An information technology foundation for fostering interdisciplinary oceanographic research and analysis

Edward Armstrong[1], Mark Bourassa[2], Thomas Cram[3], Jocelyn Elya[4], Frank Greguska[1], Thomas Huang[5], Joseph Jacob[5], Zaihua Ji[6], Yongyao Jiang[7], Yun Li[7], Lewis McGibbney[1], Nga Quach[1], Shawn Smith[4], Vardis Tsontos[8], Brian Wilson[5], Steven Worley[9], and Chaowei Phil Yang[10]

[1]NASA Jet Propulsion Laboratory
[2]Florida State Univ
[3]National Center for Atmospheric Research
[4]Florida State University
[5]Jet Propulsion Laboratory
[6]Nat'l Ctr for Atmospheric Research
[7]George Mason University Fairfax
[8]NASA/JPL
[9]NCAR
[10]George Mason Univ.

November 21, 2022

Abstract

Before complex analysis of oceanographic or any earth science data can occur, it must be placed in the proper domain of computing and software resources. In the past this was nearly always the scientist's personal computer or institutional computer servers. The problem with this approach is that it is necessary to bring the data products directly to these compute resources leading to large data transfers and storage requirements especially for high volume satellite or model datasets. In this presentation we will present a new technological solution under development and implementation at the NASA Jet Propulsion Laboratory for conducting oceanographic and related research based on satellite data and other sources. Fundamentally, our approach for satellite resources is to tile (partition) the data inputs into cloud-optimized and computation friendly databases that allow distributed computing resources to perform on demand and server-side computation and data analytics. This technology, known as NEXUS, has already been implemented in several existing NASA data portals to support oceanographic, sea-level, and gravity data time series analysis with capabilities to output time-average maps, correlation maps, Hovmöller plots, climatological averages and more. A further extension of this technology will integrate ocean in situ observations, event-based data discovery (e.g., natural disasters), data quality screening and additional capabilities. This particular activity is an open source project known as the Apache Science Data Analytics Platform (SDAP) (https://sdap.apache.org), and colloquially as OceanWorks, and is funded by the NASA AIST program. It harmonizes data, tools and computational resources for the researcher allowing them to focus on research results and hypothesis testing, and not be concerned with security, data preparation and management. We will present a few oceanographic and interdisciplinary use cases demonstrating the capabilities for characterizing regional sea-level rise, sea surface temperature anomalies, and ocean hurricane responses.

National Aeronautics and Space Administration

An information technology foundation for fostering interdisciplinary oceanographic research and analysis (OS21C-1583)

Edward.M.Armstrong@jpl.nasa.gov[1], Mark A. Bourassa[2], Thomas Cram[3], Jocelyn Lee Elya[2], Frank R. Greguska III[1], Thomas Huang[1], Joseph C. Jacob[1], Zaihua Ji[3], Yongyao Jiang4, Yun Li[4], Lewis J. McGibbney[1], Nga Quach[1], Shawn R. Smith[2], Vardis M. Tsontos[1], Brian D. Wilson[1], Steven J. Worley[3] and Chaowei Phil Yang[4]
(1) NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States, (2) Florida State University, Center for Ocean-Atmospheric Prediction Studies, Tallahassee, FL, United States, (3) National Center for Atmospheric Research, Boulder, CO, United States, (4) George Mason University Fairfax, Fairfax, VA, United States

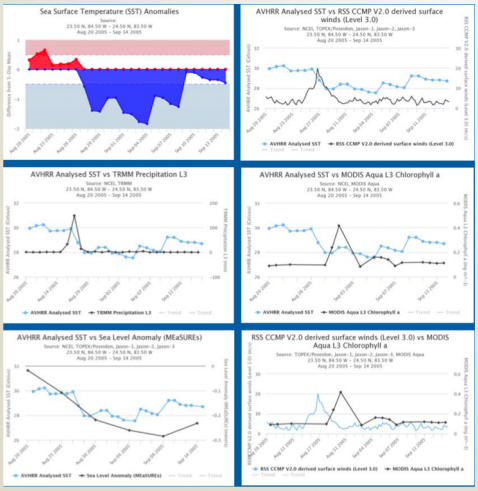
Introduction

Before complex analysis of oceanographic or any earth science data can occur, it must be placed in the proper domain of computing and software resources. In the past this was nearly always the scientist's personal computer or institutional computer servers. The problem with this approach is that it is necessary to bring the data products directly to these compute resources leading to large data transfers and storage requirements especially for high volume satellite or model datasets. Here we present a new technological solution under development and implementation for conducting oceanographic and related research based on satellite data and other sources. **Fundamentally, our approach for satellite resources is to tile (partition and chunk) the data inputs into cloud-optimized and computation friendly databases that allow distributed computing resources to perform on demand and server-side computation and data analytics**. This technology, known as NEXUS, has already been implemented in several existing NASA data portals to support oceanographic, sea-level, and gravity data time series analysis with capabilities to output **time-average maps, correlation maps, Hovmöller plots, climatological averages and more**. A further extension of this technology integrates ocean *in situ* observations, event-based data discovery (e.g., natural disasters), data quality screening and additional capabilities. It is an open source project known as the Apache Science Data Analytics Platform (SDAP) (https://sdap.apache.org), and colloquially as OceanWorks (https://oceanworks.jpl.nasa.gov), and is funded by the NASA AIST program. **It harmonizes data, tools and computational resources for the researcher allowing them to focus on research results and hypothesis testing, and not be concerned with security, data preparation and management.**

See poster **OceanWorks: Enabling Interactive Oceanographic Analysis in the Cloud with Multivariate Data (IN11C-0636)** for technical and architecture details and oral presentation **Lessons Learned in Creating Big Science Data Analysis Solutions for the Cloud (IN33A-02) (Invited).** Here we focus only on capabilities and use cases.

Ocean Use case – Hurricane Katrina

- Hurricane Katrina was a major category 5 hurricane in 2005 that impacted the oceanography and biology of coastal Florida before landfall near New Orleans, LA
- OceanWorks reproduces (within minutes) some of the key research results of Liu et al. (2009)., **showing the interplay and correlations of satellite derived ocean wind, precipitation, SST, sea level and ocean chlorophyll A**
- Liu et al., *A study of a Hurricane Katrina–induced phytoplankton bloom using satellite observations and model simulations.* JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 114, C03023, doi:10.1029/2008JC004934, 2009



Hurricane Katrina 2005

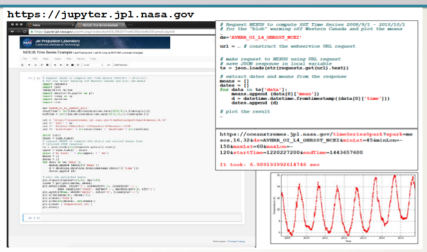
Ocean mixing and cooling leading to a phytoplankton bloom

Analysis challenges for Ocean Interdisciplinary Research

- Data resides in different formats (netCDF, HDF, ASCII etc.)
- Data in different grids or projections
- Data in different portals or APIs
- Data not easy to subset
- Data cannot easily be referenced to a climatology
- Sparse tools for online data analysis and visualization
- Match up of satellite to in situ data not performed systematically

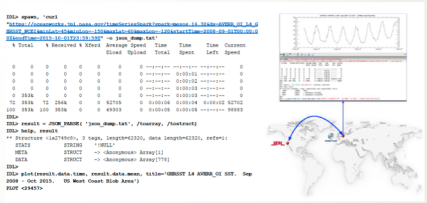
Capabilities – API and Jupyter notebook friendly

- OceanWorks in built on a RESTful API
- All capabilities (plotting, data selection, analysis) accessible via scripts and environments
- **Can be called directly from Python or IDL scripts (or Matlab, R etc.)**
- **Both these examples show the time evolution in SST of the North American West "Blob"**



Python in Jupyter Notebook

Time evolution in SST of the North American West "Blob" using the popular AVHRR_OI Level 4 SST dataset



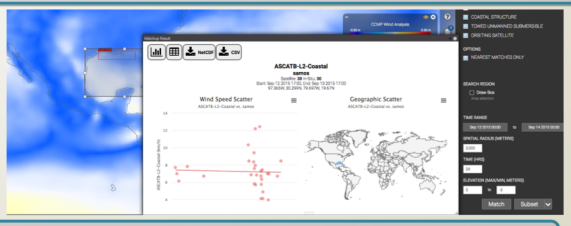
IDL

Interacting with OceanWorks' analysis at JPL from EUMETSAT (Germany) using IDL

Satellite/In situ Match Ups

Satellite to in situ observations match ups

- Satellite to in situ instruments including drifting and moored buoys, ships, gliders and other oceanographic assets from the SAMOS, ICOADS and SPURS *in situ* data collections.
- User defined time/space match-up criteria to satellite data from ocean wind, SST and salinity
- An OceanWorks service. Federated queries and responses from distinct satellite and in situ data providers.
- Key point: **Difference in satellite (ASCAT-B) and in situ (SAMOS) wind observations more than expected likely due to proximity to land affecting satellite measurements**



Additional Capabilities

- Improved integrated data layer search and discovery
- Event based data discovery
- Supports distributed architecture to reduce data movement
- Extensions for working with GIS layers (future)
- Extensions for working with ocean model data (future)
- Custom map-reduce algorithms via Python NEXUSCLI library (future)

These activities were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.