# Storage Challenges in Ocean Modeling

Symposium of the Center for Network and Storage Enabled Collaborative Computer Science

Brian K. Arbic, University of Michigan

May 19, 2017

# Outline

- Animations
- My background working in Africa (it's relevant; you'll see…)
- Motivation for ocean modeling
- Scale of storage problem
- Example science results
- Examples of network- and storage-enabled collaborations
- Relevance of network- and storage-enabled collaborations for science in Africa

UNIVERSITY OF MICHIGAN

# Animations

# Joseph Ansong and Brian Arbic in 1992


UNIVERSITY OF MICHIGAN

# Motivation: Why model the ocean?

- Inherent fascination
- Complement to in-situ and satellite measurements
- Impact on atmosphere on wide range of time scales
- National security
- Fisheries management
- Impacts of sea level rise, storm surges, etc. on coastal towns and cities

UNIVERSITY OF MICHIGAN

# The scale of the storage problem

- Ocean variability spans spatial scales from ~1 mm (the dissipation scale) up to basin scales.
- The number of spatial degrees of freedom is about $1.33*10^{18}$ $m^3$ (volume of the ocean)/1 $mm^3$ ~ $1.33*10^{27}$
- Ocean variability spans timescales from ~1 ms (acoustic timescales) to 4.6 billion years (the age of the earth)→ about $1.4*10^{20}$ temporal degrees of freedom
- There are seven dynamical variables in ocean models
  - Three components of velocity (u,v,w)
  - Temperature T, Salinity S, Density ρ
  - Pressure p

# The scale of the storage problem

- An “ultimate” computer that could simulate all of these scales would require ~5 * $10^{33}$ petabytes of storage.

UNIVERSITY OF MICHIGAN

- That’s assuming single precision output.

- CPU and memory demands would also be considerable.

UNIVERSITY OF MICHIGAN

# Some less fanciful examples

- HYCOM (HYbrid Coordinate Ocean Model) and MITgcm (MIT general circulation model) output is being used by the US Navy and NASA for operational and satellite-planning applications.
- These two simulations are unique in that they contain tidal and atmospheric forcing concurrently.
- High-resolution HYCOM experiments:
  - One year of hourly output at ~4 km, 41 vertical layer resolution is ~0.5 petabytes
- High-resolution MITgcm experiments:
  - One year of hourly output at ~2 km, 90 vertical level resolution is ~3 petabytes

UNIVERSITY OF MICHIGAN

# Another twist: Ensembles

- Due to chaos, ensembles are often used to forecast the ocean.
- The current Navy prediction system requires ~260 TB of model output to be stored every day, just to launch the next forecast.
- Problem will get worse when horizontal and vertical resolution is increased.

UNIVERSITY OF MICHIGAN

# Example science results…

UNIVERSITY OF MICHIGAN

# Comparison of barotropic tides in early HYCOM solutions to satellite altimeter-constrained TPXO model (Shriver et al. 2012)



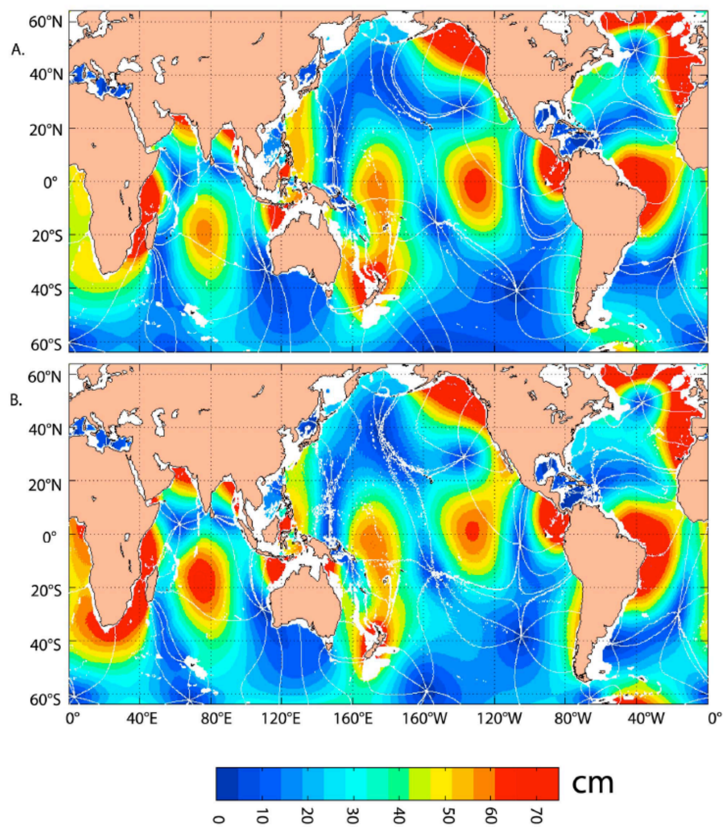



**Figure 2.** Amplitude (cm) of $M_2$ surface tidal elevation in (a) TPXO7.2 (an update to that described by *Egbert et al.* [1994]), a barotropic tide model constrained by satellite altimetry, and (b) HYCOM simulations in which the tide is unconstrained by satellite altimetry. Lines of constant phase plotted every 45° in Figures 2a and 2b are overlaid in white.

UNIVERSITY OF MICHIGAN

# Comparison of internal tides in early HYCOM solutions to along-track altimetry (Shriver et al. 2012)

024 SHRIVER ET AL.: BAROTROPIC AND INTERNAL TIDES IN HYCOM C1002

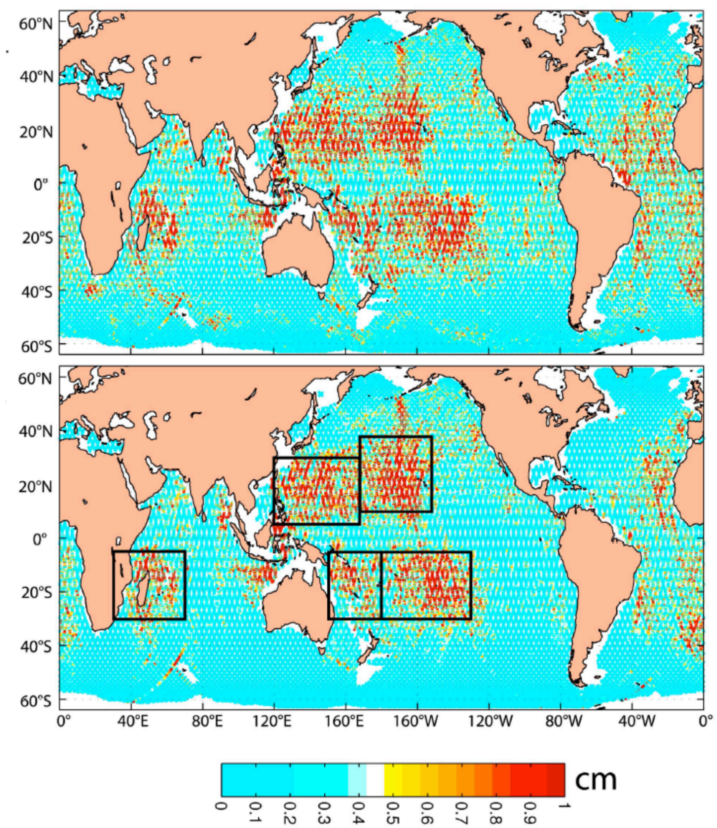



**Figure 7.** The $M_2$ internal tide amplitude from the (a) altimetric-based and (b) HYCOM tidal analyses. The five subregions denoted by black boxes in Figure 7b are used to compute the area-averaged amplitudes in Table 2.

UNIVERSITY OF MICHIGAN

# Impact of damping on low-mode internal tides (Ansong et al. 2015)

AGU Journal of Geophysical Research: Oceans 10.1002/2015JC010998

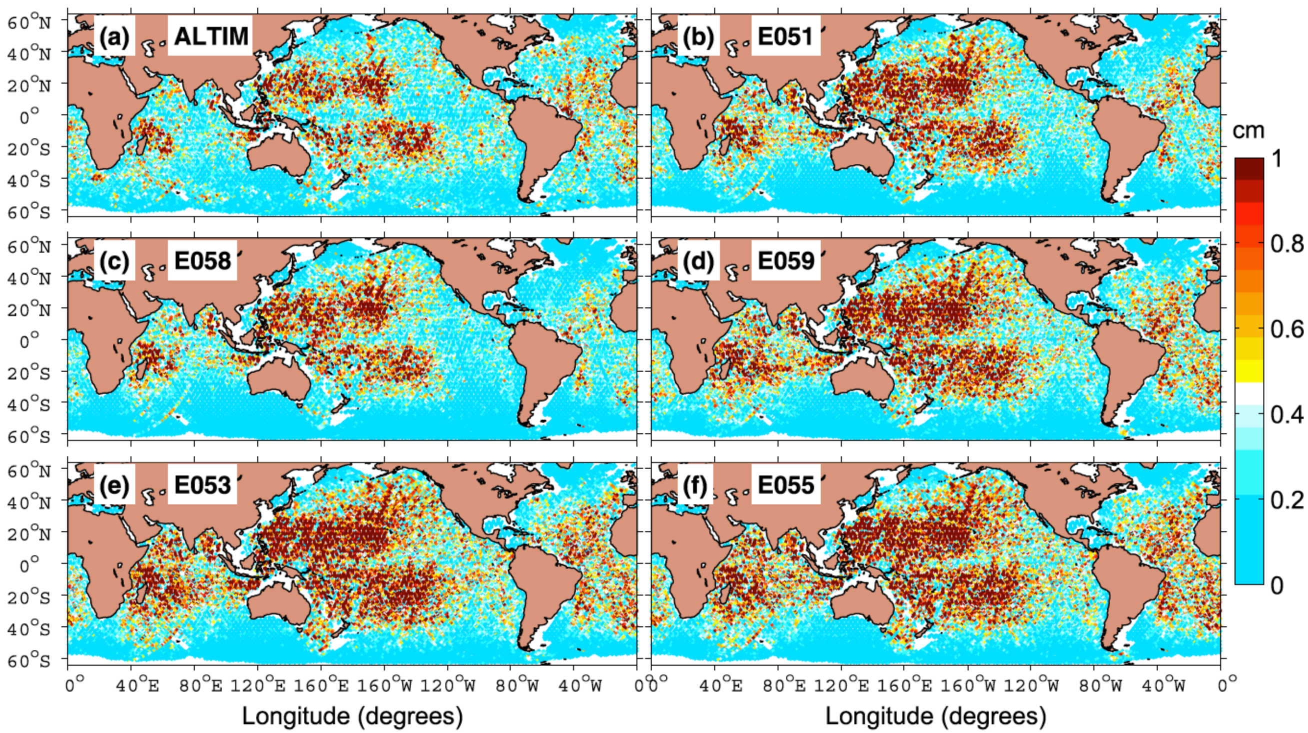
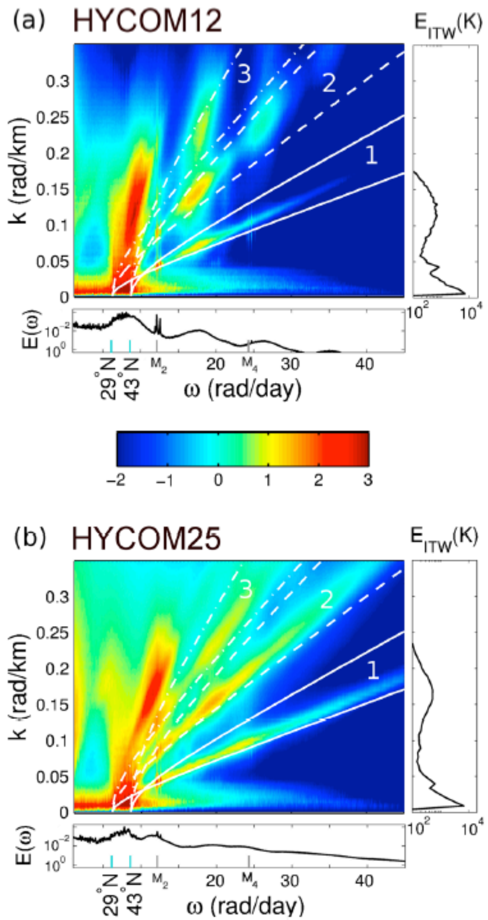


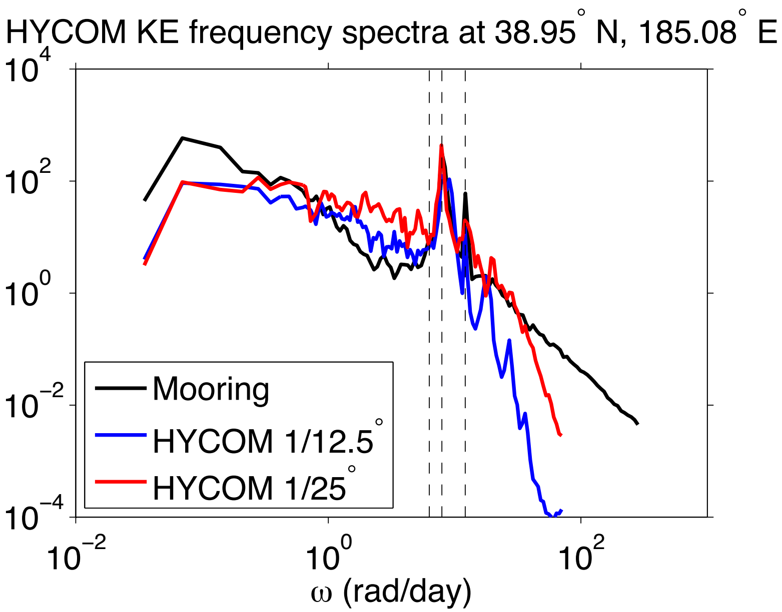

**Figure 5.** Amplitude (cm) of $M_2$ internal tide in (a) along-track altimeter-based analyses, and in HYCOM simulations (b) E051; with wave drag (scale factor = 0.5) applied to the bottom flow, (c) E058; with wave drag (scale factor = 1.0) applied to the bottom flow, (d) E059; with wave drag (scale factor = 1.0) applied to only the barotropic flow, (e) E053; without wave drag, (f) E055; without wave drag but with quadratic bottom friction increased by about 100 times along the continental shelves. The amplitudes of the HYCOM simulations are computed from 3 months of SSH output.

# Demonstration of an internal gravity wave spectrum in HYCOM in North Pacific region (Müller et al. 2015)

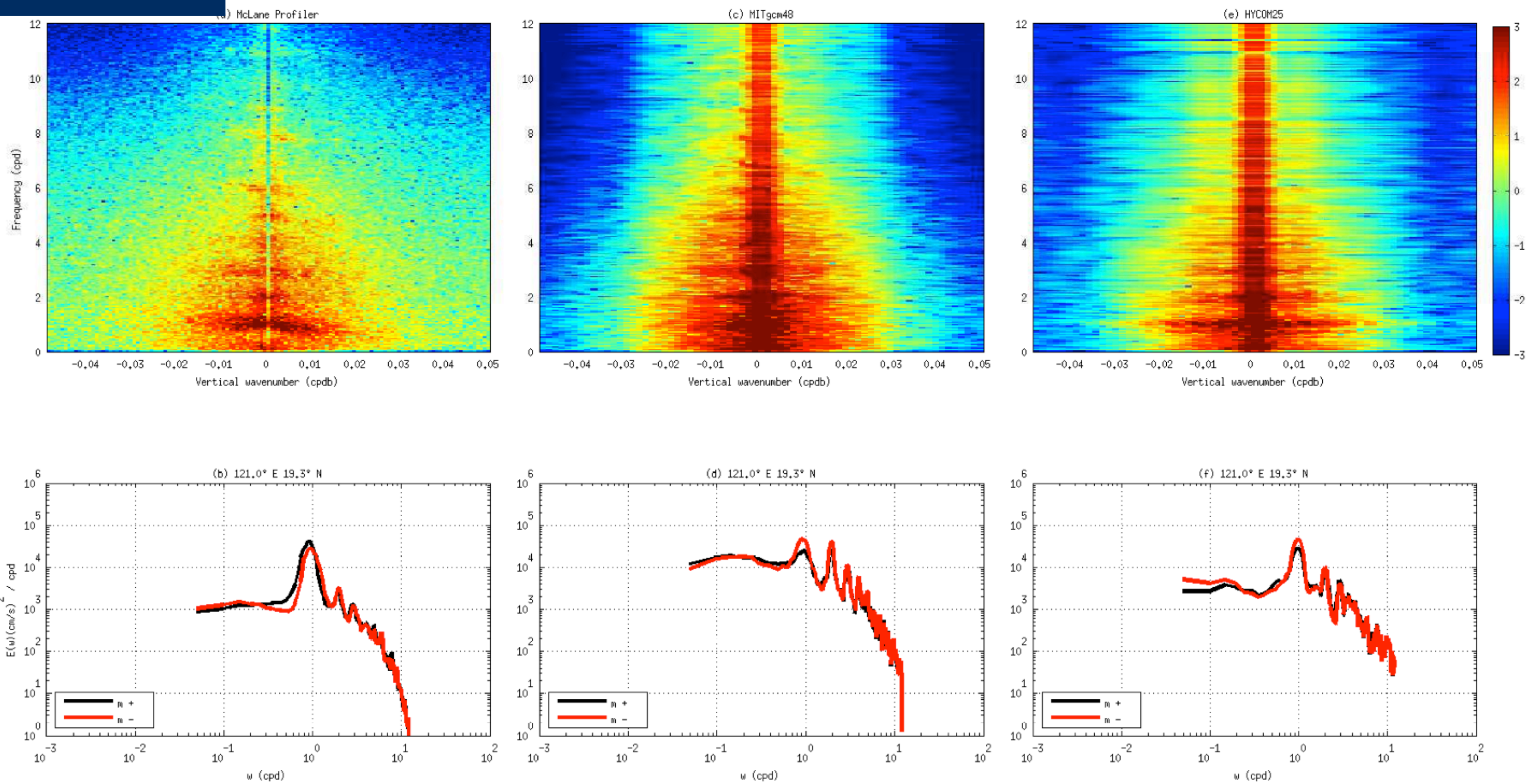
Kinetic energy frequency-horizontal wavenumber spectrum




Kinetic energy frequency spectrum



# Vertical wavenumber-frequency spectrum in observations vs. MITgcm vs. HYCOM (Ansong et al., in preparation)



- My own students and postdocs, as well as many collaborators across the world, want access to these simulations.
- The Navy simulation output is held on restricted access computers.

- How does one solve this problem of network-enabled collaborations?

UNIVERSITY OF MICHIGAN

# Examples of network- and storage-enabled collaborations

- UM students/postdocs obtain security clearance→Analyze HYCOM output on Navy computers
- UM students/postdocs obtain account on NASA machines→Analyze MITgcm output on NASA computers

UNIVERSITY OF MICHIGAN

# Examples of network- and storage-enabled collaborations

- Collaborate with scientists at SciNet→Run and store MITgcm simulations on their supercomputer
- HYCOM output→OSiRIS→The world
  - Collaborators on Navy projects
  - Collaborators in Germany
  - Collaborators from SRI in Ann Arbor

UNIVERSITY OF MICHIGAN

# Relevance for science in Africa

- Marine issues are important for every continent.
- The global ocean needs to be measured globally.
- But there aren't many oceanographers in Africa, and African scientists are not well represented in global oceanography conferences.

# Relevance for science in Africa

- I know of only one large computer cluster in sub-Saharan Africa (outside of South Africa).
- How can the brainpower of ~1/6$^{th}$ of the world's population be harnessed to help us solve scientific problems that require large supercomputers and storage silos?

- ANSWER: Network- and storage-enabled collaboration!

UNIVERSITY OF MICHIGAN

# Relevance for science in Africa

- Joseph Ansong will return to Ghana as a faculty member in mid-July.
- Access to cutting-edge storage facilities would allow his students to write publishable papers based on analysis of cutting-edge simulations, at minimal cost.

UNIVERSITY OF MICHIGAN

# Summary

- Ocean modeling is a big problem.
- Storage is a big problem for ocean modeling.
- Network- and storage-enabled collaborations are of central importance to the ocean modeling community.
- Network- and storage-enabled collaborations would allow African scientists to participate in the global scientific enterprise at an affordable cost.

UNIVERSITY OF MICHIGAN