DATASET LIFECYCLE POLICY

Edward Armstrong [1], Eric Tauer [1]

(1) Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
Email: edward.m.armstrong@jpl.nasa.gov

ABSTRACT

The presentation focused on describing a new dataset lifecycle policy that the NASA Physical Oceanography DAAC (PO.DAAC) has implemented for its new and current datasets to foster improved stewardship and consistency across its archive. The overarching goal is to implement this dataset lifecycle policy for all new GHRSST GDS2 datasets and bridge the mission statements from the GHRSST Project Office and PO.DAAC to *provide the best quality SST data in a cost-effective, efficient manner, preserving its integrity so that it will be available and usable to a wide audience.*

1. Dataset Lifecycle Policy

The primary motivation for the PO.DAAC with respect to the implementation of the policy is to ensure consistency across the data holdings with regard to metadata and formats, data quality and maturity, and to ensure requirements for internal data management best practices are followed. Impacts on data, operations, tools and distribution are assessed through the collection of various metrics. The primary components of the lifecycle are defined by a series of documents designed to collect these lifecycle policy metrics (Fig. 1). Some of the metrics are related to internal procedures to document system requirements such as impacts on operations, and tools and distribution (e.g., the System Impact Assessment document), but of fundamental importance to the data provider is a document known as the Submission Agreement. This document is part of the lifecycle "quality gate" designed to improve the capturing of data quality and descriptions. Although the document contains sections to establish the respective expectations between the data provider and the PO.DAAC with regards to data latency, tools and services availability, support and distribution requirements, it more importantly contains sections for the provider to document and improve the quality characterization of their dataset including data uncertainty assessment and validation results, and well as the processing lineage and algorithm description. Components of these sections could come from published literature, project validation results, or project algorithm description documents.

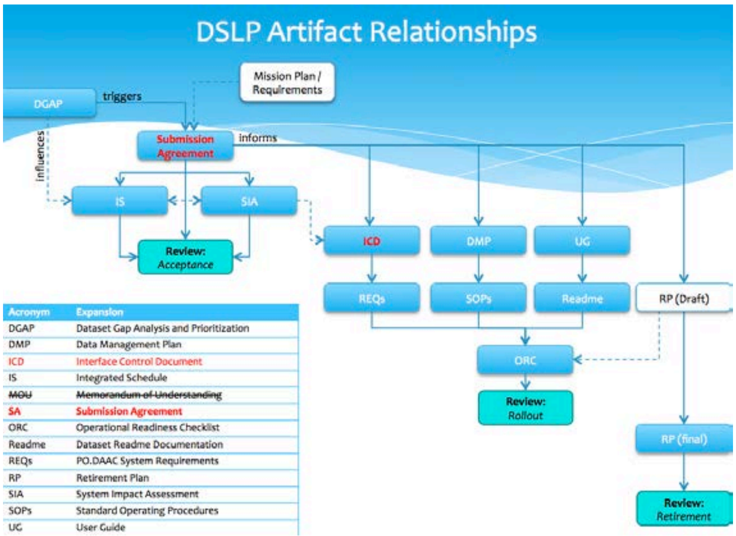
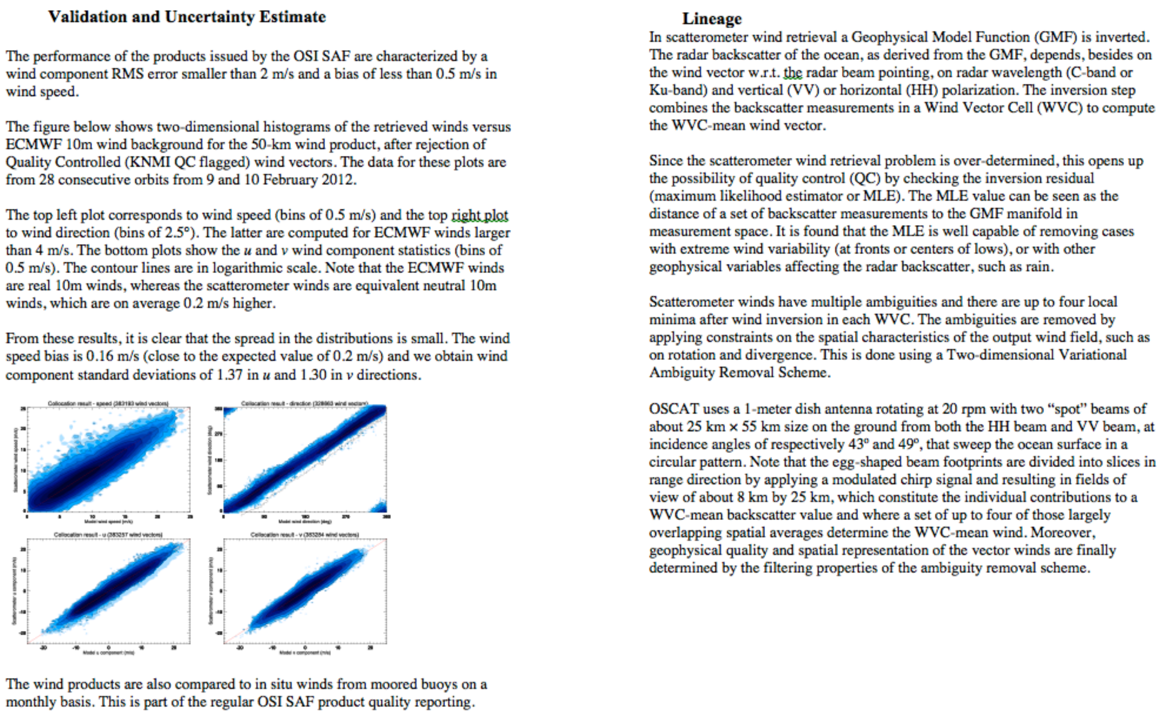


Figure 1. Example of the various facets and documents in the hierarchy of the PO.DAAC dataset lifecycle policy (DSLP). Of importance to a data provider is the Submission Agreement to document data quality.

An example of the populated data quality components in the Submission Agreement is seen in Figure 2 for a Oceansat-2 scatterometer dataset:



Validation and Uncertainty Estimate

The performance of the products issued by the OSI SAF are characterized by a wind component RMS error smaller than 2 m/s and a bias of less than 0.5 m/s in wind speed.

The figure below shows two-dimensional histograms of the retrieved winds versus ECMWF 10m wind background for the 50-km wind product, after rejection of Quality Controlled (KNMI QC flagged) wind vectors. The data for these plots are from 28 consecutive orbits from 9 and 10 February 2012.

The top left plot corresponds to wind speed (bins of 0.5 m/s) and the top right plot to wind direction (bins of 2.5°). The latter are computed for ECMWF winds larger than 4 m/s. The bottom plots show the *u* and *v* wind component statistics (bins of 0.5 m/s). The contour lines are in logarithmic scale. Note that the ECMWF winds are real 10m winds, whereas the scatterometer winds are equivalent neutral 10m winds, which are on average 0.2 m/s higher.

From these results, it is clear that the spread in the distributions is small. The wind speed bias is 0.16 m/s (close to the expected value of 0.2 m/s) and we obtain wind component standard deviations of 1.37 in *u* and 1.30 in *v* directions.

The wind products are also compared to in situ winds from moored buoys on a monthly basis. This is part of the regular OSI SAF product quality reporting.

Lineage

In scatterometer wind retrieval a Geophysical Model Function (GMF) is inverted. The radar backscatter of the ocean, as derived from the GMF, depends, besides on the wind vector w.r.t. the radar beam pointing, on radar wavelength (C-band or Ku-band) and vertical (VV) or horizontal (HH) polarization. The inversion step combines the backscatter measurements in a Wind Vector Cell (WVC) to compute the WVC-mean wind vector.

Since the scatterometer wind retrieval problem is over-determined, this opens up the possibility of quality control (QC) by checking the inversion residual (maximum likelihood estimator or MLE). The MLE value can be seen as the distance of a set of backscatter measurements to the GMF manifold in measurement space. It is found that the MLE is well capable of removing cases with extreme wind variability (at fronts or centers of lows), or with other geophysical variables affecting the radar backscatter, such as rain.

Scatterometer winds have multiple ambiguities and there are up to four local minima after wind inversion in each WVC. The ambiguities are removed by applying constraints on the spatial characteristics of the output wind field, such as on rotation and divergence. This is done using a Two-dimensional Variational Ambiguity Removal Scheme.

OSCAT uses a 1-meter dish antenna rotating at 20 rpm with two "spot" beams of about 25 km × 55 km size on the ground from both the HH beam and VV beam, at incidence angles of respectively 43° and 49°, that sweep the ocean surface in a circular pattern. Note that the egg-shaped beam footprints are divided into slices in range direction by applying a modulated chirp signal and resulting in fields of view of about 8 km by 25 km, which constitute the individual contributions to a WVC-mean backscatter value and where a set of up to four of those largely overlapping spatial averages determine the WVC-mean wind. Moreover, geophysical quality and spatial representation of the vector winds are finally determined by the filtering properties of the ambiguity removal scheme.

Figure 2. The Validation and Lineage sections describing the data quality for Oceansat-2 wind scatterometer (OSCAT).

2. Conclusion

This Submission Agreement as part of the dataset lifecycle policy is meant to be a first step to assess the dataset quality and can be eventually leveraged to improve GHRSST ISO 19115 metadata records (using data quality DQ_ and lineage LE_ objects) as well. It can also potentially be used to improve dataset selection from the user perspective. After plenary discussion it was agreed that new GHRSST datasets should strive to adopt this lifecycle approach including the Submission Agreement.

Acknowledgements

This work was carried out at the NASA Jet Propulsion Laboratory, California Institute of Technology. Government sponsorship acknowledged. Government sponsorship acknowledged.