A Generalized Distributed Data Match-Up Service in Support of Oceanographic Applications

Vardis M. Tsontos[1] , Shawn R. Smith[2], Thomas Huang[1], Steve Worley[3, 2], Benjamin Holt[1], Zaihua Ji[3], Adam Stallard[2], Jocelyn Elya[2] and Mark A. Bourassa[2]

[1] **Jet Propulsion Laboratory, Pasadena, CA, USA**
[2] **Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, FL, USA**
[3] **National Center for Atmospheric Research, Boulder, CO, USA**
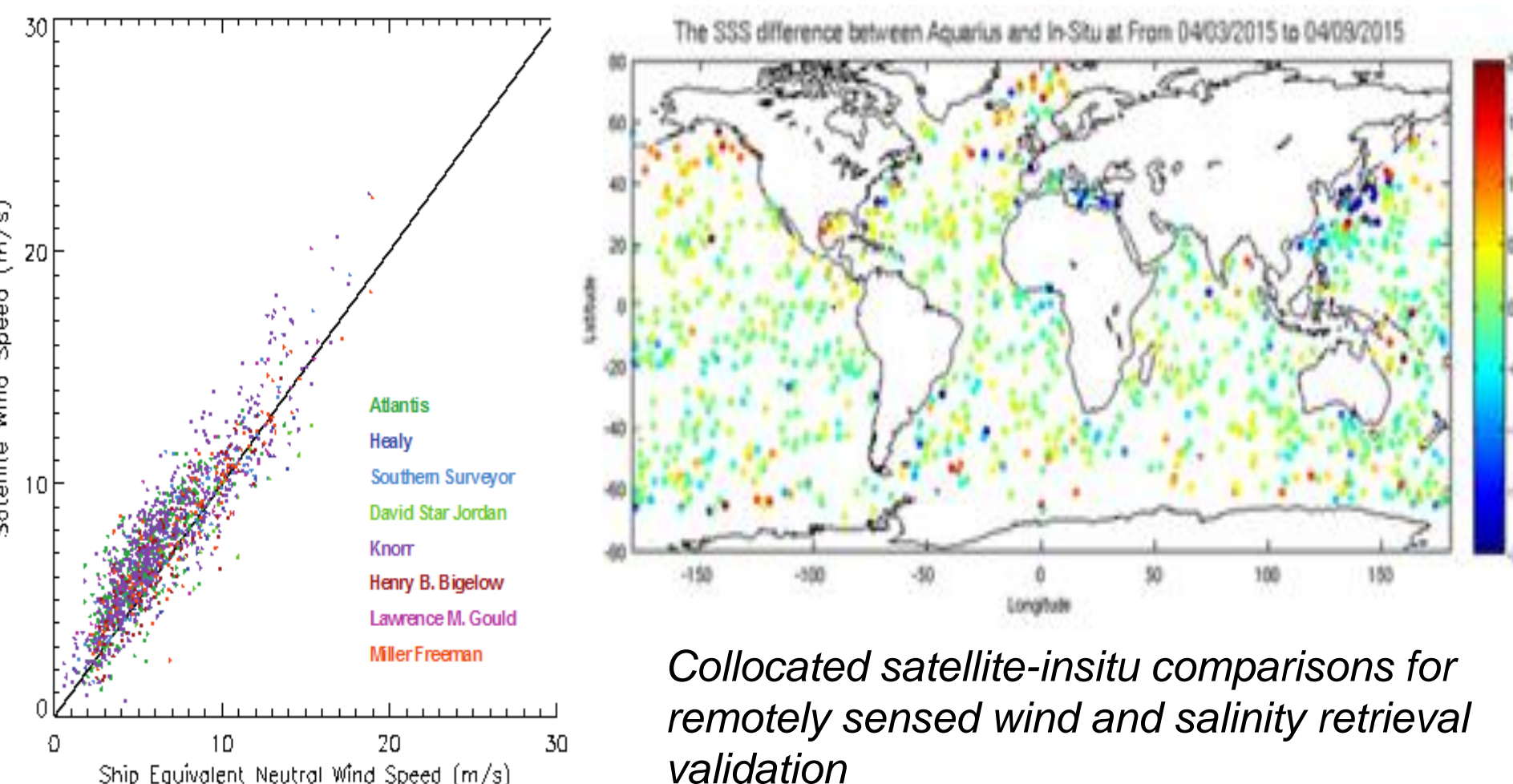
Contact: vtsontos@jpl.nasa.gov

I. ABSTRACT

Oceanographic applications increasingly rely on the integration and colocation of satellite and field observations providing complementary data coverage over a continuum of spatio-temporal scales. Here we report on a collaborative venture between NASA/JPL, NCAR and FSU/COAPS to develop a Distributed Oceanographic Match-up Service (DOMS). The DOMS project aims to implement a technical infrastructure providing a generalized, publicly accessible data collocation capability for satellite and in situ datasets utilizing remote data stores in support of satellite mission cal/val and a range of research and operational applications. The service will provide a mechanism for users to specify geospatial references and receive collocated satellite and field observations within the selected spatio-temporal domain and matchup window extent. DOMS will include several representative in situ and satellite datasets. Field data will focus on surface observations from NCAR's International Comprehensive Ocean-Atmosphere Data Set (ICOADS), the Shipboard Automated Meteorological and Oceanographic System Initiative (SAMOS) at FSU/COAPS, and the Salinity Processes in the Upper Ocean Regional Study (SPURS) data hosted at JPL/PO.DAAC. Satellite data will include JPL ASCAT L2 12.5 km winds, the Aquarius L2 orbital dataset, MODIS L2 swath data, and the high-resolution gridded L4 MUR-SST product. Importantly, while DOMS will be developed with these select datasets, it will be readily extendable for other in situ and satellite data collections and easily ported to other remote providers, thus potentially supporting additional science disciplines. Technical challenges to be addressed include: 1) ensuring accurate, efficient, and scalable match-up algorithm performance, 2) undertaking colocation using datasets that are distributed on the network, and 3) returning matched observations with sufficient metadata so that value differences can be properly interpreted. DOMS leverages existing technologies (EDGE, w10n, OPeNDAP, relational and graph/triple-store databases) and cloud computing. It will implement both a web portal interface for users to review and submit match-up requests interactively and underlying web service interface facilitating large-scale and automated machine-to-machine based queries.

II. Collocation Service Need

A wide user community seeks to match satellite to in situ observations to meet goals that include

- Satellite algorithm calibration, validation, and/or development
- Decision support for planning future field campaigns
- Scientific investigations to support process studies, data synthesis, etc.

Currently, matched datasets are created using one-off programs that require satellite and in situ data to be housed on local computers.

Collocated satellite-insitu comparisons for remotely sensed wind and salinity retrieval validation

III. DOMS Architecture

- DOMS will infuse common data access services at FSU, NCAR, and JPL.
  - Extensible Data Gateway Environment (EDGE) – a data aggregation service that supports OpenSearch, metadata export, and W10N protocol
  - Pomegranate – an implementation of the W10N specification
- Prototype will test searches across data stored using THREDDS and SQL, NoSQL, and graph databases.
- DOMS is designed to be extensible
  - Incorporate other oceanographic data types
  - Integrate data from additional data providers
  - Support matchups for terrestrial observation
 - Future matching between satellites and/or model datasets.
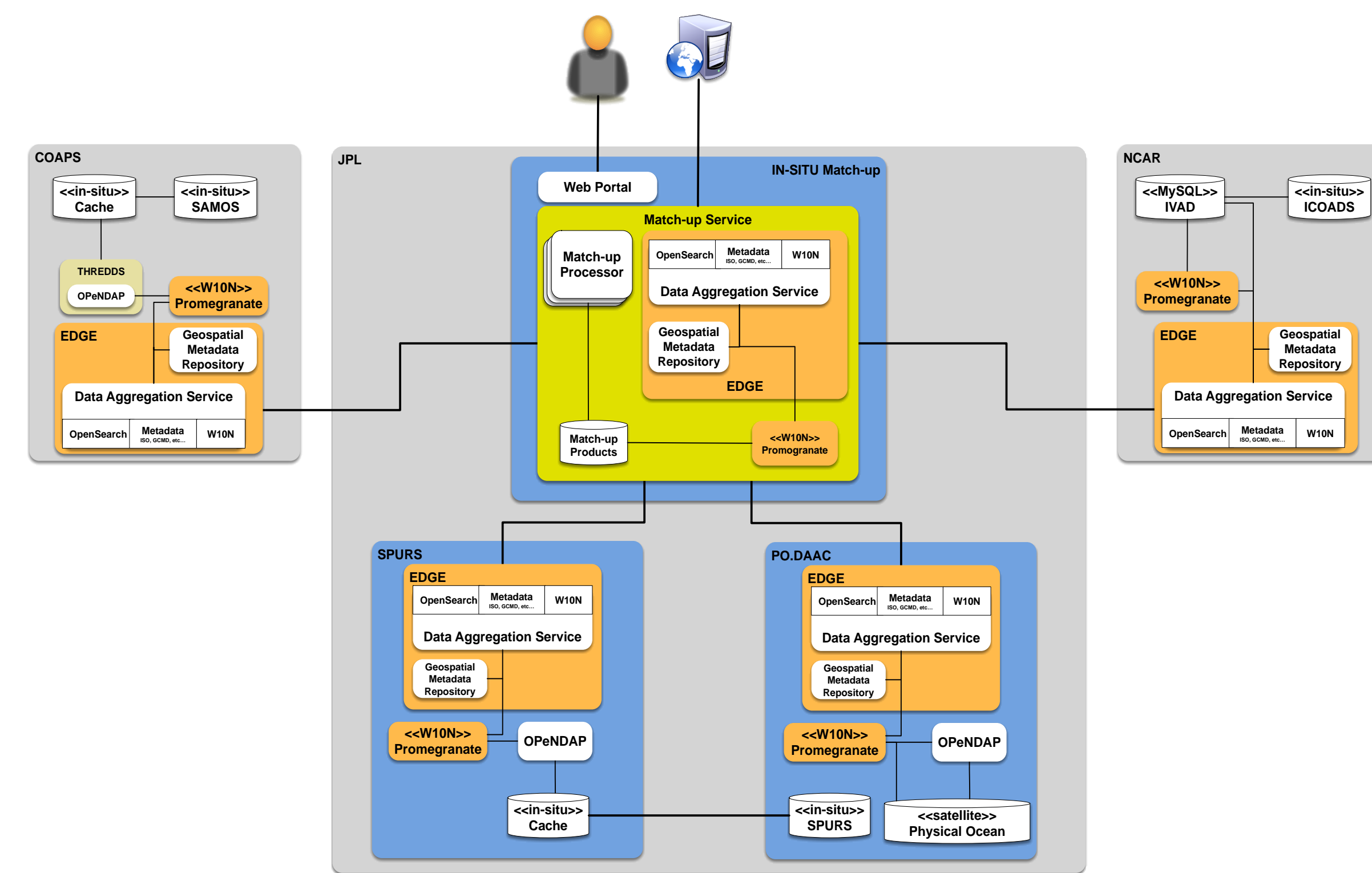
IV. Data Providers & Datasets

FSU: SAMOS

- Shipboard Automated Meteorological and Oceanographic System (SAMOS) initiative provides high-quality underway data from research vessels.
- Hosted at FSU/COAPS.
- ~30 vessels participating in 2015
  - Vessels operated by WHOI, SIO, U. Hawaii, U, Washington, U. Alaska, BIOS, NOAA, USCG, USAP, IMOS, SOI, LUMCON
  - ~30-40K one-minute observations/month/vessel
- Data include routine navigation (position, course, heading, speed), meteorology (wind, air temperature, humidity, pressure, rainfall, radiation), and oceanography (sea temperature and salinity).
- All data undergo scientific quality control.

SAMOS Data Density: 2005-2014
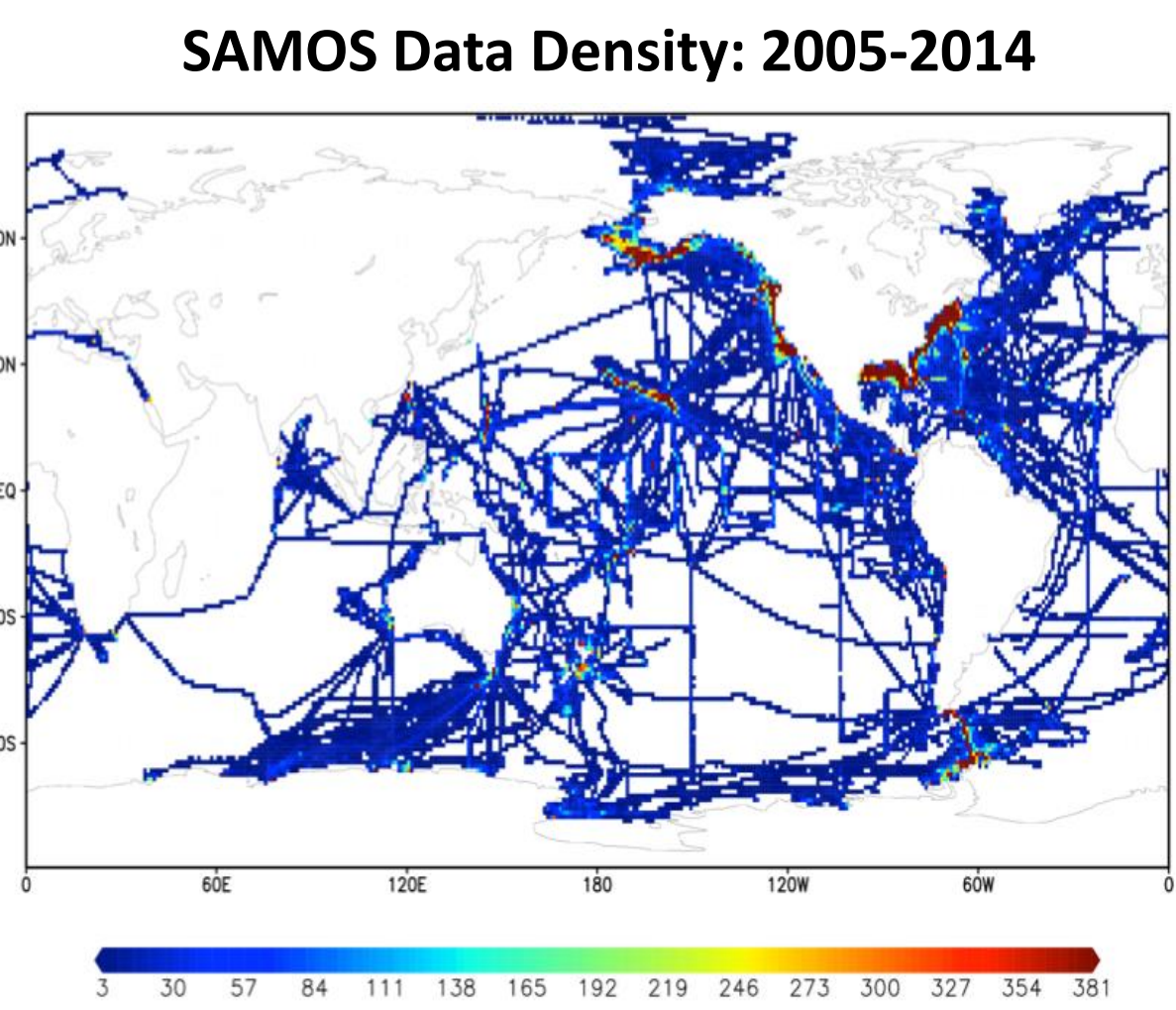
NCAR: ICOADS
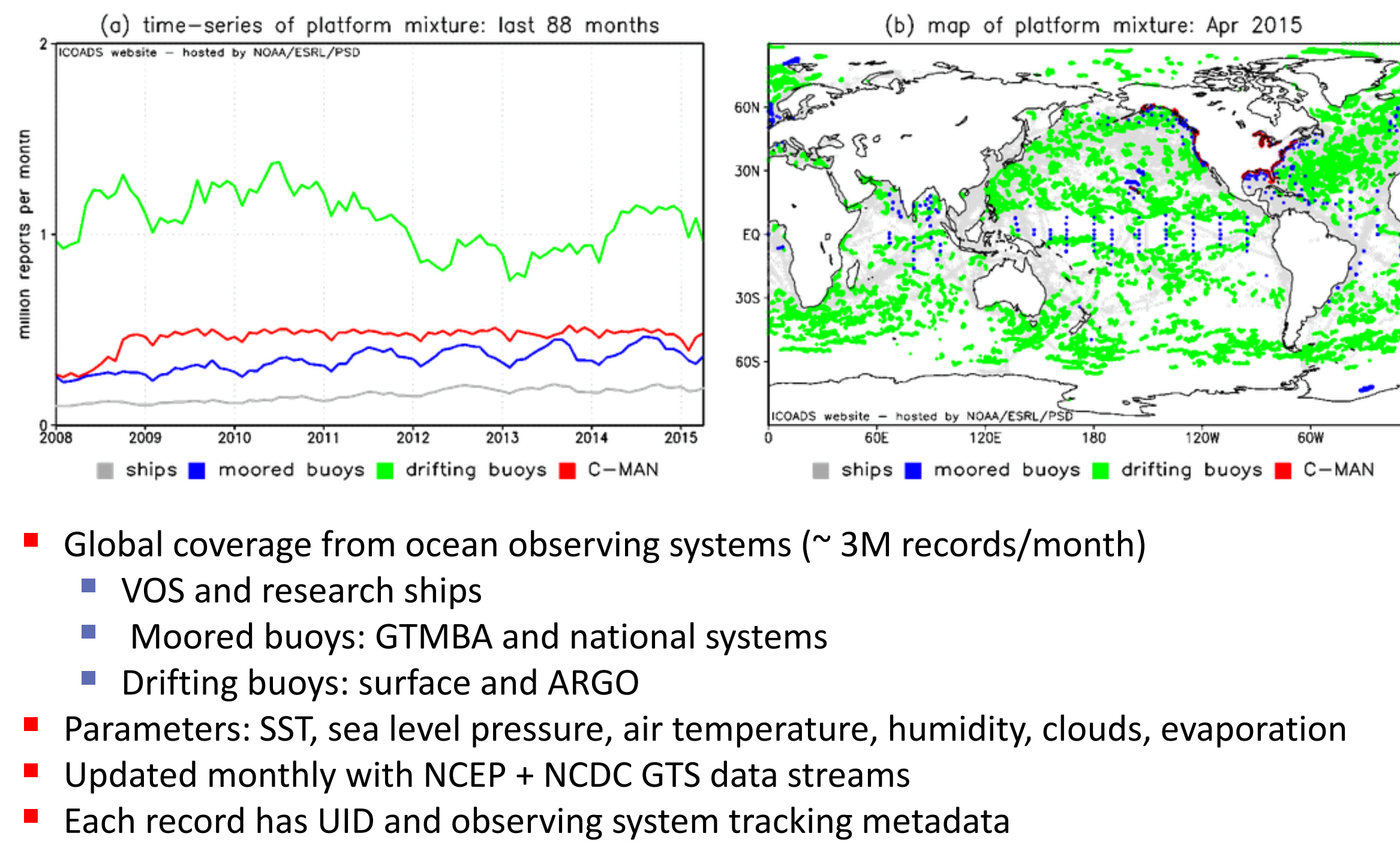
- Global coverage from ocean observing systems (~ 3M records/month)
  - VOS and research ships
  - Moored buoys: GTMBA and national systems
  - Drifting buoys: surface and ARGO
- Parameters: SST, sea level pressure, air temperature, humidity, clouds, evaporation
- Updated monthly with NCEP + NCDC GTS data streams
- Each record has UID and observing system tracking metadata

JPL: SPURS

- NASA-funded oceanographic field campaigns/science salinity process studies:
  - SPURS-1: N. Atlantic (2012-13) : salinity max region
  - SPURS-2: Eastern Equatorial Pacific (15-16): high precipitation/low evaporation region
- Advanced sampling technologies deployed in a nested design within a 900 x 800-mile² study area centered at 25˚N, 38˚W.
- SPURS converted 15 natively heterogeneous formats to NCEI NetCDF standard
  - Archived at the PO.DAAC, http://podaac.jpl.nasa.gov/spurs
- DOMS will integrate data from both SPURS campaigns
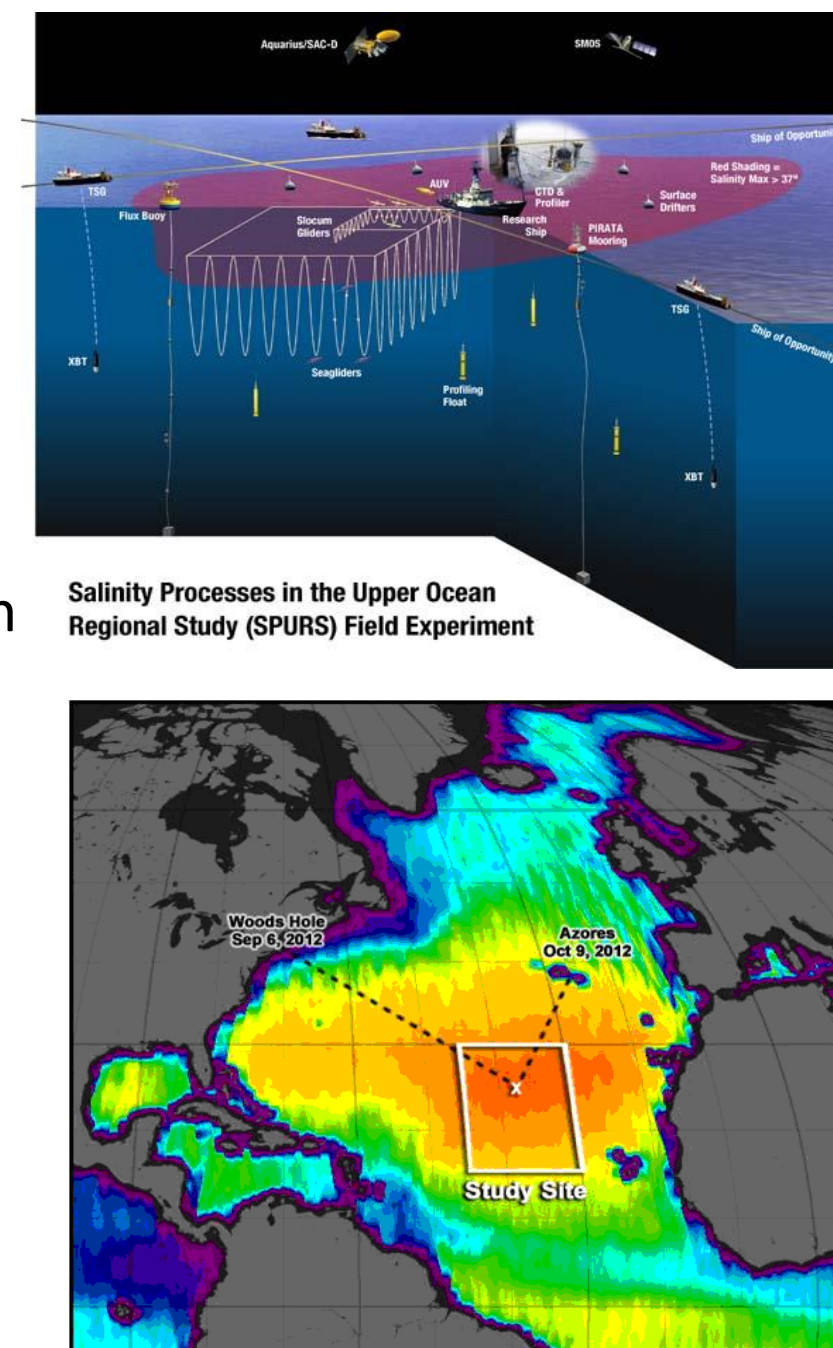
Satellite Datasets

- Satellite data hosted PO.DAAC (Physical Oceanography Distributed Active Archive Center)
- DOMS prototype will use
  - Aquarius L2 v3.0 100 km – Sea Surface Salinity
  - ASCAT L2 25 km – Wind speed and direction
  - MODIS L2 P 1 km + MRU SST 1 km daily – Sea Surface Temperature
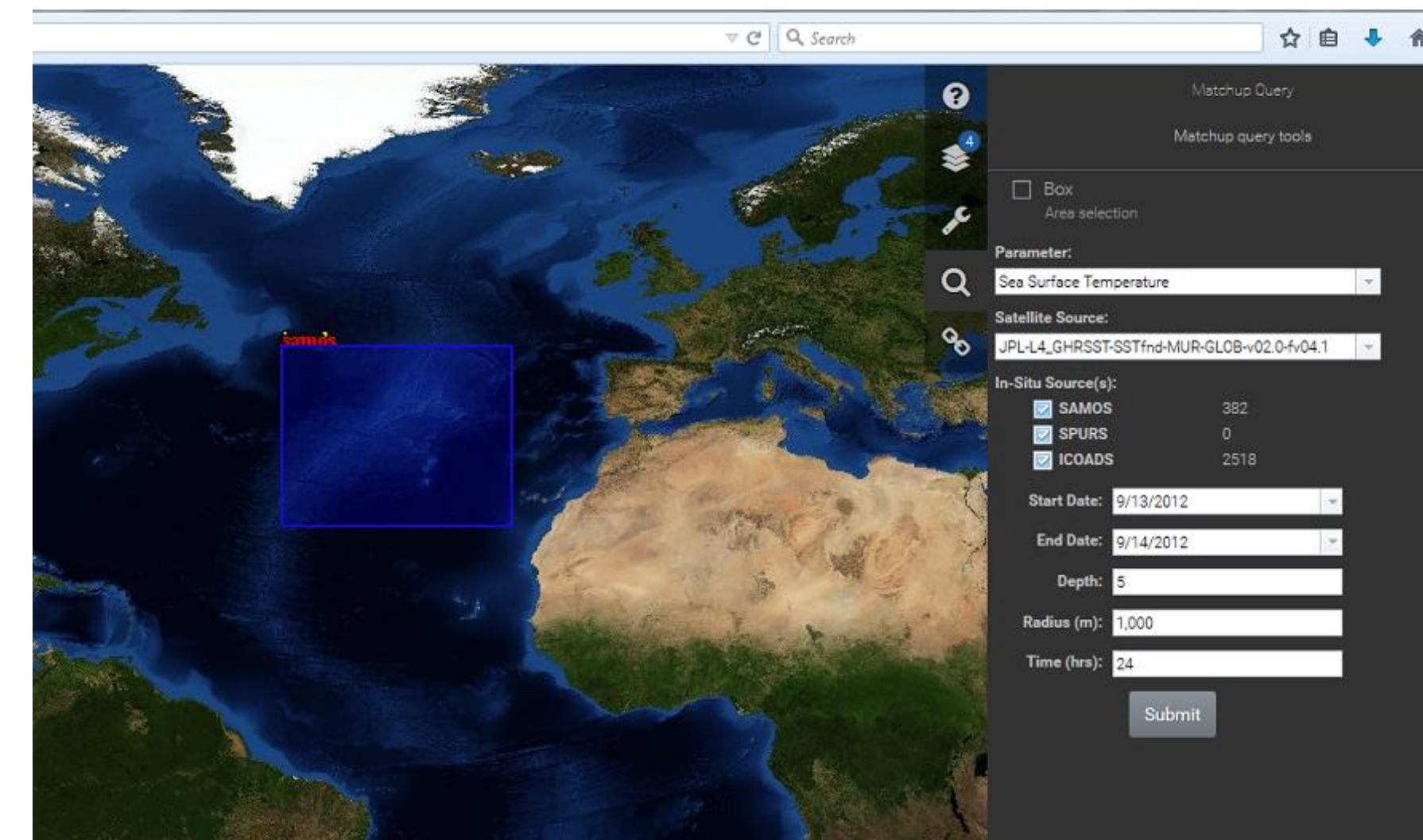- Prototype will explore match ups to both swath & gridded datasets
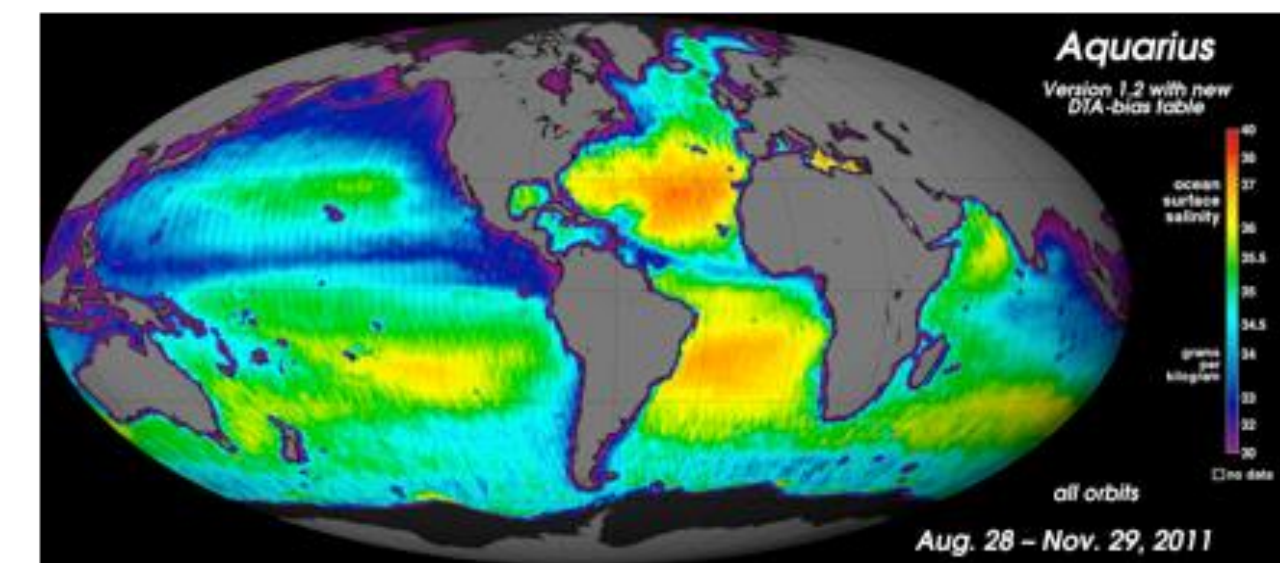
V. DOMS User Interfaces

Graphical User Interface (GUI)

DOMS will provide a web portal, graphical user interface for users to browse and to submit match-up requests interactively.

- To be hosted at JPL
- DOMS will provide flexible filtering and query specification by:
  - Instrument, sensor, parameter, provider
  - Matchup criteria: spatio-temporal domain (in x,y,z,t) and search radii/tolerances
- Interface will allow users to "test/evaluate" searches by returning metadata only, creating visualizations, and then follow with full matched dataset

Prototype DOMS GUI for Matchup query Specification

Collocated Recordset Output and Query metrics

Web service Application Programming Interface (API)

Additionally, DOMS will provide an underlying Web service interface (API) for machine-to-machine matchup operations to enable scalable data processing by external applications and services:

- Web service API support for Metadata, Subsetting, and Match-up queries
- Support for in situ to in situ, satellite to in situ, satellite to satellite collocation
- Tools & documentation will be provided to aid users in developing proper syntax for web service queries

Example DOMS metadata query URI

```
SPURS Data Service
  – Sample query: curl
    "https://deepdata.jpl.nasa.gov/ws/search/spurs?startTime=2012-08-
    01T00:00:00Z&endTime=2013-10-31T23:59:59Z&bbox=-45,15,-30,30"
  – Processed datasets
    • SPURS1_CTD
    • SPURS1_ADCP
    • SPURS1_ARGO
    • SPURS1_SEAGLIDER
    • SPURS1_MOORING_WHOI
  – 13,817,392 documents (each with sst, wind, and sss measurements)
Satellite Data Service
  – Sample query: curl -X GET
    'https://deepdata.jpl.nasa.gov/nexus/doms/datainbounds?ds=JPL-L4_GHRSST-
    SSTfnd-MUR-GLOB-v02.0-fv04.1&ny=DEV&minLat23.93&maxLat=25.2&minLon=-
    38.74&maxLon=-37.33&startTime=1345280400000&endTime=1345366800000'
  – Dataset: GHRSST Multi-scale Ultrahigh Resolution (MUR) – Over the SPURS region
    • Metadata index – 585,792 documents
    • Data tiling – 591,887 tiles
```

Example DOMS matchup query URI

```
  – In-situ to In-situ match-up: % curl -X GET
    'http://doms.jpl.nasa.gov:8080/doms?s1=samos&s2=spurs&s=2012-
    09-29T00:00:00Z&e=2012-09-29T23:59:59Z&b=-45,15,-
    30,30tt=3600&dt=504rt=100'
  – Satellite to In-situ match-up: % curl -X GET
    'http://doms.jpl.nasa.gov:8080/doms?s1=GHRSST&s2=spurs&s=2012-
    09-29T00:00:00Z&e=2012-09-29T23:59:59Z&b=-38.74,23.93-
    37.33,25.2&tt=3600&dt=50&rt=100'
```

VI. Search Criteria

- Via the user interface or web service, the following options will exist to refine one's query
  - Parameter to match – salinity, sea temperature, or winds
  - Date + Time range – ISO 8601 UTC
  - Horizontal domain – Latitude and longitude box
  - Vertical domain above/below sea level (Constrained in prototype to ~+/- 20 m)
  - Data source (e.g., which satellite vs. which in situ datasets)
  - Spatial and temporal tolerance for locating a match (e.g., within 3 hr and 50 km)
- Since most datasets used by DOMS will also have quality control flags, the system is being designed to
  - Provide data filtered by the host using documented analysis of QC flags as a default
  - Allow the user to the option to receive all data, regardless of QC flags

VII. Technical Challenges

- Ensuring that the match-up (parallel KD-tree) algorithms perform with sufficient speed to return desired information to the user.
- Performing data matches using datasets that are distributed on the network.
- Returning actual observations for the matches [e.g., salinity] with sufficient metadata so the value difference can be properly interpreted

VIII. Future Vision

- Operationalize DOMS and extend the network of in situ ocean data providers and satellite datasets encompassed
- Infuse the DOMS technology at other NASA DAACs and Agencies.
- Leverage DOMS for terrestrial collocation applications

Acknowledgments