# The CCPP-ARM Parameterization Testbed (CAPT): A Strategic Plan

**(Last update 28 April 2003)**

## 1. 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 change scenarios (e.g. global warming, ozone depletion, changes in land use, etc.). Hence, there is a compelling need to systematically improve the performance of GCMs in simulating the present climate, and thereby to reduce their uncertainties in predicting climate change (e.g. Goody et al. 2002). This endeavor requires a two-pronged strategy that entails

- comparison of GCM simulations with observations over a broad range of time scales in order to diagnose the details of the associated simulation errors, and
- reduction of these errors by improving representations of key physical processes and feedback mechanisms, thereby enhancing the physical realism of GCM climate simulations.

Much of our collective understanding of current strengths and weaknesses of GCM climate simulations comes from periodic assessments of the Intergovernmental Panel on Climate Change (IPCC 2001). These, in turn, are largely based on the findings of model intercomparison projects organized chiefly by the World Climate Research Programme (e.g. WCRP 2001).

In particular, a wealth of model-performance information is provided by the Atmospheric Model Intercomparison Project (AMIP) in which some 30 GCMs, with common forcing by observed ocean boundary conditions, are simulating the present climate (Gates 1992). On distilling the many technical details of the first AMIP experiment (e.g. Gates et al. 1999), an intriguing result emerges: the climate simulations of GCMs designed mainly for numerical weather prediction (NWP) are surprisingly "competitive" with those of GCMs designed expressly for climate studies. This outcome arguably stems from NWP modelers' adoption of a standard diagnostic protocol, which is summarized as follows.

GCMs designed for numerical weather prediction (hereafter, "NWP GCMs") are developed so as to optimize their short-range (~ 0-72 hour) forecast performance, as evaluated against either observations or NWP analyses that are synthesize observations and model-derived quantities. However, even though an NWP GCM usually is run at comparatively high resolution (~ 0.5x0.5-degree grid and ~ 30+ vertical layers), it cannot explicitly resolve many phenomena (e.g. individual clouds and associated radiation, precipitation and related microphysics, turbulent boundary layer processes, etc.). These effects must be parameterized in terms of the model's resolved variables, and deficiencies in these parameterizations are identified by diagnosing discrepancies between

1) model forecasts and NWP analyses,
2) model forecasts and observations that are not assimilated in the NWP analyses (e.g. clouds, precipitation, etc.), and
3) NWP analyses and observations.

The efficacy of modifying selected parameterizations is assessed according to whether such changes increase the skill of the GCM's short-range forecasts and/or improve the NWP analyses. If these parameterization changes also improve model performance at time scales longer than the deterministic forecast range of ~15 days (owing to reductions in systematic model biases), the new schemes are included in the next-generation NWP GCM. When this diagnostic protocol is followed rigorously, the new GCM also has a higher probability of performing well in simulating climate (seasonal, inter-annual, and multi-annual) time scales, provided that the new parameterizations function satisfactorily at the coarser resolutions (~3x3-degrees and ~15 vertical layers) typical of GCMs designed for climate simulation (hereafter, "climate GCMs").

Owing to the demonstrated improvements in model performance resulting from applying such a diagnostic protocol to NWP GCMs, the WCRP's Working Group on Numerical Experimentation has advocated a 'Transpose AMIP' project (WGNE 1999) to encourage adoption of similar methods for coarse-resolution climate GCMs. Practical impetus for this effort now is being provided by a new U.S. Department of Energy (USDOE) initiative: the CCPP-ARM Parameterization Testbed (CAPT).

CAPT will apply the first component of the NWP diagnostic protocol (comparison of model forecasts with analyses) to climate GCMs that are supported by the USDOE Climate Change Prediction Program (CCPP). In addition, CAPT will implement the second component of this diagnostic protocol (comparison of GCM forecasts with observations not assimilated in the analyses) by employing USDOE Atmospheric Radiation Measurement (ARM) observations and similar field data for model evaluation. Moreover, CAPT will foster closer collaborations between GCM developers and parameterization specialists, especially those funded under the CCPP and ARM Programs. Finally, CAPT will supply evaluation data and diagnostic software in a form that can be conveniently applied to climate GCMs.

The remainder of this paper discusses the CAPT strategy, first by explaining its scientific rationale in Section 2, then by outlining the proposed diagnostic protocol in Section 3, and finally by elaborating relevant technical details in Section 4. A brief summary is given in Section 5.

## 2. CAPT: Scientific Basis

### *2.1 Perspective*

The standard approach in developing a climate GCM is to focus on whether a given parameterization favorably impacts the model's climate statistics and/or their perceived departures from the observed climate, such as can be estimated from relatively sparse spatio-temporal sampling. However, such an approach limits accurate identification of specific parameterization deficiencies, since the GCM's climate statistics reflect compensating errors in the simulation of atmospheric dynamics as well as many different physical processes.

Instead, CAPT is advocating a different approach for parameterization testing: the use of high-frequency (~6 hourly) NWP analyses both to realistically initialize a coarse-resolution climate GCM and to evaluate the accuracy of its subsequent short-range weather forecasts. The rationale is that *most of the GCM's forecast error can be attributed to parameterization* *deficiencies, once the model's dynamical state is initialized realistically*. This is especially true when comparing analysis-initialized GCM forecasts with observations that are not assimilated in the NWP analysis (e.g. radiative and turbulent fluxes, clouds, precipitation). Such unassimilated observations can be obtained, for example, from ARM field data or satellite measurements.

There are other reasons as well to adopt this more physically based diagnostic protocol. Because forecast skill can be readily quantified using standard NWP metrics, the effects of modifying a parameterization can be objectively determined. In addition, the rich variety of weather phenomena allows GCM parameterizations to be comprehensively tested. Moreover, the short-range behavior of the GCM *is* relevant for climate simulation, since systematic model errors often arise within the first few days of a simulation, and then just grow larger with time.

In principle, a climate GCM that demonstrates generally enhanced short-range forecast skill also should show improved simulation of climate statistics, since these are aggregates of the detailed evolution of the model, rather than stochastically predicted quantities. Such improvements must be demonstrated in practice, though, by first testing the parameterization in GCM predictions beyond the deterministic range of ~15 days, and then in long climate simulations. To the degree that connections exist between forecasting errors and climate errors, the testing of GCM parameterizations also will proceed more efficiently by focusing attention first on many short forecasts rather than a few long climate simulations.

Hence, CAPT's premise is that application of high-frequency NWP analyses to evaluate the weather forecasts of climate GCMs is an effective technique for

1) identifying deficiencies in model parameterizations,
2) providing insights into the causes of these shortcomings, and
3) quantifying the impact of changes made to the models.

In applying this diagnostic protocol, CAPT's overriding goal is to improve GCM performance--as manifested initially in short-range forecasts, but ultimately in climate simulations.

### *2.2 NWP analyses*

It is evident that NWP analyses (remapped to the coarser resolution of a climate GCM) will play a central role in the CAPT diagnostic protocol. The typical analysis is generated by a four-dimensional data assimilation (FDDA) system that applies variational mathematics to optimally estimate global weather from ingested surface, radiosonde, and satellite observations (Haltiner and Williams 1980, Daley 1991). This data assimilation also entails the application of an NWP GCM--often the same model as that used for weather forecasting.

The success of the CAPT strategy hinges on the accuracy of the NWP analyses to be used for evaluating the weather forecasts of the climate GCM. Hollingsworth et al. (2002), for example, have shown that the short-range forecasts of a representative NWP GCM track observations of atmospheric state variables (i.e. pressure, temperature, moisture, and momentum fields) with an accuracy that falls within current measurement uncertainties. Thus, at least in observation-rich regions, today's operational NWP analyses (and, by extension, multi-decadal reanalyses) are demonstrably reliable references for accurately identifying errors in GCM weather forecasts.

Moreover, the existence of high-quality NWP analyses makes it feasible to initialize a climate GCM by methods that do not require developing a full-blown FDDA system for the model (see Section 4.2). If the initial state of the GCM is consistent with both the model parameterizations and the NWP analyses and is also close to dynamical balance, the initialization noise associated with gravity waves should be relatively small. In that event, the forecast errors will grow gradually (i.e. without sudden/sharp breaks) as the prediction period increases, and most of the error growth that exceeds that of the inherent (in part, resolution-dependent) predictability error will be attributable to deficiencies in the model parameterizations.

Although NWP analyses have previously been applied to the evaluation and development of climate GCMs (e.g. Jeuken et al.1996, POTENTIALS 1999), such studies have focused on analysis-forced integrations of these models. CAPT's innovation is to recognize the value of diagnosing a climate GCM's weather forecasts, wherein the parameterizations are free to interact with each other and with the model dynamics. Even though the forecast skill of a coarse-resolution climate GCM is likely to be less than that of a fine-resolution NWP GCM, *relative* improvement in the climate GCM's forecasts still should be attainable by making appropriate changes in the model's physical parameterizations. Thus, CAPT's objective is not to improve the short-range forecasts of a climate GCM *per se*, but rather to use weather forecasting as a context for parameterization testing.

### *2.3 Ancillary evaluation data*

However, an NWP analysis is not sufficient to evaluate the GCM's short-range forecast *in all respects*: although the analysis is an optimal estimate of atmospheric state variables (given the available weather observations), it cannot furnish precise checks on physical forcings such as radiation and its interaction with clouds, convective processes, and turbulent fluxes. That is, although an NWP analysis includes estimates of such forcings, these depend strongly on the physical parameterizations of the analysis GCM, and so are only indirectly related to the assimilated observations.

Hence, for *independent* evaluation of GCM physical parameterizations, high-frequency satellite data and field observations such as those provided by the ARM Program (Stokes and Schwartz 1994) are indispensable--indeed, their practical value for identifying GCM parameterization problems has been amply demonstrated (e.g. Webb et al. 2001, Morcrette 2002). Of course, the physical consequences of changing particular parameterizations also can be assessed using these ancillary data. Moreover, field observations of state variables provide a local check on the NWP analyses.

## 3. CAPT: Diagnostic Protocol

An overview of the proposed CAPt diagnostic protocol is illustrated by Figure 1. As a first step, an appropriate regional case study is selected, based on availability of high-frequency evaluation data (NWP analyses, satellite data, and field observations) that can effectively test a candidate parameterization implemented in the climate GCM. (Such candidates may include, for example, radiation, convection, or cloud-formation schemes.)

The climate GCM's state variables first are globally initialized from NWP analyses that are remapped to the coarser resolution of the climate GCM and temporally interpolated for times pertinent to the case study. Climate GCM forecasts then are generated, with data archived at regular intervals. At each forecast interval, the departures of the GCM state variables from the remapped NWP analysis data will be quantified by standard forecast skill scores. These metrics also may be computed for forecasts that are stratified according to different synoptic regimes for the case study region so as to identify model errors that are more obvious for particular atmospheric conditions.

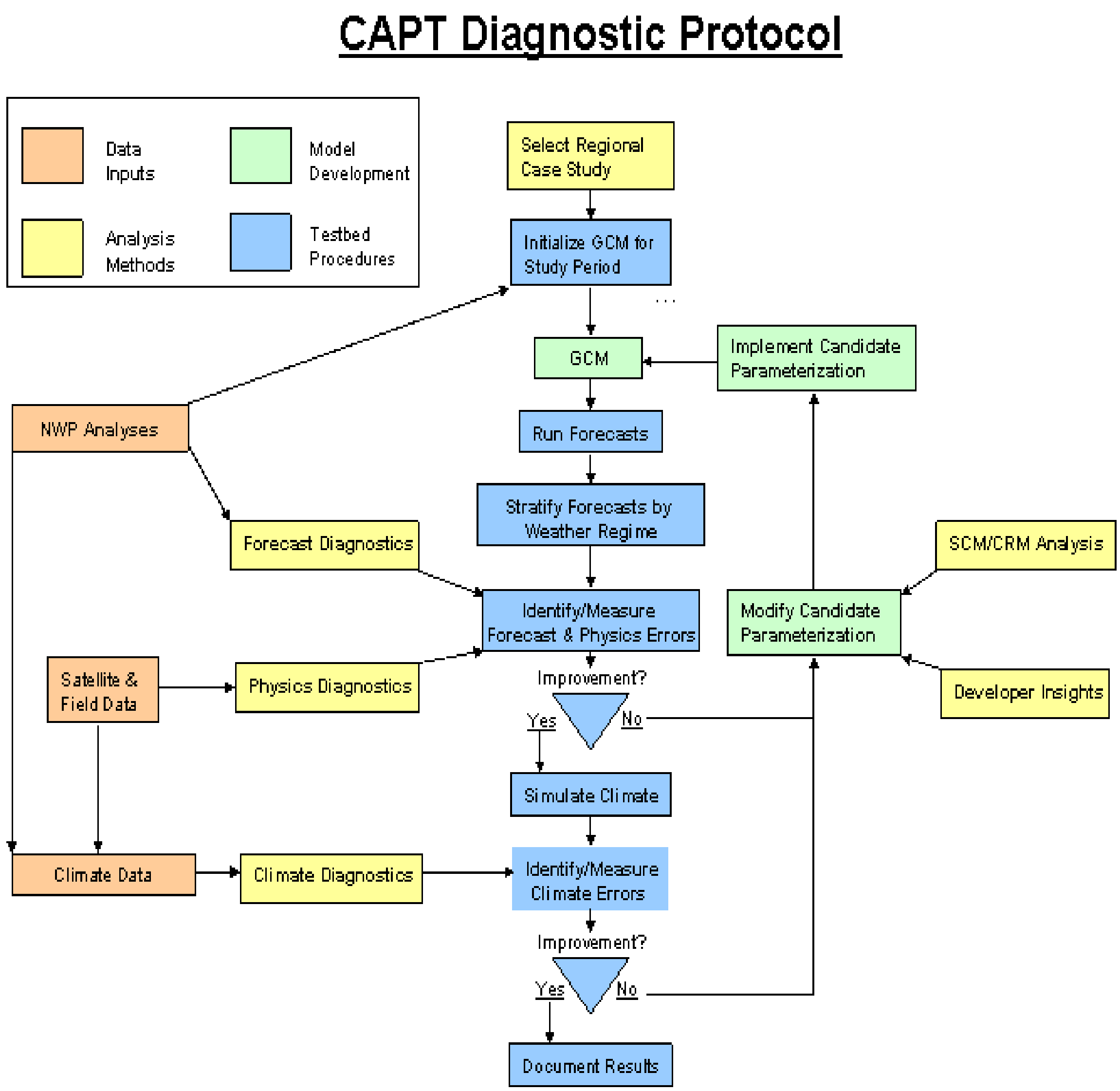


**Figure 1: Schematic depiction of the CAPT diagnostic protocol.**

If the candidate parameterization does not improve the overall forecast skill of the GCM (e.g. relative to forecasts made with a "standard version" parameterization), the model developers will need to formulate potential correctives. To guide this process, the case study observations of physical forcings can be used to infer connections between the GCM forecast errors and physics errors, and in turn to suggest needed modifications in the candidate parameterization. At this stage, other analysis tools such as single-column models (SCMs) and/or cloud-resolving models (CRMs) also may be enlisted in developing an improved parameterization.

Whether these parameterization changes translate into actual model improvement will be assessed by repeating all of the above diagnostic procedures for a new set of GCM forecasts. General improvement in model forecast skill then will be tested further in climate simulations, where the relevant evaluation will be determined from the statistics of the appropriate high-frequency data over the simulation period. A need for additional parameterization changes may, of course, become obvious in these longer simulations. When general improvements in the climate statistics of the GCM have been demonstrated, the details will be documented in technical reports and/or peer-reviewed publications that are coauthored by all participants. The entire process then may be repeated for another candidate parameterization.

## 4. CAPT: Technical Details

A prototype of the CAPT diagnostic protocol is presently being implemented for the NCAR Community Atmospheric Model Version 2 (CAM2). In this preliminary phase of the work, many details are yet to be settled, so it is possible only to convey a general sense of the technical issues to be addressed. These are organized according to the elements of Figure 1, as follows.

### *4.1 Evaluation data*

CAPT will use the high-frequency (6-hourly) ERA-40 and NCEP R2 reanalyses (ECMWF 2002, Kanamitsu et al. 2002), remapped to the coarser GCM resolution (e.g. spectral T42 with 26 vertical levels for CAM2), for primary evaluation of the model weather forecasts. Such a remapping is also necessary in using the reanalyses as target data for initializing these forecasts (see Section 4.2).

Ancillary (field and satellite) data, available at 6-hourly and higher frequencies (in some cases, at frequencies comparable to GCM time steps), will provide independent checks on the model forecast, and on other variables that are more directly related to parameterized processes. Continuous field observations are available at selected points, notably at the ARM sites (ARM 2002) in the U.S. Southern Great Plains (SGP), the North Slope of Alaska (NSA), and the Tropical West Pacific (TWP).

The most comprehensive set of high-frequency observations are presently available during sporadic intensive observation periods (IOPs), but ARM increasingly is providing continuous records of these data sets. Selected examples at the ARM SGP site include:

- Radiosonde soundings of temperature and humidity at 3-hourly frequencies at the SGP central facility (CF) near Lamont, Oklahoma, as well as at 4 neighboring stations. (These can provide independent local checks on NWP analyses.)
- Wind profiler measurements at hourly frequencies;
- Solar Infrared Radiation Station (SIRS) measurements of surface upwelling/downwelling longwave and shortwave irradiances at 1-minute frequencies;

- Energy Budget Bowen Ratio (EBBR) measurements of surface latent and sensible heat fluxes at 30-minute frequencies;
- Microwave Radiometer (MWR) measurements of column precipitable water and total cloud liquid water at 5-minute frequencies;
- Surface Meteorological Observation Stations (SMOS) and Oklahoma and Kansas Mesonet stations (OKM and KAM) observations of surface meteorology (precipitation, pressure, winds, temperature, and relative humidity) at 5 to 30-minute frequencies.

In addition, collocated Geostationary Operational Environmental Satellite (GOES) temperature and dew point retrievals are available at 30-minute frequencies.

Similar types of point field data also are available at a few other locations where international Global Energy and Water Cycle Experiment (GEWEX) Continental-Scale Experiments have been conducted (CSE 2002). There also are plans to expand this observational network to some 30 sites during the 2003-2004 GEWEX Coordinated Enhanced Observing Period (CEOP 2002).

Other potentially relevant evaluation data include coordinated satellite, aircraft, and surface measurements that have been collected during one-time field campaigns. Many of these have been centralized by the GEWEX International Satellite Cloud Climatology Project (ISCCP), in particular for several locations and time periods that are under intensive investigation by GEWEX Cloud System Study participants (GCSS 2002).

To make these point observations fully relevant for model evaluation, they need to be aggregated to the scale of a GCM grid box. To this end, some of the ARM SGP data has been subjected to objective variational analysis (Zhang and Lin1997, Zhang, et al. 2000). This method uses the domain-averaged surface precipitation and latent and sensible heat fluxes as well as the surface and top-of-atmosphere (TOA) radiative fluxes to constrain the atmospheric variables, so that heat, moisture, and momentum are conserved. Thus, the resulting derived data are dynamically and thermodynamically consistent.

Another complication of using such ancillary data for model evaluation is that a "forward model" (e.g. Morcrette 1991, Klein and Jakob 1999) often must be applied to translate the GCM's output variables into relevant observables (e.g. band radiances, cloud reflectivies, etc.). For example, the ISSCP cloud simulator (Webb 2002) is currently being incorporated in the CAM2 model. In addition, modifications of the model's standard configuration (e.g. increasing the frequency of the radiation calculations and of the archiving of model output) may be necessary to match the frequency of the observations.

### *4.2 Model initialization*

In order to isolate a climate GCM's parameterization-related errors, proper initialization of the climate GCM (initially, CAM2) is crucial for minimizing noise resulting from an unbalanced land/atmosphere initial state and for producing needed unobserved variables (e.g. cloud water and ice, soil moisture and temperature profile, etc.). Hence, CAPT is currently investigating the relative merits of two standard techniques--"nudging" and "forecast-analysis"--for initializing a model's atmospheric state variables using an NWP analysis as target data

Nudging (e.g. Jeuken et al. 1996) attempts to steer atmospheric variable $\alpha$ toward that of the corresponding analysis variable $\alpha_0$ (that is interpolated to the model's horizontal/vertical grid) by adding a Newtonian relaxation term to the relevant prognostic equation:

$$D\alpha/Dt = F(\alpha, x, t) + (\alpha_0 - \alpha)/\tau$$

Here F includes all spatio-temporal forcings, and the relaxation time constant $\tau$ may be specific to each nudged variable $\alpha$ (e.g. including atmospheric temperature T, winds u/v, surface pressure Ps, and humidity q). Nudging is currently implemented at every model time step for six months prior to the first forecast (see schematic in Figure 2a). This lengthy integration serves to spin up "slow" CAM2 land variables (e.g. soil moisture/temperature and snow cover) to a state in which the associated surface radiative and turbulent fluxes are consistent with the nudged model atmosphere. Aside from this long spin-up period, the chief disadvantage of implementing the nudging technique is that the model prognostic equations must be modified to include the necessary relaxation terms.

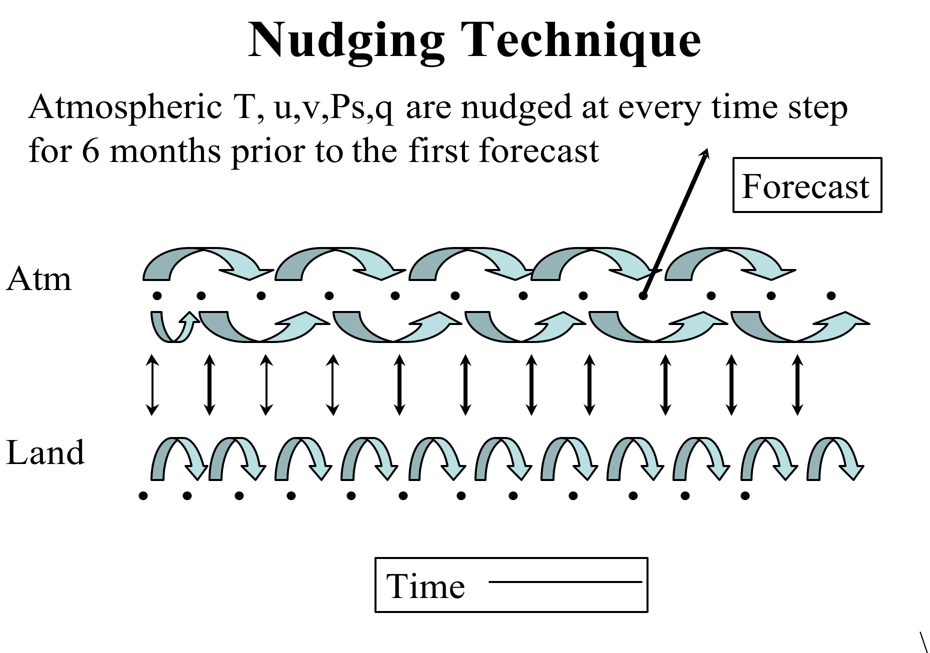


**Figure 2a: Schematic depiction of the nudging initialization technique.**

Such model code changes are not needed in implementing the forecast-analysis technique (e.g. Harrison et al. 1999). In this approach, the model's forecast (including unobserved variables) is compared directly to the NWP analysis at six-hour intervals, and a fractional difference between forecast and analysis (the "increment") is added repeatedly to the evolving model atmospheric state. (When the *entire* difference between the forecast and the analysis is added, the forecast-analysis technique is identical to "full-field insertion".) As in the nudging technique, the model's slow land variables adjust so that the surface fluxes are consistent with the updated atmospheric state (Figure 2b). A disadvantage of forecast-analysis is that the model's atmospheric state variables must be mapped to the analysis data grid, and the increment mapped back to the model grid. In addition, it may be necessary to apply a digital filter in order to damp gravity waves which may be more extensive when initializing by the forecast-analysis technique (Lynch and Huang 1992, Polavarapu et al. 2000).

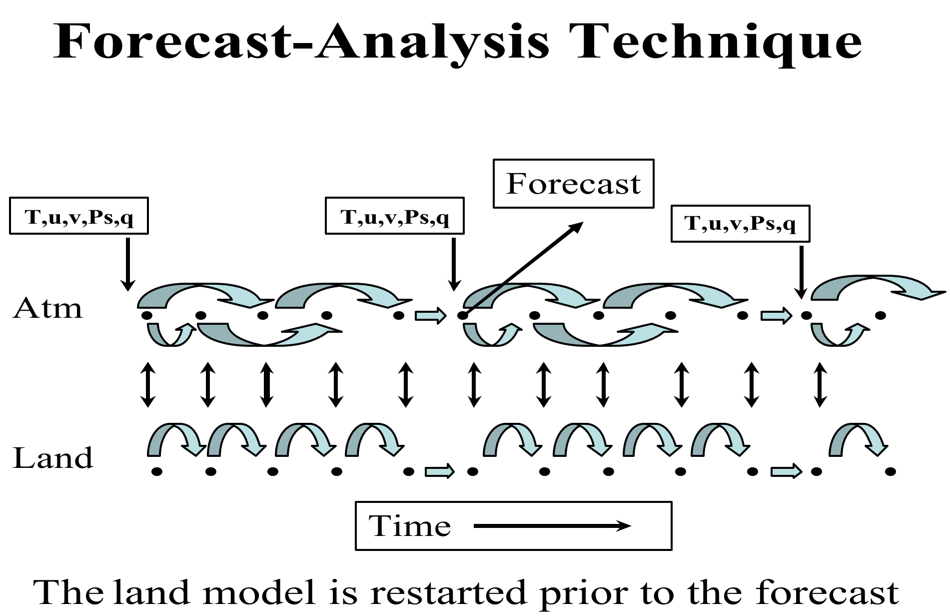


**Figure 2b: Schematic depiction of the forecast-analysis initialization technique.**

CAPT is currently applying both of these initialization techniques in studies designed to estimate the magnitude of initialization errors in different contexts. For example, a "perfect model" study is estimating the irreducible part of the CAM2 initialization error by using the model's own outputs as the target data set. Preliminary results indicate the need to nudge the model for as long as ~ 6 months, depending on the seasonally dependent adjustment time of the slower land and snow processes. In addition, a "perfect analysis" study is providing estimates of the minimum initialization error to be expected when the NCEP R2 reanalysis is used for initializing the associated NCEP GCM.

A "practical initialization" study is also underway to assess the relative merits of nudging vs. forecast-analysis methods for initialization of the model atmosphere using the ERA-40 reanalysis as target data. Thus far, the principal technical issue concerns how to appropriately remap the reanalysis surface pressure and net downward radiation to the coarser model resolution, given the associated large differences in topographic elevations. Following Morcrette (2002), initialization noise is judged to be tolerable if a 30+ day time series formed by concatenation of a sequence of forecasts started at fixed intervals apart show qualitatively similar features, without evincing sudden/sharp breaks.

Another unresolved issue is whether the CAM2 land model (Bonan et al. 2002) should be initialized in a different way than in the current practice of "slaving" it to the atmospheric model.

Here, at least three different approaches might be pursued. Using the initialization of the soil moisture profile as an example, these methods are summarized below, in ascending order of implementation difficulty:

- Remapping a reanalysis soil moisture profile to that of the climate model, subject to maintaining equivalent soil moisture availabilities (personal communication, P. Viterbo, M. Best, and H. Douville).
- Scaling of the mean and variance of the reanalysis soil moisture profile, where the scaling coefficients are estimated from a multi-year run of the climate model made with observed sea surface temperatures as ocean boundary conditions (personal communication, M. Kanamitsu).
- Specifying soil moisture profiles from a multi-annual off-line or coupled integration of the climate model's land scheme forced by estimates of observed terrestrial precipitation and surface insolation (personal communication P. Dirmeyer, R. Koster, and K. Mitchell).

In practice, it will be necessary to strike a compromise between a land initialization scheme that can be easily implemented versus one that yields a close initial land-atmosphere balance, but at excessive computational cost.

### *4.3 Model forecasts*

Although much useful diagnostic information can be gleaned in the course of initializing the model, intrinsic behaviors are more fully revealed in forecasts. For example, the observed propensity of the CAM2 atmosphere to dry excessively when nudged toward the NWP reanalyses is even more evident in its forecasts of atmospheric moisture at the ARM SGP site (apparently related to incorrect forecast of precipitation events and/or amounts). This pattern is evident whether the forecasts are compared to the corresponding nudged initial states (Figure 3) or to ARM field observations (Figure 4). These results are thought to be linked to deficiencies in the CAM's convective triggering mechanism over land (Xie et al. 2002).

The current CAPT practice is to generate 3-day (0-72 hour) CAM2 forecasts, with data archived every 3 hours. A new 3-day forecast is started at 00Z for each day of the time period of interest (e.g. during an ARM IOP), where the model atmosphere is initialized by application of either nudging or forecast-analysis techniques.

Both the magnitude of forecast errors and their growth rate are of diagnostic value. These aspects of the forecast can be quantified using standard NWP metrics, for example by computing the temporal variation of the mean bias and the root-mean-square (RMS) error (yielding error amplitude information) or that of the anomaly correlation coefficient (yielding error pattern information).

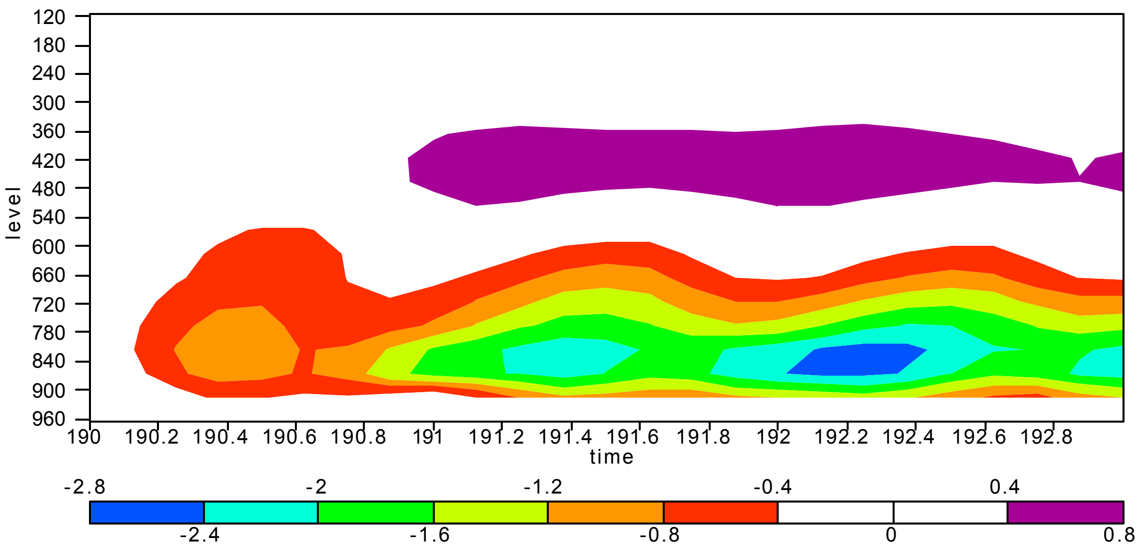


**Figure 3: Evolution over three-day intervals (for arbitrarily numbered days 190-193) of ensemble-mean differences between forecasts of the vertical profile of CAM2 specific humidity and nudged-run profiles (on pressure levels in hPa) for the ARM SGP site. The ensemble-mean difference is computed over 10 three-day forecasts and corresponding nudged runs during the period 1-10 July of the 1997 IOP. Note the anomalous drying of the lower troposphere (negative forecast-minus-nudged humidity differences, denoted by cooler colors) that intensifies with time over the three-day interval.**

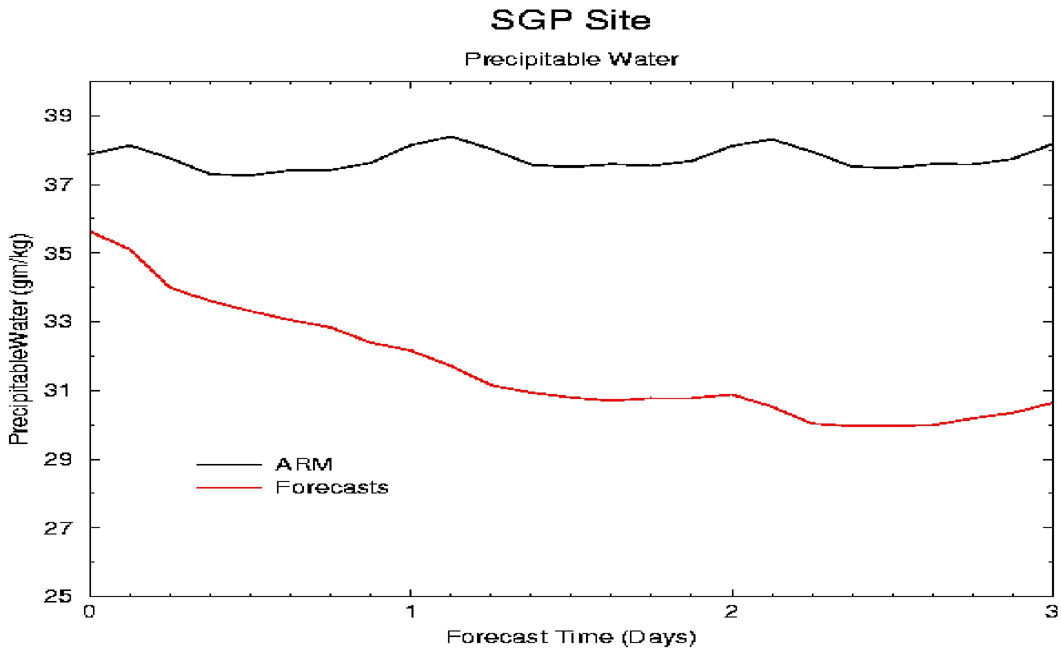


**Figure 4: Ensemble-mean of 29 three-day forecasts of precipitable water at the ARM SGP site during June/July 1997 and corresponding mean 3-hourly observations. Note that the model erroneously forecasts atmospheric drying and an attenuated diurnal cycle.**

Figure 5, for example, shows spatially averaged anomaly correlations of 3-day forecasts of 500 hPa heights by the CAM2 model versus remapped 6-hourly ERA-40 reanalysis data. Here, a different 3-day forecast is generated at 00Z for each day of the 19 June-18 July 1997 ARM IOP, where the initial conditions are obtained by nudging the CAM2 atmospheric state variables and surface pressure toward the corresponding remapped ERA-40 reanalysis data.

The anomaly correlations of different forecasts in Figure 5 decrease fairly smoothly with time, suggesting that the initial CAM2 500 hPa heights are sufficiently balanced by the nudging to reduce the initialization noise in this variable to a tolerable level. However, the anomaly correlations of separate forecasts are of different initial magnitudes and fall off at varying rates with time, implying a sensitivity to synoptic conditions. This sensitivity will be exploited in diagnosing the model's parameterizations.

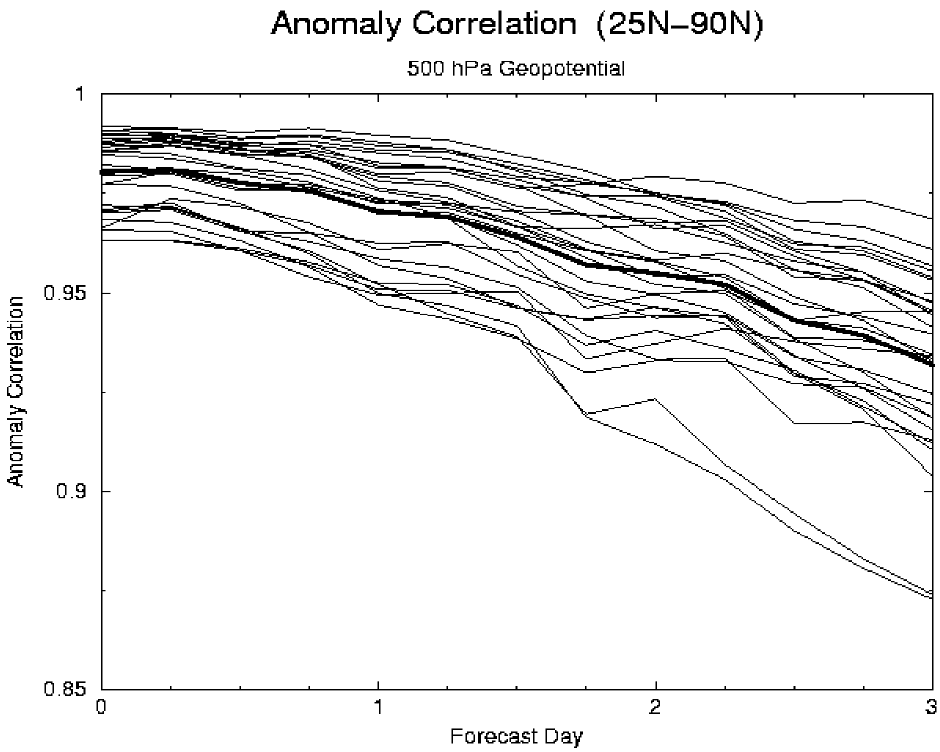


**Figure 5: Spatially averaged (over latitudes 25N-90N) anomaly correlations of three-day CAM2 forecasts of 500 hPa height versus ERA-40 reanalysis during June/July 1997. The ensemble-mean anomaly correlation is indicated by the darker line.**

### *4.4 Model parameterization diagnosis*

The attribution of forecast errors to deficiencies in model parameterizations and the subsequent correction of these schemes are central to the improvement of climate GCMs. Since most parameterization deficiencies will be more starkly revealed under certain weather conditions, CAPT will diagnose CAM2 forecasts that are stratified according to regional synoptic conditions (e.g. clear vs. cloudy, dry vs. wet, summer vs. winter cases, etc.).

Because of the highly nonlinear character of GCMs, it is rarely easy to draw connections between forecast errors in the atmospheric state variables and the physical forcings that are governed by the model's parameterizations. Thus, CAPT will analyze CAM2 forecasts that are coincident with case studies (e.g. short-term field campaigns, ARM IOPs, etc.) where available high-frequency field data can augment the reanalyses by providing independent checks on model physics and state variables. CAPT analysis of such case studies will attempt to identify a recurring connection between a forecast error in a state variable and anomalous physical forcing that can be tied to particular deficiencies in a selected parameterization.

When a problematical connection of this type is found, CAPT will report the forecast/forcing error metrics and related phenomenological details to GCM developers and collaborating parameterization specialists so that they have a basis to formulate potential correctives for the relevant parameterization. In addition, before a modified parameterization is implemented in the climate GCM, parameterization developers may test it in a single-column and/or cloud-resolving model (SCM and CRM), and in a simplified (e.g. linearized, aqua-planet, etc.) GCM (e.g. Xie et al. 2002, Xu et al. 2002).

Once a new scheme is implemented in the climate model, it will be extensively tested to determine whether a reduction in forecast errors results for the case study in question, as well as for other initial conditions. A parameterization that produces global reductions in forecast errors will also need to be tested in long integrations so that its effects on the simulation of different climate processes can be assessed. Moreover, even if a new parameterization is successful in reducing certain types of climate errors, it is likely that the GCM's other parameterizations will need to be "retuned" before its climate simulation displays overall improvement. In that event, this new version of the climate model will be the starting point for testing additional parameterization changes.

## 5. Summary

CAPT is motivated by the historical experience that it is exceedingly difficult to unravel parameterization deficiencies solely by diagnosing a GCM’s climate statistics, which reflect systematic biases resulting from the convolution of nonlinear/nonlocal interactions of many different schemes. Instead, the starting point of the CAPT protocol is to focus on the short-range forecasts of the climate GCM, and to closely compare these against well-sampled observations provided by NWP analyses and satellite/field data as the first criterion for assessing model performance/improvement.

Thus, new parameterizations will be considered for testing in GCM climate simulations only after they improve the model's weather forecasts. It should not be expected, however, that the transition from the short-range to climate scales will be entirely straightforward--further model adjustments may be needed before overall improvement in the climate simulation is evident. Hence, the CAPT protocol should not be viewed as a panacea, but only as one element in a hierarchy of techniques (including, for example, diagnostics based on single-column, cloud-resolving, and simplified global models) to bolster the observational/ scientific foundations of climate model development. Nevertheless, it is anticipated that important insights on improving climate GCMs will flow from adopting this NWP-inspired methodology.

*Acknowledgments*

This work was performed under the auspices of the U.S. Department of Energy by the University of California Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48.

**References**

ARM, 2002: ARM cloud and radiation test bed sites. Accessible online at http://www.arm.gov/docs/sites.html .

Bonan, G.B., K.W. Oleson, M. Vertenstein, S. Levis, X. Zeng, Y. Dai, R.E. Dickinson, and Z-L. Yang, 2002: The land surface climatology of the Community Land Model coupled to the NCAR Community Climate Model. *J. Climate* (submitted).

CEOP, 2002: Coordinated Enhanced Observing Period. Accessible online at http://www.gewex.org/ceop.htm.

CSE, 2002: Locations of present/future GEWEX continental-scale experiments. Accessible online at http://www.gewex.org/cseslocation.html.

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

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

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

Gates, W.L., J.S. Boyle, C. Covey, C.G. Dease, C.M. Doutriaux, R.S. Drach, M. Fiorino, P.J. Gleckler, J.J. Hnilo, S.M. Marlais, T.J. Phillips, G.L. Potter, B.D. Santer, K.R. Sperber, K.E. Taylor, and D.N. Williams, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP). *Bull. Amer. Meteor. Soc.*, **80**, 29-55.

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

Goody, R., J. Anderson, T. Karl, R. Balstad Miller, G. North, J. Simpson, G. Stephens, and W. Washington, 2002: Why monitor the climate? *Bull. Amer. Meteor. Soc.*, **83**, 873-878.

Haltiner, G.J., and R.T. Williams, 1980: Numerical Prediction and Dynamic Meteorology, John Wiley and Sons, Inc., second edition, 477 pp.

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.

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 .

IPCC, 2001: Climate Change 2001: The Scientific Basis, J. Houghton, Y. Ding, D.J. Griggs, M. Noguer, P.J. Van Der Linden, and D. Aiaosu (eds.), Cambridge University Press, Cambridge CB@1BR, UK, 944 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*., **D12**, 101, 16939-16950.

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.

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.

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

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

Morcrette, J-J., 1991: Evaluation of model-generated cloudiness: Satellite-observed and model-generated diurnal variability of brightness temperature. *Mon. Wea. Rev*., **119**, 1205-1224.

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.

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

POTENTIALS, 1999: Project on tendency evaluations using new techniques to improve atmospheric long-term simulations: Final Report. Available online at

http://www.dmi.dk/pub/POTENTIALS/Final/Final.pdf .

Stokes, G.M., and S.E. Schwartz, 1994: The Atmospheric Radiation Measurement (ARM) Pro gram: Programmatic background and design of the cloud and radiation test bed. *Bull. Amer. Meteor. Soc*., **75**, 1201-1221.

WCRP, 2001: Annual review of the World Climate Research Programme and report of the twenty-second session of the Joint Scientific Committee. WMO/TD No. 1096, World Meteorological Organization, Geneva, Switzerland. Also available online at http://www.wmo.ch/web/wcrp/documents/jsc22rpt.pdf .

Webb, M., 2002: Using the ISCCP simulator to evaluate midlatitude cloud regimes. In the Proceedings of the 2002 Atmospheric Model Intercomparison Project (AMIP) International Workshop (in press), 12-14 November 2002, Toulouse, France.

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, 1999: Discussion of the 'Transpose AMIP' Project. In Report of the Fourteenth Session of the CAS/JSC Working Group on Numerical Experimentation, CAS/JSC WGNE Report No. 14, pp. 7-8, WMO/TD-No. 964, 1999.

Xie et al., 2002: Intercomparison and evaluation of cumulus parameterizations under summertime midlatitude continental conditions. *Quart. J. Roy. Meteor. Soc*., **128**, 1095-1135.

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

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

Zhang, M.H., J.L. Lin, R.T. Cederwall, J.J. Yio, and S.C. Xie, 2000: Objective analysis of ARM IOP data: Method and sensitivity. *Mon. Weather Rev*., **129**, 295-311.

**UCRL-MI-150662-DR**