In Situ Quality Flags in the Distributed Oceanographic Match-Up Service

A Component of the OceanWorks Science Data Analytics Platform

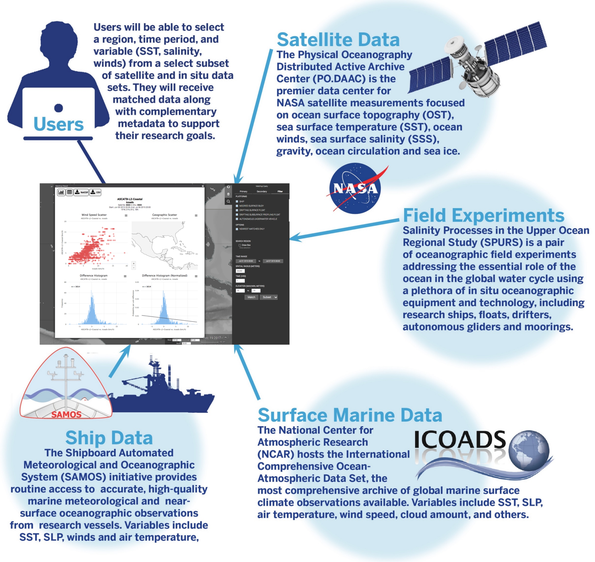
Shawn R. Smith[1] (srsmith@fsu.edu), Jocelyn Elya[1], Homer McMillian[1], Thomas Huang[2], Vardis Tsontos[2], Edward M. Armstrong[2], Elizabeth Yam[2], Maya DeBellis[2], Nga Quach[2], Frank Greguska[2], Joseph Jacob[2], Steven Worley[3], Thomas Cram3, and Zaihua Ji[3]

[1]Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, FL, USA, [2]Jet Propulsion Laboratory/California Institute of Technology, Pasadena, CA, USA, [3]National Center for Atmospheric Research, Boulder, CO, USA

Overview

The **Distributed Oceanographic Match-Up Service (DOMS)** is a collaborative effort between COAPS, NCAR, and NASA JPL. DOMS provides a mechanism for users to input a series of geospatial references for satellite observations and receive the in situ observations that are matched to the satellite data within a selectable temporal and spatial search domains. The service currently has an application program interface (API) and graphical user interface (GUI) available. DOMS directly addresses the ESIP theme to "increase the use and value of Earth science data and information", specifically in support of NASA's Earth Science mission.

Over the past year, DOMS capabilities have been enhanced to support filtering data subsets using in situ data quality flags. This new capability allows further refinement of user queries for matched satellite and in situ data. Additionally, DOMS is being integrated into the OceanWorks science data analytics platform, hosted by the PO.DAAC, and undergoing testing in the AIST Managed Cloud Environment.

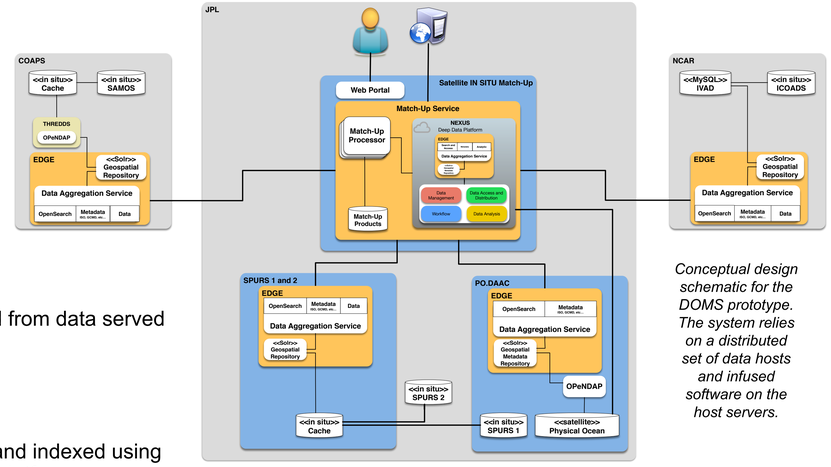


Why is DOMS Needed?

- 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
  - Investigations to support process studies, data synthesis, etc.
- The first instance of DOMS focused on algorithm calibration/validation.
- DOMS eliminates the need for one-off match-up programs that require satellite and in situ data to be housed on one's local computer

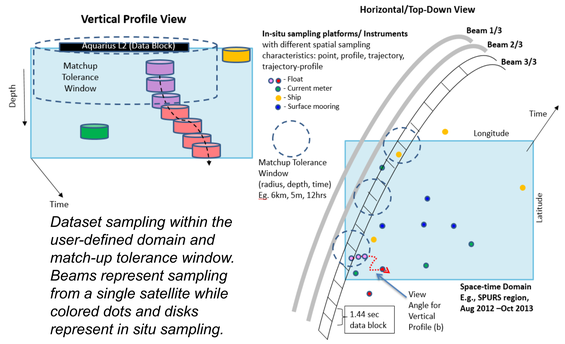
DOMS Architecture

- DOMS infuses common data access services at FSU, NCAR, and JPL.
  - Data indexing using Apache Solr
  - Extensible Data Gateway Environment (EDGE) – a data query and aggregation service that supports OpenSearch and metadata export
- In situ data are indexed from data served via
  - FSU – THREDDS
  - NCAR – MySQL
  - JPL – NoSQL
- Satellite data are tiled and indexed using the NEXUS deep data platform
- DOMS is designed to be extensible.
  - Incorporate other oceanographic data types
  - Integrate data from additional data providers



Conceptual design schematic for the DOMS prototype. The system relies on a distributed set of data hosts and infused software on the host servers.

Search Domain & Match-Up Tolerances



Dataset sampling within the user-defined domain and match-up tolerance window. Beams represent sampling from a single satellite while colored dots and disks represent in situ sampling.

- Queries facilitated by indexing the following in Solr
  - Parameter – salinity, sea temperature, or winds
  - Temporal search domain – ISO 8601 UTC
  - Horizontal search domain – latitude/longitude box
  - Vertical search domain above/below sea level
  - Data source
    - Satellite: JPL SMAP L2B v2.0 salinity; ASCAT-B L2 Coastal 12.5 km winds; AVHRR OI L4 GHRSST 0.25˚ and MUR L4 1 km daily sea surface temperature
    - In situ: ICOADS Release 3.0, SAMOS, SPURS-1, 2
  - Platform type (ship, orbiting satellite, etc.)
  - Device type (CTD, current profiler, radiometer, etc.)
  - Mission (Aquarius, ASCAT, MODIS, SAMOS, etc.)
  - Data quality flag - Mapped to IODE standard
- Users also specify spatial and temporal match-up tolerances for locating a match (e.g., within 1 hours and 30 km)

Supporting In Situ Data Quality Filtering

- Each in situ source has a unique quality control (QC) system which needs to be mapped to a common standard.
- DOMS standard applies the IODE primary flags (Table 1, columns 1-3)
- SAMOS mapping
  - SAMOS uses a parametric alphabetic flagging scheme, one flag per value.
  - These flags can be directly mapped to the IODE scheme (Table 2).
  - Original SAMOS flags can be provided as part of metadata string.
- ICOADS mapping
  - ICOADS has an array of 25 indicators including trimming, QC, land-locked, and source exclusion flags
  - Simple one-to-one flag mapping will not work
  - ICOADS applies Enhanced Trimming rules to set IODE flags (Table 1, column 4).
  - 25 original indicators are included in a metadata string with each data record
- SPURS only includes "good" data so all values are mapped to IODE=1.
- Each data provider includes IODE flags within the Apache Solr index.
- The next step is to modify EDGE templates at each institution so DOMS will be able to filter data based on the IODE flags
  - New options will be added to DOMS application programming interfaces and the web user interface

Table 1: IODE primary level flags and ICOADS mapping assignments

Value	*Primary level flag short name*	*Definition*	*ICOADS Mapping*
1	Good	Passed documented required QC tests	Passed Enhanced Trimming quality check
2	Not evaluated, not available or unknown	Used for data when no QC test performed or the information on quality is not available	Not applicable, not used
3	Questionable / suspect	Failed non-critical documented metric or subjective test(s)	Not applicable, not used
4	Bad	Failed critical documented QC test(s) or as assigned by provider	Failed Enhanced Trimming quality check
9	Missing data	Used as place holder when data are missing	Set when data is missing

Table 2: Definitions of the alphabetic flags used in the SAMOS quality control procedures and the mapping to the IODE standard.

IODE Flag	*SAMOS Flag*	*Definition*
4	B	Original data were out of a physically realistic range bounds outlined.
4	D	Data failed the T>=Tw>=Td test. In the free atmosphere, the value of the temperature is always greater than or equal to the wet-bulb temperature, which in turn is always greater than or equal to the dew point temperature.
3	E	Data failed the resultant wind re-computation check. When a dataset includes the platform's heading, course over the ground, and speed over the ground along with platform relative wind speed and direction, a program re-computes the earth relative wind speed and direction. A failed test occurs when the difference between the reported and re-computed wind direction is >20 (or >2.5 m/s for wind speed).
3	F	Platform velocity unrealistic. Determined by comparing sequential latitude and longitude positions.
3	G	Data are greater than 4 standard deviations from the climatological means (da Silva et al. 1994). The test is only applied to pressure, temperature, sea temperature, relative humidity, and wind speed.
3	H	Discontinuity (step) found in the data. Flags assigned to the maximum and minimum points in the discontinuity.
1	I	Interesting feature found in the data. Examples include: hurricanes passing stations, sharp seawater temperature gradients, strong convective events, etc.
4	J	Data are of poor quality by visual inspection, DO NOT USE.
3	K	Data suspect/use with caution – this flag applies when the data look to have obvious errors, but no specific reason for the error can be determined.
4	L	Vessel position over land based on reported latitude and longitude.
4	M	Known instrument malfunction.
3	N	Signifies that the data were collected while the vessel was in port. Typically these data, though realistic, are significantly different from open ocean conditions.
3	Q	Questionable – observation reported as questionable/uncertain in consultation with vessel operator (use with caution).
4	S	Spike in the data. Usually one or two sequential data values (sometimes up to 4 values) that are drastically out of the current data trend. Spikes for many reasons including power surges, typos, data logging problems, lightning strikes, etc.
1	Z	Data passed evaluation.

Vision for the Future

- Up next for DOMS is to fully integrate the technology into the OceanWorks science data analytics platform
  - Developing a common user interface
  - Supporting large job management (e.g., matching requests for entire satellite missions, full ocean basins)
  - Refining CSV and netCDF output formats for both data subsets and matched data to include quality flags
- Expanding data quality filtering to satellite datasets
  - Leverage the Virtual Quality Screening Service developed by JPL
- Further enhancement of DOMS being considered include the following:
  - Supporting satellite-to-satellite and in situ-to-in situ data matching,
  - Supporting satellite/in situ to numerical model matching, and
  - Including additional high-priority science datasets.

Acknowledgements

The DOMS project is supported via NASA's Earth Science Technology Office from the Advanced Information Systems Technology program. Separate awards presently supporting DOMS at FSU (80NSSC17K0417), NCAR (NNX15AG22G), and JPL.