EVALUATING PARAMETERIZATIONS IN GENERAL CIRCULATION MODELS:
Climate Simulation Meets Weather Prediction

BY THOMAS J. PHILLIPS, GERALD L. POTTER, DAVID L. WILLIAMSON, RICHARD T. CEDERWALL, JAMES S. BOYLE, MICHAEL FIORINO, JUSTIN J. HNILO, JERRY G. OLSON, SHAOCHENG XIE, AND J. JOHN YIO

Numerical weather prediction methods show promise for improving parameterizations in climate GCMs.

INTRODUCTION. Climate simulations performed with general circulation models (GCMs) are widely viewed as the principal scientific basis for developing policies to address potential future global climate change (Houghton et al. 2001). In order to reduce uncertainties in these GCM projections of future climate, there is a compelling need to improve the simulation of processes that produce the *present* climate. This undertaking demands close attention to systematic errors in GCM simulations.

Systematic errors are persistent (average) departures of the model solution from an appropriate observational standard. For example, the GCM sys-

tematic climate error is defined by the departure of the simulated climate statistics from the observed (e.g., by a difference in monthly means). For GCMs used in numerical weather prediction (NWP), an additional and relevant systematic error is defined instead by the mean departure of the model forecast from observations at short time scales. As the length of the forecast increases, the systematic forecast error approaches the systematic climate error (Palmer 1999).

Thus, in order to enhance GCM performance ongoing interdependent efforts are needed to diagnose the details of model systematic errors by comparing GCM simulations with available observations over a range of time scales, and to reduce these systematic errors by improving the representation of key processes, and thereby increase the accuracy of GCM simulations relative to available observations.

For many years, the Working Group on Numerical Experimentation (WGNE) of the World Climate Research Programme (WCRP) has coordinated efforts to diagnose systematic errors in atmospheric GCMs (AGCMs). WGNE, for example, has sponsored conferences on the characteristics and causes of systematic errors in GCMs designed both for NWP and climate applications (WGNE 1988; Jasper and Meighen 2000). WGNE also has promoted collaborative initiatives such as the Atmospheric Model Intercomparison Project (AMIP; Gates 1992; Gates et al. 1999) to ana-

AFFILIATIONS: PHILLIPS, POTTER, CEDERWALL, BOYLE, FIORINO, HNILO, XIE, AND YIO—Lawrence Livermore National Laboratory, Livermore, California; WILLIAMSON AND OLSON—National Center for Atmospheric Research,* Boulder, Colorado
*The National Center for Atmospheric Research is sponsored by the National Science Foundation
CORRESPONDING AUTHOR: Thomas J. Phillips, Lawrence Livermore National Laboratory, Mail Code L-103, P.O. Box 808, Livermore, CA 94551
E-mail: phillips14@llnl.gov

lyze systematic climate errors in AGCM simulations made with prescribed ocean boundary conditions. The WCRP Working Group on Coupled Modelling (WGCM) is now extending the diagnosis of systematic errors to climate simulations of coupled ocean–atmosphere GCMs (OAGCMs) through initiatives such as the Coupled Model Intercomparison Project (CMIP; Meehl et al. 2000).

Ongoing efforts to reduce GCM systematic errors entail both enhancements of the resolution at which the model state variables (e.g., for AGCMs, the pressure, temperature, moisture, and wind fields) are predicted, and fundamental improvements in the parameterizations of unresolved subgrid-scale processes (e.g., radiation, clouds, convection, precipitation microphysics, turbulent fluxes, and diffusion). In climate simulations, the parameterizations are crucially important for correct representation of relevant processes, while the computational costs of increasing resolution are very high. Hence, parameterization development usually is emphasized over resolution enhancement as the chief means of reducing systematic errors in GCMs designed for climate simulation (hereafter, "climate GCMs").

The deciding factor in choosing a new parameterization for a climate GCM is whether its inclusion brings the simulated climate into closer agreement with the observed statistics. However, there are inherent limitations in evaluating GCM parameterizations only in the climate-simulation mode. First, due to sampling limitations, the observed climate statistics only roughly approximate the statistics of the global climate system—to greater or lesser degree—depending on the process of interest (e.g., Kistler et al. 2001). Moreover, because the GCM climate state reflects compensating errors in the simulation of many nonlinear processes, it is very difficult to attribute these errors to particular parameterization deficiencies. In such a context also, the parameterizations are driven by an unrealistic large-scale state, so that it is difficult to evaluate their performance objectively (Schubert and Chang 1996).

For these reasons, climate GCM developers have adopted process-oriented approaches that employ high-frequency local observations for evaluating parameterizations. Some effects of introducing a new GCM parameterization can be assessed, for example, within the framework of single-column models (SCMs) or cloud-resolving models (CRMs; e.g., Betts and Miller 1986; Krueger 1988; Krueger et al. 1995; Randall et al. 1996; Xie et al. 2002; Xu et al. 2002; Randall et al. 2003). The strength of this approach is that the column parameterizations are driven by an evolving large-scale dynamical state that is specified from observations.

However, because all relevant high-frequency dynamical forcings for the atmospheric column must be specified, there are only a limited number of observational cases at a few locations that can be studied with an SCM or CRM. Feedbacks from the column parameterizations to the large-scale dynamics also cannot be represented by these models.

In order to include such feedbacks, some model developers (e.g., Jeuken et al. 1996; Kaas et al. 1999) retain the full GCM, but continuously relax the simulated mass and momentum fields toward a six hourly (6-h) global weather analysis. The objective is to constrain the GCM large-scale dynamics close to reality so that the resulting mean short-term tendency errors are attributable mainly to parameterized physical processes. Diagnosis of these tendency errors then can yield insights that might suggest possible ways to reduce climate systematic errors.

Operational NWP centers follow yet another methodology in developing parameterizations for fine-resolution weather-forecasting GCMs: the model dynamics are allowed to evolve freely and to interact fully with the parameterizations, so that all the forcings and feedbacks are generated by the GCM (e.g., Jakob 2003). In this approach, the state variables of the forecast GCM are first initialized by a data assimilation system (DAS) that usually is built around the GCM itself. After ingestion of all available observations (e.g., surface, radiosonde, aircraft, and satellite measurements), the DAS applies variational methods to produce an optimal analysis of the global weather that defines the initial conditions for the forecast GCM (Daley 1991; Kalnay 2003). In addition, the DAS provides departures of the model from observations that can be used as a further guide to parameterization errors.

Given an accurate analysis, it is reasonable to assume that the model state remains close to "truth" in the early period of the forecasts, so that the systematic forecast error can be attributed largely to parameterization deficiencies. This systematic error is estimated from differences between the mean (at fixed elapsed times) of a sequence of short-range (~5 day) forecasts of state variables and corresponding NWP analyses of weather observations. In addition, errors in parameterized model variables (e.g., radiative and turbulent fluxes, cloud properties, precipitation, etc.) are estimated from field observations or other data that are not ordinarily assimilated by the DAS (e.g., Mace et al. 1998; Miller et al. 1999). Based on developers' insights gleaned from these differences, selected model parameterizations then are modified so as to ameliorate the perceived deficiencies. These scheme changes also are evaluated in short-range GCM forecasts to determine whether

they reduce the model's forecast systematic errors. If that is the case, the new parameterizations usually also are evaluated in model integrations beyond the deterministic forecast range of ~15 days to determine whether they reduce climate systematic errors.

In view of the benefits of applying this NWP development methodology to fine-resolution forecast models, some scientists (e.g., Hollingsworth et al. 1998; Miller 1999; Jakob 2003) have advocated adoption of analogous procedures for developing parameterizations in coarse-resolution climate GCMs. To realize this goal, alternatives to NWP procedures that rely on a GCM-compatible DAS must be developed, since many climate models lack such a resource. Working relationships also need to be forged between GCM developers and parameterization specialists who do not always share the same institutional affiliations.

The U.S. Department of Energy (USDOE) is well-positioned to foster these scientific collaborations because of the support it provides to GCM developers through the Climate Change Prediction Program (CCPP) and to parameterization specialists through the Atmospheric Radiation Measurement (ARM) Program. Moreover, the extensive high-frequency ARM field data that have been collected over the last decade (Stokes and Schwartz 1994; Ackerman and Stokes 2003) are potentially very useful for evaluating GCM parameterizations (e.g., Morcrette 2002). Hence, the USDOE has established a new joint initiative, the CCPP-ARM Parameterization Testbed (CAPT) in order to support implementation of the NWP methodology in climate GCMs, and to facilitate the needed scientific collaborations as well.

The remainder of this article reports on the progress of CAPT to date. The section titled "Premise and Protocol" discusses the scientific premise of the project and an outline of the steps in the CAPT diagnostic protocol. The section titled "Technical Details" elaborates technical aspects of implementing a prototype of the CAPT protocol for version 2.0 of the Community Atmosphere Model (CAM2), an AGCM developed under the auspices of the National Center for Atmospheric Research (NCAR; Collins et al. 2003; Kiehl and Gent 2004). The "Results" section presents preliminary results of applying the CAPT methodology in this climate model, and the concluding section briefly summarizes the main points.

PREMISE AND PROTOCOL. CAPT is promoting a diagnostic approach that is new for climate models that are not associated with operational forecast centers: the use of 6-h global NWP analyses and high-frequen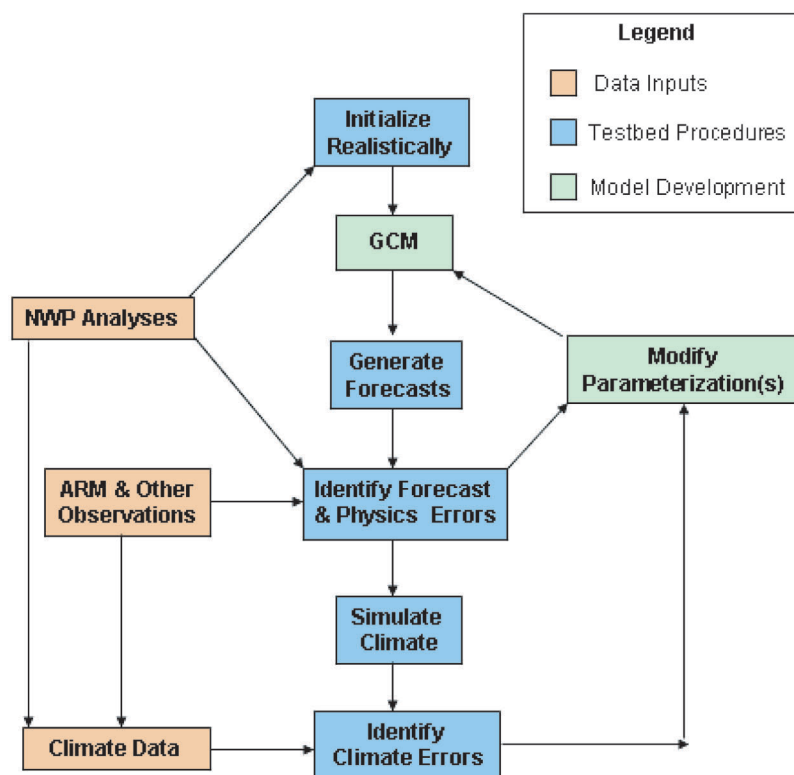cy unassimilated observations of parameterized variables (such as provided by ARM) to evaluate short-range weather forecasts made with climate GCMs that are initialized realistically.

The CAPT premise is that, as long as the dynamical state of the forecast remains close to that of the verifying analyses, the systematic forecast errors are predominantly due to deficiencies in the model parameterizations. It is then appropriate to compare parameterized variables with available observations collected under the same dynamical conditions, and to interpret their differences as indications of parameterization shortcomings. In themselves, these differences do not automatically determine a needed parameterization change, but they can provide developers with insights as to how this might be done. Then if changing the parameterization is able to render a closer match between parameterized variables and the evaluation data, and if this change also reduces the systematic forecast errors or any compensating errors that are exposed, the modified parameterization can be regarded as more physically realistic than its predecessor.

The basic elements of the CAPT protocol are illustrated in Fig. 1. First the climate GCM is initialized without recourse to a DAS, but with its atmospheric state specified from actual synoptic conditions, while also being in approximate dynamical balance (see details in the section titled "Initialization Procedures"). Next, the climate model is run in a short-range forecast mode, and these predictions are compared against the actual evolving atmospheric state, as determined both from NWP analyses and unassimilated observations of parameterized variables. Differences between the model predictions and these evaluation data are diagnosed in order to learn more about the parameterization deficiencies, and to aid the developers in establishing needed changes. The efficacy of modifying the parameterizations then can also be evaluated in a short-range forecasting framework.

However, the overriding goal is not that the climate GCM produce the "best" weather forecast, but only a good approximation thereof, so that the parameterizations respond to a realistic large-scale state. Thus, even though the weather forecasts of a coarse-resolution climate GCM may be inferior to those of a fine-resolution NWP model, *relative* decreases in systematic error are still indicative of improved parameterizations in the climate GCM. Moreover, the rich variety of weather phenomena allows the model parameterizations to be tested and diagnosed over a wide range of conditions, and at much less computational expense than is required in the climate-simulation mode. In CAPT, therefore, weather fore-

CAPT Protocol

Legend
- Data Inputs
- Testbed Procedures
- Model Development

Initialize Realistically → GCM → Generate Forecasts → Identify Forecast & Physics Errors → Simulate Climate → Identify Climate Errors → Modify Parameterization(s)

NWP Analyses

ARM & Other Observations

Climate Data

FIG. 1. Flow diagram of the CAPT protocol.

demonstrated in progressively longer (extended-range, seasonal, interannual, decadal, etc.) simulations. GCM parameterizations that are improved at short time scales also may require some further "tuning" of free parameters in order to achieve radiative balance in climate mode. Parameterization evaluation in climate simulations is therefore a necessary part of the CAPT protocol (Fig. 1).

TECHNICAL DETAILS. Several technical aspects of the CAPT protocol, as applied thus far to the CAM2 model, are elaborated here.

Evaluation data. The efficacy of the CAPT methodology depends crucially on the accuracy of current NWP analyses. For instance, earlier attempts (e.g., Williamson and Daley 1986) to diagnose parameterizations by using analyses to evaluate the weather forecasts of climate GCMs were thwarted by the strong influence of the NWP model on the analysis, which had especially negative impacts on the accuracy of the analyzed atmospheric moisture and parameterized variables.

NWP analyses now are much better approximations of the actual atmospheric state, as shown by recent findings (Simmons and Hollingsworth 2001; Hollingsworth et al. 2002) that representative operational short-range weather forecasts can track atmospheric observations with an accuracy that lies within current measurement uncertainties. Hence, in observation-rich regions (e.g., continental United States and Europe), the analyses from a modern NWP operational DAS (and, by extension, multidecadal reanalyses) can be regarded as reliable references for identifying errors in GCM short-range forecasts. CAPT therefore is using the latest 6-h reanalyses of the European Centre for Medium-Range Weather Forecasts (ECMWF ERA-40; ECMWF 2002) and of the National Centers for Environmental Prediction (NCEP–DOE R2; Kanamitsu et al. 2002) as the main data for global evaluation of the CAM2 short-range weather forecasts.

NWP reanalyses are not sufficient, however, to evaluate all aspects of a GCM forecast, since they cannot furnish precise checks on physical forcings.

casting is viewed as a context for evaluating and analyzing climate GCM parameterizations, and not as an end in itself.

But will the CAPT methodology enhance the performance of the GCM in climate simulations? In principle, yes: modified parameterizations that reduce systematic forecast errors should also improve the simulation of climate statistics, which are just aggregations of the detailed evolution of the model. In fact, connections between forecast errors and climate errors are often observed in practice. An example of this in CAM2 is an anomalous split in the intertropical convergence zone (ITCZ), which manifests itself very early in the model forecasts, and then grows more pronounced with time (Fig. 2). Some systematic climate errors develop more slowly, however. An example in CAM2 is a cold bias in the tropical tropopause temperature that sets up gradually, presumably because the controlling processes have long natural time scales. It follows that slow climate errors such as these are not as readily amenable to examination by a forecast-based approach.

Thus, once parameterization improvements are provisionally indicated by better short-range forecasts, enhancements in model performance also must be

(Estimates of these forcings from current reanalyses strongly depend on the parameterizations of the analysis GCM.) Thus, ancillary high-frequency local observations such as the ARM field data are indispensable for independent evaluation of GCM parameterizations. Moreover, field observations of state variables can corroborate the NWP analyses in identifying local forecast errors.

ARM field data at 6-h and shorter intervals (in some cases, at intervals comparable to a GCM time step of 30 min) are available at sites in the U.S. Southern Great Plains (SGP), the North Slope of Alaska (NSA), and the tropical west Pacific (TWP; ARM 2002). The most comprehensive high-frequency observations are supplied during sporadic intensive observation periods (IOPs) at the ARM SGP site, such as April 1997 and June–July 1997 (Table 1).

In contrast to methods relying on SCMs or CRMs, CAPT can utilize other datasets that are not as comprehensive as those of ARM, such as coordinated satellite, aircraft, and surface measurements that have been collected during field campaigns. For example, datasets of this type have been centralized for investigations conducted by participants in the Global Energy and Water Cycle Experiment (GEWEX) Cloud System Study (GCSS 2002; Randall et al. 2003). Similar field data at some 30 other sites are being collected during the 2003–04 GEWEX Coordinated Enhanced Observing Period (CEOP; Bosilovich and Lawford 2002).

To make such observations fully relevant for model evaluation, the data should be aggregated to the scale of a GCM grid box. Hence, observations of forcings in different locations at the ARM SGP site have been spatially averaged, and atmospheric state variables also have been subjected to objective variational analysis to ensure overall conservation of heat, moisture, and momentum (Zhang and Lin 1997; Zhang et al. 2001). On the other hand, some GCM pa-

rameterized variables (e.g., cloud properties) need to be translated into quantities that can be compared more readily with observations. For example, CAPT currently is exploring use of the International Satellite Cloud Climatology Project (ISCCP) simulator (Klein and Jakob 1999; Webb et al. 2001) for this purpose.

Initialization procedures. Like many climate models, CAM2 lacks a compatible DAS (although a community data assimilation test bed is under development—see online information at www.cgd.ucar.edu/DART), and so it is necessary to devise simple alternatives to standard NWP initialization procedures. Because of their high accuracy, NWP reanalyses can be used directly to initialize the model. It is also desirable to use atmospheric reanalyses from diverse NWP analysis models, so as to estimate the sensitivity of the CAM2 parameterizations to this difference. Hence, CAPT is using both the ECMWF ERA-40 and NCEP–DOE R2 reanalyses to initialize the CAM2 model. This entails a three-dimen-
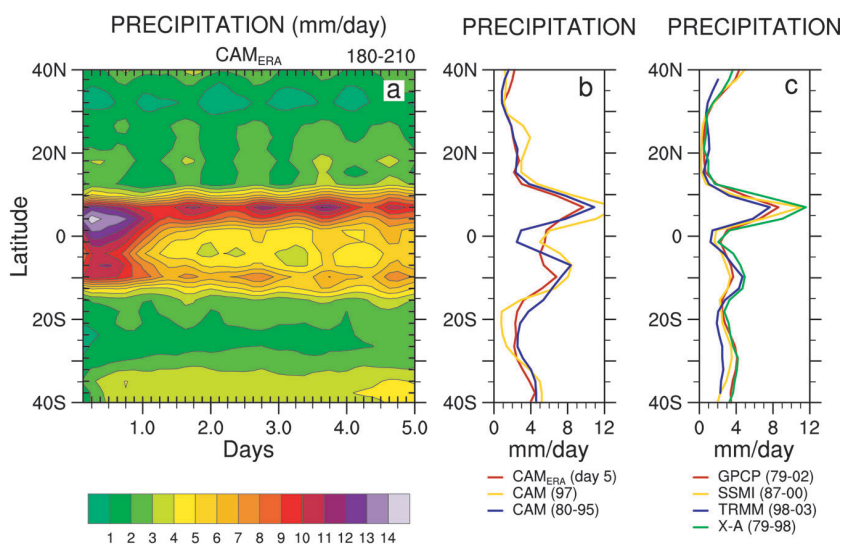
FIG. 2. Depiction of tropical Pacific precipitation in the CAM2 model. (a) A contour plot of the CAM2 mean 5-day forecast of 3-h accumulations of longitudinally averaged (between 180° and 210°E) tropical Pacific precipitation for Jun 1997. (The mean is based on thirty 5-day forecasts, initiated at 0000 UTC each day.) Here, the precipitation pattern indicates that a split ITCZ develops early in the mean forecast. (b) It is seen that the CAM2 split ITCZ grows more pronounced with time, as evinced by the 180°–210°E zonal average of 24-h precipitation in the mean 5-day forecast and in the monthly mean for Jun 1997, as well as in a 1980–95 Jun climatology, both from AMIP simulations. Note that the southern branch of the zonal-average split ITCZ in CAM2 precipitation in (b) is anomalously intense compared to various estimates of observed zonal-average precipitation climatologies, as shown in (c) for the Global Precipitation Climatology Project (GPCP, for the period 1979–2002), the Special Sensor Microwave Imager (SSM/I, for 1987–2000), the Tropical Rainfall Measuring Mission (TRMM, for 1998–2003), and the Xie–Arkin (X–A, for 1979–98) datasets. In all cases, precipitation is given in units of mm day[-1].

TABLE 1. Available observations during ARM IOPs at the SGP site. Note: NOAA denotes the National Oceanic and Atmospheric Administration, GOES the Geostationary Operational Environment Satellite, SIRS the Solar Infrared Radiation Station, EBBR the Energy Budget Bowen Ratio, MWR the Microwave Radiometer, while *U, V, T,* and RH represent horizontal winds, temperature, and relative humidity, respectively. LWT and SWT are the top-of-atmosphere longwave and shortwave radiative fluxes. CLDTOT, CLDHGH, CLDMED, and CLDLOW are the total, high-, middle-, and low-level cloud amounts, respectively. LWS and SWS are the surface longwave and shortwave radiative fluxes; LH is the surface latent heat flux, SH the surface sensible heat flux, PW the column precipitable water, CLW the column cloud liquid water, and PREC the surface precipitation; and *U*s, *V*s, *P*s, *T*s, and RHs are the surface *u* and *v* wind components, pressure, air temperature, and relative humidity, respectively.

Instruments	Sampling interval	Measured fields
ARM radiosondes	3 h	*U, V, T*, RH
NOAA wind profilers	1 h	*U, V*
GOES	30 min	LWT, SWT, CLDTOT, CLDHGH, CLDMED, CLDLOW
SIRS	1 min	LWS, SWS
EBBR	30 min	Surface LH, SH
MWR	5 min	Column PW and CLW
Surface mesonet stations	5–30 min	Surface PREC, *T*s, *U*s, *V*s, *P*s, RHs
Cloud radar, micropulse lidar	10 s–1 min	Cloud occurrences and properties

sional mapping of finer-resolution reanalysis data to the coarser (spectral T42/L26) CAM2 resolution.

Operational NWP centers routinely map state variables from high-resolution operational analyses to lower resolutions in order to provide initial conditions for their ensemble forecast systems, and for developmental research forecasts. These mapping procedures also have been applied when exchanging analyses between operational centers, for example, to investigate the effect of the initial conditions on specific forecasts (e.g., Harrison et al. 1999). CAPT has successfully adopted the relevant NWP algorithms (White 2001), even though these normally are not applied for resolutions as coarse as those of typical climate GCMs.

As noted previously, the objective is to obtain a good estimate of the atmosphere–land initial conditions, rather than a state that gives the best forecast. Although operational centers might still include a mass-momentum balancing phase ("initialization," in NWP parlance) after the resolution change, forecasts with CAM2 initialized from both the ECMWF ERA-40 and NCEP–DOE R2 reanalyses are found to be relatively

noise free. Should initialization noise arise in specific cases, a temporal digital filter could easily be included (Lynch and Huang 1992; Polavarapu et al. 2000).

For a prototype implementation, initial values of the parameterized variables that are predicted based on previous values (e.g., cloud water in CAM2) are obtained via spinup procedures that are described below in conjunction with the land initialization. These methods are presently adequate because the time scales for the adjustment of the parameterized variables are relatively fast, and the model errors are currently very large. As a future refinement, CAPT will need to develop mapping procedures for these prognostic parameterized variables as well.

Initialization of the land is particularly problematic because it is difficult to map discrete and discontinuous land variables between different resolutions, especially when there may be dissimilar definitions of soil types and variables in the systems that are involved in this mapping. Thus far, CAPT has applied two procedures to spin up land and atmospheric parameterized variables. Both allow the land model

(and parameterizations) to interact with and respond to the forcing from the atmospheric model, which is constrained to follow the evolution of the observed atmosphere. CAPT designates these two methods as "forecast/analysis" and "nudging."

The forecast/analysis method periodically updates (e.g., at 6-h intervals) the atmospheric state variables with the interpolated analyses, and lets the coupled land–atmosphere system evolve until the next update time. This is akin to the current 6-h update cycle of an NWP DAS (such as that used for the NCEP/NCAR reanalysis—see Kalnay et al. 1996), except that here the atmospheric assimilation phase is replaced by the mapping of a high-resolution reanalysis to the climate model grid.

The nudging method involves the addition of terms to the atmospheric equations to relax predicted state variables toward the reanalysis at a specified (e.g., 6 h) time scale. This procedure has been used, for example, to generate a smooth start for NWP model forecasts (Hoke and Anthes 1976).

In order to indicate the minimum period needed for either of these spinup methods to converge to the best state that it can provide, "perfect model" experiments (i.e., using the outputs of the CAM2 as input to the forecast/analysis initialization) have been performed. In these simulations, the CAM2 soil moisture, for example, spins up to "correct" values in a few months, except when snow is present. Hence, these simple spinup methods might be generally applicable in tropical regions, as well as in warm-season observational periods at the ARM SGP site and other midlatitude locations. Spinup methods appropriate for snow-covered regions will require further development.

In mapping from analyses to the climate GCM, there is some (probably small) risk of obtaining a false response to a parameterization change. However, in order to produce an initial atmospheric state that is model-consistent, it would be necessary to use a DAS that is based on the GCM, thereby excluding from consideration many climate models that lack this capability. Moreover, to produce the best estimate of the initial state, the DAS would need to be run at higher resolution than is typical for climate simulation, since even the large scales in NWP analyses are improved with enhanced resolution.

Model forecasts. The current CAPT practice is to generate 5-day (0–120 h) GCM forecasts for each day during the time period of interest (e.g., an ARM IOP), and to archive the forecast data at intervals that match the sampling of the field observations (e.g., at 3-h in-

tervals for comparison with ARM variational analysis data). For each forecast, the model atmosphere is initialized by applying either the nudging or forecast/analysis methods described previously.

Then the mean (at fixed elapsed times) of a sequence of forecasts initialized on different days is computed. This mean forecast may be calculated from model predictions that are initiated on consecutive days, or alternatively from forecasts that are stratified according to similar initial conditions, so as to assess the sensitivity of the model parameterizations to particular synoptic or seasonal conditions (Jakob 2003).

In order to estimate the GCM systematic forecast error, the difference between the mean forecast and corresponding evaluation data is examined. Both the magnitude and pattern of this systematic error are of diagnostic value, and these can be quantified using standard NWP metrics defined by the World Meteorological Organization (WMO 1999). For example, the mean bias and root-mean-square (rms) statistics provide information on error amplitudes, while anomaly correlations supply error-pattern information.

RESULTS. Here, selected results of applying the CAPT protocol to the CAM2 model are shown, so as to illustrate the concepts discussed in previous sections.

First, in order to verify that simple initialization procedures are able to produce a large-scale dynamical state that is close to that of the verifying analyses, the skill of the CAM2 forecasts of the 500-hPa height field is evaluated, following established guidelines (WMO 1999). For example, the mean anomaly correlation (AC) of these forecasts (a commonly accepted measure of forecast skill) can be computed, where the verification anomalies are defined by the departures from a 30-yr monthly mean climatology, such as that of ECMWF ERA-40 reanalysis for the period 1970–99. (The AC calculations are found to be insensitive to the choice of climatology.)

Figure 3 shows the AC decay (mean AC as a function of forecast day) of the CAM2 model forecasts, initialized from both the ECMWF ERA-40 and the NCEP–DOE R2 reanalyses, during the April and June–July 1997 ARM IOPs. These are compared with the AC decay of analogous forecasts from the models that generated the ECMWF ERA-40 and NCEP–DOE R2 reanalyses, where values less than 0.6 indicate an absence of useful forecast skill.

In general, the CAM2 forecasts of hemispheric-scale 500-hPa heights are seen to be surprisingly "competitive" with those from the two NWP models. In particular, the AC decay of the CAM2 in the first two forecast days is small, implying that its dynamical state

remains close to those of the reanalyses during the early part of the forecast. (Inspection of weather maps that include the ARM SGP site during these periods indicates that the CAM2 forecasts of 500-hPa heights also are skillful at synoptic scales.) The skill scores of the CAM2 forecasts show greater separation in the Southern Hemisphere (Figs. 3b and 3d), indicating a generally stronger dependence on initial conditions, and a somewhat higher accuracy of ECMWF ERA-40 relative to the NCEP–DOE R2. As expected, the AC decay of all forecasts is more rapid during boreal summer (Fig. 3c) when midlatitude synoptic control is weaker and forecast skill is more strongly influenced by physical processes. The comparatively low decay

of the ECMWF ERA-40 model skill in this season thus attests to the quality of its physical parameterizations.

To place these results in perspective, Table 2 compares the CAM2 scores for the Northern Hemisphere midlatitudes at day 5 with two operational NWP models for the April 1997 IOP period and for subsequent Aprils in the years 2000 and 2003 for the operational models. (CAM2 forecasts have not yet been made for these years.) The higher accuracy of the more recent ECMWF ERA-40 implied by Fig. 3 reflects, in part, the NWP technological advances made since 1997, as suggested by the skill trend of the operational models (Simmons and Hollingsworth 2001). Moreover, the fact that the CAM2 5-day AC in Table 2 is higher than representative operational NWP models in 1997 demonstrates that a coarse-resolution climate GCM can make skillful forecasts of the large-scale synoptic flow when initialized with an accurate analysis.

However, it is found that the relatively high skill of CAM2 forecasts of the 500-hPa heights does not generally carry over to predictions of large-scale atmospheric moisture, which is tied more directly to the model's physical parameterizations. This model shortcoming is found locally as well, for example, in evaluating a sequence of CAM2 daily forecasts of atmospheric relative humidity at the ARM SGP site during the period 19–25 June 1997 (Fig. 4). Here it is seen that the temporal variation of the vertical profile of relative humidity obtained from the ARM measurements (Fig. 4a) and from the ECMWF ERA-40 reanalysis (Fig. 4b) are quite similar. Relative to these evaluation data during most days of this period, the CAM2 forecasts a lower troposphere that is too dry and an upper troposphere that is too moist.
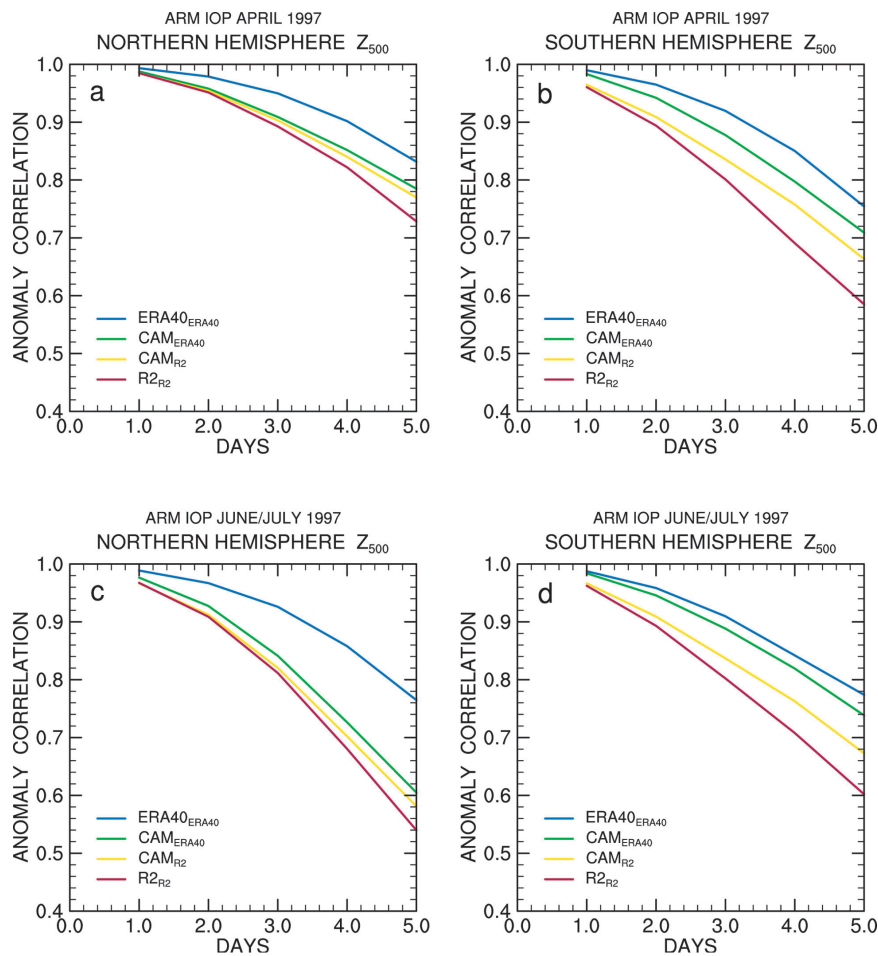
FIG. 3. Mean AC for a sequence of forecasts of 500-hPa heights made with three GCMs as a function of forecast day during (a), (b) the Apr 1997 ARM IOP and (c), (d) Jun–Jul 1997 IOP. In all cases, the AC is interpolated to a common 2.5° global grid and spatially averaged (with cosine-latitude weighting) over extra-tropical portions of the Northern and Southern Hemispheres (20°–90°N and 20°–90°S, respectively). Results are shown for the ECMWF ERA-40 reanalysis model initialized with its own analyses (blue); the NCEP–DOE R2 reanalysis model initialized with its own analyses (red); the CAM2 model initialized with ECMWF ERA-40 (green) and NCEP–DOE R2 reanalyses (yellow).

TABLE 2. Anomaly correlation at day 5 of CAM2 forecasts of Northern Hemisphere 500-hPa heights for the Apr 1997 ARM IOP period compared with that of forecasts from the NCEP and ECMWF operational models. CAM.ERA-40 denotes the CAM2 model initialized with ECMWF ERA-40 reanalysis; NCEP ops and ECMWF ops are the respective operational model scores, which also are shown for comparison in Apr 2000 and 2003. (CAM2 forecasts are not available in Apr 2000 and 2003.)

Model	Apr 1997	Apr 2000	Apr 2003
CAM.ERA-40	0.83	NA	NA
NCEP ops	0.74	0.73	0.84
ECMWF ops	0.78	0.84	0.89

The anomalous CAM2 relative humidity profile on 19–22 June is indicative of a model systematic error, as revealed both by the mean 5-day model forecast relative to ARM observations during June–July 1997 (Fig. 5a) as well as by the CAM2 June–July climatology relative to that of the ECMWF ERA-40 and the NCEP–DOE R2 reanalyses (Fig. 5b). Given the relatively skill-

ful model forecast of large-scale dynamics during this period (Fig. 3c), these results imply that there are both forecast and climate systematic errors associated with the CAM2 moist physics parameterizations.

It is just such model deficiencies that are appropriate to study further in the CAPT framework. For instance, the overly dry CAM2 lower troposphere
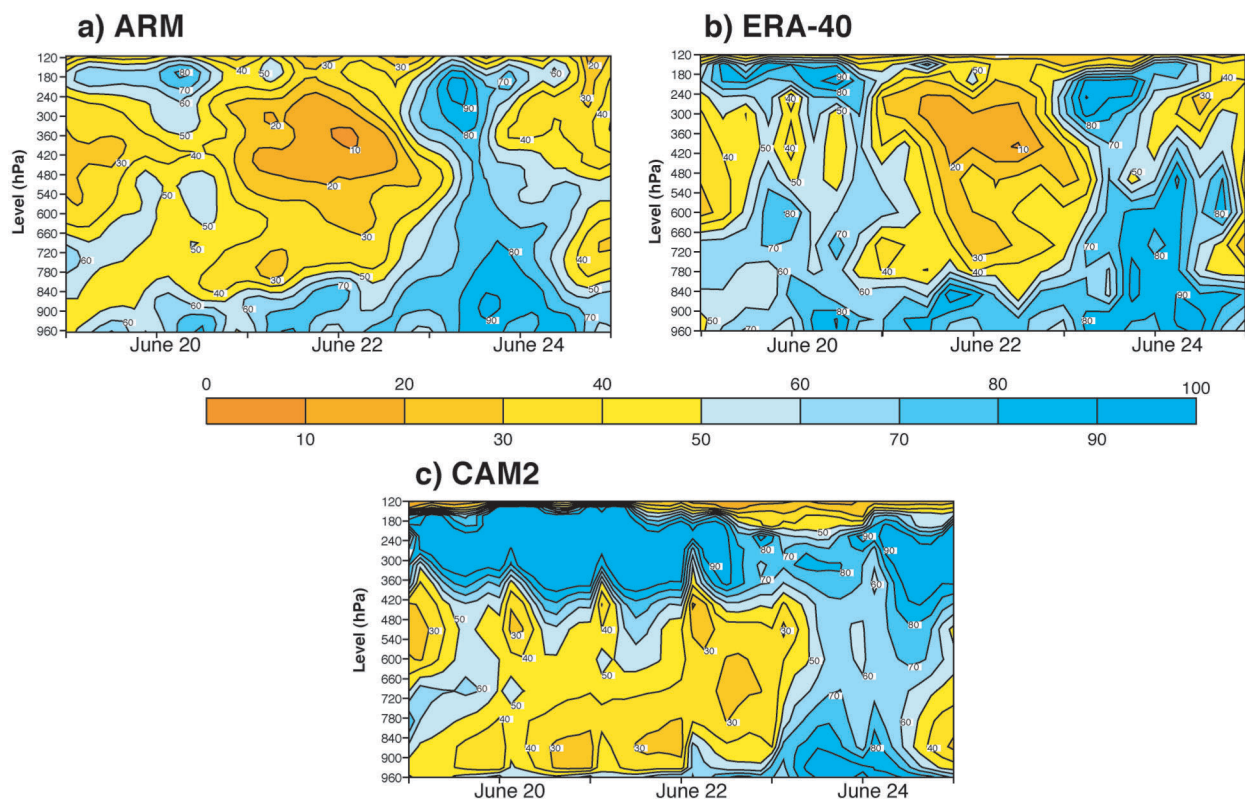
FIG. 4. Plots of the vertical profile of atmospheric relative humidity (%) at the ARM SGP site are shown at 3-h intervals for the period 19–25 Jun 1997, as obtained from (a) ARM observations, (b) the ECMWF ERA-40 reanalysis, and (c) a sequence of CAM2 forecasts that are initialized at 0000 UTC each day and valid for the period 0300–2400 UTC (but with the 0000 UTC value shown for 19 Jun supplied by the 2400 UTC forecast for 18 Jun). Note the apparent diurnal cycle in the relative humidity profile in (c) is, in actuality, evidence of the rapid departure of the CAM2 forecasts from a realistic humidity profile after their initialization at 0000 UTC each day.
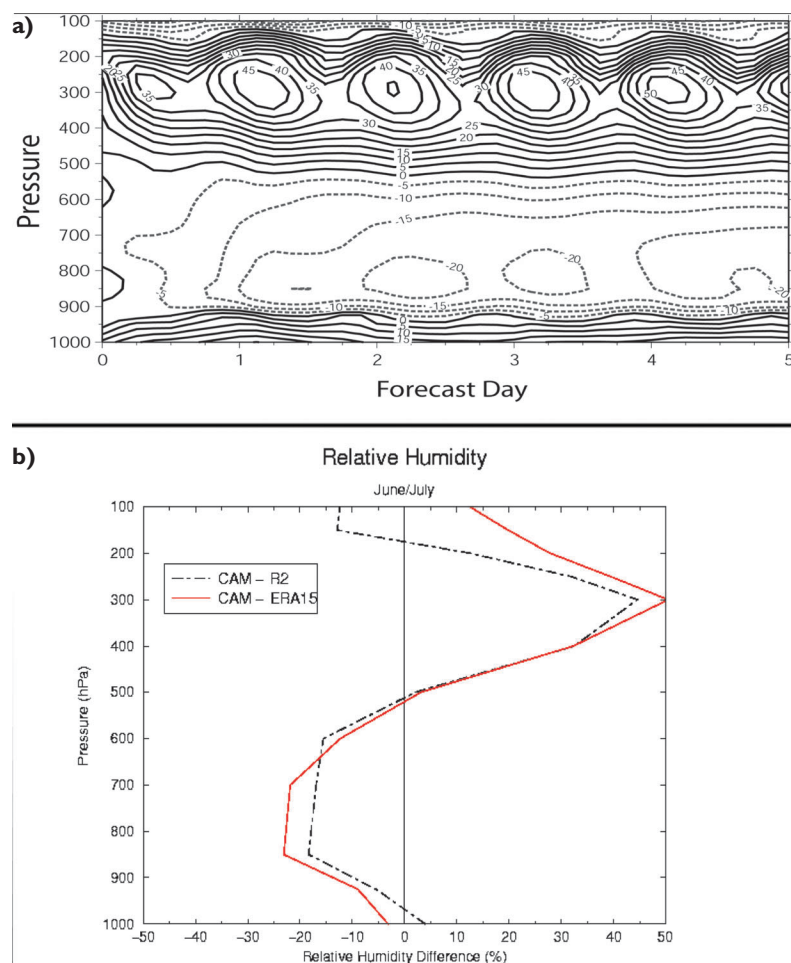
during June–July 1997 at the ARM SGP site is consistent with the model's propensity to rain out moisture nearly every day, rather than in the episodic bursts that are observed (Fig. 6b). In contrast, the agreement between CAM2 precipitation forecasts and observations is generally much better during the April 1997 IOP (Fig. 6a), when large-scale advective forcing is a more significant contributor to the column moisture balance (analysis not shown).

Fig. 5. Systematic forecast and climate errors in CAM2 predictions of the vertical profile of atmospheric relative humidity at the ARM SGP site relative to ARM observations (%) are displayed. (a) The evolution of the mean difference between CAM2 5-day forecasts of the relative humidity profile (initialized at 0000 UTC each day of the Jun–Jul 1997 IOP) and the corresponding ARM observations. Note the rapid growth of CAM2 relative humidity errors early in the mean 5-day forecast and their subsequent diurnal variation, especially in the upper troposphere. (b) Differences between a 10-year Jun–Jul climatology of the CAM2 relative humidity profile (generated in an AMIP simulation) and corresponding climatologies obtained from the ECMWF ERA-40 (solid line) and NCEP–DOE R2 reanalyses (dashed line). Note similarities in the vertical distribution of the CAM2 systematic errors (overly dry lower troposphere above the boundary layer, and overly moist upper troposphere) at both forecast and climate time scales.

This seasonal sensitivity implies there may be deficiencies in the CAM2 parameterization of convective precipitation (although this is certainly not the only possible source of error). In particular, the characteristics of the CAM2 precipitation displayed in Fig. 6b are reminiscent of problems previously identified in the triggering mechanism of the model's Zhang–McFarlane (1995) deep convection scheme when implemented in an SCM (Xie and Zhang 2000). These deficiencies were alleviated by replacing the standard trigger that is based on positive convective available potential energy (CAPE) with one based on the rate of dynamic CAPE (DCAPE) generation by large-scale advective tendencies of temperature and moisture.

The DCAPE convective triggering mechanism has been implemented in the CAM2, and the effects on its forecasts analyzed (Xie et al. 2004). In the modified model, CAPE can accumulate before convection occurs, and thus stronger but less frequent precipitation events are produced, yielding somewhat better agreement with ARM observations (Fig. 6c). Xie et al. (2004) show that the DCAPE convective trigger also reduces systematic errors in the CAM2 forecasts of atmospheric humidity (e.g., Fig. 5a) and other variables.

Further evaluation of the performance of the DCAPE convective trigger in an AMIP climate simulation is currently in progress. This, as well as analysis of other facets (e.g., clouds, surface fluxes, and components of energy and moisture budgets) of the forecasts and the climate of the standard CAM2 model, will be presented in forthcoming journal articles (Boyle et al. 2004, manuscript submitted to *J. Geophys Res.*; Williamson et al. 2004, manuscript submitted to *J. Geophys Res.*).

SUMMARY. CAPT is motivated by the experience of GCM developers that it is very difficult to identify particular parameterization defi-
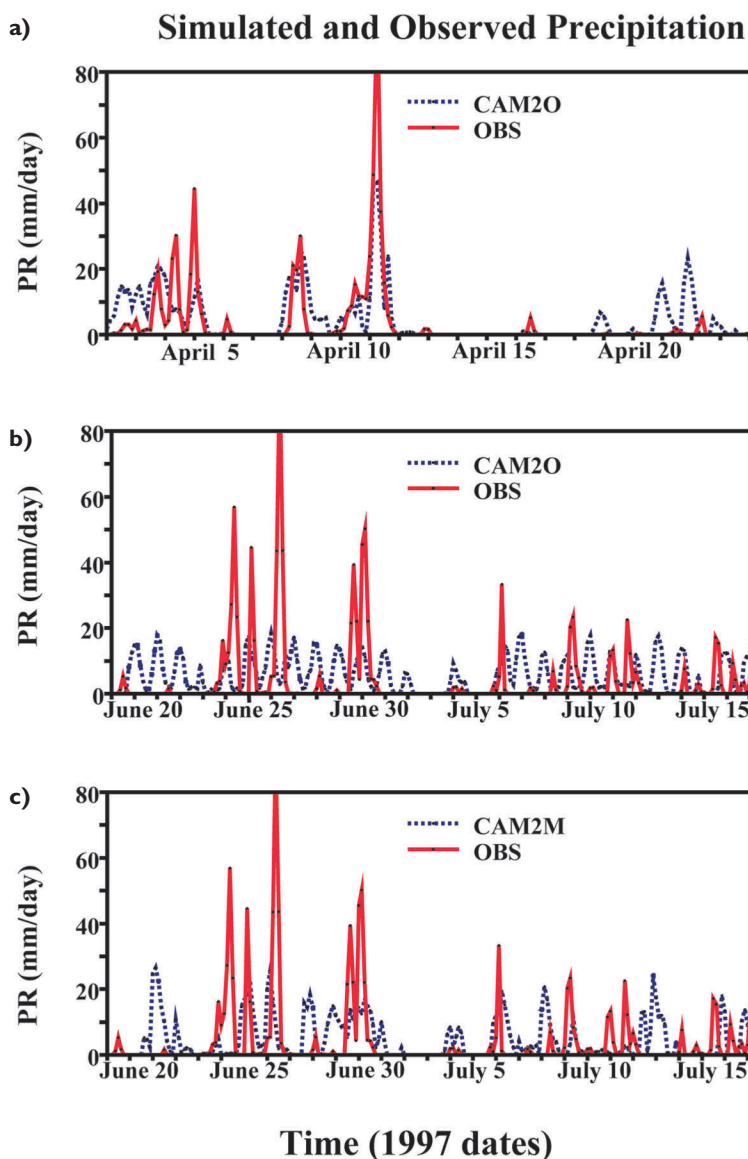
ciencies solely by analyzing a model's climate statistics, which reflect compensating errors resulting from the nonlinear interactions of many different processes. The CAPT premise is that studying climate GCMs in a weather-forecasting framework is generally a more effective way to identify parameterization deficiencies and to gain insights on how these might be ameliorated.

If a modified parameterization is able to reduce systematic forecast errors (defined relative to high-duality observations and NWP analyses), it then can be regarded as more physically realistic than its predecessor. Whether the systematic climate errors of the model also decrease must be demonstrated in practice, however. Slowly developing systematic climate errors, for example, probably will remain resistant to significant reduction by such a forecast-based method.

Thus, CAPT is not a panacea for improving climate GCM parameterizations at all time scales, but just one choice from a "tool kit" that may also include, for example, SCMs, CRMs, and simplified GCMs. Nonetheless, we expect that insights obtained from adopting this NWP-inspired methodology will contribute significantly to the general improvement of GCM climate simulation.

FIG. 6. Comparisons of 3-h accumulations of precipitation (mm day^{-1}) at the ARM SGP site, as observed (solid line), and as forecast by the CAM2 (dashed line), where the large-scale atmospheric state of the model was reinitialized each day from ECMWF ERA-40 reanalysis. The comparison is for (a) the Apr 1997 IOP, and (b) the Jun–Jul 1997 IOP; in both cases, the forecasts are made with the standard version of CAM2 (denoted as CAM2O) that employed the Zhang–McFarlane parameterization of deep convection. (c) Forecasts of Jun–Jul 1997 precipitation made with a version of the CAM2 (denoted as CAM2M) that included a modified (DCAPE) convective triggering mechanism.

REFERENCES

Ackerman, T. P., and G. M. Stokes, 2003: The Atmospheric Radiation Measurement Program. *Phys. Today,* **56,** 38–44.

ARM, cited 2002: ARM cloud and radiation test bed sites. [Available online at www.arm.gov/sites/.]

Betts, A. K., and M. J. Miller, 1986: A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX, ATEX, and arctic airmass data sets. *Quart. J. Roy. Meteor. Soc.,* **112,** 693–709.

Bosilovich, M. G., and R. Lawford, 2002: Coordinated Enhanced Observing Period (CEOP) international workshop. *Bull. Amer. Meteor. Soc.,* **83,** 1495–1500.

Boyle, J., D. Williamson, R. Cederwall, M. Fiorino, J. Hnilo, J. Olson, T. Phillips, G. Potter, and S. Xie, 2004: Diagnosis of CAM2 in NWP configuration. J. *Geophys. Res.,* submitted.

Collins, W. D., and Coauthors, cited 2003: Description of the NCAR Community Atmosphere Model (CAM2). [Available online at www.ccsm.ucar. edu/models/atm-cam/docs/description/index.html.]

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

ECMWF, cited 2002: ECMWF re-analysis ERA. [Available online at www.ecmwf.int/research/era.]

Gates, W. L., 1992: AMIP: The Atmospheric Model Intercomparison Project. *Bull. Amer. Meteor. Soc.,* **73,** 1962–1970.

——, and Coauthors, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP). *Bull. Amer. Meteor. Soc.,* **80,** 29–55.

GCSS, cited 2002: GEWEX Cloud System Study data integration for model evaluation. [Available online at http://gcss-dime.giss.nasa.gov/.]

Harrison, M., T. N. Palmer, D. S. Richardson, and R. Buizza, 1999: Analysis and model dependencies in medium-range ensembles: Two transplant case studies. *Quart. J. Roy. Meteor. Soc.,* **125,** 2487–2515.

Hoke, J. E., and R. A. Anthes, 1976: The initialization of numerical models by a dynamic-initialization technique. *Mon. Wea. Rev.,* **104,** 1551–1556.

Hollingsworth, A., M. Capaldo, and A. J. Simmons, 1998: The scientific and technical foundation of the ECMWF strategy, 1999–2008. European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, UK, 36 pp.

——, P. Viterbo, and A. J. Simmons, 2002: The relevance of numerical weather prediction for forecasting natural hazards and for monitoring the global environment. ECMWF Tech. Memo. 361, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, United Kingdom, 29 pp. [Available online at www.ecmwf.int/publications/ library/ecpublications/_pdf/tm361.pdf.]

Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson, Eds., 2001: *Climate Change 2001: The Scientific Basis.* Cambridge University Press, 881 pp.

Jakob, C., 2003: An improved strategy for the evaluation of cloud parameterizations in GCMs. *Bull. Amer. Meteor. Soc.,* **84,** 1387–1401.

Jasper, J. D., and P. J. Meighen, Eds., 2000: Model systematic errors. *Extended Abstracts, 12th Annual BMRC Modelling Workshop,* Melbourne, Victoria, Australia, WCRP/WGNE, BMRC Research Rep. 80, 288 pp.

Jeuken, A. B. M., P. C. Siegmund, and L. C. Heijboer, 1996: On the potential of assimilating meteorological analyses in a global climate model for the purpose of model validation. *J. Geophys. Res.,* **101** (D12), 16 939–16 950.

Kaas, E., A. Guldberg, W. May, and M. Deque, 1999: Using tendency errors to tune the parameterization of unresolved dynamical scale interactions in atmospheric general circulation models. *Tellus,* **51A,** 612–629.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation, and Predictability.* Cambridge University Press, 341 pp.

——, and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

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. Amer. Meteor. Soc.,* **83,** 1631–1643.

Kiehl, J. T., and P. R. Gent, 2004: The Community Climate System Model, version 2. J. *Climate,* in press.

Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.,* **82,** 247–267.

Klein, S. A., and C. Jakob, 1999: Validation and sensitivities of frontal clouds simulated by the ECMWF model. *Mon. Wea. Rev.,* **127,** 2514–2531.

Krueger, S. K., 1988: Numerical simulation of tropical cumulus clouds and their interaction with the subcloud layer. *J. Atmos. Sci.,* **45,** 2221–2250.

——, G. T. McLean, and Q. Fu, 1995: Numerical simulation of the stratus-to-cumulus transition in the subtropical marine boundary layer. Part II: Boundary-layer circulation. *J. Atmos. Sci.,* **52,** 2851–2868.

Lynch, P., and X. Y. Huang, 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.,* **120,** 1019–1034.

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.

Meehl, G. A., G. J. Boer, C. Covey, M. Latif, and R. J. Stouffer, 2000: The Coupled Model Intercom-

parison Project (CMIP). *Bull. Amer. Meteor. Soc.,* **81,** 313–318.

Miller, M., 1999: Discussion of the 'Transpose AMIP' project. Report of the Fourteenth Session of the CAS/JSC Working Group on Numerical Experimentation, R. Newsom, Ed., CAS/JSC WGNE Rep. 14, WMO Tech. Doc. 964, 7–8.

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.

Morcrette, J.-J., 2002: Assessment of the ECMWF model cloudiness and surface radiation fields at the ARM SGP site. *Mon. Wea. Rev.,* **130,** 257–277.

Palmer, T. N., 1999: Predicting uncertainty in forecasts of weather and climate. ECMWF Tech. Memo. 294, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, United Kingdom, 64 pp.

Polavarapu, S., M. Tanguay, and L. Fillion, 2000: Four-dimensional variational data assimilation with digital filter initialization. *Mon. Wea. Rev.,* **128,** 2491–2510.

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. Climate,* **9,** 1683–1697.

——, and Coauthors, 2003: Confronting models with data: The GEWEX Cloud Systems Study. *Bull. Amer. Meteor. Soc.,* **84,** 455–469.

Schubert, S., and Y. Chang, 1996: An objective method for inferring sources of model error. *Mon. Wea. Rev.,* **124,** 325–340.

Simmons, A., and A. Hollingsworth, 2001: Some aspects of the improvement in skill of numerical weather prediction. ECMWF Tech. Memo. 342, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, United Kingdom, 35 pp. [Available online at www.ecmwf.int/publications/library/ecpublications/_pdf/tm342.pdf.]

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.

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. *Climate Dyn.,* **17,** 905–922.

WGNE, 1988: Workshop on systematic errors in models of the atmosphere. CAS/JSC Working Group Rep. 12, WMO Tech. Doc. 273, Geneva, Switzerland, 382 pp.

White, P. W., Eds., cited 2001: FULL-POS postprocessing and interpolation. IFS Documentation Part VI: Technical and Computational Procedures (CY23R4). [Available online at www.ecmwf.int/research/ifsdocs_old/TECHNICAL/index.html.]

Williamson, D. L., and R. Daley, 1986: The effect of model climate drift on medium range weather prediction. *Extended Abstracts, Int. Symp. on Short- and Medium-Range Numerical Weather Prediction,* Tokyo, Japan, WMO Tech. Doc. 114, 553–555.

——, J. Boyle, R. Cederwall, M. Fiorino, J. Hnilo, J. Olson, T. Phillips, G. Potter, and S.C. Xie, 2004: Moisture and temperature balances at the ARM Southern Great Plains Site in forecasts with the CAM2. *J. Geophys. Res.*, submitted.

WMO, 1999: Commission for basic systems abridged final report with resolutions and recommendations. Rep. WMO 893, Geneva, Switzerland, 184 pp. [Available online at www.wmo.ch/web/www/reports/CBS-Ext98.pdf.]

Xie, S. C., and M. H. Zhang, 2000: Impact of the convective triggering function on single-column model simulations. *J. Geophys. Res.,* **105,** 14 983–14 996.

——, and Coauthors, 2002: Intercomparison and evaluation of cumulus parameterizations under summertime midlatitude continental conditions. *Quart. J. Roy. Meteor. Soc.,* **128,** 1095–1135.

——, M. Zhang, J. S. Boyle, R. T. Cederwall, G. L. Potter, and W. Lin, 2004: Impact of a revised convective triggering mechanism on Community Atmosphere Model, Version 2, simulations: Results from short-range weather forecasts. *J. Geophys. Res.*, **109** (D14102), doi:10.1029/2004JD004692.

Xu, K.-M., and Coauthors, 2002: An intercomparison of cloud-resolving models with the ARM measurement summer 1997 IOP data. *Quart. J. Roy. Meteor. Soc.,* **128,** 593–624.

Zhang, G. J., and N. A. McFarlane, 1995: Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centre general circulation model. *Atmos.–Ocean,* **33,** 407–446.

Zhang, M. H., and J. L. Lin, 1997: Constrained variational analysis of sounding data based on column-integrated budgets of mass, heat, moisture, and momentum: Approach and application to ARM measurements. *J. Atmos. Sci.,* **54,** 1503–1524.

——, ——, R. T. Cederwall, J. J. Yio, and S. C. Xie, 2001: Objective analysis of ARM IOP data: Method and sensitivity. *Mon. Wea. Rev.,* **129,** 295–311.