Contribution ID: **11** Type: **Presentation**

Revealing and examining the tempestuous Global Ocean through a multi-petabyte virtual ocean archive.

Friday, 19 May 2017 15:00 (30 minutes)

This talk will explore how computational science and evolving network and storage capabilities, together with ongoing improvements in remote and in-situ sensing, may be poised, possibly like never before, to have significant impacts on global ocean research. Simultaneous improvements across network, storage, computation and sensing technologies are beginning to create a new lens through which to view, explore and understand some of the key mathematics and observations used to describe and reason about physical, chemical and biological aspects of the Earth's oceans.

Specifically this presentation examines a global one-kilometer horizontal resolution numerical ocean computation that embraces network and storage enabled computational science based approaches. The computation and some of its applications will be described. Some of the key network, storage and computational science technology ingredients that enable the work will be outlined.

The computation examined is work that was recently undertaken using the NASA Pleadies computer. It is one of a new generation of ocean computations that include representations of tidal forcings and realistic synoptic meterology. Including these aspects, at kilometer scale resolution, captures more of the rich dynamics present and observed in the real ocean. This qualitatively increases fidelity of the spatial and temporal variability represented numerically.

Our calculation is initialized from a data constrained estimate of the real-world, large-scale global ocean state. It is driven with boundary conditions taken from high-resolution, data assimilating weather models. The domain is fully global. Interestingly, from a network and storage enabled computational science perspective, we chose to take a uniquely ambitious approach to storing and distributing the simulation solution. We sampled and archived computation state to a storage subsystem at hourly frequency and at full global resolution for a full year. This created a new and novel resource for ocean research. It is multi-petabyte in size and has global coverage.

The resulting set of more than 10ˆ15 spatially and temporally varying numerical values is supporting a variety of interesting and insightful studies. Many of these would not be easily possible without the underlying network and storage cyberinfrastructure. Advanced cyberinfrastructure underlies archive creation, enables distribution of sizable sub-samples from the archive, and provides tools used in multiple subsequent research studies.

High spatial and temporal storing of the computation more readily reveals an ocean that is teeming with turbulent vorticies and wave motions globally. A series of eye catching visualizations illustrate this. They show what the ocean would look like to eyes that could discriminate components vorticity and density surfaces, instead of visible light!

Examining local regions in frequency wave number space, the stored solution provides notably more complete comparison with theoretical predictions and historical observations than previous generation ocean models. This increased fidelity, combined with the rich sampling archive, is allowing the effort to help guide and support focussed observational field campaigns both at specific locations and globally.

High spatial and frequency capture also allows us to explore new directions in developing statistical relations between readily observable ocean fields and features of interest that are not as directly observable. One example of this, is trying to reduce the stochastic uncertainty due to the ocean internal wave field that impacts acoustic travel time estimates. Underwater acoustics is a potentially powerful tool for measuring the ocean and for creating fully mobile sub-surface networks. It is notoriusly challenging in part because of inherent low bandwidth, but also in part because of the complicated time dependent nature of the ocean as a transmission media. We will illustrate how network and storage enabled approaches can be leveraged in this context. Leveraging these approaches allows us to develop new ways to determine aspects of the internal wave field statistics in a more complete manner. This work draws on the application of statistical methods prevalent in machine learning/big-data communities. Using those methods we can develop various semi-empirical regressions between observable fields and internal wave statistics. Application of these sorts of methods is fundamentally enabled by increasingly robust storage and network cyberinfrastructure technologies.

Another example application looks at the role of high spatio-temporal frequency processes in shaping marine microbial patterns in the ocean. Microbial communities in the ocean form the base of the food chain and play a major, but uncertain, role in Earths carbon, oxygen and nitrogen balance. Marine microbial community structure and ecosystem dynamics remain an area of active research. A highly sampled global fluid solution with spatial and temporal resolution down to scales of kilometers and hours support new ways to explore possible ideas on governing mechanisms for these communties. Recent work in this context will be illustrated.

Finally, we will also sketch briefly the network and storage technologies employed. We will describe approaches for storing data at adequate rates and for disseminating the solution across national networks. The approaches are allowing us to begin to share solutions widely, to local/regional facilities and to cloud services including Dropbox, AWS and Azure. The technical lessons from this exercise show great promise. They provide an illustration of the potential that future ongoing hyperconnected cyberinfrastructure investments could unleash - especially if key technologies are made more routine and implemented generally in a sufficiently interoperable, capable and cost-effective manner.

Primary author: HILL, Chris (MIT)

Presenter: HILL, Chris (MIT)

Session Classification: Science Use Cases

Track Classification: Science Use-Cases