# How Can We Advance Our Weather and Climate Models as a Community?

BY ROBERT E. DICKINSON, STEPHEN E. ZEBIAK, JEFFREY L. ANDERSON, MAURICE L. BLACKMON, CECELIA DE LUCA, TIMOTHY F. HOGAN, MARK IREDELL, MING JI, RICKY B. ROOD, MAX J. SUAREZ, AND KARL E. TAYLOR

An argument is made for greater emphasis on shared infrastructure and commonality in codes and data.

As principals in a U.S. common modeling infrastructure working group, we report on directions toward a more organized approach to the building of software that underlies modeling and data analyses. An overall software infrastructure would separate the scientific and computational aspects of comprehensive climate and weather prediction models. Hence, scientists would be able to more effectively contribute to core modeling activities and could be better supported by computational scientists and computer vendors.

**AFFILIATIONS:** DICKINSON—School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia; ZEBIAK—International Research Institute for Climate Prediction, Lamont-Doherty Earth Observatory, Palisades, New York; ANDERSON—NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey; BLACKMON AND DE LUCA—National Center for Atmospheric Research, Boulder, Colorado; HOGAN—Naval Research Laboratory, Marine Meteorological Division, Monterey, California; IREDELL AND JI—NOAA/National Centers for Environmental Prediction, Boulder, Colorado; ROOD—NASA Goddard Space Flight Center, Greenbelt, Maryland; SUAREZ—Climate and Radiation Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland; TAYLOR—Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California

**CORRESPONDING AUTHOR:** Dr. Robert E. Dickinson, School of Earth and Atmospheric Sciences, Georgia Institute of Technology, 221 Bobby Dodd Way, Atlanta, GA 30332-0340
E-mail: robted@eas.gatech.edu

In final form 20 September 2001


**U.S. MODELERS SUGGEST A NEED FOR INFRASTRUCTURE.** Climate and numerical weather prediction involve computer programs that have become exceedingly complex because of the complexity both of the natural systems they describe and of the computational systems they use. The diversity of approaches toward research in these fields has ensured continuing improvements in basic understanding of their natural processes. However, a diversity of approaches to computer programs and storage of data has not always been equally beneficial. Modelers in the United States and elsewhere have begun to realize that greater uniformity of codes and better isolation of scientific issues from software issues would facilitate research. Likewise, common data standards would accelerate progress by facilitating broader exchange and analysis of model output.

Currently, individual groups and institutions implement their own versions of physical parameterizations by interpreting advances reported in the scientific literature. Such advances, however, represent the work of many people. Substantial resources and considerable expertise are needed to implement these advances in models. Also, modeling institutions typically develop modeling codes totally independently (although the old culture—with multiple groups within each major institution—has been substantially curtailed).

In August 1998, various U.S. modelers met at Camp Springs, Maryland, for a workshop sponsored by the National Science Foundation (NSF) and National Centers for Environmental Prediction (NCEP). They explored ways operational weather forecasters and climate researchers in the United States could collaborate better, and identified the need for developing infrastructure software to be common to all modeling groups. This infrastructure would not only help large modeling centers exchange technology but would also get a wider community to share research advances. The National Research Council (NRC 2001) has since reiterated this need.

**RETHINKING THE MODELING PROCESS.** Extensive communication already occurs between the modeling centers and others through workshops and meetings. However, besides these we need substantially new approaches to model code development. How can a university scientist directly improve a specific parameterization of a major center's model so that a modeling center can concentrate its resources on issues important to it and rely on colleagues elsewhere to improve other aspects of the model? Ideally, all groups would use software whose only distinct features were those dictated by particular institutional or group scientific interests or technical requirements. For example, operational numerical weather prediction (NWP) emphasizes efficient use of computational resources while research modeling facilitates innovation. Some climate research groups primarily work toward seasonal to interannual predictions, whereas others focus on anthropogenic climate change; some of the latter researchers value the synthesis of land physical and biological processes more than others do.

All models have three-dimensional geography with atmospheric components that are built about dynamical cores coupled to physical parameterizations and to land, ocean, and ice boundary conditions. The complexity, and hence intended realism, of these parameterizations and the modeling of the land, ocean, and ice have varied substantially between different modeling systems as a consequence of different objectives and resources. The latter would be less of a limitation with more widespread access to the most advanced codes.

Dynamical cores for the atmospheric model are initially developed in isolation from model physical processes. This allows well-designed tests to establish how accurately the dynamical models can recover known solutions of the equations. Likewise, most physical parameterizations are developed in a one-dimensional column representation to allow careful testing against observational data. Thus, individual scientists or small groups will continue to be the primary source of the underlying dynamics and physics of models.

By contrast, software development for use of complete coupled models involves the 3D elements of the models. The computation for individual elements is commonly distributed over different processors of massively parallel computer system. Efficient use of individual processors requires software that avoids imbalanced workloads or clogged communications channels. With the available vendor software and limited access to software specialists, modelers have difficulty achieving such efficiencies.

Yet all models require similar software. There is no more need to totally remake each model than there is a need for every modeling group to write their own FORTRAN compiler.

**INITIATIVES.** The NSF–NCEP workshop recommended that U.S. climate and NWP modelers take a lead in advancing a common modeling infrastructure, including standards, guiding principles, focusing efforts on widely required software, distributed ownership of core models, and a sharing of heritage, including documentation and journal publications. To further this objective, the workshop asked Stephen Zebiak and Robert Dickinson to cochair an ad hoc U.S. working group on common modeling infrastructure.

The working group held half a dozen meetings from October 1998 to February 2000. The community interest was remarkable. At least several dozen modelers and software specialists attended each meeting, volunteering their own time and travel resources to contribute to this effort (Table 1). Substantial planning and several prototypes were developed before further progress slowed due to the ad hoc and pro bono basis of the group.

The working group recognized that data standardization could simplify access to model output, exchange of data between modeling groups, and the transfer of data to people who deal with impacts and applications. This would simplify the development of analysis software and reduce duplication of effort, providing economy of sale in diagnosing and evaluating model performance.

Some modelers may need to use their native formats in initial production of model output. The NWP community commonly uses the GriB World Meteorological Organization (WMO) standard (e.g., DSS 2000); the climate modeling community has converged on versions of Network Common Data Form (netCDF). As a result, a single data standard thus does not appear immediately practical. The working group instead began developing and promoting a set of common metadata (i.e., quantitative and descriptive information about primary data). The working group defined this metadata standard by merging two netCDF metadata standards developed separately at the national Center for Atmospheric Research (NCAR) and the Program for Climate Model Diagnosis and Intercomparison (PCMDI)/Hadley Centre. Recently, the authors of the two standards have proceeded to merge the standards to a hybrid called "NetCDF Climate and Forecast (CF) Metadata Conventions" (Eaton et al. 2001).

**Table 1. Institutions participating in common modeling infrastructure working group.**

| |
|---|
| Center for Ocean–Land–Atmosphere Studies (COLA) |
| Geophysical Fluid Dynamics Laboratory (GFDL) |
| International Research Institute for Climate Prediction (IRI) |
| Los Alamos National Laboratory (LANL) |
| Lawrence Livermore National Laboratory (LLNL) |
| NASA/Goddard Space Flight Center (GSFC) |
| National Center for Atmospheric Research (NCAR) |
| National Centers for Environmental Prediction (NCEP) |
| Naval Research Laboratory (NRL) |
| Universities (including Colorado State University, The Florida State University, Georgia Institute of Technology, George Mason University, Massachusetts Institute of Technology, The University of Arizona, University of California, Los Angeles) |

An application program interface (API) is necessary to facilitate the use of the new standard with visualization and analysis tools and translation to and from other "native" formats. Such a data translator between CF and GriB is needed most urgently.

The working group also developed a draft set of standards for physics coding, working toward "plug compatibility" of dynamical cores and toward common couplers and common support tools such as model astronomical calculations. They set up a repository for model physics at the Goddard Space Flight Center (GSFC) that includes a single column model test bed with a standard interface to model physics packages.

The prototype software repository was very useful for enabling modeling groups to rapidly share parameterizations. It would be very desirable if the outcome of the working groups' efforts were a more robust and well-supported facility, expanded to incorporate all aspects of numerical environmental modeling.

**A SOFTWARE FRAMEWORK.** Software frameworks encourage modelers to reuse software components and design. Such frameworks have been developed previously for applications in computational continuum mechanics involving solutions of partial differential equations (e.g., Baden 1996; Brown et al. 1999). Prototype frameworks for climate modeling have also been developed at the Geophysical Fluid Dynamics Laboratory (GFDL) GSFC, and the Massachusetts Institute of Technology (MIT). These frameworks handle tasks such as gridded data decomposition and load balancing, communications, input/output, and grid transformations.

In frameworks, functionality is layered. For example, the communications methods include low-level primitive operations optimized for various computational platforms, a midlevel library of index-based communications such as transposes, and high-level routines to aggregate data for the efficient transfer of groups of data fields. Such an infrastructure may consist of two main functional areas: first, a coupling mechanism for regridding, interpolation and communication of gridded, distributed data; and second, a set of general-purpose utility routines for use of both the coupling and the application codes. In this way, two major goals of frameworks are achieved. The generic coupling mechanism facilitates the first major goal of a framework and interoperability of the integrated computational components. The utility routines include input/output, performance profiling, time management, and error handling. Together, they provide usability, the second major goal of a framework. Modelers might choose to use a small subset of the framework, or much of it, in developing their codes. Ideally, existing codes should be able to use portions of the framework (say, the coupling mechanism, or the performance profiling) without extensive rewriting.

As an outgrowth of the discussions of the working group on common modeling infrastructure, the

National Aeronautics and Space Administration (NASA) High Performance Computing and Communications Program (HPCC) developed and released a Cooperative Agreement Notice that called for proposals to develop community based software frameworks for earth system models (Fischer 2001; Steitz 2001). DeLuca (2001) reports on initial plans.

**A NEW INTERFACE WITH INNOVATION.** Admittedly, maintaining innovation in the science underlying these prediction efforts will continue to require multiple, seemingly parallel development efforts at multiple institutions. Individual research groups will continue to require multiple approaches to archiving data generated by models or required for model boundary conditions, initialization, or validation.

But a more organized approach is needed to build the software for modeling and data analysis. Such a software infrastructure would help manage the complexity inherent in both the scientific and computational aspects of comprehensive models, in part by separating the development and maintenance of these different aspects. An infrastructure would make the experiences of individual scientists, working on specific applications, broadly relevant to the field. A controlled software environment would ensure that individuals everywhere could be useful to core modeling at national labs.

A more uniform software infrastructure (including organized code sharing, common data translators, and the like) would enable computational scientists and computer vendors to provide solutions for larger markets. The interface between the scientific and computational communities should compensate for the otherwise increased complexity and reduced support for scientific applications of present-day software environments. In particular, an improved interface would facilitate the efficient use of today's massively parallel high-performance computational platforms.

## REFERENCES


Baden, S. B., 1996: Applied Computing Review (1) software infrastructure for nonuniform scientific computations on parallel processors. *Appl. Comput. Rev.,* **4,** 7–10.

Brown, D. L., W. D. Henshaw, and D. J. Quinlan, cited 1999: Overture: An object-oriented framework for solving partial differential equations on overlapping grids. [Available online at http://www.llnl.gov/CASC/Overture/henshaw/overturePublications/node2.html.]

DeLuca, C., 2001: What is the Earth System Modeling Framework? [Available online at http://www.esmf.ucar.edu.]

DSS, 2000: A guide To WMO GRIB. [Available online at http://dss.ucar.edu/dos/formats/grib/gribdoc/front_page.html.]

Eaton, B. J. Gregory, R. Drach, K. Taylor, and S. Hankin, 2001: Net CDF Climate and Forecast (CF) Metadata Conventions. [Available online at http://www.cgd.ucar.edu/cms/eaton/netcdf/CF-current.htm.]

Fischer, J., 2001: Frameworks. [Available online at http://ess.gsfc.nasa.gov/frameworks.html.]

NCR, 2001: *Panel on Improving the Effectiveness of U.S. Climate Modeling, Board on Atmospheric Sciences and Climate Commission on Geosciences, Environment, and Resources and National Research Council. Improving the Effectiveness of U.S. Climate Modeling.* National Academy Press, 128 pp.

Steitz, D. E., 2001: NASA climate modeling spurs next computing revolution. NASA press release, 1 June 2001. [Available online at http://spacelink.nasa.gov/NASA.News/NASA.News.Releases/Previous.News.Releases/01.News.Releases/01-06.News.Releases/01-06-01.Computer.Models.Study.Climate.]