

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 model 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

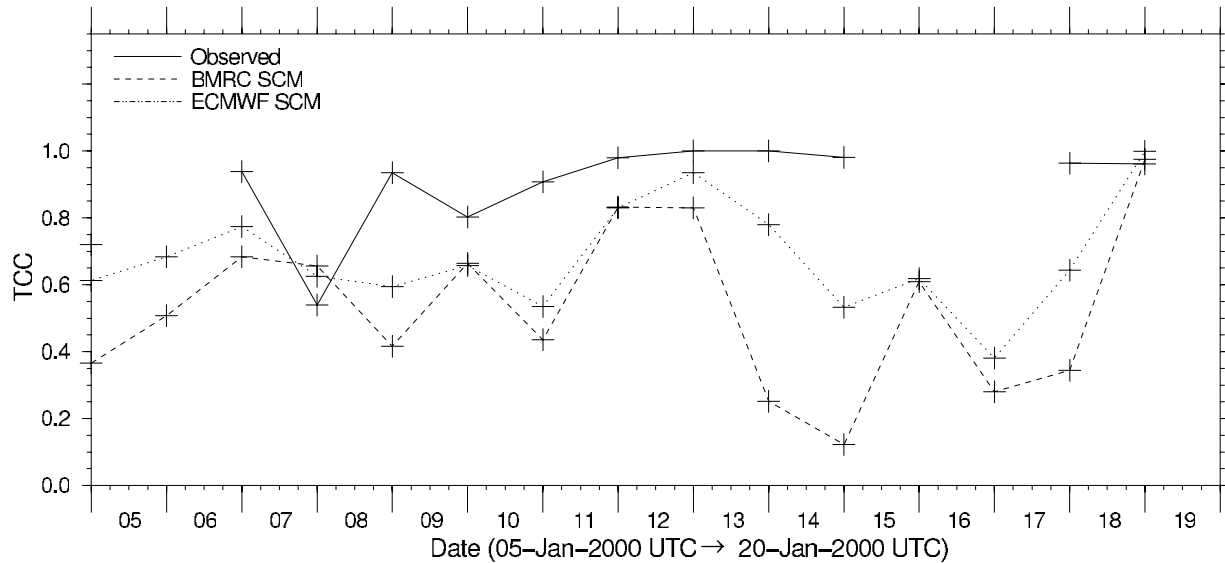


Figure 1. Time series of total cloud cover observations and predictions at Manus Island. The ECMWF and BMRC curves show the ensemble mean 12 hour predictions, valid at 0000 UTC. The observations are derived from GMS-5 data valid at 0000 UTC.

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:

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

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]:

$$RV = \frac{MSE_{climatology} - MSE}{MSE_{climatology}} \tag{2}$$

where MSE is the mean square error of the SCM forecasts being validated and $MSE_{climatology}$ 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,

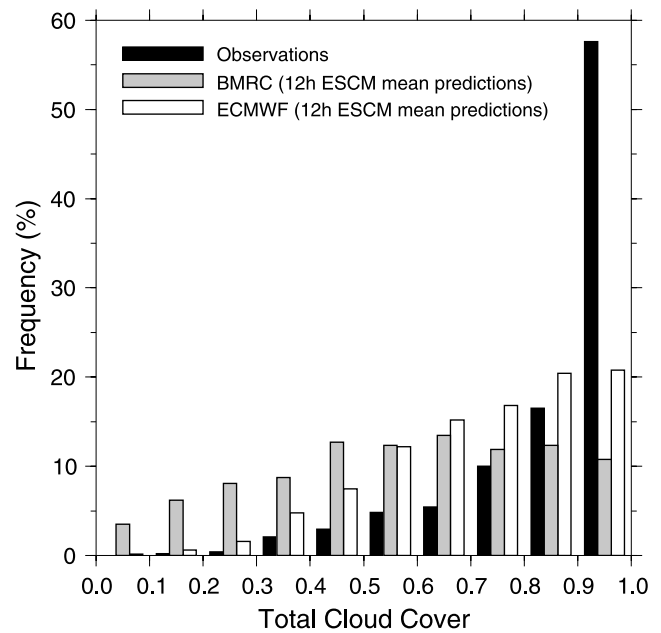


Figure 2. PDFs for the predicted and observed 0000 UTC TCC at Manus Island. The model results show the ensemble mean of the 12 hour forecasts. The results are displayed in 0.1 wide bins.

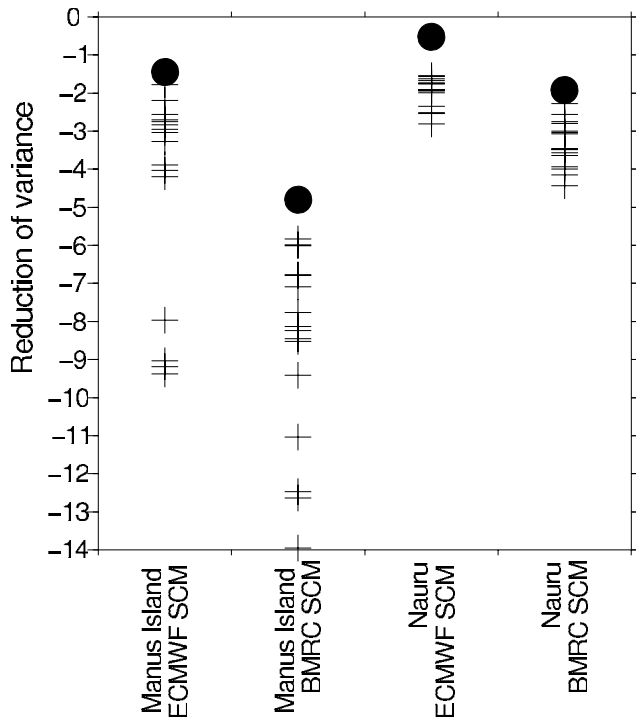


Figure 3. Reduction of variance (RV) for the ECMWF and BMRC SCM twelve hour forecasts of TCC, valid at 0000 UTC. The crosses show the RV for the individual ensemble members, and the circles show the RV for the ensemble mean.

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,

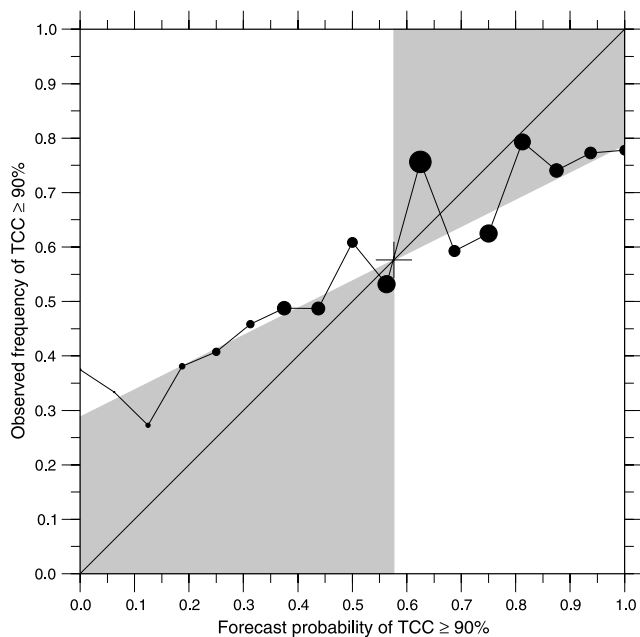


Figure 4. Reliability diagram for 12 hour forecasts of TCC greater than or equal to 0.9 from the ECMWF ESCM at Manus Island. The forecasts are valid at 0000 UTC. The size of the circles show the relative frequencies of forecasts with probabilities of $\frac{0}{16}, \frac{1}{16}, \dots, \frac{16}{16}$. 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:

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

$$BSS = \frac{BS_{climate} - BS}{BS_{climate}} \tag{4}$$

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-

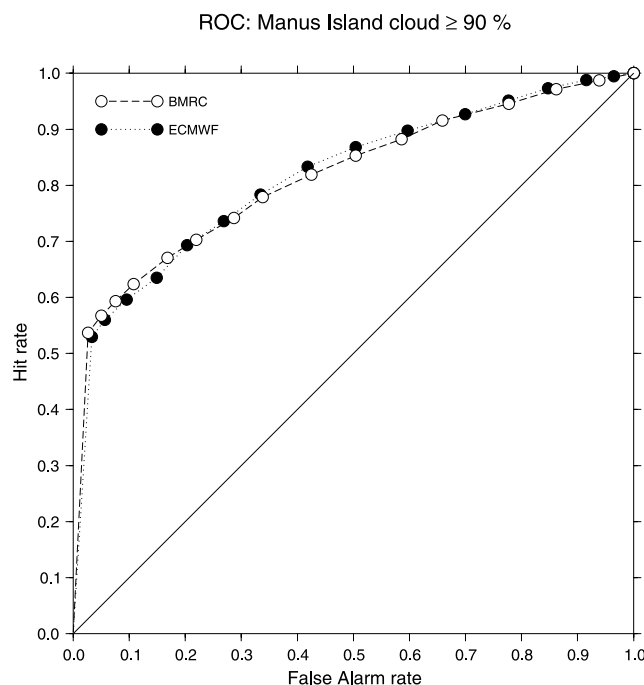


Figure 5. ROC diagram for probabilistic forecasts of TCC exceeding 0.9 for each of the SCM ensembles at Manus Island.

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

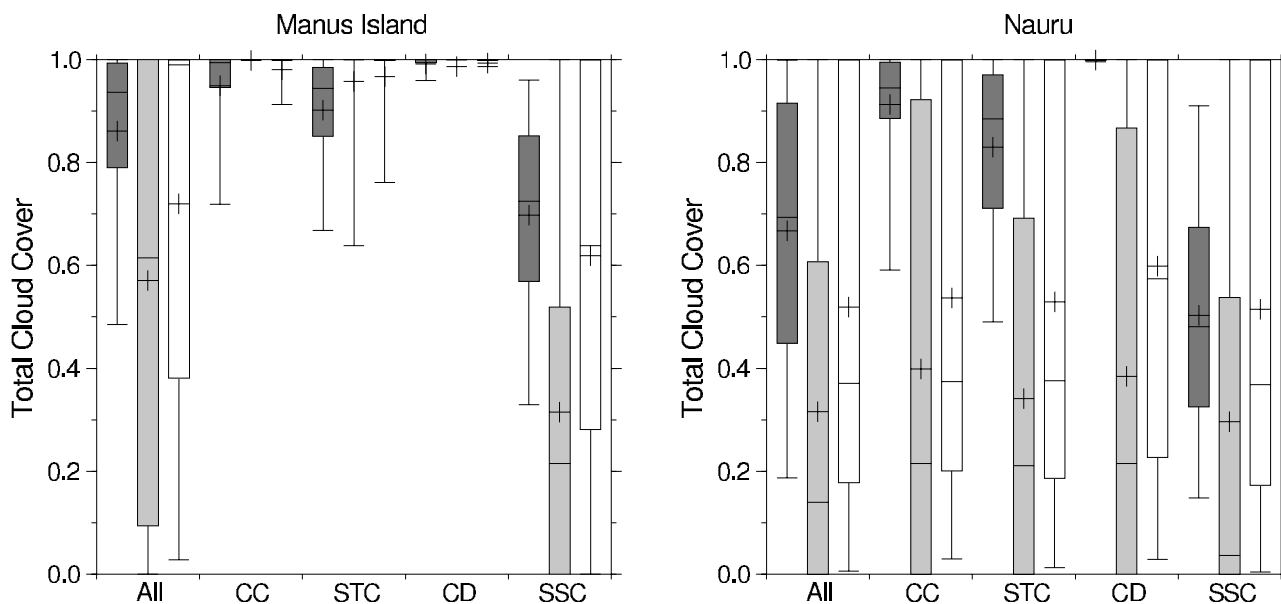


Figure 6. Box and whisker diagrams showing the TCC PDFs, for each of the observed and modelled regimes at Manus Island (left) and Nauru (right). The dark grey bars show the observed 0000 UTC TCC PDFs, while the light grey and white bars show the PDFs for the 12 hour ensemble mean forecasts of TCC valid 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.

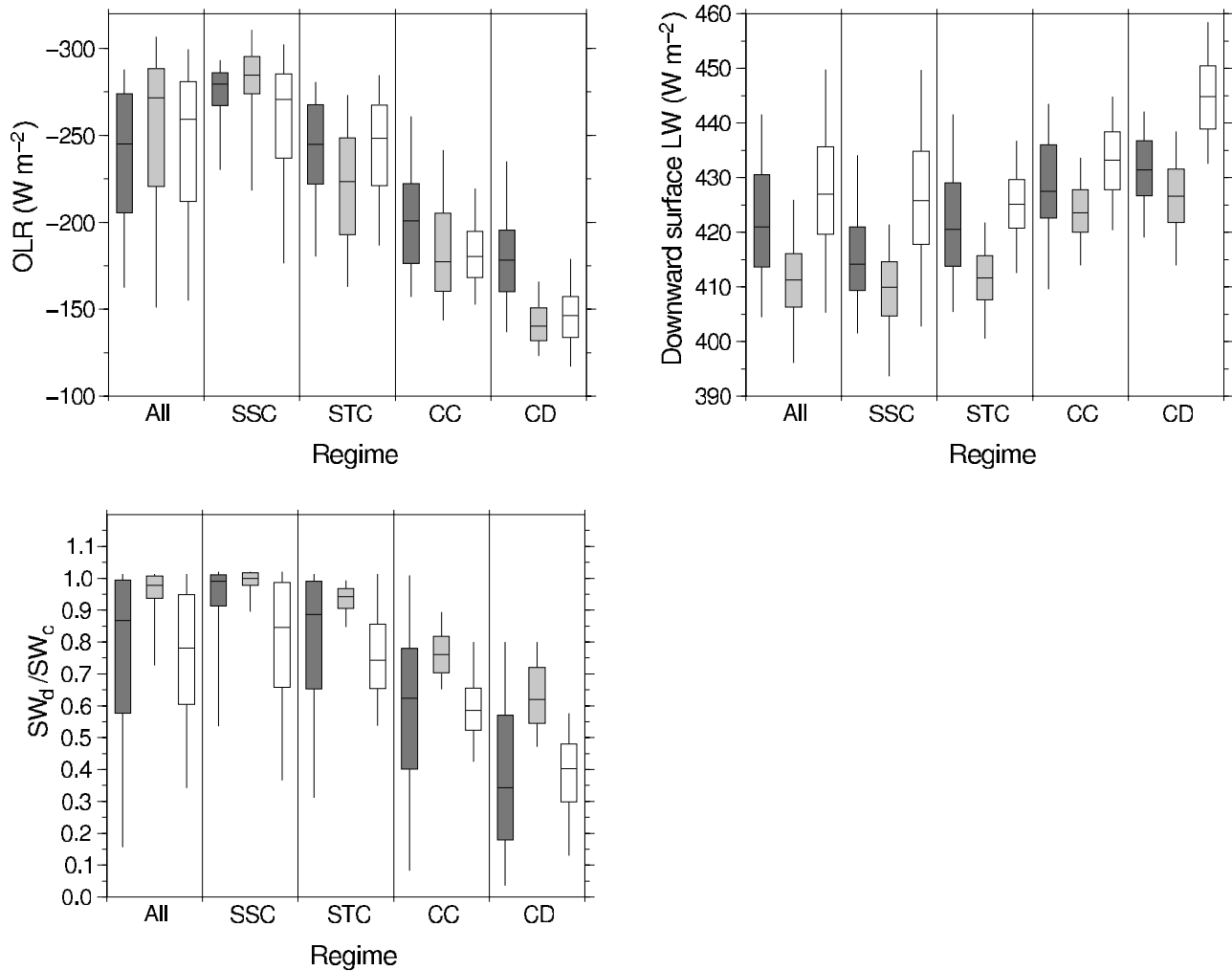


Figure 7. Radiative properties of the observed (dark grey bars) and modelled (light grey bars: BMRC SCM 12 hour ensemble mean predictions, white bars: ECMWF SCM 12 hour ensemble mean predictions) cloud regimes. The top left panel shows the OLR at the TOA, the top right panel shows the downward long-wave 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

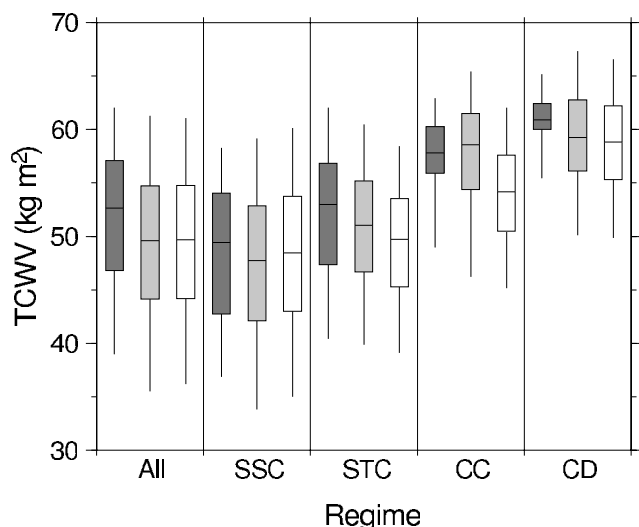


Figure 8. As for Figure 7, but for total column water vapor (TCWV).

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

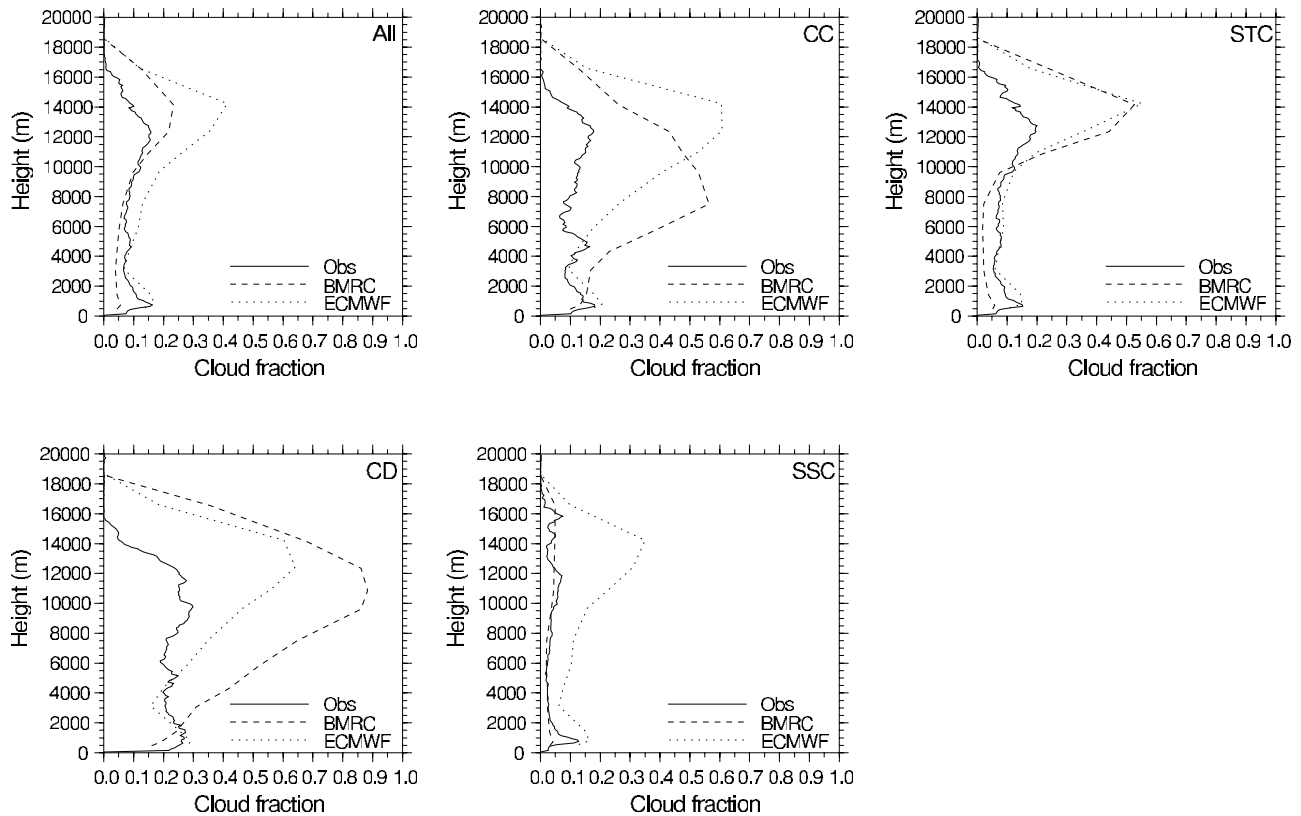


Figure 9. Vertical cloud profiles at Manus Island. Observations are from 0000 UTC ARSCL (Active Remotely Sensed Clouds Locations) data. The SCM profiles are averages of the 12 hour ensemble mean forecasts valid at 0000 UTC during 1999 and 2000 for each regime.

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

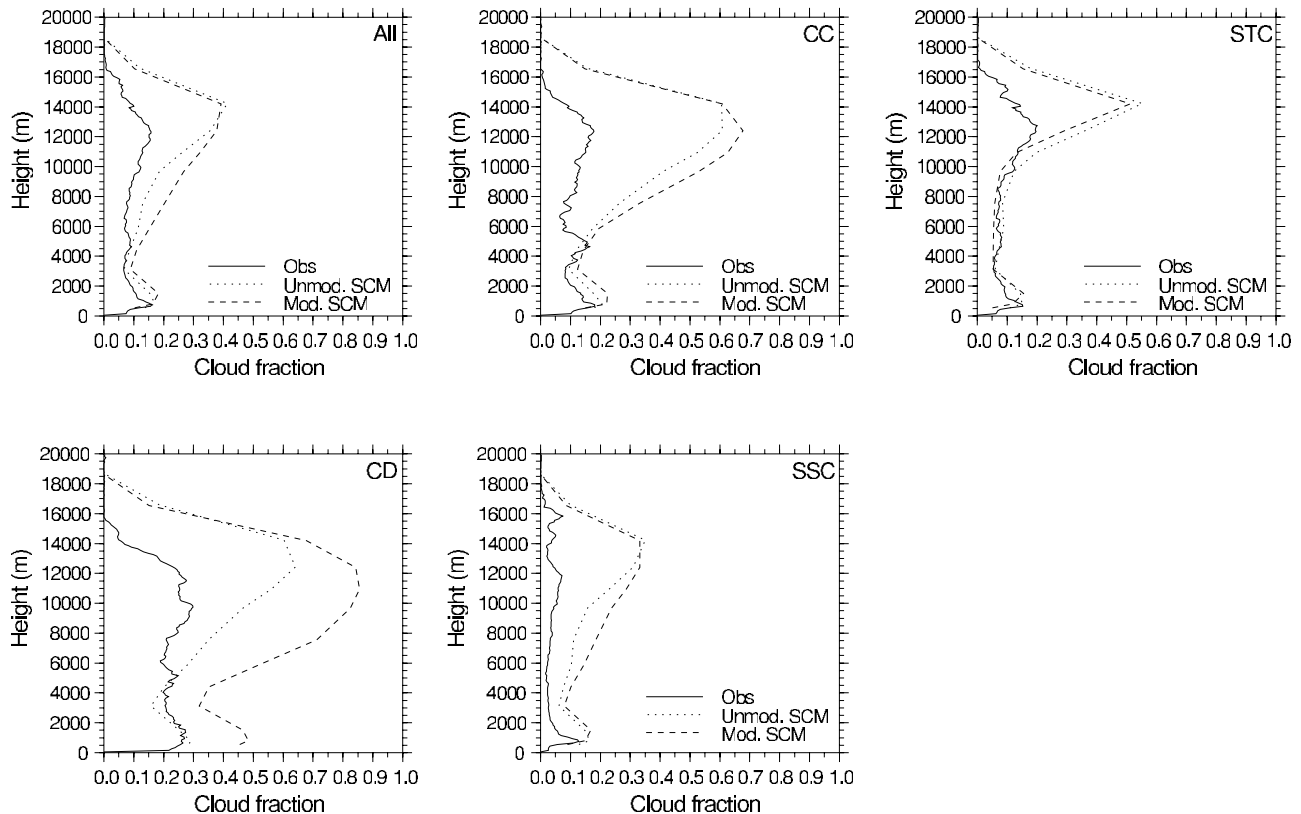


Figure 10. Comparison of vertical cloud profiles at Manus Island from the unmodified and modified ECMWF SCMs. Observed profiles are included for comparison. The SCM profiles are averages of the 12 hour ensemble mean forecasts valid at 0000 UTC during 1999 and 2000 for each cloud regime.

“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)
- Rosow, 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. 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.
- 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 CSRМ 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.