

JFF2009 report Scientific management

-- Absolutely a draft outline..... merely a dumping of ideas!!! Actively being edited.

scientific management of community resources

Information management is a growing problem across all sciences and much good can come from improvements to model and data availability, interfaces, and documentation... this section is somewhat of a wish list made up by a pair of folk who don't necessarily know the state of the art

Introduction

One of the issues that appeared regularly throughout this forum is the dichotomy that currently exists between climate science and weather prediction. Both specialty areas are founded on the same scientific principles of atmospheric dynamics and physics, but the motivation and resource constraints required of vastly different time and spatial scales have created different ways of operating. To help address the matter of how to implement the previously highlighted issues, it is useful to consider a description of best practices for management of weather and climate data and modeling systems. The ultimate goal in simulating the earth system is to provide, through effective management, a flexible system of models and data infrastructure that empowers innovative use of community resources.

One of the focus groups address the need to manage competing effects of duration of simulation, ensemble size, domain size, resolution, and especially, complexity. The relationship between these is illustrated by Figure (revise and include figure in Jabonlowski's notes)

There exists a noticeable movement toward flexible design, object-oriented, modular, in data storage and retrieval, and model design. There has been work towards making individual models modular, but not modularity across models, especially and particularly involving interoperability of weather and climate models. Some of this lack of cross-model modularity is due to fundamental differences between model parameterizations. This is illustrated in the attempts by the WRF community to incorporate CAM parameterizations.

http://www.mmm.ucar.edu/projects/global_cores/ESM_phys_comp_20090205.pdf

and:

http://www.mmm.ucar.edu/projects/global_cores/ESM_plan_20080207.pdf

Another example to think about is that of Hurrell et al BAMS 2009 submitted

As scientists relying on computational tools, we start with limited resources and are forced to make decision that could pre-determine outcomes of model simulations. Resource limitations will always exist, but need to be conscious of how to best manage complexity in a flexible manner.

? manage seamless, unified, or both.

interfaces

Online interfaces to complex modeling environments allow users to "play" with models to allow a better intuitive understanding for how model process respond and interact. This could also foster collaboration between national centers and independent scientists by allowing independent scientists to run models at national centers and learn/speak the same language as these centers... err...

Users can specify control parameters such as ensemble domain size, complexity, duration, all while balancing cost. Feedback from small users to national centers. Facilitating a large number of small users to interact with large/complex models provides more eyeballs to analyze data and utilizes the power of "crowd-sourcing".

Distribution

Even the best model and data simulations are useless if they are not readily available to the scientific community at large. To that end, distribution systems are required that simplify the process of accessing these data in multiple standardized formats. In addition, because the output data sets are often extremely large (~Terabytes), it is necessary to provide tools for sub-setting in space and time of individual variables.

The ideal distribution center should be available both from an online GUI (ala Google Earth) and programmatically. The availability of such a resource is not currently limited by technological factors, but by our current lack of prioritization for such a system. Currently there is little reward to scientists for providing such interfaces to their models, as a result other researchers are required to spend a great deal of time learning new data formats and...

Documentation (metadata)

Even with improved model and data availability, the scientific utility of these resources is often limited by the lack of documentation. Many man-years are spent on the user end trying to understand poorly documented models. This is a waste of resources compared to the (relatively) few man-hours required to document these models in the first place. For example, the NCAR CCSM modeling community produces detailed tech-notes with each new version of their model, while this documentation is time consuming, it makes the CCSM model and data infinitely more useful. In contrast, for many models, documentation is limited to published papers which often have limited space.

Again, the ideal documentation could even be built into a GUI interface in which a user clicks on a model diagram and is able to drill down into model processes as well as comparing different model parameterizations.

~~fifth paragraph: some examples of best practices: notedf, dodds, thredds, this should just be included within the previous paragraphs.~~

~~where does this go? seamless vs unified modeling systems~~

~~decisions regarding high resolution deterministic "nature run" or lower resolution millennial simulations.~~

JFF2009 report section

Wiki Documents

- [mining hi-res](#)

Attachments

File	Modified
------	----------

No files shared here yet.