POINT

Open Source Climate Model Development Is Worth It

Isaac Held

A fully open source climate-modeling effort could be of great pedagogical value and maybe even of direct scientific importance by providing a toolbox for active researchers and people new to the field.

THE CLIMATE SYSTEM is complex by any measure. The atmospheric component in a climate model, for example, is designed to simulate the distinctive turbulent flow of an ideal gas in a thin layer on a rapidly rotating sphere bounded by a complicated land/ocean/ice surface, transporting tracers that in turn modify the flow through phase changes, chemical reactions, and radiative fluxes. But despite this complexity, climate modeling is approachable in ways that many branches of modern science are not, based as it is in large part on classical physics and chemistry. The way we evaluate simulation quality isn't particularly mysterious either, given the familiarity most of us have with atmospheric and, to a lesser extent, oceanic phenomena.

Moreover, you can contribute to a climate model without mastering the entire system. Indeed, any given modeling center might have specialists on atmospheric radiative transfer, computational fluid dynamics, stratospheric ozone chemistry, ocean carbon cycling, land vegetation modeling, sea ice dynamics, and so on, as well as on the phenomenology of atmospheric and oceanic flows on the small scales that can't be explicitly resolved in global models but whose effects on resolved scales must still be accounted for. Few of these researchers are intimately familiar with the modeling system as a whole. Because it's difficult to bring all the needed expertise into a single location, most modeling groups today have a virtual flavor, with many off-site collaborators playing key roles.

This diverse set of specialists is necessary but not sufficient. A reductive approach to climate modeling has its limits, thus we need a holistic understanding to guide the development effort that will ultimately create scientifically useful climate models. Examples of the kinds of questions that arise include, "The dominant period of our El Niño is two to three years, but it should be three to five years—is the problem in the atmosphere or the ocean?" and, "The track of storms over the Atlantic Ocean in winter doesn't have a sufficient southwest-to-northeast slope, so the storms travel south of England rather than hitting Scotland. Is the source of this problem in the distribution of tropical precipitation, the simulation of the sea ice boundary, or in the model's fluid dynamical core?" A holistic understanding of the system can prevent model development from being too random a walk in the space of possible models.

So what are the goals of an open source climate-modeling effort? I've discussed elsewhere (www.gfdl.noaa.gov/bibliography/related_files/ih501.pdf) the essential role for model hierarchies if we are to understand the climate system—not only for the subjective feeling of satisfaction that this understanding provides, but to optimize our climate simulations efficiently. The different elements in this hierarchy are designed to build our

continued on p. 64

0740-7459/11/$26.00

Should Climate Models Be Open Source?

David Randall

Community-based climate model development can be facilitated by coding the model in a modular fashion, but it can cause problems because the real climate system isn't modular.

CLIMATE MODELS ARE developed by teams of experts organized under central management. The members of the modeling team itself, including its managers, are *insiders*. Their expertise includes climate science (atmospheric science, oceanography, and so on) and software engineering.

What does it mean for a climate model to be open source? At a minimum, the source code should be freely available for download by people outside the modeling team, who I will call *outsiders*. This "open source of the first kind" enables outsiders with access to sufficient computing power and savvy to compile and run the model, and to create a modified version of it.

A much greater degree of openness comes if support is provided to outsiders, and if changes created by outsiders are sometimes incorporated back into the centrally managed repository. In that case, we would have a *community model*, of which the Community Earth System Model (CESM; www.cesm.ucar.edu), managed by the US National Center for Atmospheric Research, is the best-known example. A community model is open source of the second kind, meaning that all community models are open source, but not all open source models are community models.

Simply running a model without modifying it is straightforward, but making sense of the results is not. Are the results realistic? Do they make physical sense? How or to what extent do they fit with what's in the preexisting literature? What numerical experiments could and should be performed to shed light on the key scientific questions of the day? The expertise required to answer these questions goes far beyond the technical skills needed to run the model. Making scientifically meaningful improvements to the model is even more challenging, and testing such changes is very time-consuming. In many cases, improvements to one model component uncover problems with other components.

Proposed changes to a community model must be rigorously vetted through a formal oversight process, such as the one the CESM uses. Management of community models tends to be hierarchical; for example, it can have an advisory panel, a steering committee, and a swarm of working groups. The users are distinct from the managers, but multiple levels of user authorization can exist, some more powerful than others. Voilá! A bureaucracy.

Climate models have many coupled components. From a management as well as software engineering viewpoint, the concept of plug-compatible model components, or modules, is almost irresistibly attractive. From a physical viewpoint, however, the possibility of modularity is illusory because the physical climate system itself isn't modular. In fact, coupling model components is one of the most important, subtle, and neglected aspects of model development. Here's a timely example: cumulus clouds are strongly coupled to

continued on p. 64

0740-7459/11/$26.00 © 2011 IEEE

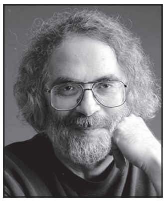
POINT

continued from p. 62

understanding of the climate system (and our models of it) in stages. Manipulating a high-end simulation is a bit like biologists trying to understand human diseases by experimenting with human models only, when E. coli to species from mice to zebra fish have proven essential to understanding human biology. However, the problem facing climate science is that nature hasn't provided us with a natural hierarchy of climate models. We need to develop this hierarchy on our own, and an open source development effort would be ideal for doing so; in fact, it's difficult to imagine how such an effort could succeed in building comprehensive models without creating a hierarchy of well-documented simpler models in the process. This hierarchy in turn would provide the material necessary for a powerful approach to educating students and scientists from other fields about climate modeling.

Many people with diverse backgrounds have formed opinions on the quality of the science of climate change and have organized into online communities to share these opinions. This level of interest isn't surprising, given the possibility of significant and long-lasting changes in our climate that would affect everyone on the planet in one way or another and the massive changes in our energy economy that might be required to avoid disruptive climate change. I'm convinced that amid this cacophony of voices and differences in style, much of it ill-informed and even mean-spirited, there's clear evidence of a reservoir of expertise in science, engineering, mathematics, and computer science that, if tapped more effectively, could have a substantial positive impact on the communication of the science to the scientifically literate public. A serious open source climate-modeling enterprise could be a valuable way to tap into this reservoir.

But could such an effort have a positive impact on the science itself? I think the answer is yes, most plausibly if the effort is aimed at developing a climate-modeling toolbox that could be used as a resource for providing specific improvements in code architecture or algorithms for existing researchers, or as a starting point for those wishing to enter the field. A toolbox perspective, rather than a vision of building a model that will somehow manage to be superior to others around the world, would also be consistent with the pedagogical imperative of building a hierarchy of models. Plus, it would be fun trying.



ISAAC HELD is a senior research scientist at NOAA's Geophysical Fluid Dynamics Laboratory and Lecturer with rank of Professor at Princeton University. His research interests include climate change, atmospheric dynamics, and geophysical turbulence. Held has a PhD from Princeton in atmospheric and oceanic sciences. He is a member of the National Academy of Sciences and a recipient of the Rossby Medal of the American Meteorological Society. Contact him at Isaac.Held@noaa.gov.

COUNTERPOINT

continued from p. 63

the turbulent boundary layer near the ground, and this coupling is crucial for the day-night cycle of precipitation over land. In many models, however, Scientist A formulates the cumulus component, Scientist B develops the boundary layer component, and the coupling between them falls through the cracks—which might explain why models have trouble simulating the day-night cycle of precipitation over land. The point is that the illusion of modularity discourages modelers from giving an appropriate amount of attention to the physical couplings among climate processes.

To ensure realistic coupling among processes, to maintain balance among components in terms of complexity and computational cost, and to provide a basis for setting goals and priorities, a climate model needs a well-designed global architecture. Despite the fact that this essay is appearing in *IEEE Software*, I'm not talking about a software architecture (although it's needed as well). Instead, I'm talking about a scientific architecture, built on our current understanding of the workings of the climate system as a whole. Creating and maintaining a scientific architecture is a critical, daunting, and easily neglected scientific function for management. This is true for any modeling team, open source or not.

Open sourcing of the second kind makes the establishment of a scientific architecture more difficult because a community model must accommodate diverse contributions from a wide range of participants, many of whom hardly know each other. The path of least resistance is Design by Committee. A subtle danger is that a model's software architecture could be mistaken for a scientific architecture.

But imagine performing a cost-benefit analysis of open source climate models of the second kind, relative to more conventional modeling approaches. Benefits would include the potential for valuable contributions from all corners of the climate science community. Costs would include the financial expense of supporting a large

RANDALL RESPONDS

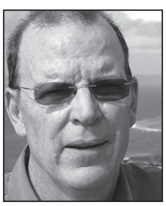
Isaac and I agree that the climate models I call "open source of the first kind" are a good thing. The Community Earth System Model, for example, has made it possible for students and other researchers to perform climate simulations, including numerical experiments designed to test ideas. This has been a strongly positive development for climate science as a whole.

We also agree on another important point: I say that a climate model needs a "global scientific architecture," and Isaac says that "a holistic understanding of the climate system is needed to guide the design effort." These are much the same points, I think.

Our strongest disagreement is on modularity. I say that modularity is not a sensible goal because the real physical system is not modular, and modularity encourages modelers to neglect or ignore important physical interactions between physical subsystems. In contrast, Isaac says that modularity is useful because it allows individual scientists to work on single components of a model without having to learn the other components.

Modularity also relates to standardized model components or modeling frameworks, and I will close with a brief comment on that topic. Standardized model components can be useful, but only if they are "small" components. Here's an analogy: the many manufacturers of the international auto industry use standard nuts, bolts, tires, and fuels because everyone benefits. But they don't use standard engines, transmissions, or body styles because those parts are key to automotive design innovations. Similarly, climate models can benefit from the use of standardized components for small and scientifically uncontroversial ingredients such as the Earth's orbital parameters, the Clausius-Clapeyron equation, various physical constants, and some types of model output. Standard grids, cloud parameterizations, or vertical coordinates would be more problematic because those parts are key to modeling innovations.

and far-flung user group, the difficulty of establishing a scientific architecture, and the operational penalties that come with bureaucracy. For open source of the second kind to be successful, the benefits must outweigh the costs.



DAVID RANDALL is a professor of atmospheric science and director of the Center for Multiscale Modeling of Atmospheric Processes at Colorado State University. His research interests include the roles of clouds and turbulence in climate science and numerical methods for climate modeling. Randall received the Meisinger Award of the American Meteorological Society and NASA's Medal for Distinguished Public Service. Contact him at randall@atmos.colostate.edu.

HELD RESPONDS

One main difference in our two perspectives is my focus on creating a hierarchy of models that would be useful pedagogically. And a pedagogically useful and easily accessible hierarchy is important for a field like climate modeling, in which so many outside of the field take an interest. I've constructed aspects of this hierarchy over the years, but I'm sure that these are far from optimal in a variety of ways, including transparency and portability, in addition to being limited by a focus on my own research interests. What I have learned from the construction process itself has been essential to whatever success I have had in my own research, and to my comfort level with using or analyzing the results from more comprehensive models constructed by large teams of researchers. I think others would develop a deeper understanding of climate modeling by participating in, or watching, the construction of a hierarchy of models.

The advantage of thinking along these lines is that, if we find (as David expects) that building a fully comprehensive modern climate model is too challenging, we might still have created something useful along the way.