

SAMOS Versions and Quality Control

Drafted: 1 June 2017

This document provides a brief overview of the data flow, data processing versions, and quality control procedures applied to Shipboard Automated Meteorological and Oceanographic System (SAMOS) observations from research vessels recruited by the Marine Data Center (MDC) at the Florida State University.

Processing Workflow and Version Control

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 ship and instrumental metadata provided to the MDC by each operator.

The data then 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 (version 100)** data file. On a 10-day delay from the observation date, **intermediate (version 200)** 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 **research-quality (version 300)** data files. 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. 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 (Smith et al. 2009). As a final note on SAMOS versions, sometimes it is necessary to modify existing version 100, 200, or 300 files. In these cases, the respective version will be incremented by 001 (e.g., v201, v202, or v301, v302) and these files are also distributed and archived.

*The MDC staff recommends using the **research-quality dataset** whenever it is available and the **intermediate** version for vessels where the research-quality product is not produced. The preliminary product is recommended primarily for operational activities that require data access soon after the data are received at the MDC (for those with requirements to use the data prior to the 10-day delay when the intermediate product is created).*

Quality Control Procedures

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 QC 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; http://samos.coaps.fsu.edu/html/samos_quality_flag.php). Initial tests verify that (1) the vessel is positioned over water by comparing the vessel latitude and longitude to a 2-minute gridded global relief dataset (ETOPO2; U.S. Dept. Commerce 2006; flag=L), (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 $\pm 4\sigma$ 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 \geq wet-bulb temperature \geq 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 a data file arriving at FSU and results in a **version 100 preliminary** file.

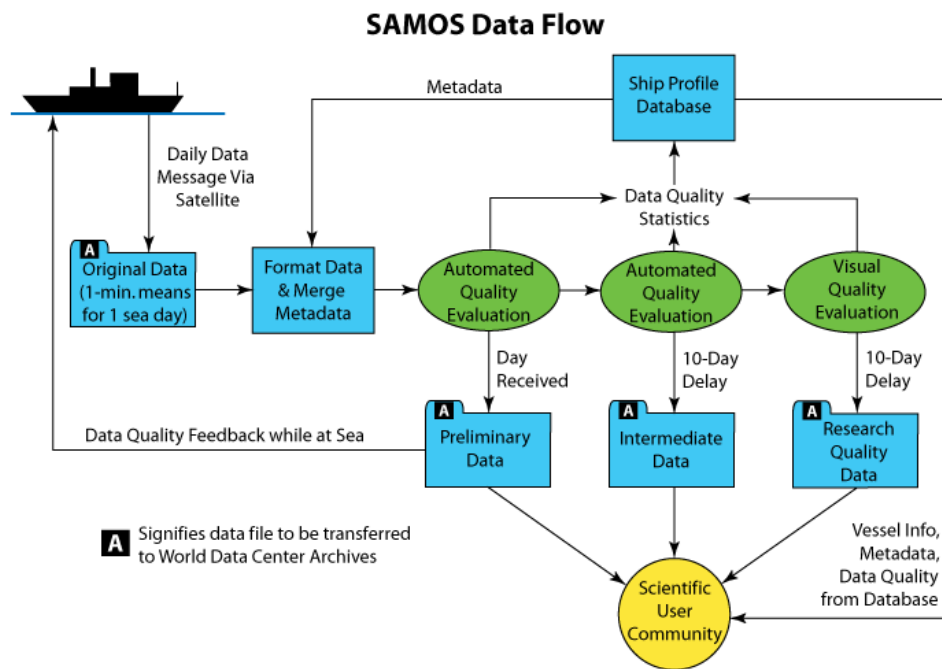


Figure 1. Flow of one-minute sampling rate SAMOS observations from the vessel, through the MDC, and on to the archives and user community.

Merging multiple files for a given ship and day to create an **intermediate (version 200)** file 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 file 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 QC 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 outlined in Table 1. 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. Visual QC, which is only done for select vessels, results in **research-quality (version 300)** files.

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 (Abbreviation)	Parameter	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

The SAMOS QC system is based on the procedures used during the World Ocean Circulation Experiment and the original QC documentation is available at <http://coaps.fsu.edu/woce/docs/qchbook/qchbook.htm>.

References

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.

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.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Geophysical Data Center, 2006: *2-minute Gridded Global Relief Data (ETOPO2v2)*. Available from <http://www.ngdc.noaa.gov/mgg/fliers/06mgg01.html>.