And now for something completely different …

(data management?)
HYCOM Data Management & Services

Ashwanth Srinivasan (RSMAS)
Steve Hankin (PMEL)

A community of contributors, including …
Peter Cornillon, OPeNDAP PI (URI)
Mike Clancy, US GODAE Server (FNMOC)

IPRC, Unidata, MBARI, COLA, DOE/PCMDI,
NOAA/GFDL&CDC&PFEL

Jon Callahan, Joe Mclean, Kevin O’Brien
(U.Washington/JISAO)
Roland Schweitzer, Yonghua Wei (contractors)
Ansley Manke (NOAA/PMEL)
James Gallagher, Dan Holloway, et. al. (URI)
Phil Sharfstein, Dave Dimitriou (FNMOC)
HYCOM Data Management & Services

Part I. The Requirements, the Vision, the Tools
Steve Hankin, NOAA/PMEL

Part II. Current System and Capabilities
Ashwanth Srinivasan, RSMAS

Part IIb. - Product serving activities at the IPRC
Peter Hacker

Part III. Plenary discussion: goals and priorities
Outline

Part I.
• Quick Review
• Data Management Requirements
• Around the corner (FDS)

Part II.
  Current System and Capabilities

Part III.
  Plenary data management priorities
Live Access Server (LAS)

Quick Review

Web access to model outputs and observations from distributed sites
Compare variables from different models and data products.
Compute simple analyses
Ocean obs via LAS (incl. custom “constraints”)
GTS Ocean obs on US GODAE Server

Live Access to USGODAE Data

Select your desired view (geometry of output) and then set the 4-D region (for lat-long-time) and a variable.

Select view: Latitude-longitude
Select output: Surface values plot (GIS)
Select region: Full Region

Select time range: 24-Oct-2004 to 25-Oct-2004

Apply: Age of Observation

HYCOM Nat'l Meeting
Steve Hankin
LAS -- an “Information Product Server”

- XML Metadata contains the “intelligence”
- Back end applications do the real work
- OPeNDAP provides remote data access
OPeNDAP: network access to data and “semantic metadata”
“Scripted” access to LAS

Query available data sets:

> lasls http://cpu/LAS

Query variables in data set “model_1”:

> lasls http://cpu/LAS model_1

Query space-time domain:

> lasls http://cpu/LAS model_1 sst

Request a subset of data as a file: ("asc" for ASCII format)

> lasget -x 20:60 -y 20:60 -t 11-Dec-2000 -f asc
http://cpu/LAS model_1 sst
HYCOM information & data portal

Information access

Uniform data access

Live Access Server
Outline

Part I.
• Quick Overview (LAS & OPeNDAP)
• Data Management Requirements
• Around the corner (FDS)

Part II.
  Current System and Capabilities

Part III.
  Plenary data management priorities
Our goal:

- Create a framework for sharing data that permits HYCOM modeling projects (and others) to subset, browse, analyze, and inter-compare model outputs and compare to observations.
Why is this important?

• Facilitate collaboration between partners (especially basin scale ➔ coastal models)

• Facilitate model validation

• Reach other users; increase the visibility of the project
Outline

Part I.
• **Quick Overview (LAS & OPeNDAP)**
• **Data Management Requirements**
• **Around the corner**

Part II.
  - Current System and Capabilities

Part III.
  - Plenary data management priorities
FDS
“Ferret Data Server”

• OPeNDAP output for …
  – any variable available from LAS
  – with uniform metadata standards

• Built on a Java framework from COLA
  (called ‘Anagram’)
FDS

• We can configure an LAS to present HYCOM output variables, regridded to the Levitus grid (2001) …
On-the-fly regridding from curvilinear to rectilinear coordinates
Similarly from hybrid-\(Z\) to rectilinear
FDS

- The rectilinear-grid “view” of that data is now available via OPeNDAP for a number of purposes …
HYCOM data at the desktop
(for Matlab, IDL, Ferret, GrADS, …)

The remote dataset is just a “filename”
http://server/FDS/my_HYCOM_dataset

The dataset will be exactly as LAS presents it.
Collaborating “sister” servers

Comparison between datasets becomes straightforward:
Servers request regridded data from one another via FDS.
HYCOM data at the desktop
(for Matlab, IDL, Ferret, GrADS, …)

You can ask FDS to perform simple analyses:

Vertical average of variable “TEMP”

\[ \text{http://server/FDS/}_{\text{expr}}\{\text{my_output}\}{\text{Tave=TEMP[Z=AVE]}} \]

(reduces data volumes transferred over the net)
FDS

• FDS employs “delayed analysis”
  – It appears that the whole dataset has been transformed
  – In reality the selected subset is transformed on-demand
• Arbitrary regridding on-demand, too (with cautions)
A batch job has been submitted to create the data subset you have requested.

The file you have requested will be available in 1 hour(s). Please click on the following URL after that time to receive the file:

ftp://ferret.pmel.noaa.gov/special_request/lasBatch/fb30c1b12b08e81479409464ba3af2.nc

This file will only be available for 24 hours after created.

If you would like to be notified when the file is ready, please enter your email address in the form below:

[Submit Email]
Outline

Part I.
- Quick Overview (LAS & OPeNDAP)
- Data Management Requirements
- Around the corner (FDS)

Part II.
Current System and Capabilities
Ashwanth Srinivasan

Part III.
Plenary data management priorities
Outline

Part I.
• Quick Overview (LAS & OPeNDAP)
• Data Management Requirements
• Around the corner (FDS)

Part II.
Current System and Capabilities
Ashwanth Srinivasan

Part III.
Plenary: data management priorities
Plenary Discussion: Data Management Priorities

- Immediate priorities
  - To make HYCOM basin-scale outputs readily usable by coastal modeling groups
    - Curvilinear? hybrid-Z? rectilinear?
    - Formats (CF ....)? Format for finite element modeling sites?
  - How much data kept on-line ?
  - Are we succeeding ?
  - How to gather feedback ?
Plenary Discussion: Data Management Priorities

• 2nd tier priorities
  – A “data portal” for all project participants?
    • Basin-scale and regional data, model outputs, and descriptions
    • How can we make it easy/tempting/compelling to get ALL models on-line?
  – Other suggestions? How else can we facilitate coastal modeling activities?
    • Who and what specific needs?
Plenary Discussion: Data Management Priorities

• ‘Next’ Priorities
  – What reference data sources for comparisons?
    • Moored time series? GTS surface obs? Climatologies? Met models? Other ocean models?
  – Model-data comparison techniques?
    • Sparse data techniques? Dense data techniques? Error fields? Correlations?
  – Comparison between neighboring domains?
  – Ensemble comparison techniques?
  – Metrics useful for comparison?
Plenary Discussion:
Data Management Priorities

• ‘Next’ Priorities
  – “3 time axis” data management
  – Special user interfaces? Education? GIS-style?
  – Are there other applications to make compatible with the HYCOM framework?

  – Other suggestions?
    (Break out of the box!)
Plenary Discussion: Data Management Priorities

“Technical” priorities

- Visualizations of time series, profiles, sections, ... from curvilinear data? (done)
- Graphics on ‘native’ curvilinear coordinates?
- Automate synchronization of Web portal with model outputs
- Higher efficiency storage techniques (e.g. HDF with “chunking” and compression)?
Plenary Discussion: Data Management Priorities

• Suggestions:
  – .push capability?
  – .LAS production of native HYCOM format
  – outputs available daily?
  – Mercator projection of maps?
  – How to encourage regional modeling sites to make data available (via OPeNDAP, LAS, FTP, or other)?