



# Data Management Plan

---

## Background

The MARACOOS Data Management and Communication (DMAC) system is a robust, services-oriented system that includes tools to 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), and Environmental Research Division's Data Access Program (ERDDAP); (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).

Distributed data systems have heterogeneous storage and distribution methods and therefore present difficulties for monitoring and enforcing consistent processes and standards. For this reason, MARACOOS has adopted a centralized data aggregation approach, in the form of a regional Data Assembly Center (DAC), for model, *in situ* and remotely sensed observation data. The data storage system includes servers at a secure offsite server farm (THREDDS) and in the Azure cloud (ERDDAP).

Our centralization efforts include management of data generated by MARACOOS partners such as Rutgers and Atlantic Shores Offshore Wind, as well as data collected by non-partner organizations such as the Hudson River Environmental Conditions Observing System (HRECOS), Maryland Department of Natural Resources (DNR) and National Estuarine Research Reserve System (NERRS), and the Chesapeake Bay Interpretive Buoy System (CBIBS). Where applicable, MARACOOS data will continue to be delivered to national data assembly centers (DACs), such as the GliderDAC and the National HF Radar Network, and, as described elsewhere in this document, relevant data are already or in the process of being archived at the National Centers for Environmental Information (NCEI).

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 many federal sources of *in-situ* observation data (NDBC, CO-OPS, NERRS, USGS, CBIBS, HRECOS) are accessible via MARACOOS, these sources are managed by federal partners and are not discussed further in this document. For CBIBS specifically, while the MARACOOS DMAC team played a large role in the design and implementation of the system,

ultimately it runs on federal infrastructure and we use the federal API to access the data. In the five years since the MARACOOS DMAC plan was last submitted, HRECOS transitioned from a regional to a federal source. The HRECOS data is now served by a combination the USGS National Water Information System (NWIS) and NEERS, depending on the station.

The DMAC system facilitates discovery, access, and understanding of regional and relevant national observation and model data. All MARACOOS services are listed in the IOOS catalog, allowing them to be discovered by users. Data can be accessed via Data Access Protocol (DAP) using a THREDDS or ERDDAP and visualized in the [MARACOOS OceansMap](#) platform.

The MARACOOS DMAC system uses a combination of the RPS Environmental Data Server (EDS) and THREDDS to manage gridded data, including HF radar, Satellite, and model hindcasts, nowcasts, and forecasts.

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 buoy (CBIBS), glider, and HF radar data. More recently we updated our QARTOD approach to handle profile data from both ADCPs and floating LiDAR in order to maintain compliance with QAQC requirements for the Atlantic Shores Offshore wind buoys.

The MARACOOS DMAC team keeps abreast of IOOS and U.S. IOOS Program Office data management protocols via regular communication with Program Office personnel, other Regional Association (RA) DMAC teams, and community-based discussions.

## **Roles and Responsibilities**

The MARACOOS DMAC team manages and operates a robust services-oriented system that includes tools to aggregate, quality-control, and monitor a variety of observational and numerical model assets. The DMAC team includes personnel from both Rutgers and RPS. RPS is an industry partner and leads the MARACOOS data management efforts.

### **Mike Crowley (Rutgers) – MARACOOS Technical Director**

Mr. Crowley brings 29 years of ocean observation experience to the MARACOOS team, the last 11 years acting as the Technical Director of the Rutgers University Center for Ocean Observing Leadership and MARACOOS. Additionally, he was the Program Manager for the NSF OOI Cyber-Infrastructure (CI) data management team from 2014-2018. Michael oversees the day-to-day management of observatory operations, including data acquisition, management, as well as modeling and product development.

### **Kelly Knee (RPS ASA) – MARACOOS DMAC Lead**

Ms. Knee has overall management and coordination responsibility for the DMAC team, including technical issues, scheduling, and budgeting. She serves as the primary point of contact between MARACOOS, the DMAC team, and the data providers. She is also responsible for design and review of updates and additions to the DMAC system.

### **Cheryl Morse (RPS) – Operations and Maintenance Lead**

Ms. Morse is responsible for operations and maintenance of all MARACOOS DMAC hardware and software as well as implementation of updates and additions to the system. This includes management of the THREDDS and ERDDAP servers, WMS services, registration of datasets with the IOOS Catalog, and coordination with data providers. Ms. Morse also monitors both data integration and service availability and together with Ms. Knee, coordinates resolution of any issues that may arise.

The *curriculum vitae* for each individual are provided [here](#). The procedures used to evaluate the capability of each individual are described [here](#).

## **Implementation of Data Management Protocols**

MARACOOS DMAC personnel maintain regular communication with the U.S. IOOS Program Office through a variety of mechanisms including in-person meetings, phone and video calls, mailing lists, webinars, email conversations, and GitHub repositories. MARACOOS also regularly contributes to and participates in the IOOS Annual DMAC meeting. The continuous communication ensures that the DMAC team is aware of all new practices and protocols, as promulgated by the IOOC and the IOOS Program Office and understands how and when to implement them.

The MARACOOS servers and data management software are maintained by the MARACOOS DMAC team so data management protocols can be applied as soon as practicable, limited only by resource restrictions (personnel time, budget, or server capabilities). The MARACOOS DMAC team plans to implement all new protocols as soon as possible and within one year of adoption by IOOS. Once the new protocol is received by the DMAC lead, an implementation plan, including a more detailed schedule, will be outlined and executed by the DMAC team.

## **MARACOOS Data Streams**

As each data stream provided via MARACOOS is unique, this section describes each MARACOOS data source individually. In total, ten data streams are described. These include (1) High Frequency Radar – Surface Currents, (2) High Frequency Radar – Waves (3) Gliders, (4) Advanced Very High-Resolution Radiometer (AVHRR), (5) Cold Water Preserving Sea Surface Temperature Product, (6) Moderate Resolution Imaging Spectroradiometer (MODIS), (7) GOES, (8) Maryland Department of Natural Resources (DNR) water quality monitoring stations, (9) Atlantic Shores Offshore Wind buoys, and (10) Weatherflow weather stations. The descriptions for each data stream include the data flow (certification requirements Section f2), quality control procedures (certification requirements Section f3), adherence to the NOAA Data Sharing Procedural Directive (certification requirements Section f4), and the existing or planned archiving procedure (certification requirements Section f6).

### **High Frequency Radar – Surface Currents**

The Mid-Atlantic High Frequency (HF) Radar Network consists of 40 radars operated by eight separate organizations (Old Dominion University, Rutgers - The State University of New Jersey, University of Connecticut, University of Massachusetts (Dartmouth), Woods Hole Oceanographic Institute, University of North Carolina, University of Delaware and University of Rhode Island). The network provides hourly

measurements of ocean surface currents (representative of water depths of 0.3 – 2.5 m) within ~200 km of the coast, an area encompassing more than 190,000 km<sup>2</sup> of the ocean's surface, at resolutions of 1 km, 2km, and 6km.

### 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 diagram documenting the flow of data from the radar stations is shown in Figure 1. The diagram 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 2 (the radial generation step) 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 are then passed on to the NOAA National HF Radar Network. The process of adding quality control checks is underway and expected to be completed in Fall 2021. Once this process is completed the quality-controlled radials will be provided to the NOAA National HF Radar Network. Level 3 and 4 occur at the regional and national data assembly centers where the radial vector data are combined to make total vector products.

From the provided radials, NOAA 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 the [CORDC THREDDS](#), [NDBC THREDDS](#), or [FTP \(radials\)](#). The RPS Environmental Data Server (EDS) also collects the National HF Radar Network product to support United States Coast Guard (USCG) search and rescue operations. The data is harvested from the national [THREDDS](#) server. 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, 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](#).

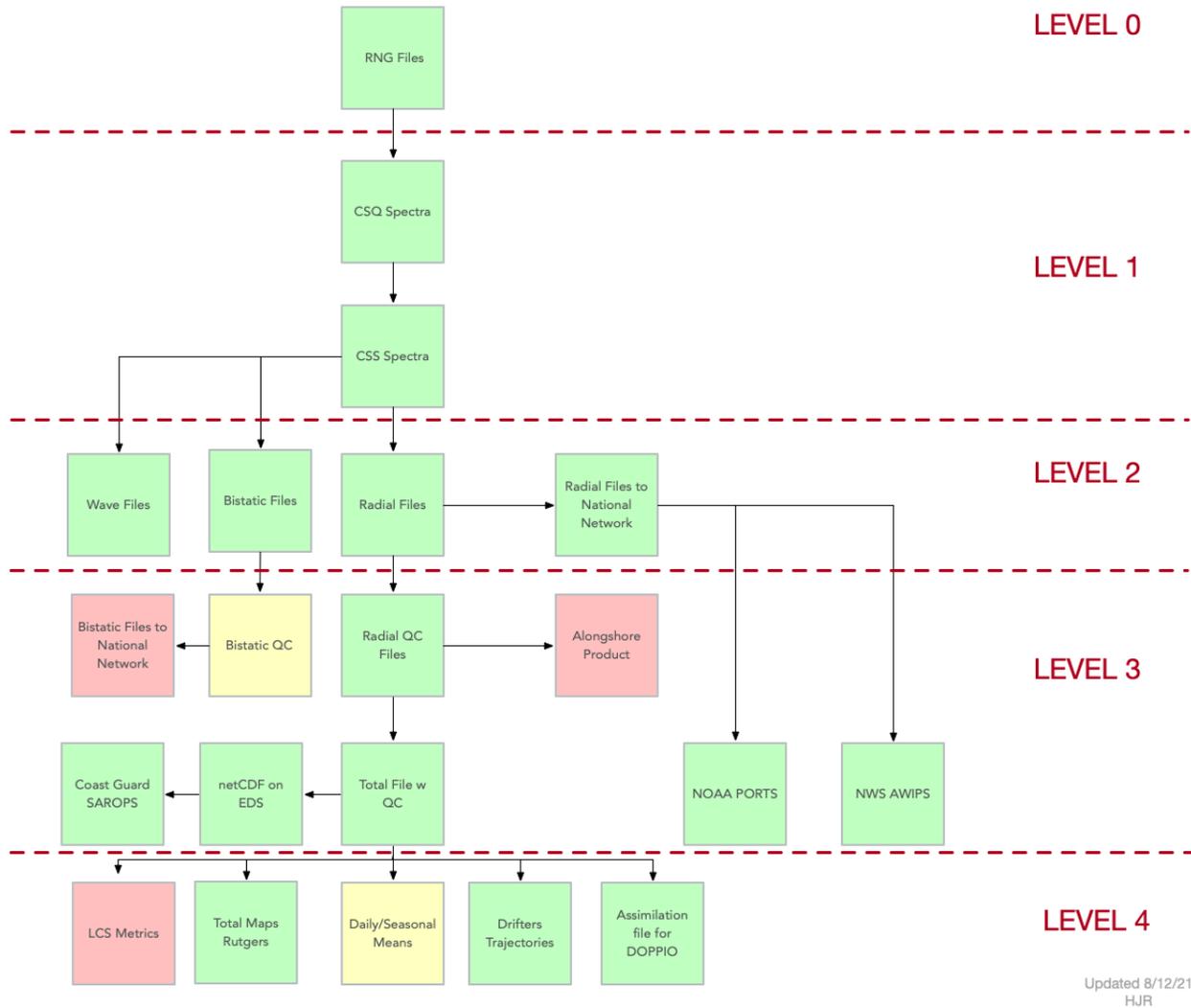
To satisfy additional user requests, MARACOOS also produces an alternative 6 km product that uses quality-controlled radials and an optimal interpolation process ([Kohut et al., 2012](#)) to combine the radial vectors into a total vector product. The total vector data are then transformed into NetCDF, further quality controlled (no modification of the data – flags are added) and made publicly available on a Rutgers [ERDDAP](#) server. One benefit of this alternative 6 km product is that it will accept data that is older than 24 hours. The National HF Radar Network will accept radial data older than 24 hours but it will not regenerate totals to include this missing data.

The total vector data are harvested from the Rutgers [ERDDAP](#) server on an hourly basis by the RPS EDS so that they can be made available to the USCG for use in the Search and Rescue Optimal Planning System (SAROPS). The EDS harvests the NetCDF files directly from ERDDAP and stores the data in the EDS Data Store as hourly NetCDF files. As with harvesting of the national total vector product, no transformations or modifications of the data are made during this process, and all variables, including

flags, are retained. The EDS makes the data available via OGC compliant WMS services and also provides OPeNDAP access via [THREDDS](#).

Detailed documentation on the processing scripts and code can be found [here](#).

## Rutgers HF Radar Processing Flow Chart



**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 green boxes indicate products that are in production, yellow indicates products in development and red indicates planned products.**

The total vector data are also used to create four seasonal climatology datasets. These datasets represent decadal climatology of Mid-Atlantic optimal interpolation totals product from December 2007 through November 2016 for fall, winter, spring, and summer. Seasonal means of totals were created for each year (September to November) with a requirement of 50% coverage at error <0.60 per gridpoint, and annual season means were averaged over the 10 years with a requirement of coverage in at least

5/10 years per gridpoint. These datasets are served directly from the [Rutgers THREDDS](#) server and made available via OPeNDAP and the MARACOOS WMS.

## 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 in the process of being applied during Level 2, after radials are generated and before they are passed to the National Network. Currently the National Network is unable to accept the radial file with QARTOD flags, so we have written a script that will remove the vectors that fail the QARTOD tests and this is the data file that will be sent to the National Network. This will be implemented in the fall of 2021.

Quality assurance and control procedures followed include:

- [A Plan to Meet the Nation's Needs for Surface Current Mapping](#), US IOOS, May 2015
- [Deployment & Maintenance of a High Frequency Radar for Ocean Surface Current Mapping: Best Practices](#), SCOOS & CeNCOOS, February 2008

In addition to the quality control described above, 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 and total vector data. All seven of the required tests have been implemented; these include syntax, signal-to-noise ratio, maximum radial velocity threshold, maximum total velocity threshold, valid location, data density, and geometry threshold tests. The radial count test and radial spatial median test have also been implemented. The MARACOOS product includes radials from SECOORA stations located in North Carolina's Outer Banks. Those stations perform additional radial metric level quality control on site. These tests are listed in the HF radar QARTOD manual as Test 2 (Cross Spectra Covariance Matrix Eigenvalues), Test 3 (Single and Dual Angle Solution - Direction of Arrival (DOA) Metrics (magnitude)) and Test 4 (Single and Dual Angle Solution - Direction of Arrival (DOA) Function Widths (3 dB)). The quality-controlled radials are used as input to the total vector product produced by MARACOOS.

## 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 data via [THREDDS](#), and [FTP \(radials\)](#). The [THREDDS](#) instance provides OPeNDAP, WCS, NetCDF, and WMS access as well as ISO metadata. 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 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 a [THREDDS server](#) via OPeNDAP.

The MARACOOS total vector data is available on an [ERDDAP](#) server at Rutgers, which provides access to the data directly and via GridDAP. ERDDAP also allows users to easily generate plots and review metadata directly through the ERDDAP web interface. The 6km data is also available via OPeNDAP from the [EDS THREDDS server](#). HF radar data for the region can also be visualized using the [MARACOOS OceansMap](#) platform. There are four HF radar products that are harvested hourly and available in OceansMap:

1. IOOS HF Radar East and Gulf Coast 6 km - National HF Radar Network 6km total vector product, harvested by EDS and visualized using the EDS WMS service
2. IOOS HF Radar East and Gulf Coast 2 km - National HF Radar Network 2km total vector product, harvested by EDS and visualized using the EDS WMS service
3. MARACOOS HF Radar Mid-Atlantic 6km – MARACOOS total vector product, harvested from Rutgers ERDDAP by EDS and visualized using the EDS WMS service
4. IOOS National HF Radar 1km - National HF Radar Network total vector product, visualized using the National HF Radar Network tiles (GetFeatureInfo unavailable for this layer)

The HR radar seasonal climatology datasets are available via OPeNDAP from the [Rutgers THREDDS](#) and via the [MARACOOS WMS](#) server. These four products include:

1. HFR Fall Seasonal Climatology – derived from 10 years of the MARACOOS 6km total vector product
2. HFR Spring Seasonal Climatology – derived from 10 years of the MARACOOS 6km total vector product
3. HFR Summer Seasonal Climatology – derived from 10 years of the MARACOOS 6km total vector product
4. HFR Winter Seasonal Climatology – derived from 10 years of the MARACOOS 6km total vector product

### **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 will continue to reach out to NCEI about archiving 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.

MARACOOS intends to have the archiving process in place for the MARACOOS total vector product by December 2022. The DMAC team has initiated the archiving process with NCEI, has an Advanced Tracking and Resource Tool for Archive Collections (ATRAC) account with an active MARACOOS project, and has requested review of sample files and data submission workflow.

## High Frequency Radar - Waves

The primary measurement of an HF radar are ocean surface currents. Under certain conditions an ancillary measurement is wave height, period, and direction, and surface wind direction. Of the 42 radars in the Mid-Atlantic High Frequency Radar Network, six also measure waves along the New Jersey coast. It is envisioned that these measurements can supplement traditional buoy measurements and provide additional value for validating wave models.

### Data Flow

Hourly wave data is generated at five of the MARACOOS HF radar stations. The stations are currently running Release 8 Update 5 CODAR SeaSonde software. Both the WVLR and WVLM SeaSonde wave file types are generated. The WVLR provides wave measurements at several range bins while the WVLM file is an average wave measurement over all the range cells. Table 1 provides a list of the SeaSonde software tools and version numbers utilized in the wave processing.

Table 1: SeaSonde wave processing tools and version as of August 11, 2021.

SeaSonde Wave Tool	Version
WaveModelFilter	3.0.2
SpectraToWavesModel	11.8.0
WaveModelForFive	11.5.0
WaveModelArchiver	12.1.6
AnalyzeSpectra	10.9.8

The wave data (WVLR and WVLM) is telemetered back to Rutgers via the **rsync\_waves.sh** script that runs at 20 minutes past the hour.

The ASCII CODAR wave files are converted to netCDF via the **realtime\_waves\_to\_netcdf.sh** that runs at 50 minutes past the hour. That shell script calls a Python script **realtime\_waves\_to\_netcdf.py** to convert the CODAR file to netCDF.

The netCDF data files are stored as monthly files and made available by the Rutgers [ERDDAP](#) server. MARACOOS harvests the HF radar wave data from the Rutgers ERDDAP hourly by making ERDDAP requests that return JSON data. The JSON feed is used to create compliant daily netCDF files which are then hosted on the [MARACOOS ERDDAP](#) server.

### Quality Control

MARACOOS follows and applies a variety of quality assurance and quality control procedures to the wave data before passing it to the ERDDAP server. The wave measurements require a signal to noise ratio

(Test 4 of the QARTOD Wave manual) of 5 dB in order to be processed. We plan on adding additional QARTOD tests to the wave data this coming year.

### Data Sharing

The HF radar waves data 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 International Organization for Standardization (ISO) 19115 metadata.

MARACOOS makes the data available on the [MARACOOS ERDDAP](#) where it is available via TableDAP. ERDDAP also allows users to easily generate plots and review metadata directly through the ERDDAP web interface. The HFR Wave Data can be visualized using the [MARACOOS OceansMap](#) platform, where data from each of the five stations is available in the observation station. OceansMap allows users to view the station locations on a map, plot time-series data, and compare the observations to wave models.

### Archiving

MARACOOS is already in the process of archiving the MARACOOS total vector product and plans to include the HF radar waves in the same submission. MARACOOS intends to have the archiving process in place for both products by December 2022.

### Gliders

The Mid-Atlantic Glider Network is coordinated through a consortium of glider operators working within MARACOOS. The MARACOOS glider operators include Rutgers University, University of Delaware, Virginia Institute of Marine Sciences, University of Connecticut, University of Rhode Island, the University of Massachusetts Dartmouth. The primary measurements provided by the gliders are temperature and conductivity. Derived parameters include salinity and density. Increasingly, additional variables such as oxygen, CDOM, and chlorophyll are also being collected.

The glider operations are coordinated by the team (institutions listed above) 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, state, and corporate 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.

## 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 and a Wiki describing the Glider DAC in detail is available [here](#). 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:

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, which includes vertical cross section data

<http://oceansmap.maracoos.org>

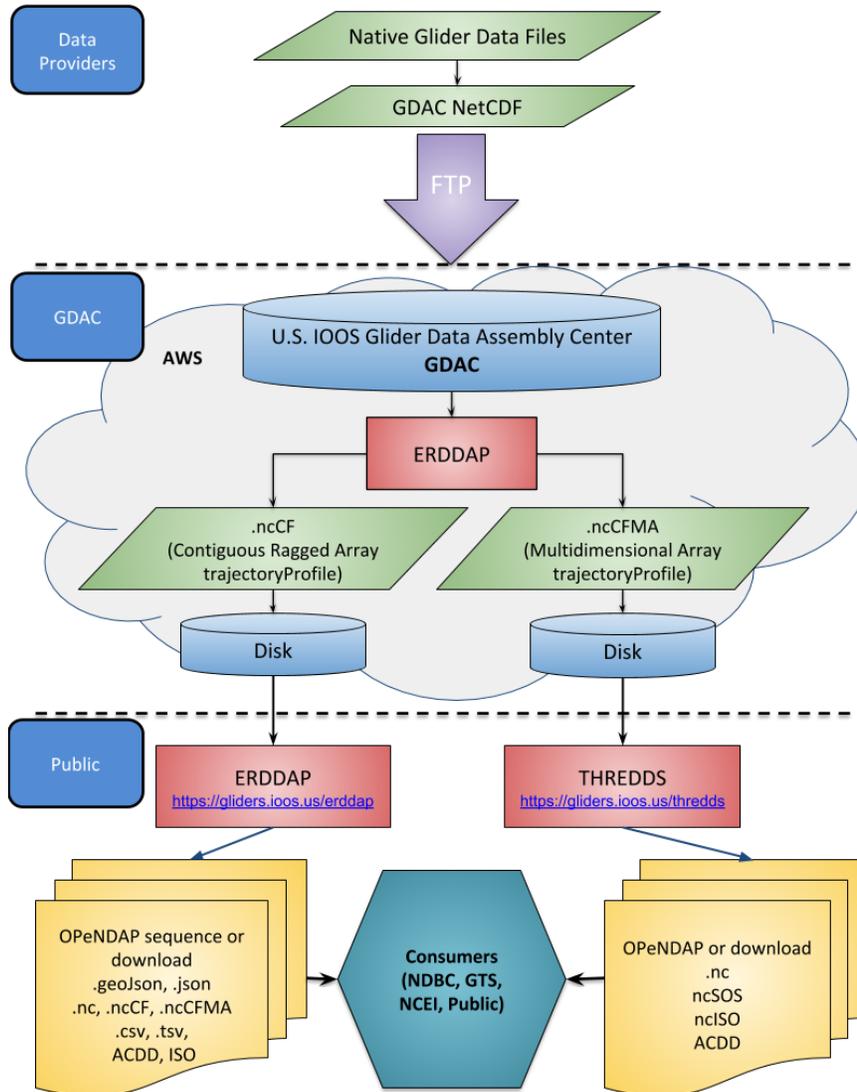


Figure 2: Diagram of data flow from gliders to the IOOS glider DAC.

### Quality Control

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

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

In addition, MARACOOS is in the process of reimplementing IOOS QARTOD Standards, described in IOOS QARTOD [Manual for Quality Control of Temperature and Salinity Data Observations from Gliders](#). MARACOOS is in the process of transitioning code from Matlab to python and once complete QARTOD will be reimplemented on glider data.

The following tests will be applied:

- Timing Test (Required)
- Syntax Test (Required)
- Location Test (Required)
- Gross range Check (Required)
- Climatology Test (Required)
- Pressure test (Required)
- Spike Test (Strongly Recommended)
- Rate of Change Test (Strongly Recommended)
- Flat Line Test (Strongly Recommended)

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.

### **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, however the GliderDAC is currently evaluating use of THREDDS and whether ERDDAP alone is sufficient to meet the needs of the community. Metadata is accessible via both endpoints. The DAC also provides visualizations of Glider data using the [IOOS Glider Map](#).

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

## Archiving

Quality-controlled glider data are uploaded to the NOAA Glider DAC. Each dataset must adhere to the [NCEI authority tables](#) agreed between the GliderDAC and NCEI teams for institutions, instruments, platforms, and projects. The glider data are archived by NCEI directly from the DAC. The timing of archive submission is determined by the DAC.

## Advanced Very High-Resolution Radiometer (AVHRR)

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 (NOAAb, 2016). 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 or less) than most federal products (4km resolution). The data stream is accessed by ocean models and used for regional research and product development.

## Data Flow

The MARACOOS AVHRR data stream is coordinated through the University of Delaware, with a backup sub-regional technology center at Rutgers University. The hardware for these system currently includes two dual-use L/X-Band dishes at the University of Delaware and 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 misses 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 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 the single pass AVHRR data 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.

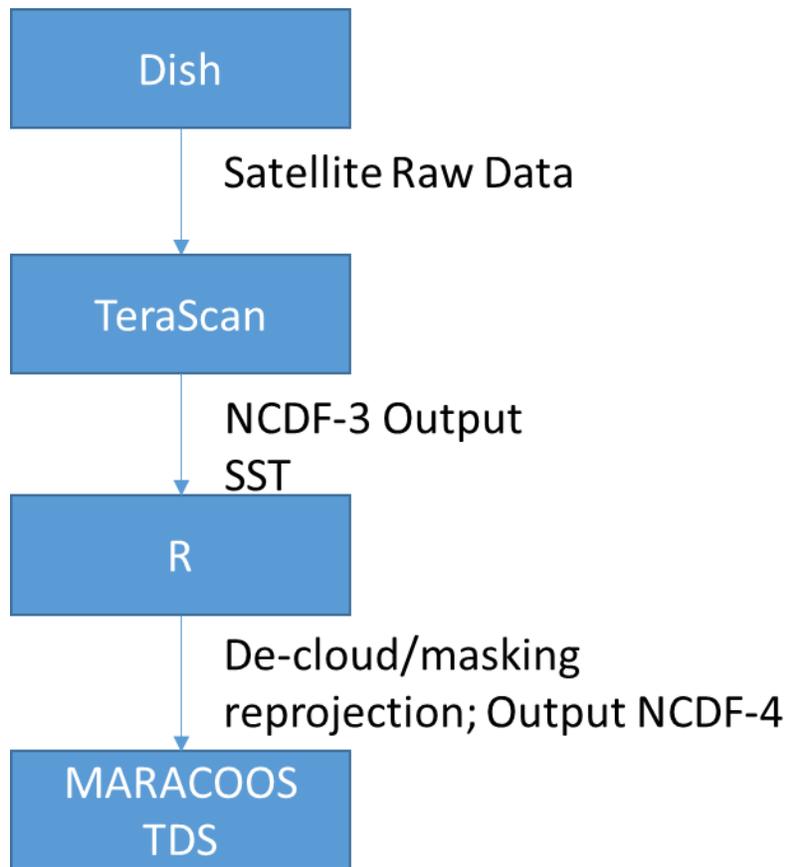


Figure 3: Data flow for MARACOOS AVHRR SST Products

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, encompassing files from 2013 to present
    - ii. Single Pass Individual Files: individual satellite passes, organized in yearly folders from 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, organized in yearly folders 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, organized in yearly folders 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, organized in yearly folders from 2013 to present
  - b. Unmasked – Regional experimental cloud mask is provided, but not applied
    - i. Single Pass Aggregation: aggregations of single passes from 2013 to present
    - ii. Single Pass Individual Files: individual satellite passes from 2013 to present

## 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 the climatology test if the temperature is more than 5 degrees colder than the climatologies calculated by 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 test, climatology test, 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 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. Figure 4 shows the RMS error for each station. Figure 5 shows the bias for each station. The RMS, bias, and observation number values for each buoy are summarized in [Appendix 1](#), Tables 1 through 3 for NDBC operated buoys and Tables 6 through 8 for non-NDBC operated buoys. 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.

After elimination of pixels with flags greater than 3, RMS and bias statistics were calculated for the 24-hour and weekly data. The results of these calculations are provided in [Appendix 1](#), Tables 4 and 5 for NDBC buoys and Tables 9 and 10 for non-NDBC buoys.

### RMS Error for eliminating data with flag>3

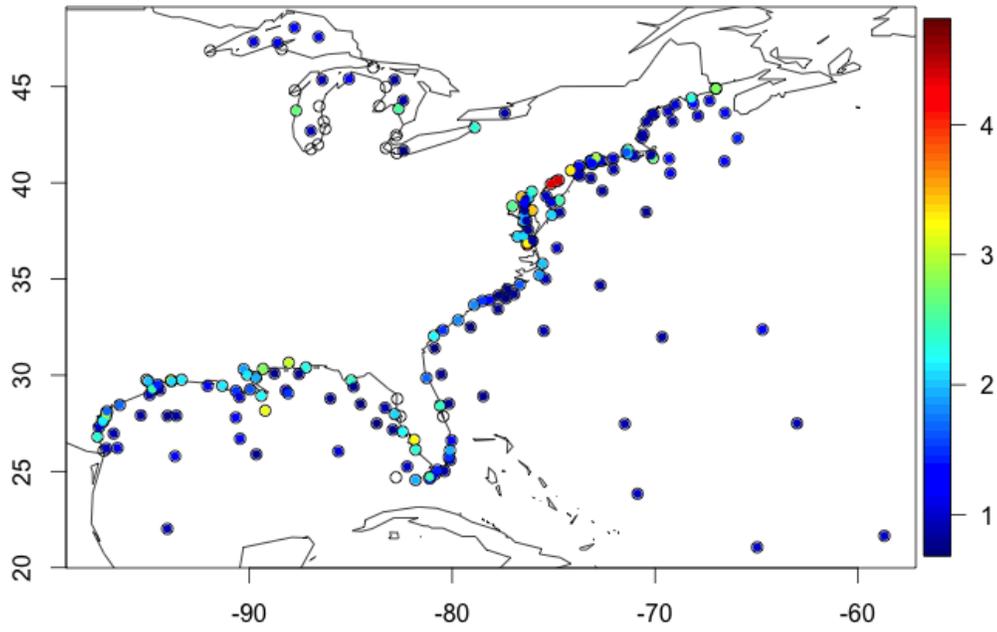


Figure 4: RMS Error ( $^{\circ}\text{C}$ ) for the 205 NDBC stations tested in the AVHRR analysis

### Bias (Buoy-Satellite Deg C) for eliminating data with flag>3

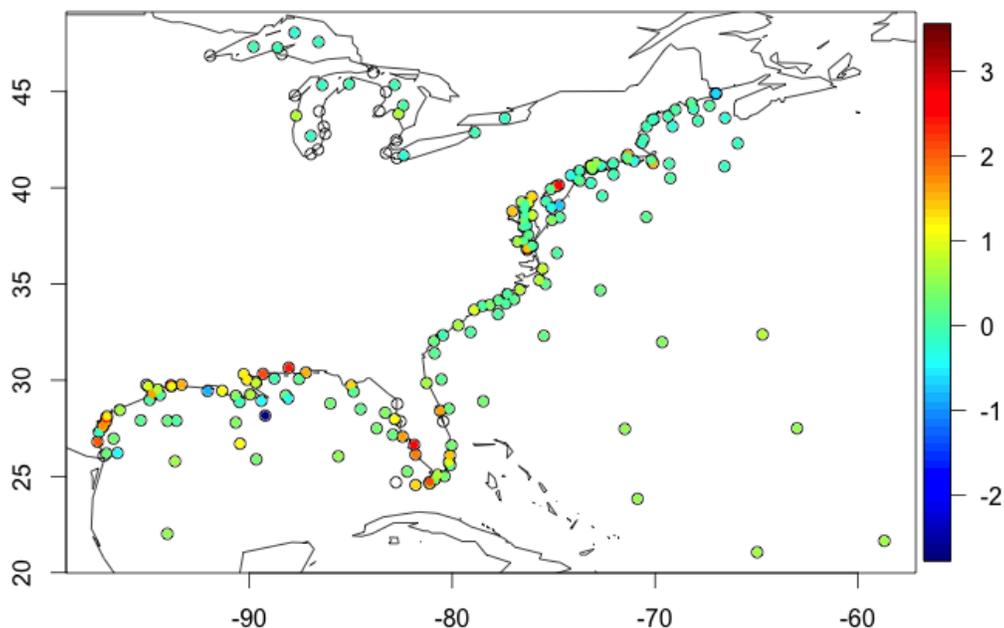


Figure 5: Bias ( $^{\circ}\text{C}$ ) for the 205 NDBC stations tested in the AVHRR analysis. Bias was computed as Buoy-Satellite. Values greater than zero indicate that the buoy measurements were warmer than satellite estimates.

### Data Sharing

The AVHRR 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

A number of AVHRR SST products are stored on the [MARACOOS TDS](#) and available via OPeNDAP. 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. ISO metadata is provided for each aggregation.

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, encompassing files from 2013 to present
    - ii. Single Pass Individual Files: individual satellite passes, organized in yearly folders from 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, organized in yearly folders 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, organized in yearly folders 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, organized in yearly folders from 2013 to present
  - b. Unmasked – Regional experimental cloud mask is provided, but not applied
    - i. Single Pass Aggregation: aggregations of single passes from 2011 to present
    - ii. Single Pass Individual Files: individual satellite passes from 2011 to present

The AVHRR SST data products are also available as WMS products via the MARACOOS [WMS](#). Available products include (note: hyperlink above provide access to the GetCapabilities document; list below provides titles for individual products):

1. [AVHRR unmasked](#): Aggregation without experimental regional mask applied
2. [AVHRR1](#): Aggregation of 1-day composites with experimental regional mask applied
3. [AVHRR3](#): Aggregation of 3-day composites with experimental regional mask applied
4. [AVHRR3](#): Aggregation of 7-day composites with experimental regional mask applied

The AVHRR satellite data products can be visualized using the [MARACOOS OceansMap](#) platform. Products available from OceansMap include the 1-, 3-, and 7-day composites (with mask applied).

### **Archiving**

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

1. A rolling store of at least 3 years of Single Pass AVHRR SST files with the experimental regional mask applied
2. Individual AVHRR SST files without the experimental regional mask applied for at least the current year
3. A rolling store of at least 3 years 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 annually and adjusted based on available hardware resources. As the national data from which the MARACOOS products are derived are archived in perpetuity by the NOAA Comprehensive Large Array Data Stewardship System (CLASS), any data removed from the MARACOOS store can be recreated if necessary.

## Cold-Water Preserving Sea Surface Temperature Products

Most satellite Sea Surface Temperature (SST) cloud-removal algorithms are ineffective at preserving cold water, frequently flagging it as clouds especially when the cold water appears after a very rapid drop in temperature. Development of products that can remove clouds while preserving cold water is necessary for early identification of features frequently observed in the Mid Atlantic Bight such as upwelling and storm cooling, and thus can also help improve regional wind model forecasts.

### Data Flow

The existing cold-water preserving SST product is a daily coldest pixel composite generated based on AVHRR data received on the Rutgers satellite dish. Each image is declouded based on reflectivity in the visible spectrum and a daily composite is created from the coldest data at each gridpoint from all images available during daylight hours. These daily composite netCDF files are hosted on the Rutgers [ERDDAP](#) server. The files are then harvested from the Rutgers ERDDAP into the [MARACOOS THREDDS](#), where they are made available as an aggregated product covering the period 2018 through present.

Due to persistent issues with AVHRR data, as well as the higher frequency of regional SST images being returned by GOES-16, MARACOOS is transitioning away from the existing product to one that uses hourly GOES SST images accessed from NOAA via Google Bucket and shared on the University of Delaware THREDDS and ERDDAP servers. This data is declouded following the spike filter method described in Murphy et al. (2021), composited into a daily product, and then gap-filled using the reconstruction method of Data Interpolating Empirical Orthogonal Functions (DINEOF). This method is still under development and subject to change, and is not yet being generated in real-time. When complete the final product will be harvested from University of Delaware's THREDDS server and made available on MARACOOS servers. The expected timeframe for transition is Spring 2022.

## Quality Control

Because the cold-water preserving sea surface temperature products are derived from quality controlled MARACOOS sources (currently AVHRR, eventually GOES) no additional quality control is applied other than a visual check by satellite data subject matter experts.

## Data Sharing

The cold-water preserving sea surface temperature product 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. MARACOOS provides access to an aggregation of the product via [THREDDS](#) so the data is available via OPeNDAP. MARACOOS also provides a [WMS](#) service the layer title is RU-AVHRR-Coldest. The products can also be visualized using the [MARACOOS OceansMap](#) platform.

## Archiving

MARACOOS stores cold-water preserving sea surface temperature on MARACOOS infrastructure and makes them available on the [MARACOOS THREDDS](#) as long-term aggregations. This aggregation includes individual files from 2018 to present.

The amount of cold-water preserving sea surface temperature stored by MARACOOS is re-evaluated annually and adjusted based on available hardware resources. As the national data from which the MARACOOS products are derived are archived in perpetuity by NASA any data removed from the MARACOOS store can be recreated if necessary.

## Moderate Resolution Imaging Spectroradiometer (MODIS)

The MARACOOS Moderate Resolution Imaging Spectroradiometer (MODIS) data stream is coordinated through the University of Delaware, where it is collected and distributed in near-real time. MODIS is a key instrument aboard the Aqua satellite. MODIS-Aqua views the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands, or groups of wavelengths (NASAa, 2016). MARACOOS provides a subset of MODIS ocean color data covering the Gulf of Mexico and the Eastern Seaboard at higher resolution (1 km or less) than most federal products (4km resolution). The data stream is accessed by ocean models and used for regional research and product development.

## Data Flow

The MARACOOS MODIS 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. 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 dish misses a MODIS pass, the ocean color data are filled in from the [NASA Ocean Color Data Distribution](#) site.

The data flow from raw MODIS data to processed NetCDF-4 files on the [MARACOOS TDS](#) is illustrated in Figure 6. All data processing is done using NASA software. The raw MODIS satellite data collected by MARACOOS is copied from the X-Band dish to the acquisition machine, where it is transferred to the

processing server via Secure Copy Protocol (SCP). The raw data is then processed into Level 0 data using the Direct Readout Laboratory's Real-Time Satellite Telemetry Processing System (RT-STPS). The Ground-Based Attitude Determination (GBAD) Ephemeris and Attitude Data Converter is used to create ephemeris and attitude files from the Level 0 data for geolocation. The NASA Ocean Color Biology Group (OCBG) SeaDas software is then used for processing the Level 0 data into Level 1a, Level 1b, and Level 2 data, producing 1-, 3-, and 8-day composites at 1 km resolution. Finally, using a series of R scripts the data are projected to latitude and longitude coordinates and converted to NetCDF-4. The NetCDF-4 files are pushed to the [MARACOOS TDS](#).

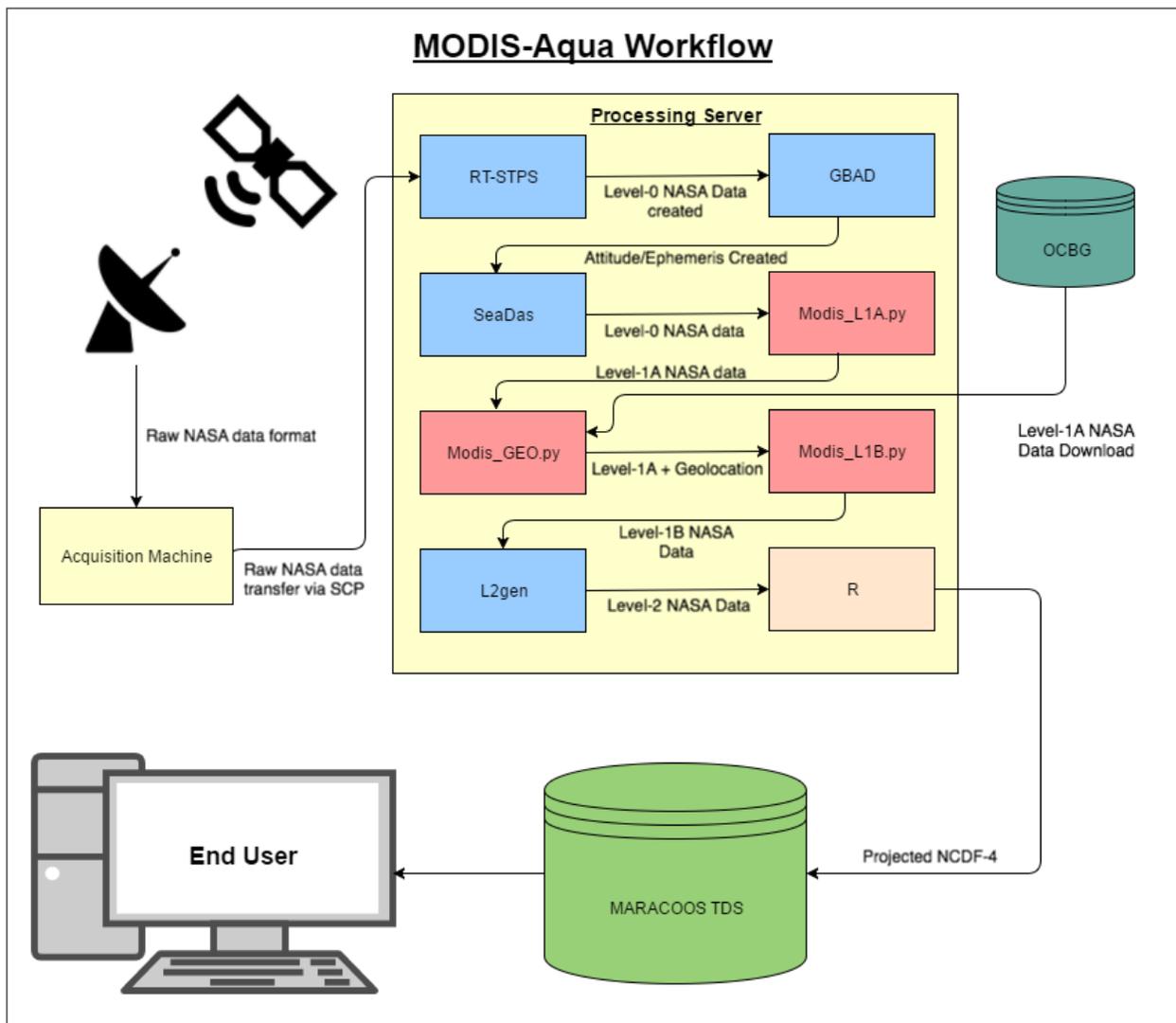


Figure 6: Data flow for MARACOOS MODIS Products

A number of MODIS products are provided via the MARACOOS TDS. These include 1-, 3-, and 8-day composites and aggregations of those composites. Composites are created by averaging each time-step in a dataset over a period of time. For example, a 3-day composite is the gridded average of available

values over the last three days. Aggregations are single, long-term datasets, rather than individual composite files. Individual satellite passes are not provided for the MODIS source.

The MARACOOS TDS stores the MODIS data in two separate directories [REALTIME](#) and [ARCHIVE](#). The REALTIME directory provides symlinks to 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 composite files and composite aggregations in yearly directories.

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

1. [REALTIME](#)
  - a. 1-, 3-, and 8-day composites
2. [ARCHIVE](#)
  - a. 1-day-Aggregation: aggregation of 1-day composites, organized in annual folders from 2002 to present
  - b. 1-Day Individual Files: individual 1-day composites, organized in annual folders from 2002 to present
  - c. 3-day-Aggregation: aggregation of 3-day composites, organized in annual folders from 2002 to present
  - d. 3-Day Individual Files: individual 3-day composites, organized in annual folders from 2002 to present
  - e. 8-day-Aggregation: aggregation of 8-day composites, organized in annual folders from 2002 to present
  - f. 8-Day Individual Files: individual 8-day composites, organized in annual folders from 2002 to present

### Quality Control

MODIS data sets are quality controlled using standard algorithms provided by the NASA OCBG through their SeaDAS software. The SeaDAS quality control flags are used to process data, which are distributed to users in near-real time as described above. The SeaDAS software is updated as necessary and calibration coefficients are updated as they become available (daily). If the processing systems fail, the DMAC team is notified via email that a MODIS pass was missed and that the MODIS datasets have not posted to the MARACOOS THREDDS server. The missed data is acquired from the [NASA Ocean Color Data Distribution](#) site. The historical data is then reprocessed and quality controlled in the same manner as data acquired directly by MARACOOS.

### Data Sharing

The AVHRR 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

A number of MODIS products are provided via the [MARACOOS TDS](#) and the [MARACOOS ERDDAP](#). These include 1-, 3-, and 8-day composites and aggregations of those composites. Composites are created by averaging each time-step in a dataset over a period of time. For example, a 3-day composite is the gridded average of available values over the last three days. Aggregations are single, long-term datasets, rather than individual composite files. Individual satellite passes are not provided for the MODIS source on the MARACOOS TDS because the composites are created by the University of Delaware process before the data is transferred to the TDS.

The MARACOOS TDS stores the MODIS data in two separate directories [REALTIME](#) and [ARCHIVE](#). The REALTIME directory provides symlinks to 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 composite files and composite aggregations in yearly directories.

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

1. [REALTIME](#)
  - a. Individual 1-, 3-, and 8-day composites
2. [ARCHIVE](#)
  - a. 1-day-Aggregation: aggregation of 1-day composites, organized in annual folders from 2002 to present
  - b. 1-Day Individual Files: individual 1-day composites, organized in annual folders from 2002 to present
  - c. 3-day-Aggregation: aggregation of 3-day composites, organized in annual folders from 2002 to present
  - d. 3-Day Individual Files: individual 3-day composites, organized in annual folders from 2002 to present
  - e. 8-day-Aggregation: aggregation of 8-day composites, organized in annual folders from 2002 to present
  - f. 8-Day Individual Files: individual 8-day composites, organized in annual folders from 2002 to present

### **Archiving**

MARACOOS stores some historical MODIS data on MARACOOS infrastructure and makes them available on the [MARACOOS TDS](#). These data include individual MODIS files for the 1-, 3-, and 8-day composites from 2002 to present.

The amount of MODIS data stored by MARACOOS is re-evaluated annually and adjusted based on available hardware resources. As the national data from which the MARACOOS products are derived are archived in perpetuity by NASA using their OceanColor Web, any data removed from the MARACOOS store can be recreated if necessary.

## GOES

NOAA's Geostationary Operational Environmental Satellites (GOES) provide imagery and measurements of the western hemisphere. They hover continuously over one position on the surface, approximately 35,800 km (22,300 miles) above the Earth. They provide a constant vigil for the atmospheric "triggers" for severe weather conditions such as tornadoes, flash floods, hail storms, and hurricanes. They are regularly animated on television weather forecasts and websites.

### Data Flow

GOES-R Sea Surface Temperature (SST) satellite data is downloaded from a [NASA JPL THREDDS](#) server every hour. The SST data are subset to the mid-Atlantic region of interest (16-52N, 50-100W), projected into latitude and longitude coordinates, supplemented with additional metadata, and converted to a daily composite for distribution via the [University of Delaware THREDDS](#). The GOES daily composite is then harvested by MARACOOS and made available by the [MARACOOS THREDDS](#) and [MARACOOS ERDDAP](#), where it is made available as an aggregated daily composite. Daily anomaly composites are also created based on the daily composite and a climatology of the mid-Atlantic region. These derived products are stored on the [University of Delaware THREDDS](#) and made available via OPeNDAP.

### Quality Control

The GOES-R SST product is provided by the Group for High Resolution Sea Surface Temperature (GHR SST) and the National Oceanic and Atmospheric Administration. The data is distributed by NASA JPL. NOAA monitors the performance of the product and validates against drifters and tropical moorings.

### Data Sharing

The GOES-R SST 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

MARACOOS provides access to an aggregation of the daily composite via both [THREDDS](#) and [ERDDAP](#). Composites are created by averaging each time-step in a dataset over a period of time. The daily composite is comprised of the gridded average of available values from the previous calendar day. Aggregations are single, long-term datasets, rather than individual composite files. The derived GOES anomaly product is available from the [University of Delaware THREDDS](#) server via OPeNDAP; RPS provides [WMS](#) access to the dataset (the layer title is GOES\_SST)

### Archiving

MARACOOS stores historical GOES-R SST daily -composites on MARACOOS infrastructure and makes them available on the [MARACOOS THREDDS](#) and [MARACOOS ERDDAP](#) as long-term aggregations. These data include individual GOES-R SST files for the 1-day composites from 2018 to present.

The amount of GOES-R data stored by MARACOOS is re-evaluated annually and adjusted based on available hardware resources. As the national data from which the MARACOOS products are derived are archived in perpetuity by NASA JPL using their TDS, any data removed from the MARACOOS store can be recreated if necessary.

## 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 (calculated from specific conductance and temperature), dissolved oxygen, turbidity (NTU/FNU), fluorescence and total chlorophyll (used to estimate chlorophyll a), 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.

### 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/QCed 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/QCed data generated after continuous monitoring sonde replacement.

MARACOOS harvests data from the Maryland DNR Shallow Water Monitoring database onto the [MARACOOS SOS THREDDS](#) server in NetCDF format. The data is not transformed or modified during this process.

The MARACOOS SOS THREDDS server stores individual daily files for each station and provides time-aggregations of the entire data store by station. The NetCDF files are then made available via [OPeNDAP](#) and [ERDDAP](#) and include ISO 19115-2 metadata.

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 will be linked to the MARACOOS ERDDAP server. Time-series data can also be downloaded directly from OceansMap in CSV format.

## 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. Calibration and post-calibration procedures for the individual sondes are fully described in the [Maryland DNR QAPP](#).

Telemetered data are examined daily by DNR staff to provide a basic level of QUALITY CONTROL. 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. The error codes used are described in Appendix 11 of the [QAPP](#). No data are discarded. Only data considered reliable are published in reports or on the web.

## 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 THREDDS and ERDDAP servers. Maryland DNR also 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 THREDDS server in NetCDF format. The 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 ERDDAP and include ISO 19115-2 metadata. The [MARACOOS OceansMap](#) platform includes the Maryland DNR data and allows 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 ERDDAP. Time-series data can also be downloaded directly from OceansMap in CSV format.

## 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 (Appendix 11 of Maryland DNR QAPP) 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.

MARACOOS is still in the process of archiving the Maryland DNR data. The provisional data collected in near real-time by MARACOOS will be archived daily from the NetCDF files available in the MARACOOS WAF. The provisional nature of the data will be documented within the NetCDF files and metadata. The DMAC team will review these files and the associated metadata to ensure that they meet the requirements of the NCEI 2.0 templates.

MARACOOS intends to have the archiving process in place for Maryland DNR data by March 31st 2022. The DMAC team has already initiated the archiving process with NCEI and has a SIF in place that will cover this dataset.

## Atlantic Shores Offshore Wind

Atlantic Shores Offshore Wind has a lease area approximately 15 miles off the coast of New Jersey and currently holds a contract with the New Jersey Board of Public Utilities to develop 1,510 MW of offshore wind energy. Atlantic Shores is a 50:50 partnership between Shell New Energies and EDF Renewables. The project has been ongoing since 2018 and in 2019 survey efforts were undertaken to support siting analysis. The survey data collected helped Atlantic shores to select six locations in the lease area for metocean data collection – ASOW-1 through ASOW-6. The buoys assist in understanding wind, ocean, and weather conditions to support development and construction in the ASOW lease area, but also provide wide-ranging value to a broader community of stakeholders including mariners and the modeling community.

So far, buoys have been deployed at three of these locations, ASOW-1, ASOW-4, and ASOW-6. The ASOW-6 buoy was deployed from 29 December 2020 through 14 May 2021. The ASOW-1 and ASOW-4 buoys came online on 14 May 2021 and are presently collecting data. Details on data flow and parameters collected are provided below.

### Data Flow

MARACOOS harvests the ASOW data from an XML feed that Fugro pushes to RPS's ftp server. Data is pushed to the FTP server where it is harvested every 30 minutes by RPS and converted to compliant (CF-1.7, ACDD-1.3, IOOS 1.2) daily netCDF files. During the conversion from XML to netCDF no other transformations or modifications of the data are made. Once the netCDF file is created, RPS runs QARTOD on the file, flags the data, and makes the updated netCDF file available to users via [ERDDAP](#). Several netCDF files and subsequent ERDDAP datasets are created for each buoy to capture both the timeseries and timeseries profile data. These include:

1. ASOW-1 (14 May 2021 through present)
  - a. timeseries data for air pressure, air temperature, humidity, water temperature, salinity, pressure, wave height, period, and direction, wind speed, wind direction, and wind gust

- b. current timeseries profile data for sea water speed and direction from 3 to 47 meters
- 2. ASOW-4 (14 May 2021 through present)
  - a. timeseries data for air pressure, air temperature, humidity, water temperature (at 1, 2, and 32 m), salinity, pressure, wave height, period, and direction, and wind gust
  - b. wind timeseries profile data for speed and direction from 10 to 250 meters
  - c. current timeseries profile data for speed and direction from 3 to 36 meters
- 3. ASOW-6 (29 December 2020 through 14 May 2021)
  - a. timeseries data for air pressure, air temperature, humidity, water temperature (at 1, 2, and 32 m), salinity, pressure, wave height, period, and direction, and wind gust
  - b. wind timeseries profile data for speed and direction from 10 to 250 meters
  - c. current timeseries profile data for speed and direction from 3 to 36 meters

All ASOW data can be visualized in [MARACOOS OceansMap](#), where data tables, timeseries, and profile plots can be created. The ASOW data is also available in the Validate Tool where it can be used to create on-the-fly validations of spatially and temporally overlapping models.

MARACOOS has requested data flow details from Fugro detailing flow from the buoys themselves to the XML feed and will add these details to the data management plan in Fall 2021.

### Quality Control

MARACOOS received unverified data from ASOW and runs QARTOD on the netCDF files after the data is converted from XML to netCDF. Flag values for each test are added to the file, but all original data is retained. The following tests have been implemented:

1. Currents: gross range and flat line
2. Winds: gross range, spike, flat line, rate of change, climatology
3. Air pressure: gross range, rate of change, spike, flat line, climatology
4. Air temperature: gross range, rate of change, spike, flat line, climatology
5. Relative humidity: gross range, rate of change, spike, flat line
6. Sea water temperature: gross range, rate of change, spike, flat line, climatology
7. Sea water salinity: gross range, rate of change, spike, flat line, climatology
8. Sea water pressure: gross range, rate of change, spike, flat line, climatology
9. Significant wave height: gross range, rate of change, spike, flat line
10. Mean wave period: gross range, rate of change, spike, flat line
11. Wave direction: gross range, spike, flat line
12. Wind gust: gross range, spike, flat line, climatology

The tests run on each parameter are used to calculate an aggregate flag that is the primary flag for each parameter.

Fugro also has a quality control process in place for the ASOW data. This process does not run in real-time and thus does not apply to the data currently available on ERDDAP. MARACOOS has been in discussions with Fugro about added verified data to ERDDAP when it is available. When this additional

data is received and processed, we will update this data management plan to document the additional QAQC processes from Fugro.

### **Data Sharing**

The ASOW 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 ASOW data is available via the MARACOOS [ERDDAP](#) server and also displayed (real-time data only) on the ASOW [website](#).

MARACOOS harvests data from XML files pushed by Fugro to the RPS FTP site and stores it on the Azure server that hosts the MARACOOS ERDDAP instance. The data is stored in NetCDF format as individual daily files, which are aggregated by station for access in ERDDAP. The data are accessible via and ERDDAP and include ISO 19115-2 metadata. The [MARACOOS OceansMap](#) platform includes the ASOW data and allows users to query any station to obtain both current and historical values, create time-series and profile plots, and compare the observations with overlapping model data. Each data stream is linked back to its source, in the case the MARACOOS ERDDAP. Time-series data can also be downloaded directly from OceansMap in CSV format.

### **Archiving**

MARACOOS is still in the process of archiving the ASOW data. The provisional data collected in near real-time by MARACOOS will be archived daily from the NetCDF files available in the MARACOOS WAF. The provisional nature of the data will be documented within the NetCDF files and metadata. The DMAC team will review these files and the associated metadata to ensure that they meet the requirements of the NCEI 2.0 templates. If Fugro provides verified data to MARACOOS we will update the archiving procedure to ensure that these data are also archived with NCEI.

### **Weatherflow**

Weatherflow operates the largest industrial grade private observing network in the world. They own and operate 450 stations along the Atlantic, Pacific, Gulf, and Great Lakes coasts that report real-time weather as frequently as every minute. The Weatherflow stations are a combination of a proprietary Scylla data logger and commercial sensors; they have independent power and reliable communication infrastructure. All stations record wind speed, wind direction, air temperature, and barometric pressure. Some stations also monitor relative humidity, rain accumulation, rain rate, visibility, soil temperature, sea surface temperature, surface currents, water level, and wave heights.

### **Data Flow**

Data from Weatherflow stations is transmitted to Weatherflow's data center in Florida in near real-time via cellular modem. All transmissions are encrypted and each observation record is validated by a checksum. If transmission fails, the data is retransmitted. Data is collected at Weatherflow stations is

also stored on a Scylla data logger, which can store up to one year of data in non-volatile memory. If communications go down data can be transmitted at a later date or manually recovered.

Once the data arrives at the data center it is processed and logged on a file server. The raw data is then inserted into a database, processed with Weatherflow's strict QC algorithms and stored in a separate database for access by users. Publicly available data can be accessed by external users via a Sensor Observation Service (SOS) provided by Weatherflow.

Data from the SOS feed is harvested hourly by MARACOOS, where it is converted to compliant (CF-1.7, ACDD-1.3, IOOS 1.2) daily netCDF files. During the creation of the NetCDF no other transformations or modifications of the data are made. Once the files are created they are made available to users via [ERDDAP](#). There are a total of ten stations harvested by MARACOOS, each represented by an individual dataset in ERDDAP. Each station includes wind speed, wind direction, and wind gust.

All Weatherflow data can be visualized in [MARACOOS OceansMap](#), where data tables and timeseries, plots can be created. The Weatherflow data is also available in the Validate Tool where it can be used to create on-the-fly validations of spatially and temporally overlapping models.

### Quality Control

Weatherflow enforces a strict set of quality control procedures on all station data. These procedures are as least as strict as QARTOD and meet the requirement for quality-controlled data. Weatherflow is currently in the process of updating its quality control documentation so that it can be shared. When those details are made available to MARACOOS they will be included in this data management plan, likely in Fall 2021.

### Data Sharing

The Weatherflow 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.

MARACOOS harvests data from the Weatherflow SOS feed and stores it on the Azure server that hosts the MARACOOS ERDDAP instance. The data is stored in NetCDF format as individual daily files, which are aggregated by station for access in [ERDDAP](#). The data are accessible via [ERDDAP](#) and include ISO 19115-2 metadata. The [MARACOOS OceansMap](#) platform includes the Weatherflow data and allows users to query any station to obtain both current and historical values, create time-series and compare the observations with overlapping model data. Each data stream is linked back to its source, in the case the MARACOOS ERDDAP. Time-series data can also be downloaded directly from OceansMap in CSV format.

### Archiving

The Weatherflow station data is archived independently by Weatherflow. The archive process, which compresses the raw logs and transfers them to separate facility, occurs regularly. Transfer occurs through a secure, encrypted VPN tunnel and data at the second facility is stored on hard drives and regularly backed up on removable optical drives.

## References

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

Murphy, S.C., L. Nazzaro, J Simkins, M.J. Oliver, M. Crowley, T.N. Miles, (2021) "Persistent upwelling in the Mid-Atlantic Bight detected using gap-filled, high-resolution satellite SST", *Remote Sensing of Environment*, vol. 262.

NASAa, 2016. Moderate Resolution Imaging Spectroradiometer, <http://modis.gsfc.nasa.gov/about/>. Date accessed, 14 July, 2016.

NASA Jet Propulsion Laboratory (NASAb, 2016). Multi-scale Ultra-high Resolution Sea Surface Temperature, <http://mur.jpl.nasa.gov/>. Date accessed, 14 July, 2016.

NOAA Office of Satellite and Product Operations (NOAA), 2016. Polar-orbiting Operational Environmental Satellites (POES), <http://www.ospo.noaa.gov/Operations/POES/>. Date accessed, 14 July, 2016.