# Enabling a Community Environment for Advanced Oceanic Modeling

Eric Chassignet, Scott Doney, Robert Hallberg, Dennis McGillicuddy, and James McWilliams

## Contents



## 1. Executive Summary

Ocean models are increasingly important tools for addressing a broadening and increasingly interdisciplinary range of oceanic questions and the range of expertise that needs to be embodied in the models surpasses what any one person, or even small groups of people, possesses. This White Paper calls for the National Science Foundation to support the development of a Community Environment for Advanced Ocean Modeling, addressing the Nation's physical, biogeochemical, and ecological ocean modeling and data assimilation needs. Such an Environment will serve a number of critical purposes. It will:

- Accelerate the assessment of ocean models and modeling techniques and the development of superior ocean models, both through coordinating efforts and by enabling a broader range of scientists to contribute new ideas and see them translated into widespread use.

- Facilitate the exchange of ideas and capabilities among diverse groups of scientists, particularly between oceanographic subdisciplines and beyond the bounds of direct physical interaction.

- Provide easier access to "best practice" ocean models and the documentation underlying these models for scientists, students, and other users without requiring extensive prior training.

- Minimize the challenges that the scientist may encounter when computing on high end computer platforms.

- Provide a stable financial and organizational base for the maintenance and continued development of ocean models.

The National Science Foundation's leadership in supporting the development of a Community Environment for Advanced Ocean Modeling would greatly increase the effectiveness of the diverse range of scientific studies that depend on ocean models, and speed our fundamental understanding of the ocean's role in the Earth system.

## 2. Introduction

Oceanic circulation models are essential tools for understanding, assessing, and predicting global currents, heat content, and material distributions, as well as their role in climate and the Earth system. Much of the uncertainty associated with the prediction of climate can be ascribed to an imperfect knowledge of the ocean and its mechanisms for mitigating or exacerbating changes in the atmosphere and cryosphere. Oceanic predictions rely both on the ability to initialize a model to agree with observed conditions and on the model's ability to accurately evolve this initial state.

The overarching finding of the Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure (2003) “Revolutionizing Science and Engineering Through Cyberinfrastructure” is that a new age has dawned in scientific and engineering research, pushed by continuing progress in computing, information, and communication technology; and pulled by the expanding complexity, scope, and scale of today’s research challenges. The capacity of this technology has crossed thresholds that now make possible a comprehensive “cyberinfrastructure” on which to build new types of scientific and engineering knowledge environments and organizations and to pursue research with increased efficacy. The U.S. Commission on Ocean Policy Report (2004) acknowledges the need for and promise of such an ocean cyberinfrastructure when it recommends (Rec. 28-2) that:

> “NOAA and the U.S. Navy should establish a joint ocean and coastal information management and communications program ... [that] should create a research and development component ... to generate new models and forecasts in collaboration with Ocean.IT, taking full advantage of the expertise found in academia and the private sector.”

The fact that modeling of the ocean is of interest to many national agencies (i.e., NSF, ONR, NOAA, NASA, DOE, etc.) has been recognized with the funding of several overarching ocean modeling projects under the National Oceanographic Partnership Program (NOPP), a collaboration of fifteen federal agencies that provides leadership and coordination of national oceanographic research and education initiatives. Different agencies have different modeling needs, but the Community Environment is an area where NSF can contribute substantially.

As stated in the Information Technology Infrastructure Plan to Advance Ocean Sciences (2002), community models are emerging in all aspects of oceanic research, from circulation and climate to ecosystems, sedimentation, and tectonics. There are several classes of numerical circulation models that have achieved a significant level of community management and involvement, including shared development, regular user interaction, and ready availability of software and documentation via the worldwide web. The numerical codes are typically maintained within university user groups or by government laboratories and their development primarily resulted from individual efforts, rather than a cohesive community effort. While this was appropriate in a previous era when oceanic modeling was a smaller enterprise and computer architectures were simpler, the limitations of this approach are increasingly apparent in an era when increasingly diverse demands are being placed upon oceanic models. The time has come for the development of modeling capabilities to become a coherent community effort that both systematically advances the models and supports widespread access. It is also important to state that one cannot separate the effective development of oceanic ecological and biogeochemical models from that of the physical circulation model, and these extended modeling capabilities need to be an integral part of this community effort. Moreover, testing of these models with observations requires advanced inverse methods and data assimilation techniques which must be linked to this effort from the outset.

Within the field of oceanic hydrodynamic modeling, there are several classes of numerical models that can be distinguished by their respective approaches to spatial and temporal discretization, vertical coordinate treatment, and parameterization. The choice of the vertical coordinate system in an oceanic model remains one of the most important aspects of its design (Chassignet et al., 2006). In practice both the representation and parameterization of the processes not resolved by the model grid and the general structure of the space-time discretization of the underlying model equations are often directly linked to the vertical coordinate choice. Oceanic circulation models traditionally represent the vertical dimension in a series of discrete intervals in either a depth, density, or terrain-following units, and recent model comparisons exercises have shown that the use of any single vertical coordinate representation cannot be optimal everywhere in the ocean. For example, isopycnic (density tracking) layers are best in the interior stratified circulation, levels at constant fixed depth or pressure are best used to provide high vertical resolution near the surface within the mixed layer, and the terrain-following levels are often the best choice in coastal regions with complex topography. There is now a general agreement among physical modelers that a hybrid approach to generalized vertical coordinates (i.e., that mimics different coordinate types in different regions and conditions) is highly desirable for more skillful simulations (Griffies et al., 2000).

This white paper argues for the development of a community environment for advancing and evaluating the algorithmic elements for oceanic models, focusing first on a generalized vertical coordinate but encompassing all components of the models. The paper is based on extensive previous deliberations among physical and biogeochemical modelers (e.g., ongoing development efforts at NOAA/GFDL and CCSM/NCAR/LANL and a draft plan for the Hybrid Ocean Model Environment (HOME, 2005)). Such an environment would

- Create a common code base, based on powerful software frameworks such as the Earth System Modeling Framework (ESMF), to allow a synthesis of different algorithmic elements. This code base will permit diversity while developing a common language and mechanism for absorbing and evaluating novel methodologies.

- Provide a test bed to allow for exploration of the merits of different approaches in representing the important model elements, leading to recommendations for best practices.

- Provide estimates of model uncertainty by performing, for a given configuration, ensemble calculations with a variety of algorithms, vertical-coordinate and other discretizations, parameterizations, etc.

- Include the core algorithms for evaluation of current practices in marine ecological and biogeochemical modeling.

- Make available standard data sets to facilitate comparison to observations as well as algorithm development and testing.

- Facilitate linkage with inverse methods for testing models with observations, as well as data assimilation techniques for use in prediction and in the state estimation problem.

- Encourage collaboration among model developers to accelerate the pace of designing and testing new algorithms.

- Provide rapid community access to model advancements.

It is important to make the distinction between the proposed community ocean modeling environment and a single community ocean model. The modeling environment would provide a common, interchangeable code base with minimized restrictions on the algorithms that can be contributed or selected for a specific model application. Whereas many model algorithm developers would find a single community ocean model to be stifling, a community modeling environment should dramatically invigorate the development of new and superior ocean modeling techniques. An environment will offer a much broader range of options than would be possible with a single monolithic model. This diversity of options is critical for selecting the most appropriate configuration for any particular oceanic application.

The proposed oceanic modeling environment will provide dramatically more user- and developer-friendly models, and it is an indispensable step toward a longer-term vision of modeling science and practice. There are several key ocean-science cyberinfrastructure issues that are addressed by enabling this environment:

- Consolidation of many near-duplicate oceanic models into a single paradigm embodied in the code base.

- Separation of computational platform implementation software from the design and maintenance of the model codes (though of course they are closely coupled in model performance and the environment will be built in such a manner that the computational challenges that the scientist may face are minimized).

- Development of a single, standardized interface for coupling with ecological, biogeochemical and other models (Rothstein et al., 2006).

- Possible extension to new vertical and horizontal coordinate approaches, including local grid resolution refinement (Fringer et al., 2006).

- Improvement of the ability to intercompare and share model runs, setups, initial conditions, etc.

- Improvement of the ability to evaluate models' performance by facilitating comparisons to observations.

- Accelerated development of superior oceanic models by "breeding" the best algorithmic combinations from existing models and facilitating the introduction of new techniques into a broad suite of widely used oceanic models.

- Assurance of more rapid adoption of new computational technology by the oceanic modeling community by being able to efficiently adapt the common code base.

## 3. The 10-Year Vision

The 10-year vision is to have a broad unification of physical, ecological, and biogeochemical oceanic modeling tools and practices by collecting the expertise of the current sigma-, geopotential-, and isopycnic/hybrid- vertical-coordinate models in a single open and multi-disciplinary software framework. This will allow the greatest possible flexibility for users and synergies for model developers.

The environment will promote exploration of novel modeling concepts, more rapid improvement of multi-scale physical, ecological and biogeochemical models, and a stable base for the development of new application services built around a core model framework that can be maintained at the cutting edge of the science. It will also provide a framework for experimentation and rapid implementation of improvements in the parameterization of unresolved processes in oceanic models.

The environment will furnish the capability to interchange, combine, and modify choices of vertical coordinate, physical parameterizations, numerical algorithms, parameter settings, and so on. This is in contrast with the usual single model consisting of usually a fixed set of parameterizations and algorithms, perhaps with some restricted freedom in the setting of parameters, but with very limited user options to experiment with the model architecture. It is indeed essential to maintain and extend the diversity of available algorithms. The diverse collection of techniques is the gene pool of future oceanic models, and a rich pool provides the best prospect for selecting the models that are optimal for answering specific questions about processes of interest. By comparing the performance of a rich array of configurations, the community will be able to breed oceanic models that are most skillful at representing the broad assortment of processes that are important in the simulation of a system as complicated as the ocean. It will also provide an estimate of the model uncertainty by giving an envelope of solutions resulting from different choices in numerical algorithms; vertical, horizontal, and temporal discretizations; and parameterizations.

The emergence of a common modeling environment for oceanic modeling will provide concrete benefits to the nation's need to develop the next generation of scientists. The models that are presently used for oceanic predictions or climate studies are much more complicated than is often appropriate for many pedagogical purposes or for idealized studies. As a result, many students are not exposed to the models that are used in practice. A simplified selection of the full code-base for illustrative or idealized simulations will be included along with a very user-friendly interface for the widespread use of the base code in graduate oceanographic education.

The ultimate goal of this initiative is to transform oceanic models into the "handy, graceful tools, easily and promptly applicable to any well posed scientific question, usable by anyone anywhere, and with well established uncertainty estimates" that are called for in the World Ocean Circulation Experiment (WOCE) final report (Hallberg and McWilliams, 2001).

## 4. Enabling the Environment

The 2003 report on Cyberinfrastructure clearly states that NSF must institute a broad and deep program that supports the true needs of all the science and engineering missions within NSF by committing to make the fruits of cyberinfrastructure research and development (as well as related work from other agencies and companies) available in an integrated fashion to facilitate new approaches to scientific and engineering research. The environment as described in this white paper is clearly an integrated approach that will facilitate ocean science research.

As already stated in the introduction, US oceanic modelers have already begun discussions about whether they could effectively channel the currently disparate efforts into new community oceanic modeling environments [*e.g.*, future model development at NOAA/GFDL and NCAR/LANL; white paper on HOME; etc.]. It is also very important to note that, at this point in time, core oceanic model development support is waning across the country and that manpower is in short supply. The effort invested in oceanic model development varies greatly from one place to the other and are in many places subcritical. The oceanic research community therefore needs to act in order to advance its oceanic modeling capacities and remain on the forefront of science. We are well aware of how challenging it will be to skillfully manage the personal interactions, scientific content, and code structural complexity involved in this environment. But leaving them fragmented and largely unmanaged, as is now the common practice, is not a responsible approach.

The oceanic modeling community needs to take several concrete steps in order to enable the above-described community modeling environment. The proposed steps are:

1. Evaluation and quantification of the diverse range of existing physical, ecological, and biogeochemical oceanic modeling practices.

2. Assessment of what the ocean science community really needs while ensuring excellence and diversity. This can be achieved via town-hall forums at national meetings and/or workshops.

3. Development of an implementation plan.

4. Carrying out of the implementation plan all the while ensuring that the physical, ecological and biogeochemical oceanic modeling activities become a truly coherent and multidisciplinary community effort.

In practice, enabling the environment will require close collaborations between modelers as well as software engineers. An example of such a collaboration will be the implementation in the generalized coordinate environment of generic modules for incorporating ecological and biogeochemical processes. For many applications, it will suffice to provide a user interface to specify an arbitrary number of tracers, their interactions, and boundary conditions, such that the tracers can be advected and diffused using the same numerical operators used for temperature and salinity in the physical model. Alternatively, individual-based models (IBMs) can be used to track the characteristics of individuals through time in three dimensions, facilitating assessment of the large-scale consequences of local interactions. The main purpose of the collaboration will be to ensure that the proper physical tools (i.e., numerical advection schemes compatible with the physical model) are available to construct interdisciplinary models. A similar type of collaboration will also be established with data assimilation specialists to facilitate the development of the tangent linear and adjoint of the physical model. While code generation of the adjoint is usually based on a few basic principles which permit the establishment of simple construction rules, one has to exercise care in the case of non differentiable algorithms.

Enabling the environment will also require a sustained financial and organizational base that will allow scientists to work on continuing model evolution. We envision an organizational structure that would consist of a small core of dedicated scientists that will be responsible for coordination and management of the environment while ensuring that the whole community (especially the major modeling centers) remains engaged. Community oversight will be provided via an advisory panel, open workshops, and regular reviews. It will also be important to strike the right balance between the provision of services to the community and the algorithmic development/improvement of the environment. More specifically, development of the environment will consist of the design, creation, and distribution of the code base as well as the establishment of best practices. The environment will provide the necessary interface for biogeochemical modeling and data assimilation, but it will not initially branch out into the building of specific interdisciplinary models or data assimilation techniques.

In summary, this new environment will provide:

- A unified, stable, production-level common code base using generalized vertical coordinates.

- A single framework for interfacing with biology, geochemical, or other Earth systems.

- Diversity and ability to include novel techniques that will extend the skill and applicability of oceanic models.

We hope that this white paper provides enough information to give an indication of the potential power of this type of environment and that it will be a good starting place for discussion within the National Science Foundation.

## 5. References

Chassignet, E.P., H.E. Hurlburt, O.M. Smedstad, G.R. Halliwell, A.J. Wallcraft, E.J. Metzger, B.O. Blanton, C. Lozano, D.B. Rao, P.J. Hogan, and A. Srinivasan, 2006: Generalized vertical coordinates for eddy-resolving global and coastal circulation forecasts. *Oceanography*, **19**, 118-129.

Fringer, O.B. J.C. McWillams, and R.L. Street, 2006: A new hybrid model for coastal simulations. *Oceanography*, **19**, 64-77

Griffies, S.M., C. Böning, F.O. Bryan, E.P. Chassignet, R. Gerdes, H. Hasumi, A. Hirst, A.M. Treguier, and D. Webb, 2000: Developments in circulation Climate Modelling. *Ocean Modelling*, **2**, 123-192.

Hallberg, R., and J.C. McWilliams, 2001: Objective 9: To improve numerical models for the diagnosis, simulation, and prediction of the general circulation of the circulation. *2001 U.S. WOCE Implementation Report*, **13**, U. S. WOCE Office, 60-64.

HOME, 2005: *Implementation Plan for the Hybrid Ocean Modeling Environment, ftp://oceanmodeling.rsmas.miami.edu/eric/HOME/HOME_Implementation_Plan.pdf.*

Information Technology Infrastructure Plan to Advance Ocean Sciences, 2002: *http://www.geo-prose.com/projects/pdfs/oiti_plan_lo.pdf*

Rothstein, L.M., J.J. Cullen, M. Abbott, E.P. Chassignet, K. Denman, S.C. Doney, H. Ducklow, K. Fennel, M. Follows, D. Haidvogel, E. Hoffman, D.M. Karl, J. Kindle, I. Lima, M. Maltrud, C. McClain, D. McGillicuddy, M.J. Olascoaga, Y. Spitz, J. Wiggert, and J. Yoder, 2006: Modeling Ocean Ecosystems: The PARADIGM Program. *Oceanography*, **19**, 22-51.

Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, 2003: *Revolutionizing Science and Engineering Through Cyberinfrastructure, http://www.communitytechnology.org/nsf_ci_report/report.pdf.*

U. S. Commision on Ocean Policy, 2004: *An Ocean Blueprint for the 21$^{st}$ Century, Final Report*.