

EXECUTIVE SUMMARY: EARTHCUBE WORKSHOP RESULTS

Mohan Ramamurthy, Unidata/UCAR, Fuqing Zhang, Penn State U., and Russ Schumacher, Colorado State U.
17-18 December 2012

Earth Cube Workshop Title: Shaping the Development of EarthCube to Enable Advances in Data Assimilation and Ensemble Prediction

Introduction (field(s)/area(s) of interest and purpose, number of participants): Data intensive science has rapidly emerged as the Fourth Paradigm of scientific discovery after empirical, theoretical, and computational methods. This is particularly true in the area of data assimilation and ensemble prediction. This workshop was held to shape the development of EarthCube from the perspectives of the mesoscale modeling, data assimilation, and ensemble prediction communities and help in building an infrastructure that makes it easy to integrate and use observations and model output from disparate sources, support distributed modeling and data assimilation activities and share the resulting data, allow investigators to perceive linkages that today are obscured because data formats are incompatible, increase data transparency and ease-of-use, and reduce “time to science and publications.”

There were 72 registered participants, and they came from all sectors of the atmospheric science community (academia, government, and private sector), and from geographically distributed universities, research labs, and organizations that provide data to the atmospheric research community. Since the workshop was held in Boulder, CO, where NCAR and NOAA/ESRL labs are located, the workshop benefited from a large number of local participants.

SCIENCE ISSUES AND CHALLENGES

- 1. Important science drivers and challenges:** Participants identified several high-priority science questions that will be the focus of interdisciplinary efforts during the next 5-15 years (list 3 to 6).
 - What are the limits of predictability in the atmosphere? What are the sources of uncertainty/errors, and how do they feed into predictability?
 - What observations are critically needed to enhance atmospheric predictions, and where? What is the optimal configuration of the observation network?
 - What are the appropriate types, combinations, and configurations of parameterization schemes for high-resolution mesoscale models? How can the errors and biases in these parameterizations be quantified and corrected?
 - What is the optimal ensemble configuration to accurately predict the distribution of possible outcomes? How many ensemble members are needed and how should the ensembles be initialized?
 - What are the advantages and disadvantage of variational versus ensemble-based data assimilation techniques, as well as different types of hybrid approaches?
 - What are the most effective ways to post-process ensemble forecasts to achieve reliable and calibrated probabilistic predictions?
- 2. Current challenges to high-impact, interdisciplinary science:** Several themes emerged as consistent challenges faced within/across the involved discipline(s) (list 3 to 6).

- Significant barriers exist in using the data efficiently or integrating them into data assimilation or ensemble prediction systems. Today, there is too much overhead to doing research efficiently – e.g., setting up one's data and analyzing it. Rarely are there tools that really reduce this overhead.
- The scientific community lacks easy-to-use common cyberinfrastructure frameworks, data format standards, sufficient metadata for observations, and methods/tools for quality controlling observations, mining of large volumes of data, visualization, and verification.
- While many good facilities exist in this field (e.g., Unidata, DTC, and DART), they sometimes operate in silos and their activities and services are not always well coordinated or integrated.
- Lack of a central repository for finding, accessing, and using data and software.
- Significant spin-up time for students in preparing, using, processing, and analyzing data. While similar challenges exist for researchers, such problems are particularly acute for students who have a limited time before they graduate.
- Barriers to collaboration between closely linked disciplines; e.g., Atmospheric Sciences, Computer Science, Mathematics and Statistics;

TECHNICAL INFORMATION/ISSUES/CHALLENGES

1. **Desired tools, databases, etc. needed for pursuing key science questions with brief elaboration:**
 - Centralized data repositories and services that link existing and future data systems. For example, a centralized community repository could be created for data submission and sharing.
 - Advanced software, tools kits, and services for quality control, in-depth data analysis, visualization, verification, and mining of data (observational and model output). These tools and services need to be user-friendly and accessible by the whole scientific community.
 - Common data formats and frameworks for assimilation, modeling, analysis and visualization.
 - Common data assimilation framework; currently, each assimilation system uses its own framework for data I/O, processing, and running algorithms.
 - Collaboration tools, platforms, and frameworks (e.g., Wiki for data)
 - Server-side processing tools for data processing, analysis, visualization

COMMUNITY NEXT STEPS

1. **List of what your community needs to do next to move forward how it can use EarthCube to achieve those goals:**

- A pilot project on coordinated, distributed national ensemble prediction that involves universities that are interested in participating
- Developing a prototype system that links data sets/systems together, involving the most used projects like the reanalysis data sets; Develop a system that works seamlessly, and then expand to include other data sets/systems
- Continued discussion with the goal of developing a concrete plan for greater coordination of ongoing and future programs and facilities that serve the data assimilation and prediction communities, and developing a next-generation testbed facility to advance the science.
- PI meetings to leverage and expand communication, and enhance data sharing, and facilitate sustained interactions
- Entrain current undergraduate and graduate students into research and educational activities related to “big data”, ensemble prediction and data assimilation, and EarthCube, to move these initiatives forward for the future scientific workforce.
- Reach out to other geoscience communities, including climate, space weather, oceanography, hydrology, and air-quality, as well as the computer and information science communities.