

AN IMPROVED STRATEGY FOR THE EVALUATION OF CLOUD PARAMETERIZATIONS IN GCMs

BY CHRISTIAN JAKOB

Cloud regime diagnostics build the bridge from climate model evaluation to case studies.

Great effort is continually being made to improve the parameterization of clouds in global circulation models (GCMs), and as a result the physical realism of cloud parameterizations has been improved considerably. Almost all GCMs now use a prognostic equation to represent the evolution of cloud condensate (e.g., Sundqvist 1978; del Genio et al. 1996; Fowler et al. 1996). The treatment of cloud cover is more diverse, ranging from simple diagnostic relations (e.g., Sundqvist 1978; Slingo 1987; Smith 1990) to a fully prognostic treatment (e.g., Tiedtke 1993). Through evaluation studies, model developers want to be able to expose flaws in these parameterizations, and, if possible, reveal reasons for those flaws. Evaluating the “model clouds” against their real-life counterparts, however, is becoming increasingly difficult because the models themselves are becoming increasingly complex. Current GCMs are among the

most complex of physical models, not only because they describe a large number of processes but also because they include nonlinear interactions.

Many studies have evaluated the representation of clouds and their radiative effects in GCMs. Their approaches vary widely, but most are evaluations of either the model climate or case studies. As we shall see, there is an unfortunate gap between these two approaches. In the numerous studies using one or more of the techniques we outline below, a lack of coherence in the application of these techniques to the same model is clearly visible. Most likely this is because each of the techniques as such requires substantial resources. However, this lack of coherence—a lack of strategy—when evaluating cloud parameterizations, has led to a considerable dilution of efforts.

Results of a number of recent studies will be used in an illustrative fashion to highlight some of the evaluation techniques in use today and to propose how a technique of compositing by dynamical regime might bridge the gap between model climate and case study and thereby provide new insight into cloud parameterization. Most of the studies used here have been carried out with various versions of the European Centre for Medium-Range Weather Forecasts (ECMWF) global forecast model, which applies the cloud parameterization of Tiedtke (1993) with recent modifications described in Jakob (2001). After assessing the current evaluation techniques we will propose

AFFILIATION: JAKOB—ECMWF, Shinfield Park, Reading, United Kingdom

CORRESPONDING AUTHOR: Christian Jakob, Bureau of Meteorology Research Centre, GPO Box 1289K, 150 Lonsdale St., Melbourne 3001, Victoria, Australia

E-mail: c.jakob@bom.gov.au

DOI: 10.1175/BAMS-84-10-1387

In final form 17 March 2003

©2003 American Meteorological Society

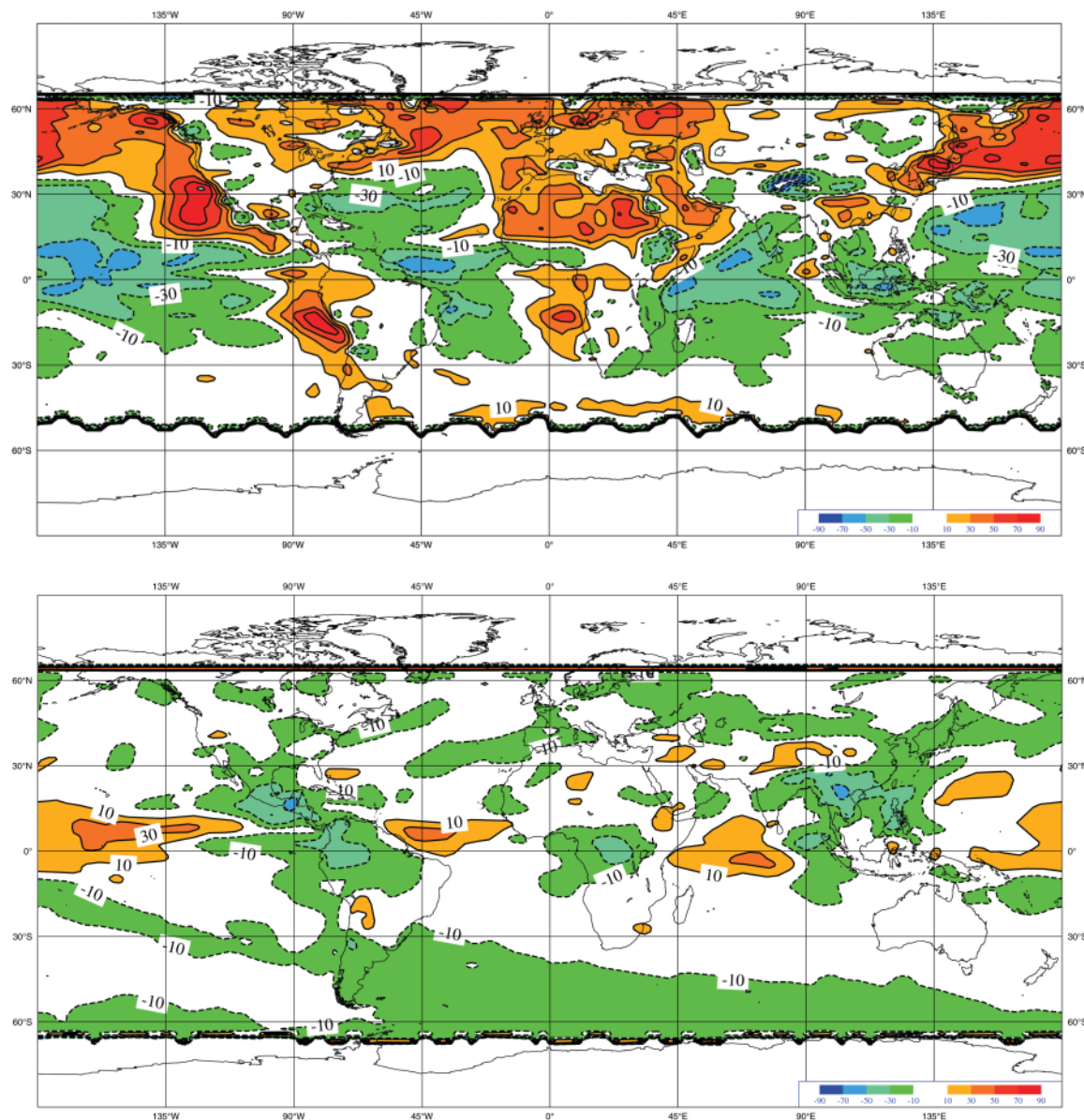
a strategy for cloud evaluation in GCMs that integrates most of those techniques—including compositing—into a coherent procedure.

EVALUATING THE MODEL CLIMATE. *Broadband radiative fluxes.* Clouds are included in GCMs mainly because they interact with radiation. Not surprisingly one of the most common ways to evaluate cloud parameterizations is to compare radiative fluxes produced by the model to those observed by satellites at the top of the atmosphere. The broadband flux measurements gathered during the Earth Radiation Budget Experiment (ERBE; Barkstrom and Smith 1986) are frequently used for such comparisons.

In such a comparison for a version (CY18R6) of the ECMWF model (Fig. 1), several regions of erro-

neous top of the atmosphere (TOA) radiation emerge for both shortwave and longwave radiation. TOA shortwave radiation is overestimated (too little reflection) in the extratropics, predominantly over the oceans, over the eastern parts of the subtropical oceans, and over the Sahara. It is underestimated over most of the deep Tropics and over the western parts of the subtropical oceans. The outgoing longwave radiation (OLR) is underestimated (positive difference) over much of the tropical ocean, strongly overestimated over the tropical continents, and overestimated to a lesser extent in the extratropics.

The overall model radiative fluxes per se do not indicate anything about the radiative effect of the model clouds. The errors could be caused not only by erroneous model clouds, but also by incorrect surface



albedo affecting shortwave radiation or incorrect surface temperatures or water vapor distribution affecting longwave radiation.

Because some error patterns coincide with regions dominated by particular cloud types, however, we may suspect problems with the description of the radiative effect of these clouds. For instance, TOA shortwave radiation is greatly overestimated in regions of extensive stratocumulus off the west coasts of landmasses in the subtropics. Given what we know about the albedo of the sea surface, it is more likely that these errors identified in Fig. 1 are related to errors in the radiative behavior of clouds. On the other hand, the large error in shortwave radiation over the Sahara, a region with almost no clouds, probably indicates a problem in the description of surface albedo.

Cloud radiative forcing. A better variable to compare is the cloud radiative forcing also derived from ERBE observations (e.g., Ellis 1978). The cloud radiative forcing at the TOA can easily be derived in a model comparing clear-sky radiative fluxes to the all-sky fluxes.¹

Comparing Fig. 2 to Fig. 1 confirms that clouds are the major source of radiative error at the top of the model atmosphere. This is not too surprising, since we know much more about clear-sky radiative transfer than about the representation of clouds and their interaction with radiative fluxes. The model clouds reflect too much solar radiation over the deep tropi-

cal oceans and the western parts of the subtropical oceans and too little over the eastern part of the subtropical oceans and in the extratropics. They also have excessive effects on longwave radiation over tropical oceans. These comparisons make it possible to assess the net effect of model clouds. One of the major drawbacks of the technique is that this net effect is the result of many parameters. The shortwave cloud radiative forcing in the trade cumulus regions, for instance, could be overestimated because cloud fractions are too high, cloud liquid water contents are too large, assumed particle sizes are too small, broken cloud effects in the radiation parameterization are misrepresented, or a combination of these. From the perspective of cloud modelers, this is an extremely dissatisfying evaluation since it provides only very limited guidance for future development. All one can learn is where, geographically, the general problems are.

Cloud fraction. One improvement is to evaluate the specific predictions of the cloud parameterization, such as the model cloud fraction. Figure 3 shows that the model underestimates cloud fraction in the stratocumulus regions off the west coast of the subtropical continents, underestimates cloud fraction over the extratropical oceans, and overestimates cloud fraction over the tropical oceans. These errors are consistent with the errors detailed in the cloud radiative forcing. However, no obvious error in cloud fraction over the trade cumulus areas could explain the large errors in cloud radiative forcing there.

While the use of total cloud cover provides an evaluation more directly relevant to cloud parameterization, many problems remain. Direct observations of cloud cover by surface observers (e.g., Warren et al. 1986, 1988) that are mostly limited to land- and satellite-derived cloud cover, such as from the International Satellite Cloud Climatology Project (ISCCP), are indirect and involve complex algorithms. The most serious drawback of model climate evaluation, however, is that, in a long model integration, feedbacks between different model errors occur. An apparently poor representation of clouds may therefore be caused not by a failing of the cloud parameteriza-

¹ The derivation of cloud radiative forcing from the model is different from the derivation from data, in that the clear-sky radiation in the model is calculated for cloudy columns by just ignoring the cloud variables, but still using the water vapor and temperature profiles of a cloudy column. In contrast, the cloud radiative forcing in the data is derived by comparing cloudy columns with neighboring (both in space and time) truly clear-sky columns. The difference introduced this way can amount to a few watts per square meter (Cess and Potter 1987; Cess et al. 1992) and needs to be considered in cases of small model errors. As will be shown later, the errors of the model in many regions exceed 10 W m^{-2} so that this effect should not affect the conclusions drawn here.

FIG. 1. Difference in (top) shortwave and (bottom) longwave radiation at the top of the atmosphere between a model integration and ERBE observations for Jun–Aug 1987. Upward fluxes are taken as negative so that a negative difference indicates a too-strong upward flux, that is, too much reflection in the shortwave case and too-high emission in the longwave part of the spectrum. Positive differences are shown as solid contours, negative differences as dashed. The contour interval is 20 W m^{-2} with shading starting at 10 W m^{-2} . The model integration is carried out with CY18R6 of the ECMWF model at T63L31 resolution. The initial date is 1 May 1987. Sea surface temperature (SSTs) are time varying and prescribed.

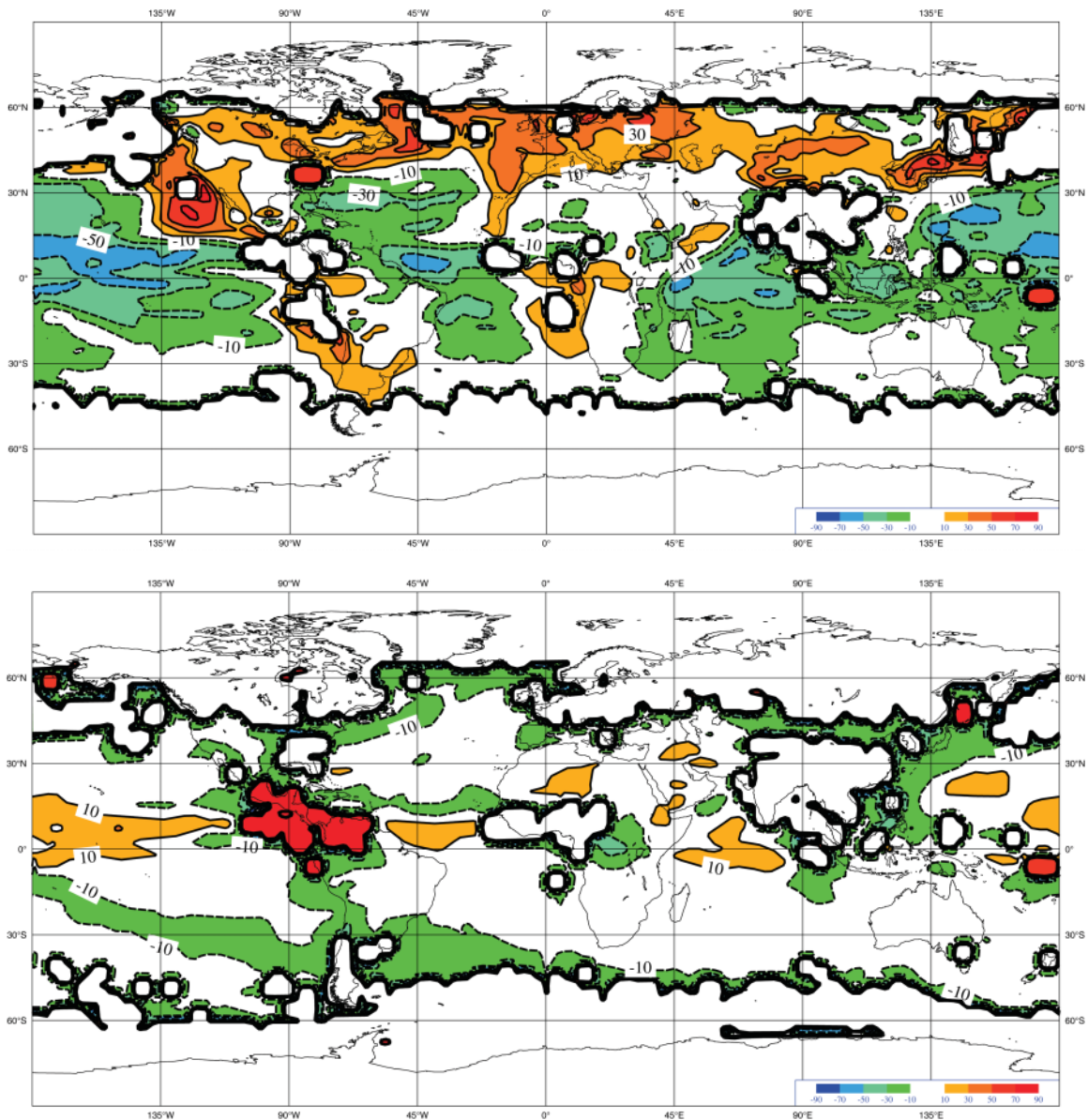


Fig. 2. Difference in (top) shortwave and (bottom) longwave cloud radiative forcing at the top of the atmosphere between a model integration and ERBE observations for Jun–Aug 1987. The model integration is carried out with CY18R6 of the ECMWF model at T63L31 resolution. The initial date is 1 May 1987. SSTs are time varying and prescribed. White areas surrounded by heavy contouring denote missing data.

tion but, for instance, by errors in the model’s large-scale circulation.

The climate of short-range NWP. Using short-range numerical weather prediction (NWP) to evaluate clouds in GCMs avoids the problem of building up large systematic model errors. That is because the large-scale circulation in such forecasts is strongly controlled by initial conditions generated with data assimilation systems.

Jakob (1999) used this approach to evaluate the ECMWF cloud parameterization (Fig. 4). Although the version of the model in Fig. 4 (the ECMWF re-analysis, or ERA; Gibson et al. 1997) differs from that shown in Fig. 3, and although the averaging periods are significantly different, the main error patterns (and in some areas even the magnitude of the errors) in total cloud cover as seen in Figs. 3 and 4 are very similar. This suggests that many of the errors in total cloud cover in the model climate are likely due to

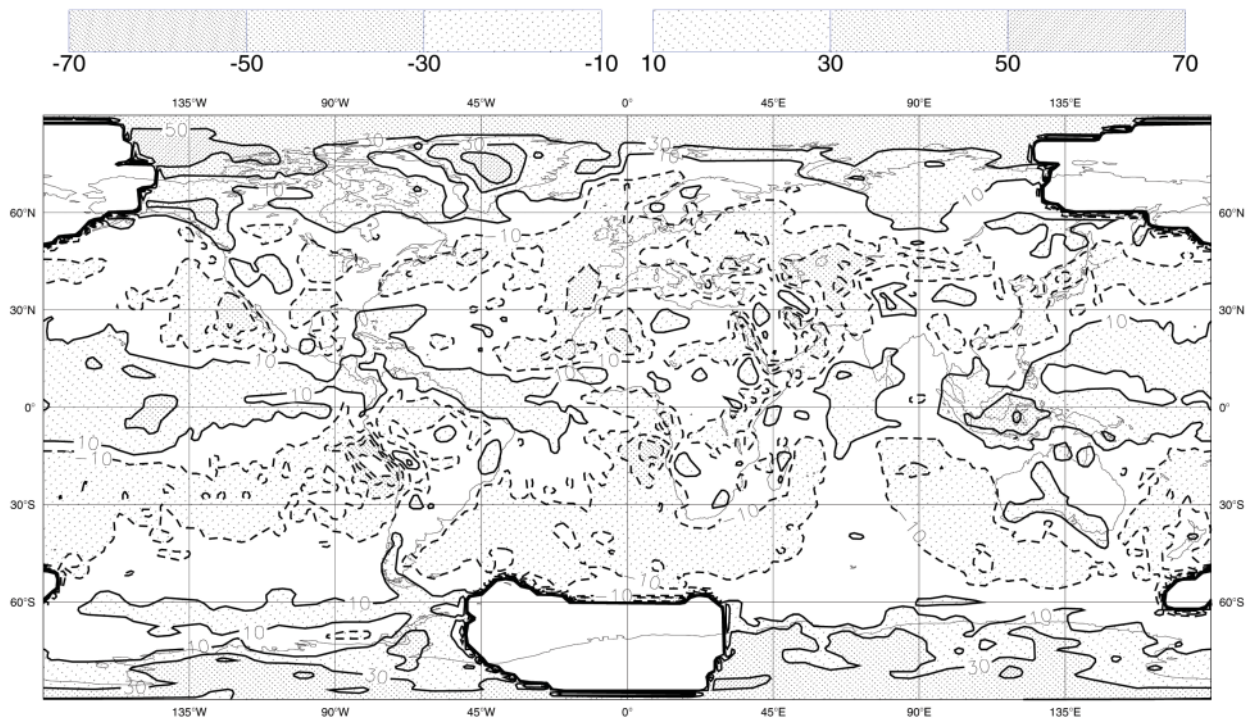


FIG. 3. Difference in total cloud cover between a model integration and ISCCP (Rossow and Schiffer 1983) observations for Jun–Aug 1987. The model integration is carried out with CY18R6 of the ECMWF model at T63L31 resolution. The initial date is 1 May 1987. SSTs are time varying and prescribed. White areas surrounded by heavy contouring denote missing data.

problems in the cloud or related (e.g., convection) parameterizations.

Another obvious comparison using short-range forecasts is that to cloud observations routinely collected by observers on the ground and distributed regularly via the Global Telecommunications System (GTS). Many NWP centers make this type of comparison daily. These routine statistics should reflect changes in cloud parameterization. For instance (in Fig. 5), there is a significant reduction in both mean error and standard deviation of cloud cover in April 1995. Not surprisingly, this can be traced back to the introduction of the cloud parameterization of Tiedtke (1993) into the operational ECMWF model.

The results of routine monitoring, like the results of climate simulations, guides the development of cloud parameterization by exposing general problem areas. The big advantage over climate simulations is that, through the use of short-range forecasts, model errors can be more easily ascribed to the parameterization itself since the large-scale flow is captured more realistically.

THE USE OF CASE STUDIES. Though the long-term studies above are valuable for studying the rep-

resentation of clouds in models, these studies cannot provide crucial insights into the reasons for the model failures. One can, of course, speculate on why certain model climate features exist, and use intuition and trial and error to correct the model shortcomings. However, without a clear understanding of the causes of the errors, this speculative approach may introduce even more errors that just happen to compensate for the already existing ones and hence give a better end result. In the extreme case, this approach might lead to an extensive “tuning” exercise in which adjustable model parameters are modified until a satisfactory end result is achieved.

This is clearly not a desirable model development strategy—even if necessary at times. It is therefore desirable to develop evaluation techniques that allow the study of the intrinsic workings and failings of the cloud parameterization. The most widely used methods in this context are case studies. The two most common model types used in these studies are NWP models and single-column models (SCM), which we will now examine.

NWP. In global NWP the full GCM can predict the state of the atmosphere globally on a daily basis for

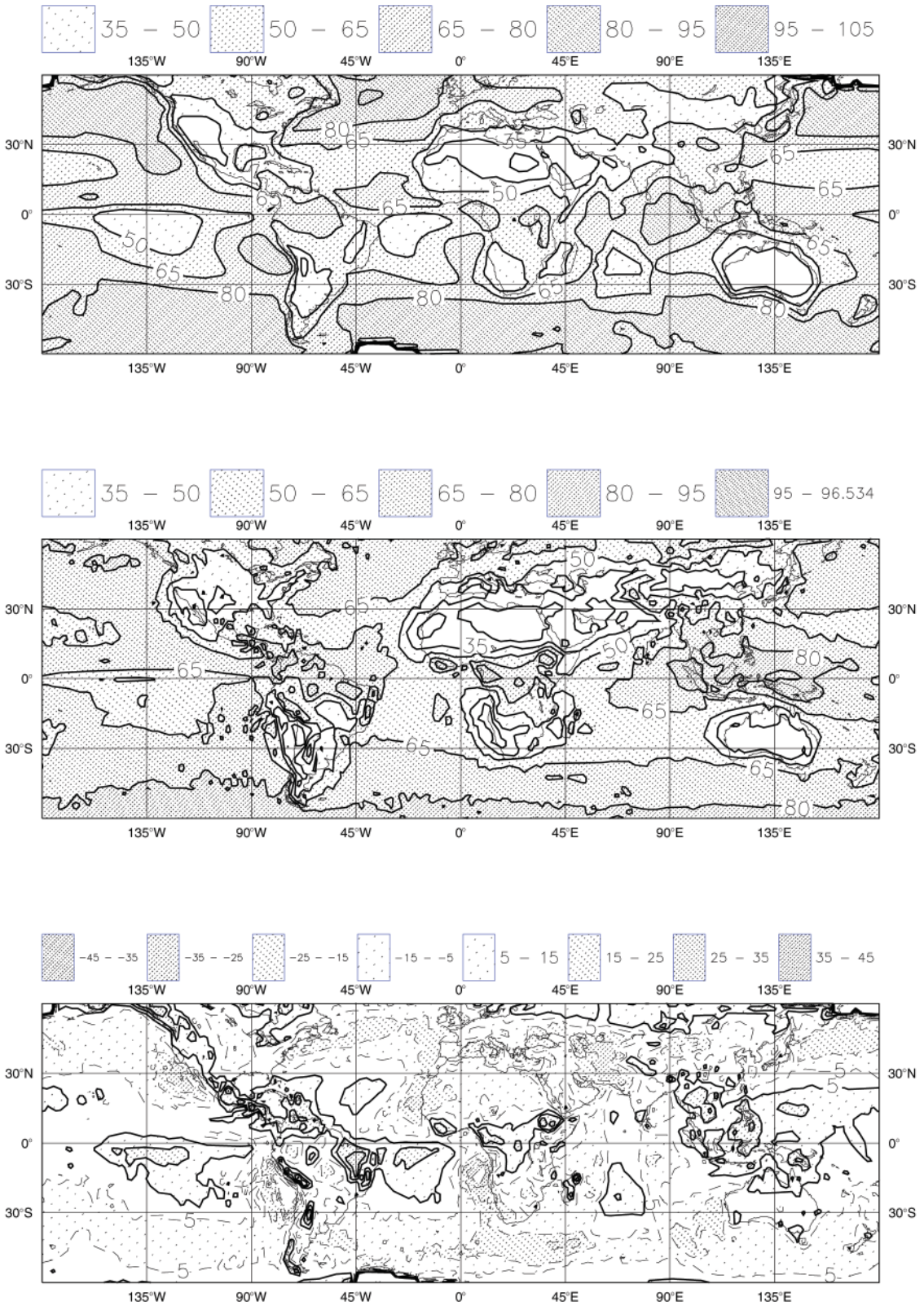


FIG. 4. Annual mean of total cloud cover averaged from Jul 1983 to Dec 1990 for (top) ISCCP, (middle) ERA, and (bottom) ERA minus ISCCP. Positive differences are depicted by thick solid lines, negative by thin dashed lines. From Jakob (1999).

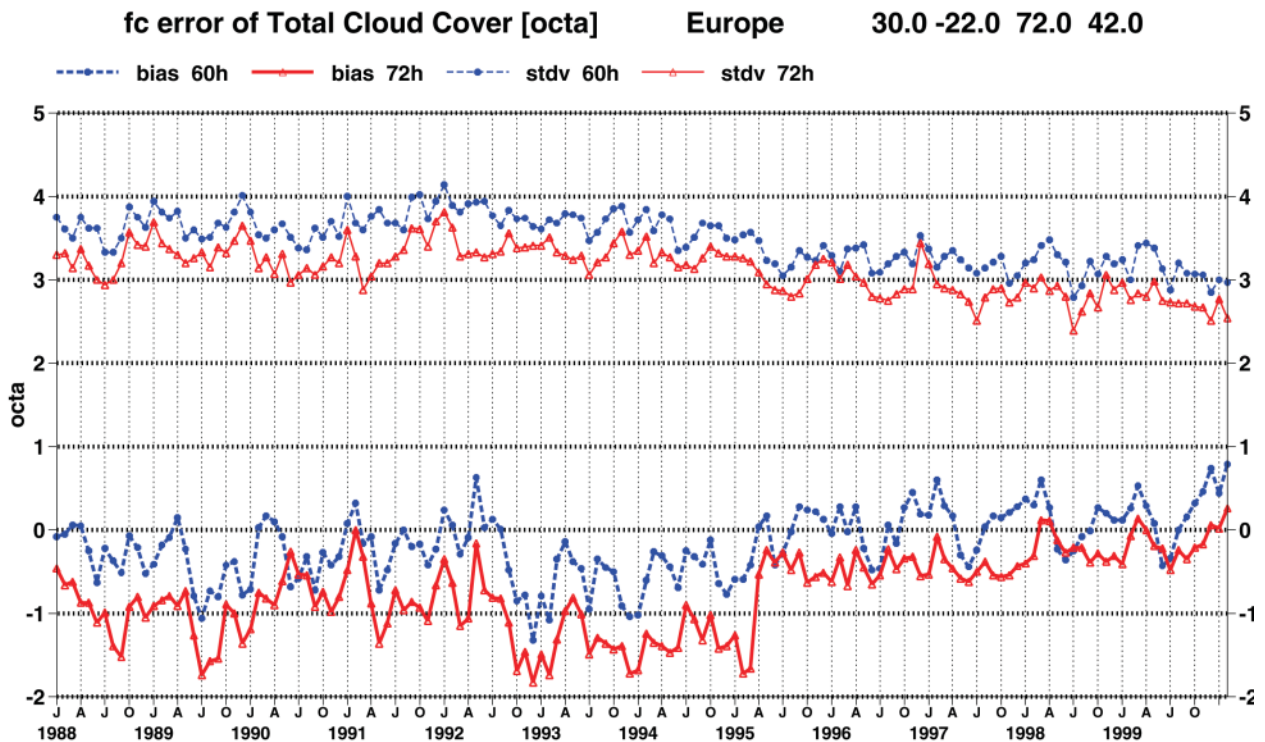


FIG. 5. Time series of short-range forecast model errors in total cloud cover (in octa) for Europe (the exact averaging area is indicated in the top-right corner) when compared to synoptic observations. Shown are monthly means of the mean error (thick lines) and the standard deviation (thin lines) for all operational 60-h forecasts (valid at 0000 UTC; dashed) and 72-h forecasts (valid at 1200 UTC; solid).

several days ahead. Each day many “case studies” are available from such a forecast. In order to find suitable observations for useful comparisons, it is desirable to minimize the influence of errors in other parts of the model. Therefore, it is common to choose forecasts of less than three days in which the numerical solutions are known to be most accurate.

One can evaluate the numerical predictions for a cloud parameter, such as total cloud cover, as they become available from the operational forecasts. Figure 6 shows a typical example for such an evaluation comparing model cloud cover to reports from ground observers. At first glance the agreement between model and observations is quite striking, with the major cloud features well captured. Closer inspection reveals several shortcomings, especially over southeastern Europe. The evaluation shown in Fig. 6 is far from comprehensive, but it can serve a monitoring purpose if regularly applied to operational NWP forecasts as is done at ECMWF and other NWP centers.

Of course, NWP evaluations are not restricted to the use of operational products. Model simulations during dedicated observational campaigns are desirable, since they allow evaluation of several versions of a parameterization or even completely different sets

of parameterizations. A number of studies using this kind of data have been carried out (Mace et al. 1998; Miller et al. 1999; Beesley et al. 2000; Hogan et al. 2001) and provide insight into a model’s performance.

Single-Column Modeling. A full NWP system for case studies is expensive, however, and it is still cumbersome to store and retrieve all the necessary information to gain insight into possible parameterization errors. It is therefore desirable to further simplify case studies. One such simplification is the use of so-called single-column models—a relatively inexpensive and computationally efficient evaluation method. Rather than using a fully three-dimensional GCM, a single column is “extracted” and the results of the model in only this column are considered. This is possible in part because physical parameterizations in GCMs are assumed to be locally applicable and therefore only require information at a single grid point and no direct interaction between model grid columns.

With only information in a single model column, the intricate details of a parameterization are easily explored. The information (such as advection) from neighboring grid cells, normally provided by the full GCM, needs to be prescribed in an SCM (Randall and

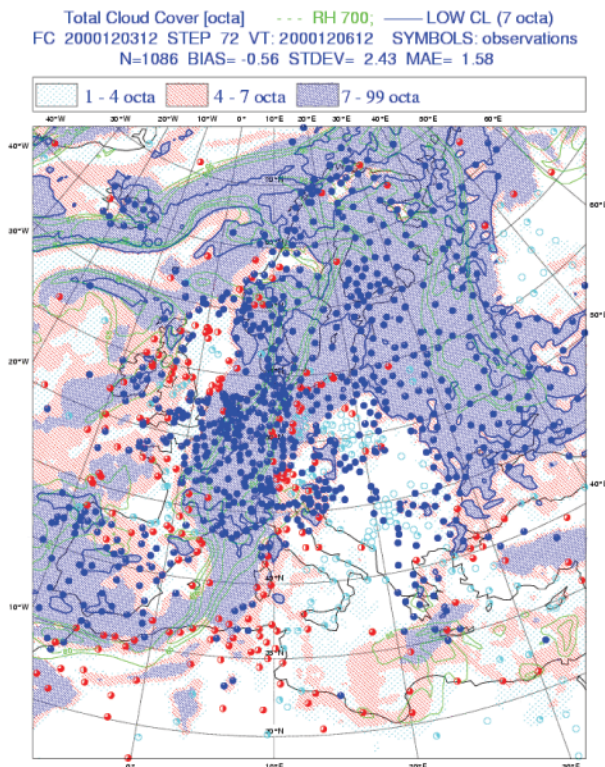


FIG. 6. Comparison of the ECMWF 72-h forecast of total cloud cover (shaded) with observations from the synoptic network (station symbols). The verifying time of the forecast is 1200 UTC 6 Dec 2000.

Cripe 1999). This prescription of the boundary conditions prevents errors that result from feedback within the full GCM. If the boundary conditions are perfect, all errors in the SCM are due solely to the parameterization (an advantage as long as one does not want to study the errors caused by the feedbacks).

The major difficulty in the SCM approach is to find suitable observations with enough information to derive the boundary conditions and to evaluate the parameterization in question. Very few datasets are suitable for SCM studies.² A parameterization tested with only a few cases, which have a high risk of being unrepresentative, may work in the single column but not in the full GCM.

Two major activities over the last 5 to 10 yr aim to improve the usefulness of SCMs in parameterization development, particularly cloud and convection parameterizations. The first is to gather more observations suitable for SCM studies. At the forefront of this activity is the Atmospheric Radiation Measurement program (ARM; Stokes and Schwartz 1994). ARM is collecting quasi-continuous data related to clouds in the Southern Great Plains (SGP); at Barrow, Alaska; and at Manus and Nauru islands in the tropical western Pacific. The observations are by design single column observations, obviously useful in SCM studies. Several studies have used mainly data from the SGP site (e.g., Randall and Cripe 1999).

A second activity aims to increase the number and therefore representativeness of available SCM case studies through the use of high-resolution “cloud-resolving models” (CRM), which enable detailed simulation of cloud processes. Their resolution depends on the type of cloud under study: from several meters (horizontal and vertical) for some boundary layer cloud studies to several hundred meters for some studies of deep convective systems. This effort is carried out by the World Climate Research Program’s Global Energy and Water Cycle Experiment (GEWEX) Cloud System Study (GCSS; GEWEX Cloud System Science Team 1993). CRMs that have been proven to simulate observations can provide the “truth” against which an SCM can be evaluated. GCSS has made many model intercomparisons involving both CRMs and SCMs in order to achieve this goal (e.g., Bechtold et al. 1996; Moeng et al. 1996; Bretherton et al. 1999; Bechtold et al. 2000; Redelsperger et al. 2000; Ryan et al. 2000).

Figure 7 shows an example of results from GCSS studies. The figure highlights both the strength and intrinsic difficulties of the GCSS approach so far. It is undoubtedly useful to use a variety of state-of-the-art CRMs. Assessing the spread in the results gives one confidence or, as in the case of Fig. 7, caution on using CRM results as a surrogate for observations. It is difficult to argue that the SCMs perform considerably worse in simulating the vertical distribution of cloud

² Examples of suitable datasets include data for modeling the Lagrangian evolution of the marine cloud-topped PBL gathered during the Atlantic Stratocumulus Transition Experiment (ASTEX; Albrecht et al. 1995), a number of datasets for the study of shallow cumulus clouds collected during the Barbados Oceanographic and Meteorological Experiment (BOMEX; Holland and Rasmusson 1973; Nitta and Esbensen 1974), and the Atlantic Trade-Wind Experiment (ATEX; Augstein et al. 1973), and a few datasets for the study of penetrative convection derived during the Global Atmospheric Research Program’s (GARP) Atlantic Tropical Experiment (GATE; e.g., Houze and Betts 1981), during the Tropical Ocean Global Atmosphere Coupled Ocean–Atmosphere Response Experiment (TOGA COARE; Webster and Lukas 1992), and more recently in the Atmospheric Radiation Measurement (ARM) program.

cover, but given the large spread in the CRM results and the absence of observations, this should not necessarily be interpreted as an indication that the SCMs are correct.

Figure 7 also highlights another danger, the so-called “intercomparison trap.” Simply using a large number of models and comparing the results will not automatically lead to the improvement of any of them. It is better to assess and test hypotheses about cloud parameterization with the full power of the GCSS framework. The points raised here have been recognized by the GCSS and its strategy has been revised (GCSS Science Plan, see online at www.gewex.org/gcss.html). Despite this criticism, one should not underestimate the achievements of programs like ARM and GCSS. The observational, cloud-scale and large-scale modeling communities have been brought closer together to jointly address problems in cloud parameterization; a large library of case studies

for various cloud types is now available, and a protocol exists for their use. More and more knowledge on the use of CRMs to investigate cloud parameterization is emerging. Overall SCMs provide valuable information for cloud parameterization development, particularly as part of a wider strategy of evaluating cloud parameterizations.

Use of NWP models can provide a natural extension to the SCM approach. If data are available at a single point for a given period of time, it is of course feasible to extract the forecast model results at the same time for comparison with observations. The use of the full GCM has two advantages. There is no need to prescribe the boundary conditions, and also the parameterization under investigation is working in

the environment it is designed for. However, use of a full GCM can add expense and the uncertainties of the other model components. Use of operational forecasts solves the first problem and use of short-range forecasts minimizes the second.

COMPOSITES—THE MISSING LINK?

The evaluation of the cloud parameterization with either highly averaged information or information from individual case studies has some serious drawbacks. The results of a comparison between a cloud-related parameter and observations can be completely unrelated to the quality of the parameterization. For instance, in Fig. 3 it is entirely possible that the vertical velocity at the top of the PBL is incorrect, thus leading to

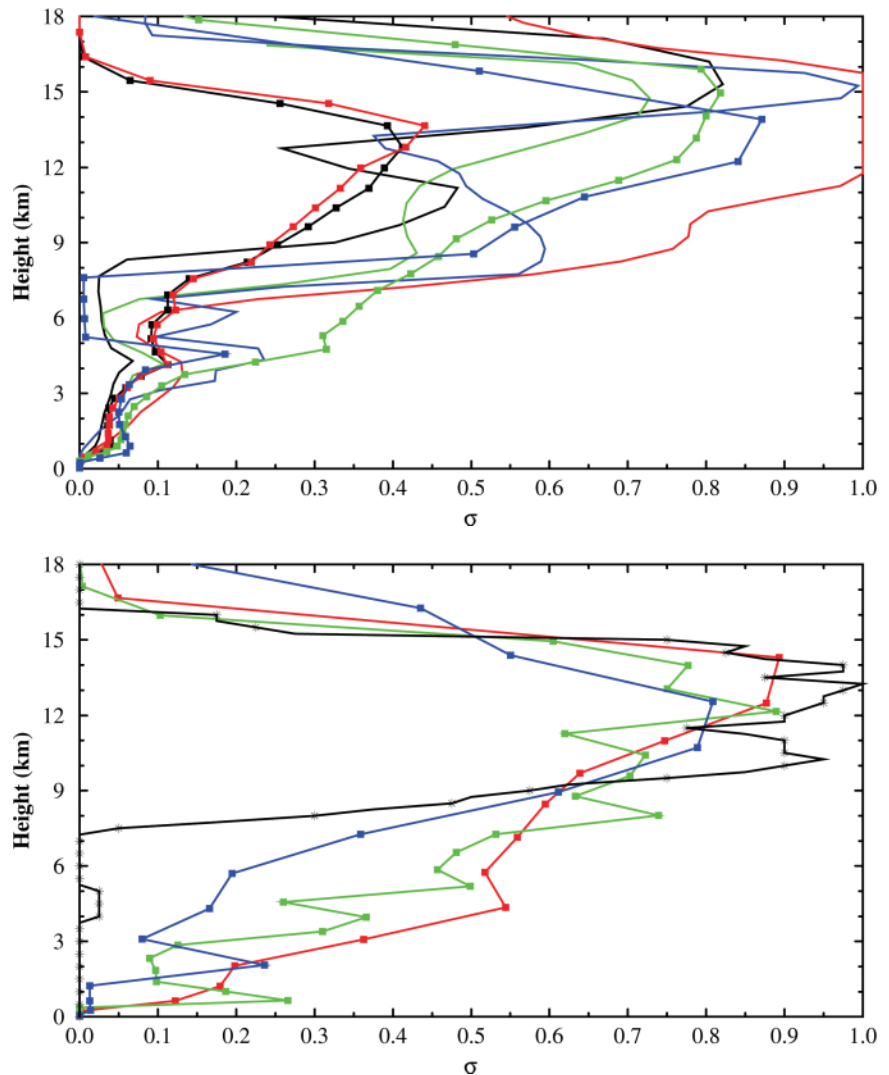


FIG. 7. Six-day average of cloud fraction as a function of height from simulations for a period of TOGA COARE (20–25 Dec 1992) with a number of (top) CRMs and (bottom) SCMs (Krueger et al. 2000, personal communication; see also online at <http://www.met.utah.edu/skrueger/gcss/case2.html>)

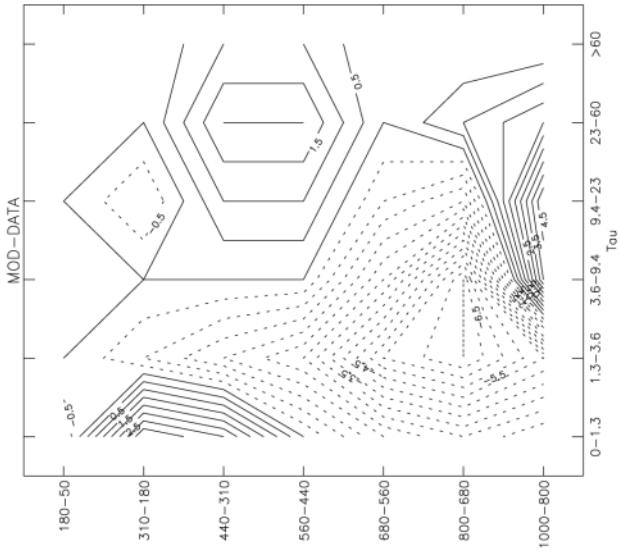
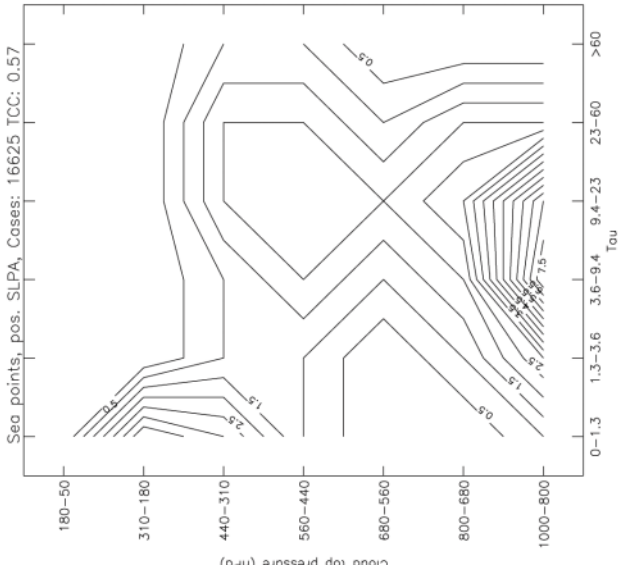
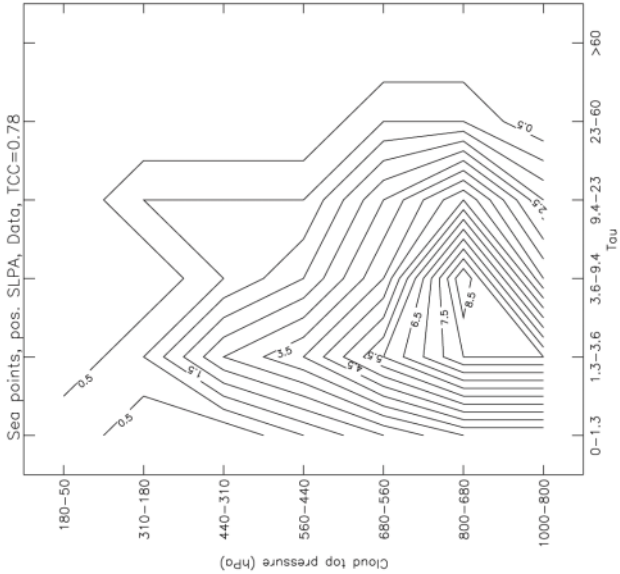
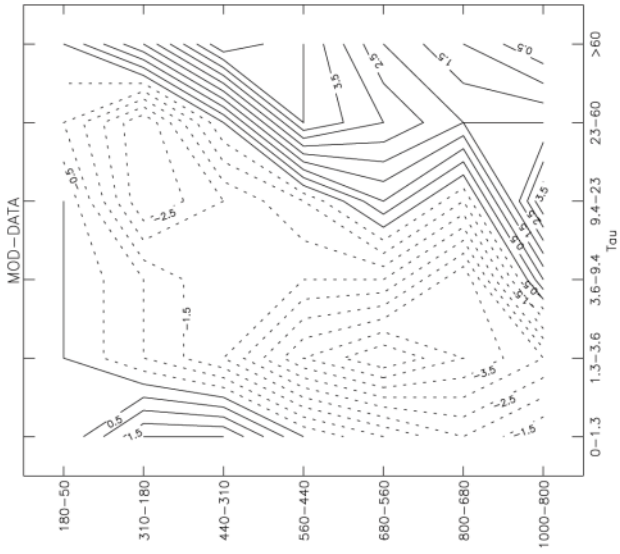
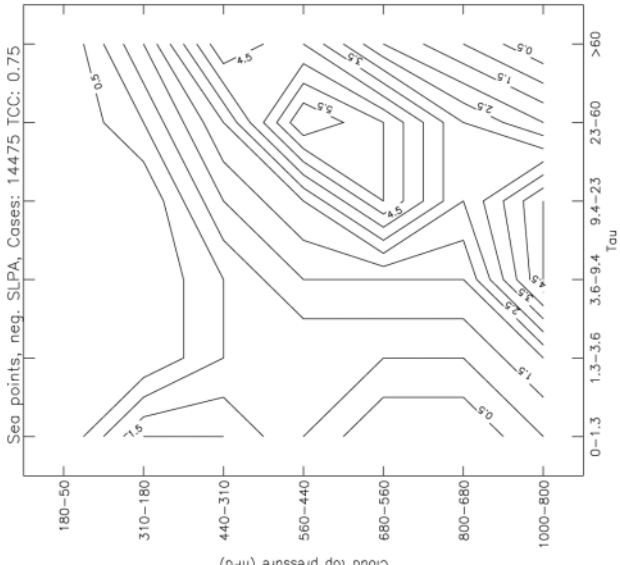
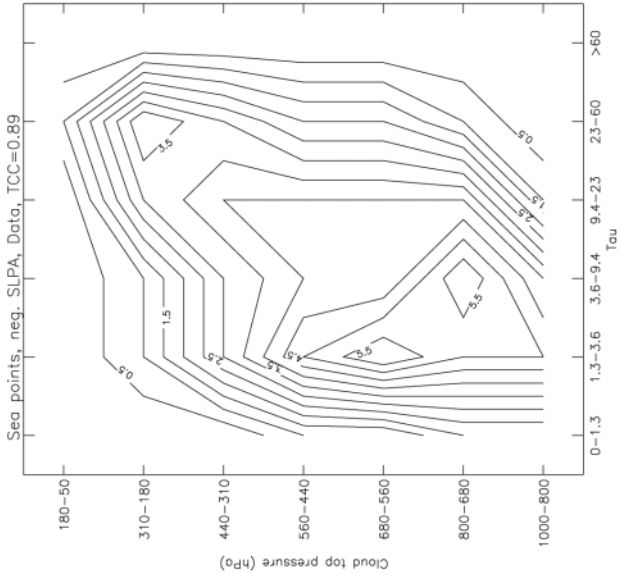


FIG. 8. Distribution of cloud-top pressure vs cloud optical thickness for the ocean areas of 30° to 60°N in the month of Apr 1992. The data is stratified into local negative sea level pressure anomalies (from the monthly mean) of (top) more than -5 hPa and positive sea level pressure anomalies of (bottom) more than +5 hPa. Distributions are derived from (left) ISCCP DI data, and (middle) ECMWF T106L31 12- and 24-h forecasts. (Right) The difference between model and data.

underestimation of cloud fraction off subtropical coasts. Since there are no observations of vertical velocity (although some studies of model subsidence rates have been carried out by Betts et al. 1995), it is difficult to assess whether this is really the case. It is obvious that comparisons based on the model climate alone can shed little light on what might be wrong with the cloud parameterization scheme.

In regions predominantly covered by one cloud type, such as the above-cited stratocumulus areas, the problems mentioned can be partly overcome by using averages of short-range forecasts to build up the “model climate.” But what about regions of large variance in cloud amount and type, such as over the extratropical oceans? Here, case studies are preferable but the choice of the cases for study is far from trivial.

A possible way to reconcile the model climate and case study approaches is to find “more intelligent” ways of averaging the data, so that the general characteristics of certain cloud systems remain intact even when a large number of cases is included in the average.

Clouds over the North Atlantic and North Pacific. An example of such an approach follows an idea of Tselioudis et al. (2000) and is based on data provided by ISCCP. Up to this point, only monthly mean values of total cloud fraction derived in this project have been considered here. ISCCP provides much more than that (Rossow and Schiffer 1991; Rossow and Schiffer 1999), such as cloud-top pressure and cloud optical thickness and the joint statistical distribution of these parameters. Tselioudis et al. (2000) survey these distribution functions for the Northern Hemisphere extratropical oceans as a function of the dynamical regime. They use the simplest indicator, namely surface pressure, to define three “dynamical” regimes as anomalously low, normal, and anomalously high pressure. Even with this extremely simple classification they find remark-

ably different cloud distributions for each regime. An example is shown on the left side of Fig. 8.

The difference in the “cloud” distribution between the panels is quite marked. In the below-average pressure regime (top), three predominant types of cloud appear: thick high-top clouds (very likely associated with frontal systems), medium-high-top thin clouds (most likely altostratus and altocumulus), and low clouds of medium optical thickness (associated with cloudiness at the top of the PBL). In the above-average regime the last becomes the predominant cloud type; thick, high-top clouds are virtually absent. One would expect this type of cloud distribution in the subsidence regions of high-pressure systems over the oceans and ahead of and behind extratropical cyclones.

Norris and Weaver (2001) and Tselioudis and Jakob (2002) use this technique to evaluate GCM performance not only in simulating mean cloud properties but also observed cloud structure differences between dynamical regimes. The middle and right of Fig. 8 show the result of such a comparison using short-range forecasts from the ECMWF model. The model is probed in exactly the same way as the data. [Local pressure anomalies based on the model results are calculated and the cloud-top pressure versus optical thickness distribution are derived. The technique for deriving these distributions from the model cloud fields is described in Klein and Jakob (1999, hereafter KJ99)].

A number of important differences between model and data are evident. First of all the total cloud cover is underestimated by about 15% for negative anomalies and by 20% for positive pressure anomalies. The clouds in negative anomalies are optically too thick and their tops are too low, compared to the data. When repeating the model analysis using the physical instead of the radiative cloud top (not shown), the latter effect disappears, indicating that the model is producing cloud tops at the right height but that the top parts of the clouds are optically too thin. This points to a possible deficiency in the ice water content.

In the positive pressure anomalies, the model cloud tops appear to be too low and the clouds are too thick optically. Also, the model produces both optically thick midlevel to high-top clouds and thin high-top clouds that are not observed. It is very possible that the cloud-top pressure error for low clouds in this regime is due to the difficulty of determining the exact cloud top in the data for clouds at the top of planetary boundary layers capped by an inversion. This can lead to a misinterpretation of the height associated with the measured brightness temperature.

However, the overestimation of optical thickness is most likely a true model problem.

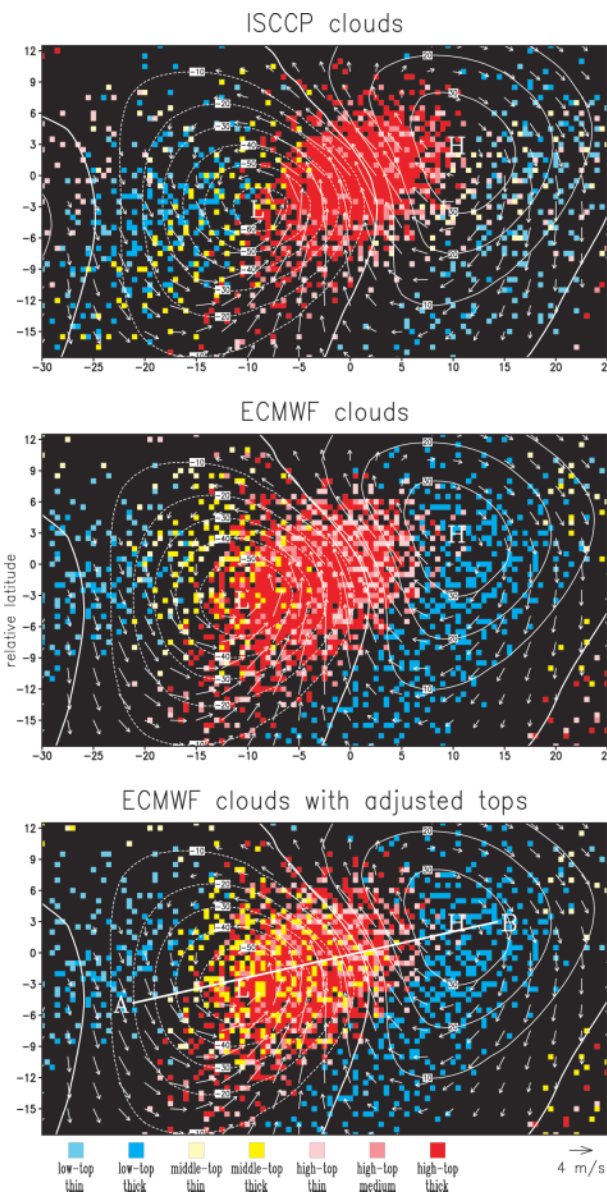
The model has an intriguing tendency to underestimate total cloud cover and seems to “compensate” by producing clouds that are too thick optically. The errors identified here, although for a different period, are nevertheless consistent with the underestimation of the reflection of solar radiation and shortwave cloud radiative forcing that was pointed out in Figs. 1 and 2. Naturally, the comparison here is for illustrative purposes only and is far from comprehensive. Although simple, the approach proves very useful. First, it combines cloud fraction information with radiative effects of the clouds present by studying optical thickness. Second, by splitting the dataset with a “dynamic” criteria, regimes in which model errors are particularly large can be identified and investigated further, possibly in an SCM. Finally, although short-range forecasts have been used here, this is not necessary. Dynamical criteria can be entirely defined by the model data, as was done here. This makes the technique useful not only for NWP models but for any GCM.

Validating clouds associated with extratropical cyclones. More complex compositing techniques can reveal cloud parameterization problems in even greater detail. For instance, KJ99 studied cloud structure in North Atlantic extratropical cyclones with a technique based on an idea of Lau and Crane (1995, hereafter LC95). LC95 identify from ISCCP data the optically thickest clouds occurring at given locations over a number of years. Each maximum-optical-thickness point (in time and space) is then considered the center of a relative coordinate system. When all maxima with surrounding points are centered onto this coordinate system, the result is a composite of the spatial distribution of the observed cloud and other meteorological fields, as shown in the top of Fig. 9.

In Fig. 9a, the relative low pressure center is southwest of the clouds with maximum optical thickness. A large shield of high-top thick clouds—normally associated with warm fronts—is northeast of the low pressure center. Middle-top thick clouds extend out of the low pressure region to the southwest. Ahead and behind the composite cyclone the cloud fields are dominated by low-top medium-thick clouds.

KJ99 have used short-range forecasts from ERA to generate the same picture of cloud distribution from the ECMWF model (using the same dates, locations, and analysis techniques as LC95). Figure 9b shows the model results when using physical cloud top, while Fig. 9c shows the results using the radiative cloud top to define cloud categories. As is evident from the fig-

ure, the model can reproduce the overall distribution of cloudiness around the cyclone quite well, perhaps with the exception of the cloud band extending southwestward from the low pressure center. On closer inspection, however, similar errors to those identified above in the pressure anomaly composites appear. The high-top clouds are optically too thin, leading to large errors in cloud-top height when using the radiative cloud top (which is what a satellite would most likely identify). The low-top clouds are optically too thick—in particular ahead of the cyclone. KJ99 identified the microphysical assumptions for ice settling as one of the major sensitivities for the simulation of the high-top cloud thickness. However, none of their sensitivity studies were able to reduce the error in low-top cloud optical thickness.



The two studies briefly summarized above demonstrate the usefulness of composite averaging in the evaluation of cloud parameterizations. Other more recent examples of this approach can be found in Webb et al. (2001), Norris and Weaver (2001), and Tselioudis and Jakob (2002). By averaging over a large number of cases so that key dynamical and hence cloud structures remain intact, it is possible to identify not only the deficiencies of the model cloud representation but also the dynamical environment in which they occur. This provides the first clues for possible model errors, which then can be investigated further. Compositing can thus play a central role in a strategy that can bring coherence to the application of different techniques for the evaluation of cloud parameterizations.

A STRATEGY FOR CLOUD PARAMETERIZATION EVALUATION.

At the core of our proposed strategy (outlined in Fig. 10) is the attempt to link the evaluation of the model climate to the selection of case studies through the use of compositing techniques. The evaluation of the model climate normally reveals geographical areas in which clouds and/or their effects are not correctly represented. As pointed out above, it is virtually impossible to infer reasons for the observed errors from such studies. Those can normally only be discovered in detailed case studies. But how should a case be chosen, such that it is typical for the model error? This is where compositing observations and model results using some criterion that describes the main mechanisms in cloud generation and/or maintenance should prove useful. By applying compositing techniques similar to those outlined above, not only is a first link to the possible causes for model problems established but also the typical model error is revealed. From the

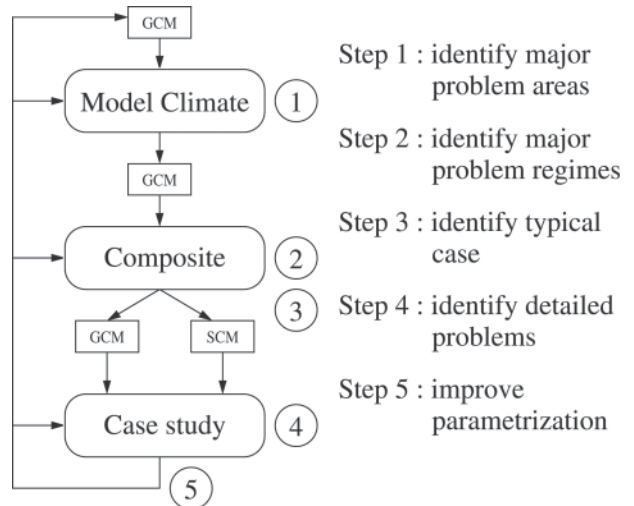


FIG. 10. A strategy for the evaluation of cloud parameterizations in GCMs.

(hopefully) considerable number of cases entering each composite average, one can then select those for which the model error is close to the mean error in the composite. That ensures that the following case study represents a typical model behavior rather than an extreme one. The case study can be carried out either with the full GCM, for example, in an NWP environment, or with a corresponding SCM. After improving the parameterization, it is, of course, necessary 1) to repeat the entire validation process to test the performance of the new parameterization in all aspects of the model and 2) to identify the next target for improvement.

A crucial component of the strategy proposed here is the availability of long-term datasets that are comprehensive enough to facilitate the use of either an NWP model or an SCM and that contain enough relevant information for the model evaluation. While there are still a number of problems to overcome, large strides in that direction have been made in particular through the generation of long-term cloud and radiation datasets in the ARM program; through improvements in NWP data assimilation systems, which bear fruit in the many reanalysis projects that have been carried out; and through the continuous collection and analysis of satellite products in programs such as ISCCP. The combination of these and many other available datasets not mentioned here should enable the use of the ideas presented here, even though an application of the entire philosophy outlined in Fig. 10 has not been achieved yet.

Finally, it is worthwhile pointing out that progress in cloud parameterization will ultimately never result

FIG. 9. (a) Distributions of 1000-hPa horizontal wind (arrows, see scale at bottom right) and geopotential height (contours, interval 10 m) from ERA analyses, and various cloud types (color pixels) from ISCCP observations as shown in LC95. The ordinate (abscissa) of the coordinate system corresponds to displacements in degrees from the reference location. Inside each $2.5^\circ \times 2.5^\circ$ area the presence and abundance of a certain cloud type is indicated by plotting a number of randomly scattered pixels with the color designated to the cloud type (see legend). Each pixel represents a 1% positive deviation from the background field as estimated as the 5-day average centered on the key dates. (b) As in Fig. 9a, but from 24-h ERA forecasts using physical cloud-top pressure; (c) as in Fig. 9b, but using emissivity-adjusted cloud-top pressure. From KJ99.

from evaluation studies of any kind, with or without a strategy. It can only be achieved by applying the knowledge about model errors gained from evaluation studies to develop novel ideas and from testing those using the strategy outlined here.

REFERENCES

- Albrecht, B. A., C. S. Bretherton, D. W. Johnson, W. H. Schubert, and A. Shelby Frisch, 1995: The Atlantic Stratocumulus Transition Experiment-ASTEX. *Bull. Amer. Meteor. Soc.*, **76**, 889–904.
- Augstein, E., H. Riehl, F. Ostapoff, and V. Wagner, 1973: Mass and energy transports in an undisturbed Atlantic trade-wind flow. *Mon. Wea. Rev.*, **101**, 101–111.
- Barkstrom, B. R., and G. L. Smith, 1986: The Earth Radiation Budget Experiment: Science and implementation. *Rev. Geophys.*, **24**, 379–390.
- Bechthold, P., S. K. Krueger, W. S. Lewellen, E. van Meijgaard, C.-H. Moeng, D. A. Randall, A. van Ulden, and S. Wang, 1996: Modelling a stratocumulus-topped PBL: Intercomparison among different one-dimensional codes and with large eddy simulation. *Bull. Amer. Meteor. Soc.*, **77**, 2033–2042.
- , and Coauthors, 2000: A GCSS model intercomparison for a tropical squall line observed during TOGA-COARE. Part II: Intercomparison of SCMs and with CRM. *Quart. J. Roy. Meteor. Soc.*, **126**, 865–888.
- Beesley, J. A., C. S. Bretherton, C. Jakob, E. Andreas, J. Intrieri, and T. Uttal, 2000: A comparison of ECMWF model output with observations at SHEBA. *J. Geophys. Res.*, **105**, 12 337–12 349.
- Betts, A. K., C. S. Bretherton, and E. Klinker, 1995: Relation between mean boundary layer structure and cloudiness at the R/V Valdivia during ASTEX. *J. Atmos. Sci.*, **52**, 2752–2762.
- Cess, R. D., and G. L. Potter, 1987: Exploratory studies of cloud radiative forcing with a general circulation model. *Tellus*, **39A**, 460–473.
- , —, W. L. Gates, J. J. Morcrette, and L. Corsetti, 1992: Comparison of general circulation models to Earth Radiation Budget Experiment data: Computation of clear-sky fluxes. *J. Geophys. Res.*, **97D**, 20 421–20 426.
- del Genio, A. D., M.-S. Yao, W. Kovari, and K. K.-W. Lo, 1996: A prognostic cloud water parameterization for global climate models. *J. Climate*, **9**, 270–304.
- Ellis, J. S., 1978: Cloudiness, the planetary radiation budget and climate. Ph.D. thesis, Colorado State University, 129 pp.
- Fowler, L. D., D. A. Randall, and S. A. Rutledge, 1996: Liquid and ice cloud microphysics in the CSU general circulation model. Part I: Model description and simulated microphysical processes. *J. Climate*, **9**, 489–529.
- GEWEX Cloud System Science Team, 1993: The GEWEX Cloud System Study (GCSS). *Bull. Amer. Meteor. Soc.*, **74**, 387–399.
- Gibson, J. K., P. Källberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano, 1997: ECMWF Re-Analysis Project Report Series: 1. ERA description. Tech. Rep. 1997, ECMWF, 72 pp.
- Hogan, R. J., C. Jakob, and A. J. Illingworth, 2001: Comparison of ECMWF winter-season cloud fraction with radar-derived values. *J. Appl. Meteor.*, **40**, 513–525.
- Holland, J. Z., and E. M. Rasmusson, 1973: Measurements of atmospheric mass, energy and momentum budgets over a 500-kilometer square of tropical ocean. *Mon. Wea. Rev.*, **101**, 44–55.
- Houze, R. A., Jr., and A. K. Betts, 1981: Convection in GATE. *Rev. Geophys. Space Phys.*, **19**, 541–576.
- Jakob, C., 1999: Cloud cover in the ECMWF reanalysis. *J. Climate*, **12**, 947–959.
- , 2001: The representation of cloud cover in atmospheric general circulation models. Ph.D. thesis, Ludwig-Maximilians-University Munich, 193 pp.
- Klein, S. A., and C. Jakob, 1999: Validation and sensitivities of frontal clouds simulated by the ECMWF model. *Mon. Wea. Rev.*, **127**, 2514–2531.
- Lau, N.-C., and M. W. Crane, 1995: A satellite view of the synoptic-scale organization of cloud properties in midlatitude and Tropical circulation systems. *Mon. Wea. Rev.*, **123**, 1984–2006.
- 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.
- Miller, S. D., G. L. Stephens, and A. C. M. Beljaars, 1999: A validation survey of the ECMWF prognostic cloud scheme using LITE. *Geophys. Res. Lett.*, **26**, 1417–1420.
- Moeng, C.-H., and Coauthors, 1996: Simulation of a stratocumulus-topped PBL: Intercomparison among different numerical codes. *Bull. Amer. Meteor. Soc.*, **77**, 261–278.
- Nitta, T., and S. Esbensen, 1974: Heat and moisture budget analyses using BOMEX data. *Mon. Wea. Rev.*, **102**, 17–28.
- Norris, J. and C. Weaver, 2001: Improved techniques for evaluating GCM cloudiness applied to the NCAR CCM3. *J. Climate*, **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**, 24 527–24 546.

- Redelsperger, J. L., and Coauthors, 2000: A GCS model intercomparison for a tropical squall line observed during TOGA-COARE. Part I: Cloud-Resolving Models. *Quart. J. Roy. Meteor. Soc.*, **126**, 823–863.
- Rossow, W. B., and R. A. Schiffer, 1983: The International Satellite Cloud Climatology Project (ISCCP): The first project of the World Climate Research Program. *Bull. Amer. Meteor. Soc.*, **64**, 779–784.
- , and —, 1991: ISCCP cloud data products. *Bull. Amer. Meteor. Soc.*, **72**, 2–20.
- , and —, 1999: Advances in understanding clouds from ISCCP. *Bull. Amer. Meteor. Soc.*, **80**, 2261–2287.
- Ryan, B. F., and Coauthors, 2000: Simulation of a cold front using cloud-resolving, limited-area, and large-scale models. *Mon. Wea. Rev.*, **128**, 3218–3235.
- Slingo, J. M., 1987: The development and verification of a cloud prediction scheme for the ECMWF model. *Quart. J. Roy. Meteor. Soc.*, **113**, 899–927.
- Smith, R. N. B., 1990: A scheme for predicting layer clouds and their water content in a general circulation model. *Quart. J. Roy. Meteor. Soc.*, **116**, 435–460.
- Stokes, G. M., and S. E. Schwartz, 1994: The Atmospheric Radiation Measurement (ARM) Program: Programmatic background and design of the Cloud and Radiation Test Bed. *Bull. Amer. Meteor. Soc.*, **75**, 1201–1221.
- Sundqvist, H., 1978: A parameterization scheme for non-convective condensation including prediction of cloud water content. *Quart. J. Roy. Meteor. Soc.*, **104**, 677–690.
- Tiedtke, M., 1993: Representation of clouds in large-scale models. *Mon. Wea. Rev.*, **121**, 3040–3061.
- 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**, 4781, doi:10.1029/2002JD002259.
- , Y. Zhang, and W. B. Rossow, 2000: Cloud and radiation variations associated with northern midlatitude low and high sea level pressure regimes. *J. Climate*, **13**, 312–327.
- Warren, S. G., C. J. Hahn, J. London, R. M. Chervin, and R. L. Jenne, 1986: Global distribution of total cloud cover and cloud type amounts over land. NCAR Tech. Note NCAR/TN-273;plSTR, 29 pp.
- , —, —, and —, 1988: Global distribution of total cloud cover and cloud type amounts over the oceans. NCAR Tech. Note NCAR/TN-317+STR, 42 pp.
- Webb, M., C. Senior, S. Bony, J.-J. Morcrette, 2001: Combining ERBE and ISCCP data to assess clouds in three climate models. *Climate Dyn.*, **17**, 905–922.
- Webster, P. J., and R. Lukas, 1992: TOGA COARE: The Coupled Ocean–Atmosphere Response Experiment. *Bull. Amer. Meteor. Soc.*, **73**, 1377–1416.