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

[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.

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 intercomparison 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	1000−10 hPa	Surface−10 hPa	1000−10 hPa
	23 pressure + surface	15 pressure + surface	29 sigma	15 pressure + surface
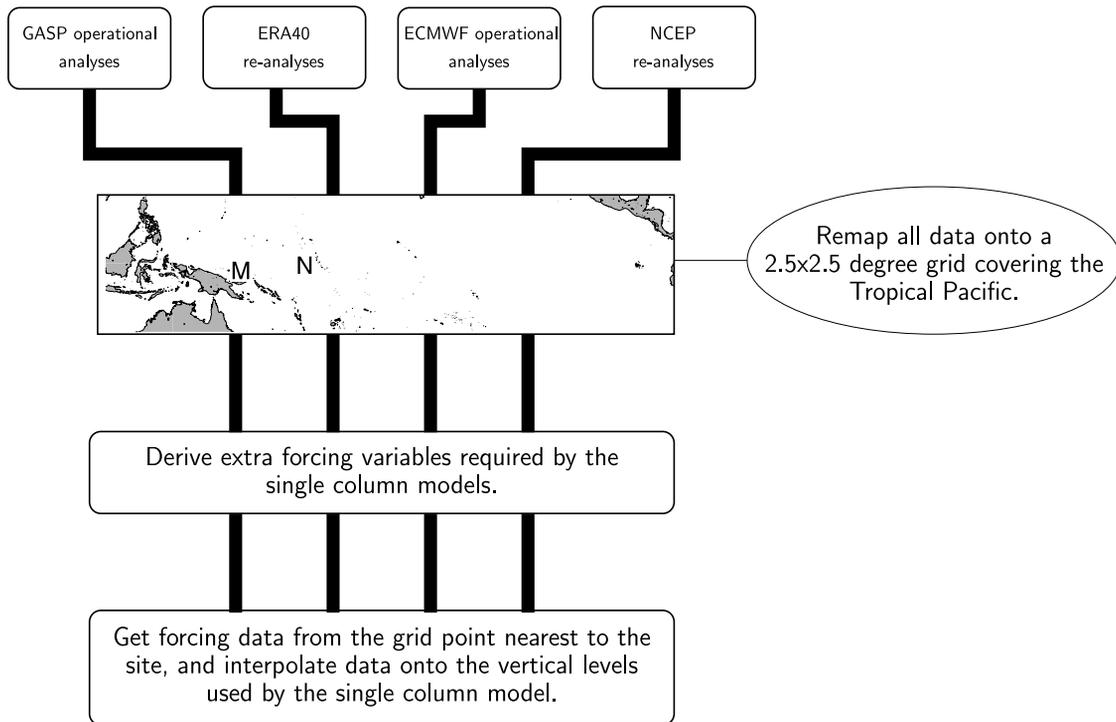
Figure 1. Derivation of SCM forcing data from NWP analyses. The locations of Manus Island (M: 2.058°S, 147.425°E) and Nauru (N: 0.521°S, 166.916°E) are also shown.

[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

$$A_T = -\mathbf{V} \cdot \nabla T \tag{1}$$

$$A_q = -\mathbf{V} \cdot \nabla q \tag{2}$$

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
MWR	Hourly	Observations contaminated by rainfall removed.
Surface Met.	Hourly	Observations of temperature, rain etc.
GMS-5	Hourly	Starting in June 1999.
		Quantities such as outgoing long wave radiation,
		cloud top temperature, cloud cover and optical
		thickness derived from GMS-5 IR and VIS channels.

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, $2-3$ m s^{-1} and $5-15\%$. 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-
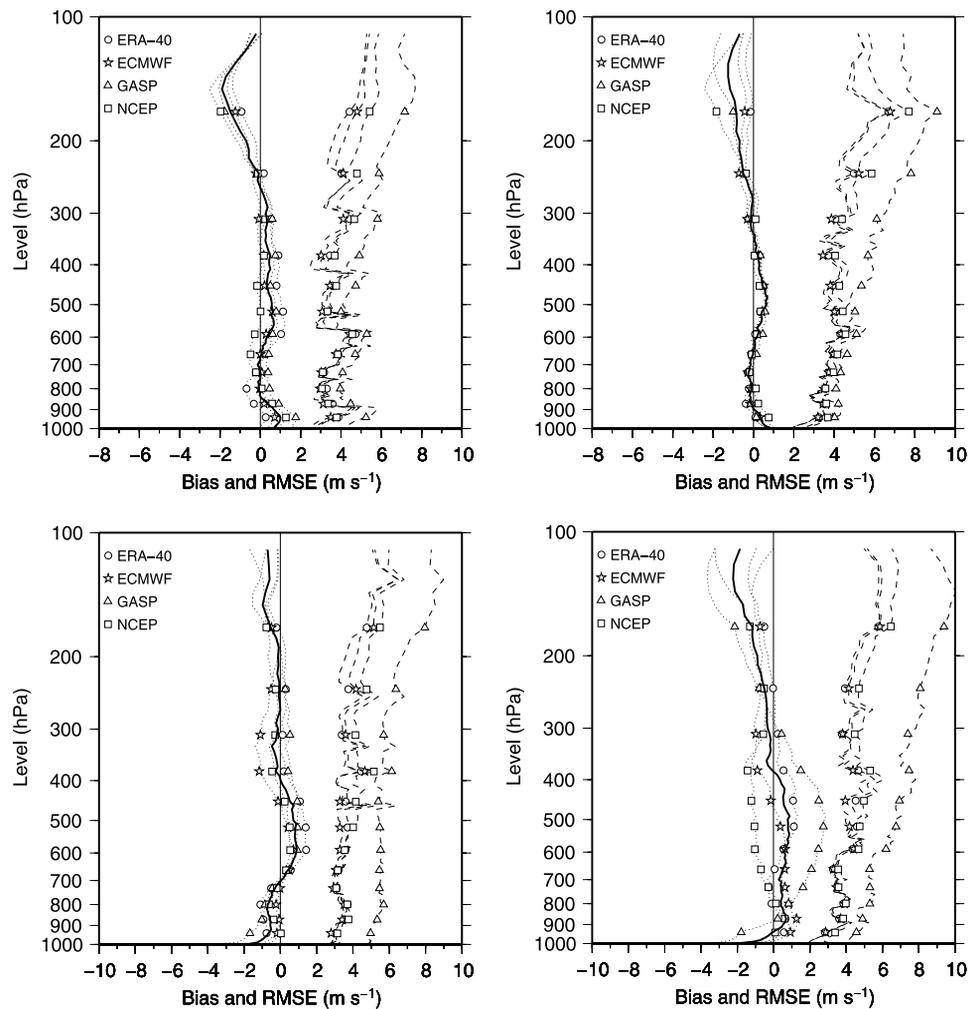
Figure 2. Bias (dotted lines) and RMSE (dashed lines) of U (top panels) and V (bottom panels) forcing data at Manus Island (left panels) and Nauru (right panels). The solid lines show the mean bias for all the models. Validations were made using all available radiosondes wind observations during 1999 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.
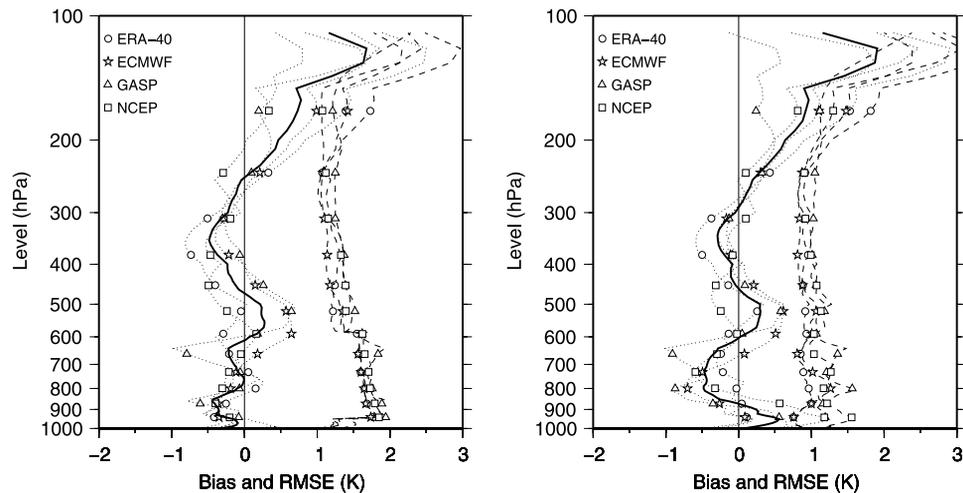
3.2.2. Temperature

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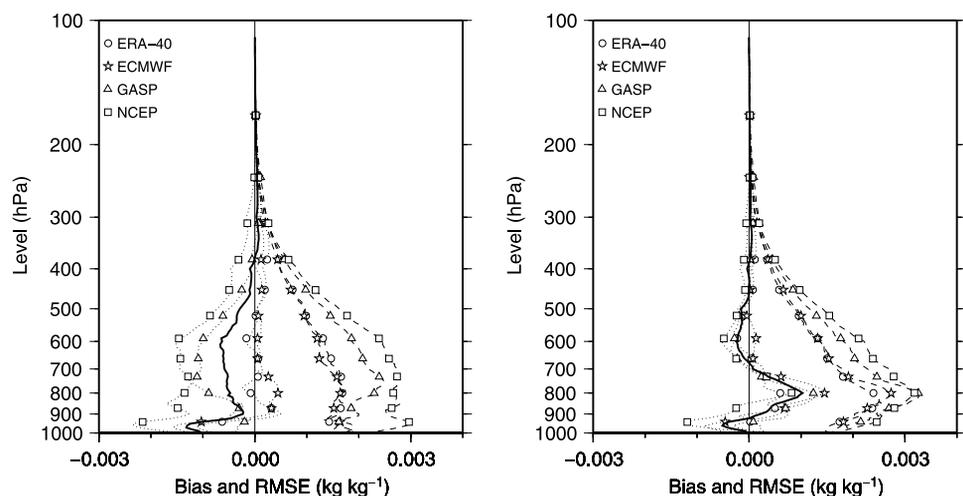
Figure 4. Bias (dotted lines) and RMSE (dashed lines) of specific humidity data at Manus Island (left) and Nauru (right). The solid line is the mean bias for all the models. Validations were made using all available corrected radiosonde data (as described in the text) during 1999 and 2000.

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**		

[a]Units 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.
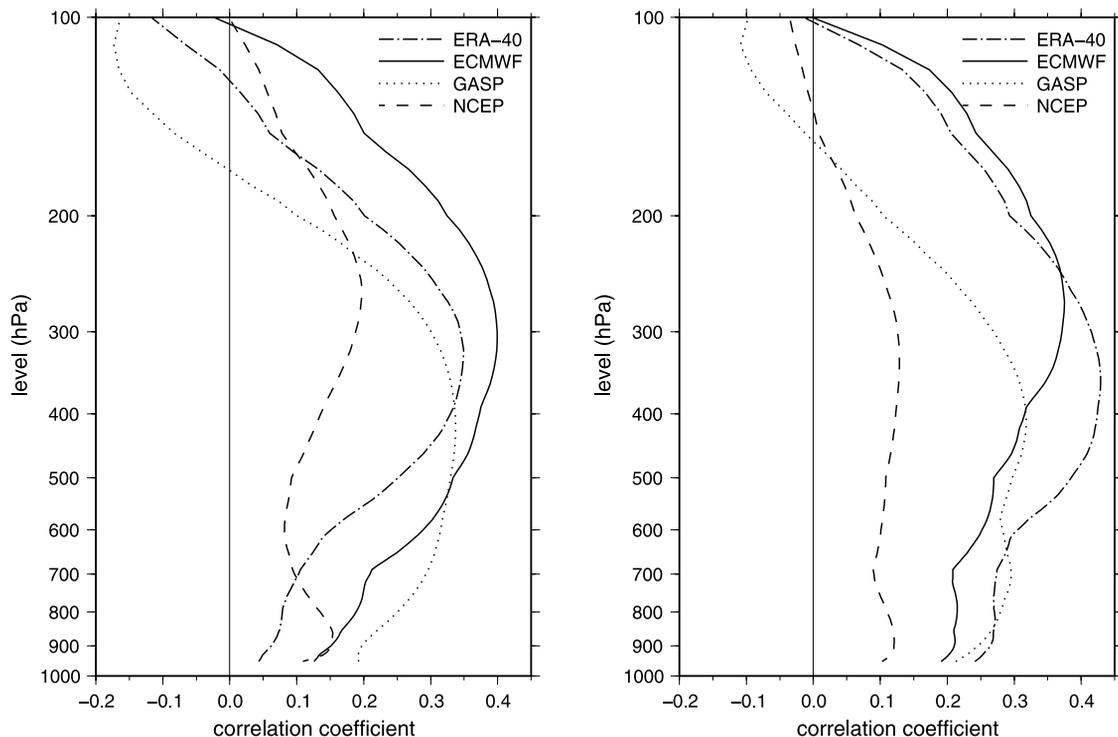
3.3.2. Advective Tendencies

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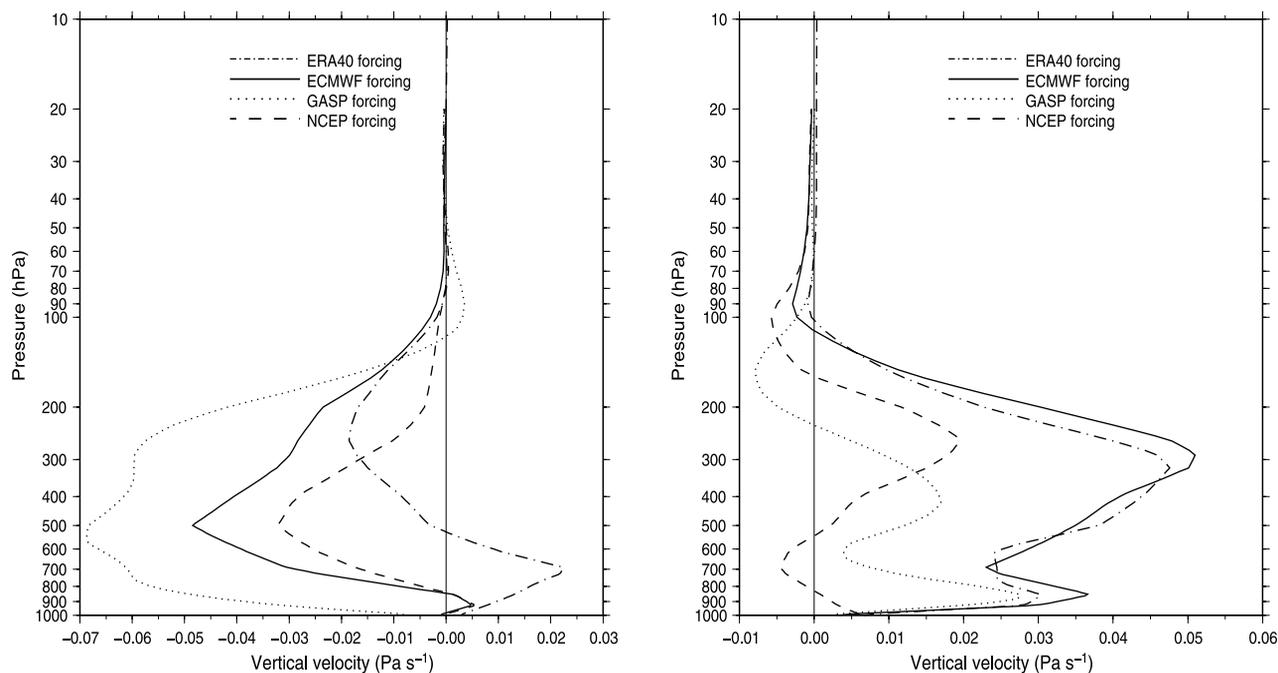
Figure 5. Average vertical velocity data at Manus Island (left) and Nauru (right).

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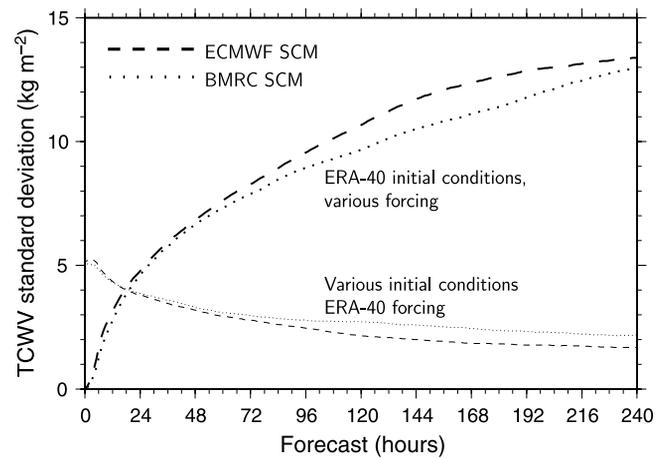
Figure 7. Standard deviation of SCM ensemble predictions of total column water vapor (TCWV) at Manus Island. The ensembles were constructed from model runs using four different initial conditions and ERA-40 forcing data (thin lines), and model runs using ERA-40 initial conditions and four different forcing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

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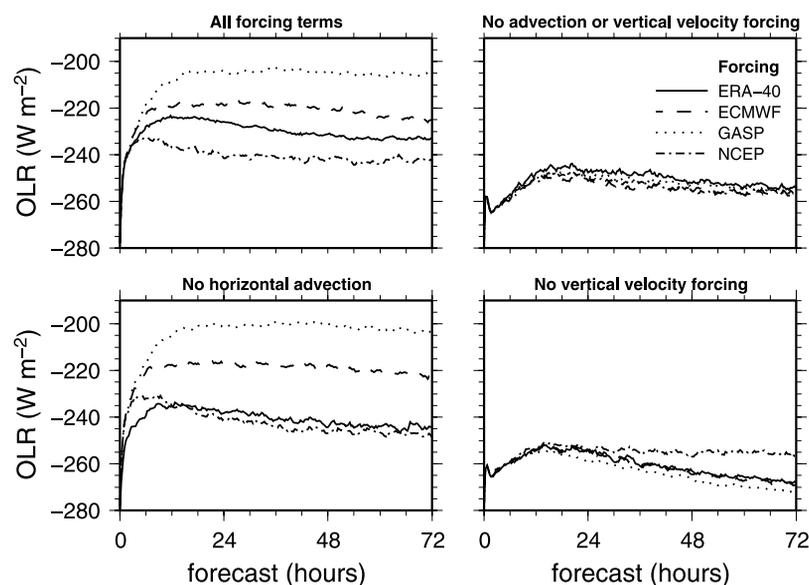
Figure 8. Average outgoing longwave radiation (OLR) at Manus Island during January and February 1999 for various configurations of the ECMWF SCM.

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%

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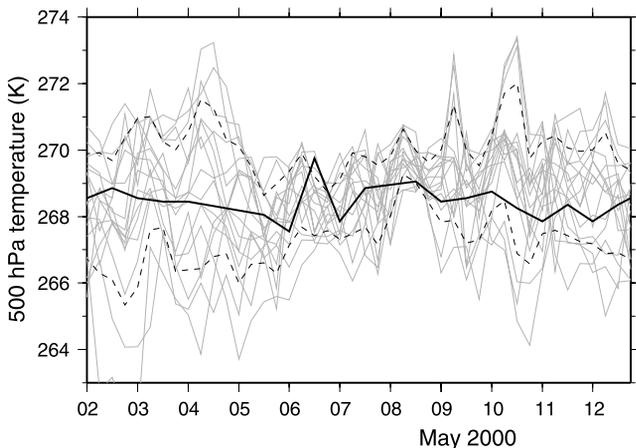
Figure 9. Time series of 12 hour forecasts of 500 hPa temperature at Manus Island from an ensemble of ECMWF SCM runs during May 2000 (thin grey lines). The dashed lines show the values one standard deviation either side of the ensemble mean, and the solid black line shows the observed temperature.

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
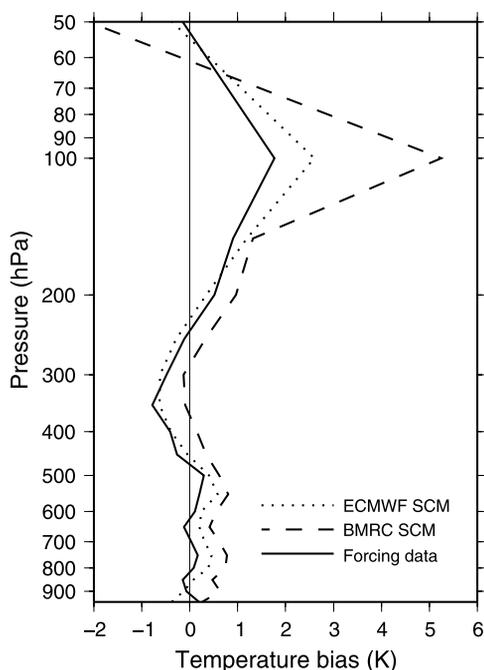
Figure 10. Bias for the 12 hour ensemble mean prediction of temperature at Manus Island. The dotted line shows the bias for the ECMWF SCM and the dashed line shows the bias for the BMRC SCM. The solid line indicates the mean bias of the forcing data sets for the times when the 12 hour ensemble predictions are valid.
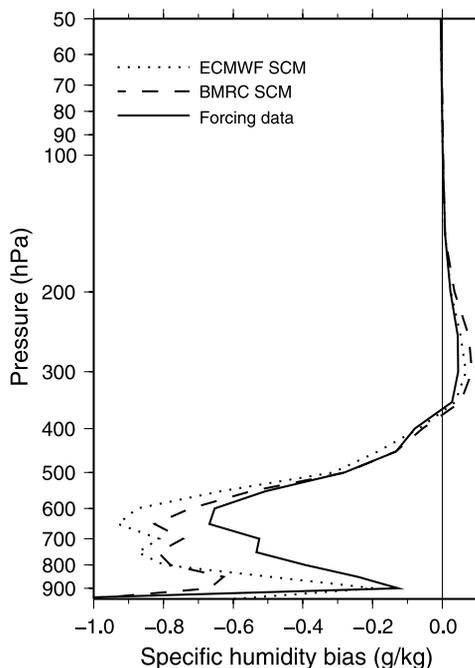
Figure 11. As for Figure 10, except for the specific humidity biases.

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

[a]The 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
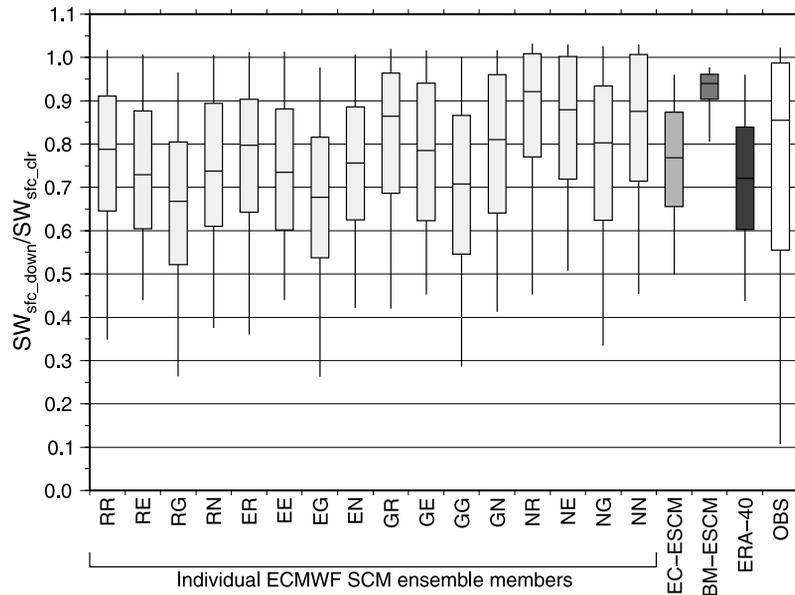
Figure 12. Ratio of predicted downward solar radiation at the surface to clear sky downward solar radiation at Manus Island. Data are for 14 hour forecasts valid at 02 UTC for the complete 1999–2000 period. The whiskers show the 5th and 95th percentiles, the boxes show the upper and lower quartiles, 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.

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 GCSS 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. Aiyyer (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. Embery, 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.

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)