Shipboard Automated Meteorological and Oceanographic System (SAMOS) Initiative

Marine Data Center, Center for Ocean-Atmospheric Prediction Studies, The Florida State University
samos@coaps.fsu.edu

Last updated: 6 November 2019.

1. General Description of Data to be Managed

1.1. Name data collection Project: Shipboard Automated Meteorological and Oceanographic System (SAMOS) Initiative.

1.2. Summary description of the data: Since 2005, the SAMOS initiative has been collecting 1-minute average navigational, meteorological, and oceanographic observations (derived from higher frequency, several per minute up to 1 Hz, sensor measurements) from select research vessels. These averages are produced at 1-minute intervals onboard each participating vessel and are delivered in daily ship-to-shore email messages to the Marine Data Center (MDC) at the Florida State University (FSU). A SAMOS consists of a computerized data logging system that continuously records navigation (ship's position, course, speed, and heading); meteorological (winds, air temperature, pressure, moisture, rainfall, and radiation); and near ocean surface (sea temperature and salinity) parameters while a vessel is underway (A full list of observations requested from recruited vessels is available at https://samos.coaps.fsu.edu/html/parameters.php.

1.3. Is this a one-time data collection, or an ongoing series of measurements? SAMOS is an ongoing series of measurements, updated daily via satellite communication/email service between the vessel operator and the MDC at the Florida State University (FSU).

1.4. Actual or planned temporal coverage of the data: May 2005 to present.

1.5. Actual or planned geographic coverage of the data: Global Oceans (-90 to 90 N, 0 to 360 E), with highest sampling density in the oceans surrounding North America.

1.6. Type(s) of data: Digital numeric data and associated metadata.

1.7. Data collection method(s): research vessels

2. Point of Contact for this Data Management Plan

2.1. Name: Mr. Shawn R. Smith

2.2. Title: Senior Research Associate

2.3. Affiliation or facility: MDC, Center for Ocean-Atmospheric Prediction Studies (COAPS), FSU

2.4. E-mail address: srsmith@fsu.edu

2.5. Phone number: 850-644-6918

3. Responsible Party for Data Management

3.1. Name: Mr. Shawn R. Smith

3.2. Position Title: Senior Research Associate

4. Resources

4.1. Have resources for management of these data been identified? Yes. The SAMOS Initiative at

FSU is base funded via a cooperative agreement (NA16OAR4320199) from the Climate Program Office, Ocean Observing and Monitoring Division of the National Oceanographic and Atmospheric Administration (FundRef#100007298) via a subaward (191001.363513.01D) from the Northern Gulf of Mexico Cooperative Institute administered by the Mississippi State University and the U. S. National Science Foundation's Oceanographic Instrumentation and Technical Services Program (grant # OCE-1447797). Starting in 2013, the Schmidt Ocean Institute (SOI) provided contract funding to recruit the *RV Falkor* to the SAMOS initiative.

5. Data Lineage and Quality

5.1. Processing workflow of the data from collection or acquisition to making it publicly accessible: The flow of SAMOS observations from the vessel to the SAMOS data center (Figure 1) begins with the operator sending all one-minute data records from the previous day to the MDC at 0000 UTC via an e-mail protocol (note: the vessels contributing to SAMOS from New Zealand and Australia post their data to a THREDDS server at the Australian Bureau of Meteorology and the MDC pulls the data from their server). SAMOS uses a custom key:value paired comma-separated value format for data transmission. Each operator encodes one-minute average observations, derived from higher sampling frequency instrumental observations, into the SAMOS format using their vessel's data acquisition software. Once received by the MDC, these observations are converted into a standard network common data form (netCDF) that is augmented with detailed ship and instrumental metadata provided to the MDC by each operator.

At this point, the data undergo a series of scientific data quality control (QC) processes. The first QC process (see below) is fully automated and results in what the MDC calls a preliminary data file. On a 10-day delay from the observation date, intermediate files are automatically created by merging all preliminary files received for a given ship and observation day. This delay allows for receipt of delayed or corrected files from the RV. Finally, a select set of ships (including all recruited NOAA vessels) undergo visual QC to create a research-quality dataset. Preliminary, intermediate, and research quality netCDF files are made publically accessible via the MDC as soon as they are produced via web, ftp, and THREDDS services (see below). Each month, the original data received from the vessel and all three levels of SAMOS-quality processed files are packaged for each ship and submitted to the National Centers for Environmental Information – Maryland (See below).

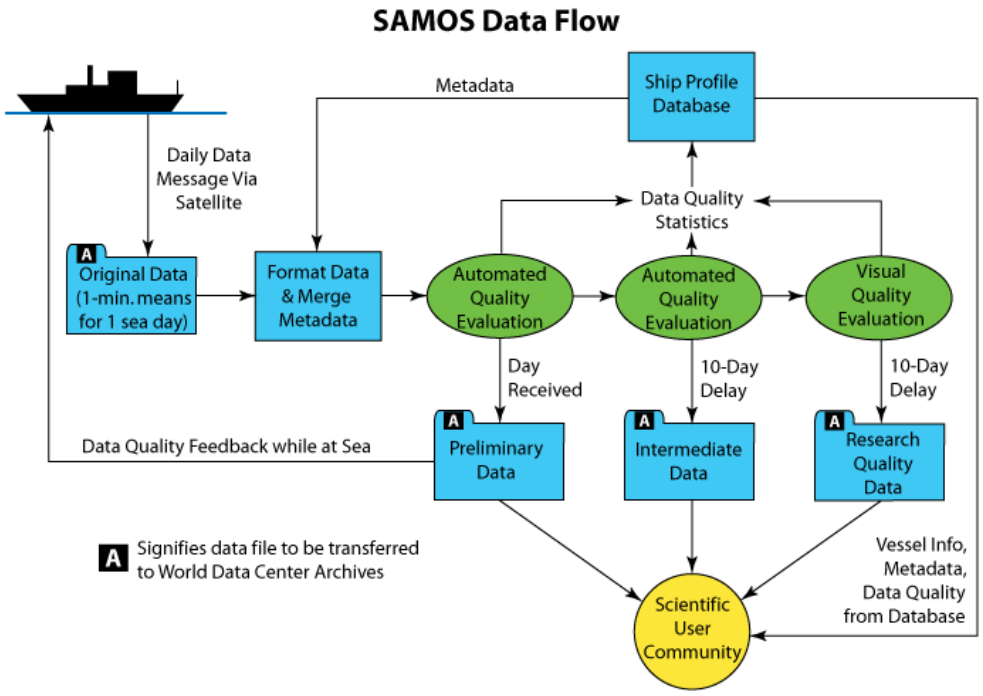

Figure 1. Flow of one-minute sampling rate SAMOS observations from the vessel, through the MDC, and on to the archives and user community.

5.2. Quality control procedures employed: SAMOS data QC begins with verifying that the original file came from a recruited vessel and is in the proper key:value format. Once verified, the data are converted to SI units (if necessary), checked for temporal sequence, and blended with ship and instrumental metadata (e.g., instrument height, units, sensor make/model) from the SAMOS database. This first netCDF version of the observations undergoes automated evaluation to apply flags to the data. SAMOS uses a hierarchical, parametric A-Z quality control scheme (e.g., each value can have only one flag; https://samos.coaps.fsu.edu/html/samos_quality_flag.php). Initial tests verify that (1) the vessel is positioned over water (flag=L) by comparing the vessel latitude and longitude to a two arc-minute ETOPO2 (National Geophysical Data Center [NGDC] 2006) for data processed prior to 31 May 2017, following which the topographic data set was changed to the one arc-minute ETOPO1 (Amante and Eakins 2009), (2) the vessel speed between sequential positions as calculated on a great-circle arc is not greater than 15 m s^{-1} (a realistic speed for a research vessel; flag=F) and (3) the observations are within realistic physical limits (Table 1, flag=B). The pressure, air and sea temperature, wind speed, and relative humidity are also flagged when they exceed ±4σ from a monthly climatology (da Silva et al. 1994; flag=G). The climatology test also uses a minimum standard deviation threshold in data sparse areas (e.g., Southern Ocean) where da Silva et al. (1994) has unrealistically small standard deviations. Another test ensures that the relationship of air temperature ≥ wet-bulb temperature ≥ dew-point temperature is not violated (flag=D; although this test is not commonly used in SAMOS because moisture data is primarily measured as relative humidity). Finally, true winds are recalculated according to Smith et al. (1999) - using the reported vessel course over ground, speed over ground, heading, and relative wind direction and speed – and compared to the reported true wind values. Flags (E) are applied to the reported true winds when the speed (direction) differs by more than 2.5 m s-1 (20˚). This entire process occurs within one to three minutes of the e-mail arriving at FSU.

Table 1. Limits outside of which SAMOS flags the listed parameters with a bounds (B) flag. Air and sea temperatures apply latitude-dependent boundaries in polar (-60 to -90 or 60 to 90 degree N), mid-latitude (-30 to -60 or 30 to 60 degree N), and tropical (-30 to 30 degree N) bands. Ranges are designed to flag 'likely' errors, but do include some realistic values. For example pressure can dip to 880 hPa in a hurricane, but the likelihood of a ship being at that location is extremely low).

SAMOS Parameter (Abbreviation)	Lower Bound	Upper Bound	Units
Latitude (lat)	-90	90	Degrees North
Longitude (lon)	0	359.9999	Degrees East
Speed over ground (PL_SPD)	0	15	ms^{-1}
Course over ground (PL_CRS)	0	360	Degrees
Heading (PL_HD)	0	360	Degrees
True wind direction (DIR)	0	360	Degrees
True wind speed (SPD)	0	40	ms^{-1}
Pressure (P)	950	1050	hPa
Relative humidity (RH)	0	100	percent
Air temperature (T)	-30 (polar) -10 (mid-latitude) 10 (tropical)	15 (polar) 40 (mid-latitude) 40 (tropical)	˚C ˚C ˚C
Sea temperature (TS)	-2 (polar) -2 (mid-latitude) 15 (tropical)	15 (polar) 30 (mid-latitude) 35 (tropical)	˚C ˚C ˚C

Merging data files for a given ship and day to create the intermediate files removes temporal duplicates between multiple files using the QC flags applied to the preliminary files. Duplicates are resolved through a series of tests that first determine whether the data values are exact or different. When they differ, the first test retains the value with the "best" preliminary QC flag. Best flag hierarchy for position data (latitude, longitude) is Z>F>L and for other parameters (sea temperature, humidity, etc.) is Z>G>E>B>D, where Z is the flag used for data that do not fail any QC tests. If the flags on the data values are identical, the second duplicate resolution test compares the values in question to the 30-minute mean centered on the duplicate time, retaining the value closest to the mean. Failure to resolve the duplicate at this stage results in all duplicate values being removed for the time in question and the situation being stored in a processing log (a compromise to allow automation of the merge process).

Visual QC checks on intermediate files for select vessels are completed by a trained meteorological data quality analyst using the SAMOS Visual Data Assessment Tool (SVIDAT). The analyst reviews all observations and has the option to remove flags applied by the automated QC and/or add new flags based on the analyst's experience. In general, visual QC will only involve the application of quality control flags to identify discontinuities (H), interesting features (I), obviously erroneous values (J), suspicious/suspect values (K), known instrument malfunctions (M), occurrences of the vessel being in port (N) and spikes (S). Quality control flags J, K, and S are the most commonly applied by visual inspection, with K being the catchall for the various issues common to most vessels, such as (among others) steps in data due to platform speed changes or obstructed platform relative wind directions, data from sensors affected by stack exhaust contamination, or data that appears out of range for the vessel's region of operation. M flags are primarily assigned when there has been communication with vessel personnel in which they have dictated or confirmed there was an actual sensor malfunction. Port (N) flags are reserved for the latitude and longitude parameters and are rarely used, in an effort to minimize over-flagging. The primary application of the port flag occurs when a vessel is known to be in dry dock. The port flag may also be applied, often in conjunction with flags on other parameters, to indicate that the vessel is

confirmed (visually or via operator) in port and any questionable data are likely attributable to dockside structural interference, although this practice is traditionally only used in extreme cases. The I flag is optionally used to identify meteorologically interesting values (e.g., pressure minima associated with a frontal passage or tropical cyclone). SAMOS data analysts may also apply Z flags to data, in effect removing flags that were applied by automated QC. For example, B flagging is dependent on latitude and occasionally a realistic value is assigned a B flag simply because it occurred very close to a latitude boundary. This happens with sea temperature from time to time in the extreme northern Gulf of Mexico – TS values of 32˚C or 33˚C are not unusual there in the summer, but portions of the coastline are north of 30˚ latitude and thus fall into a region where such high temperature are coded as "out of bounds." In this case the B flags would be removed by the data analyst and replaced with good data (Z) flags.

The SAMOS QC system is based on the procedures used during the World Ocean Circulation Experiment and the original QC documentation is available at https://coaps.fsu.edu/woce/docs/qchbook/qchbook.htm. Further details can be found in Smith et al. (2018).

6. **Data Documentation**

6.1. Metadata. SAMOS collects a broad suite of vessel- and instrument-specific metadata. The metadata specification was based on a combination of the metadata requirements for the Voluntary Observing Ship Climate (VOSClim; https://www.ncdc.noaa.gov/data-access/marineocean-data/vosclim/ship-metadata) schema and metadata included in the International Comprehensive Ocean-Atmosphere Data Set (ICOADS; Freeman et al 2016). Metadata are submitted to the MDC from the operators initially when a vessel is recruited using forms available on the SAMOS website (https://samos.coaps.fsu.edu/html/participate.php). The operators update their metadata periodically by either resubmitting these forms to the MDC or via graphical metadata user interface on the SAMOS website. Metadata submitted to SAMOS are stored in a MySQL database, where all instrumental metadata are date tracked and versioned. This database is used to augment the SAMOS netCDF files that contain the actual navigational, meteorological, and oceanographic data.

Once submitted to NCEI for archival, additional metadata records are prepared by NCEI for each archive package according to NCEI procedures and outlined in a data submission agreement between the MDC and NCEI. This includes creating FGDC and ISO-19115-2 metadata for each ship and month of SAMOS data submitted to NCEI.

6.2. Name of organization or facility providing metadata hosting: MDC, COAPS (original metadata) and NCEI-Maryland (for archive metadata).

6.3. URL of metadata folder or data catalog:

6.3.1. Via SAMOS web site: https://samos.coaps.fsu.edu/html/meta.php

6.3.2. Via NCEI geoportal: https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.nodc:COAPS-SAMOS

6.4. Process for producing and maintaining metadata: See 6.1 above.

7. **Data Access**

7.1. Do these data comply with the Data Access directive? Yes. The MDC at FSU has a long standing policy to ensure free and open access to underway marine meteorological observations. For SAMOS, the policy ensures that all data provided to FSU will be redistributed to the user community and national archive centers without any restrictions or "holds." Data providers are requested to notify the SAMOS data center in writing (samos@coaps.fsu.edu) if they have any concerns with this free and open data policy.

7.2. Name of organization of facility providing data access: MDC, COAPS, FSU

7.3. Data access methods or services offered:

7.3.1. HTTP - (By date) https://samos.coaps.fsu.edu/html/data_availability.php; (By cruise, when available): https://samos.coaps.fsu.edu/html/cruise_data_availability.php

7.3.2. THREDDS – https://www.coaps.fsu.edu/thredds-listing

7.3.3. FTP - ftp://ftp.coaps.fsu.edu/samos_pub/data

7.4. Approximate delay between data collection and dissemination*:* Preliminary quality controlled data are available within minutes of receipt at the MDC, intermediate files are available on a 10-day delay from the date of receipt, and research quality data are typically available within a few weeks of receipt.

8. **Data Preservation and Protection**

8.1. Actual long-term data archive location: NCEI-Maryland – Archived SAMOS data are accessible via https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.nodc:COAPS-SAMOS and have been assigned a digital object identifier by NCEI (Smith et al. 2009).

8.2. Data storage facility prior to being sent to an archive facility: MDC, COAPS, FSU.

8.3. Approximate delay between data collection and submission to an archive facility: Two months. Archive packages for each SAMOS ship are created approximately 11 days following the end of the month of data observation. NCEI typically acquires and archives these data approximately 30 days following the creation by the MDC of each archive package. We are working with NCEI to reduce the delay between archive package creation at the MDC and final archival at NCEI.

8.4. How will the data be protected from accidental or malicious modification or deletion prior to receipt by the archive? All original SAMOS data files and subsequent quality-controlled netCDF files are held with multiple copies on RAID6 storage disks which provide redundant storage at the MDC. The data are also backed up daily via a ZFS snapshot and we are working to re-establish external tape/disk backups. The SAMOS SQL database is dumped every 4 hours to a separate backup storage server. Plans are being developed to provide for off-site (e.g., a data center outside of Florida) replication of the SAMOS data infrastructure to support disaster recovery.

The archive packages submitted to NCEI also provide an off-site back-up of mission critical files. These packages include the original data received from the vessel and all versions of the quality controlled data. Every file in the archive package is listed in a file manifest with MD5 checksums, which ensures that the data package is received at NCEI (and or returned to the MDC) without loss of any digital information.

9. **References**

Amante, C. and Eakins, B. W. (2009). ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA. doi:https://doi.org/10.7289/V5C8276M

da Silva, A. M., C. C. Young, and S. Levitus, 1994: *Atlas of Surface Marine Data, Vol. 1: Algorithms and Procedures*. NOAA Atlas NESDIS 6, U. S. Department of Commerce, NOAA, 83pp. Available from: NOAA/NODC, Customer Service, E/OC, 1315 East-West Highway, Silver Spring, MD 20910.

Freeman, E., S. D. Woodruff, S. J. Worley, S. J. Lubker, E. C. Kent, W. E. Angel, D. I . Berry, P. Brohan, R. Eastman, L. Gates, W. Gloeden, Z. Ji, J. Lawrimore, N. A. Rayner, G. Rosenhagen, and S. R. Smith, 2016: ICOADS Release 3.0: a major update to the historical marine climate record. *Int. J. Climatol*. doi:10.1002/joc.4775

National Geophysical Data Center. (2006) 2-minute Gridded Global Relief Data (ETOPO2) v2. National Geophysical Data Center, NOAA, https://doi.org/10.7289/V5J1012Q

Smith, S. R., M. A. Bourassa, and R. J. Sharp, 1999: Establishing more truth in true winds. *J. Atmos. Oceanic Technol*., **16**, 939-952.

Smith S. R., J. J. Rolph, K. Briggs, and M. A. Bourassa, 2009: Quality-controlled underway oceanographic and meteorological data from the Center for Ocean-Atmospheric Prediction Studies (COAPS) – Shipboard Automated Meteorological and Oceanographic System (SAMOS). National Oceanographic Data Center, NOAA. doi:10.7289/V5QJ7F8R.

Smith, S. R**.**, K. Briggs, M. A. Bourassa, J. Elya, and C. R. Paver, 2018: Shipboard automated meteorological and oceanographic system data archive: 2005–2017. *Geosci Data J*., **5**, 73–86. https://doi.org/10.1002/gdj3.59