Archiving Numerical Models of Biogeochemical Dynamics

PAGE 431

No adequate community-wide mechanism or set of standards currently exists to ensure the long-term reproducibility of results from numerical modeling in biogeochemical research. Various investigators maintain copies of code and supporting materials through ad hoc methods, and some models and modeling studies are not archived by any mechanism other than the published manuscripts describing specific research results. This state of affairs invites trouble.

To illustrate this situation, one of the authors of this article was recently involved in producing a state-of-the-art review on the topic of the effects of climate change on forests. The literature on this topic contains a rich array of results from dynamic models of global or regional vegetation produced by forcing these models with climate change scenarios.

These results stem from various versions of different models used with different, often elaborate climate change cases. Several papers compare model results with one another, sometimes with very different predictions under equivalent conditions.

Because the specific models and scenarios are not accessible from the literature or through any other practical mechanism, it is not possible to answer simple but fundamental questions such as, "Is the model-predicted response of vegetation more or less sensitive to a 2° warming than was predicted 20 years ago?"

Reproducibility of experimental results is a fundamental criterion for credible scientific research. Scientists working with physical and biological systems take great care to describe the methods and materials central to their experimental design, the observational environment, and other conditions that might influence experimental outcome. Traditionally, for laboratory and field studies, a description of an experimental result in the literature is not considered complete unless it is accompanied by a description of methods detailed enough that an expert practitioner could replicate the experiment.

In contrast, this same level of methodological detail is only rarely provided for in numerical simulation or modeling studies. This is true in spite of the fact that numerical experimental systems are inherently reproducible; unlike a physical experiment, which may be subject to many sources of uncontrollable variability, a valid numerical result can always be replicated exactly. This article reviews several factors responsible for the current state of affairs, and it presents a set of recommended "best practices" aimed at raising the standards for reproducibility in numerical modeling studies through the use of a dedicated archive for numerical models and modeling studies.

The lack of methodological detail in numerical modeling studies hinders not only expert practitioners attempting to understand or replicate experimental results obtained by their colleagues. It also hinders modelers revisiting their own studies some time after publication (speaking, with humility, from the experience of this article's authors). While numerical modeling results are inherently reproducible, at least two aspects of numerical models and modeling studies make it difficult to achieve reproducibility in practice.

First, the only complete description of many numerical models is the model itself, together with the computational environment in which it is executed, including specifications of input parameters and boundary conditions. Although there are certainly examples of numerical models that can be completely described in a typical published manuscript, the concern here is with the large class of numerical models that are not amenable to typical publishing methods due to the complexities of their software environments and input data structures, and to journal page limits.

This problem is often addressed by describing key algorithms and parameters in appendices to published modeling studies. This practice serves a useful purpose, highlighting the model developer's own interpretation of the critical dependencies, but such descriptions rarely represent the complete logical flow of a model.

A second practical problem for reproducibility of modeling results is related to the tendency of numerical models to evolve over time. Challenges to model predictions lead to new parameterizations, representation of new processes, and software modifications that support new boundary conditions and diagnostic output requirements. Modeling studies may refer to a model by name, including one or more standard citations, but in many cases the exact model version used for a particular study differs in significant ways from the versions used in cited studies.

Reproducibility for these models is contingent on being able to trace the evolution of model versions over time, and requires that a particular published study be tied to an identifiable and retrievable model version. To the extent that published studies can be associated explicitly with a model version, it becomes feasible in cases where only minor modifications are made to build on previous publications by referring to the previous version and to the exact modifications made for a new study.

It is clear that some additional mechanism beyond the traditional published manuscript is required to address the practical problems associated with reproducibility of numerical modeling results. One obvious solution is an electronic archive of models that can store and provide access to the model software, boundary conditions, and documentation with sufficient detail to allow an experienced practitioner to replicate key research results.

The authors propose here a system that is designed to serve two distinct purposes: First, the requirements for a long-term, stable archive for release or benchmark model versions are addressed, including the associated documentation as well as example inputs and outputs. Such versions may or may not be associated with a particular published manuscript, but they can serve as detailed, identifiable, and retrievable points of reference for subsequent model implementations.

Second, the archiving of specific model implementations that connect to a specific published research result is proposed. Model elements proposed for archiving include exact software, boundary conditions, parameterizations, and, in some cases, also analysis routines required to reproduce specific experimental results associated with a published manuscript.

In line with what has been proposed here, an implementation of these recommendations, including release and manuscript archive entries for several different models, has now been established at the Oak Ridge National Laboratory Distributed Active Archive Center for Biogeochemical Dynamics (http://daac.ornl.gov/model_intro.shtml). An additional discussion of the principles of reproducibility as applied to numerical modeling results is given by *Buckheit and Donoho* [1995].

Release or Benchmark Model Archive

Many model developers or development groups periodically make benchmark versions of their source code available to the research community through a public release. Such benchmark versions are sometimes associated with a particular publication, but in other cases the release may consist of posting the software for download on a publicly accessible Web site and circulating a notification through mailing lists.

Largely on the basis of current practices, the authors propose the following set of standards that should be met to qualify for inclusion of benchmark models in an archive:

1. A unique model name and/or version number, an archiving date, and contact information (individual or group). [Required for basic archive integrity]
2. Complete source code, including code compilation (build) utilities for at least one operating system type. The model should build without error on the target platforms, as tested by the model developers prior to archiving. [Required]
3. Example model inputs, detailed documentation describing the application of the model using the example inputs, and example outputs at a level of detail sufficient to judge whether a new installation of the model produces answers identical to the archived version. [Required]
4. Documentation of model process representation, presented as detailed commenting within the code or as a companion technical

By P. E. Thornton, R. B. Cook, B. H. Braswell, B. E. Law, W. M. Post, H. H. Shugart, B. T. Rhyne, and L. A. Hook

reference manual. [Recommended]

5. A description of model lineage, if applicable. This description preferably references earlier archived benchmark and manuscript models, indicating primary relationships between earlier archived models and the current model. [Recommended]

Manuscript Model Archive

Experimental results in a published modeling manuscript are a critical scientific contribution, but an equally important component of the scholarship is embodied in the numerical model itself. Contributions to a manuscript model archive are a way to ensure that this scholarship survives with due credit ascribed, and that it can be evaluated fairly at arbitrary points in time as new observations become available and as new hypotheses are formed.

Because the primary purpose of an entry in the manuscript model archive is to connect a specific published research result such as a table or figure with a particular model implementation [cf. *Buckheit and Donoho,* 1995], the criteria are different than those suggested for the more general purpose of archiving benchmark model versions. For a manuscript model archive entry, the authors propose the following standards for inclusion in the archive:

1. Explicitly related to at least one published manuscript, with a scientific point of contact who is also an author of the manuscript. [Required]
2. Complete source code, including build utilities for the platform on which published results were generated. [Required]
3. All model inputs required to generate key research results contained in associated publication. In cases where model inputs are themselves archived data sets, a citation to the archived data set should substitute for a copy of the input data. [Required]
4. All post-processing or analysis routines required to generate the targeted research results. For the case of analysis routines that are written within a publicly available software package (either commercial software or another archived model), the user-written routines and a definitive reference to the software package (e.g., name and version) should be provided. [Recommended]
5. Detailed instructions describing the process of model execution and analysis of model output that results in the targeted research result. [Recommended]
6. A description of model lineage that relates the manuscript model implementation to benchmark model versions and describes process representations and parameterizations that have been modified. [Recommended]

Archive Description

In order to meet the expectation of long-term stability for reference and retrieval of archived benchmark models, the authors propose the following set of standards that applies to the creation and maintenance of the archive itself.

1. Housed at an institution, such as the Oak Ridge National Laboratory Distributed Active Archive Center, with a mission that includes long-term preservation of electronic information in support of research.
2. Adequate staffing to interact with model developers and provide some degree of quality control on new benchmark model versions as they are being added to the archive (e.g., checking uniqueness of model name/version number, consistency of metadata formatting, tests of model build and execution on common platforms).
3. Provide convenient access to the archive through search and retrieval mechanisms, and maintain statistics on contributions to and retrievals from the archive over time.
4. Provide initial point of contact for technical aspects of model retrieval, build, and execution on common platforms, with tracked forwarding of research-related questions to model developers.
5. Develop evaluation criteria based on citations in the scientific literature for legacy model versions that should be targeted for archiving. Publicize the availability of the archive to encourage submission of new benchmark model versions as a standard practice among targeted community of model developers. Provide a mechanism for periodic publication of new and current archive holdings to the relevant research community.
6. Provide guidelines for the citation of models retrieved from the archive and subsequently used in a published study. These guidelines need to ensure that credit is attributed both to the model developers providing the primary scholarship, and to the center providing the archiving service.

These recommended "best practices" for archiving models will be critical for presentation of results in large research programs such as the U.S. Climate Change Science Program, and they will support accountability for research results and the long-term viability of modeling research programs.

Acknowledgments

The ideas on model archiving presented here grew from discussions at the Oak Ridge National Laboratory Distributed Active Archive Center User Working Group meetings over the past several years. We acknowledge the input of the current and ex officio User Working Group members. Support for the generation of example manuscript archive entries was provided to P. E. Thornton by NASA Grant W-19953.

References

Buckheit, J. B., and D. L. Donoho (1995), WaveLab and reproducible research, in Wavelets and Statistics, edited by A. Antoniadis and G. Oppenheim, pp. 55–81, Springer, New York.

Author Information

Peter E. Thornton, Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, Colo., E-mail: thornton@ucar.edu; Robert B. Cook, Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; E-mail: cookrb@ornl.gov; Bobby H. Braswell, Complex Systems Research Center, University of New Hampshire, Durham; E-mail: rob.braswell@unh.edu; Beverly E. Law, College of Forestry, Oregon State University, Corvallis; E-mail: bev.law@oregonstate.edu; Wilfred M. Post, Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; E-mail: postwmiii@ornl.gov; Herman H. Shugart, Department of Environmental Sciences, University of Virginia, Charlottesville; E-mail: hhs@virginia.edu; B. Timothy Rhyne, Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; E-mail: rhynebt@ornl.gov; Leslie A. Hook, Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; E-mail: hookla@ornl.gov.