“Revealing and examining” the tempestuous Global Ocean through a multi-petabyte virtual ocean archive, that is “powered” by emerging network and storage enabled computational technology and thinking.

Chris Hill,

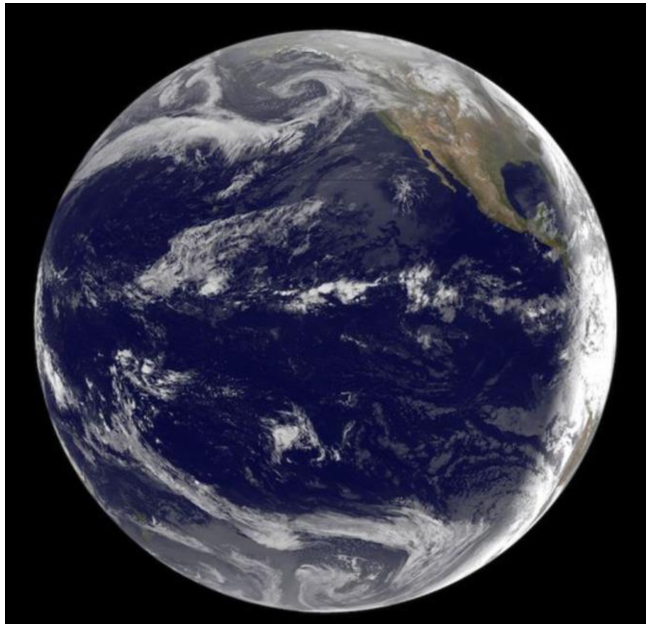
Principal Researcher, MIT

Director of Research Computing, MIT

This talk

- Look at some novel ocean modeling work that is leveraging “storage and network” thinking and pushing the limits of computational ocean science
  - Background on ocean and ocean modeling
  - Describe modeling activity – that is one of the largest ocean computations undertaken.
  - Show some examples - that illustrate role of network and storage.
  - Describe some of the network and storage infrastructure pieces.

The Oceans – why we care in general



- covers 70% of Earth
- no past ocean, possibly no life?
- imagine changing the present ocean,
  - a lot different atmosphere (marine-microbial cycles are central to carbon cycle e.g. deep ocean stores >1000 times carbon of atmosphere, oxygen, nitrogen cycles etc…)
  - a potentially different climate. Ocean carries solar heat poleward, modulates mid-latitude temperate zones etc…, supports hydrological cycle etc..
  - and there's sea-level of course

No environment = no startups, no technology transfer etc….

Ocean math

- Basic physical equations are relatively well established. Simple, approximate, models exist (for physics, chemistry and biology)
  - Navier-Stokes (19C), Bousinessq fluid. RHS (tides, winds, heating, cooling, precip, evap, unresolved turbulence)

  - Well known properties
    - Rotating frame on sphere ➔ western boundary currents e.g. Gulf Stream (Stommel 48).
    - El Nino/La Nina (Cane/Zebiak, 77)
    - Wind driven and density driven elements.

Basic fluid equations (on a rotating surface)

$$\frac{D\vec{\mathbf{v}}_h}{Dt} + \left(2\vec{\Omega} \times \vec{\mathbf{v}}\right)_h + \nabla_h \phi = \boxed{\mathcal{F}_{\vec{\mathbf{v}}_h}}$$

$$\frac{D\dot{r}}{Dt} + \hat{k} \cdot \left(2\vec{\Omega} \times \vec{\mathbf{v}}\right) + \frac{\partial \phi}{\partial r} + b = \boxed{\mathcal{F}_{\dot{r}}}$$

momentum (horizontal/geopotential and vertical)

$$\nabla_h \cdot \vec{\mathbf{v}}_h + \frac{\partial \dot{r}}{\partial r} = 0$$

momentum

$$b = b(\theta, S, r)$$

density/buoyancy

$$\frac{D\theta}{Dt} = \boxed{\mathcal{Q}_\theta}$$

temperature

$$\frac{DS}{Dt} = \boxed{\mathcal{Q}_S}$$

salt(s)

$$\frac{DC_i}{Dt} = \boxed{Q_c}$$

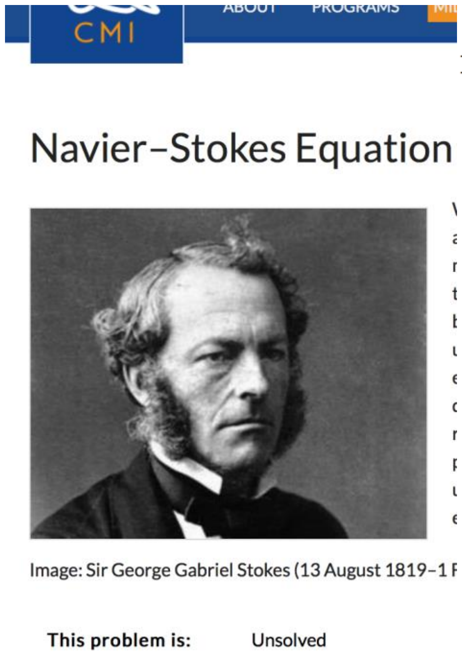
dynamically inactive chem/bio

Non-linear system, turbulent, stratified.

$\boxed{}$ RHS (tides, winds, heating, cooling, precip, evap, unresolved turbulence, chem, bio…)

Ocean modeling

- We don’t know “solutions” to equations that describe the ocean (even after 200years…)



EXISTENCE AND SMOOTHNESS OF THE NAVIER–STOKES EQUATION

CHARLES L. FEFFERMAN

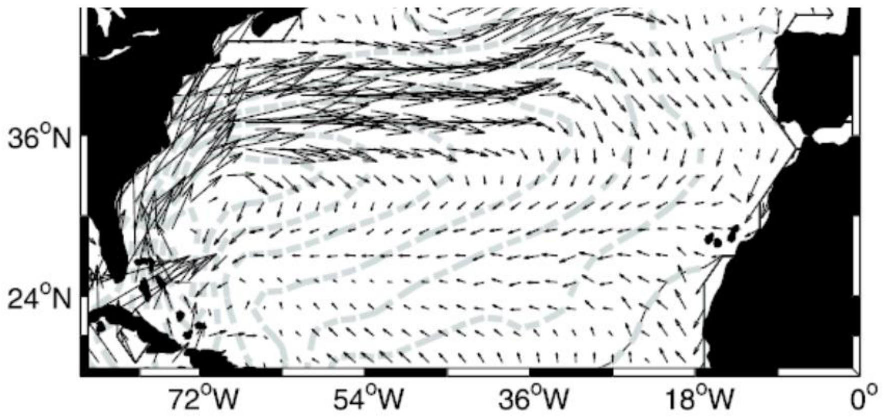
$1M prize….

- so computing can help
- (we also don’t know exact equations for other pieces, e.g. biology)

- Complexity/dimensionality emerge when we want to computationally look **in detail** at local←→global scale interactions, understand how small scale processes balance out overall behavior (e.g. maintaining stratification, nutrient and chemical gradients – turbulence closures – microbial life)
  - network and storage enabling ideas can help this more tractable, and broaden audience and impact.
- recently we have been exploring applying this sort of approach with one of the most highly resolved global ocean (and sea-ice) dynamics models to date. Our model uses large scale parallel computing (30,000 - 100,000 cores), and solves a discrete form of the ocean equations (http://mitgcm.org)

Brief historical context

First large-scale ocean circulation models were developed about 50 years ago.



A numerical investigation of the oceanic general circulation

By KIRK BRYAN and MICHAEL D. COX, *Geophysical Fluid Dynamics Laboratory, Environmental Science Services Administration, Washington, D.C.*

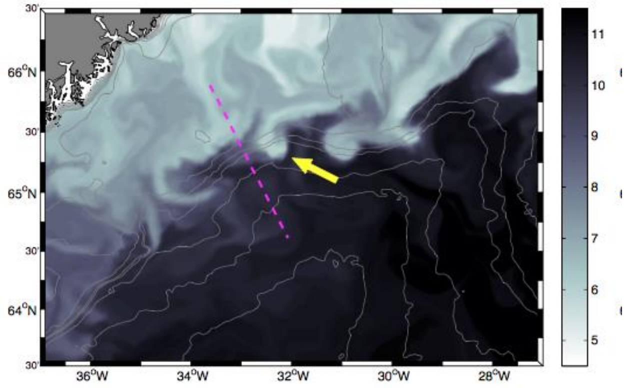
(Manuscript received October 5, 1965)

For a long time (40 years) models were accurate depictions of some gross features, but were easy to distinguish from observations.

Around 2010 some large region models began to have enough resolution (about 1km) and processes that they could compare quite well with observations ("*Turing test*").

Hydrostatic and non-hydrostatic simulations of dense waters cascading off a shelf: the East Greenland case

Marcello G. Magaldi[a,b], Thomas W. N. Haine[b]



Around same time a more complete set of observing systems began to come online (remote sensing and in-situ)

A network and storage enabled perspective.

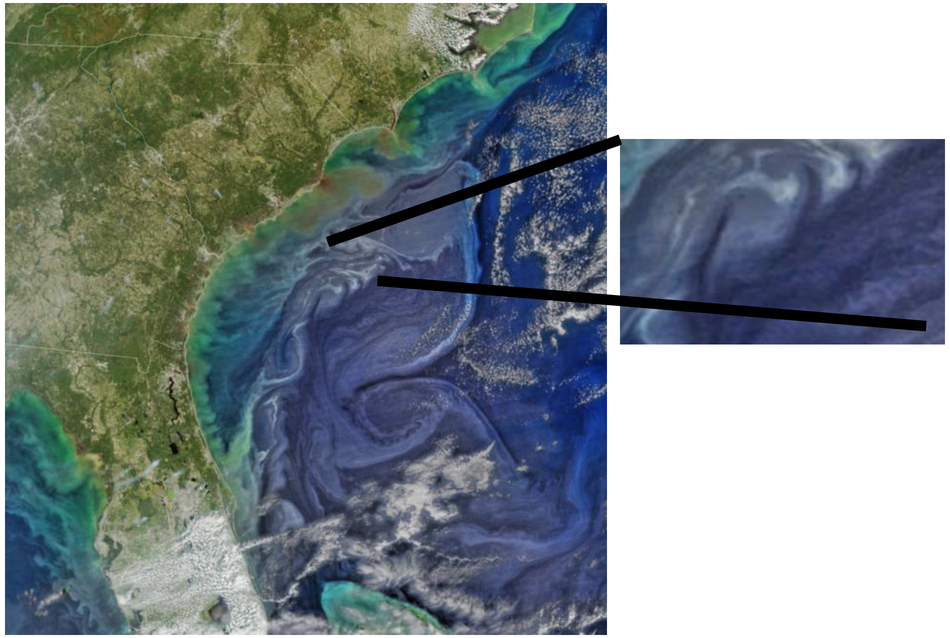
- Global ocean modeling at 1km resolution, 90 vertical levels
  - multi-year and short term free simulations. 30,000 – 100,000 x86 cores at NASA Ames.
  - initialized from data assimilated estimates.
  - driven with observational fields.
  - sub-models at 250m and below (ultimately non-hydrostatic dynamics, wave breaking, impacts on mixing and on nutrient availability and ecosystem dynamics).

- Network and storage enabling
  - powered by emerging “network and storage” enabled tech to support high-spatial and temporal frequency recording for analysis.
  - enables saving of full state every hour ➔ ($O(10^{15})$ numbers per year)
  - fully open and shared realistic, virtual ocean for a wide range of studies.

- Status
  - One+ year “proof of concept”, ~3PiB archive created.
  - Initial analyses, significantly closer to passing some “Turing tests”.
  - Sharing infrastructure partially in place, maturing all the time (including under the OSIRIS and NESE projects).

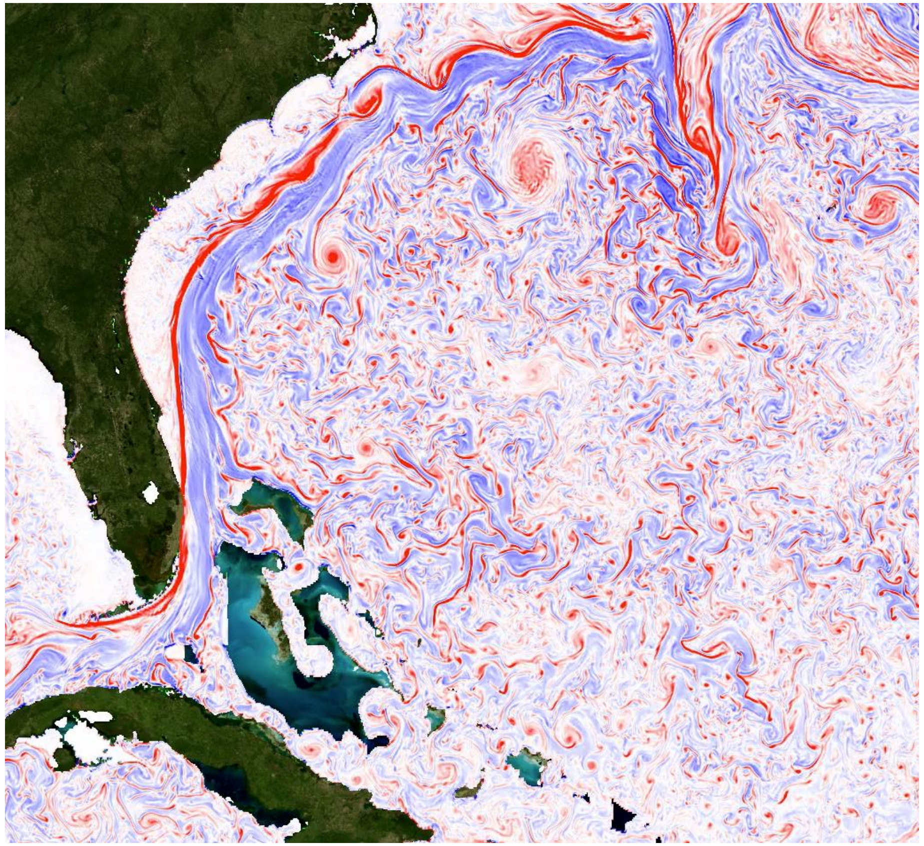
Some examples - ocean sub-mesoscale dynamics.

- Signatures of vorticity in phytoplankton/color (satellites)



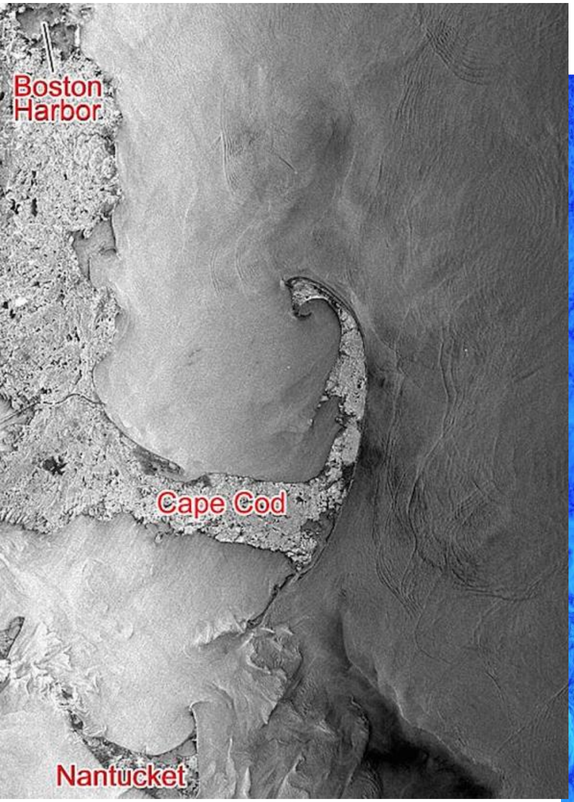
Enhanced ocean color from space is one way to discern coherent vortex structures. The size and character aligns well with numerical vorticity field.

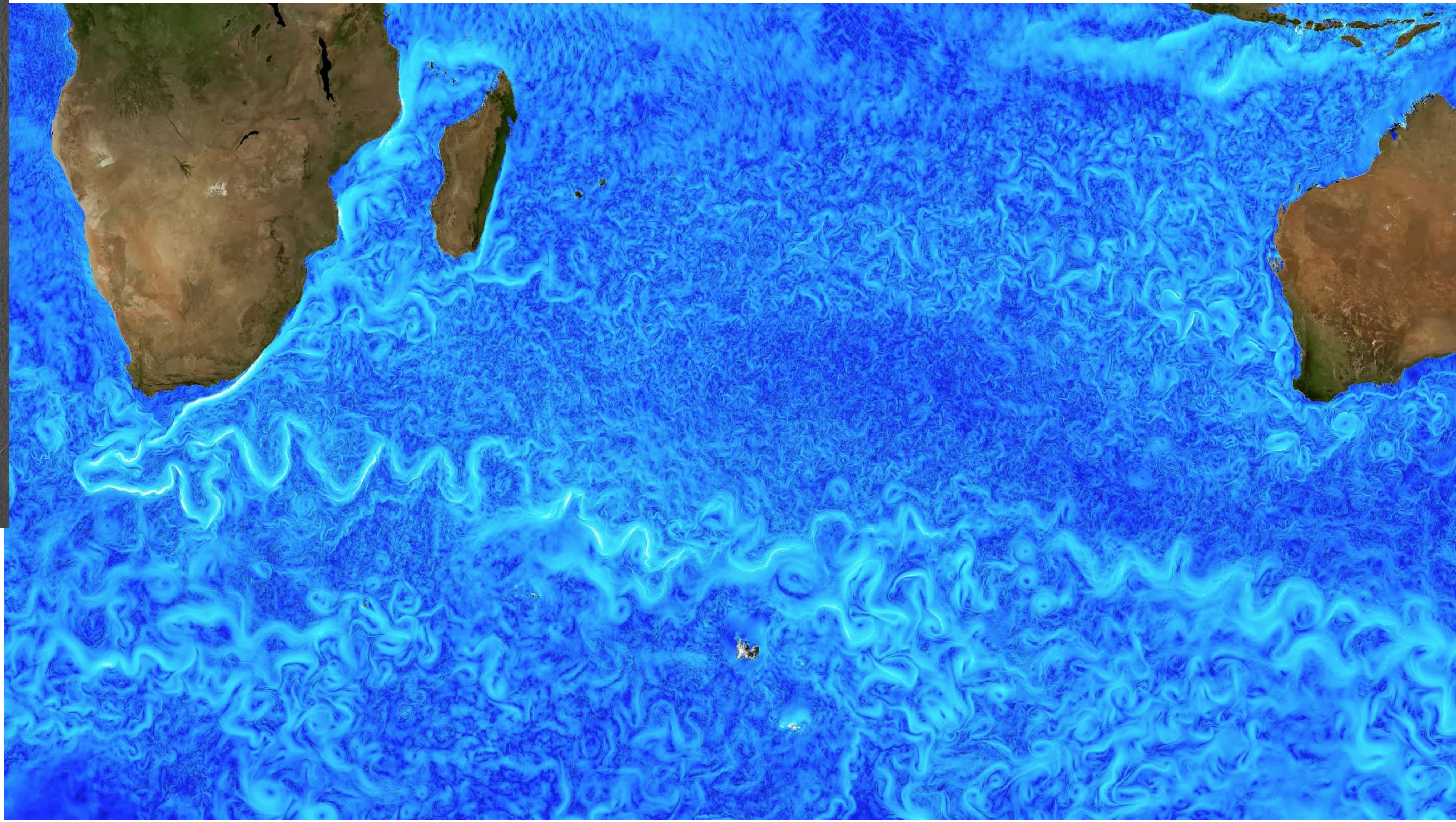
- Vertical component of relative vorticity model view



From virtual view like this we can begin to understand the local and macro-scale role of ”sub-mesoscale eddies” in the ocean. Effective “viscosity”, “diffusion” – setting ecosystem characteristics.

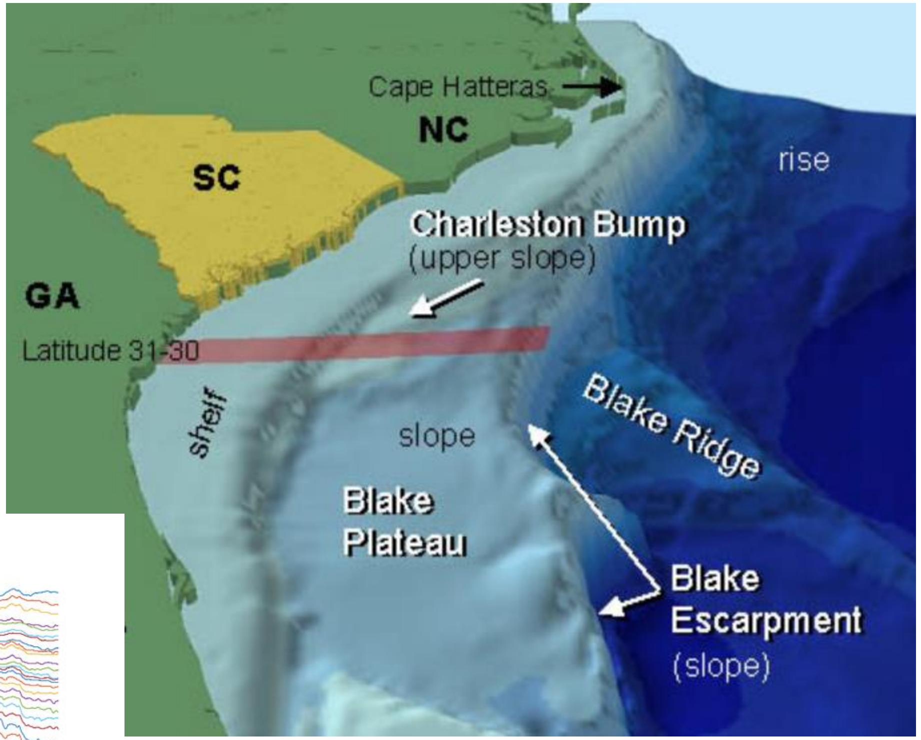
Internal wave field



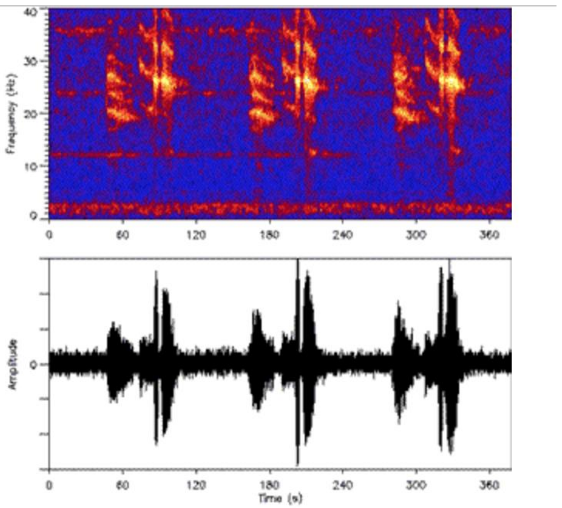


Internal waves and acoustics

internal tide and and wave supporting models, stored at high frequency, can be used to revisit active and passive acoustic monitoring. Emerging low cost measurement devices + models at right resolution can explore new tomographic possibilities.

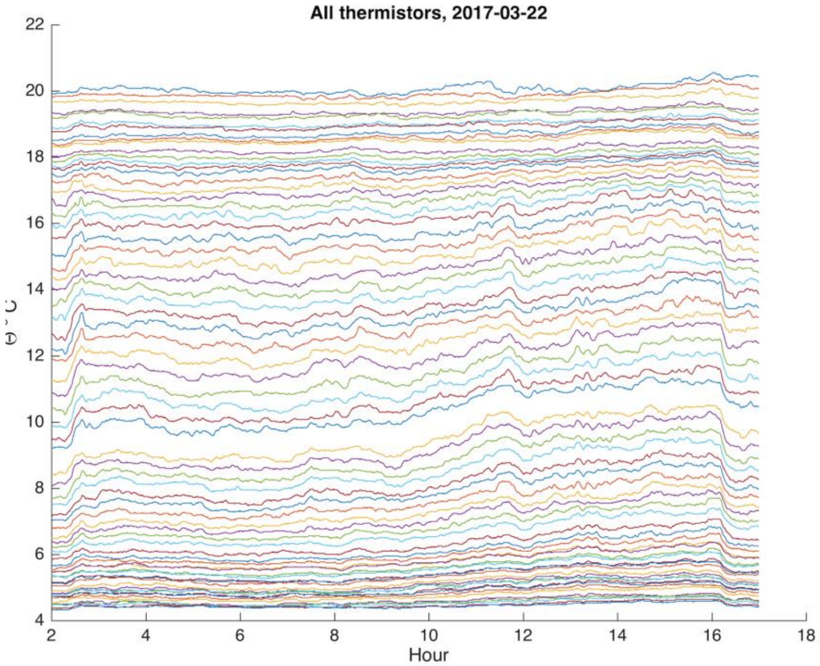


Ongoing passive and active field and modeling experimentation.



Sub-surface sonogram. Ocean quite rich in acoustic "information" c.f. EM signals in extra-terrestrial science, ambient seismic and tremor in solid Earth.

Sub-surface high-spatial, temporal thermistor trace.



Spectra – frequency/wave number

KE spectra comparison with current meter

With these modeling approaches (resolution, tides, high-frequency sampling and storing) can get much improved representations of historical current meter data statistics.

Can go back to some pioneering observational work and begin to catch up!

Classic obs (~1970)

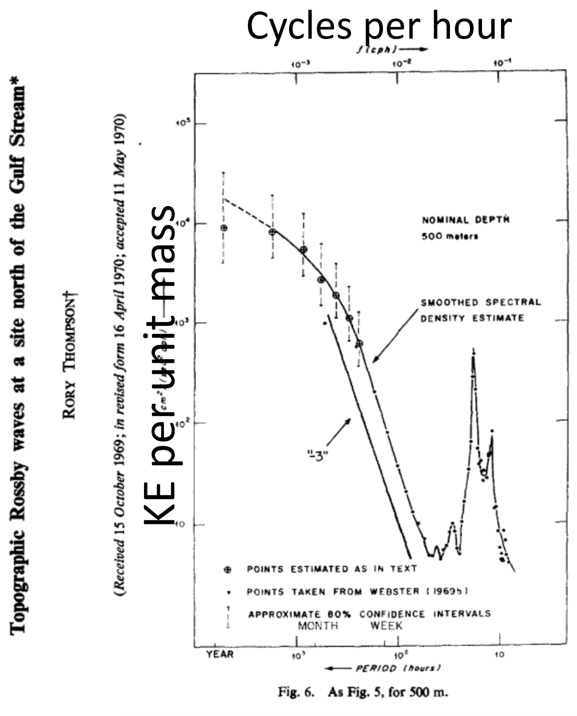
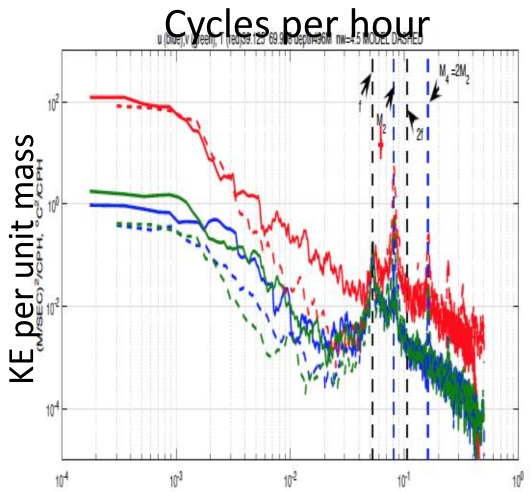
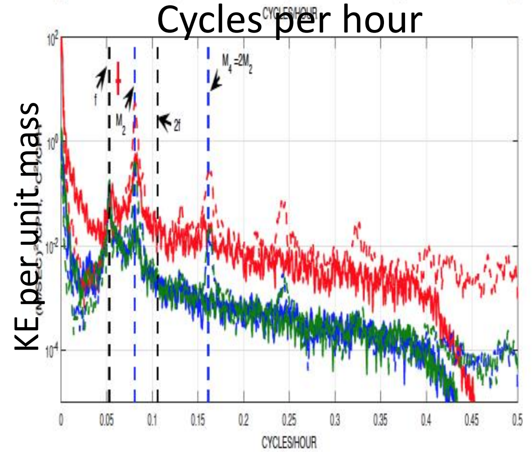


Fig. 6. As Fig. 5, for 500 m.

Woods Hole "Site D" (42° 9.71'N, 70° 38.4'W) current meter spectra comparison.

model





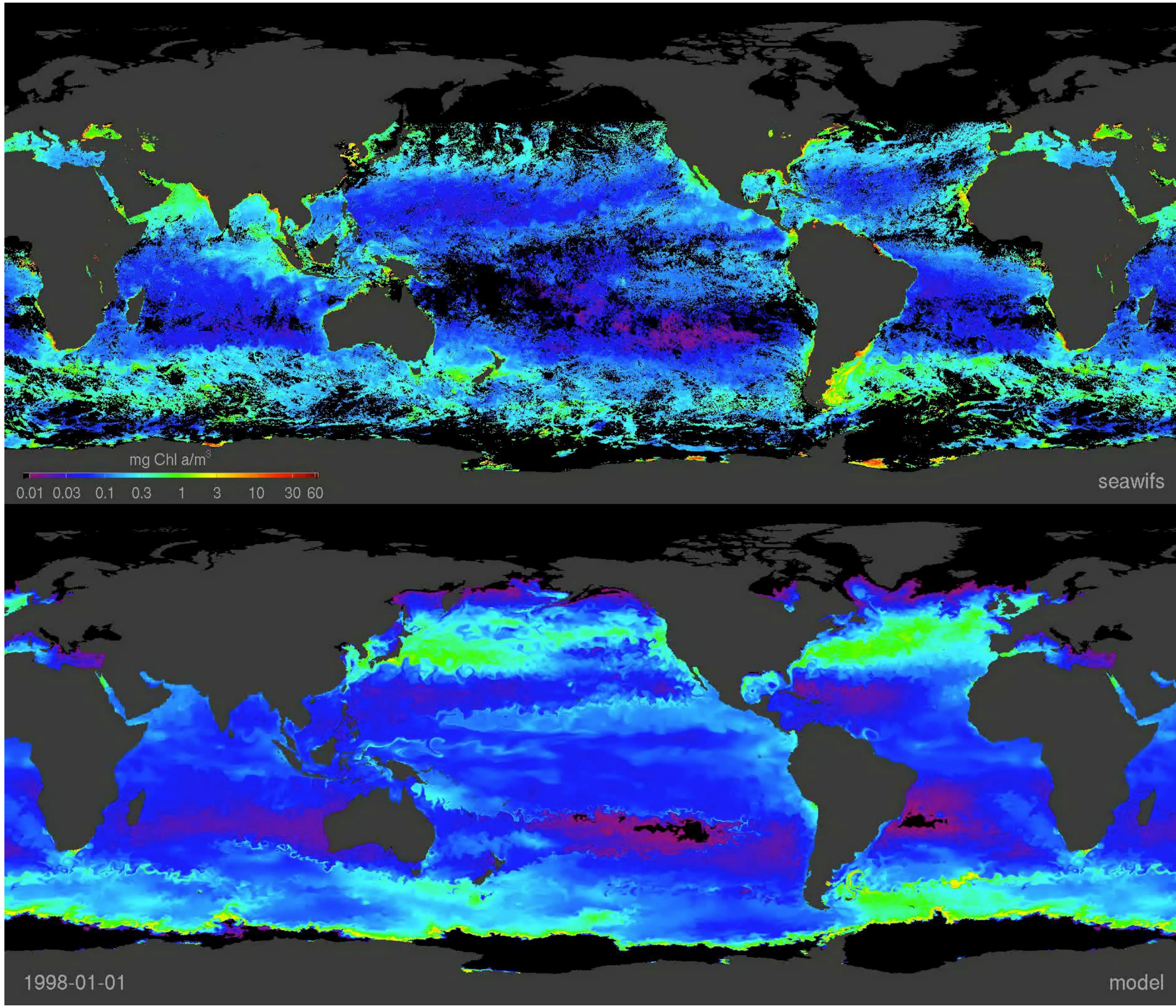
Model (dashed) and obs (solid) spectra

(*cf Carl Wunsch*)

Bio/ecosystems

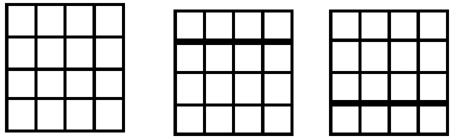
Marine microbial biomass is significant and its dynamics not well understood. It plays a role in a wide number of air-sea gas exchanges and in maintaining base food web in ocean.

One major uncertainty is interplay between ecosystem diversity and ocean dynamics across scales.



Statistical approaches

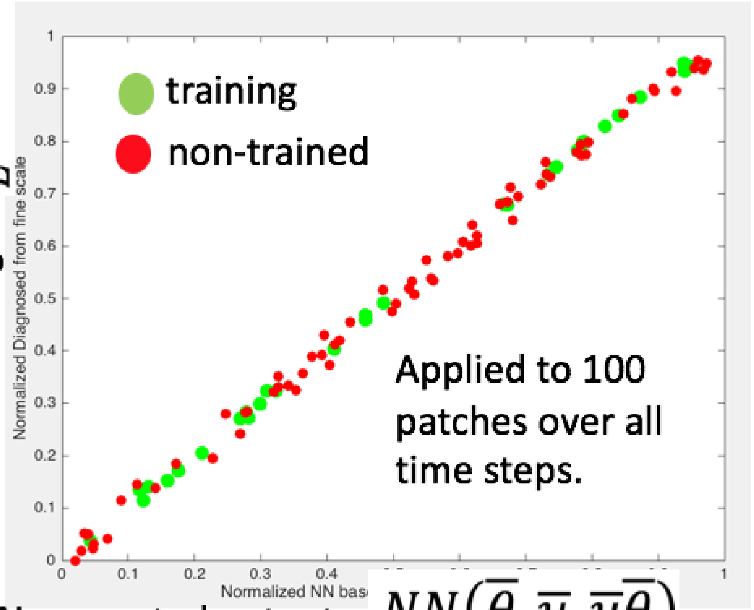
A small test - NN trained on slightly coarsened stencil to "fit" integrated fine scale term.



Network is trained on 16 point average of velocity in one direction and temperature as input. Fine-scale line integral around area boundary is used as target output.

Emulate this

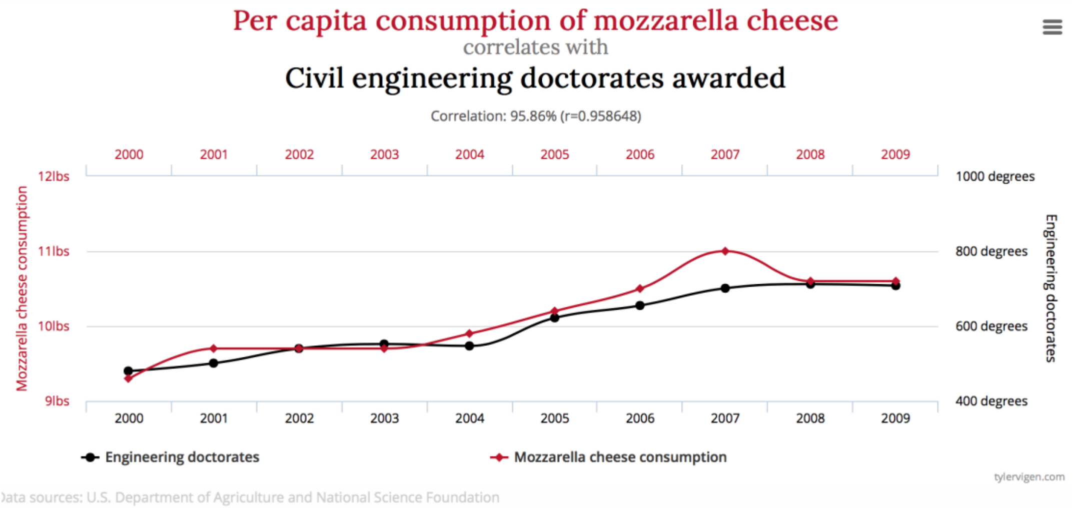
Normalized target $\frac{1}{L}\oint u_f \theta_f$



Using info at coarsened resolution.

NN computed output $NN(\overline{\theta},\overline{u},\overline{u\theta})$

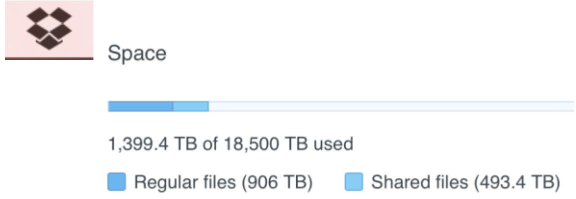
xkcd



spuriouscorrelations.com

Very promising but, "correlation vs causation" challenges…

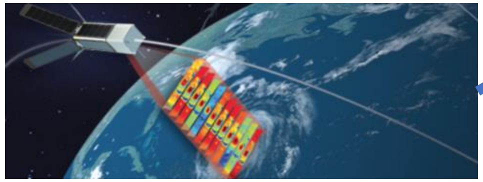
Technologies



OSIRIS About

PROJECT NESE

- Sharing
  - dropbox, NESE/OSIRIS etc….
- Storage
  - End to end pipelines (generation to curation)



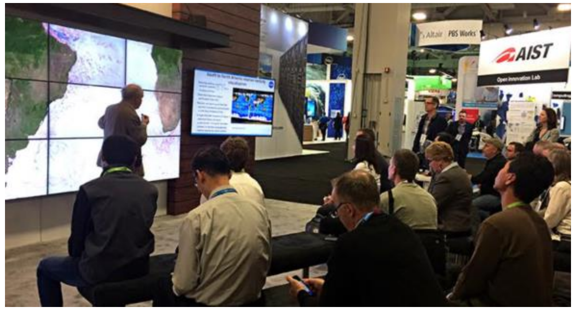
404 : Not Found

You are requesting a page that does not exist!

v

Google Earth Engine

- Viz
  - storage and networking enable



Sharing

Streamlined portal in development, to be backed by dropbox (for now), NESE/OSIRIS tech.

Already 25+ projects using, including

Award Abstract #1520825

Hazards SEES: Advanced Lagrangian Methods for Prediction, Mitigation and Response to Environmental Flow Hazards

NSF Org: AGS Div Atmospheric & Geospace Sciences

MATHEMATICS & PHYSICAL SCIENCES | LIFE SCIENCES | AUTISM RESEARC

Home » Funding » Funding Opportunities » Life Sciences » Simons Collaboration on Ocean Processes ar Investigator Awards

Simons Collaboration on Ocean Processes and Ecology Investigator Awards

Defense Advanced Research Projects Agency > Program Information

Positioning System for Deep Ocean Navigation (POSYDON)

Dr. Lisa Zurk

LLC4320 Portal

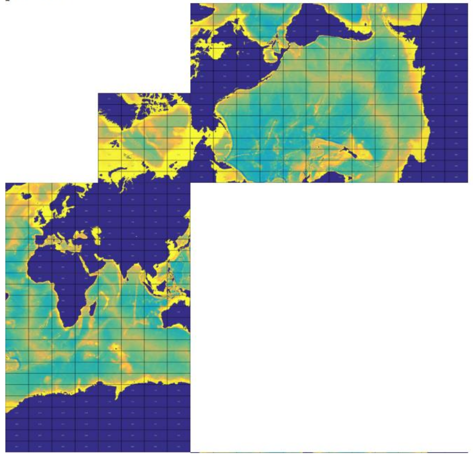
Home

LLC Model Output Fetch

Tooltip testing

LLC Model Output Fetch

Click on a tile in the image below to bring up a menu controlling output to fetch.

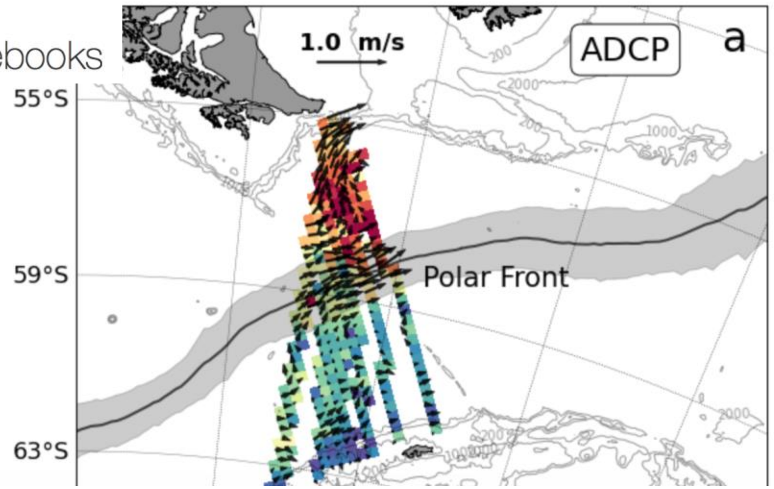


```
cb.set_label(u'$(<u^2>+<v^2>)/2$ [m$^2$ s$^{-2}$]',fontsize=17)
cb.set_ticks(np.array([0.005,0.04,0.08]))
cb.ax.tick_params(labelsize=17)

ig('figs/adcp_var26.',dpi=300,bbox_inches='tight')
g('figs/adcp_var26.eps',format='eps',bbox_inches='tight')
```

nbviewer

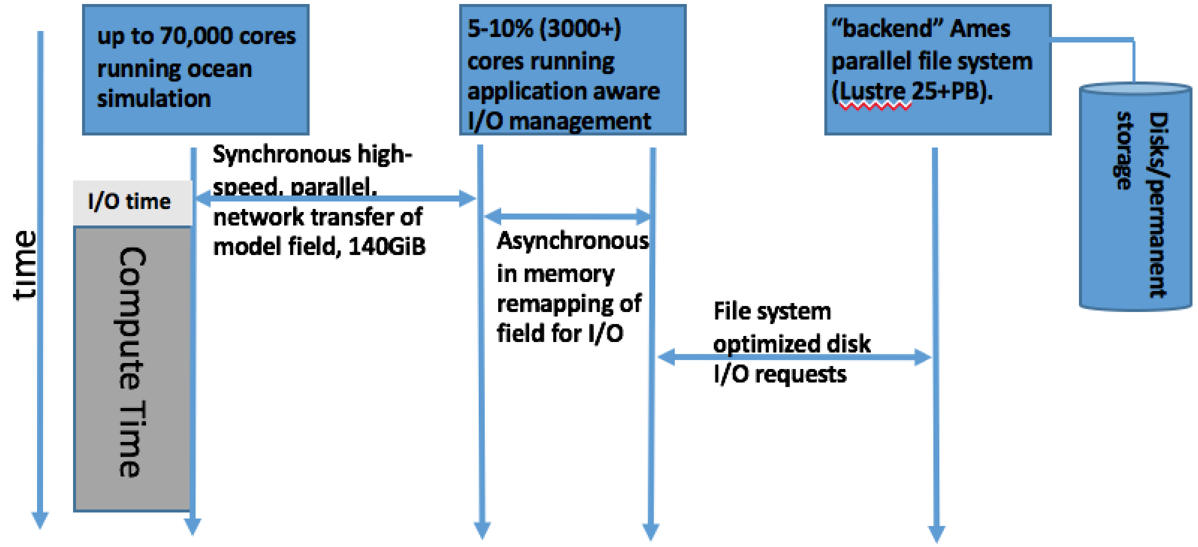
A simple way to share Jupyter Notebooks



Storage and network tools of the trade

Output saving enabled by parallel, asynchronous active I/O layer development and NASA Ames. This offloads I/O preparation to an intermediary that can map between compute data structure layout and shared file system layout.

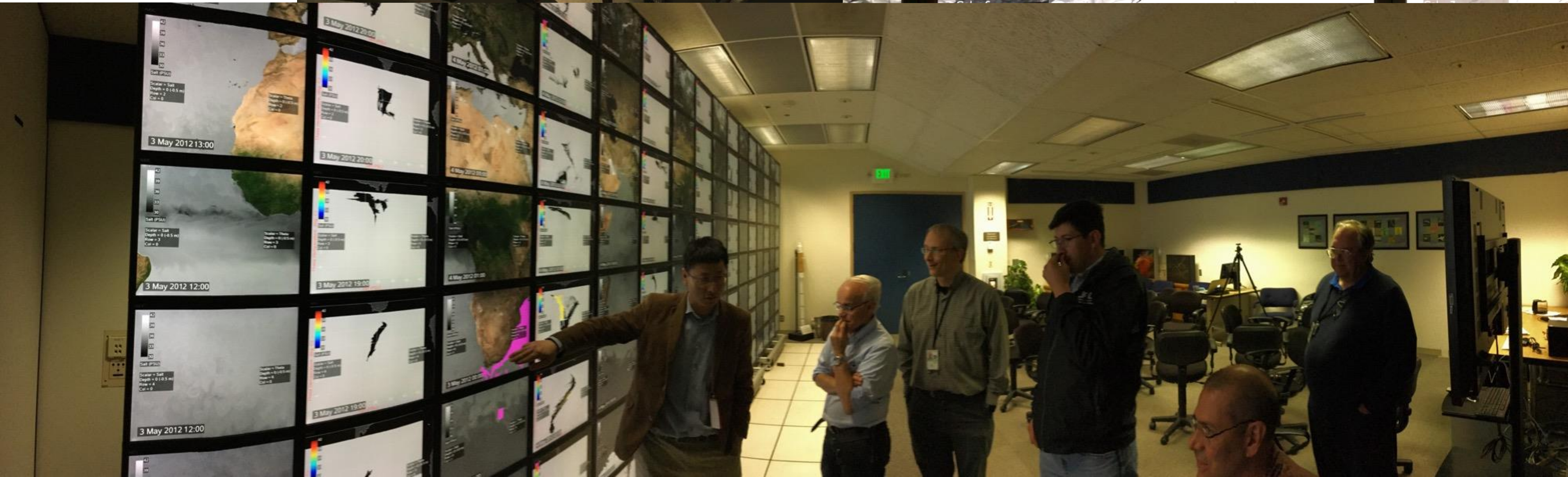
e.g.



This I/O layer layer is customized to support the activity with sufficient efficiency. Generic solutions/tools all had too slow I/O and negative computer system impacts.

- netcdf, python, OpenDAP
  - Temporal-geospatial extraction
- Data transfer nodes
  - ssd's, FIONA's, ….
- Human time and expense
  - Nicole etc….
- Integration challenges
  - NASA - no globus
  - Globus doesn't speak netcdf/OpenDAP extract
  - DTN nodes depend on modality, perfSonar to disk-disk gap.
  - NESE, OSIRIS, NDS good steps in right direction
  - Long-term curation, brokering partnership with libraries and big-data monopolies – very unclear clear who pays and how, need some experiments here too etc…

Storage and network enables analytic viz.

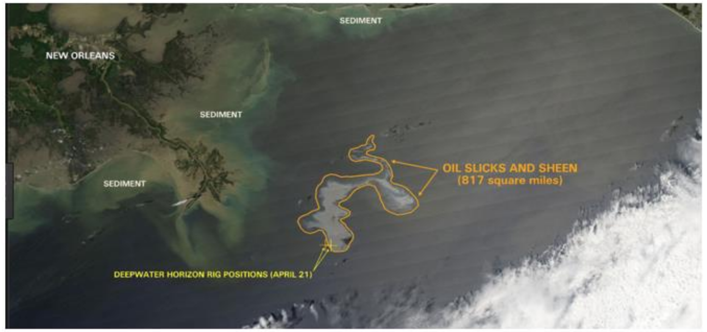


Summary

- Continued scale up of interoperable storage and network “enablement” CI is transformative for sharing geophysical synthetic and sensor data Earth system info.
- Our ocean activity is still somewhat in its infancy, but potential for many publications per year (c.f. our long running open modeling efforts) from these approaches is clear. Some suggestion that key *Turing tests* are within reach – for first time.
- Definitely exciting to see comprehensive cyberinfrastructure enhancements/innovations that help boost utility.
    - Some credits…. Dimitris Menemenelis +, NASA JPL. NASA Ames viz, compute/storage and network team (Chris Henze, Bob Ciotti, Bill Thigpen, Nicole Boscia etc..). MITgcm, ECCO, Darwin projects – (NSF, DOE, DARPA, ONR, NASA, NOAA, Unidat, Moore Foundation, Simons Foundation, Commonwealth of Mass etc…). Brian Arbic.

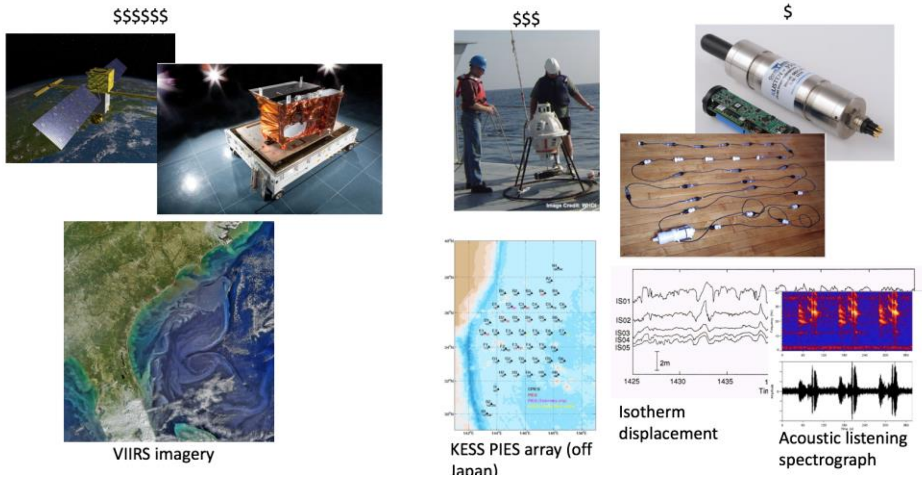
Sea-level in Boston/Cambridge/Somerville

Better global to coastal planning.



Better response.

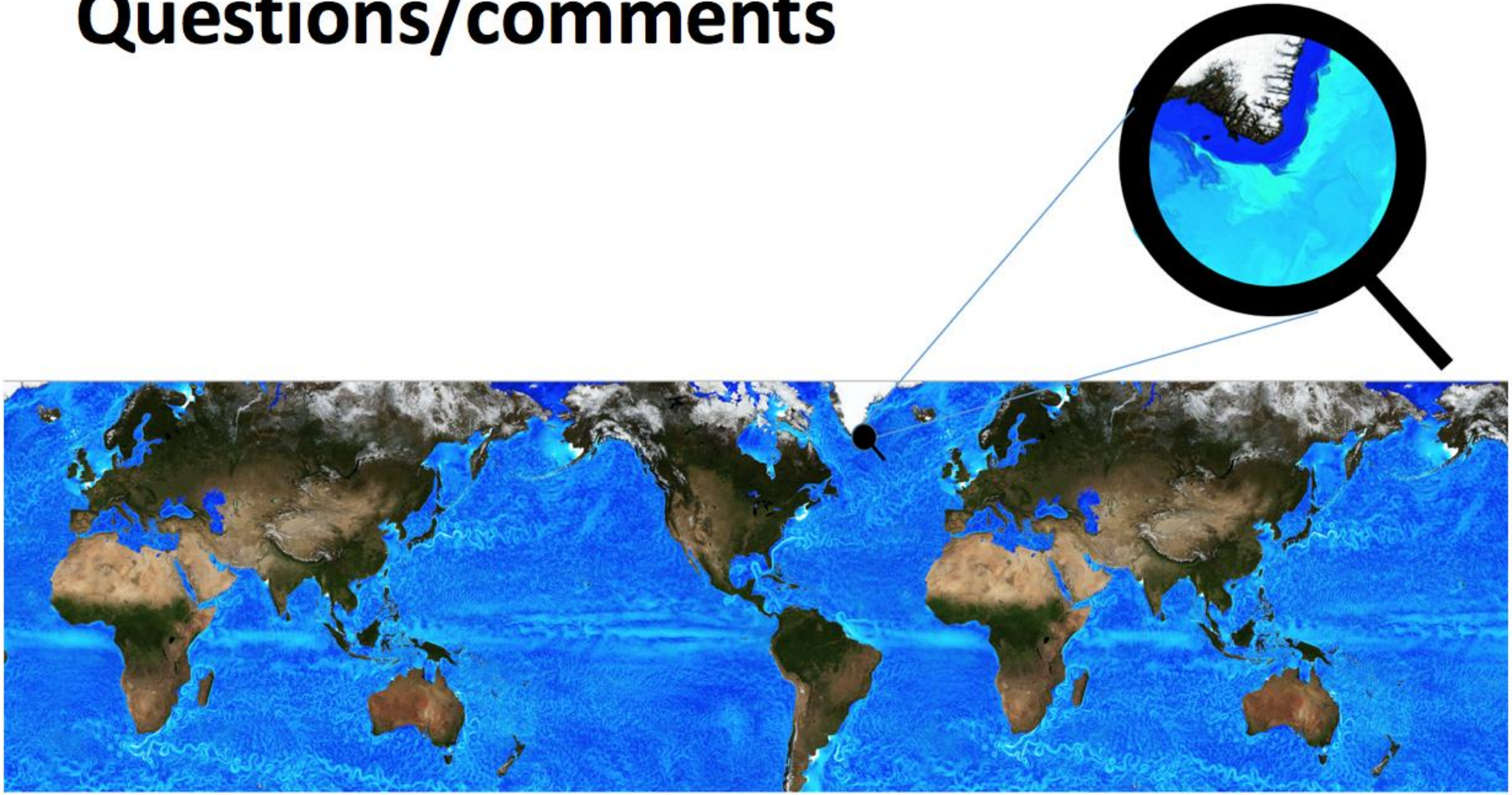
Next generation O(km) observing systems



Deeper cyber-physical capabilities.

Thank you

Questions/comments



Context

Better World

Less Disease | Less Hunger
More Knowledge | Cooler gadgets
...

Research

Physics | Earth Science
Atrophysics | Mathematics
Biology | Social Science
Genomics | Economics
...

Research Infrastructure

Submarines | Sequencers
Satellites | Pencil and Paper
Telescopes | Drilling Rigs
Electron Microscopes | Cyberinfrastructure

→

Better World

Less Disease | Less Hunger
More Knowledge | Cooler gadgets
...

Research

Physics | Earth Science
Atrophysics | Mathematics
Biology | Social Science
Genomics | Economics
...

Research Infrastructure

Submarines | Sequencers
Satellites | Pencil and Paper
Telescopes | Drilling Rigs
Electron Microscopes

Cyberinfrastructure

Buildings Libraries