Performance Metrics for GODAE

MAIN QUESTION TO ADDRESS: Reanalysis versus forecast products (different spatial and time resolution, different communities)

Internal milestones:

- accessibility of data
- standardization of quality control measures
- mechanism for feedback between data providers and assimilators
- quantifiable metrics for model and assimilation intercomparisons
- performance baseline required for demonstration products
- global nature of products
- links to user constituencies

External measures:

- availability and accessibility of products [e.g., number of bytes per day downloaded by users - .gov, .edu., .com]
- utility of products - feedback from science community, FNMOC, NOAA, and international partners, International Linkages Task Group [e.g, number of scientific papers published using GODAE products; comparison of assimilation products with observation-only products]
- quantified improvements in operational products through GODAE partnerships [predictability enhancement, cross-validation statistics]
- impact on other disciplines [e.g, number of partnerships established with e.g., biogeochemical or coastal communities]

Summary of the Easton working group (Hurlburt, Roemich, Chassignet)

A. DATA QUALITY

B. DATA DELIVERY

C. OBSERVING SYSTEM

D. END PRODUCT

- distinguish between internal and external needs

- should quantify the impact of the globally assimilated data

- high spatial and temporal coverage (mostly surface fields: SSH, SST, drifters, information on the mesoscale)

- low spatial and temporal coverage (ARGO; large scale information, water mass distribution, heat and salt content)

1) comparison to unassimilated data, i.e., Bermuda station, high resolution XBT section, tide gauges, ARGO float trajectories, choke points (Florida Straits, Drake passage, ...), ice coverage, ...

2) comparison between products, i.e., performance of one system in comparison to others (examples: western boundary current paths, transports, overturning circulation, heat and salt content, water mass formation, variability, ...)

3) forecast skills: how well is the model retaining the information from the assimilated data?

4) fit to the assimilated data, i.e., error diagnostics from the data assimilation scheme

5) sensitivity to forcing, observing system, ..., i.e., define the envelope of the model's response (examples: NCEP vs ECMWF, OSSE M-F Twin experiments, less assimilated data (SSH only for example), importance of geoid/mean SSH), ...)

6) can we quantify the added value of assimilating data? Example: Are we able to better estimate quantities such as heat/salt content?

Metrics for GODAE

Classifications of metrics:

1. Operational utility measures
2. Model climatology, mean and variability statistics, including spectral character
3. Dynamical performance
4. Forecast skill
5. Fit to assimilated data, including error statistics of the assimilation
6. Descrepancy between the model and data just prior to assimilation
7. Comparison to unassimilated data
    a. Same quantity (field)
        i. Same data type
        ii. Different data type
    b. Ability to constrain a sparsely observed field or a field for which the data is not assimilated
8. Real-time metrics vs metrics applied to applications using historical data
9. Metrics measuring the ability to represent oceanic features vs point by point statistics
10. Observing system assessment vs data-assimilative model assessment
11. Assessment of data impact, atmospheric and oceanographic

Steve Piotrowicz's letter

Hello,

The subject of measuring performance arose at last week's GODAE meeting. Many of you are aware that measuring the performance of Government has become a major focus in the United States. Measuring performance is now legislatively mandated by the "Government Performance and Results Act of 1993" (referred to as the GPRA). It requires all Agencies in the U.S. Government to develop Strategic Plans, Performance Plans, and evaluate their performance. These plans, and the evlauation process, are now part of the budget process. Developing Strategic Plans and Performance Measures for research has been quite a learning process for those Agencies supporting research. In my opinion, we are still groping for solutions, however, the GPRA is now a management tool in the budget process that we must address. Discussions regarding the GPRA have appeared in Science Magazine, been raised at all levels of the Federal Government, and amongst University Research Managers. There is no single solution.

In NOAA, each component of our Strategic Plan approaches the issue differently. In some cases you can be quantitative, in some cases you cannot. I have been involved with the issue of developing performance measures for climate. NOAA's performance is audited every year by a commercial accounting firm and we have been somewhat successful in our approaches with them in terms of evaluating long-term research.

The National Academy of Sciences has been looking at the issue and has recently published two reports on the subject. Both of these reports address the crucial issue of the long-term nature of research. They are both available for viewing on the National Academy of Sciences web page but they cannot be downloaded (they are for sale).

The most recent of these reports is entitled "Evaluating Federal Research Programs: Research and the Government Performance and Results Act (1999)" published by the Committee on Science, Engineering and Public Policy. The Panel was made up of Academy members, very senior industry managers, etc. I have not found it very useful. About the only conclusion or any importance is that applied and basic research should have different evaluation criteria. This is actually is very applicable (my opinion) to GODAE. The URL for this report is: http://www.nap.edu/readingroom/enter2.cgi?0309064309.html Click on "Image Version" to read it.

The other report (I feel) is quite useful. It is entitled "Assessing the Value of Research in the Chemical Sciences: Report of a Workshop (1998)" and was published by the Chemical Sciences Roundtable of the Board on Chemical Sciences and Technology. I participated in the Workshop that developed this report. The Workshop consisted of Government, Industry, and Academic research managers. I think the presentations by industry (for example, IBM and Lucent Technologies) and the historical analysis of DuPont illustrate the range of approaches. I found the concept of the "Technology Value Pyramid" (Chapter 5) quite helpful. I would also suggest you look at the U.S. Department of Energy's approach (Chapter 10). The

URL is: http://www.nap.edu/readingroom/enter2.cgi?0309061393.html Click on the image version to read it.

You can order the reports as well.

Thank you, Steve

Stephen R. Piotrowicz NOAA/OAR/ERL - R/E SSMC-3, Rm. 11411 1315 East-West Highway Silver Spring, Maryland 20910 Voice: 301-713-2465 Ext. 124 Facsimile: 301-713-4023 E-mail: Steve.Piotrowicz@noaa.gov