10.1

THE CCPP-ARM PARAMETERIZATION TESTBED (CAPT): EVALUATING CLIMATE MODELS IN A WEATHER FORECASTING FRAMEWORK

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[#]

[#] Lawrence Livermore National Laboratory, Livermore, California

[+] National Center for Atmospheric Research, Boulder, Colorado

1. INTRODUCTION

Climate simulations performed with general circulation models (GCMs) are widely viewed as the principal scientific basis for proposed policies to address potential future global climate change. 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.

To further enhance GCM performance, ongoing interdependent efforts are needed:

1) to compare GCM simulations with observations over a range of time scales in order to diagnose the details of the associated model systematic errors; and

2) to reduce these systematic errors by improving the representation of key physical processes, and thereby increase the accuracy of GCM simulations relative to available observations.

In practice, the reduction of GCM systematic errors entails both increases in the resolution at which the model state variables (e.g. for atmospheric GCMs, the pressure, temperature, moisture, and wind fields) are predicted, and improvements in the parameterizations of unresolved sub-grid processes (e.g. radiation, clouds, convection, precipitation microphysics, turbulent fluxes and diffusion). In GCMs designed for climate simulations (hereafter, "climate GCMs"), parameterization development is especially important for correct representation of relevant processes.

The deciding factor in choosing a new parameterization for a climate GCM is whether its inclusion brings the simulated climate statistics into closer agreement with those observed. However, there are inherent limitations to evaluating GCM parameterizations exclusively in climate-simulation mode. Because the GCM climate state reflects compensating errors in the simulation of many nonlinear dynamical and physical processes, it is very difficult to unravel deficiencies in specific parameterizations. In this context also, an unrealistic large-scale climate state is driving the parameterizations, so that it is difficult to evaluate them objectively.

Operational numerical weather prediction (NWP) centers follow a different methodology for developing parameterizations in GCMs designed for weather forecasting. The state variables of the forecast GCM are first initialized by a data assimilation system (DAS) which usually is built around the GCM itself. After ingestion of all available observations (e.g. surface, radiosonde, 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 (Kalnay 2003). Given an accurate analysis, the model state should remain close to "truth" in the early period of the forecasts, so that the systematic forecast error can be attributed largely to parameterization deficiencies.

The systematic forecast error is estimated from differences between the mean of a sequence of short-range (~ five-day) forecasts and evaluation data that include NWP analyses as well as observations of parameterized variables (e.g. radiative and turbulent fluxes, cloud properties, precipitation, etc.) that are not assimilated by the DAS. Modified parameterizations are similarly evaluated in short-range forecasts to determine whether they reduce the model's high-frequency systematic errors. If that is the case, the new parameterizations usually are evaluated in model integrations beyond the deterministic forecast range of ~ 15 days to determine whether they also reduce low-frequency systematic errors.

The Working Group on Numerical Experimentation of the World Climate Research Programme (WCRP/WGNE) has advocated an approach similar to this NWP methodology for evaluation of parameterizations in climate GCMs. Practical support for this effort now is being provided by a joint initiative of the U.S. Department of Energy Climate Change Prediction Program (CCPP) and the Atmospheric Radiation Measurement (ARM) Program: the CCPP-ARM Parameterization Testbed (CAPT).

**Corresponding author address:* Thomas J. Phillips, Mail Code L-103, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, California 94551
e-mail: phillips14@llnl.gov

2. THE CAPT DIAGNOSTIC PROTOCOL

CAPT is promoting a diagnostic approach which is new for climate GCMs that are not associated with operational forecast centers: the diagnosis of short-range weather forecasts made with a climate GCM that is initialized realistically.

The CAPT premise is that, as long as the evolving dynamical state of the GCM forecast remains close to that of the verifying NWP weather analyses, the systematic errors in the forecast of atmospheric state variables are predominantly due to deficiencies in the GCM parameterizations. Under these circumstances, it is also appropriate to compare the parameterized variables of the GCM with available high-frequency observations collected under the same dynamical conditions, and to modify the relevant parameterizations so as to better match such observations. Finally, if the modified parameterizations are able to reduce the systematic forecast errors, it is probable that the GCM climate simulation will improve as well.

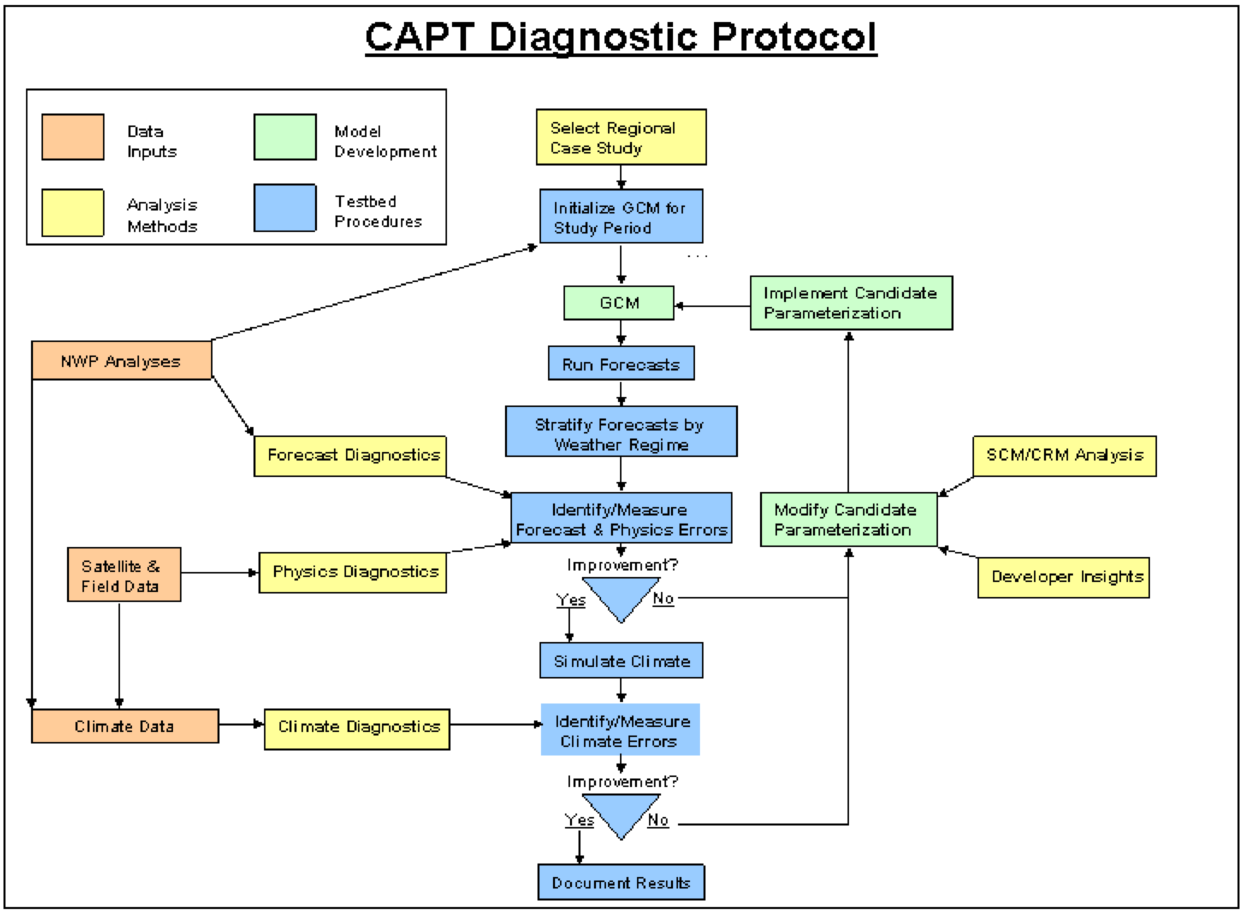

Figure 1: Flow diagram of the CAPT diagnostic protocol.

The basic elements of the CAPT diagnostic protocol are illustrated in Figure 1. First the climate GCM is initialized so that its atmospheric state approximates synoptic conditions at a specified time, while also being dynamically balanced (see details in Section 3). 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 reanalyses and unassimilated observations of parameterized variables such as provided by ARM data (Ackerman and Stokes 2003). Differences between the model predictions and these evaluation data are diagnosed in order to learn more about the performance of the model parameterizations, and to suggest needed changes. The efficacy of modifying the parameterizations then also can be evaluated in a short-range forecasting framework.

However, the overall 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 model may remain inferior to those of a fine-resolution NWP model, relative decreases in systematic error are still indicative of improved parameterizations. Moreover, the rich variety of weather phenomena allows the model parameterizations to be evaluated over a wide range of conditions, and at much less computational expense than is required in climate-simulation mode. In CAPT, therefore, weather forecasting is viewed as a context for learning more about climate GCM parameterizetions, 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 aggregates of the detailed evolution of the model. This improvement of the climate statistics must be demonstrated in practice, however. We acknowledge, for example, that slowly developing GCM systematic errors may not be especially amenable to resolution by the CAPT methodology. Thus, parameterization evaluation in climate simulations is a necessary element of the CAPT diagnostic protocol (Figure 1). Even when new parameterizations improve model performance at short time scales, further adjustments at climate scales may be necessary. For example, some "tuning" of the free parameters of a new model scheme may be needed in order to achieve radiative balance in a climate simulation.

3. TECHNICAL DETAILS

Here we elaborate on several aspects of the CAPT diagnostic protocol, as it has been applied thus far to version 2.0 of the Community Atmosphere Model (CAM2) which was developed under the auspices of the National Center for Atmospheric Research (NCAR) (Collins et al. 2003).

3.1 Evaluation Data

The efficacy of the CAPT methodology depends crucially on the accuracy of the verifying NWP analyses of the weather. Current NWP analyses are very good approximations of the actual atmospheric state, as shown by recent findings (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 U.S. and Europe), the analyses from modern NWP operational DAS's (and, by extension, multi-decadal reanalyses) can be regarded as reliable references for identifying errors in GCM short-range forecasts. We therefore are using the latest high-frequency (6-h) reanalyses of the European Centre for Medium-Range Weather Forecasts (ECMWF ERA-40 reanalysis,ECMWF 2002) and of the National Centers for Environmental Prediction (NCEP/DOE R2 reanalysis, Kanamitsu et al. 2002) as the main reference data for global evaluation of the CAM2 short-range weather forecasts.

However, current NWP reanalyses are not sufficient to evaluate all aspects of a GCM forecast, since they cannot furnish precise checks on physical forcings. Thus, ancillary high-frequency observations such as the ARM field data are indispensable for *independent* evaluation of GCM parameterizations. The most comprehensive data at 6-h and higher frequencies (in some cases, at frequencies comparable to a GCM time step of 30 minutes) are available at the ARM site in the U.S. Southern Great Plains (SGP) during intensive observation periods (IOPs) such as June/July 1997.

3.2 Initialization Procedures

The CAM2 model currently lacks a DAS (although a community data assimilation testbed is under development--see http://www.cgd.ucar.edu/DART), and so it is necessary to devise a simple alternative to standard NWP initialization procedures. Because of their high accuracy, we currently are using both the ECMWF ERA-40 and NCEP/DOE R2 reanalyses for the CAM2 initialization. However, this entails a three-dimensional mapping of finer-resolution reanalysis data to the coarser (spectral T42/L26) CAM2 resolution. We have successfully adopted the relevant NWP algorithms (White 2001) for this mapping.

For our prototype implementation, initial values for the prognostic parameterized variables (e.g. cloud water in the CAM2 model) are successfully obtained via a spin-up procedure that is also used for the land initialization. The initialization of the land is especially problematical because it is difficult to map discrete and discontinuous land variables between different resolutions. Thus far, we have applied two model spin-up procedures, both of which allow the land model and prognostic parameterized variables to interact with and respond to the forcing from the atmospheric model which is constrained to follow the evolution of the observed atmosphere. We refer to 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. 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. In the future, we will experiment with other procedures for initialization of the land and the parameterized atmospheric variables.

3.3 Model Forecasts

The current CAPT practice is to generate five-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 frequencies for comparison with relevant ARM data). For each forecast, the model atmosphere is initialized by applying either the nudging or forecast/analysis methods described previously. We then compute the mean, at some elapsed time, of a sequence of forecasts that are initialized on different days. (This mean forecast may be calculated from forecasts that are initiated on consecutive days, or instead 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.) We also compute the difference between the mean forecast and corresponding evaluation data, as a way of estimating the GCM systematic forecast error.

4. RESULTS

Here we present selected results of applying the CAPT protocol to the CAM2 model that illustrate concepts previously discussed.

In order to verify that our simple initialization procedures were able to produce a large-scale dynamical state that was close to that of the verifying analyses, we first evaluated the skill of the CAM2 forecasts of the 500 hPa height field, following established guidelines (WMO 1999).

For example, we computed the mean anomaly correlation (AC) of these forecasts (a commonly accepted measure of forecast skill), where the verification anomalies were calculated from a thirty-year monthly mean climatology of the ECMWF ERA-40 reanalysis for the period 1970-1999. (The AC calculations were fairly insensitive to the choice of climatology.)

Figure 2 shows the AC decay (spatially averaged 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 June/July 1997 ARM IOP. These are compared with the AC decay of analogous forecasts from the models that generated the ECMWF ERA-40 and NCEP/DOE R2 reanalyses.

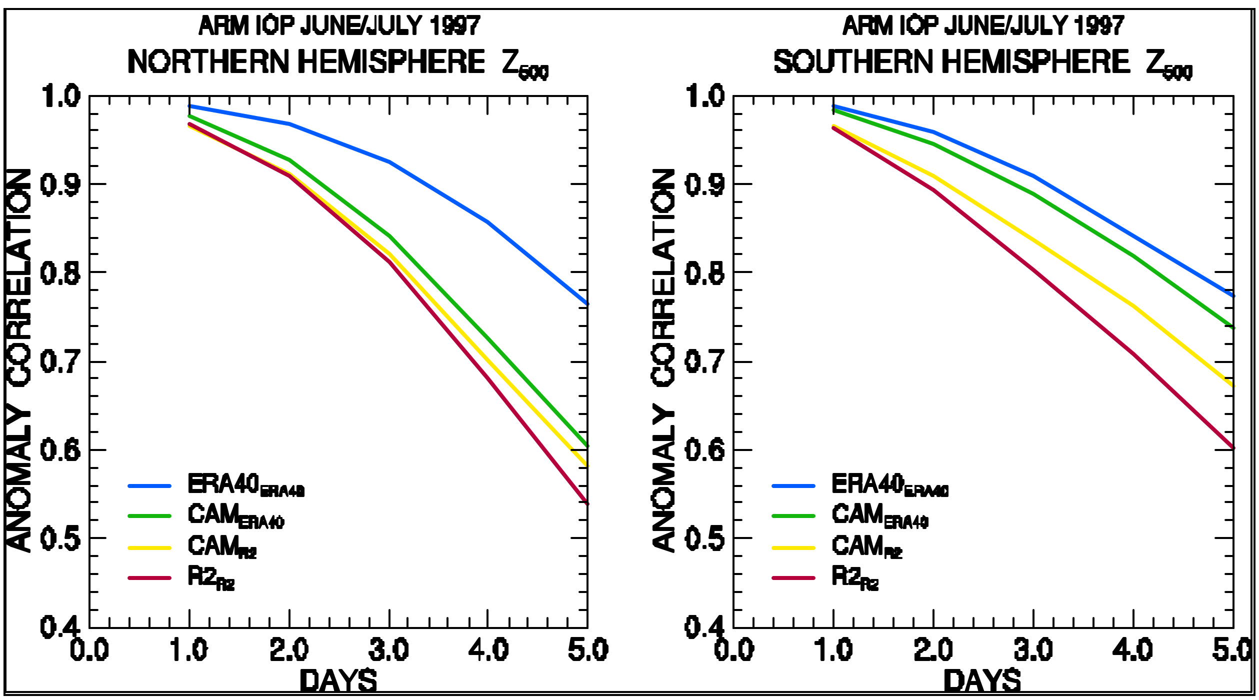

Figure 2: Mean anomaly correlation (AC) for a sequence of forecasts of the 500 hPa geopotential height field made with three GCMs as a function of forecast day during the June/July 1997 IOP. In all cases, the AC is interpolated to a common 2.5-degree global grid and spatially averaged (with cosine-latitude weighting) over the mid-latitudes of the Northern and Southern Hemispheres (20 N-90N and 20S-90S, 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 reanalyses (green); and the CAM2 model initialized with NCEP/DOE R2 reanalyses (yellow). Note that AC values less than 0.6 indicate the effective absence of forecast skill.

In general, the CAM2 forecasts of 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. As expected, the AC decay of all forecasts is more rapid in boreal summer, when mid-latitude synoptic control is weaker and forecast skill is more strongly influenced by physical processes.

However, CAM2 forecasts of large-scale atmospheric moisture (not shown) generally are not as skillful as those of the 500 hPa heights. This characteristic model behavior is illustrated locally as well (Figure 3).

It is seen that a CAM2 forecast of the atmospheric relative humidity profile for the period 19-25 June 1997 (Figure 3c) predicts a lower troposphere that is mostly too dry, and an upper troposphere that is mostly too moist, relative to both ARM observations (Figure 3a) and ECMWF ERA40 reanalysis (Figure 3b). The mean of 20 five-day forecasts during June/July 1997 (Figure 3d), implies that this is a systematic summer pattern in CAM2. Given the relatively skillful model forecast of large-scale dynamics during June/July 1997 (Figure 2), this result indicates that there are probably deficiencies in the CAM2 parameterizations of atmospheric moist processes

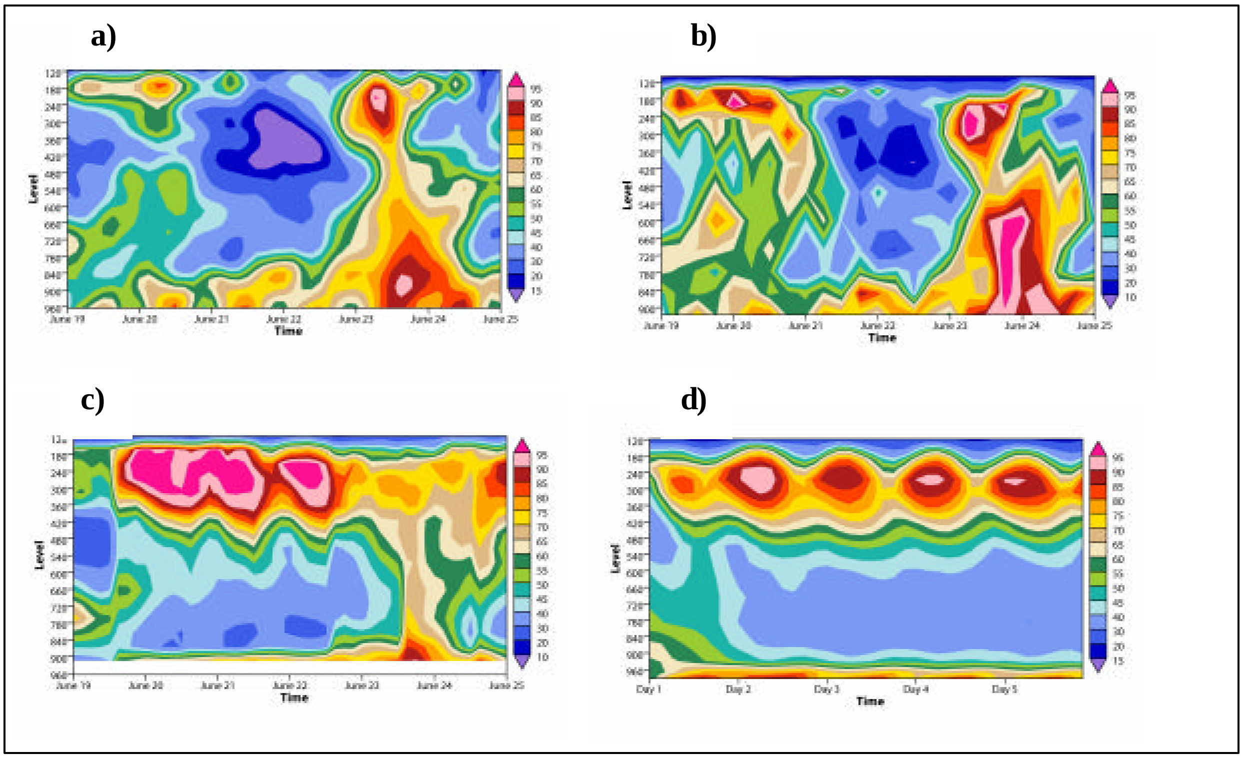

Figure 3: Time-height plot of relative humidity (in %) at the ARM SGP site for 19-25 June 1997, as obtained from a) ARM measurements, b) ECMWF ERA-40 reanalysis, and c) a CAM2 forecast. In d) the mean of 20 five-day CAM2 forecasts during June-July 1997 is shown. Note the tendency of the model mean forecast in d) to revert to a systematically dry lower troposphere and moist upper troposphere after being initialized at the start of each forecast day with the ERA-40 reanalysis.

The overly dry lower CAM2 troposphere is consistent with the tendency of the model to rain out moisture nearly every day during the June/July 1997 IOP at the ARM SGP site, rather than in the episodic bursts that are observed (Figure 4, top). In contrast, the agreement between CAM2 precipitation forecasts and observations was generally much better during the April 1997 IOP (not shown), when large-scale advective forcing was a more significant contributor to the column moisture balance.

This apparent seasonal sensitivity implies there may be a deficiency in the CAM2 parameterization of convective precipitation, although this is certainly not the only possible source of error. These phenomena are reminiscent of problems in the triggering mechanism of the model's deep convection scheme (Zhang and McFarlane 1995) that Xie et al. (2002) have identified previously in experiments with single-column models (SCMs) using ARM SGP observations. In that instance, the deficiencies were alleviated by replacing the standard trigger based on positive convective available potential energy (CAPE) with one based on the rate of generation of dynamic CAPE (DCAPE) by large-scale advective tendencies of temperature and moisture (Xie and Zhang 2000).

Implementation of the DCAPE convective triggering mechanism in the CAM2 model also has been evaluated in the CAPT framework. Because

CAPE can accumulate before convection occurs in the modified CAM2, stronger but less frequent precipitation events are produced by the new scheme, yielding generally better agreement with ARM data (Figure 4, bottom panel). Further evaluation of the performance of the new convective scheme in an AMIP climate simulation is currently in progress.

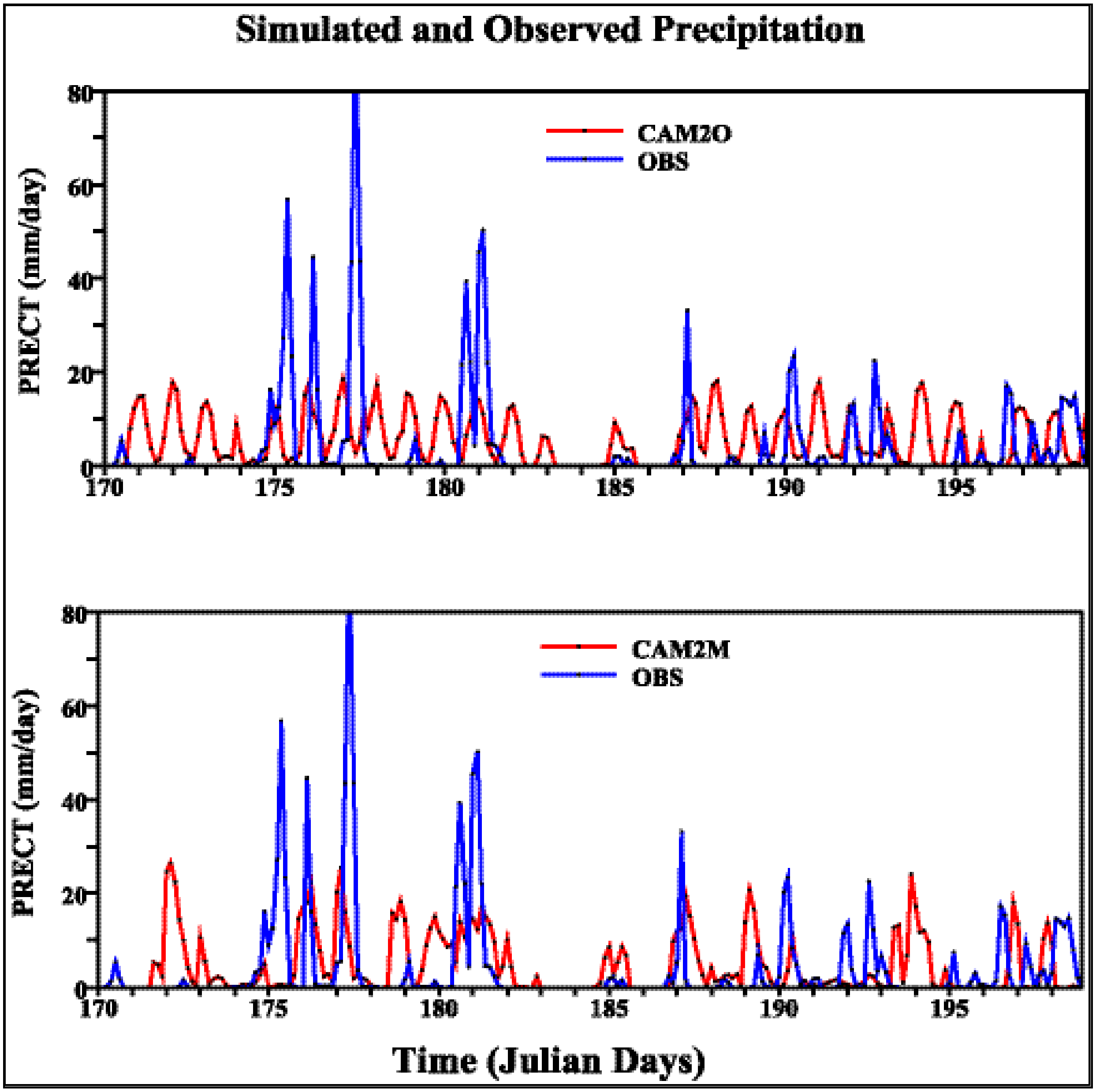

Figure 4: CAM2 forecasts of 24-h precipitation (initiated at 00Z each day) during June/July 1997 (blue) compared with precipitation observed at the ARM SGP site (red), both in units of mm/day. In the top panel, the forecasts are made with the standard version of the CAM2 model (denoted as CAM2O) which uses the Zhang-McFarlane parameterization of deep convection. The bottom panel shows forecasts of June/July 1997 precipitation made with a modified CAM2 model (denoted as CAM2M) that includes a modified convective triggering mechanism based on dynamical convective available potential energy (DCAPE). See text for further details.

5. SUMMARY

CAPT is motivated by the experience of model developers that it is very difficult to unravel GCM parameterization deficiencies solely by diagnosing the simulated climate, which includes systematic errors resulting from nonlinear interactions of many different processes. Our premise is that studying climate GCM parameterizations in a weather-forecasting framework is an effective way to identify their deficiencies and gain insights on their amelioration.

The overriding goal of CAPT is to improve the performance of model parameterizations as manifested by reduced GCM systematic errors, first in short-range weather forecasts, but ultimately in climate simulations. We acknowledge, however, that slowly developing systematic climate errors may remain resistant to significant reduction by the CAPT methodology.

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

Acknowledgments

This work was performed under the auspices of the U.S. Department of Energy (USDOE) Office of Science, Biological and Environmental Research (BER) program by the University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48. In addition, this work was partially supported at the National Center for Atmospheric Research (NCAR) by the Climate Change Prediction Program (CCPP), which is also administered by the USDOE Office of Science, BER program. NCAR is sponsored by the National Science Foundation.

References

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

Collins, W.D., J.J. Hack, B.A. Boville, P.J. Rasch, D.L. Williamson, J.T. Kiehl, B. Briegleb, J.R. McCaa, C. Bitz, S-J. Lin, R.D. Rood, M. Zhang, and Y. Dai, 2003: Description of the NCAR Community Atmosphere Model (CAM2). Accessible online at http://www.ccsm.ucar.edu/ models/atm-cam/docs/description/index.html.

ECMWF, 2002: ECMWF re-analysis ERA. Accessible online at http://www.ecmwf.int/research/era .

Hollingsworth, A., 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, March 2002. Also accessible online at http://www.ecmwf.int/publications/ library/ecpublications/pdf/tm361.pdf .

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

Kanamitsu, M. W. Ebisuzaki, J. Woolen, 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.

White, P.W. (ed.), 2001: FULL-POS Postprocessing and Interpolation, in IFS Documentation Part VI: Technical and Computational Procedures (CY23R4). European Centre for Medium-Range Weather Forecasts, Reading, UK.

WMO, 1999: Commission for Basic Systems Abridged Final Report with Resolutions and Recommendations, Extraordinary Session Karlsruhe, 30 September to 9 October 1998, Report WMO-No. 893, Geneva, Switzerland.

Xie, S., K-M. Xu, R.T. Cederwall, P. Bechtold, A.D. Del Genio, S.A. Klein, D.G. Cripe, S.J. Ghan, D. Gregory, S.F. Iacobellis, S.K. Krueger, U. Lohmann, J.C. Petch, D.A. Randall, L.D. Rotstayn, R.C. J. Somerville, Y.C. Sud, K. Von Salzen, G.K. Walker, A. Wolf, J. J. Yio, G.J. Zhang, and M.H. Zhang, 2002: Intercomparison and evaluation of cumulus parameterizations under summertime midlatitude continental conditions. *Quart. J. Royal Meteor. Soc.*, **128**, 1095-1135.

Xie, S., and G.J. Zhang, 2000: Impact of convective triggering function on single-column model simulations. *J. Geophys. Res.*, **105**, 14983-14996.

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.