

Data QA/QC in the IOOS Mid-Atlantic Regional Association Coastal Ocean Observation System

Michael Crowley, M. Smith, L. Nazzaro, H. Roarty, J. Kohut, S. Glenn
Dept. of Marine & Coastal Sciences
Rutgers University
New Brunswick, NJ, USA
crowley@rutgers.edu

Kelly Knee
RPS-ASA
55 Village Square Drive
South Kingstown, RI, USA
Kelly.knee@rpsgroup.com

Matt Oliver
College of Earth, Ocean & Environment
University of Delaware
700 Pilottown Road
Lewes, DE, USA
moliver@udel.edu

Gerhard F. Kuska
MARACOOS
318 South College Ave.
Newark, DE, USA
kuska@maracoos.org

Abstract— The Mid-Atlantic Regional Association Coastal Ocean Observing System (MARACOOS) is one of the eleven Regional Associations (RAs) comprising the coastal network of the U.S. Integrated Ocean Observing System (US IOOS). MARACOOS involves participants from academia, government, the private sector, and non-profit entities, and covers the ocean and estuaries from Cape Cod, MA to Cape Hatteras, NC.

The high quality of MARACOOS-served data is predicated on the quality assurance and quality control (QA/QC) and data flow of the internal and external data streams that MARACOOS makes available through an interactive map/data display (oceansmap.maracoos.org) as well as through THREDDS servers. These data are from (1) High Frequency Radar (HF-Radar), (2) Ocean Gliders, (3) the Advanced Very High Resolution Radiometer (AVHRR) satellite imagery, (4) the Hudson River Environmental Conditions Observing System (HRECOS), and (5) the Maryland Department of Natural Resources (Maryland DNR). This documented and sustained QA/QC elevates MARACOOS-served data to be federally-equivalent—a pre-requisite to becoming a certified Regional Information Coordination Entity (RICE), which MARACOOS achieved in December 2016.

Keywords— U.S. IOOS, MARACOOS, QA/QC, QARTOD, HF-Radar, ocean gliders, satellite, IOOS certification, DMAC

I. INTRODUCTION

Along the Mid-Atlantic coastal region (Cape Cod, MA to Cape Hatteras, NC), the Mid-Atlantic Regional Association Coastal Ocean Observing System (MARACOOS) is the front line for federally-certified ocean and coastal observing, data management and provision of quality controlled data from a variety of sources to address societal needs.

MARACOOS, established in 2004, is a 501(c)3 association of 25 universities, 21 state and federal government agencies, 18 private sector partners, and 12 non-governmental entities that, along with the other 10 regional associations across the country, comprise the non-federal element of the U.S. Integrated Ocean Observing System (US IOOS).

MARACOOS' ability to achieve and consistently maintain federal data quality standards led to its certification as a

Regional Information Coordination Entity (RICE) in December 2016. This government seal of approval assures users of MARACOOS data that these data meet or exceed federal standards for quality.

A. Federal Certification through U.S. IOOS

The Integrated Coastal Ocean Observing System (ICOOS) Act of 2009 directs NOAA to “certify and integrate RICEs into the U.S. IOOS.” Integration into the System formally “establishes the role of the RICE in the U.S. IOOS and ensures that the data collected and distributed by the RICE are managed according to the best practices, as identified by NOAA.” <https://ioos.noaa.gov/about/governance-and-management/certification/>. A key part of a RICE's system is a Data Management and Communications Plan, documenting how a RICE maintains and controls data quality and distribution (Dept. of Commerce, Ref. 1).

II. DATA ACCESS

The MARACOOS Data Management and Communication (DMAC) system is a robust, services-oriented system that includes tools to auto-aggregate, quality-control, and monitor data from a variety of observational and numerical model assets. The system includes (1) management of *in situ* and remotely-sensed observation and model data; (2) establishment of data-model connections for assimilation and comparison; (3) implementation and management of standards-based data access services such as Thematic Real-time Environmental Distributed Data Services (THREDDS), Open-source Project for a Network Data Access Protocol (OPeNDAP), Web Mapping Services (WMS), in particular sciWMS, and Sensor Observation Services (SOS), in particular ncSOS; (4) provision of data in approved common data formats such as Network Common Data Form (NetCDF); (5) provision of infrastructure for hosting data and services; (6) delivery of data to operational users; and (7) coordination of data archiving efforts with the National Centers for Environmental Information (NCEI).

MARACOOS has adopted a centralized data aggregation approach, in the form of a regional Data Assembly Center (DAC), for model and remotely-sensed observation data. MARACOOS is in the process of implementing a centralized approach for *in situ* observation data, which will involve harvesting recent observation data onto MARACOOS infrastructure. Our centralization efforts include management of data generated by MARACOOS partners such as Rutgers University and the University of Delaware, as well as data collected by non-partner organizations such as the Hudson River Environmental Conditions Observing System (HRECOS) and Maryland Department of Natural Resources (DNR). Where applicable, MARACOOS data will continue to be delivered to national data assembly centers (DACs) (GliderDAC and the National High Frequency RADAR [HF RADAR] Network). By aggregating data into the regional and national DACs, we can better manage data quality, distribution, and services to help enable open data sharing. While several federal sources of *in-situ* observation data (NDBC, CO-OPS, NERRS, USGS, CBIBS) are accessible via MARACOOS, these sources are managed by federal partners and are not discussed further in this paper.

The DMAC system facilitates discovery, access, and understanding of regional and relevant national observation and model data. All MARACOOS services are registered with National Geophysical Data Center (NGDC) and listed in the IOOS catalog, allowing them to be discovered by users. Data can be accessed via Data Access Protocol (DAP), WMS, and SOS, and visualized in the [MARACOOS OceansMap](#) platform.

The MARACOOS DMAC system uses a combination of the Environmental Data Server (EDS) and THREDDS to manage gridded data. The MARACOOS DMAC system also supports quality assurance. MARACOOS is a leader in the implementation of Quality Assurance of Real-Time Oceanographic Data (QARTOD), having implemented QARTOD processes for glider and HF RADAR data.

III. DATA STREAMS AND QA/QC

As each data stream provided via MARACOOS is unique, this section describes a selection of MARACOOS data sources individually. In total, five data streams are described. These include (1) High Frequency Radar (HF Radar), (2) Ocean Gliders, (3) Advanced Very High Resolution Radiometer (AVHRR), (4) the Hudson River Environmental Conditions Observing System (HRECOS), and (5) the Maryland Department of Natural Resources (DNR) water quality monitoring stations. The descriptions for each data stream include the data flow, quality control procedures, adherence to the NOAA Data Sharing Procedural Directive, and the existing or planned archiving procedure.

A. HF-Radar

The Mid-Atlantic High Frequency Radar Network consists of approximately 46 radars operated by seven separate organizations (Old Dominion University, Rutgers - The State University of New Jersey, University of Connecticut, University of Massachusetts (Dartmouth), University of North Carolina (Chapel Hill), University of Rhode Island, and Woods Hole Oceanographic Institution). The network provides hourly measurements of ocean surface currents (i.e., representative of water depths of 0.3 – 2.5 m) within ~250 km of the coast, an area encompassing more than 190,000 km² of the ocean’s surface, at resolutions of 1 km, 2km, and 6km.

1) HF-Radar Data Flow

At each radar station, ocean surface currents within the station’s coverage zone are measured and reported in terms of radial velocities (i.e., the component of velocity directed towards the station) on an hourly basis. A flow chart documenting the flow of data from the radar stations is shown in Figure 1.

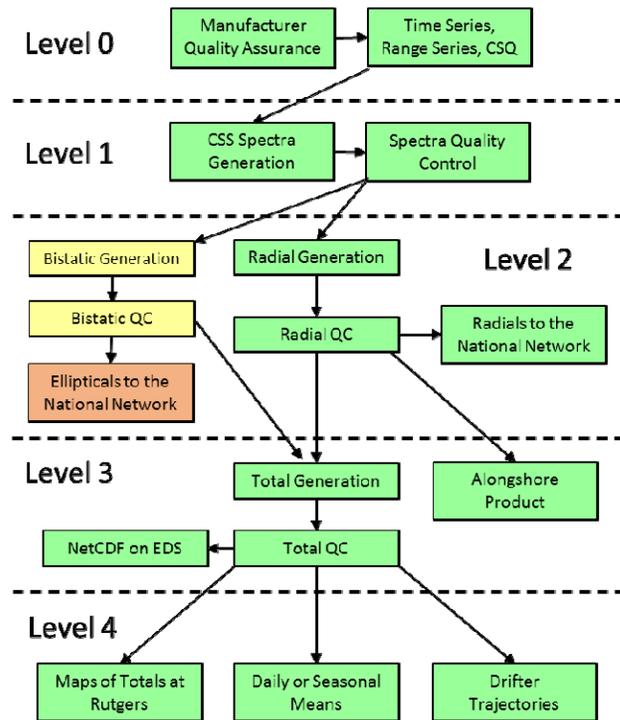


Figure 1: Diagram showing the flow of data from initial measurement at the radar station (top) to derived product from the surface current measurement (bottom).

The flow chart is broken into five levels following the convention from the [NASA Earth Science Reference Handbook](#) for describing Standard Data Products (SDPs). Level 0, 1 and the radial generation step of Level 2 all occur at the individual radar station. The radial vector data are transferred electronically from the station to a regional data assembly center at Rutgers, where they undergo quality control checks. Once the radials are quality controlled they

are sent to the [NOAA National HF Radar Network](#). NOAA then produces a total vector product via the unweighted least-squares method. This product is available from [National HF Radar Network](#), where it can be viewed on a map, embedded in a third party map via the [mapping API](#) or accessed via [THREDDS](#) or [FTP](#). The EDS also collects the National HF RADAR Network product to support U.S. Coast Guard search and rescue operations. The data is harvested from the national [THREDDS](#) server, published as an Open Geospatial Consortium (OGC) compliant WMS service and made available on the [OceansMap THREDDS server](#) via OPeNDAP. While no changes to the data are made during the EDS harvesting process, only the variables needed to support search and rescue operations are maintained in the EDS NetCDF file. For HF RADAR data these include only the components of velocity (u, v) plus the time and latitude and longitude dimensions. No unit transformations are made and the original grid is maintained. The EDS makes the data available via OGC compliant WMS services and also provides OPeNDAP access via [THREDDS](#).

To satisfy additional user requests, MARACOOS also produces an alternative product that uses optimal interpolation (Kohut et al., 2012, Ref. 2) to combine the radial vectors into a total vector product. The algorithm uses a search area that is tailored to the geometry of the Mid-Atlantic Bight shelf to generate ASCII files of the total vectors. The total vector data are then transformed into NetCDF, without any other modification of the data, and made publicly available on a Rutgers [THREDDS](#) server.

The total vector data are harvested from the Rutgers [THREDDS](#) server on an hourly basis by RPS ASA's EDS so that they can be made available to the U.S. Coast Guard for use in the Search and Rescue Optimal Planning System (SAROPS). The EDS harvests the NetCDF files directly via OPeNDAP and stores the data in the EDS Data Store as daily NetCDF files. As with harvesting of the national total vector product, no transformations or modifications of the data are made during this process, however the components of velocity (u, v) are the only variables retained. The EDS makes the data available via OGC compliant WMS services and also provides OPeNDAP access via [THREDDS](#).

2) HF-Radar Quality Control

MARACOOS follows and applies a variety of quality assurance and quality control procedures to the data before passing it to the National HF RADAR Network. As described in Figure 1, the quality control is applied during Level 2, after radials are generated and before they are passed to the National Network. The procedures followed include:

- [A Plan to Meet the Nation's Needs for Surface Current Mapping](#), US IOOS, April 2013
- Deployment & Maintenance of a High Frequency Radar for Ocean Surface Current Mapping: Best Practices, February 2008
 - [PDF Document](#)
 - [Wiki Document](#)

- [Mid Atlantic Regional Coastal Ocean Observing System \(MARACOOS\) HF Radar Quality Control/Quality Assurance Document](#), April 2008
- [Mid Atlantic Regional Coastal Ocean Observing System \(MARACOOS\) HF Radar Quality Control/Quality Assurance Document](#), April 2010
- [MARACOOS Quality Assurance and Quality Control for High Frequency Radar Systems](#), April 2012
- [Rutgers University Standard Operating Procedure for SeaSonde HF Radar Quality Control](#), May 2015
- [QA/QC & Related Practices at CODAR](#), June 2012

In addition to the quality control described in the above documents, for the MARACOOS product the procedures outlined in the [US IOOS HF RADAR QARTOD QC manual](#) (May 2016) are also being applied to the radial data. All seven of the required tests are being implemented; these include syntax, signal-to-noise ratio, maximum threshold, maximum radial and total velocity threshold, valid location, data density, and geometry threshold tests. Inclusion of the quality-controlled radials in the total vector product produced by MARACOOS is currently being evaluated. MARACOOS expects that the quality-controlled radials will be included in the total vector product by the end of 2017.

3) HF-Radar Data Sharing

Both total vector products (those created by the National HF RADAR Network and those created by MARACOOS) adhere to the NOAA Data Sharing Procedural Directive by providing access to near real-time data via a variety of access points at no cost and in a format that is machine-readable, based on open standards, and includes International Organization for Standardization (ISO) 19115 metadata.

The National HF RADAR Network total vector product is available from their [website](#). The site provides an embedded map allowing users to view the vectors. It also provides direct access to the numeric data via [THREDDS](#), [FTP](#), and downloadable shapefiles. The [THREDDS](#) instance provides OPeNDAP, WCS, NetCDF, and WMS access as well as ISO metadata. The [FTP](#) provides access to the most recent three days of data. The data can also be visualized by downloading Keyhole Markup Language (KML) files or using the [mapping API](#), which provides access to tiles of the data.

The RPS ASA EDS also collects the National HF RADAR Network product to support USCG search and rescue operations. The data are harvested from the national [THREDDS](#) server, published as a WMS service and made available on the [OceansMap THREDDS server](#) via OPeNDAP.

The MARACOOS total vector data is available on [THREDDS](#) server at Rutgers, which provides OPeNDAP access. The 6km data is also available via OPeNDAP from the [OceansMap THREDDS server](#).

HF Radar data for the region can also be visualized using the [MARACOOS OceansMap](#) platform. There are five HF RADAR products available in OceansMap:

- HF Radar USEC - National HF Radar Network total vector product, harvested by EDS and visualized using the EDS WMS service
- HF Radar MARACOOS – MARACOOS total vector product, harvested from Rutgers by EDS and visualized using the EDS WMS service
- UCSD HF Radar 1 km - National HF Radar Network total vector product, visualized using the National HF Radar Network tiles (GetFeatureInfo unavailable for this layer)
- UCSD HF Radar 2 km - National HF Radar Network total vector product, visualized using the National HF Radar Network tiles (GetFeatureInfo unavailable for this layer)
- UCSD HF Radar 6 km - National HF Radar Network total vector product, visualized using the National HF Radar Network tiles (GetFeatureInfo unavailable for this layer)

4) HF-Radar Archiving

Quality-controlled radial velocity data are uploaded to the National HF Radar Network, where it is processed into total vector products as described above. The total vector products are archived by NCEI directly from the national server. The timing of archive submission is determined by the National Network.

MARACOOS is currently working with NCEI to archive the MARACOOS total vector product. The product will be archived daily from NetCDF files available in the MARACOOS Web Accessible Folder (WAF). The MARACOOS DMAC team is currently working with the NetCDF files to ensure that they meet the requirements of the NCEI 2.0 templates.

B. Ocean Gliders

The Mid-Atlantic Glider Network is coordinated through a consortium of glider operators working within MARACOOS. The MARACOOS glider operators include Rutgers University, the University of Delaware, the University of Maryland, The Virginia Institute of Marine Sciences, and the University of Massachusetts Dartmouth. The primary measurements provided by the gliders are temperature and conductivity. Derived parameters include salinity and density.

The glider operations are coordinated by the team (institutions listed previously) to ensure the MARACOOS glider operations serve regional needs. The team also informs the timing and mission flight strategy for the gliders. Currently MARACOOS funds ~20% of the glider flights operating within the Mid-Atlantic. The other glider flights are funded by other federal and state programs, which determine the gliders’ flight mission strategy and timing. The advantage for these programs is that the MARACOOS community provides a centralized glider community with a wide range of technical expertise and hardware resources that are shared.

1) Glider Data Flow

To facilitate coordination of gliders across the IOOS Regional Associations and the nation, the IOOS office has developed the [Glider Data Assembly Center \(DAC\)](#). The Glider DAC allows users to submit glider data to a centralized location from which it is made widely available via web services and archived with NCEI. MARACOOS glider data flows into the IOOS Glider DAC. A flow chart documenting the flow of data through the DAC is shown in Figure 2.

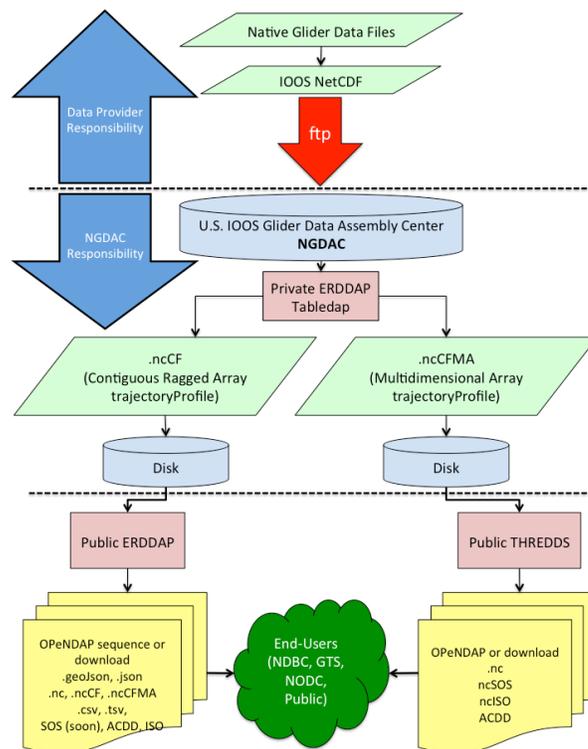


Figure 2: Diagram of data flow from gliders to the IOOS Glider DAC.

During surfacings, binary data files are sent from the glider via satellite. These files are converted to ASCII (one for each segment of the deployment) and combined into a single file containing all data from the deployment. During this step, individual profiles are defined and converted to NetCDF. Glider data are transferred to the Glider DAC after the profiles are written to NetCDF. To ensure the files are appropriately formatted they are assessed for compliance using the [IOOS NetCDF compliance checker](#). Uploading the data to the Glider DAC requires that the deployments have been registered, which requires acquisition of a Registration ID and a World Meteorological Organization (WMO) ID. WMO IDs are requested from the Glider DAC team. Deployments are registered using the [Glider DAC Provider Page](#). Data is then uploaded via FTP. Data is then made available via Environmental Research Division's Data Access Program (ERDDAP) and THREDDS within 2 hours of being uploaded. Current data access end-points are found at:

- o ERDDAP: <http://data.ioos.us/gliders/erddap/tabledap/>

- THREDDS: <http://data.ioos.us/gliders/thredds/catalog/deployments/catalog.html>
- IOOS Glider Map: <http://gliders.ioos.us/map>
- MARACOOS OceansMap: <http://oceansmap.maracoos.org>

2) *Glider Quality Control*

MARACOOS follows and applies the Quality Assurance (QA)/Quality Control (QC) procedures outlined in the following documents to all gliders:

- [Deployment Procedures](#)
- [Daily Monitoring Procedures](#)
- [Glider Piloting](#)
- [Oxygen Calibration](#)

In addition, MARACOOS is in the process of implementing IOOS QARTOD Standards, described in IOOS QARTOD standards [Manual for Quality Control of Temperature and Salinity Data Observations from Gliders](#). Implementation is expected to be complete by Summer 2017. The following required tests are applied: Timing, Syntax, Location, Gross Range, Climatology and Pressure. The following strongly recommended tests are also applied: Spike, Rate of Change and Flat line.

The timing, syntax, and location tests are performed on the binary data files sent from the glider during surfacings. Any binary files passing those tests are converted to ASCII files (one for each segment of the deployment), which are combined into one file containing all data from the deployment. During this step, individual profiles are defined.

As new segments are added to a full deployment file, the gross range, climatological, spike, rate of change, flat line, and monotonic pressure tests are applied to any new data; flag values for each test are added to the file, but all original data is retained. A NetCDF file for each individual profile is then created containing the original data and corresponding flags.

3) *Glider Data Sharing*

Through the Glider DAC, the Mid-Atlantic Glider Network adheres to the NOAA Data Sharing Procedural Directive by providing access to near real-time data via a variety of access points at no cost and in a format that is machine-readable, based on open standards, and includes ISO 19115 metadata.

The Glider DAC provides direct access to the Mid-Atlantic Glider Network data via [ERDDAP](#) and [THREDDS](#). [ERDDAP](#) allows users to subset and plot data on the fly and access data directly via OPeNDAP. The [THREDDS](#) instance provides access to the glider data via OPeNDAP and SOS. Metadata is also accessible via [THREDDS](#). The DAC also provides visualizations of Glider data using the [IOOS Glider Map](#).

The [MARACOOS OceansMap](#) platform includes the Atlantic glider data and provides access to the tracks, associated time frames, attributions, and cross-section plots for each parameter.

4) *Glider Archiving*

Quality-controlled glider data are uploaded to the NOAA Glider DAC. The glider data are archived by NCEI directly from the DAC. The timing of archive submission is determined by the DAC.

C. *AVHRR Satellite Imagery*

The MARACOOS Advanced Very High Resolution Radiometer (AVHRR) data stream is coordinated through the University of Delaware, where it is collected and distributed in near-real time. AVHRR is aboard the Polar-orbiting Operational Environmental Satellites (POES) system. POES provides daily global coverage by making nearly polar orbits 14 times per day and provides visible, infrared, and microwave data, which is used for a variety of applications such as cloud and precipitation monitoring, determination of surface properties, and humidity profiles (NOAA, 2016, Ref. 3). MARACOOS provides a subset of AVHRR sea surface temperature (SST) data covering the Gulf of Mexico and the Eastern Seaboard at higher resolution (1 km) than most federal products (4km). The data stream is accessed by ocean models and used for regional research and product development.

1) *AVHRR Data Flow & Data Sharing*

The MARACOOS AVHRR data stream is coordinated through the University of Delaware, with a sub-regional technology center at Rutgers University. The hardware for this system currently includes one dual-use L/X-Band dish at the University of Delaware and a second dual-use L/X-Band dish at Rutgers. Computing infrastructure includes two reception computers at Rutgers University and two reception computers, three processing nodes, and one computer cluster at the University of Delaware. If the dishes miss an AVHRR pass, the data is filled in using the [NOAA CLASS](#) system.

The data flow from raw AVHRR SST data to the processed NetCDF-4 files stored on the [MARACOOS THREDDS Data Server \(TDS\)](#) is illustrated in Figure 3. The raw SST data is downloaded from the NOAA AVHRR satellites using MARACOOS' L-Band capabilities. It is processed with standard SeaSpace TeraScan software, using default coefficients for cloud masking, which produces NetCDF-3 files. Using a series of R scripts developed by the University of Delaware, the data is then projected into latitude and longitude coordinates, the experimental regionally-specific mask is generated, and the data is converted to NetCDF-4. The standard TeraScan outputs and experimental regionally-specific mask product are pushed as NetCDF-4 files to the [MARACOOS TDS](#). As single AVHRR pass comes into the [MARACOOS TDS](#), two processes are run:

- 1) Masking/De-Clouding: The NetCDF Operators (NCO) tool, NetCDF Weighted Averager (ncwa) is used to apply a mask to the input (unmasked file) and flag all points that have failed any test run by University of Delaware. This process is run anytime a new unmasked file is added to the system
- 2) Time Compositing: For each time interval (1-, 3-, and 7-day) an average in time (by pixel) is done using the NCO tool, NetCDF Ensemble Averager (ncea, now known as nces). Files within each interval are determined by moving back in time from the time of the satellite pass specified in the filename, and the average is output to a file corresponding to those two times. This is done every time a new masked file is written by the masking/de-clouding process.

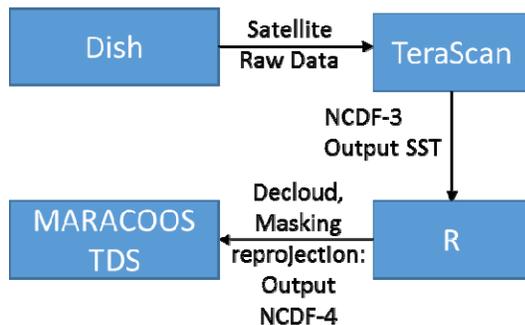


Figure 3. Data flow for MARACOOS AVHRR SST products.

A number of AVHRR SST products are provided via the MARACOOS TDS. These include individual passes, composites, and aggregations. Composites are created by averaging each time-step in a dataset over a period of time. For example, a 7-day composite is the gridded average of available values over the last seven days. Aggregations are single, long-term datasets, rather than individual satellite passes or composite files. The aggregations are virtual datasets, the raw data are stored as individual files by satellite pass or composite. As described above, for the AVHRR source, time composites are created from the individual satellite passes directly on the THREDDS server.

The MARACOOS TDS stores the AVHRR data in two separate directories, REALTIME and ARCHIVE. The REALTIME directory provides symlinks to aggregations of the last 8 days of data stored in the ARCHIVE directory. The symlinks are updated every 30 minutes. This arrangement allows users to access real-time data from short aggregations without stressing system resources by accessing long-term aggregations. The ARCHIVE directory stores both individual single pass and composite files and single pass and composite aggregations in yearly directories.

The full list of available products available on the MARACOOS TDS includes:

1. REALTIME
 - a. Aggregations of the 1-, 3-, and 7-day composites
 - b. Aggregation of the unmasked single pass

2. ARCHIVE
 - a. Masked - AVHRR SST with the experimental regional cloud mask applied (the masking process is described in the Quality Control section)
 - i. Single Pass Aggregation: Annual aggregations of single passes, 2013 to present
 - ii. Single Pass Individual Files: individual satellite passes, 2013 to present
 - iii. 1-day-Aggregation: annual aggregations of 1-day composites from 2013 to present
 - iv. 1-Day Individual Files: individual 1-day composites from 2013 to present
 - v. 3-day-Aggregation: annual aggregations of 3-day composites from 2013 to present
 - vi. 3-Day Individual Files: individual 3-day composites from 2013 to present
 - vii. 7-day-Aggregation: annual aggregations of 7-day composites from 2013 to present
 - viii. 7-Day Individual Files: individual 7-day composites from 2013 to present
 - b. Unmasked – Regional experimental cloud mask is provided, but not applied
 - i. Single Pass Aggregation: aggregations of single passes for the current year
 - ii. Single Pass Individual Files: individual satellite passes for the current year

2) AVHRR Quality Control

MARACOOS applies a variety of quality control procedures to the masked AVHRR SST product before publishing it on the MARACOOS TDS. For each satellite pass, three statistical tests are applied to each pixel. The results of these tests are coded into eight different data quality flags. These tests include:

1. Time Test: A measurement fails the time test if it is more than 2 degrees cooler than the mean of the satellite passes in the previous 72 hrs.
2. Climatology Test: A measurement fails if the temperature is more than 5 degrees colder than the climatologies from the NSIPP AVHRR Pathfinder and Erosion Global 9km SST Climatology.
3. Sequential time and climatology test: In this test, the previous 72 hours of satellite passes are first passed through a climatology test, then passed through the time test. It is a sequential test, rather than two independent tests.

If any measurement fails one of these three tests, all neighboring measurements within 3km are also automatically flagged. The flags for the time, climatology, and sequential test are 1, 2 and 4 respectively. The integer sum of these tests tells the user which of the three tests any pixel failed.

Interpretation of the flags:

1. 0 = passed all tests 0+0+0
2. 1 = failed only time test 1+0+0=1

3. 2 = failed only climatology test $0+2+0=2$
4. 3 = failed time test and climatology test $1+2+0=3$
5. 4 = failed only sequential test $0+0+4=4$
6. 5 = failed time test and sequential test $1+0+4=5$
7. 6 = failed climatology test and sequential test $0+2+4=6$
8. 7 = failed time test, climatology test and sequential test $1+2+4=7$

Pixels that failed the time test could have flags of 1, 3, 4, 7; pixels that failed the climatology test could have flags of 2, 3, 6, 7; pixels that failed the sequential test could have flags of 4, 5, 6, 7.

The experimental regionally-specific cloud masking algorithms were developed by comparing satellite SST measurements with NDBC SST measurements. Sea surface temperature data was downloaded for 205 NDBC sites in the mid-Atlantic region using NOAA's [Coastwatch](#) site for the time period from Jan 1, 2006 – September 14, 2014. While the data from NDBC is initially quality controlled, it was further quality controlled by visual inspection of each data set. The hourly buoy data was then interpolated to the satellite pass time. For each site, a root mean square (RMS), bias and observation number statistic was calculated. The results of these comparisons suggest that the regional de-clouding algorithms should eliminate any pixel with a flag greater than 3. Thus, only pixels that passed all tests (flag = 0), failed only the time test (flag = 1), failed only the climatology test (flag = 2) or failed both the time and climatology test (flag = 3) are kept in the regional de-clouding process.

3) AVHRR Archiving

MARACOOS has discussed the possibility of archiving AVHRR products with NCEI and expects that the data will eventually be archived by NCEI. While NCEI is interested in archiving these data products eventually, they are not prepared to do so at this time and no formal process has been initiated.

MARACOOS stores some historical AVHRR data on MARACOOS infrastructure and makes it available on the [MARACOOS TDS](#). These data include:

1. A rolling 3-year store of Single Pass AVHRR SST files with the experimental regional mask applied
2. Individual AVHRR SST files without the experimental regional mask applied for 2016
3. A rolling 3-year store of the individual 1-, 3-, and 7-day composite files with the regional mask applied

The amount of AVHRR data stored by MARACOOS is re-evaluated semi-annually and adjusted based on available storage. As the national data from which the MARACOOS products are derived are archived in perpetuity, data removed from the MARACOOS store can be recreated if necessary.

D. HRECOS

HRECOS has operated a collection of fixed, continuous, meteorological and water quality monitoring stations on the Mohawk and Hudson Rivers since 2008. The system is operated by a consortium of government, academic, and private institutions. All stations operate year-round except the Mohawk at Ilion station due to seasonal drawdown in the canal, and other stations that risk damage due to ice conditions. The system continuously monitors a total of 15 parameters, including water temperature, specific conductivity, salinity, dissolved oxygen, depth, water elevation (for those stations that are surveyed to an elevation datum), pH, turbidity, air temperature, barometric pressure, relative humidity, wind speed, wind direction, radiation, and rainfall, recording data at 15-minute intervals.

1) HRECOS Data Flow

Data are recorded at each water quality sonde and weather station by a data logger and transmitted to the HRECOS database via satellite, cellular modem, wireless transmitter, or direct internet connection. At all stations, individual sensors record data internally at 15-minute intervals. Once ingested to the database, daily rainfall accumulations, water elevation relative to the North American Vertical Datum of 1988 (NAVD88), dew point, and corrected depth are calculated in near-real time and automatic flags are applied (flagging system described below). The near real-time data are then made available on the [HRECOS Current Conditions page](#).

In December 2013, MARACOOS began harvesting data from the HRECOS database onto the [MARACOOS SOS](#) server where it is stored in NetCDF format and made accessible via OPeNDAP and ncSOS. The data is harvested hourly using a Python script that uses a RESTful interface to retrieve and parse data provided by the Stevens Institute of Technology. The script maps the data provided to CF 1.6 accepted standard names and then creates NetCDF and NCML aggregation files that are both NCEI 2.0 and CF 1.6 compliant. In the process of converting the data to a compliant NetCDF format, wind speeds and gusts are converted from knots to m/s, air pressures are converted from Hg to Pa, and conductivity is converted from mS/cm to S/m. The scripts also check that all parameters are present in each of the data requests and that every time step is available for each parameter. If a time step or parameter is missing from the data response, the NetCDF is filled with FillValues (-999.9).

The SOS server stores individual daily files for each station and provides time-aggregations of the entire data store by station. The NetCDF files can be accessed via OPeNDAP and ncSOS and include ISO 19115-2 metadata. The THREDDS instance is monitored by RPS ASA; any dataset more than 30 hours old is reported and RPS ASA personnel follow up with HRECOS to determine where the data flow failure is occurring.

MARACOOS OceansMap displays the HRECOS stations. OceansMap allows users to query any station to obtain values, create time-series plots, and compare the observations with overlapping model data. Data can also be accessed directly via OceansMap. Each data stream is linked back to its source, in the case of HRECOS the stream is linked to the MARACOOS SOS server. Time-series data can be downloaded directly from OceansMap in comma separated value (CSV) format. During the conversion from NetCDF to CSV no other transformations or modifications of the data are made.

Currently MARACOOS only collects real-time data from HRECOS. As the NCEI archiving process proceeds, the MARACOOS DMAC team will consider how to also provide access to the quality-controlled annual datasets.

2) HRECOS Data Sharing

The HRECOS data stream adheres to the NOAA Data Sharing Procedural Directive by providing access to near real-time data via a variety of access points at no cost and in a format that is machine-readable, based on open standards, and includes ISO 19115 metadata.

The HRECOS data is available both from HRECOS directly and via the MARACOOS SOS server. HRECOS shares their data directly at <http://www.hrecos.org/>. This site provides a map of all the stations, tools for plotting and downloading for current conditions at each station (<http://hudson.dl.stevens-tech.edu/hrecos/d/index.shtml>), and the ability to download final flagged data (http://www.hrecos.org/index.php?option=com_content&view=article&id=143&Itemid=54). The final data is uploaded to the HRECOS website as annual files approximately six months after the end of the year. Final data is inserted into the HRECOS database for public download within approximately six months of the end of each quarter. Quarterly data for the current year is available from the HRECOS Current Conditions page. The titles of plots available from this page indicate whether the data has been verified. Annual data can be downloaded from the HRECOS Historical Data store, which provides final data by station and year.

MARACOOS harvests near-real time data directly from the HRECOS database and stores it on the MARACOOS SOS server in NetCDF format. The SOS server stores individual daily files for each station and provides aggregations of the entire store by station. The NetCDF files can be accessed via OPeNDAP and ncSOS and include ISO 19115-2 metadata. Currently MARACOOS only collects real-time data from HRECOS.

The MARACOOS OceansMap platform includes the HRECOS data and allows users to query both current and historical values at any station, create time-series plots, and compare the observations with overlapping model data. Each data stream is linked back to its source, in the case of HRECOS the stream is linked to the MARACOOS SOS

server. Time-series data can also be downloaded directly from OceansMap in CSV format.

3) HRECOS Quality Control

The HRECOS QAPP describes the objectives and criteria and specifies the measures and corrective actions used to ensure data quality. HRECOS strives to record and accept at least 90% of total possible observations and specifies precision and accuracy criteria for each parameter (see Table 4 of the HRECOS QAPP). Water quality sensors and weather stations are recalibrated if they do not meet the standards specified in the QAPP. Standards are in place both for the measured values and their associated timestamps. As described below, HRECOS has both an automated real-time flagging process and a manual quarterly process in place.

The data flags used for the automated real-time data flagging process are summarized in Table 1 and apply to all parameters. The seasonal averages are calculated based on the previous year's data.

TABLE I. AUTOMATED HRECOS DATA FLAGS

Flag Value	Definition
0	Acceptable data
5	Data that demonstrate a significant increase or decrease from the previous value
6	Flat-lined data
30	Data outside three standard deviations of the seasonal mean
40	Data outside four standard deviations of the seasonal mean
100	Data outside the range of the instrument

TABLE II. HRECOS FINAL FLAG VALUES

Flag Value	Definition
0	Data determined to be acceptable after a final review by the site manager
10000	Suspicious data according to a final review by the site manager
20000	Corrected data
500000	Rejected data according to a final review by the site manager

The recorded data is also reviewed and flagged as necessary on a quarterly basis. Within six months of the end of the quarter the HRECOS coordinator adds missing data, recalculates automatic flags using actual seasonal averages, and compiles data for review by station operators and research assistants. Station operators and/or research assistants then review the automatic flags following the Water Quality data review and editing Standard Operating Procedure (HRECOS SOP 2012.05) and the Weather data review and editing Standard Operating Procedure (HRECOS SOP 2012.06) and

makes a final determination for the public record by applying the final flag values described in Table 2. When the quarterly flagging procedures are completed they are uploaded to the HRECOS database.

4) HRECOS Archiving

The HRECOS archival package contains both the provisional and the verified data. The file names, metadata, and documentation clearly indicate whether the data files contain provisional or verified data. All HRECOS parameters follow the same procedure for archiving.

The provisional data is archived daily from the NetCDF files available in the MARACOOS WAF. The provisional nature of the data is documented within the NetCDF files and metadata. The DMAC team worked closely with NCEI to review these files to ensure that they meet the requirements of the NCEI 2.0 templates.

The historical verified data has also been provided to NCEI. As new verified data is made available (expected annually) the archive will be updated with these files. These verified CSV files will be harvested by the MARACOOS DMAC team and converted to the same NetCDF file format used for the provisional data. The annual NetCDF files will be stored in the MARACOOS WAF along with the aggregated provisional files.

E. Maryland DNR

The Maryland Department of Natural Resources (DNR) has operated a collection of fixed, continuous, water quality monitoring stations within the Chesapeake Bay and its tidal tributaries since 1998. The geographic distribution of sites varies from year to year as stations have been added to the system or, in some cases, relocated. The stations are typically deployed from April through October, with a subset of stations deployed year-round insofar as ice conditions allow. The stations continuously measure a total of nine parameters, including water temperature, specific conductance, salinity, dissolved oxygen, turbidity, fluorescence and total chlorophyll, pH, and depth, recording data at 15-minute intervals. At a subset of stations, blue-green algal concentrations and/or chlorophyll a have been measured and recorded during some periods of time.

1) Maryland DNR Data Flow

Data are available from both autonomous logging and telemetered (i.e., real-time) stations. All monitoring data are maintained in a Shallow Water Monitoring database as part of Maryland DNR's Water Quality Monitoring Program.

At all stations, individual sondes record data internally at 15-minute intervals. Sondes are retrieved, calibrated, and replaced during site-servicing visits that range in frequency from bi-weekly to monthly, depending upon the station. Data are uploaded directly to a personal computer upon the return of a sonde to the field office following its retrieval. After the

data from a deployment have undergone QA/QC checks, the QA/QC'd data files are transferred to an internal network directory where they are automatically inserted into DNR's Shallow Water Monitoring database.

At telemetered stations, data is transmitted hourly to a server located at the Chesapeake Bay Program office which then uploads data to the DNR's Shallow Water Monitoring database. To ensure the quality of the data in the Shallow Water Monitoring database, the real-time (telemetered) data is replaced with the QA/QC'd data generated after continuous monitoring sonde replacement.

MARACOOS harvests data from the Maryland DNR Shallow Water Monitoring database onto the MARACOOS SOS server in NetCDF format.

The MARACOOS SOS server stores individual daily files for each station and provide time-aggregations of the entire data store by station. The NetCDF files are available via OPeNDAP and ncSOS and include ISO 19115-2 metadata. The THREDDS instance is monitored by RPS ASA. Any dataset more than 30 hours old is reported and RPS ASA personnel follow up with Maryland DNR to determine where the data flow failure is occurring.

The MARACOOS OceansMap platform is used to display the Maryland DNR stations. OceansMap allows users to query any station to obtain values, create time-series plots, and compare the observations with overlapping model data. Data can also be accessed directly via OceansMap. Each data stream is linked back to its source. In the case of Maryland DNR data, the stream is linked to the MARACOOS SOS server. Time-series data can also be downloaded directly from OceansMap in CSV format.

Currently, MARACOOS only collects real-time data from Maryland DNR. As the NCEI archiving process proceeds, the MARACOOS DMAC team will consider how to also provide access to the quality controlled datasets.

2) Maryland DNR Quality Control

Sondes are retrieved, calibrated and replaced at intervals ranging from bi-weekly to monthly to minimize the potential for data degradation due to biofouling.

Telemetered data are examined daily by DNR staff to provide a basic level of QC. Automated programs alert DNR staff when equipment malfunctions and probe failures may be occurring at real-time continuous monitoring stations. To ensure data quality, telemetered data is replaced with the QA/QC'd data generated following routine sonde retrieval, calibration and replacement.

At the end of the monitoring season, the DNR staff conducts additional QA/QC procedures to discover problem data which may not have been detected earlier. Staff members plot all of the data and then thoroughly research any outliers or other unusual values. In cases where data are determined, by the Quality Assurance officer, to be unreliable, the reason(s) values are determined to be unacceptable are

documented with error codes and comments, as detailed in the Maryland DNR QAPP. Unreliable data are flagged with error codes. No data are discarded. Only data considered reliable are published in reports or on the web.

3) Maryland DNR Data Sharing

The Maryland DNR data stream adheres to the NOAA Data Sharing Procedural Directive by providing access to near real-time data via a variety of access points at no cost and in a format that is machine-readable, based on open standards, and includes metadata.

The Maryland DNR data is available both from Maryland DNR directly and via the MARACOOS SOS server. Maryland DNR shares their data directly via their [Eyes on the Bay](#) website.

MARACOOS harvests data directly from the Maryland DNR Shallow Water Monitoring database and stores it on the MARACOOS SOS server in NetCDF format. The SOS server stores individual daily files for each station and provides aggregations of the entire data archive by station. The NetCDF files are accessible via OPeNDAP and ncSOS and include ISO 19115-2 metadata. The [MARACOOS OceansMap](#) platform includes the Maryland DNR data and allow users to query any station to obtain both current and historical values, create time-series plots, and compare the observations with overlapping model data. Each data stream is linked back to its source in the case of Maryland DNR data the stream is linked to the MARACOOS SOS server. Time-series data can also be downloaded directly from OceansMap in CSV format.

4) Maryland DNR Archiving

All monitoring data are maintained in a Shallow Water Monitoring database as part of Maryland DNR's Water Quality Monitoring Program. Data that are rejected during QA/QC checks are never deleted. Data determined to be unreliable are flagged with an appropriate error code rather than deleted. This ensures that all raw data are stored and available for future re-evaluation. Techniques and technologies may be developed which will enable utilization of data that have been previously rejected. After the data from a deployment have undergone QA/QC checks, the QA/QC'd data files are transferred to an internal network directory where they are automatically inserted into DNR's Shallow Water Monitoring database. Redundant backup copies of all raw data and QA/QC'd data are also maintained.

The Maryland DNR archival package is still being completed and will contain both the provisional and the verified data. The file names, metadata, and documentation

will clearly indicate whether the data files contain provisional or verified data.

The provisional data are archived daily from the NetCDF files available in the MARACOOS WAF. The provisional nature of the data is documented within the NetCDF files and metadata. The DMAC team is currently reviewing these files and the associated metadata to ensure that they meet the requirements of the NCEI 2.0 templates.

The verified data will be archived annually upon completion of the manual QA/QC procedures conducted by Maryland DNR staff. The monitoring season's data will be re-harvested by the MARACOOS DMAC team and converted to the same NetCDF file format used for the provisional data. This data will be stored as monthly NetCDF files are stored in the MARACOOS WAF along with aggregated provisional files.

CONCLUSION

The federal certification of MARACOOS as a RICE elevates the status of its data and their underlying data management and stewardship procedures, assuring users of MARACOOS data (whether from government, private sector, or the public) that these data can be trusted to be of the highest quality for any applications of the data in models or data products.

ACKNOWLEDGMENTS

There were many contributors from the numerous groups involved with MARACOOS. Special thanks go to J. Kerfoot, C. Haldeman, D. Aragon, O. Schofield, T. Miles, M. Breece, D. Haulsee, C. Thoroughgood, W. Boicourt, D. Gong, W. Brown, E. Handel, C. Evans, T. Garner, T. Fake, and N. Verona.

REFERENCES

- [1] DEPARTMENT OF COMMERCE, National Oceanic and Atmospheric Administration. 15 CFR Part 997. Docket No. 120813326-4163-02 . RIN 0648-BC18. *U.S. Integrated Ocean Observing System; Regulations to Certify and Integrate Regional Information Coordination Entities*. U.S. Integrated Ocean Observing System Program Office (IOOS), National Oceanic and Atmospheric Administration (NOAA), Commerce.
- [2] Kohut, J.T., *Roarty, H.J.*, Randall-Goodwin, E., Glenn, Lichtenwalner, C.S., (2012) "Evaluation of two algorithms for a network of coastal HF radars in the Mid-Atlantic Bight", *Ocean Dynamics*, vol. 62, 953-968.
- [3] NOAA Office of Satellite and Product Operations (NOAA), 2016. Polar-orbiting Operational Environmental Satellites (POES), <http://www.ospo.noaa.gov/Operations/POES/>. Date Accessed, July 14, 2016.
- [4]