sets at both Manus Island and Nauru. Results described later suggest that the GASP forcing data probably show too much ascent at both Manus Island and Nauru. At Nauru, the ERA-40 and ECMWF forcing data are similar, but at Manus Island the ERA-40 forcing data show more subsidence than the ECMWF forcing data. It is unknown why the ERA-40 and ECMWF operational models differ, but it could be partially a result of the different resolutions that the models are run at. Additionally, there were changes to the operational model during the two year period studied here, which may account for some of the differences. Finally, the ECMWF operational model may not have had as much data available for assimilation as the ERA-40 model, because of operational time constraints.

[34] Vertical velocity is, at least partially, correlated to outgoing long wave radiation (OLR) [Charlock *et al.*, 1989]. In the tropics this is due to strong ascent in deep convection and the associated cold (low OLR) anvils. Figure 6 shows the correlation coefficients for the vertical velocity forcing data and instantaneous satellite OLR measurements at Manus Island and Nauru. While it is clear that OLR is correlated to vertical velocity, especially in the 250–400 hPa layer, the correlation coefficient is not par-

ticularly large. In the tropics, cirrus may extend over large regions, even where there is no active convection [Fu *et al.*, 1990]. This high cirrus can lower the OLR significantly, and may explain the relatively poor correlation between OLR and vertical velocity at Manus Island and Nauru. Nevertheless, the NCEP vertical velocities are so poorly correlated with OLR at all levels, that this may indicate a high level of uncertainty in the vertical velocity forcing data from this source.

[35] Jakob *et al.* [2005] compared vertical velocity from the four analysis data sets at Manus Island for each of four cloud regimes they found in the TWP. Their cloud regimes were derived from satellite optical thickness and cloud top pressure data. Two of the regimes were convective, and two were suppressed. All the four data sets showed more ascent in the convective regimes than in the suppressed regimes. However, only the ERA-40 vertical velocity data showed average ascent in the convective regimes and descent in the suppressed regimes. The NCEP and GASP data showed ascent in all four of their regimes. In particular, the GASP data showed quite large ascent values, even in the suppressed regimes. These results are consistent with the findings presented here, and suggest the GASP and NCEP data have larger uncertainties than the ERA-40 and ECMWF data.

3.3.2. Advective Tendencies

[36] There are no observations of the horizontal advective tendencies of temperature and specific humidity. However, as described earlier, these forcing terms are derived from temperature, humidity and wind analyses, which have known uncertainties. Therefore, it is possible to calculate the uncertainties in the advective tendencies from the uncertainties in the terms used to derive them.

[37] A Monte Carlo technique was used to estimate the uncertainties in the advective tendencies. The process involved simulating uncertainties in the base terms used

Figure 5. Average vertical velocity data at Manus Island (left) and Nauru (right).

Figure 6. Correlation between OLR and vertical velocity at Manus Island (left) and Nauru (right).

to derive the advective tendencies (temperature, specific humidity and wind components) with random, but known, Gaussian error terms. The resulting errors in the calculated advective tendencies were then noted. The standard deviations of the Gaussian error terms were based on the RMSE scores for temperature, specific humidity and wind reported in the previous sections. This procedure was repeated a large number of times, to obtain estimates of the error distributions for the advective tendencies.

[38] Table 5 shows the uncertainties for the 1000 hPa temperature and specific humidity advective tendencies at Manus Island and Nauru. The mean temperature and specific humidity, and mean temperature and specific humidity advective tendencies are also shown. It must be noted that the uncertainties shown in the table are only approximate estimates. The actual uncertainties at any time will vary according to the values and uncertainties in the base variables used to derive the advective tendency terms.

[39] The first important point to note is that the mean advective tendencies are small compared to the mean

temperature and specific humidity. This is because the gradients of temperature and specific humidity tend to be quite small in the tropics. The second point to note is that the relative uncertainties in the advective tendencies are large. The main source of these large uncertainties is the gradient calculation. The uncertainties in temperature and specific humidity at adjacent grid points are comparable to the differences between the values at adjacent grid points. Consequently the gradient derivation contains a large uncertainty, irrespective of which numerical technique is used to derive it. However, because the advective tendencies are small in the tropics, the absolute uncertainties in these terms are not extreme.

[40] Finally, it should be noted that the uncertainties calculated in this section do not represent a systematic bias. Although it is possible to get excessively large or small advective tendencies at any particular time during a SCM run, it is unlikely that every forcing term used during the complete SCM run will be either too large or too small. In the next section it will be shown that

Table 5. Mean Temperature and Specific Humidity (First Column); Mean Temperature and Specific Humidity Advective Tendencies (A.T. Mean: Second Column); and RMSE for the Temperature and Specific Humidity Advective Tendencies (A.T. RMSE: Last Column) at Manus Island and Nauru

Variable	Mean	A.T. Mean	A.T. RMSE
<i>Manus Island</i>			
1000 hPa Temperature	300 K	0.1 K day ⁻¹	4 K day ⁻¹
1000 hPa Specific humidity	17 g kg ⁻¹	-0.4 g kg ⁻¹ day ⁻¹	4 g kg ⁻¹ day ⁻¹
<i>Nauru</i>			
1000 hPa Temperature	300 K	-0.2 K day ⁻¹	4 K day ⁻¹
1000 hPa Specific humidity	17 g kg ⁻¹	-0.2 g kg ⁻¹ day ⁻¹	4 g kg ⁻¹ day ⁻¹

uncertainties in the SCM results arising from uncertainties in the advective tendency terms are not as large as uncertainties arising from uncertainties in the vertical velocity forcing term.

4. Assessment of ESCM Uncertainties

[41] In the preceding section it was shown that the forcing and initial condition data derived from NWP analyses contain significant uncertainties. These uncertainties are both model and variable dependent. For example, while the GASP specific humidity data have a smaller bias and RMSE than the NCEP data, the GASP wind data have a larger RMSE than the NCEP model. In some cases it is difficult to quantify the uncertainties. For example, it is clear that some or all of the models contain large vertical velocity uncertainties. However, it is not obvious which of the models is most accurate.

[42] Difficulties such as those described above suggest that it may not be feasible to choose a single “best” NWP forcing for use by the SCM. As mentioned earlier, an alternative approach is to run an ensemble of SCMs using initial condition and forcing data derived from the various NWP analyses. One of the benefits of this approach is that an ensemble of SCM runs forced with different NWP data is likely to contain more information than any single SCM run. Indeed, a multi-model ensemble might outperform an ensemble constructed by adding perturbations to a single forcing data set. For example, *Ziehmann* [2000] showed that a four member multi-model ensemble performed better in most forecast aspects than the larger single-model ECMWF ensemble prediction system.

[43] The issue of uncertainty in the initial conditions and forcing data is not restricted to forcing data derived from NWP products. Even when forcing data are derived from observations, they contain errors, including instrumental errors and errors due to the representativeness of the observational array [*Mapes et al.*, 2003]. As *Hack and Pedretti* [2000] showed, SCMs are sensitive to even quite small uncertainties in the initial conditions and forcing data. Therefore, they also suggested an ESCM approach was useful.

[44] We now turn our attention to how uncertainties in the initial condition and forcing data affect the SCM runs. Errors in SCM results can be attributed to three causes; errors in the initial conditions, errors in the forcing data, and model errors. In this context, model errors refer to a range of possible errors, including errors in the physical parameterisations and numerics of the SCM. Since one of the main purposes of running SCMs is to test physical parameterisations, and identify model errors, it is helpful to quantify the SCM uncertainties which are caused by other factors, including uncertainties in the forcing data and initial conditions. This will be investigated next.

4.1. Initial Versus Boundary Condition Uncertainties

[45] To quantify the relative influence of uncertainties in the initial and boundary conditions (forcing data), two SCM ensembles were run at Manus Island. The first ensemble consisted of runs started every six hours (at 00:00, 06:00, 12:00 and 18:00 UTC) during January and February 1999.

Figure 7. Standard deviation of SCM ensemble predictions of total column water vapor (TCWV) at Manus Island. The ensembles were constructed from model runs using four different initial conditions and ERA-40 forcing data (thin lines), and model runs using ERA-40 initial conditions and four different forcing data sets (thick lines).

At each time, four separate runs were made. Each run was initialised with ERA-40 initial conditions, but was forced with either ERA-40, ECMWF, GASP or NCEP data. Therefore, the total number of model runs in this ensemble was 944 (four models started four times per day for fifty eight days). The second ensemble also contained 944 SCM runs, but in this case each run was forced with ERA-40 data and initialised with either ERA-40, ECMWF, GASP or NCEP data. The whole experiment was carried out using both the ECMWF and BMRC SCMs.

[46] Figure 7 shows the standard deviation of the total water vapor column as a function of the forecast time for the ECMWF and BMRC SCMs. Firstly, it is worth noting that the ensemble standard deviation gets quite large very quickly. To put these values into context, the average total water vapor column at Manus Island, as measured by MWR is 52 kg m^{-2} .

[47] Secondly, the standard deviation for the ensembles with different forcing conditions exceeds the standard deviation of the ensemble with different initial conditions after about 18 hours. This suggests that the influence on the SCM of uncertainties in the forcing data exceeds the influence of uncertainties in the initial conditions after less than one day. It is also interesting to note that the standard deviation of the ensemble with only one source of forcing data gets smaller with increasing forecast time. It appears that the initial conditions continue to have an influence on the forecasts, even for forecasts as long as six to nine days, but this influence gradually diminishes. This finding is consistent with the results of *Hack and Pedretti* [2000]. They showed that by varying the initial conditions, bifurcations in the SCM results were often obtained, but these collapsed after a number of days, suggesting the model was equilibrating around a state determined by the forcing data alone.

[48] Finally, it is worth commenting that the curves in Figure 7 for the BMRC and ECMWF SCMs are similar.

Table 6. Time Taken (in Hours) for the Standard Deviation of the SCM Runs With Various Forcing Data but the Same Initial Conditions to Exceed the Standard Deviation of Runs With Various Initial Conditions but the Same Forcing^a

Variable	ECMWF	BMRC
Total water vapor column	18	20
500 hPa specific humidity	18	19
850 hPa specific humidity	16	15
500 hPa temperature	13	10
850 hPa temperature	12	12
Clear sky OLR	12	11
Total scene OLR	7	7
Convective rain	7	6
Stratiform rain	5	5

^aThe table has been organized so that the longest times are at the top.

This suggests that the responses of these different SCM models to uncertainties in the initial conditions and forcing data are similar. To highlight this point further, Table 6 shows the number of hours taken until the standard deviation of the ensemble with ERA-40 initial conditions and various forcing data sources exceeds the standard deviation of the ensemble with only ERA-40 forcing but different initial conditions, for a variety of variables. The times for both models are quite similar. This is a useful result, because it raises the possibility that the sensitivity of other SCMs to uncertainties in the initial condition and forcing data may be similar to the two SCMs studied here.

[49] Another interesting point to note in Table 6 is that the variables most heavily influenced by moist processes are more rapidly affected by uncertainties in the forcing data. For example, uncertainties in the forcing data appear to have more influence on the rainfall predictions than uncertainties in the initial conditions for predictions longer than about six hours. This further highlights the important influence of the forcing data on the SCM results. It also

raises the question of the relative influence of the various forcing terms on the SCM predictions. This will be investigated in the next section.

4.2. Sensitivity of ESCM to Forcing Terms

[50] As noted above, after less than one day the SCM runs are predominantly influenced by the forcing data. It is useful to investigate the relative influence of the various forcing terms on the SCM predictions. To do this, the ECMWF SCM was run in four configurations every six hours at Manus Island during January and February 1999. The first model configuration was a “standard” model run, utilising all the forcing terms (vertical velocity, and horizontal advective tendencies of temperature and specific humidity). The second configuration had these forcing terms turned off (effectively the model was run with the vertical velocity and advective tendencies set to zero). The third configuration had the horizontal advection forcing turned off, and the final configuration had the vertical velocity forcing turned off.

[51] Figure 8 shows the average forecasts of OLR from the aforementioned SCM configurations. The results from the model configuration which only uses horizontal advection forcing (that is, no vertical velocity forcing) are similar to the results from the model runs with no forcing. In contrast, the results for the model configuration which only uses vertical velocity forcing are similar to the results for the “standard” SCM runs (where all the forcing terms are used). This suggests that the vertical velocity has more influence on the SCM than the advective forcing terms. This is not surprising, because the advection of temperature and moisture in the tropics tends to be quite small.

[52] As shown in Figure 8, the vertical velocity forcing acts to reduce the OLR. Table 7 shows that this is mostly due to larger amounts of high cloud cover in the model configurations which include vertical velocity forcing. This is expected, because Manus Island is in a convectively

Figure 8. Average outgoing longwave radiation (OLR) at Manus Island during January and February 1999 for various configurations of the ECMWF SCM.

Table 7. Average High, Mid, and Low Level Cloud Cover at Manus Island During January and February 1999 for 12–72 Hour Forecasts From Each of the Four Model Configurations^a

Model Configuration	Low Cloud	Middle Cloud	High Cloud
All forcing terms	35%	20%	80%
No advection or vertical velocity forcing	40%	20%	50%
No horizontal advection forcing	35%	20%	75%
No vertical velocity forcing	30%	15%	50%

^aCloud cover values are rounded to the nearest 5%.

active region. Including the vertical velocity forcing triggers convection in the SCM. As described earlier (Figure 5), vertical velocities derived from the GASP analyses show larger average ascent values than the other NWP models. This is reflected in the lower OLR values for the model configurations which included GASP vertical velocity forcing data.

5. Use of the ESCM to Identify Model Errors

[53] The foregoing results highlighted the sensitivity of the SCM to initial conditions and forcing data, especially vertical velocity. However, it must be borne in mind that SCM model errors (for example, errors in physical parameterisations and numerics) could also be causing some of the uncertainties. In fact, the forcing which leads to SCM predictions which are closest to observations may not be the “best” forcing, because it is possible that uncertainties caused by the forcing data simply cancel the model errors. Considering that one of the major aims of single column modeling is to identify model errors, it would be wrong to restrict the study to using SCM and forcing configurations which make the best predictions with respect to observations.

[54] The aim of this section is to ascertain if model errors can be distinguished from errors which are caused by uncertainties in the forcing and initial condition data. There are two aspects to this problem. Firstly, we simply need to show that model errors can be distinguished from errors resulting from uncertainties in the initial condition or forcing data. If this is not possible, then there is little prospect that the SCMs can be used to test physical parameterisations. However, identifying model errors in the SCMs is not sufficient if one of the main motivations of running SCMs is as a test bed for GCM parameterisations. This leads to the second problem, which is to show that errors observed in the GCM can also be reproduced in the SCM. The final part of this section presents some results which suggest that it is indeed possible to reproduce errors seen in the GCM in SCMs.

[55] To test if it is possible to identify model errors in the SCMs (irrespective of whether these errors are observed in the GCM or not), two ensembles were constructed. The first ensemble was comprised of 12 hour forecasts from sixteen ECMWF SCM runs at Manus Island. Twelve hour forecasts were chosen to allow the models time to “spin up” and develop clouds. The sixteen SCM runs were initialised and forced with various combinations of initial and forcing data. For example, one ensemble member was initialised and forced with ERA-40 data, the next was initialised with ERA-40 data and forced with NCEP data and so on. New

model runs were started every six hours. The configuration of the second ensemble was similar to the first one, but was comprised of BMRC SCM runs instead of ECMWF SCM runs.

5.1. Ensemble Predictions of Temperature and Specific Humidity

[56] It is useful to focus on the ensemble predictions of prognostic variables, such as temperature and specific humidity. Model errors which affect the SCM’s prediction of these variables will obviously also affect the prediction of diagnostic quantities such as OLR.

[57] As an example, Figure 9 shows a time series of the 12 hour forecasts of 500 hPa temperature at Manus Island for each ECMWF SCM ensemble member during the period 2–13 May 2000. Also drawn on the figure are the values one standard deviation either side of the ensemble mean, and the observed 500 hPa temperature.

[58] It is clear that there is a large variation between the different ensemble members, presumably as a result of uncertainties in the initial conditions and forcing data. However, despite this, a number of interesting points can be noted. Firstly, the observed temperature almost always lies in the range of values covered by the ensemble, and usually within one standard deviation of the ensemble mean. Secondly, there are periods when it appears that the ensemble is more skillful at predicting the temperature. For example, the spread of the ensemble is significantly less during the period 6–9 May than in the preceding and following days.

[59] Referring to Figure 9, it is not immediately obvious if model errors can be distinguished from errors caused by uncertainties in the initial conditions and forcing data. To filter out the large amount of noise in the ensemble predictions, the ensemble mean and its associated statistics (bias and RMSE) were investigated. The statistics described in this section were calculated from all model runs during 2000 when observations were also available. The model runs were restricted to this year because the observational data are more complete than in 1999.

[60] Figures 10 and 11 show the bias of the ensemble mean prediction of temperature and specific humidity from the ECMWF and BMRC SCMs. The mean bias of the initial condition data for the same times as the 12 hour SCM predictions are valid is also shown for comparison. A number of interesting points can be seen. For example, the temperature biases for both SCM ensembles are quite small throughout most of the troposphere, except near the tropopause. The ECMWF SCM bias is closer to the initial condition bias than the BMRC SCM bias. This latter point is interesting, because it suggests the model bias (as

Figure 9. Time series of 12 hour forecasts of 500 hPa temperature at Manus Island from an ensemble of ECMWF SCM runs during May 2000 (thin grey lines). The dashed lines show the values one standard deviation either side of the ensemble mean, and the solid black line shows the observed temperature.

opposed to biases in the initial condition or forcing data) is larger for the BMRC SCM than the ECMWF SCM.

[61] The bias results for the specific humidity predictions have quite different characteristics than for the temperature. Firstly, as was shown earlier, the average of the four initial condition data sets for Manus Island exhibits a reasonably large negative bias in the low to middle troposphere. Not

Figure 10. Bias for the 12 hour ensemble mean prediction of temperature at Manus Island. The dotted line shows the bias for the ECMWF SCM and the dashed line shows the bias for the BMRC SCM. The solid line indicates the mean bias of the forcing data sets for the times when the 12 hour ensemble predictions are valid.

Figure 11. As for Figure 10, except for the specific humidity biases.

surprisingly, this bias carries over into the results for both the ECMWF and BMRC SCMs. In fact, the difference between the SCM bias (for 12 hour predictions) and the initial condition bias tends to be less than the initial condition bias alone. This is consistent with the results in Table 6 which show it takes up to 19 hours for uncertainties in the forcing data to have a greater influence than uncertainties in the initial conditions on the 500 and 850 hPa specific humidity SCM predictions.

[62] While the preceding results are informative, it is important to note that the differences between the biases for the SCM predictions are quite small. For example, the 12 hour predictions of 500 hPa temperature have biases of 0.7 K and 0.4 K for the BMRC and ECMWF SCMs respectively. To put this into context, the RMSE scores for the predictions from the two SCMs are 1.2 K and 0.8 K respectively. This raises the question of whether the differences between the biases shown in Figures 10 and 11 are statistically significant. This will be addressed in the next section.

5.2. Statistical Analysis of ESCM Results

[63] In the preceding section, results were presented which suggest the biases for predictions by the ECMWF and BMRC SCMs are different. If these differences are significant, then we have clearly identified an error in either the BMRC or ECMWF SCMs, or both, which is not directly caused by uncertainties in the initial condition or forcing data. This is because the initial condition and forcing data used in the two ensembles are the same, and any differences in the predictions are likely caused by differences in the SCM physics or numerics.

[64] It is possible to test if the differences in the biases shown in Figures 10 and 11 are statistically significant. Tables 8 and 9 summarise the results of t-tests comparing the mean ECMWF SCM bias to the mean initial condition

Table 8. Bias and RMSE for the Ensemble Mean of the 12 Hour BMRC and ECMWF SCM Predictions of Temperature^a

Level, hPa	N	Bias/RMSE, K			t-test		
		BMRC	ECMWF	IC	I-B	I-E	B-E
850	521	0.4/1.0	0.02/0.8	−0.2/0.7	11.7	4.5	7.4
700	514	0.5/1.1	0.3/0.9	−0.03/0.7	9.8	5.5	4.6
500	513	0.7/1.2	0.4/0.8	0.3/0.6	5.9	2.7	3.5
250	514	0.3/0.9	−0.3/0.8	−0.1/0.6	9.3	3.7	11.4
100	478	5.3/2.6	2.6/2.2	1.8/1.9	23.4	5.7	17.1

^aThe bias and RMSE for the mean of the four initial condition data (IC) sets are also shown for comparison. N is the population sample size. The last three columns show the t-values for t-tests comparing the mean initial condition bias (I), the mean ECMWF SCM bias (E), and the mean BMRC bias (B).

bias, the mean BMRC SCM bias to the mean initial condition bias, and the mean ECMWF SCM bias to the mean BMRC SCM bias. It is commonly accepted that t-values greater than about two indicate that the means are statistically different.

[65] At this point it is worth noting that a number of assumptions were required in order to perform the t-tests. For example, it was assumed the biases for the models and forcing data are normally distributed. Kolmogorov-Lilliefors tests [see, e.g., *Thiébaux*, 1994] showed that at many vertical levels the bias distributions were approximately normal. Visual inspection showed that those distributions which failed this test for normality, such as the ECMWF SCM 850 hPa temperature biases, were still reasonably close to normality for the purposes of this study. Additionally, there is likely to be some serial correlation in the data sets being studied. That is, the bias from one model run to the next will not be totally independent. This causes the “effective population size” of the data set to be somewhat less than the actual population size, as discussed by *Trenberth* [1985]. It is difficult to estimate the effective population size. However, the minimum time between successive members of the populations is twelve hours (because the radiosonde observations required to calculate the bias were only made twice per day). Furthermore, quantities such as temperature and specific humidity are significantly affected by convective processes with time scales less than twelve hours. Therefore, it is assumed that serial correlation will not be too serious a problem. In any case, the t-tests should give a reasonable first approximation of whether the model biases are different from the forcing data biases.

[66] There is a very clear difference between the biases for the ECMWF SCM 12 hour temperature predictions, the initial condition temperature data, and the BMRC SCM temperature predictions. It is conceivable that the differences between the SCM 12 hour temperature biases and the initial condition temperature biases are the result of uncertainties in the forcing terms, for example the advective

tendency of temperature. However, there is also a statistically significant difference between the ECMWF and BMRC SCM predictions. This implies that there is a detectable model error in at least one of the models, because the data used to initialise and force both models were the same. The BMRC SCM probably produces a larger model error than the ECMWF SCM, because its average bias is larger. While these results do not tell us what the cause of the model errors are, they do suggest that model errors can be isolated from errors arising because of uncertainties in the initial condition or forcing data.

[67] On the other hand, as was alluded to earlier, the differences between the specific humidity biases shown in Figure 11 are not always statistically significant. In particular, the difference between the ECMWF and BMRC biases are not statistically significant throughout the lower and middle troposphere. However, above approximately 250 hPa, the BMRC SCM is significantly moister than the ECMWF SCM. This difference is likely related to the treatment of upper tropospheric clouds in both models. The ECMWF SCM uses a prognostic cloud scheme, whereas the BMRC SCM uses a diagnostic cloud scheme. In the diagnostic scheme, condensate detrained from deep convection is evaporated into the upper model layers, producing relatively high humidity values, as can be seen in the BMRC column of Table 9. In contrast, when condensate is detrained in the ECMWF model, it can remain as cloud droplets, or precipitate into lower model levels [*Tiedtke*, 1993]. The effect of this treatment of upper tropospheric humidity has been shown by *Stephens et al.* [1998].

[68] The preceding discussion highlights the ability of the ESCM approach to identify model errors. It is possible to identify model errors in both the temperature and specific humidity predictions. In the case of specific humidity, the errors detected by the ensemble are consistent with findings of previous studies. While the cause of the temperature bias was not ascertained, the results showed that it is very likely due to a model error, and not the result of uncertainties in

Table 9. As for Table 8, Except Comparing the 12 Hour Forecasts of Specific Humidity

Level, hPa	N	Bias/RMSE, g kg ^{−1}			t-test		
		BMRC	ECMWF	IC	I-B	I-E	B-E
850	341	0.7/1.8	0.5/1.7	0.3/1.5	3.2	2.1	1.1
700	335	0.8/1.5	0.8/1.4	0.5/1.4	1.9	2.6	0.7
500	322	0.3/1.1	0.4/1.0	0.3/0.9	0.2	0.5	0.7
250	326	0.08/0.07	0.05/0.06	0.04/0.05	6.7	1.6	5.1
100	301	0.003/0.0007	0.001/0.0006	0.001/0.0006	27.4	0.8	27.9

Figure 12. Ratio of predicted downward solar radiation at the surface to clear sky downward solar radiation at Manus Island. Data are for 14 hour forecasts valid at 02 UTC for the complete 1999–2000 period. The whiskers show the 5th and 95th percentiles, the boxes show the upper and lower quartiles, and the median. The individual ensemble members are labeled according to which initial condition and forcing data were used: G = GASP, E = ECMWF, N = NCEP, R = ERA-40. For example, ensemble member EG was initialised with ECMWF data and forced with GASP data.

the initial condition or forcing data. Given that there are model errors in the temperature and specific humidity predictions, it is likely that there are also detectable model errors in some of the derived fields, such as OLR. This will be investigated in a later study.

5.3. Comparison of GCM Predictions With ESCM Predictions

[69] As discussed earlier, one of the main motivations for running SCM simulations is to test GCM parameterisations. It is therefore important that errors observed in the GCM can be reproduced in the SCM.

[70] To test if GCM errors can be reproduced using the ESCM approach, predictions of the down-welling solar radiation measured at the surface at Manus Island were investigated. The down-welling solar radiation was chosen because it should be a useful variable to test the model's representation of clouds and radiation. Importantly, continuous measurements of down-welling solar radiation are made at the ARM sites.

[71] Figure 12 shows box-whisker diagrams summarising predictions during 1999 and 2000 of the ratio of the down-welling solar radiation at the surface to the value that would be measured in clear sky conditions. Box-whisker diagrams are shown for 14 hour predictions valid at 02Z (approximately local noon at Manus Island) from the ECMWF ESCM, the individual ECMWF SCM ensemble members, and the ERA-40 GCM. The observed values are also shown. Finally, for comparison, the BMRC ESCM results are shown.

[72] The first important point to note in Figure 12 is that the ERA-40 model under-predicts the down-welling solar radiation. This was also noted by Jakob [2004]. Additionally, the model does not reproduce the spread observed in

the observations. It should be borne in mind that the model results are probably representative of a larger spatial area than the observations. This may partially explain the larger spread in the observations. However, this issue should have no impact on the comparisons between the various models, described below.

[73] It is obvious that there is a large variation in the results from the different SCM ensemble members. Some of the SCM runs clearly do not reproduce the errors observed in the GCM. For example, the SCM initialised with NCEP data and forced with ERA-40 data over-predicts the down-welling radiation. In this case, the dry bias in the NCEP data may be causing the model to under-predict cloud cover. Nevertheless, the ESCM mean is quite similar to the GCM results. This is encouraging, because it indicates that the ESCM approach may be able to reproduce errors observed in the GCM, even though some of the individual SCM simulations do not.

[74] Finally, it is worth comparing the BMRC-ESCM results with the ECMWF-ESCM and ERA-40 results. The BMRC-SCM quite clearly has different error characteristics than the other models. This is probably caused by differences in the physical parameterisations used in the models.

[75] The foregoing results highlight that the ESCM approach appears to be able to reproduce errors observed in the GCM, potentially making it useful for testing GCM parameterisations. However, the results do not explain which aspects of the model's parameterisations are causing the observed errors. This will be investigated in later work.

6. Conclusions

[76] This paper describes the use of initial condition and forcing data derived from four NWP analysis data sets in the

ECMWF and BMRC SCMs at Manus Island and Nauru. Not surprisingly, some of the initial condition and forcing data derived from the analyses contain large uncertainties. For example, the initial condition data derived from NCEP reanalyses have a significant dry bias. Furthermore, there is evidence that the NCEP and GASP vertical velocity forcing data contain large uncertainties.

[77] It was found that the SCM simulations are quite sensitive to the uncertainties in the initial condition and forcing data. For SCM predictions beyond about one day, the results are more affected by uncertainties in the forcing data than uncertainties in the initial conditions. In particular, the vertical velocity forcing term has a greater impact on the SCM results than the advective tendency terms in the tropics. Nevertheless, using an ensemble approach, it was shown that it is possible to distinguish between SCM model errors and errors arising from uncertainties in the initial conditions and forcing data. For example, the ESCM approach is able to identify a significant high level moist bias in the BMRC SCM, which is probably attributable to the model's treatment of upper tropospheric clouds. Importantly, the comparison of down-welling solar radiation predictions from the ERA-40 model and the ECMWF ESCM showed the ESCM approach was able to reproduce errors seen in the GCM.

[78] It is worth noting that this study mainly concentrated on simple measures of ensemble skill, such as the ensemble mean. There are a range of other measures and techniques, such as the Brier score, reliability diagrams and relative operating characteristics (ROC) which are commonly used to validate ensemble predictions, and potentially contain more information than is present in the ensemble mean alone [e.g., Atger, 1999]. These will be utilised in future work.

[79] In summary, this study has shown that an ESCM approach using initial condition and forcing data derived from NWP analyses is capable of identifying SCM model errors. This finding is significant, because it demonstrates the feasibility of running SCM simulations in areas such as the TWP where there are insufficient observations to derive the required forcing data. Furthermore, using the ESCM approach with NWP derived forcing data, it is also possible to make SCM runs over extended periods of time. This enables model parameterisations to be tested in the full range of environmental regimes that occur at a particular location, something which is difficult to achieve when studying SCM simulations for a limited number of short duration cases. An in depth ESCM study over the complete two year period at both Manus Island and Nauru is the subject of ongoing work, and the results will be reported later.

[80] **Acknowledgments.** We would like to thank Greg Roff for his efforts in porting and setting up the BMRC SCM on the workstation used in this project. Greg Roff and Peter May also provided helpful comments and suggestions on this paper. We are grateful for support from the U.S. Department of Energy under grant DE-FG02-03ER63533 as part of the Atmospheric Radiation Measurement Program.

References

Ackerman, T. P., and G. M. Stokes (2003), The Atmospheric Radiation Measurement Program, *Phys. Today*, Jan., 38–44.
 Atger, F. (1999), The skill of ensemble prediction systems, *Mon. Weather Rev.*, 127, 1941–1953.

Bechtold, P., et al. (2000), A GCM model intercomparison for a tropical squall line observed during TOGA-COARE IIL Intercomparison of single-column models and a cloud-resolving model, *Q. J. R. Meteorol. Soc.*, 126, 865–888.
 Bergman, J. W., and P. D. Sardeshmukh (2003), Usefulness of single column model diagnosis through short-term predictions, *J. Clim.*, 16, 3803–3819.
 Charlock, T. P., F. G. Rose, and K. M. Cattany-Carnes (1989), Cross correlations between the radiation and atmospheric variables in a general circulation model and in satellite data, *Mon. Weather Rev.*, 117, 212–220.
 Emanuel, K. A., and M. Živković-Rothman (1999), Development and evaluation of a convection scheme for use in climate models, *J. Atmos. Sci.*, 56, 1766–1782.
 Fu, R., A. D. D. Genio, and W. B. Rossow (1990), Behavior of deep convective clouds in the tropical Pacific deduced from ISCCP radiances, *J. Clim.*, 3, 1129–1152.
 Gregory, D., J.-J. Morcrette, C. Jakob, A. C. M. Beljaars, and T. Stockdale (2000), Revision of convection, radiation and cloud schemes in the ECMWF Integrated Forecasting System, *Q. J. R. Meteorol. Soc.*, 126, 1685–1710.
 Grell, G. A. (1993), Prognostic evaluation of assumptions used by cumulus parameterizations, *Mon. Weather Rev.*, 121, 764–787.
 Hack, J. J., and J. A. Pedretti (2000), Assessment of solution uncertainties in single-column modeling frameworks, *J. Clim.*, 13, 352–365.
 Jakob, C. (2004), Cloud regimes, Model evaluation and model development, *GEWEX News*, Nov., 6–8.
 Jakob, C., G. Tselioudis, and T. Hume (2005), The radiative, cloud and thermodynamic properties of the major tropical western Pacific cloud regimes, *J. Clim.*, in press.
 Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter (2002), NCEP-DOE AMIP-II Reanalysis (R-2), *Bull. Am. Meteorol. Soc.*, 83, 1631–1643.
 Lesht, B. M. (2004), Balloon-Borne Sounding System (BBSS) Handbook, *Tech. Rep. TR-029*, Atmos. Radiat. Meas. Program, Washington, D. C., Nov. (Available at <http://www.arm.gov/instruments/instrument.php>)
 Mapes, B. E., P. E. Ciesielski, and R. H. Johnson (2003), Sampling errors in rawinsonde-array budgets, *J. Atmos. Sci.*, 60, 2697–2713.
 Mather, J. H., T. P. Ackerman, W. E. Clements, F. J. Barnes, M. D. Ivey, L. D. Hatfield, and R. M. Reynolds (1998), An atmospheric radiation and cloud station in the tropical western Pacific, *Bull. Am. Meteorol. Soc.*, 79, 627–642.
 Minnis, P., W. L. Smith Jr., D. P. Garber, J. K. Ayers, and D. R. Doelling (1995), Cloud properties derived from GOES-7 for the spring 1999 ARM intensive observing period using version 1.0.0 of the ARM Satellite Data Analysis Program, *NASA Tech. Rep.*, RP 1366.
 Nagarajan, B., and A. R. Aiyer (2004), Performance of the ECMWF operational analyses over the tropical Indian Ocean, *Mon. Weather Rev.*, 132, 2275–2282.
 Nordeen, M. L., D. R. Doelling, P. Minnis, M. M. Khaier, A. D. Rapp, and L. Nguyen (2001), GMS-5 satellite-derived cloud properties over the tropical western Pacific, in *Proceedings of 11th ARM Science Team Meeting*, Atmos. Radiat. Meas. Program, Washington, D. C. (Available at <http://www.arm.gov/publications/proceedings/conf11/title.stm>)
 Randall, D. A., and D. G. Cripe (1999), Alternative methods for specification of observed forcing in single-column models and cloud system models, *J. Geophys. Res.*, 104(D20), 24,527–24,545.
 Randall, D. A., K.-M. Xu, R. J. C. Somerville, and S. Iacobellis (1996), Single-column models and cloud ensemble models as links between observations and climate models, *J. Clim.*, 9, 1683–1697.
 Randel, W. J., F. Wu, and D. J. Gaffen (2000), Interannual variability of the tropical tropopause derived from radiosonde data and NCEP reanalyses, *J. Geophys. Res.*, 105(D12), 15,509–15,523.
 Revercomb, H. E., et al. (2003), The ARM Program's water vapor intensive observation periods: Overview, initial accomplishments, and future challenges, *Bull. Am. Meteorol. Soc.*, 217–236.
 Roff, G. (2004), The BAM single column model, technical report, Bur. of Meteorol. Res. Cent., Melbourne, Victoria, Australia.
 Seaman, R., W. Bourke, P. Steinle, T. Hart, G. Embury, M. Naughton, and L. Rikus (1995), Evolution of the Bureau of Meteorology's Global Assimilation and Prediction System. Part I: Analysis and initialisation, *Aust. Meteorol. Mag.*, 44, 1–18.
 Stephens, G. L., C. Jakob, and M. Miller (1998), Atmospheric ice - A major gap in understanding the effects of clouds on climate, *GEWEX News*, 8(1), 1–8.
 Stokes, G. M., and S. E. Schwartz (1994), The Atmospheric Radiation Measurement (ARM) Program: Programmatic background and design of the cloud and radiation test bed, *Bull. Am. Meteorol. Soc.*, 75, 1201–1221.
 Thiébaux, H. J. (1994), *Statistical Data Analysis for Ocean and Atmospheric Sciences*, 247 pp., Elsevier, New York.

- Tiedtke, M. (1993), Representation of clouds in large-scale models, *Mon. Weather Rev.*, *121*, 3040–3061.
- Trenberth, K. E. (1985), Persistence of daily geopotential heights over the Southern Hemisphere, *Mon. Weather Rev.*, *113*, 38–53.
- Trenberth, K. E., and C. J. Guillemot (1998), Evaluation of the atmospheric moisture and hydrological cycle in the NCEP/NCAR reanalyses, *Clim. Dyn.*, *14*, 213–231.
- Turner, D. D., B. M. Lesht, S. A. Clough, J. C. Liljegren, H. E. Revercomb, and D. C. Tobin (2003), Dry bias and variability in Vaisala RS80-H radiosondes: The ARM experience, *J. Atmos. Oceanic Technol.*, *20*, 117–132.
- Uppala, S. (2002), ECMWF Reanalysis 1957–2001, ERA-40, in *Proceedings of Workshop on Re-analysis, 5–9 November 2001*, pp. 1–10, Eur. Cent. for Medium-Range Weather Forecasts, Reading, UK.
- Wang, J., H. L. Cole, D. J. Carlson, E. R. Miller, K. Beierle, A. Paukkunen, and T. K. Laine (2002), Corrections of humidity measurement errors from the Vaisala RS80 radiosonde—Application to TOGA COARE data, *J. Atmos. Oceanic Technol.*, *19*, 981–1002.
- Xie, S., R. T. Cederwall, M. Zhang, and J. J. Yio (2003), Comparison of SCM and CSRM forcing data derived from the ECMWF model and from objective analysis at the ARM SGP site, *J. Geophys. Res.*, *108*(D16), 4499, doi:10.1029/2003JD003541.
- 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.
- Zhang, M. H., and J. J. 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.
- Zhou, X., M. A. Geller, and M. Zhang (2001), Tropical cold point tropopause characteristics derived from ECMWF reanalyses and soundings, *J. Clim.*, *14*, 1823–1838.
- Ziehmann, C. (2000), Comparison of a single-model EPS with a multi-model ensemble consisting of a few operational models, *Tellus, Ser. A*, *52*, 280–299.

T. Hume and C. Jakob, Bureau of Meteorology Research Centre, GPO Box 1289K, Melbourne, Vic 3001, Australia. (t.hume@bom.gov.au)

Ensemble single column model validation in the tropical western Pacific

Timothy Hume¹ and Christian Jakob¹

Received 11 September 2006; revised 9 January 2007; accepted 14 January 2007; published 19 May 2007.

[1] Single column models (SCMs) are useful tools for the evaluation of parameterizations of radiative and moist processes used in general circulation models (GCMs). SCM applications have usually been limited to regions where high-quality observations are available to derive the necessary boundary condition or forcing data. Recently, researchers have developed techniques for deriving SCM forcing data from other data sets, such as NWP (numerical weather prediction) analyses. The uncertainties inherent in these forcing data products have an unknown and possibly significant effect on SCM runs. This paper shows how an ensemble SCM (ESCM) approach can be used to minimize the uncertainty in SCM simulations resulting from uncertainties in the forcing data. Some innovative evaluation techniques have been applied to ESCM runs at the tropical western Pacific Atmospheric Radiation Measurement (ARM) program sites at Manus Island and Nauru. These techniques, making use of traditional ensemble verification methods and objectively determined cloud regimes, are shown to be able to highlight parameterization deficiencies and provide a useful tool for testing new or improved model parameterizations.

Citation: Hume, T., and C. Jakob (2007), Ensemble single column model validation in the tropical western Pacific, *J. Geophys. Res.*, 112, D10206, doi:10.1029/2006JD008018.

1. Introduction

[2] Radiative and moist processes play a critical role in the Earth's climate system. In climate models the parameterizations of these processes remain a major source of uncertainty; consequently, much effort has been put into improving them. For example, one of the primary objectives of the U.S. Department of Energy's Atmospheric Radiation Measurement (ARM) program is to develop and test parameterizations describing moist and radiative processes, with the intent of incorporating them into general circulation models (GCMs) [Stokes and Schwartz, 1994; Ackerman and Stokes, 2003].

[3] Modern GCMs are very complicated, containing parameterizations of many physical processes. The interactions between the various components of a GCM are often nonlinear. Furthermore, running a GCM requires significant computational resources. Consequently, GCM evaluation can be an onerous task. Several approaches have been developed to simplify the testing of GCM parameterizations. One popular technique is single column modelling. A single column model (SCM) is comprised of a single vertical column of grid points. Interactions with neighboring columns are prescribed by lateral boundary conditions, also referred to as forcing data [Randall *et al.*, 1996].

[4] The simple design of an SCM makes identification of errors in the model's parameterizations easier than in a full

GCM. If the initial and boundary conditions are perfect, then SCM errors can be attributed solely to the physical parameterizations operating in the vertical column. For example, Jakob [2003] describes a strategy for evaluating GCM cloud parameterizations, which utilizes single column modelling and objective weather or climate regimes. A number of researchers have tested, and developed or modified, parameterizations based on SCM evaluations [e.g., McFarquhar *et al.*, 2003; Iacobellis *et al.*, 2003; Zhang and Lohmann, 2005].

[5] One difficulty with single column modelling is the specification of the forcing data, such as the vertical velocity and advective tendencies of temperature and moisture. SCM errors caused by uncertain boundary conditions may mask errors caused by parameterization deficiencies. Therefore most SCM studies have been restricted to locations and times when high-quality observations are available to derive the forcing data [e.g., Randall and Cripe, 1999]. More recently, researchers have derived forcing data from NWP profiles constrained by surface and top of the atmosphere observations [e.g., Xie *et al.*, 2003, 2004]. With all these approaches, the issue of the forcing data quality remains. Even when observations are used, instrument and sampling errors, such as unresolved spatial variability within a radiosonde array, can introduce significant uncertainties into the derived forcing data [Mapes *et al.*, 2003].

[6] To address the issue of uncertainty in the forcing data, Hume and Jakob [2005, hereinafter referred to as HJ05] used an ensemble SCM (ESCM) approach. They first created four forcing data sets, derived from four different NWP analyses, for the tropical western Pacific (TWP) ARM

¹Bureau of Meteorology Research Centre, Melbourne, Victoria, Australia.

sites at Manus Island and Nauru. These forcing data sets were then used to run ensembles of the ECMWF and Bureau of Meteorology Research Centre (BMRC) SCMs for Manus Island and Nauru. HJ05 showed that using this ensemble approach, uncertainties in the SCM results which were caused by errors in the model physics could be distinguished from errors caused by uncertainties in the forcing data sets.

[7] While HJ05 have hypothesized that the ESCM approach can find model problems, that claim is based on very few results. The purpose of this paper is twofold: (1) to extend the HJ05 results and (2) to introduce some innovative evaluation techniques that, when used in conjunction with the ESCM approach, can highlight parameterization deficiencies. First, in section 3, some simple validations of the ESCM method are studied. Important questions which are addressed in this section include: (1) how well does the ESCM method compare with single SCM runs, and (2) how sensitive is the method to outliers in the forcing data? Then, in section 3.2, a number of validation techniques, which can only be applied to ensembles, are investigated. As discussed earlier, *Jakob* [2003] proposed a regime-based evaluation methodology. In section 4 we investigate if this methodology can be applied to the ESCM approach. Finally, in section 5, a simple example shows how these validation techniques can be used to test modified model parameterizations.

2. ESCM Runs

[8] The validations presented in this paper are for two separate SCM ensembles. The first ensemble consists of sixteen ECMWF SCM runs at Manus Island and Nauru, initialized every 6 hours (at 0000, 0600, 1200, and 1800 UTC) during 1999 and 2000. The second ensemble is similar to the first but uses the BMRC SCM. The ECMWF SCM used in this work is the single column version of the model used for the ERA-40 reanalysis [Uppala *et al.*, 2005]. The BMRC SCM is the single column version of the Bureau of Meteorology's GCM and is described by *Roff* [2005]. While the SCMs used in this study are perhaps more frequently used by the climate community, there are no reasons preventing the ESCM technique being used to test single column versions of operational forecasting models.

[9] Each member of the BMRC and ECMWF ensembles was initialized with one of four initial condition data sets and forced with one of four forcing data sets (making a total of 16 ensemble members). The initial condition and forcing data sets were derived from ERA-40 [Uppala *et al.*, 2005], operational ECMWF analyses [Gregory *et al.*, 2000], operational analyses from the Australian Bureau of Meteorology's Global Assimilation and Prediction System (GASP) model [Seaman *et al.*, 1995], and National Centers for Environmental Prediction (NCEP) reanalyses [Kanamitsu *et al.*, 2002], respectively. The horizontal resolution of forcing data were 2.5 degrees. The NWP analyses did not contain sufficient information to initialize clouds in the SCMs. Therefore the SCM runs initially start with clear skies and are allowed to "spin-up" for 12 hours. HJ05 showed this was sufficient time to develop realistic clouds in the model. Further details describing the derivation and validation of the initial and forcing data sets are found in HJ05.

[10] As shown in HJ05, results from individual SCM ensemble members vary significantly. For example, the ECMWF SCM runs initialized with ERA-40 data and forced with GASP data underestimate the downward surface solar radiation, while the runs initialized with NCEP data and forced with ERA-40 significantly overestimate the same quantity. Nevertheless, the ESCM mean is quite close to the observed value. Furthermore, it was shown that errors in the ECMWF ESCM mean are similar to the errors in the full ECMWF model (which does not require the forcing data used for the SCM runs). These results suggest the ESCM approach is capable of identifying model errors, distinct from errors resulting from uncertainties in the forcing data.

[11] HJ05 highlighted that the ESCM technique is potentially capable of identifying model errors. However, only a limited number of results were presented; it is not clear how much additional information about model problems can be obtained using this method compared to traditional single model run validations. Therefore the following sections apply some well-known NWP ensemble validation techniques to an SCM problem for the first time, in an attempt to identify specific model problems.

3. ESCM Validations

3.1. Use of the Ensemble Mean

[12] A common ensemble validation technique is to compare the ensemble mean to observations. This was the approach taken in HJ05. One advantage of using the ensemble mean instead of a single model run selected from the ensemble is that the influence of outliers is lessened. Moreover, it is well known that the mean of several model predictions is often more skillful than any of the individual model predictions [e.g., *Ziehmann*, 2000]. On the other hand, outliers in an ensemble may contain useful information about the inherent uncertainties in the predictions, which is lost when only the mean is studied.

[13] Figure 1 shows a short time series for the ensemble mean of 12 hour SCM predictions of total cloud cover (TCC), valid at 0000 UTC, at Manus Island. To derive TCC from cloud cover at each model level, the ECMWF SCM uses a maximum-random cloud overlap assumption, while the BMRC SCM uses a random overlap assumption. The observed TCC is also plotted in Figure 1. The TCC observations are derived from GMS-5 (Geostationary Meteorological Satellite) satellite data using the method of *Minnis et al.* [1995], as described by *Nordeen et al.* [2001], and represent the average TCC in a 3 degree square box covering Manus Island. It is not obvious if the SCM predictions exhibit any skill. Sometimes (e.g., 8 January, 0000 UTC), the observations correspond quite well with the predictions; however, there are many times when high TCC values are observed and quite low values are predicted by both SCM ensembles, suggesting the SCMs may be negatively biased.

[14] To help understand the model results, Figure 2 shows the PDFs for the 12 hour ESCM mean predictions of TCC and the observed TCC values. The PDFs cover the 1999–2000 period. Both SCMs underestimate the frequency of TCC greater than about 0.9 and overestimate the frequency of lower cloud cover amounts. It should be noted that the

Figure 1. Time series of total cloud cover observations and predictions at Manus Island. The ECMWF and BMRC curves show the ensemble mean 12 hour predictions, valid at 0000 UTC. The observations are derived from GMS-5 data valid at 0000 UTC.

spatial resolutions of the models and observations are similar; therefore the differences in the TCC PDFs are probably attributable to model deficiencies.

[15] The preceding PDFs highlight some deficiencies in the SCM results. However, even if the modelled and observed PDFs were the same, this would not necessarily indicate a great level of skill in the predictions. It is possible for a model to produce the correct PDF for TCC but have no skill at predicting when the cloud cover occurred. Therefore it is often desirable to measure the performance of a forecasting system against a known “no-skill” forecasting system, such as climatology. In this paper, a modified version of the cloud verification method presented by Jakob *et al.* [2004] is used. Whereas Jakob *et al.* [2004] used point observations of cloud cover, which were either 0 or 1, the cloud cover observations used in this study are for a three degree square box covering Manus Island and can assume any value between 0 and 1.

[16] The mean square error (MSE) is a convenient score for verifying cloud forecasts. It is defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2 \quad (1)$$

where f_i is the forecast cloud cover, o_i is the observed cloud cover, and n is the number of forecasts being validated. The MSE for a perfect forecasting system is 0. In the case of “no-skill” climatological forecasts, f_i is always the same and is equal to the long term average cloud cover.

[17] The reduction of variance (RV) is a commonly used score for comparing a forecasting system against no-skill climatological forecasts [Stanski *et al.*, 1989]:

$$RV = \frac{MSE_{climate} - MSE}{MSE_{climate}} \quad (2)$$

where MSE is the mean square error of the SCM forecasts being validated and $MSE_{climate}$ is the mean square error for the climatological forecasts. A perfect forecast system will have a RV of 1; however, this is unlikely to be achievable in practice. If the forecasts are worse than climatology, the RV will be negative.

[18] Figure 3 shows the RV for 12 hour SCM forecasts of TCC valid at 0000 UTC for Manus Island and Nauru. A number of interesting points are evident in the figure. First,

Figure 2. PDFs for the predicted and observed 0000 UTC TCC at Manus Island. The model results show the ensemble mean of the 12 hour forecasts. The results are displayed in 0.1 wide bins.

Figure 3. Reduction of variance (RV) for the ECMWF and BMRC SCM twelve hour forecasts of TCC, valid at 0000 UTC. The crosses show the RV for the individual ensemble members, and the circles show the RV for the ensemble mean.

at both Manus Island and Nauru, and for both SCMs, the ensemble mean scores better than any individual ensemble member. As mentioned earlier, this is a well-known advantage of ensemble prediction systems (EPS). Most importantly, it is also apparent from the figure that some of the ensemble members are habitual outliers and score poorly in the validation. Despite this, the ensemble mean and median are rather insensitive to these outliers. This highlights the utility of the ESCM approach in the absence of perfect forcing data. If only one of the forcing data sets was used, there would be no way to tell if it belonged to the outlier runs evident in Figure 3. Finally, the SCMs appear to be performing worse than climatology because the RV is negative in all the cases. At first sight, this is a rather disappointing result. However, it is well known that the skill of a forecast system cannot usually be judged by a single scalar score [e.g., Buizza, 2000]. The next section will show the SCM ensembles exhibit some skill compared to climatology, when validated using other methods. It is worth noting that Jakob *et al.* [2004] also reported negative skill scores when validating a cloud system model (CSM) and the ECMWF operational model against radar observations at the ARM Southern Great Plains site in Oklahoma. Finally, it is clear that the ECMWF SCM is performing better than the BMRC SCM. This result supports HJ05's finding that the use of the ESCM technique does provide information about the model performance. If the results were entirely dominated by features in the forcing data sets, the model results would be expected to lie much closer to each other.

3.2. Probabilistic Forecast Validation Techniques

[19] The preceding section presented some validations of the ESCM mean and showed the ensemble mean was more skillful than any of the individual ensemble members. However, validations of the ensemble mean potentially ignore useful information contained in the spread of the ensemble. In this section, some probabilistic forecast validation techniques are applied to the ESCM predictions. It is worth noting that this is the first application of such techniques in studies using SCMs.

[20] It is straightforward to generate probabilistic forecasts from an ensemble. All that is necessary is to bin the model predictions and count the proportion of ensemble members which produced predictions in each bin. Ideally, the probabilistic forecasts will not have any systematic errors. For example, if a large number of forecasts of 50% probability of TCC between 0.8 and 0.9 are collected, it would be hoped that the observed TCC was between 0.8 and 0.9 in 50% of the cases (for clarity, probabilities are reported as percentages, and cloud amounts as a decimal fraction between 0 and 1). A forecasting system which meets this criteria is said to be reliable.

[21] Reliability diagrams are useful tools for studying ensemble forecasting systems [e.g., Wilks, 2006]. The horizontal axis of a reliability diagram shows the forecast probability of an event occurring, while the vertical axis shows the observed frequency of the event, given the forecast probability. A perfectly reliable forecasting system has all the points lying on a diagonal line extending from the bottom left to top right corners of the diagram. By definition, a climatological forecast is perfectly reliable. Figure 4 shows a reliability diagram for the ECMWF ESCM predictions of TCC greater than or equal to 0.9 at Manus Island. The 0.9 threshold is somewhat arbitrary, and was chosen because approximately half the observations exceed 0.9, ensuring numerical stability when various skill scores are calculated. Furthermore, it is clear from Figure 2 that the model seriously underpredicts cloud cover exceeding about 0.9. Similar results to those described in this section are obtained with different thresholds. It is clear that the ESCM forecasts are not perfectly reliable. For example, when the ESCM predicts a 6.25% probability of TCC exceeding 0.9, approximately 35% of the observations exceed 0.9. In a later section, it will be shown that this underprediction of TCC > 0.9 events is caused by the SCMs underestimating the occurrence of deep convection and overestimating the occurrence of relatively less cloudy, suppressed regimes.

[22] Another important characteristic of an ensemble prediction system (EPS) is resolution. Resolution represents the extent to which observed outcomes differ, given differing forecasts. Climatological forecasts are a common example of a reliable forecasting system with no resolution. It is clear from Figure 4 that the ESCM does in fact show a moderate level of resolution because increasing forecast probabilities are generally associated with an increase in the observed frequency of TCC exceeding 0.9. However, it should be noted that a reasonably high proportion of the probabilistic forecasts lie near the climatological forecast on the reliability diagram.

[23] Often it is useful to have a single scalar measure comparing the EPS against a reference forecasting system,

Figure 4. Reliability diagram for 12 hour forecasts of TCC greater than or equal to 0.9 from the ECMWF ESCM at Manus Island. The forecasts are valid at 0000 UTC. The size of the circles show the relative frequencies of forecasts with probabilities of $\frac{0}{16}, \frac{1}{16}, \dots, \frac{16}{16}$. The cross represents climatological forecasts. The shaded region indicates points which make a positive contribution to the Brier score.

such as climatology. The Brier score (BS) [Brier, 1950] and Brier skill score (BSS) are frequently used for this purpose:

$$BS = \frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2 \quad (3)$$

$$BSS = \frac{BS_{\text{climate}} - BS}{BS_{\text{climate}}} \quad (4)$$

where n is the number of observations, f_i is the forecast probability of TCC exceeding 0.9, and o_i is the observed occurrence of TCC exceeding 0.9. o_i is either 0 or 1, for observed TCC less than 0.9 and observed TCC exceeding 0.9, respectively. The reference climatological forecast used for the Brier skill score is simply the relative frequency of occurrence of observed TCC greater than 0.9 during the 1999–2000 period. The Brier score and Brier skill score are the probabilistic forecast equivalent to the mean square error and reduction of variance scores defined in the previous section.

[24] The Brier skill scores for the ECMWF and BMRC ESCM TCC > 0.9 forecasts at Manus Island are -0.5 and -1.2 , respectively. For both models, the skill scores are negative, suggesting the SCM forecasts are less skillful than climatology. This is consistent with the results presented in the previous section, which showed the ESCM mean TCC forecasts were less skillful than climatology. Finally, as also shown in the previous section, the BMRC ESCM is less skillful than the ECMWF ESCM at predicting TCC > 0.9.

[25] The Brier score is sensitive to both reliability and resolution of the forecasts [e.g., Wilks, 2006; Jakob *et al.*, 2004]. Sometimes it is informative to focus on the resolution of a forecasting system because this reflects the potential skill of a perfectly calibrated forecasting system. A useful tool for highlighting EPS resolution is the relative operating characteristics (ROC) diagram [e.g., Mason, 1982; Harvey *et al.*, 1992]. Forecasts with poor reliability, but good resolution may still show good validation results when a ROC diagram is used. On a ROC diagram, climatological forecasts lie on the diagonal line from the bottom left to top right corner. Forecasts which lie above this line are considered to be more useful than climatology.

[26] Figure 5 shows the ROC curves for predictions of TCC exceeding 0.9 at Manus Island, from both the ECMWF and BMRC ensembles. Using this validation technique, it is clear that both the BMRC and ECMWF ESCM forecasts of TCC are more skillful than climatology. A frequently used scalar score, based on the ROC diagram, is the area beneath the ROC curve (ROCA). If ROCA exceeds 0.5 (the area beneath the diagonal climatology line), the forecasts are considered skillful. The ROCA scores for both ensembles at Manus Island are 0.82; this clearly indicates the skill of the ensembles compared to climatology. It is interesting to note the ROCA for the ECMWF and BMRC ESCMs is the same, yet the BMRC model has a lower BSS. This suggests the resolution of both models is similar, but the BMRC model is less reliable.

[27] While ROC diagrams are a useful tool for highlighting the resolution of a forecasting system, they are insensitive to reliability. For example, an error in a physical parameterization may introduce a systematic bias which affects the reliability but does not greatly affect the resolu-

Figure 5. ROC diagram for probabilistic forecasts of TCC exceeding 0.9 for each of the SCM ensembles at Manus Island.

Table 1. Relative Frequency of Occurrence, %, of the 0000 UTC Observed and Model-Derived Cloud Regimes at Manus Island and Nauru

Regime	Manus Island			Nauru		
	Observed	ECMWF	BMRC	Observed	ECMWF	BMRC
SSC	28.9	72.2	60.8	57.7	84.9	80.0
STC	46.1	15.6	32.1	33.4	13.3	19.2
CC	15.6	7.0	3.2	8.2	1.1	0.3
CD	9.4	5.2	4.0	0.7	0.7	0.5

tion of the forecast. A ROC diagram may not be able to identify such errors. Therefore it is important to use ROC diagrams in conjunction with reliability diagrams or other scores which are sensitive to forecast reliability.

4. Cloud Regime Validations

[28] One weakness of the validations presented in the preceding sections is that even though it is possible to identify model errors (as distinct from errors caused by uncertainty in the forcing data), it is difficult to ascertain the physical causes of these errors. This is perhaps inevitable because the validations cover a 2 year period, during which, several different weather or climate regimes might be expected to occur. Errors which may be present in one particular weather or climate regime may be compensated, or compounded, by errors at other times. A traditional approach to overcome this problem is to perform case studies and identify the situations when particular errors occur. One difficulty with case studies is that there is a degree of subjectivity in selecting the cases to validate. An alternative approach is to objectively split the period being studied into distinct weather or climate regimes, and perform validations for each regime [e.g., *Tselioudis et al.*, 2000; *Norris and Weaver*, 2001; *Tselioudis and Jakob*, 2002; *Jakob et al.*, 2005].

[29] As discussed earlier, we are interested in studying errors in model representations of moist and radiative processes, particularly clouds. Therefore it is useful to study the model behavior during periods when cloud regimes with distinct characteristics are observed. For example, in the tropics there is the possibility that different model errors occur during convective regimes associated with large amounts of deep cloud than in suppressed regimes with relatively little cloud, or only thin, high cirrus. This study makes use of the objective cloud regimes identified by *Jakob and Tselioudis* [2003] for the tropical western Pacific (TWP). Jakob and Tselioudis applied a clustering algorithm to International Satellite Cloud Climatology Project (ISCCP) cloud top pressure and optical thickness histograms, derived from satellite data, to identify four distinct cloud regimes which affect the TWP. The regimes included a suppressed regime with shallow clouds (SSC), a suppressed regime dominated by thin cirrus (STC), a convective regime with a small coverage of deep convective clouds and high cirrus clouds (CC), and a regime dominated by deep convective clouds and thick anvils (CD). Each of the four cloud regimes were shown to have different thermodynamic and radiative characteristics, suggesting that they represent physically distinct weather or climate regimes [*Jakob et al.*, 2005].

4.1. Model Cloud Regimes

[30] An obvious first test is to see how frequently the SCMs predict each of the four cloud regimes identified by *Jakob and Tselioudis* [2003]. As mentioned above, the observed regimes were derived from ISCCP histograms which show the frequency of clouds of given combinations of optical thicknesses and cloud top pressure [*Rossow and Schiffer*, 1991]. To produce model ISCCP histograms, cloud top pressure and optical thickness were first derived from the 12 hour ESCM predictions valid at 0000 UTC, using the ISCCP simulator code [*Klein and Jakob*, 1999; *Webb et al.*, 2001]. The ISCCP simulator used a maximum-random cloud overlap assumption for the ECMWF predictions and a random overlap assumption for the BMRC forecasts. The resulting model ISCCP histograms were then classified according to which of the four observed ISCCP regimes they were closest to (using a Euclidean distance metric). This procedure, which classifies the model ISCCP histograms according to which observed ISCCP regime they are closest to, has several advantages. First, it requires fewer computational resources than the clustering algorithm which *Jakob and Tselioudis* [2003] used to generate the observed regimes. It also ensures that there are the same number of model regimes as observed regimes, making comparison between the model and observations easier. A possible disadvantage of this procedure is that it forces the model histograms into one of the four observed regimes, when in fact the model predictions may be so different from observations that a different set of model regimes would be more appropriate. For example, *Williams et al.* [2005] used the KMEANS algorithm to cluster the simulated histograms in the same manner as the observed histograms were clustered. For some models, this resulted in a different number of model-derived regimes than observed regimes. Nevertheless, the model regimes derived using the clustering method were usually similar to the observed regimes. Therefore it is anticipated that the simpler method adopted in this study for classifying the simulated ISCCP histograms will not adversely affect the results.

[31] Table 1 shows the relative frequency of occurrence for the 0000 UTC observed and model-derived cloud regimes at Manus Island and Nauru. It is notable that both SCMs overestimate the frequency of the SSC regime and underestimate the frequency of the convective regimes (CC and CD). It is not clear why the model-derived cloud regime frequencies differ from the observed frequencies. It is very likely that errors in the model physics are partially responsible for differing frequencies; this will be examined in more detail in the next section. Another possibility is that some of the differences are caused by deficiencies in the forcing data. For example, the frequency of the STC regime is underestimated at both Manus Island and Nauru by the SCMs. It is possible that some of the high cirrus observed in the STC regime originates from outside the model domain and is advected over the site. The SCMs are unable to model advected cloud because advective terms for cloud condensate are not included in the forcing data sets.

4.1.1. Total Cloud Cover Validations

[32] The preceding section showed that both the ECMWF and BMRC SCMs underestimated the frequency of occurrence of the STC, CC, and CD cloud regimes and overestimated the frequency of the SSC regime. Predicting the

Figure 6. Box and whisker diagrams showing the TCC PDFs, for each of the observed and modelled regimes at Manus Island (left) and Nauru (right). The dark grey bars show the observed 0000 UTC TCC PDFs, while the light grey and white bars show the PDFs for the 12 hour ensemble mean forecasts of TCC valid at 0000 UTC from the BMRC and ECMWF SCMs, respectively. The boxes show the median, upper and lower quartiles. The whiskers show the 5th and 95th percentiles; the bold crosses show the mean.

correct frequency of cloud regimes is a particularly stringent test; the model has to correctly predict the cloud tops and optical thicknesses, and their distribution in the grid columns. Notwithstanding the incorrect cloud regime frequencies, it is interesting to investigate some other properties of the model-derived cloud regimes. For example, when the model “thinks” it is in a particular cloud regime, are its predictions of other variables consistent with what is expected from observations? Tests such as these may be helpful in identifying regime-dependent model errors.

[33] Figure 6 shows the observed and modelled TCC PDFs at Manus Island and Nauru for each of the four cloud regimes. To test if there is a significant difference between the ECMWF and BMRC PDFs, the ECMWF and observed PDFs, and the BMRC and observed PDFs for each cloud regime, Kolmogorov-Smirnov (K-S) tests were applied. The K-S test is a nonparametric test which can be used to test whether two samples are drawn from different distributions [Wilks, 2006]. The differences between the PDF means, and whether the PDFs are drawn from different distributions, are summarized in Table 2.

[34] Several points can be noted from Figure 6 and Table 2. At Manus Island, for the complete period when the validations were performed, the median ECMWF TCC is slightly higher than the median observed TCC, yet the mean ECMWF cloud cover is considerably less than observed. This is because the modelled total cloud cover distribution is not symmetric, with large numbers of cloud cover values near the extreme of 1. Importantly, the K-S tests show the ECMWF and BMRC TCC distributions are different for almost all the cloud regimes. The different ECMWF and BMRC TCC distributions must be caused by differences in the model physics because the forcing data

sets applied were the same. This highlights that the ESCM technique is capable of identifying the influence of the different model physics on the SCM predictions, a feature that will be investigated further below.

[35] At Manus Island, most of the model error appears to be associated with the SSC regime. As shown in Table 1, both models predict this regime more than twice as frequently as it is observed. For the other regimes, the mean cloud cover is close to 1 for both the model predictions and observations. For the non-SSC regimes, most of the difference between observations and predictions is in the spread of the PDFs. The models almost always predict a TCC of 1, whilst observations as low as 0.9 are reasonably common for the CC and STC regimes.

[36] At Nauru, the SCMs underestimate cloud cover for all the regimes. Furthermore, the model PDFs do not vary much between the cloud regimes, though there is a clear

Table 2. Difference Between the TCC PDF Means for Each Cloud Regime^a

Regime	Manus Island			Nauru		
	E-B	E-O	B-O	E-B	E-O	B-O
CC	−0.02	0.03	0.05	0.14	−0.38	−0.51
STC	0.01	0.07	0.06	0.19	−0.30	−0.49
CD	0.01	0.00	0.00	0.21	−0.40	−0.61
SSC	0.30	−0.08	−0.38	0.22	0.01	−0.21
All	0.15	−0.14	−0.29	0.20	−0.15	−0.35

^aTCC is reported as a fraction between 0 and 1. Abbreviations are E (ECMWF 12 hour predictions of TCC valid at 0000 UTC), B (BMRC 12 hour predictions of TCC valid at 0000 UTC), and O (TCC observed at 0000 UTC). The nonitalic table entries indicate the K-S test showed the PDFs were drawn from different distributions at the 95% significance level.

Figure 7. Radiative properties of the observed (dark grey bars) and modelled (light grey bars: BMRC SCM 12 hour ensemble mean predictions, white bars: ECMWF SCM 12 hour ensemble mean predictions) cloud regimes. The top left panel shows the OLR at the TOA, the top right panel shows the downward long-wave radiation at the surface, and the bottom panel shows the downwelling solar radiation at the surface, normalized by the clear sky radiation. The normalized solar radiation shown in the bottom panel have been restricted to lie within the 0–1 range; occasionally values slightly larger than one occur due to differences between the modelled radiation data and the clear sky radiation data used for normalization. Boxes show the median, lower and upper quartiles. The whiskers show the 5th and 95th percentiles.

difference between the ECMWF and BMRC models. Because the SCMs very rarely predict anything other than suppressed conditions at Nauru (refer to Table 1), the sample sizes for the active regimes are too small to draw any conclusions about model performance in those regimes from that site.

4.1.2. Radiative Properties of the Modelled Cloud Regimes

[37] Validations of TCC have some limitations. For example, Figure 6 showed that the models and observations have a cloud cover of almost 1 for the CC, STC, and CD regimes at Manus Island. While the agreement between the model's TCC predictions and observations is encouraging, the TCC does not provide any information on the vertical distribution of clouds or cloud thicknesses. Validations of some of the SCM's radiation predictions with observations

from the Manus Island ARM site can provide useful additional information on how well the clouds are modelled.

[38] Figure 7 shows validations of outgoing long wave radiation at the top of atmosphere (OLR), downwelling long wave radiation at the surface (DLR), and downwelling shortwave radiation at the surface (normalized by the clear-sky shortwave radiation, SW_d/SW_c). The OLR observations were derived from the same satellite data set used for the TCC observations; surface radiation observations were made by radiometers at the ARM site. Estimates of the clear-sky shortwave radiation were acquired using the method of *Long and Ackerman* [2000] (the specific algorithm is described by *Long and Gaustad* [2004]). It should be noted that the SCM horizontal domain is predominantly water, while the surface radiation measurements are made from a land site, and may not be representative of the

Figure 8. As for Figure 7, but for total column water vapor (TCWV).

domain. The magnitude of any land or topography effects on the radiation measurements made at Manus Island is not clear. *McFarlane et al.* [2005] found that island effects at Nauru could affect downwelling solar radiation by as much as 8–10% and downwelling long-wave radiation by 1–2%. However, the ARM site at Manus Island is located on the downwind side of the island, so it is hoped island effects will be minimized.

[39] The ECMWF SCM appears to slightly underestimate the OLR for the SSC regime, while the BMRC SCM slightly overestimates it. In the case of the BMRC SCM, the small overestimation of OLR is probably because the model underestimates the TCC for the SSC regime. On the other hand, the small underestimation of OLR by the ECMWF model is possibly caused by the SCM producing too much high cirrus, which compensates for the increased OLR associated with the smaller cloud cover. This will be examined in more detail later. In the case of the two convective regimes (CC and CD), both SCMs predict notably smaller OLR values than the observations. Since both models and observations show a cloud cover of almost 1 for the convective regimes, the smaller predicted OLR values are consistent with the models producing higher or thicker cirrus than is observed.

[40] The ECMWF SCM overpredicts the DLR in all regimes. Two possible explanations for overpredicting the DLR are that the model predicts too much low cloud or too much low level water vapor. Figure 8 shows the total column water vapor predicted by the SCMs. For all cloud regimes, both models either predict TCWV close to the observed value or underpredict it. Therefore the likely explanation for the ECMWF model overpredicting DLR is that it predicts too much low cloud. This is borne out by the results showing that the overprediction is greater in the regimes where a lot of low clouds are present (CD and SSC) and not as severe in the regimes where cirrus predominates (STC and CC). In contrast, the BMRC SCM consistently underpredicts the DLR, which is likely due to underpredicting the low level cloud cover and water vapor.

[41] The final panel in Figure 7 shows the 0000 UTC downwelling solar radiation at the surface, normalized by the 0000 UTC clear sky downwelling solar radiation. The ECMWF SCM underestimates the solar radiation during the SSC regime. This is consistent with the DLR results which suggested the ECMWF SCM produces too much low cloud. In contrast, the BMRC SCM overestimates the solar radiation during the CC and CD regimes, suggesting it is not producing enough cloud.

4.1.3. Vertical Cloud Distribution

[42] The preceding section showed that it is possible to infer some information about how well the SCMs model clouds from the radiative properties of the model-derived cloud regimes. It is also possible to directly validate the model clouds with observations from millimeter wavelength cloud radars, which are located at all the tropical ARM sites [e.g., *Mace et al.*, 1998; *Jakob et al.*, 2004]. Data from these radars are combined with other observations from the ARM sites to produce the Active Remotely-Sensed Clouds Locations (ARSCL, *Clothiaux et al.* [2000]) product, which includes, amongst other things, estimates of the vertical cloud distribution.

[43] Figure 9 shows the observed and modelled cloud profiles for 0000 UTC at Manus Island, both for the entire period and sorted by cloud regime. The ECMWF SCM overestimates the amount of high level cloud for all regimes, and the BMRC model overestimates it for all regimes except SSC. The overprediction of high cloud cover for the convective regimes (CC and CD) is consistent with the results showing both models underpredicted the OLR for these regimes. Similarly, the overprediction of cirrus by the ECMWF model for the SSC regime explains the slight underprediction of OLR by the model for this regime.

[44] It is interesting to note that only the ECMWF model simulates the observed peak in cloud cover at the top of the boundary layer (approximately 1000–2000 m). The vertical profiles in Figure 9 also suggest that the ECMWF model overestimates the thickness of these boundary layer clouds, especially for the SSC regime. This is also consistent with the earlier results, which showed a marked overestimation of the DLR, and underestimation of solar radiation, by the ECMWF model for the SSC regime.

5. Development of Modified SCM Parameterizations

[45] As mentioned in the introduction, SCMs are useful tools for efficiently testing GCM parameterizations. It is obviously important that SCM evaluation methods, such as the one presented here, are capable of detecting the effect of modifications to the model's physical parameterizations. In the previous sections, a variety of validation techniques were presented which could clearly identify differences between the BMRC and ECMWF SCMs. However, the influence of modifications to a single parameterization in an SCM is likely to be more subtle, and hence more difficult to detect, than the differences between two quite different models.

[46] To test if the techniques presented in this paper can indeed detect the effect of parameterization modifications, a modified version of the ECMWF SCM was created. The

Figure 9. Vertical cloud profiles at Manus Island. Observations are from 0000 UTC ARSCL (Active Remotely Sensed Clouds Locations) data. The SCM profiles are averages of the 12 hour ensemble mean forecasts valid at 0000 UTC during 1999 and 2000 for each regime.

ECMWF model uses the Tiedtke mass flux scheme for cumulus parameterization [Tiedtke, 1989]. In the modified SCM, the entrainment rate for penetrative convection was doubled from $1 \times 10^{-4} \text{ m}^{-1}$ to $2 \times 10^{-4} \text{ m}^{-1}$. Doubling the entrainment rate will have two effects: (1) it will increase the mass flux in convective plumes and (2) it will increase mixing between the plumes and the environment, thereby reducing their buoyancy. The former effect will generate more cloudy mass overall in the model. The latter effect is anticipated to reduce the penetration depth of convection and will probably result in clouds forming at lower levels than in the unmodified model. We do not expect this modification to represent an improvement to the cumulus parameterization scheme, rather it is being used to test the sensitivity of the ESCM validation techniques to modified model physics.

[47] Table 3 compares the relative frequency of occurrence for the four cloud regimes in the modified and unmodified SCMs. The observed frequencies are also included for comparison. It is notable that the frequency of deep convection (CD) is considerably less in the modified SCM, and there is a slight increase in the frequency of the SSC regime. These results suggest deep convection is weakened in the modified SCM, as expected.

[48] Figure 10 compares the vertical cloud profiles for the modified and unmodified ECMWF SCMs. As noted earlier, the average cloud profile during 1999 and 2000 is dominated by the SSC regime. In the modified SCM, increased midlevel cloud cover is predicted for most of the regimes, as

expected. Interestingly, when deep convection does occur (albeit rarely) in the modified SCM, there is considerably more cloud at all levels compared to the unmodified model. This means that the convection is strong enough to overcome the reduced buoyancy, and the effect of the extra cloudy mass takes over.

[49] While the modification to the cumulus parameterization described here is somewhat contrived, it is encouraging to note that its effects are easily detected by the ESCM validation techniques and were also expected from physical considerations. While it would be interesting to study the effect of parameterization changes being made to the current operational GCMs, this poses several difficulties. Importantly, the versions of the SCMs used in this study are older than the current GCMs. Therefore it is not clear if it is appropriate to test a parameterization being modified for the latest version of the GCM in an older version of the SCM. The use of the ESCM technique to test

Table 3. Relative Frequency of Occurrence (RFO), %, of the Four Cloud Regimes at Manus Island in the Original and Modified Versions of the ECMWF SCM

Cloud Regime	Original ECMWF SCM	Modified ECMWF SCM	Observed
SSC	72.2	75.3	28.9
STC	15.6	11.4	46.1
CC	7.0	9.7	15.6
CD	5.2	3.6	9.4

Figure 10. Comparison of vertical cloud profiles at Manus Island from the unmodified and modified ECMWF SCMs. Observed profiles are included for comparison. The SCM profiles are averages of the 12 hour ensemble mean forecasts valid at 0000 UTC during 1999 and 2000 for each cloud regime.

“real” parameterization modifications to a current operational model will be the subject of later research.

[50] It should also be stressed there is no guarantee that changes seen in the SCM as a result of parameterization modifications will necessarily be seen in the full GCM, where nonlinear interactions can occur with the surrounding grid point columns. It is therefore necessary to validate the results in the full GCM before a particular modification can be claimed to be beneficial. However, HJ05 showed that the ECMWF ESCM produced similar results to the full ECMWF model, at least for some solar radiation validations.

6. Conclusions

[51] This study has described the use of the ESCM technique of HJ05 for model evaluation. For this purpose, a number of validation techniques which can be applied to the ESCM method have been investigated. First, in section 3, some simple validation techniques were applied to ESCM runs at the ARM Manus Island and Nauru sites for the 1999–2000 period. They showed that the ESCM method is consistently more skillful than single SCM runs. In particular, the ESCM technique is insensitive to outliers in the forcing data. This is important because when a single SCM run is being evaluated, it is often difficult to know if the forcing data being used are outliers or not. Indeed, even when forcing data derived from observations are used, the ESCM technique is probably beneficial. This is because

observed forcing data can contain significant uncertainties resulting from instrument and sampling errors, such as unresolved spatial variability within a radiosonde array [Mapes *et al.*, 2003]. The derivation and use of ensembles of observed forcing data for the Darwin region (in Northern Australia) during the Australian summer monsoon of 2006 is the subject of current research.

[52] A second advantage of the ESCM technique is that it allows the use of a range of NWP ensemble validation techniques, which cannot be applied to single model runs. Section 3.2 investigated some of these ensemble validation techniques, including the Brier score, reliability diagrams, and ROC curves. Each of these validation methods were able to highlight different aspects of the ESCM runs. However, it is important to note that no one validation method can fully assess the skill of a model. For example, the ROC curves suggested the ECMWF and BMRC models had similar resolution, yet the reliability diagrams and Brier score highlighted that the ECMWF model was more reliable than the BMRC model. The appropriate validation scores to use depend, to certain extent, on the applications the SCMs are being used for. From the perspective of improving or developing new model parameterizations, both reliability and resolution are important aspects of a forecasting system which should be focused on.

[53] One difficulty with applying validations on model runs from a complete 2 year period is that it is not possible to determine if model errors are associated with particular weather or climate regimes. Traditionally, case studies have

been used to investigate this. While case studies are useful, they are potentially affected by subjectivity in the selection of cases to validate. To overcome this difficulty, Jakob [2003] suggested a method whereby model runs for objectively determined cloud regimes are validated. Section 4.1 showed that this approach can be used with the ESCM method. Both the ECMWF and BMRC SCMs produce substantially lower frequencies for the TWP convective regimes identified by Jakob and Tselioudis [2003] and overpredict the occurrence of the suppressed SSC regime. While the validations for the complete 1999–2000 period show the SCMs underpredict the total cloud cover, this can to a large extent be attributed to the overprediction of the SSC regime. In contrast, for the convective regimes, the SCMs produce substantially more high cloud than is observed. Kolmogorov-Smirnov tests also showed that statistically significant differences between the ECMWF and BMRC predictions can be identified for all the TWP cloud regimes. These differences must be caused by differences in the model physics because the same forcing data were used for both ensembles. This is an important finding because it highlights the ESCM is capable of identifying errors caused by the model physics, apart from errors associated with uncertainties in the forcing data.

[54] As discussed in the introduction, a major motivation for using SCMs is as a framework for testing new or modified parameterizations. Validations on a version of the ECMWF SCM where the penetrative entrainment rate was doubled showed that the ESCM technique can clearly identify changes caused by the modified physics. Furthermore, in the simple example presented here, the influence of the modified cumulus parameterization was as expected from physical considerations. This suggests ensemble single column modelling is a useful technique for testing modifications to parameterizations in a model. For future work, we plan to use the ESCM technique to test a “real” parameterization modification in an operational model used within the Bureau of Meteorology. This will be reported on in a future publication.

[55] **Acknowledgments.** We would like to thank Greg Roff for his efforts in porting and setting up the BMRC SCM on the workstation used in this project. Greg Roff, Peter May, and Beth Ebert also provided helpful comments and suggestions on this paper. We are grateful for support from the U.S. Department of Energy under grant DE-FG02-03ER63533 as part of the Atmospheric Radiation Measurement Program.

References

- Ackerman, T. P., and G. M. Stokes (2003), The atmospheric radiation measurement program, *Phys. Today*, Jan., 38–44.
- Brier, G. W. (1950), Verification of forecasts expressed in terms of probability, *Mon. Weather Rev.*, 78, 1–3.
- Buizza, R. (2000), Accuracy and potential economic value of categorical and probabilistic forecasts of discrete events, *Mon. Weather Rev.*, 129, 2329–2345.
- Clothiaux, E. E., T. P. Ackerman, G. G. Mace, K. P. Moran, R. T. Marchand, M. A. Miller, and B. E. Martner (2000), Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites, *J. Appl. Meteorol.*, 39, 645–665.
- Gregory, D., J.-J. Morcrette, C. Jakob, A. C. M. Beljaars, and T. Stockdale (2000), Revision of convection, radiation and cloud schemes in the ECMWF Integrated Forecasting System, *Q. J. R. Meteorol. Soc.*, 126, 1685–1710.
- Harvey, L. O., K. R. Hammond, C. M. Lusk, and E. F. Mross (1992), The application of signal detection theory to weather forecasting behaviour, *Mon. Weather Rev.*, 120, 863–883.
- Hume, T., and C. Jakob (2005), Ensemble single column modelling (ESCM) in the tropical western Pacific: Forcing data sets and uncertainty analysis, *J. Geophys. Res.*, 110, D13109, doi:10.1029/2004JD005704.
- Iacobellis, S. F., G. M. McFarquhar, D. L. Mitchell, and R. C. J. Somerville (2003), The sensitivity of radiative fluxes to parameterized cloud microphysics, *J. Clim.*, 16, 2979–2996.
- Jakob, C. (2003), An improved strategy for the evaluation of cloud parameterizations in GCMs, *Bull. Am. Meteorol. Soc.*, 84, 1387–1401.
- Jakob, C., and G. Tselioudis (2003), Objective identification of cloud regimes in the Tropical Western Pacific, *Geophys. Res. Lett.*, 30(21), 2082, doi:10.1029/2003GL018367.
- Jakob, C., R. Pincus, C. Hannay, and K.-M. Xu (2004), Use of cloud radar observations for model evaluation: A probabilistic approach, *J. Geophys. Res.*, 109, D03202, doi:10.1029/2003JD003473.
- Jakob, C., G. Tselioudis, and T. Hume (2005), The radiative and thermodynamic properties of the major Tropical Western Pacific cloud regimes, *J. Clim.*, 18, 1203–1215.
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter (2002), NCEP-DOE AMIP-II Reanalysis (R-2), *Bull. Am. Meteorol. Soc.*, 83, 1631–1643.
- Klein, S., and C. Jakob (1999), Validation and Sensitivities of Frontal Clouds Simulated by the ECMWF Model, *Mon. Weather Rev.*, 127, 2514–2531.
- Long, C. N., and T. P. Ackerman (2000), Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects, *J. Geophys. Res.*, 105, 15,609–15,626.
- Long, C. N., and K. L. Gaustad (2004), The shortwave (SW) clear-sky detection and fitting algorithm: Algorithm operational details and explanations, rev. 1, *Atmos. Radiat. Meas. Progr. Tech. Rep. ARM TR-004.1*, U.S. Dept. of Energy, Washington, D. C. (Available at <http://www.arm.gov/publications/techreports.stm>)
- Mace, G. G., C. Jakob, and K. P. Moran (1998), Validation of hydrometeor occurrence predicted by the ECMWF model using millimeter wave radar data, *Geophys. Res. Lett.*, 25, 1645–1648.
- Mapes, B. E., P. E. Ciesielski, and R. H. Johnson (2003), Sampling errors in rawinsonde-array budgets, *J. Atmos. Sci.*, 60, 2697–2713.
- Mason, I. B. (1982), A model for assessment of weather forecasts., *Aust. Meteorol. Mag.*, 30, 291–303.
- McFarlane, S. A., C. N. Long, and D. M. Flynn (2005), Impact of island-induced clouds on surface measurements: Analysis of the ARM Nauru Island effect study data, *J. Appl. Meteorol.*, 44, 1045–1065.
- McFarquhar, S. Iacobellis, and R. C. J. Somerville (2003), SCM simulations of tropical ice clouds using observationally based parameterizations of microphysics, *J. Clim.*, 16, 1643–1664.
- Minnis, P., W. L. Smith Jr., D. P. Garber, J. K. Ayers, and D. R. Doelling (1995), Cloud properties derived from GOES-7 for the spring 1999 ARM intensive observing period using version 1.0.0 of the ARM Satellite Data Analysis Program, *Tech. Rep. RP 1366*, NASA, Washington, D. C.
- Nordeen, M. L., D. R. Doelling, P. Minnis, M. M. Khaier, A. D. Rapp, and L. Nguyen (2001), GMS-5 satellite-derived cloud properties over the tropical western Pacific, paper presented at 11th ARM Science Team Meeting, Atmos. Radiat. Meas. Progr., Washington, D. C. (Available at <http://www.arm.gov/publications/proceedings/conf11/index.stm>)
- Norris, J., and C. Weaver (2001), Improved techniques for evaluating GCM cloudiness applied to the NCAR CCM3, *J. Clim.*, 14, 2540–2550.
- Randall, D. A., and D. G. Cripe (1999), Alternative methods for specification of observed forcing in single-column models and cloud system models, *J. Geophys. Res.*, 104, 24,527–24,545.
- Randall, D. A., K.-M. Xu, R. J. C. Somerville, and S. Iacobellis (1996), Single-column models and cloud ensemble models as links between observations and climate models, *J. Clim.*, 9, 1683–1697.
- Roff, G. (2005), A Single Column Model incorporating BAM physics, *BMRC Res. Lett.*, 3, 6–8. (Available at http://www.bom.gov.au/bmrc/pubs/researchletter/reslett_03/reslett_03.pdf)
- Rossow, W. B., and R. A. Schiffer (1991), ISCCP Cloud Data Products, *Bull. Amer. Meteorol. Soc.*, 72, 2–20.
- Seaman, R., W. Bourke, P. Steinle, T. Hart, G. Embury, M. Naughton, and L. Rikus (1995), Evolution of the Bureau of Meteorology's Global Assimilation and Prediction System. part 1: Analysis and initialisation, *Aust. Meteorol. Mag.*, 44, 1–18.
- Stanski, H. R., L. J. Wilson, and W. R. Burrow (1989), Survey of common verification methods in meteorology, 2nd ed., *World Weather Watch Tech. Rep. 8*, WMO/TD 358, World Meteorol. Soc., Geneva. (Available at http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/Stanski_et_al.html)
- Stokes, G. M., and S. E. Schwartz (1994), The Atmospheric Radiation Measurement (ARM) program: Programmatic Background and Design of the Cloud and Radiation Test Bed, *Bull. Am. Meteorol. Soc.*, 75, 1201–1221.
- Tiedtke, M. (1989), A comprehensive mass flux scheme for cumulus parameterization in large-scale models, *Mon. Weather Rev.*, 117, 1779–1800.

- Tselioudis, G., and C. Jakob (2002), Evaluation of midlatitude cloud properties in a weather and a climate model: Dependence on dynamic regime and spatial resolution, *J. Geophys. Res.*, *107*(D24), 4781, doi:10.1029/2002JD002259.
- Tselioudis, G., Y. Zhang, and W. B. Rossow (2000), Cloud and radiation variations associated with northern midlatitude low and high sea level pressure regimes, *J. Clim.*, *13*, 312–327.
- Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc.*, *131*, 2961–3012, doi:10.1256/qj.04.176.
- Webb, M., C. Senior, S. Bony, and J.-J. Morcrette (2001), Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models, *Clim. Dyn.*, *17*, 905–922.
- Wilks, D. S. (2006), *Statistical Methods in the Atmospheric Sciences*, 2nd ed., 627 pp., Elsevier, New York.
- Williams, K. D., C. A. Senior, A. Slingo, and J. F. B. Mitchell (2005), Towards evaluating cloud response to climate change using clustering technique identification of cloud regimes, *Clim. Dyn.*, *24*, 701–719.
- Xie, S., R. T. Cederwall, M. Zhang, and J. J. Yio (2003), Comparison of SCM and CSRm forcing data derived from the ECMWF model and from objective analysis at the ARM SGP site, *J. Geophys. Res.*, *108*(D16), 4499, doi:10.1029/2003JD003541.
- 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.
- Zhang, J., and U. Lohmann (2005), A microphysical parameterization for convective clouds in the ECHAM5 climate model: Single-column model results evaluated at the Oklahoma Atmospheric Radiation Measurement Program site, *J. Geophys. Res.*, *110*, D15S07, doi:10.1029/2004JD005128.
- Ziehmann, C. (2000), Comparison of a single-model EPS with a multi-model ensemble consisting of a few operational models, *Tellus Ser. A*, *52*, 280–299.

T. Hume and C. Jakob, Bureau of Meteorology Research Centre, GPO Box 1289K, Melbourne, VIC 3001, Australia.

Observed Large-Scale Structures and Diabatic Heating and Drying Profiles during TWP-ICE

SHAOCHENG XIE

Lawrence Livermore National Laboratory, Livermore, California

TIMOTHY HUME

Centre for Australian Weather and Climate Research, Melbourne, Victoria, Australia

CHRISTIAN JAKOB

School of Mathematical Sciences, Monash University, Melbourne, Victoria, Australia

STEPHEN A. KLEIN AND RENATA B. MCCOY

Lawrence Livermore National Laboratory, Livermore, California

MINGHUA ZHANG

School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, New York

(Manuscript received 3 February 2009, in final form 6 July 2009)

ABSTRACT

This study documents the characteristics of the large-scale structures and diabatic heating and drying profiles observed during the Tropical Warm Pool–International Cloud Experiment (TWP-ICE), which was conducted in January–February 2006 in Darwin during the northern Australian monsoon season. The examined profiles exhibit significant variations between four distinct synoptic regimes that were observed during the experiment. The active monsoon period is characterized by strong upward motion and large advective cooling and moistening throughout the entire troposphere, while the suppressed and clear periods are dominated by moderate midlevel subsidence and significant low- to midlevel drying through horizontal advection. The midlevel subsidence and horizontal dry advection are largely responsible for the dry mid-troposphere observed during the suppressed period and limit the growth of clouds to low levels. During the break period, upward motion and advective cooling and moistening located primarily at midlevels dominate together with weak advective warming and drying (mainly from horizontal advection) at low levels. The variations of the diabatic heating and drying profiles with the different regimes are closely associated with differences in the large-scale structures, cloud types, and rainfall rates between the regimes. Strong diabatic heating and drying are seen throughout the troposphere during the active monsoon period while they are moderate and only occur above 700 hPa during the break period. The diabatic heating and drying tend to have their maxima at low levels during the suppressed periods. The diurnal variations of these structures between monsoon systems, continental/coastal, and tropical inland-initiated convective systems are also examined.

1. Introduction

The large-scale state of the tropical atmosphere as characterized by low-level convergence and advective

cooling and moistening plays an important role in destabilizing the atmospheric structure, initiating and maintaining deep convection. On the other hand, latent heating released from tropical convective systems is a major energy source for the large-scale circulation. By releasing latent heat and vertically redistributing sensible heat and water vapor, cumulus clouds modify the temperature and moisture structure of the environment

Corresponding author address: Shaocheng Xie, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94551-0808.
E-mail: xie2@llnl.gov

(Riehl and Malkus 1958; Yanai 1961; Yanai and Johnson 1993). Documenting the large-scale structure and latent heating profiles of tropical convective cloud systems from observations is a key step toward understanding how cumulus convection interacts with its large-scale environment. This has been one of the primary goals of a number of major field experiments conducted in the tropics (Thompson et al. 1979; Frank 1978; Frank and McBride 1989; Lin and Johnson 1996a,b; Schumacher et al. 2007). Retrieval of latent heating profiles from satellite measurements is also a major research topic of the National Aeronautics and Space Administration's (NASA) Tropical Rainfall Measuring Mission (TRMM) (Tao et al. 2006, 2007).

Latent heating is the dominant component of total diabatic heating in the tropics during convective periods. The total diabatic heating and drying can be estimated as the residuals of heat and moisture budgets of large-scale motion measured by a suitably designed sounding network. The purpose of this study is to document the characteristics of the large-scale structure and diabatic heating and drying profiles of the convective cloud systems observed during the latest major field experiment in the tropics, the Tropical Warm Pool–International Cloud Experiment (TWP-ICE). The derived diabatic heating data can also be used as an observed base for validating TRMM latent heating retrievals, which are designed for application with satellite-estimated surface rain rate and precipitation profile inputs using physically based models, which range from simple profile shapes to cloud-resolving models (CRMs). As shown in Tao et al. (2006), different retrieval algorithms may result in large differences in the retrieved latent heating profiles from TRMM measurements.

TWP-ICE was conducted in January–February 2006 in Darwin during the northern Australian monsoon season by the U.S. Department of Energy (DOE) Atmospheric Radiation Measurement Program (ARM) (Stokes and Schwartz 1994; Ackerman and Stokes 2003) and the Australian Bureau of Meteorology (BOM) in collaboration with many other agencies and universities (May et al. 2008) to further improve the understanding of the interaction of tropical convection with its environment. The major goal of TWP-ICE was to provide a comprehensive dataset that describes the structure and evolution of monsoonal deep tropical convection and the resulting cirrus clouds and their radiative impacts. In conjunction with the description of the convection, the experiment aimed to provide an extensive description of the large-scale state of the atmosphere on the scale of a typical climate model grid box ($\sim 2^\circ \times 2^\circ$). For this purpose, a comprehensive ground-based observational network, including an array of six radiosonde stations centered at the ARM Darwin site and an extensive set of

in situ and remote sensing instruments, was deployed to observe the evolution of convective cloud systems, their large-scale environment, and the detailed cloud microphysical properties (Fig. 1). One of the unique features of TWP-ICE is that radiosondes were launched with very high temporal resolution (3-hourly launches) at all of the five boundary sounding stations. Consequently, the diurnal variability of convective cloud systems can be better resolved by TWP-ICE data than in data from earlier tropical field experiments.

The northern Australian summer monsoon season usually begins in mid- to late December and ends in March. Climatologically, the monsoonal flow is characterized by low-level westerlies of equatorial origin overlain by strong easterlies in the upper troposphere, with “break” periods where low- to midlevel easterlies of continental origin prevail. Characteristics of the convection and its associated large-scale environment during the northern Australian summer monsoon season have been described in many studies. An overview of the northern Australian summer monsoon was provided by McBride (1987) and Manton and McBride (1992); Gunn et al. (1989) described the synoptic aspects of the 1987 Australian summer monsoon during the Australian Monsoon Experiment (AMEX) over the Gulf of Carpentaria; Frank and McBride (1989) performed a detailed investigation of diabatic heating and drying structures and the evolution of the life cycle of the convective cloud systems for AMEX. They found that AMEX convective systems typically showed stronger upward motion and more heating in the mid- and upper levels but weaker upward motion and less heating in the lower troposphere throughout their life cycles compared to those systems observed in Atlantic Ocean during the Global Atmospheric Research Program (GARP) Atlantic Tropical Experiment (GATE). Keenan and Carbone (1992) showed differences in the structural characteristics of tropical precipitating systems of oceanic, continental/coastal, and island origin observed in the vicinity of Darwin. Tropical island-initiated convective systems over the Darwin region were studied in Carbone et al. (2000) and were the major focus of the Maritime Continent Thunderstorm Experiment (MCTEX) held during November–December 1995 over the Tiwi Islands north of Darwin (Keenan et al. 2000).

During TWP-ICE, Darwin experienced four distinct synoptic regimes, coincident with the activity of a major Madden–Julian oscillation (MJO) event that influenced the region (see the MJO index online at <http://www.bom.gov.au/bmrc/clfor/cfstaff/matw/maproom/RMM/>). Various types of tropical convective systems of oceanic, continental/coastal, and island origin were observed during the experiment. The experiment provided a unique

FIG. 1. The ground observation network established during TWP-ICE. The analysis domain is enclosed by the five boundary sounding stations. The mainland subdomain and the Tiwi Islands subdomain are enclosed by the dashed lines. The locations of the sounding, surface radiative flux, and surface turbulence flux stations are displayed as stars, squares, and dots, respectively.

opportunity to study the structure and evolution of these tropical convective systems and their associated environments in an unprecedented level of detail. An overview of TWP-ICE and its synoptic conditions can be found in May et al. (2008).

The focus of this study is to elucidate the variations in the TWP-ICE large-scale structures as well as diabatic heating and drying profiles within the four observed regimes. We attempt to understand how the large-scale atmospheric structure interacts with the organization of the TWP-ICE convective systems and to provide a comprehensive description of the large-scale environment and diabatic heating and drying profiles for future studies on TWP-ICE. An important by-product of the present study are the derived large-scale fields (e.g., vertical velocity and advective tendencies of temperature and moisture), which can be used for heat and moisture budget studies as well as single-column modeling (SCM) and CRM studies of the observed convec-

tive cloud systems. Note that providing such large-scale fields was one of the key goals of the TWP-ICE field campaign (May et al. 2008).

Section 2 provides details about the data and data analysis issues. The synoptic conditions and associated cloud fields are briefly summarized in section 3. Section 4 describes characteristics of the analyzed mean flow and surface conditions. Section 5 examines large-scale structures and diabatic heating and drying profiles exhibited over the four distinct regimes. A further investigation of differences in these fields relative to the diurnal cycle of convection of oceanic, continental/coastal, and island origin is conducted in section 6. Section 7 summarizes the results.

2. Data and data analysis issues

The constrained variational analysis approach of Zhang and Lin (1997) is used to perform an objective analysis of

TWP-ICE data. This method was designed for deriving the large-scale vertical velocity and advective tendencies of temperature and moisture from sounding measurements over a network with a small number of stations. Diabatic heating and drying, or the apparent heat source Q_1 and apparent moist sink Q_2 as defined by Yanai et al. (1973), can then be estimated as the residuals of heat and moisture budgets of large-scale motion. The variational analysis uses domain-averaged surface and top-of-the-atmosphere (TOA) observations as constraints to adjust atmospheric state variables from soundings by the smallest possible amount to conserve column-integrated mass, moisture, and static energy so that the final analysis dataset is dynamically and thermodynamically consistent. The required observation constraints include the surface and TOA radiative fluxes, surface latent and sensible heat fluxes, and surface precipitation. The variational analysis has been successfully used to process field data collected at the ARM Southern Great Plains (SGP) (Zhang et al. 2001) and North Slope of Alaska (NSA) sites (Xie et al. 2006). It was also used for analyzing data from the TRMM Kwajalein Experiment (KWAJEX), the TRMM Large-Scale Biosphere–Atmosphere Experiment (LBA), and the TRMM South China Sea Monsoon Experiment (SCSMEX) (Schumacher et al. 2007). Details of the variational analysis method can be found in Zhang and Lin (1997) and Zhang et al. (2001). In this section, we provide some technical details about the objective data analysis of TWP-ICE observations to help users better understand the large-scale state dataset developed from this study.

The TWP-ICE sounding array (Fig. 1) provided three hourly soundings during the intensive operational period (IOP) from 22 January to 12 February 2006. The array contains four mainland stations (including the ARM Darwin site), one station on the Tiwi Islands, and one station on the Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO) *Southern Surveyor* ship stationed in the Timor Sea. The sounding array covers an area with a radius of approximately 150 km. The variational analysis technique described above was applied to the sounding IOP over the domain enclosed by the five boundary sounding stations (Fig. 1). In addition, two subanalysis domains were defined to better describe the convective systems of continental/coastal and island origin observed during the break period (respectively the mainland and Tiwi Islands subdomains shown by dashed lines in Fig. 1) since those systems occupied only small regions of the experiment domain as will be shown in the next section. Both continental/coastal and island systems were mainly driven by sea breezes, but the continental coastal systems were also influenced by several squall lines crossing Darwin in

the evening and early morning during TWP-ICE. Therefore, it would be interesting to examine similarities and differences in the structural characteristics of these two types of cloud systems. In the following discussion, the analysis domain refers to the whole TWP-ICE sounding domain unless the mainland or Tiwi Islands subdomains are explicitly indicated. The temporal and vertical resolutions of the analyses used in this study are 3 h and 25 hPa, respectively, for both the sounding domain and the two subdomains.

The variational analysis requires initial analyses of temperature, wind, and humidity for the TWP-ICE domain. These are generated from the TWP-ICE soundings and background European Centre for Medium-Range Weather Forecasts (ECMWF) analyses using an interpolation scheme described by Cressman (1959). The Cressman scheme uses a weighting function that depends on the distance between an observation station and an analysis grid point, as well as the difference between observations and the background. The interpolation is carried out for the difference field between observations and the background. If there is no measurement within a specified distance (L) of an analysis grid point, the scheme will only take the background field as the analysis. The distance L used in this study is 50 km.

The surface and TOA observations, used as the constraints in the variational analysis, were measured by the enhanced surface observation network near Darwin shown in Fig. 1 and the Japanese Multifunctional Transport Satellite (MTSAT), respectively. Among these constraint fields, the surface precipitation rates were derived from a scanning C-band Doppler radar at Gunn Point (near Darwin) using a combination of polarimetric estimators that are dominated by reflectivity at low rain rates, differential reflectivity at moderate rates, and the specific differential phase (KDP) at high rain rates. The particular algorithms are described in Bringi et al. (2001, 2002, 2004). The surface radiation was obtained from radiometer measurements using the algorithm described in Long and Ackerman (2000) and Long (2001). The surface turbulent fluxes were derived from 3D eddy covariance system measurements according to the method described in Beringer et al. (2007). The TOA radiative fluxes were from hourly MTSAT products, which were retrieved using the algorithm described in Minnis et al. (2001).

Since both the C-band polarimetric (C-POL) precipitation radar and MTSAT covered the analysis region completely, a simple arithmetic averaging was used to calculate the domain average precipitation rates and TOA radiative fluxes. For the surface radiative and turbulent fluxes, a weighted average was applied to the data collected from the limited number of surface flux

sites. The weights used in computing the area mean of the surface radiative fluxes were based on the land and ocean fraction within the analysis domain, 0.57 and 0.43, respectively. For the sensible and latent heat fluxes, the weights for data collected at each site (i.e., Darwin Harbour, Fogg Dam, Howard Springs, Daly River, and the ship) were guided by the frequencies of the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite-derived land types within the analysis domain. As shown in Fig. 1, the Daly River site is located outside the analysis domain. However, data collected from the site are important for us to obtain meaningful area-averaged heat fluxes since the Daly River site is surrounded with eucalypt woodland/grassland savanna, one of the most common land types in the Darwin region. The locations of the other four heat flux sites were selected to represent inshore waters (Darwin Harbour), typical northern floodplain (Fogg Dam), eucalypt open forest savanna (Howard Springs), and open ocean (the ship). There were quite large gaps in the surface heat flux measurements. The major missing data periods are 22–25 January 2006 and 1–6 February 2006 at Fogg Dam, 22–23 January 2006 and 1–10 February 2006 at Howard Springs, 22–31 January 2006 at Daly River, 3–7 February 2006 at Darwin Harbour, and 22–25 January 2006 at the ship. The data gaps were filled with ECMWF data, which were adjusted using linear regression equations that were derived for the times when both ECMWF and observations were available.

The Vaisala RS92 sondes used in TWP-ICE are known to have a significant daytime solar radiation dry bias and a minor temperature-dependent calibration error (Vömel et al. 2007). The dry bias is caused by solar heating of the radiosonde's humidity sensor; at the height of 15 km this bias can be as great as 50%. In this study, the dry bias and the temperature-dependent calibration errors in the original RS92 sondes were corrected using an algorithm similar to the one described in Vömel et al. (2007). A radiation dry bias correction factor, a temperature-dependent calibration correction factor, and a solar zenith angle correction factor were applied to the original RS92 sondes (Hume 2007). The largest correction is in the upper troposphere during daylight; however, small changes can also be seen at night and in the lower troposphere mainly because of the use of the temperature-dependent calibration correction factor. More details about the correction can be found in the appendix.

The subdomain large-scale state was derived using the same approach as that used for the entire sounding domain, except that the surface and TOA constraints are averaged over the corresponding subdomains. Sounding data from those stations that are within the specified

distance (50 km) of the analysis grid points were used in the analysis through the Cressman method as described earlier. As indicated in Fig. 1, the mean flow and large-scale structures over the mainland subdomain were determined mainly by the sounding data collected at Darwin, Mount Bundy, and Point Stuart, while over the Tiwi Islands subdomain they were determined mainly by the sounding data collected at Pirlangimpi, *Southern Surveyor*, and Cape Don. It is worth noting that the use of subdomain-averaged constraints helps alleviate some of the problems caused by scale aliasing in the measurements over the two smaller subdomains that are originally introduced by the reduction of the number of sounding stations used in the analysis.

It is believed that there are quite large uncertainties in the area-mean surface fluxes because of the limited number of surface flux stations and large data gaps. The small number of surface flux stations cannot well resolve the large spatial variability exhibited in the radiative and heat fluxes in the Darwin region, which contains both ocean and diverse land surface types. Several sensitivity tests were performed to examine the impact of potential uncertainties in the domain-averaged surface fluxes on the derived large-scale structures. In these tests, the area-averaged surface radiative and turbulent fluxes were derived using different weighting coefficients for the data collected from the flux stations. It was found that the derived large-scale fields from these sensitivity tests exhibited very similar temporal and vertical structures with some variations in their magnitudes. The differences between these sensitivity tests and the standard run (which is used in this study) are at least one order of magnitude smaller than typical values of the large-scale fields in the standard run (not shown). This suggests that the analyzed data are suitable for studying large-scale structures for the convection systems observed during TWP-ICE.

3. Synoptic conditions and observed clouds

As described in May et al. (2008), the Darwin region was influenced by a typical monsoonal circulation during TWP-ICE. It experienced four distinct regimes: active monsoon during 13–25 January, suppressed monsoon during 26 January–2 February, clear skies during 3–5 February, and a monsoon break period during 6–13 February. The active monsoon period was characterized by westerly monsoon flow, intensive mesoscale convective systems of mostly oceanic origin, and heavy surface rainfall. During the suppressed monsoon period, an intense low pressure system moved inland to the southwest of Darwin where it deepened, leading to strong westerlies over Darwin as is typical in the days following the

active phase of an MJO. This period was characterized by clouds associated with relatively shallow convection (cloud tops at or below 8 km) accompanied by much lower area-mean precipitation than in the preceding monsoon period. Cirrus resulting from the inland low to the south of Darwin was advected into the TWP-ICE domain. Following the dissipation of the inland low on 3 February, the Darwin region experienced three days of clear skies with no precipitation. This was followed by a monsoon break period, during which an inland heat trough dominated the large-scale conditions over north Australia. Convection during the break period was characterized by intense afternoon thunderstorms with several squall lines crossing Darwin in the evening and early morning.

The cloud and precipitation systems that occurred during TWP-ICE were well observed by the ARM surface remote sensing instruments, including a 35-GHz millimeter wave cloud radar (MMCR), micropulse lidar (MPL), laser ceilometers, and the BOM C-POL radar. Figure 2a shows the time–pressure cross section of observed frequency of occurrence of clouds at the ARM Darwin site. The figure is constructed by integrating measurements from MMCR, MPL, and laser ceilometers using the Active Remotely Sensed Clouds Locations (ARSCL) algorithm (Clothiaux et al. 2000) for the sounding IOP. The integrated products combine the capability of the MMCR to penetrate optically thick hydrometeor layers and the ability of the MPL to detect optically thin clouds (e.g., cirrus). Cloud base shown in Fig. 2a is determined by the MPL and the laser ceilometers, which are usually insensitive to precipitation particles (if they are not sufficiently large) or clutter. As indicated in Clothiaux et al. (2000), the laser ceilometers and MPL can provide quite accurate cloud-base measurements. The area-average surface precipitation rates retrieved from the C-POL radar are shown in Fig. 2b.

Examples of the horizontal distribution of the observed cloud systems and surface rainfall are displayed in Figs. 3 and 4, respectively. The pentagon in these snapshot figures denotes the TWP-ICE sounding domain. It is seen that the major cloud systems and surface rainfall associated with the active monsoon covered much of the analysis region, while those associated with intense afternoon convection and squall lines during the break period only occupied scattered regions (over the mainland and the Tiwi Islands) of the analysis domain. This suggests that the current sounding network might be too big to appropriately describe the break period systems. This motivated the analysis over the mainland subdomain and the Tiwi Islands subdomain carried out in section 6.

FIG. 2. (a) Observed cloud frequency at Darwin from the ARM surface remote sensing instruments (MMCR, MPL, and laser ceilometer) and (b) the domain-averaged surface precipitation rates from the BOM C-POL radar. The four regimes are marked in (b).

4. Mean flow and surface conditions

The large-scale environment of the Australian summer monsoon as analyzed with the variational analysis is shown in Figs. 5a–c, which display the analyzed zonal and meridional wind components and relative humidity (RH) fields, respectively. RH is calculated with respect to ice for $T < 0^{\circ}\text{C}$. Note that the analyzed data represent an average over the analysis domain as shown in Fig. 1. A 24-h running smoothing was applied to all fields for display purposes.

The analyzed winds clearly demonstrate the northern Australia summer monsoonal flow observed during TWP-ICE. The monsoon period (active, suppressed, and clear skies) was characterized by westerlies at low-to midlevels, whereas the break period was characterized by easterlies at the levels above 915 hPa. The easterly flow observed on 24 January at the levels above 915 hPa was due to the monsoon trough retreating to the north of Darwin (May et al. 2008). A strong westerly wind event was observed during the suppressed monsoon period, following the active phase of an MJO. At the same time, an easterly flow occurred in the upper

FIG. 3. Four snapshots showing the satellite images for the different cloud systems observed during (a) the active; (b) suppressed; and (c),(d) break periods. The pentagon in these figures denotes the TWP-ICE sounding domain. Local time in Darwin is 9.5 h greater than universal time. Black represents clear-sky area.

troposphere. The meridional wind component is much weaker than the zonal wind component, typically less than 5 m s^{-1} in the troposphere (Fig. 5b). The temporal and vertical structure of RH shows a strong association with the observed cloud fields (Fig. 2). The RH was usually larger than 80% throughout the troposphere during the active monsoon period. Low RH values were seen in the midtroposphere during the suppressed, clear skies, and break periods. The dry air was likely related to dry intrusions originated from higher latitudes, which are often observed over the tropical western Pacific

(Numaguti et al. 1995). The dry midtroposphere could act to suppress the convection and prevent the formation of deep cumulus clouds.

Figures 6a–d show the analyzed domain-averaged surface temperature, surface wind speed, sensible heat flux, and latent heat flux, respectively. These surface fields exhibited notable differences over the four regimes. For example, the break period had higher surface temperature and sensible heat flux but weaker surface wind and latent heat flux compared to the active and suppressed monsoon periods. It is seen that the

FIG. 4. Images of the C-POL radar precipitation for the similar times as the satellite pictures shown in Fig. 3. The pentagon in these figures denotes the TWP-ICE sounding domain. Local time in Darwin is 9.5 h greater than universal time.

sensible heat flux showed some correlation with the surface temperature, whereas the surface latent heat flux was better correlated with the surface wind speed. Both the surface wind and latent heat flux (or evaporation) were significantly enhanced following the passage of the active phase of the MJO event. The increase of surface temperature with time was consistent with an increase of net surface shortwave radiation (Fig. 7a) due to a decrease of cloud fraction especially at low levels. The reduction of clouds from the active monsoon to the break period also led to an increase of net downward shortwave and outgoing longwave radiative fluxes at TOA (Figs. 7b and 7d, respectively) and an increase of net surface upward longwave radiation (Fig. 7c).

Associated with the activity of a major MJO event over the Darwin region, the observed characteristics of the large-scale environment during the active and suppressed monsoon period of TWP-ICE show many

similarities to those observed over the tropical western Pacific warm pool region, such as those described in Lin and Johnson (1996a) for the Tropical Ocean and Global Atmosphere Coupled Ocean–Atmosphere Response Experiment (TOGA COARE) conducted over the western Pacific warm pool, in which the large-scale circulation was also largely impacted by major MJO events. For example, both TWP-ICE and TOGA COARE data show significantly increased westerlies and evaporation following the active phase of an MJO event. This indicates that the TWP-ICE data could help understand tropical convective systems that occur over the warm pool region. However, they differ significantly during the break period, where TWP-ICE had the unique continental/coastal and island-initiated convective systems that occurred in the Darwin region almost on a daily basis. Data collected from TWP-ICE provide a unique opportunity to study these tropical land systems and their associated environments.

FIG. 5. The time–pressure cross sections of the analyzed large-scale state fields over the sounding domain during TWP-ICE. (a) Zonal wind, (b) meridional wind, and (c) relative humidity (with respect to ice for $T < 0^{\circ}\text{C}$). In these figures, solid lines are for values larger than or equal to zero and dotted lines for values less than zero. In (a) and (b), positive values are shaded and in (c) values larger than 60% are shaded. Contour interval is 5 in (a) and (b) and 20 in (c). A 24-h running mean was applied to these fields.

5. Mean large-scale structures and diabatic heating and drying profiles

One of the major goals of TWP-ICE is to determine mean large-scale dynamic and thermodynamic structures and diabatic heating and drying profiles associated with tropical convective cloud systems. Details of the vertical heating profile are known to have large influence on both tropical weather and climate systems and are central to their intraseasonal variability (Yanai et al. 2000; Lin et al. 2004). In this section, we examine the evolution of the derived large-scale state and diabatic heating and drying as well as their vertical structures over the four distinct regimes.

The temporal evolution and period-averaged large-scale vertical velocity derived from TWP-ICE observations over the whole sounding domain are shown in Figs. 8a and 8b, respectively. The mean vertical velocity field during the active monsoon period showed strong upward motion through the entire troposphere with maximum around 265 hPa (Fig. 8b). In contrast, the av-

erage vertical motion was weak during the suppressed, clear, and break periods. The suppressed monsoon period was characterized by moderate upward motion below 665 hPa and above 245 hPa and weak downward motion in the midtroposphere. Compared to the suppressed period, the three clear days had stronger downward motion above 465 hPa and weaker upward motion below 665 hPa. During the break period, moderate midlevel upward motion occurred with virtually no vertical velocity below 765 hPa in its averaged profile. The characteristics of the vertical velocity field during the monsoonal period are similar to those shown in TOGA COARE (Lin and Johnson 1996a) in which the upward motion maxima were found at upper levels during the active convective period, and low-level weak upward motion or subsidence was observed during suppressed convective period when strong westerly winds prevailed. The major difference between these two datasets is that the ascending maximum of the TWP-ICE profile occurs higher (265 hPa), indicating higher clouds or stronger updrafts, than that of the TOGA

FIG. 6. Time series of (a) analyzed surface temperature, (b) surface wind speed, (c) surface sensible heat flux, and (d) surface latent heat flux, over the sounding domain during TWP-ICE. A 24-h running smoothing was applied to these fields.

COARE profile (350 hPa) during the active convection period.

Figure 9 displays the analyzed total advective temperature tendency (T_{adv_tot}) and its horizontal component (T_{adv_hori}), which are defined as

$$\begin{aligned} T_{adv_tot} &= -\mathbf{V}_h \cdot \nabla_p T - \omega \frac{\partial T}{\partial p} + \frac{\omega}{\rho C_p} \\ &= T_{adv_hori} - \omega \frac{\partial T}{\partial p} + \Delta T_{adiab}, \end{aligned} \quad (1)$$

where T is the temperature, $\mathbf{V}_h$ is the horizontal wind vector, ΔT_{adiab} is the adiabatic compression term, p is the pressure, and ω is the vertical velocity in p coordinates. The total advective temperature tendency was mainly dominated by its vertical component (including the adiabatic compress term) during the active and break periods as the horizontal advective tendency of temperature was relatively weak. For the active mon-

soon period, the entire column was cooled primarily by vertical advection with the maximum cooling between 340 and 565 hPa (Fig. 9c), whereas for the break period moderate cooling was seen only at midlevels. Unlike the active and break periods, the suppressed period experienced pronounced horizontal advective cooling between the surface and 215 hPa (Fig. 9d). The total advective cooling rate decreased with height above 765 hPa, with the largest cooling occurring in the lower troposphere (Fig. 9c). Another interesting feature seen in Fig. 9d was that the boundary layer was largely cooled by horizontal advection during the three clear days.

Figure 10 shows the temporal evolution and period-averaged total and horizontal advective tendencies of moisture, q_{adv_tot} and q_{adv_hori} , which are defined as

$$q_{adv_tot} = -\mathbf{V}_h \cdot \nabla_h q - \omega \frac{\partial q}{\partial p} = q_{adv_hori} - \omega \frac{\partial q}{\partial p}, \quad (2)$$

FIG. 7. Time series of (a) analyzed surface net shortwave radiative flux, (b) TOA net shortwave radiative flux, (c) surface net longwave radiative flux, and (d) TOA net longwave radiative flux, over the sounding domain during TWP-ICE. A 24-h running smoothing was applied to these fields.

where q is water vapor mixing ratio. For the active period, strong moistening existed through the entire troposphere with two peaks respectively at around 715 and 465 hPa. For the break period, the moistening was mainly located at midlevels and was associated with weak drying in the boundary layer. Notable midlevel drying with the peak around 615 hPa was found for the suppressed period and significant low-level drying was found over the three clear days (Fig. 10c). The drying tendencies were primarily due to the large horizontal advection seen during these two periods (Fig. 10d). The mid- and low-level horizontal dry advection would seem to be partially responsible for the substantially dry midtroposphere seen in the RH field during the suppressed monsoon period. It is interesting to see from Fig. 10d that basically all the four subperiods experienced horizontal dry advection, indicating that the moistening tendencies shown in Fig. 10c were from vertical advective tendencies of moisture.

The diabatic heating (Q_1) and drying (Q_2) derived from the variational analysis are given in Figs. 11 and 12, respectively. The Q_1 and Q_2 fields showed significant variations over the four different regimes. The heating and drying were strongest during the active monsoon period, consistent with the heaviest rain produced there. The averaged Q_1 profile of the monsoon systems exhibited a maximum between 340 and 540 hPa. The upper-level heating peak was associated with both deep convection and stratiform precipitation. The shape of the mean profile of the active monsoon cloud systems is very similar to that diagnosed from other tropical regions, for example, the heating profiles derived from the AMEX systems (Frank and McBride 1989) and the TOGA COARE systems (Lin and Johnson 1996b), which all had maximum heating in the mid- and upper troposphere.

The Q_2 profile of the monsoon systems showed two major drying layers: one above 565 hPa and another one

FIG. 8. Analyzed vertical velocity over the sounding domain during TWP-ICE. (a) Temporal evolution (dotted lines for values less than -1 hPa hr^{-1}); (b) averaged profiles for different periods.

below 665 hPa. The relative minimum in Q_2 between 565 and 665 hPa corresponds to the minimum in the total moisture advection. The double-peak structure in Q_2 is similar to that observed in tropical convective systems over the Marshall Islands (Yanai et al. 1973) and northern Australia during AMEX (Frank and McBride 1989). As suggested in Johnson (1984), the double-peak structure might be related to cumulus updrafts that contribute to the lower peak and mesoscale updrafts that result in the higher one. It is noted that the drying maximum occurred at levels just slightly lower than the heating maximum, implying a significant contribution from stratiform rain to the Q_2 profile.

Compared to the active monsoon period, the heating was moderate and was confined above 700 hPa for the break period. Below this level, weak cooling was observed. The cooling was likely related to evaporation of shallow cumulus clouds and convective and mesoscale downdrafts, which also led to moderate moistening there. The heating and drying during the suppressed period was

mainly located at low levels, consistent with the prevalence of shallow cumulus clouds. Quite strong moistening was observed at midlevels, likely associated with evaporation from the low- and midlevel clouds. For the clear days, moderate cooling was seen above 565 hPa, presumably because of clear-sky radiative cooling, and strong moistening was found around 815 hPa. Note that the “clear days” had nonprecipitating shallow convection, as indicated by ARSCL (Fig. 2a). The strong similarity between Q_2 and $q_{\text{adv_hori}}$ suggests that convection is countering the effects of horizontal dry air advection.

6. The diurnal cycle of oceanic, mainland, and island-initiated cloud systems

The 3-hourly radiosonde launches during TWP-ICE provide a unique opportunity to study the diurnal cycle of the cloud systems observed during the northern Australian monsoon season. In this section, we focus our analysis on the diurnal cycle of the large-scale structures and diabatic heating and drying profiles with those convective cloud systems that occurred under different surface boundary conditions. The convective systems examined include the intensive monsoon systems of oceanic origin that occurred between 0000 local standard time (LST) 22 January and 2400 LST 25 January during the active monsoon period and the mainland and island-initiated convective systems that were observed between 0000 LST 6 February and 2400 LST 11 February during the break period. To more appropriately describe the mainland and island-initiated convection we carried out additional analyses for the two smaller subregions shown in Fig. 1. The results of these analyses will be used here and will be referred to as the mainland domain and the Tiwi Islands domain, respectively. For the large-scale monsoon cloud systems, the analysis data over the whole sounding domain are used. In the following discussions, we use the term “mainland systems” to refer to those convective systems that occurred over the mainland domain and “island systems” to refer to those convective systems that were observed over the Tiwi Islands domain during the break period.

a. Precipitation

According to earlier studies (e.g., Nesbitt and Zipser 2003), tropical precipitation tends to peak in the afternoon over land areas because of afternoon boundary layer destabilization caused by daytime insolation and in the early morning over adjacent ocean regions associated with the long, nocturnal life cycle of mesoscale convective systems. Similar diurnal features are seen in the convective systems observed during TWP-ICE

FIG. 9. Analyzed advective tendency of temperature over the sounding domain during TWP-ICE. (a) Temporal evolution of total advective tendency of temperature; (b) temporal evolution of horizontal advective tendency of temperature. In (a) and (b), dotted lines for values less than -1 K day^{-1} . (c) Averaged total tendency profiles for different periods; (d) averaged horizontal tendency profiles for different periods.

(Fig. 13). The monsoon precipitation associated with intensive mesoscale convective systems tends to peak in the early morning hours (around 0300 LST), while the island precipitation reaches its maximum value at 1500 LST in the afternoon. In contrast, the mainland precipitation shows a semidiurnal variation, with one peak in the early morning (0600 LST) and the other one in the late afternoon (1800 LST). The daytime convective systems over the mainland and Tiwi Islands domains were mainly driven by sea breezes due to the differential heating between the ocean and the land. Shallow cloud systems tended to form around midmorning (around 0900–1000 LST) along the coast and then developed into precipitating deep convective systems in the afternoon. The time for coastal convective clouds to propagate into inland likely results in the 3-h delay for the mainland precipitation to reach its maximum compared to the island systems. The early morning peak seen in the main-

land precipitation is associated with the passage of squall lines that originated outside the measurement domain (May et al. 2008).

b. Cloud frequency

The diurnal cycle in the profiles of cloud occurrence (Figs. 14a,b) is computed from the ARM cloud measurements at the Darwin site. The single-point measurements can be considered to roughly represent the monsoon clouds (which usually covered a large fraction of the experimental domain) and the clouds associated with the coastal systems over the mainland because of their proximity to the measurement site. There were no cloud observations over the Tiwi Islands domain.

Significant diurnal variations are seen in cloud occurrence for both the monsoon and mainland systems. The monsoon system has significant amounts of clouds through the troposphere at night and relatively few clouds

FIG. 10. Same as Fig. 9 except for the advective tendency of moisture.

during the day, consistent with the diurnal variation of the monsoon precipitation as shown in Fig. 13. In contrast, the diurnal evolution of the mainland clouds displays a clear progression from shallow to deep convection during the day: shallow cumulus clouds grow atop the daytime boundary layer and then develop into deep convective clouds in the late afternoon. It is interesting to see that the mainland clouds peak at midnight with the prominent maximum value seen in upper troposphere rather than peak around the time of the maximum precipitation. This might reflect the fact that the sea-breeze precipitation has moved inland away from Darwin so that the ARM Darwin site only observes the high clouds that have advected over the site from the deep convection that is further inland. The clouds seen at night and in the early morning hours are also related to the passage of squall lines over the mainland domain. It is noted that there is a general lack of midlevel clouds in the mainland systems. This might be due to the dry

midtroposphere observed during the break period, as shown in Fig. 5c.

c. Vertical velocity

Consistent with the convective activity, Figs. 15a–c show strong diurnal variations in the large-scale vertical velocities associated with the three types of convective systems. For the monsoon systems, the large-scale environment is characterized by upward motion throughout the troposphere over the composite period. Upward motion is weak during the day. It increases rapidly in the evening (around 2100 LST) associated with the development of nocturnal deep convection. The strongest ascent is seen in the upper troposphere between 265 and 465 hPa at 0300 LST when precipitation reaches its maximum intensity. The upward motion then decreases gradually between 0300 and 1200 LST while the maximum ascent still remains above 565 hPa, corresponding to the development of stratiform clouds.

FIG. 11. Analyzed diabatic heating rate (Q1) over the sounding domain during TWP-ICE. (a) Temporal evolution (dotted lines for values less than -1 K day^{-1}). (b) Averaged profiles for different periods.

The mainland systems exhibit a layer of weak vertical motion at low levels in the midmorning (around 0900 LST) that is presumably related to the low-level convergence caused by the sea breezes. Interestingly there is a subsidence layer at midlevels in the late morning and early afternoon that contributes to confining the mainland convection to low levels and presumably also prevents convection from developing spontaneously anywhere in the domain at the initial stage. As the sea-breeze system develops, quite strong upward motion is observed in the levels above 665 hPa with weak vertical motion below. The upward motion peaks at 1500 and 2400 LST with the afternoon peak (around 315 hPa) higher than the midnight peak (around 465 hPa). It is noted that the peaks in vertical velocity occur a few hours earlier than those in precipitation for the mainland systems, indicating a time lag between the dynamic structure and the formation of significant rainfall over the mainland region.

FIG. 12. Analyzed diabatic drying rate (Q2) over the sounding domain during TWP-ICE. (a) Temporal evolution (dotted lines for values less than -1 K day^{-1}). (b) Averaged profiles for different periods.

The diurnal cycle of vertical velocity for the island systems is quite different from that of the mainland systems. The island systems display a strong diurnal variability in vertical velocity in the mid- and upper troposphere. The strongest upward motion is seen around 465 hPa at 1500 LST when the surface precipitation reaches its maximum and the downward motion peaks around 415 hPa between 0600 and 0900 LST. The upward motion remains weak and constant below 665 hPa during most of the day except for midnight and the early morning hours. This is different from the mainland systems, which show a moderate subsidence layer below 665 hPa near noon. Compared to the mainland systems, the island systems show a much deeper ascending layer in the lower troposphere during the day and a more significant subsidence layer in the mid- and upper troposphere in the midmorning (0900–1000 LST) when convection starts. The subsidence layer acts to restrict the island convection at low levels (shallow convection) at its initial stage,

FIG. 13. Diurnal cycle of surface precipitation rates (mm day^{-1}). Solid line is for monsoon systems, dashed line is for mainland systems, and dot-dashed line is for island systems.

similar to the mainland convection. This is consistent with the results from earlier studies (e.g., Simpson et al. 1993), which indicated that the island-initiated convection was characterized by widespread shallow convection at its initial stage over the islands.

d. Advective tendencies of temperature and moisture

The diurnal variation of the total advective tendency of temperature generally resembles that of the vertical velocity field with advective cooling corresponding to upward motion and advective warming corresponding to downward motion (not shown). This is consistent with the fact that the horizontal advection of temperature was very weak during these two periods.

Similarly, the diurnal variation of the total advective tendency of moisture for the monsoon systems (Fig. 16a) also resembles that of the vertical velocity field. Strong advective moistening is seen in the mid- and upper troposphere at night and weak drying is found below 800 hPa during the day. The atmospheric moistening shown in the monsoon systems is primarily due to the vertical advection of moisture because the horizontal terms (Fig. 17a) generally show advective drying in the mid- and upper troposphere, while the daytime lower-level drying is mainly associated with dry horizontal advection of moisture. In contrast, both the mainland and island-initiated systems show a weak moistening layer near the surface and a moderate drying layer above during the day in both the total and horizontal advection fields (Figs. 16b,c and 17b,c). Compared to the mainland systems, the island systems show considerably stronger total and horizontal advective moistening below 865 hPa over the composite period and weaker drying in the low and midtroposphere during the day. It is interesting to note

that there exists a minimum in total advective moistening near 600 hPa for the monsoon system (Fig. 16a). This is caused by the persistent maximum relative humidity observed at that level (not shown). The moistening minimum results in a distinct double-peak structure in the total moisture advection profile for the monsoon systems when they reach the maximum intensity between midnight and 0300 LST.

e. Heating and drying profiles

The composite profiles of the diurnal cycle of diabatic heating for the three periods are shown in Figs. 18a–c, respectively. The Q1 profiles are largely influenced by the vertical velocity. For the monsoon systems (Fig. 18a), the heating is weak during the day and reaches its maximum at 0300 LST when the strongest precipitation occurs. The maximum heating is seen in the upper troposphere between 365 and 465 hPa. In contrast, the upper-level maximum heating occurs in the afternoon around 1500 LST for the mainland systems (Fig. 18b), associated with the development of intensive afternoon deep convection. In the morning (0900–1200 LST), the heating is mainly located near the surface because of the occurrence of shallow convection. The heating and cooling seen at night are related to the passage of squall lines. Compared to the mainland systems, the island systems (Fig. 18c) also exhibit a diabatic heating layer near the surface in the morning associated with the development of shallow convection. The heating rapidly extends to the mid- and upper troposphere as convection develops and reaches its maximum in the afternoon (around 1500 LST). Diabatic cooling dominates the entire troposphere at night for the island systems. This is different from the mainland systems.

FIG. 14. Diurnal cycle of clouds. (a) Monsoon systems and (b) mainland systems. The unit is %. In the figures, values larger than 50% are shaded. Contour interval is 10 in (a) and 5 in (b).

The composite profiles of diurnal cycle of diabatic drying for the three types of systems are shown in Fig. 19. The variations of the vertical profiles of Q2 resemble those of the total advective tendency of moisture. All the systems show considerable moistening (particularly for the mainland and island systems) in the lower troposphere during the day with the peak seen around local noon. This low-level moistening is presumably due to the evaporation of shallow cumulus clouds or boundary layer processes. In the mid- and upper troposphere, the monsoon systems display a prominent maximum drying at midnight because of nocturnal deep convection, while both the mainland and island systems exhibit a pronounced maximum drying in the afternoon.

7. Summary

We have documented the large-scale structure and diabatic heating and drying profiles of the tropical convective systems observed during TWP-ICE, which was conducted in January and February 2006 in Darwin during the Australian summer monsoon season. Our analysis has focused on the variations of the structures with dif-

FIG. 15. Diurnal cycle of vertical velocity (hPa hr^{-1}). (a) Monsoon systems, (b) mainland systems, and (c) island systems. In (a)–(c), negative values are shaded. Contour interval is 5 in (a) and 2 in (b) and (c).

ferent synoptic regimes experienced in Darwin during the experiment. The monsoon behavior during the experiment was found to be modulated by the passage of an MJO event. A constrained objective variational analysis approach was used to derive the large-scale state and Q1 and Q2 profiles from original sounding measurements.

It is shown that the Darwin region had a very similar environment to that observed in other tropical regions

FIG. 16. Diurnal cycle of total advective tendency of moisture ($\text{g kg}^{-1} \text{ day}^{-1}$). (a) Monsoon systems, (b) mainland systems, and (c) island systems. In (a)–(c), negative values are shaded. Contour interval is 2 in these figures.

during the active and suppressed periods associated with the monsoon and MJO circulations. The active monsoon period was characterized by strong upward motion and large advective cooling and moistening through the entire troposphere, while the suppressed and clear periods exhibited moderate midlevel subsidence and significant

FIG. 17. Diurnal cycle of horizontal advective tendency of moisture ($\text{g kg}^{-1} \text{ day}^{-1}$). (a) Monsoon systems, (b) mainland systems, and (c) island systems. In (a)–(c), negative values are shaded. Contour interval is 1 in these figures.

low- to midlevel drying through horizontal advection. The midlevel subsidence and horizontal dry advection contributed to the considerably dryer midtroposphere observed during this period and are likely factors that limit the growth of clouds. During the break period, mean upward motion was accompanied by advective

FIG. 18. Diurnal cycle of diabatic heating rates (K day^{-1}). (a) Monsoon systems, (b) mainland systems, and (c) island systems. In (a)–(c), negative values are shaded. Contour interval is 5 in (a) and 3 in (b) and (c).

cooling and moistening located at midlevels with weak advective warming and drying at low levels. These different large-scale structures are likely important in the interaction of the large-scale environment with tropical convection.

The Q1 and Q2 profiles demonstrated considerable variations with different regimes. This appears to result

FIG. 19. Diurnal cycle of diabatic drying rates (K day^{-1}). (a) Monsoon systems, (b) mainland systems, and (c) island systems. In (a)–(c), negative values are shaded. Contour interval is 5 in (a) and 3 in (b) and (c).

from differences in the large-scale structures, cloud types, and rainfall rates between the regimes. This is consistent with Schumacher et al. (2007), who showed that large-scale flow variations, surface precipitation rates, and cloud types all played an important role in determining the Q1 profiles based on the data from the KWAJEX, TRMM-LBA, and SCSMEX field campaigns. This study shows that the strongest diabatic heating and drying

occurred during the active monsoon period when the rainfall rates were the largest. In this period, the heating was observed throughout the troposphere with maxima between 350 and 550 hPa, while the drying exhibited a distinct double-peak structure that is presumably associated with cumulus updrafts at low levels and meso-scale updrafts at high levels. These features are similar to those diagnosed from other tropical field campaigns. In contrast, the heating and drying were moderate and only appeared above 700 hPa during the break period. For the suppressed periods, the heating and drying tended to have their maxima at low levels, corresponding to the prevalence of shallow cumulus clouds during these periods.

To further investigate the observed large-scale structure and diabatic heating and drying profiles, a composite analysis has been performed to study the diurnal cycle of the monsoon systems during the active period and the mainland and island convective systems during the break period by using the unique TWP-ICE 3-h sounding data. To more appropriately describe the mainland and island systems, we have performed additional analyses over two subdomains covering the mainland and Tiwi Islands regions.

These systems exhibited significant differences in the large-scale structures and Q1 and Q2 profiles through their diurnal evolutions. Convection during the monsoon period was weak during the day, increased rapidly in the evening, and reached its maximum intensity in the early morning hours. Convection over the mainland and Tiwi Islands domains was initiated primarily by the sea-breeze circulation and confined in the lower troposphere in the midmorning and then quickly developed into deep convection and reached its maximum intensity in the afternoon. The transition from shallow to deep convection is clearly shown in both cloud and Q1 profiles of the mainland and island systems. This is supported by the differences in the large-scale structures between the monsoon systems and the mainland and island systems. The monsoon systems featured upward motion throughout the day with the maximum ascending level in the upper troposphere at 0300 LST because of the nocturnal deep convection. In contrast, the mainland and island systems showed an afternoon maximum in vertical velocity in the mid-upper troposphere. At their initial system development stages (around 0900–1000 LST), the mainland and island systems developed in an environment with an ascending layer near the surface because of the sea breeze circulation and a subsidence layer above, especially for the island systems. The subsidence layer contributes to limiting the mainland and island convective systems to low levels at their initial stages and to focusing convection along sea-breeze convergence lines later in the day.

Although the mainland and island convective systems were both driven by sea breezes, considerable differences were found in their large-scale structures and diabatic heating and drying profiles. We have shown that the island systems had a deeper layer of upward motion and stronger advective moistening near the surface than the mainland systems during the day, which resulted in a deeper low-level diabatic heating and stronger drying in the island systems than the mainland systems. It is also shown that the diurnal cycle of the mainland system was largely influenced by squall lines that passed over the experimental domain at night, which led to the secondary peak in its surface precipitation, large-scale structures, and diabatic heating and drying profiles.

It should be noted that the results shown in this study are based on a relatively small data sample and should be viewed with a degree of caution. Nevertheless, the large-scale structures and diabatic heating and drying profiles revealed in the examined convective systems generally agree with those diagnosed from other tropical field campaigns. Furthermore, the high-frequency (3 hourly) sounding data from TWP-ICE enabled us to perform a diurnal analysis of the convective systems under different surface boundary conditions. This is especially valuable for the mainland and island systems during the break monsoon period since these systems often occur on a daily basis and have much shorter lifetimes than the monsoon systems. By using the intensive sounding measurements combined with other unique observations from the enhanced surface observational network and satellite data, this study has provided a comprehensive description of the large-scale environment and diabatic heating and drying profiles of the tropical convective systems observed during TWP-ICE. It therefore provides important information for future studies on TWP-ICE.

Acknowledgments. The authors thank the two anonymous referees for their valuable comments that helped to clarify and improve the paper. We gratefully thank P. May, J. Mather, C. Long, P. Minnis, K. Johnson, M. Whimpey, J. Beringer, and N. Tapper for making the TWP-ICE data available for our use. We thank P. May and C. Schumacher for helpful discussion in defining the subdomains within the TWP-ICE sounding network. Thanks also go to A. Fridlind for her help in obtaining the domain average C-POL radar retrieved rainfall data. The observational data were obtained from the ARM program sponsored by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Environmental Sciences Division. Work at LLNL was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National

Laboratory under Contract DE-AC52-07NA27344. Work at CAWCR and Monash University was supported by the DOE ARM program under Grant DE-FG02-03ER63533 and DE-FG02-08ER64527. Work at SUNY Stony Brook was supported by the DOE ARM Program under Grant EFG0298ER62570. The data can be obtained from the ARM archive (<http://www.db.arm.gov/cgi-bin/IOP2/selectAftIOP.pl?iopName=twp2006twp-ice>).

TABLE A1. Empirical RS92 calibration factors $[C_{\text{cal}}(T)]$ and their uncertainties derived from nighttime soundings. Reproduced from Vömel et al. (2007).

Temperature (T , Celsius)	$C_{\text{cal}}(T)$	Uncertainty
0	0.98	± 0.02
-30	0.98	± 0.06
-50	0.94	± 0.03
-60	1.04	± 0.06
-70	1.13	± 0.06

APPENDIX

Correction of the Vaisala RS92 Sondes

a. Vömel et al. (2007) method

The Vaisala RS92 sondes are known to have a significant daytime solar radiation dry bias and a minor temperature-dependent calibration error. Vömel et al. (2007) developed an empirical method to correct the bias based on results from a comparison of RS92 humidity observations with reference measurements made by a cryogenic frostpoint hygrometer (CFH). As described in Vömel et al. (2007), the RS92 measured relative humidity can be corrected by using a pressure- (P) dependent radiation error correction factor $C_{\text{rad}}(P)$ and a temperature- (T) dependent calibration factor $C_{\text{cal}}(T)$. The correction factor can be written as

$$C_{\text{rh}} = \frac{\text{RH}_{\text{corr}}}{\text{RH}_{\text{TL}}} = \frac{1}{[C_{\text{rad}}(P)][C_{\text{cal}}(T)]}, \quad (\text{A1})$$

where RH_{corr} is the corrected relative humidity, RH_{TL} is the time lag corrected humidity profile described by Miloshevich et al. (2004), $C_{\text{cal}}(T)$ is given in Table A1, and $C_{\text{rad}}(P)$ can be expressed as

$$C_{\text{rad}}(P) = -0.12158 \ln(P)^2 + 1.664 \ln(P) - 4.7855. \quad (\text{A2})$$

b. Revised Vömel et al. (2007) method for TWP-ICE

The Vömel et al. (2007) method was developed for correcting RS92 soundings launched near solar noon and at night because there were no early morning or late afternoon soundings available over the period they studied. Using the 3-hourly high-frequency TWP-ICE soundings and the ARM microwave radiometer (MWR) measured precipitable water vapor (PWV) from the ship, we found that the dry bias is also dependent on the solar zenith angle with greater bias found for radiosondes launched near solar noon than those made in the

early morning or late afternoon. Therefore, the use of the near-noon relative humidity correction factors for early morning or late afternoon soundings is going to result in an overcorrection of the relative humidity. To take into account of this, a new correction factor C_{zen} is introduced into Eq. (A1):

$$C_{\text{rh}} = \frac{\text{RH}_{\text{corr}}}{\text{RH}_{\text{TL}}} = \frac{C_{\text{zen}}}{[C_{\text{rad}}(P)][C_{\text{cal}}(T)]}, \quad (\text{A3})$$

where C_{zen} is a function of the zenith angle, z :

$$C_{\text{zen}} = 1 + [\min(z, 90)] \left(\frac{C_{\text{rad}} - 1}{90} \right). \quad (\text{A4})$$

The zenith angle correction factor ensures that C_{rh} decreases linearly, from a maximum at local noon to a minimum at nighttime. At noon, RH_{corr} is the same as the daytime correction described in Vömel et al. (2007).

Table A2 shows the biases for soundings made from the ship with various corrections applied. The first column shows the biases for uncorrected soundings, the second column is for soundings corrected using the method described in Vömel et al. (2007), the third column is for soundings corrected using Eqs. (A3) and (A4), and the last column shows the number of soundings that were used in the validation. Results are shown for near-noon soundings ($z < 30^\circ$, first row), all daytime soundings (second row), and all nighttime soundings (last row).

At the ship, the uncorrected soundings show a significant daytime dry bias. When the temperature calibration and pressure-dependent radiation correction are applied, the dry bias is overcorrected, resulting in quite a large moist bias. As discussed above, this is expected because the bias correction method developed by Vömel et al. (2007) was based on soundings made near noon, whereas many of the soundings in this study were for the early morning or late afternoon, when the radiation dry bias is less. Finally, when all the corrections are applied, the PWV bias is significantly reduced, even for the near-noon soundings.

TABLE A2. Bias of PWV derived from radiosonde soundings made from the ship.

Sounding	Uncorrected	C_{rad} and C_{cal}	All corrections	No. of soundings
$Z \leq 30$	−4.4	2.1	0.8	22
Day	−2.5	3.9	1.3	51
Night	−0.7	0.3	0.3	38

Of course, the RS92 dry bias will vary from one launch to the next, as a result of local factors such as cloud cover, which are not accounted for by the correction algorithm. Nevertheless, the correction method presented in this paper provides a simple method for significantly reducing the RS92 relative humidity errors.

REFERENCES

- Ackerman, T. P., and G. M. Stokes, 2003: The Atmospheric Radiation Measurement Program. *Phys. Today*, **56**, 38–44.
- Beringer, J., L. B. Hutley, N. J. Tapper, and L. Cernusak, 2007: Savanna fires and their impact on net ecosystem productivity in North Australia. *Global Change Biol.*, **13**, 990–1004, doi:10.1111/j.1365-2486.2007.01334.x.
- Bringi, V. N., G. J. Huang, V. Chandrasekar, and T. D. Keenan, 2001: An areal rainfall estimator using differential propagation phase: Evaluation using a C-band radar and a dense gauge network in the tropics. *J. Atmos. Oceanic Technol.*, **18**, 1810–1818.
- , —, —, and E. Gorgucci, 2002: A methodology for estimating the parameters of a gamma raindrop size distribution model from polarimetric radar data: Application to a squall-line event from the TRMM/Brazil campaign. *J. Atmos. Oceanic Technol.*, **19**, 633–645.
- , T. Tang, and V. Chandrasekar, 2004: Evaluation of a new polarimetrically based Z – R relation. *J. Atmos. Oceanic Technol.*, **21**, 612–622.
- Carbone, R. E., J. Wilson, T. D. Keenan, and J. M. Hacker, 2000: Tropical island convection in the absence of significant topography. Part I: Life cycle of diurnally forced convection. *Mon. Wea. Rev.*, **128**, 3459–3480.
- Clothiaux, E. E., T. P. Ackerman, G. G. Mace, K. P. Moran, R. T. Marchand, M. Miller, and B. E. Martner, 2000: Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites. *J. Appl. Meteor.*, **39**, 645–665.
- Cressman, G. P., 1959: An operational objective analysis scheme. *Mon. Wea. Rev.*, **87**, 367–374.
- Frank, W. M., 1978: The life cycle of GATE convective systems. *J. Atmos. Sci.*, **35**, 1256–1264.
- , and J. L. McBride, 1989: The vertical distribution of heating in AMEX and GATE cloud clusters. *J. Atmos. Sci.*, **46**, 3464–3478.
- Gunn, B. W., J. L. McBride, G. J. Holland, T. D. Keenan, N. E. Davidson, and H. H. Hendon, 1989: The Australian summer monsoon circulation during AMEX Phase II. *Mon. Wea. Rev.*, **117**, 2554–2574.
- Hume, T., 2007: Radiation dry bias in the TWP-ICE radiosonde soundings. *Proc. 17th ARM Science Team Meeting*, Monterey, CA, U.S. Department of Energy, P00011. [Available online at <http://www.arm.gov/publications/proceedings/conf17/poster/P00011.pdf>.]
- Johnson, R. H., 1984: Partitioning tropical heat and moisture budgets into cumulus and mesoscale components: Implications for cumulus parameterization. *Mon. Wea. Rev.*, **112**, 1590–1601.
- Keenan, T. D., and R. E. Carbone, 1992: A preliminary morphology of precipitation systems in tropical northern Australia. *Quart. J. Roy. Meteor. Soc.*, **118**, 283–326.
- , and Coauthors, 2000: The Maritime Continent Thunderstorm Experiment (MCTEX): Overview and some results. *Bull. Amer. Meteor. Soc.*, **81**, 2433–2455.
- Lin, J., B. Mapes, M. Zhang, and M. Newman, 2004: Stratiform precipitation, vertical heating profiles, and Madden–Julian oscillation. *J. Atmos. Sci.*, **61**, 296–309.
- Lin, X., and R. H. Johnson, 1996a: Kinematic and thermodynamic characteristics of the flow over the western Pacific warm pool during TOGA COARE. *J. Atmos. Sci.*, **53**, 695–715.
- , and —, 1996b: Heating, moistening, and rainfall over the western Pacific warm pool during TOGA COARE. *J. Atmos. Sci.*, **53**, 3367–3383.
- Long, C. N., 2001: The shortwave (SW) clear-sky detection and fitting algorithm: Algorithm operational details and explanations. Atmospheric Radiation Measurement Program Tech. Rep. ARM TR-004, 26 pp. [Available online at http://www.arm.gov/publications/tech_reports/arm-tr-004-1.pdf.]
- , and T. P. Ackerman, 2000: Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects. *J. Geophys. Res.*, **105D**, 15 609–15 626.
- Manton, M. J., and J. L. McBride, 1992: Recent research on the Australian monsoon. *J. Meteor. Soc. Japan*, **70**, 275–285.
- May, P. T., J. H. Mather, G. Vaughan, C. Jakob, G. M. McFarquhar, K. N. Brown, and G. G. Mace, 2008: The Tropical Warm Pool International Cloud Experiment. *Bull. Amer. Meteor. Soc.*, **89**, 629–645.
- McBride, J. L., 1987: The Australian summer monsoon. *Monsoon Meteorology*, C. P. Chang and T. N. Krishnamurti, Eds., Oxford University Press, 203–231.
- Miloshevich, L. M., A. Paukkunen, H. Vömel, and S. J. Oltmans, 2004: Development and validation of a time-lag correction for Vaisala radiosonde humidity measurements. *J. Atmos. Oceanic Technol.*, **21**, 1305–1327.
- Minnis, P., and Coauthors, 2001: A near-real-time method for deriving cloud and radiation properties from satellites for weather and climate studies. *Proc. 11th Conf. on Satellite Meteorology and Oceanography*, Madison, WI, Amer. Meteor. Soc., 477–480.
- Nesbitt, S. W., and E. J. Zipser, 2003: The diurnal cycle of rainfall and convective intensity according to three years of TRMM measurements. *J. Climate*, **16**, 1456–1475.
- Numaguti, A., R. Oki, K. Nakamura, K. Tsuboki, N. Misawa, T. Asai, and Y.-M. Kodama, 1995: 4–5-day-period variation and low-level dry air observed in the equatorial western Pacific during the TOGA-COARE IOP. *J. Meteor. Soc. Japan*, **73**, 267–290.
- Riehl, H., and J. S. Malkus, 1958: On the heat balance in the equatorial trough zone. *Geophysica*, **6**, 503–538.
- Schumacher, C., M. H. Zhang, and P. E. Ciesielski, 2007: Heating structures of the TRMM field campaigns. *J. Atmos. Sci.*, **64**, 2593–2610.
- Simpson, J., T. Keenan, B. R. H. Simpson, and G. J. Holland, 1993: Cumulus mergers in the maritime continent region. *Meteor. Atmos. Phys.*, **51**, 73–99.

- Stokes, G. M., and S. E. Schwartz, 1994: The Atmospheric Radiation Measurement (ARM) Program: Programmatic background and design of the cloud and radiation test bed. *Bull. Amer. Meteor. Soc.*, **75**, 1201–1221.
- Tao, W.-K., and Coauthors, 2006: Retrieval of latent heating from TRMM measurements. *Bull. Amer. Meteor. Soc.*, **87**, 1555–1572.
- , R. J. Houze, and E. A. Smith, 2007: The fourth TRMM latent heating workshop. *Bull. Amer. Meteor. Soc.*, **88**, 1255–1259.
- Thompson, R. M., S. W. Payne, E. E. Recker, and R. J. Reed, 1979: Structure and properties of synoptic-scale wave disturbances in the intertropical convergence zone of the eastern Atlantic. *J. Atmos. Sci.*, **36**, 53–72.
- Vömel, H., and Coauthors, 2007: Radiation dry bias of the Vaisala RS 92 humidity sensor. *J. Atmos. Oceanic Technol.*, **24**, 953–963.
- Xie, S., S. A. Klein, M. Zhang, J. J. Yio, R. T. Cederwall, and R. McCoy, 2006: Developing large-scale forcing data for single-column and cloud-resolving models from the Mixed-Phase Arctic Cloud Experiment. *J. Geophys. Res.*, **111**, D19104, doi:10.1029/2005JD006950.
- Yanai, M., 1961: A detailed analysis of typhoon formation. *J. Meteor. Soc. Japan*, **39**, 187–214.
- , and R. Johnson, 1993: Impacts of cumulus convection on thermodynamic fields. *The Representation of Cumulus Convection in Numerical Models*, Meteor. Monogr., No. 46, Amer. Meteor. Soc., 39–62.
- , S. Esbensen, and J. Chu, 1973: Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets. *J. Atmos. Sci.*, **30**, 611–627.
- , B. Chen, and W.-W. Tung, 2000: The Madden–Julian oscillation observed during the TOGA COARE IOP: Global view. *J. Atmos. Sci.*, **57**, 2374–2396.
- Zhang, M. H., and J. L. Lin, 1997: Constrained variational analysis of sounding data bases on column-integrated budgets of mass, heat, moisture, and momentum: Approach and application to ARM measurements. *J. Atmos. Sci.*, **54**, 1503–1524.
- , —, 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.