

6 January 2010

**Ocean Data Quality Control Lecture Notes**  
**GODAE Summer School 11-22 January 2010**  
**University of Western Australia**  
**Perth, Australia**

**James A. Cummings**  
**Naval Research Laboratory**  
**Monterey, California, USA**

**Abstract:** Automated ocean data quality procedures are presented. The procedures are logically grouped into four stages of ocean data quality, which when taken together form a complete sensor-to-prediction quality control system. The main features of the ocean observing systems are presented along with sources and examples of errors that can occur in the data. Specific quality control procedures are described that test for these errors as well as more general procedures that estimate the consistency of the data across all observing systems. The importance of real-time ocean data quality control as an observing system monitoring tool is emphasized, with some specific examples of new techniques for doing this function that have been developed in numerical weather prediction systems.

**1. Introduction:** Observation data quality control is a fundamental requirement of GODAE ocean data assimilation systems. Using or accepting erroneous data can cause an invalid conclusion to be made or an incorrect analysis. Alternatively, rejecting extreme, but valid, data can miss the detection of important events. The goal of quality control, therefore, is to reduce or eliminate making the wrong decisions. Quality control must correctly identify observations that are obviously in error, as well as the more difficult process of identifying measurements that fall within valid and reasonable ranges, but nevertheless are erroneous. It is likely that decisions made at the quality control step affect the success or failure of the entire analysis/forecast system.

Ocean data quality control is best performed in stages. The first stage consists of a series of data sensibility checks. Observations failing any one of these tests are considered to have gross errors and are removed from the system. The second stage is based on a complex quality control procedure where observations are subjected to a series of tests, with the final quality control decision based on simultaneous consideration of results from all of the individual quality control tests. Observations are not rejected after failing any one of the individual quality control tests. This process uses a decision-making algorithm where the ultimate fate of the observation in the analysis/forecast system is decided (accept, reject, schedule for manual intervention). The outcome of the decision-making algorithm represents the likelihood an observation contains a random error. The third stage of ocean data quality control is performed by the analysis system itself. At this point, the gross and random error characteristics of the observations are known, and observations considered acceptable for the analysis have been selected. This third stage of the quality control is designed to protect the analysis from marginally acceptable data that have, for unknown reasons, passed the earlier stages of the quality control. It is of interest to examine observations rejected by the analysis in the third stage of the quality control to understand better the failure of the earlier quality control stages to detect and reject what turned out to be truly erroneous observations. A final stage of the quality control is done after the analysis and the

forecast have been completed as part of a system that performs routine assessment of data impact on reducing forecast error.

In this paper, various approaches, procedures, and algorithms used to quality control ocean observations are described. The emphasis is on real-time, fully automated ocean data quality control. It is beyond the scope of this paper to discuss the wide variety of delayed-mode or manual intervention quality control efforts that have been implemented. The paper is organized as follows. Section 2 describes the real-time ocean observing systems, and section 3 gives the stand alone, gross error quality control procedures that are often applied to ocean observations. Section 4 provides descriptions of external quality control data checks and brief overviews of specific sources of error in ocean observing systems. Section 5 describes how the various independent external quality control data checks can be combined in a quality control decision-making algorithm. Section 6 outlines the internal consistency checks that are used in the assimilation system itself, while section 7 describes the quality control outcomes from the data impact system. Finally, section 8 provides a summary and section 9 gives some conclusions on the interactions between ocean data quality control and observation monitoring.

**2. Ocean Observing Systems.** A wide variety of observation data types are used in GODAE assimilation systems. The data include both *in situ* measurements and measurements from space. As will be discussed, each observing system has its own unique data issues and quality control requirements. Sources of operational ocean data are described in this section.

Most GODAE centers receive their *in situ* ocean observations over the Global Telecommunication System (GTS). At the current time, data transmitted via the GTS are coded in specific data type formats which use, at most, two decimal places for measurements of temperature and salinity. Further, the existing formats do not allow for additional information about the data in the form of quality flags. However, observational data on the GTS are moving to a new binary format based on BUFR (Binary Universal Form for the Representation of data – a data format maintained by the World Meteorological Organization). In this format data values can be transmitted with more precision than the existing text-based data formats, and local tables can be added to the message that contain value added or quality assurance information from the data provider. The move to BUFR on GTS is a long process, scheduled to be completed for all ocean data types in 2016.

There is no standard way satellite oceanographic observations are distributed to GODAE assimilation systems. In some cases, there is a dedicated push from the data provider to the center. In other cases, data are placed on dedicated servers where the observations are then pulled by the center. The GODAE High Resolution SST pilot project has been instrumental in setting up data servers where satellite SST data providers transmit their SST retrievals in near-real-time in a common format. This effort has made the availability of SST data from a wide variety of satellite systems commonplace.

*2.1 Ship and Buoy Sea Surface Temperature and Salinity.* Ship sea surface temperature (SST) observation data types are identified as engine room intake, hull contact sensor, or bucket temperatures based on type codes contained in the ship reports received over the GTS. Buoy SST data types are also received from GTS and consist of fixed and drifting buoy reports.

Observing systems that report both *in situ* SST and sea surface salinity (SSS) include thermo-salinograph observations from ships are sent over GTS in TRAKOB reports.

*2.2 Satellite Sea Surface Temperature.* There are many sources of satellite SST observations. Infrared satellite sensors include the NOAA and METOP AVHRR polar orbiter global area coverage (GAC, 8.8 km) and local area coverage (LAC, 2.2 km) SST retrievals. The Geostationary Operational Environmental Satellite (GOES) infrared SST measurements have a resolution of 12 km and are available from the GOES-10, GOES-11, and GOES-12 satellites. The Meteosat Second Generation (MSG) is also a geostationary satellite and provides 12-km resolution infrared SST measurements centered over Europe. The Advanced Microwave Sensor Radiometer Earth (AMSR-E) on board the NASA Aqua satellite provides global coverage of microwave SST. The Advanced Along Track Scanning Radiometer (AATSR) on board ENVISAT provides the first routine measurements of a true skin SST.

*2.3 Sea Ice Concentration.* The Special Sensor Microwave Imager (SSM/I) and the Special Sensor Microwave Imager/Sounder (SSMIS) on board the Defense Meteorological Satellite Program (DMSP) series of satellites provide routine observations of sea ice concentration at approximately 25 km resolution. At the present time there are 3 SSM/I (F11, F13, F15) and 2 SSMIS (F16, F17) satellites providing sea ice data.

*2.4 Temperature and Salinity Profiles.* Profile observations are reported from both fixed and moving platforms. Most profile observations report only temperature, such as expendable bathythermographs (XBT) and some fixed buoys, but profiling floats (Argo), conductivity-temperature-depth (CTD) sensors, and an increasing number of fixed buoys report both temperature and salinity. Profile observations are also available from gliders that measure temperature and salinity at varying depths and positions along a dive. A glider dive consists of a descending and an ascending profile in which the latitude, longitude positions and times of the observations change with depth. The presence of temperature and salinity in some reports and the unique sampling characteristics of ocean gliders present new challenges to the quality control of ocean profile data. There are common approaches to the quality control of the various profile observing systems, but there are also unique instrumentation specific error checks that are performed.

*2.5 Altimeter Sea Surface Height.* At the present time, sea surface height anomaly (SSHA) observations are available from satellite radar altimeters on board the Jason-1, Jason-2, and ENVISAT satellites. Historically, satellite altimeters have been flown on Topex/Poseidon, Geosat and Geosat Follow-on, and the European Remote Sensing series of satellites (ERS-1 and ERS-2). Altimeter SSHA observations undergo a wide variety of geophysical corrections, inter-satellite calibrations, and along track averaging before the data are used in GODAE assimilation systems.

*2.6 Altimeter and Buoy Significant Wave Height.* Each satellite radar altimeter also provides significant wave height (SWH) and wind speed observations. SWH observations are also available from many fixed buoy locations, mainly in the northern hemisphere. The SWH observations are used in wave model assimilation systems.

**3. Preliminary Data Checks.** Typically, several preliminary gross error data checks are performed prior to the quality control of the observed values. Observations failing any one of these preliminary data checks are discarded. In some cases these preliminary data checks are performed by the data provider and the data are simply not distributed. The preliminary data checks and logic for accepting/rejecting observations at this stage of the quality control process are described below.

*3.1 Land/Sea Test.* All observation locations are checked against a global, high-resolution land/sea database. Observation locations surrounded by water in all directions are accepted, and observation locations surrounded by land in all directions are discarded. For observations very near the coast a fuzzy land/sea boundary check is used. Observation locations are accepted if, in any one direction, the otherwise over-land location is next to a water point. This fuzzy test allows for resolution errors of the land/sea database and precision errors in the reporting of observation latitude and longitude positions over the GTS. Examples of where a fuzzy land/sea test is useful include: (1) ships providing observations while parked at a dock, and (2) instruments deployed on piers very close to the coast. Note that the land/sea test is applied to observations from both fresh- and seawater locations. Remotely sensed lake temperatures are routinely provided.

*3.2 Location/Speed Test.* A location (speed) test is used to determine if the reported position of an observation is consistent with prior positions from the same platform. This test is restricted to data types that report unique call signs, such as Argo floats, surface ships, aircraft, and fixed and drifting buoys. Observations failing the location/speed test are typically scheduled for manual intervention. No automated method exists at the present time to correct erroneous reported positions. A sliding time window of the last 25 reported platform positions is used in these calculations. The speed test takes into account differences in the expected movement of airborne versus surface ship platforms, as well as a test to ensure that identification of a buoy as fixed or drifting is correct, based on the expected pattern of buoy all numeric call signs. Care is taken to ensure that the speed test does not inadvertently reject new drifting buoy observations when drifting buoy call signs are reused. The practice of reusing buoy call signs often results in very large changes from the last reported position of the failed buoy and the position of the new buoy with the same call sign. This problem is minimized by not checking locations from buoys with the same call sign that have reporting times which differ by more than 120 hours. Observations are rejected if the reported position differs by twice the expected rate of speed as computed from the recent time history of platform locations.

*3.3 Valid Value Range Tests.* Valid value range tests are applied to observed variables as well as observation locations and sampling times. Reported values of temperature are required to be greater than  $-2.5^{\circ}\text{C}$  and less than  $42^{\circ}\text{C}$ . Reported values of salinity are required to be greater than or equal to 0 PSU and less than 42 PSU. Geographic dependencies can be built into the temperature and salinity valid value range test to take into account unique oceanographic conditions in marginal seas like the Red Sea, Mediterranean Sea, and Black Sea. Observation latitudes must be between  $-90$  and  $90$  degrees, and observation longitudes must be between  $-180$  and  $180$  degrees. Observation sampling times must contain valid year, month, day, hour, minute, and second date-time information, and the combined date time group cannot formally lie in the future.

*3.4 Duplicates.* Detection of duplicate reports in ocean observational data is a difficult and recurring problem. The same data message can be received multiple times over different networks at the operational centers. One powerful method of detecting duplicates is to use the cycle redundancy check (CRC). The CRC checksum is calculated as a function of the message contents. It processes each byte of a message file. Any change, no matter how small, will produce a different CRC number. Identical CRC numbers, therefore, will indicate an exact duplicate message and one of the messages can simply be deleted. The real problem with detecting duplicates, however, is the issue of near-duplicates. These are observations with the same location, time and exact match of data type, but otherwise are different. It may be that one version has lower precision, fewer reporting levels, or a different reporting source. It is not possible to reject near-duplicates at the preliminary data check level of quality control processing. Further examination of the reported values is usually needed to make an informed decision. Often the determination of near-duplicates is not done until observations are read into the analysis. Closely spaced observations are typically thinned horizontally prior to the analysis to ensure that the covariance matrix is not ill conditioned. Near-duplicates are naturally encountered in this processing and decisions to retain or toss an observation at that point can take into account additional information about the observation, such as quality control outcomes from the external data checks described in the next section.

**4. External Data Checks.** Effective quality control is a strong function of the amount of information available. A primary purpose of the quality control system is to gather validated information about a newly received observation in order to determine the consistency of the new observation with what is known about the observed variable. Knowledge of the uncertainty of the observation and the collocated information is also needed to formulate and test hypotheses in the quality control decision-making process. This information is acquired and combined in a series of external data checks that are performed prior to the analysis. Many of these pre-analysis quality control procedures are specific to an observing system and test for instrument failure or known biases in certain data types. Other pre-analysis quality control procedures are common to more than one data type. In this category, checks against appropriate background field checks and cross validation analyses are particularly important. Pre-analysis procedures in common are described first, followed by descriptions of procedures unique to specific data types.

*4.1 Background Field Check.* Background fields used to quality control ocean data include climatology, short-term forecasts, and global or regional analyses. In all cases, appropriate background error variances must be used. Background and background error fields at the observation sampling time are interpolated to the observation location. An innovation is formed (observation minus background) and normalized by the error estimate of the background field. Assuming the errors are normally distributed, the probability the observation contains a random error is computed by,

$$p(x \leq X) = (\sigma\sqrt{2\pi})^{-1} \int_{-\infty}^X e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} dx \quad (4.1)$$

where  $x$  is the observed value,  $\mu$  is the background value,  $\sigma$  is the background error standard deviation, and  $p$  is the area to the left of  $X$  beneath the standardized normal probability curve. The error probability can be used to screen data at this stage of the quality control. Histograms and formal statistical tests, however, should be examined to show that the normalized

innovations are indeed normally distributed in order to use the probability of error values in making decisions to accept or reject observations.

*4.2 Cross Validation.* Cross validation compares observations against other nearby data. A variety of methods are used to make these comparisons. The most common method is to perform an optimum interpolation (OI) analysis at the observation location and sampling time using nearby validated data, excluding the datum being checked. The innovations for the cross validation are typically computed from climatology. It is important to ensure that cross validation checks are data driven and independent of any analysis/forecast model backgrounds. The uncertainty of the analyzed value is computed from the OI analysis error reduction of climate variability. The cross validation analyzed value and its uncertainty are then used as the background value and background error in a background field check similar to the one described in section 4.1. In the absence of any nearby valid data, the cross validation procedure simply returns climate and climate variability as the analysis and error estimates, and the cross validation check is identical to the background check using climatology. Thus, the cross validation check can be thought of as a dynamic climatology.

The background error covariances used in the cross validation procedure can be very simple, data within some specified distance from the observation being checked, or more complicated, based on the covariances used in the assimilation procedure itself. Cross validation can be applied to all observation data pairs in the quality control or it can be preceded by other data checks which first detect suspect observations. The cross validation is then performed only on the suspect observations to save on computational time. In data sparse areas the cross validation check will have limited effect. However, the continuing development of the Argo profiling float array allows the cross validation of profile observations to work well in practice. Cross validation is also useful in the quality control of altimeter SSH and SWH observations, since those data tend to be rejected along sequential segments of altimeter tracks due to phase errors in the model background fields.

*4.3 Ship and Buoy Sea Surface Temperature.* Volunteer observing ship (VOS) temperatures have very different error characteristics depending on the measurement method. Hull contact sensor measurements of temperature appear to be the most accurate followed by engine room intake and buckets. However, all ship-based measurement systems are prone to error since the on-board instruments are rarely calibrated. In general, ship-based SST measurements are noisy and tend to be warm biased from the use of engine room intake instruments. In addition, there appears to be some geographic dependence in ship SST errors, with errors higher in the Pacific than in the Atlantic.

Drifting buoy measurements of SST are very important since the buoys are globally distributed and have a relatively long life. In general, drifting buoy SST measurements are of high accuracy and high quality. Occasionally, spurious drifter locations are received, but these can easily be checked using the location speed check. Buoy SST measurements can usually be reconciled by a series of background field checks using climatology or analysis fields. Since drifting buoys and ships are identified by unique call signs the time history of individual instruments can be monitored for indications of drift and calibration errors. In this regard, the UKMO maintains a

grey list of buoy call signs that have been determined by users of the data to be of low quality. The UKMO grey list should regularly be reviewed by users of drifting buoy SST measurements.

*4.4 Satellite Sea Surface Temperature.* Infrared and microwave satellite SST retrievals measure very different properties of the sea surface, requiring unique quality control procedures.

*4.4.1 Residual Cloud Contamination.* Infrared SST measurements are derived from radiometric observations at wavelengths of  $\sim 3.7 \mu\text{m}$  and  $11\text{-}12 \mu\text{m}$ . Though the  $3.7 \mu\text{m}$  channel is more sensitive to SST, it is primarily used only for night-time measurements because of the relatively strong reflection of solar irradiation in this wavelength region during daytime, which contaminates the retrieved radiation. The infrared wavelength bands are sensitive to the presence of clouds and atmospheric water vapor. For this reason, thermal infrared measurements of SST first require atmospheric correction of the retrieved signal and can only be made for cloud-free pixels. However, cloud clearing algorithms are far from perfect and atmospheric water vapor variations are significant. In addition, satellite zenith angle plays a role in determining SST errors, since the atmospheric path length over which the retrieved radiation is observed is longer at higher zenith angles. These errors are not accounted for in the calibration of infrared satellite SST retrievals using the drifting buoy network. Accordingly, quality control or bias correction of infrared satellite SST retrievals is required prior to assimilation.

*4.4.2 Aerosol Contamination.* Satellite sea surface temperature retrievals from infrared radiometers are known to be prone to bias when significant amounts of aerosol are present in the atmosphere. Retrievals are degraded by the presence of tropospheric aerosols, as are cloud detection tests that depend upon accurate visible and infrared channel measurements. In particular, desert dust particles are large enough to attenuate and contribute to the infrared signal emitted from the ocean surface before it reaches the satellite sensor in space. Saharan dust events are common in the eastern tropical Atlantic and Mediterranean Sea. Saharan dust is lifted by convection over hot desert areas, and can reach very high altitudes; from there it can be transported over the ocean by winds, covering distances of thousands of kilometers. The dust combined with the hot dry air of the Sahara Desert has significant effects on tropical weather, especially as it interferes with the development of hurricanes. In Eastern Asia, mineral dust events originate in springtime in the Gobi Desert (Southern Mongolia and Northern China). The aerosols are carried eastward by prevailing winds, and pass over China, Korea, and Japan, sometimes as far as the western United States. The impact of atmospheric aerosols on infrared SST retrievals is a cold bias that is a global problem.

The current limiting factor for dealing with aerosol contamination in satellite SST retrievals is accurate knowledge of the characteristics and amount of the aerosol at the coincident time and location of the satellite SST retrievals. This information is available in the daytime for the anti-solar side of the scan in the visible channels of the instrument, but there is no information at night or on the solar side of scan from the effects of sun glint (75% of the data). However, aerosol transport models can be used to provide the necessary information. One model, the Navy Aerosol Analysis Prediction System (NAAPS), provides 3-hourly, aerosol optical depth (AOD) forecast products for four aerosol sources (dust, smoke, sulphate, and sea spray) at 19 different wavelengths; 14 of the NAAPS aerosol optical depth wavelengths match the wavelengths of the channels used in the satellite SST retrieval algorithm.

The wavelength dependent, global NAAPS optical depth products are used in a canonical variate analysis to detect aerosol contamination. Canonical variate analysis finds the linear combination of observed variables that maximize the ratio of between-group to within-group variation. There are five groups of infrared satellite SST retrievals: four groups are defined with varying levels of aerosol contamination and one group is free from aerosol contamination. The canonical variates are then used to discriminate between the groups. Separate canonical variate functions have been computed for day versus night retrievals and for difference geographic areas.

Let  $\mathbf{B}$  be the between group covariance matrix and  $\mathbf{W}$  the within group covariance matrix. Linear variate functions ( $\lambda$ ) are found to maximize,

$$\nu = \lambda' \mathbf{B} \lambda / \lambda' \mathbf{W} \lambda \quad (4.2)$$

which represent the ratio of between group to within group variances. Differentiating Eq. (4.2) and setting to zero gives,

$$(\mathbf{B} - \nu \mathbf{W}) \lambda = \mathbf{0}. \quad (4.3)$$

The eigenvalues of  $\mathbf{W}^{-1} \mathbf{B}$  and the corresponding eigenvectors ( $\lambda$ ) are the canonical variate functions used for discrimination. The observed channel brightness temperatures and NAAPS wavelength dependent AOD components are projected onto the canonical variates and the Euclidean distances to the projected group means is determined. The SST retrieval is classified as contaminated if the distance to a contaminated group mean is closer than the distance to the non-contaminated group mean according to,

$$\kappa = \min_j \sum_{i=1}^r [\lambda_i (x - \mu_j)]^2 \quad (4.4)$$

where  $\mathbf{x}$  is the AOD wavelength components and AVHRR channel brightness temperatures vector for a given SST retrieval,  $r$  is the number of canonical variate functions,  $\mu_j$  is the group mean vector of observed values for the contaminated and non-contaminated groups, and  $\kappa$  is the group classification code. The group assignment probability is computed assuming that group distances are chi-square distributed with  $r-1$  degrees of freedom. Satellite SST retrievals assigned to a contaminated aerosol group are either rejected or corrected (experimental).

*4.4.3 Diurnal Warming.* Surface diurnal warming events are common in the world oceans. The warming events produce near-surface thermal gradients that create daytime near-surface or warm-layer temperatures 2-4°C warmer than nighttime. Although not strictly a measurement error, combining measurements with different observation times in a daily analysis requires consideration of diurnal warming events. Knowledge of diurnal warming events, in turn, requires information on the local wind speed and surface solar radiation at the time of the SST observation. Typically, satellite SST retrievals are collocated with surface wind speed and solar radiation fields from NWP systems, which are then used to detect diurnal warming and potential skin-layer effects in the satellite SST retrieval.

*4.4.4 Microwave SST.* Due to lower signal strength of the radiation curve in the microwave region, accuracy and resolution is poorer for SST derived from passive microwave measurements as compared to SST derived from thermal infrared measurements. However, the advantage gained with passive microwave is that radiation at these longer wavelengths is largely unaffected by clouds and generally easier to correct for atmospheric effects. Phenomena which

do affect passive microwave signal return, however, are wind-generated roughness at the ocean surface and precipitation. These affects can usually be corrected for using multiple frequencies. SST measurements are primarily made at a channel near 7 GHz with a water vapor correction enabled by observations at 21 GHz. Other frequencies used for correction of surface roughness (including foam), precipitation, and what little effect clouds do have on microwave radiation, include information in the 11, 18, and 37 GHz channels. Nevertheless, rain contamination continues to be a problem at the edge of rain cells, where there is often undetected rain that cause biased SST retrievals. Land contamination is also an issue with microwave measurements. Within 50-100 km of land microwave measurements are affected by emissions from land resulting in a warm bias in coastal microwave SST. For this reason, microwave SST observations are typically not produced within 100 km of land.

*4.5 Sea Ice Concentration.* A problem with sea ice concentration retrievals from the SSSM/I and SSMIS sensors on board the DMSP series of satellites is the false indication of sea ice over the open ocean and at the ice edge. These spurious sea ice concentrations result from the presence of atmospheric water vapor, non-precipitating cloud liquid water, rain, and sea surface roughening by surface winds. While these effects are relatively minor at polar latitudes in winter, they result in serious weather contamination problems at all latitudes in summer. The various sea ice retrieval algorithms used operationally attempt to eliminate these false positive sea ice concentrations, but with limited success. Accordingly, prior to assimilation, sea ice concentrations need to be quality controlled using a weather filter. A good proxy for a weather filter is SST. If the sea ice concentration is greater than zero and the collocated SST exceeds  $4^{\circ}\text{C}$ , then it is likely the sea ice retrieval is contaminated by a weather event and should be rejected. The  $4^{\circ}\text{C}$  SST threshold is considered to be very conservative. Tests with a  $1^{\circ}\text{C}$  SST threshold resulted in spurious rejections of sea ice retrievals in the East Greenland Current during periods of rapid ice growth, when the analyzed SST fields are not accurate due to a lack of SST observations. A cross validation check will not work as a weather filter, since weather contamination is typically large scale affecting many near-by sea ice retrievals simultaneously.

Sea ice retrieval algorithms also return false positive ice conditions near land due to land contamination of the microwave signal. This bias is most evident during the summer ice melt season in the northern hemisphere when the Arctic land boundaries become ice free. A high resolution distance from land database is used to check if the retrieval is within 100 km of land. The test uses retrieval distance from land, background field anomalies, collocated SST, and a cross validation check of nearby locations to determine if positive sea ice observations near land are valid. Land contaminated sea ice retrievals are typically rejected at this point.

*4.6 Temperature and Salinity Profiles.* Profile observations are first checked for duplicate depths and strictly increasing depths. Reported levels that fail these tests are flagged and not used in the following profile quality control procedures.

*4.6.1 Instrumentation Error Checks.* Special instrument specific error tests are applied to profile observations to identify errors that have unique profile signatures. These errors include temperature inversions at the bottom of the profile, spikes in the temperature profile, and positive temperature gradients (warm bulge) in the mixed layer. The instrumentation error checks are applied iteratively until all errors are found, since a profile may have one or more of these types

of errors. Reported temperature-depth levels that contain instrumentation errors are flagged and not used in the next iteration of the instrumentation error checks. One difficulty with the current suite of profile instrumentation error checks is that the tests are designed to detect errors specific to expendable bathythermographs (XBT). Other profile data types, such as Argo floats, gliders and CTD probes, are likely to have failure modes that are different from a XBT. Automated quality control tests need to be developed to detect instrumentation errors in these data types as more experience is gained with their assimilation.

*4.6.2 Static Stability.* A static stability test is performed to detect profiles with density inversions. The reported *in situ* temperature and depth data pairs are converted to potential temperature and pressure, and potential density is computed at each pressure level using observed or derived salinity values. Salinity observations are generated for profiles that report only temperature. Salinity is computed from observed temperature values using bi-monthly climatological temperature-to-salinity regression models that have been computed on a global  $0.25^\circ$  resolution grid. The potential density profile is examined for inversions (higher density shallower than lower density), and observed temperature and salinity profile levels with inversions that exceed a minimum specified inversion threshold of  $0.025 \text{ kg}\cdot\text{m}^{-3}$  are flagged. For profiles with derived salinities, static instabilities are corrected by iteratively adjusting the derived salinity until the resulting profile is neutrally buoyant. Salinity is removed from the base of the mixed layer upward and added from that depth downward in the adjustment. The salinity correction algorithm is not applied to density inversions for profiles that observe both temperature and salinity levels, since it is difficult to determine *a priori* if the cause of the density inversion is due to the reported temperature or the reported salinity value. In that case, profile levels with density inversions are simply flagged.

*4.6.3 Vertical Gradient Checks.* A global climatology of vertical mean temperature differences and standard deviations about these means has been computed from the historical profile archive. The climatology is used to test observed vertical temperature gradients for outliers. First, the climate temperature differences and variability are interpolated to the observation location and sampling time. Second, the vertical temperature differences are converted to vertical temperature gradients and interpolated to the observed profile levels. Observed vertical temperature gradients are computed, and the difference between the observed and the expected mean vertical gradient from the climatology is standardized by the expected gradient variability,

$$z = (\Delta T_o \cdot m^{-1} - \Delta T_c \cdot m^{-1}) / \sigma \quad (4.5)$$

where  $\Delta T_o \cdot m^{-1}$  is the observed vertical gradient,  $\Delta T_c \cdot m^{-1}$  is the climate mean vertical gradient,  $\sigma$  is the variability about that mean, and  $z$  is the standardized vertical gradient variate. If the observed profile gradient exceeds  $0.2^\circ\text{C}\cdot\text{m}^{-1}$  and  $|z| > 4$ , then the profile level is flagged. Experience has shown that the vertical gradient test based on climate statistics tends to spuriously flag as erroneous profile levels associated with a strong thermocline. This problem is particularly acute in the tropics. Truly erroneous profile vertical gradients are often associated with bad temperature or salinity observations, which are detected in the spike test and background field checks described previously. Hence, at the present time, quality control flags set by the vertical gradient check are flagged and only used for informational purposes (see discussion in section 5).

*4.6.4 Profile Shape Comparisons.* Observed profiles are compared to profiles extracted from the various background fields using a profile shape quality control procedure. The procedure has the advantage of taking an overview of the entire profile. Profile levels that have previously been determined to be unreliable based on other profile quality control data checks are excluded in the profile shape quality control procedure. The shape quality control procedure computes an integrated observed-minus-predicted statistic that takes into account level thicknesses. The test statistic is calculated as,

$$\eta = \sum_k ((O_k - P_k) / \sigma_k) \cdot (z_{k+1} - z_{k-1}) / \sum_k (z_{k+1} - z_{k-1}) \quad (4.6)$$

where  $O_k$  is the observed value at level  $k$ ,  $P_k$  is the prediction (background) value at level  $k$ ,  $\sigma_k$  is the prediction error standard deviation at level  $k$ , and  $z_k$  is the depth of level  $k$ . The probability of  $\eta$  being greater than zero is computed assuming a normal probability distribution function. The shape comparison statistic is analogous to a goodness-of-fit test of two cumulative distribution functions. It identifies observed profiles with large errors relative to the background profiles. Profiles that have large temperature or salinity differences over narrow depth ranges, such as dissimilar mixed layer depths, will be considered similar. Observed profile shape must be consistent with forecast and climate background profiles in order for the profile to be accepted into the analysis.

*4.6.5 Glider Dives.* Ocean gliders are autonomous platforms which fly in a saw-tooth-sampling pattern in the upper ocean by changing their buoyancy. Depending upon configuration, gliders sample profiles of pressure, temperature, and conductivity. The gliders surface at regular intervals to transmit their observations using the Iridium network. Gliders provide both downward and upward profiles of temperature and salinity, with glider position and time varying with depth during the dive. Quality control of glider data is similar to that of single profile data, other than relaxation of the strictly increasing depths check. However, several glider specific tests are performed that are, in most cases, functions of the vertical velocity of the glider. These tests are applied to gliders with a non-pumped CTD, where the speed of flow through the conductivity cell depends upon the speed of the glider, making the thermal inertial correction speed-dependent.

*4.7 Altimeter Sea Surface Height.* The along-track altimeter data undergo an extensive series of pre-processing steps to prepare the data for use in the assimilation. The measured sea surface height (SSH) is corrected for geophysical effects (wet and dry troposphere, ionosphere, inverted barometer, and winds), and the tidal signal is removed. The corrected SSH from each satellite altimeter mission are then intercalibrated with a global crossover adjustment using Topex/Poseidon data as the reference. Next, the data are resampled every 7 km (1 sec intervals) along the tracks. A mean SSH is removed from the individual SSH measurements producing sea surface height anomalies (SSHA). The mean SSH contains both the unknown geoid signal and the mean dynamic topography over the averaging period. For most satellite missions a mean SSH calculated over a 7-year period is used, although the averaging period continues to be extended in time as the altimeter satellite missions continue. These altimeter pre-processing steps are typically performed by the data provider.

Altimeter SSHA observations are of lower accuracy or are not interpretable near the coasts due to inaccurate tidal corrections and incorrect removal of atmospheric wind and pressure effects at the sea surface in shallow water. The coastal region for altimeter data assimilation is often

defined as everywhere shallower than 400 m depth. Altimeter observations also have significant along-track correlated errors that must be taken into account in the assimilation. The along-track altimeter SSHA data are very noisy at the full 7 km resolution. Accordingly, altimeter SSHA data are smoothed along-track using a median or Lanczos filter to reduce the measurement noise. In addition, the altimeter data are often sub-sampled to remove redundant observations. Finally, altimeter SSHA measurements are scaled using a hyperbolic tangent operator by local dynamic height variability limits that have been computed from the historical profile archive. This operation essentially prevents spurious altimeter SSHA outliers from being used in the assimilation.

A final issue with altimeter SSHA measurements is the fact that different versions of the data are reported in both near real-time and in delayed mode. The real-time SSHA observations are computed using less precise, predicted orbits rather than the more precise, observed orbits, which are used in the delayed mode data. The more precise orbit information is not available for several days after real-time. Although less precise, real-time SSHA observations still have significant value in the analysis. However, when the more precise delayed mode SSHA observations are available, the corresponding real-time SSHA data should be identified and replaced by the delayed mode data. This procedure ensures that the higher quality, delayed mode SSH observations are incorporated into the altimeter SSHA data archive and used in all hindcast studies. Satellite altimeter SSHA observations are a critical data source in GODAE assimilation systems and access to the most complete, highest quality data is essential.

*4.8 Altimeter Significant Wave Height.* Comparisons with buoy data show that altimeter SWH estimates are in agreement with the *in situ* data, with standard deviations of differences of the order of 0.30 m, but the satellite data tends to slightly overestimate low SWH and to underestimate high SWH. The altimeter SWH data thus needs to be bias corrected before it is used in the assimilation. These bias corrections are generally linear and are derived from altimeter/buoy matchups and correspond to corrections of only a few percent of SWH. Altimeter SWH data can also be contaminated by sea ice or land. Elimination of these data requires a contemporaneous sea ice concentration field computed from an analysis of SSM/I and SSMIS sea ice retrievals. The land mask needs to resolve the along-track 7-km footprint of the altimeter SWH data. A suitable source for a land surface mask at this resolution is available at the USGS from the Global Land 1-km AVHRR Project.

**5. Quality Control Decision-Making Algorithms.** The quality control outcomes of the various external data checks described previously are combined in a decision-making algorithm. The outcome of the decision-making algorithm is the overall indication of observation quality, which is used to select data for the assimilation. The decision-making algorithm is applied to each observed reporting level and, in the case of profile (and glider) observations, to the entire profile in the shape comparison test. Thus, for profile observations there are two indicators of data quality: one indicator for the overall profile shape, and a second indicator for each profile level. It is important to take into consideration results from all of the external data checks before the final quality decision is made. For example, an observation could fail the climate background check while at the same time pass the forecast background check. Quality control decision-making algorithms, therefore, are necessarily complex and must combine outcomes from the

different external test results appropriately. It helps if the external test outcomes are of the same form, such as probabilities or standard normal deviates.

A quality control decision-making algorithm in use at the U.S. Navy oceanographic centers is described here. The quality control outcomes from the various external data checks are in the form of probabilities of error. The majority of these probabilities are calculated according to Eq. (4.1), assuming a normal probability density function, but probabilities are also calculated assuming chi-square or uniform distribution functions. Given a set of error probabilities the decision-making algorithm is summarized as follows:

$$\begin{aligned}
 \mathbf{p}_b &= \min(\mathbf{p}_g, \mathbf{p}_r) \\
 \mathbf{p}_d &= \min(\mathbf{p}_c, \mathbf{p}_x) \\
 \mathbf{p}_b < \tau_f, \mathbf{p}_o &= \mathbf{p}_b \\
 \mathbf{p}_b > \tau_f, \mathbf{p}_o &= \min(\mathbf{p}_b, \mathbf{p}_d)
 \end{aligned}
 \tag{5.1}$$

where  $\mathbf{p}_b$  is the composite background error probability,  $\mathbf{p}_d$  is the composite data-derived error probability,  $\mathbf{p}_g$  and  $\mathbf{p}_r$  are the global and regional forecast background error probabilities,  $\mathbf{p}_c$  and  $\mathbf{p}_x$  are the climate and cross validation error probabilities,  $\tau_f$  is the forecast error threshold probability, and  $\mathbf{p}_o$  is the overall probability the observation contains a random error. Forecast error probability threshold is typically set to 0.99 (3 standard deviations).

The algorithm first determines if the observation is consistent with the model background fields by taking the minimum error probability of the global and regional forecasts. If the minimum background error probability is less than the prescribed forecast error tolerance limit, then the algorithm returns it as the overall probability of error for the observation. However, if the minimum model background error probability exceeds the forecast error threshold, then it is compared against the data-derived error defined as the minimum of the cross validation and climatology error probabilities. The overall observation error probability is returned as the minimum of the composite background and composite data-derived errors. In this way, cross validation and climate backgrounds determine data quality only if the observation is not consistent with the forecast. Experience has shown that requiring observations to always be consistent with climate backgrounds results in spurious rejection of valid observations during extreme events.

Once the overall probability of error for an observation has been determined, output from the various specific observing system quality control tests are simply added to the error probability as integer-valued flags. The flag value for each quality control test is unique. The quality control test flags have three levels of severity, ranging from information-only (<100), to cautionary ( $\geq 100$ ), and fatal ( $\geq 1000$ ). Observations with fatal errors are not used in the analysis. Information-only flagged observations are routinely used in the analysis, but the use of cautionary flagged observations is under user control via analysis namelist options. The ultimate decision to accept an observation into the analysis, however, is always based on the underlying error probability value obtained from the decision-making algorithm. If quality control flags have been appended, the underlying probability of error can always be recovered from the summation using some simple modular arithmetic.

**6. Internal Data Checks.** Internal checks are those quality control procedures performed by the analysis system itself. These data consistency checks are best done within the assimilation algorithm since it requires detailed knowledge of the background and observation error covariances, which are available only when the assimilation is being performed. The internal data checks are the last defense of the assimilation algorithm against bad observations. Data that contain gross and random errors have hopefully been removed prior to the assimilation in the preliminary and external data checks. The purpose of the internal data checks is to decide whether any marginal observations remaining in the assimilation data set are acceptable or unacceptable. The need for quality control at this stage of the analysis/forecast system cannot be over emphasized. Any assimilation system based on the assumption of normality, no matter how sophisticated, is vulnerable to bad observations that do not fit a normal distribution. Further, since many GODAE forecasting systems use a sequential analysis-forecast cycle, it is difficult to remove the propagation of error through the forecast period that occurs when erroneous data have been assimilated. Once this happens the only option is to blacklist the bad observations and back-up and rerun the analysis-forecast cycle. This remedy will cause a delay in the production of the forecast, which can be a serious problem in operations if the forecast products are time critical.

The internal consistency checks are quite different from the cross validation procedure described in section 4. In particular, each observation is compared with the entire set of observations used in the assimilation, not just nearby observations. A metric is devised to test whether the observation innovations are likely or unlikely with respect to the specified background and observation error statistics. Once the decision to reject an observation is made it is necessary to intervene in the assimilation process to ensure that the rejected observation has no effect on the analysis. Typically, internal data checks are performed in variational analysis schemes, where the solution is obtained using iterative methods that can be interrupted and started up again.

In an observation based analysis system the analyzed increments (or correction vector) are computed according to,

$$(\mathbf{x}_a - \mathbf{x}_b) = \mathbf{B}\mathbf{H}^T (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} [\mathbf{y} - \mathbf{H}(\mathbf{x}_b)] \quad (6.1)$$

where  $\mathbf{x}_a$  is the analysis and  $\mathbf{x}_b$  is the forecast model background. In the right hand side of Eq. (6.1),  $\mathbf{B}$  is the background error covariance,  $\mathbf{H}$  is the forward operator,  $\mathbf{R}$  is the observation error covariance, and  $\mathbf{y}$  is the observation vector. The observation vector contains all of the synoptic temperature, salinity and velocity observations that are within the geographic and time domains of the forecast model grid and update cycle. When the analysis variable and the model prognostic variable are the same type, the forward operator  $\mathbf{H}$  is simply spatial interpolation of the forecast model grid to the observation location performed in three dimensions. Thus,  $\mathbf{H}\mathbf{B}\mathbf{H}^T$  is approximated directly by the background error correlation between observation locations, and  $\mathbf{B}\mathbf{H}^T$  directly by the error correlation between observation and grid locations. The quantity  $[\mathbf{y} - \mathbf{H}(\mathbf{x}_b)]$  is referred to as the innovation vector (model-data misfits at the observation locations).

Denote  $\mathbf{A} = \mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}$  as the observation symmetric positive definite matrix of Eq. (6.1). Define  $\hat{\mathbf{A}} = \text{diag}(\mathbf{A})$ . Then, define the observation vector  $\hat{\mathbf{d}} = \hat{\mathbf{A}}^{-1/2} [\mathbf{y} - \mathbf{H}(\mathbf{x}_b)]$ . The elements of  $\hat{\mathbf{d}}$  are the normalized innovations and should be distributed (over many realizations) in a normal distribution with a standard deviation equal to 1.0 if the background and observation error covariances have been specified correctly. Assuming this to be the case, tolerance limits

are defined to remove unacceptable observations. Since  $\mathbf{B}$  and  $\mathbf{R}$  are never perfectly known, it is best to use a higher tolerance value than a lower one. The test statistic is designed to reject an observation if its element of  $\mathbf{d}^\wedge$  is larger than the specified tolerance limit.

Once the clearly unacceptable observations have been removed, a final internal check is done which compares each observation with every other observation (that has passed the tolerance limit check). The procedure is a logical extension of the tolerance limit check described above. Define the vector  $\mathbf{d}^* = \mathbf{A}^{-1/2}[\mathbf{y} - \mathbf{H}(\mathbf{x}_b)]$ . The elements of  $\mathbf{d}^*$  are like those of  $\mathbf{d}^\wedge$ , dimensionless quantities normally distributed. However, because  $\mathbf{d}^*$  involves the full covariance matrix  $\mathbf{A}$ , it includes correlations between all of the observations. By comparing the vectors  $\mathbf{d}^\wedge$  and  $\mathbf{d}^*$  it can be shown which observations are inconsistent with other observations and can therefore be rejected. The  $\mathbf{d}^*$  metric should increase (decrease) with respect to  $\mathbf{d}^\wedge$  when that observation is inconsistent (consistent) with other observations, as specified by the background and observation error statistics. For example, assume there are 2 observations with high calculated tolerance limits that are also highly correlated. In this case, both of the observations would have been rejected individually on the basis of the tolerance limit check, but because of the high correlation between the observations they are determined to be self-consistent and retained in the analysis.

There are difficulties applying this internal data check in practice since it requires calculating the entire  $\mathbf{A}^{-1/2}$  matrix, which is prohibitive for very large problems. Fortunately, there are some good approximations to this calculation that can be used. However, other implementation issues remain. To reject an observation a large constant is added to the appropriate diagonal element of the  $\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}$  matrix. This modifies the matrix in such a way as to effectively prevent the rejected observation from affecting the analysis. If this operation is done during the descent iteration then the modified matrix is no longer consistent with the other vectors that have been evolving as part of the conjugate gradient solution. The descent can be restarted (very expensive) or the conjugate gradient vectors can be suitably altered to allow the descent to continue. In either case the tolerance limit and internal consistency checks can be applied multiple times during the descent as more and more observations are brought into the solution.

Modifications to this procedure can be made for extreme events when the specified background error statistics are likely to be incorrect. The error statistics in the assimilation are produced by averaging time series of innovations and forecast differences and reflect average, rather than extreme, conditions over the domain. When changes are occurring in the ocean (such as an eddy shedding event) the background error statistics are likely to be larger than normal. In this case, a tolerance limit specified too low could reject good (and very important) data. One option for dealing with this is make one pass through the tolerance limit check and compute the mode of the  $\mathbf{d}^\wedge$  values over some limited subareas of the analysis domain. The mode is a better statistic here because it is less susceptible to outliers than the mean. If the subarea mode is much greater than one, then it can be concluded that there are serious discrepancies between the observations and the background in that area. To avoid spuriously rejecting good data the tolerance limit should be increased in the subarea.

**7. Adjoint Sensitivities.** Adjoint-based observation sensitivity, initially developed in Numerical Weather Prediction as an observation-targeting tool, provides a feasible (all at once) approach to estimating observation impact for a large variety of datasets and individual observations.

Observation impact is calculated in a two-step process that involves the adjoint of the forecast model and the adjoint of the assimilation system. First, a cost function ( $\mathbf{J}$ ) is defined that is a scalar measure of some aspect of the forecast error. The forecast model adjoint is used to calculate the gradient of the cost function with respect to the forecast initial conditions ( $\partial\mathbf{J}/\partial\mathbf{x}_a$ ). The second step is to extend the initial condition sensitivity gradient from model space to observation space using the adjoint of the assimilation procedure ( $\partial\mathbf{J}/\partial\mathbf{y}=\mathbf{K}^T\partial\mathbf{J}/\partial\mathbf{x}_a$ ), where  $\mathbf{K}=\mathbf{B}\mathbf{H}^T[\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R}]^{-1}$  is the Kalman gain matrix of Eq. 6.1. The adjoint of  $\mathbf{K}$  is given by  $\mathbf{K}^T=[\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R}]^{-1}\mathbf{H}\mathbf{B}$ . The only difference between the forward and adjoint analysis codes is in the post-multiplication of going from the solution in observation space to grid space. The solver ( $\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R}$ ) is symmetric or self-adjoint and operates the same way in the forward and adjoint directions. Given an analysis sensitivity vector, observation impact is obtained as a scalar product of the observed model-data differences and the sensitivity of the forecast error to those differences. Observations will have the largest impact on reducing forecast error when the observation influences the initial conditions in a dynamically sensitive area. It is not necessary for the observation to produce a large change (i.e., innovation) to the initial conditions for it to have a large forecast impact.

If the assimilation of an observation has made the forecast issued from the analyzed state more accurate than a forecast valid at the same time but issued from a prior state, then the observation is considered to have a positive impact. All assimilated observations are expected to have beneficial impacts on correcting the initial conditions and thereby improving the forecast issued from the analysis. However, if consistent non-beneficial impacts are found for a particular data type or observing system, then that may indicate data quality control issues, such as subtle instrument drift or calibration problems that otherwise are difficult to assess when considering the data in isolation. Thus, the adjoint-based data impact procedure is an effective tool to provide quantitative diagnostics for monitoring the outcomes from all of the various stages of ocean data quality control.

**8. Summary.** Effective ocean data quality control is a difficult problem. Observations are imperfect and prone to error. Data with errors that are not described by the assimilation system through the error covariance matrices need to be eliminated prior to the analysis. Effective quality control, therefore, requires a set of pre-established, standardized test procedures, with results of the procedures clearly associated with the data values. Effectiveness in turn depends on the reliability of the standard(s) and on the choices made for measuring goodness of fit.

The need for observation quality control depends on the use being made of the observations. Users of quality controlled data sets have a wide range of views on the most appropriate standards and on the appropriate "tightness of fit" demanded by the quality control procedures (too tight increases the chance of erroneously rejecting anomalous features; too loose increases the chance of accepting bad data). Indicators of data quality, therefore, must be useful for determining if the quality controlled observations are appropriate for a particular purpose. In this paper, observation quality control is performed as a prelude to assimilation of the observations in an ocean forecast system. In many regards, the best ocean data quality scheme is that which leads to the best ocean forecast.

It is surprisingly difficult to demonstrate consistent impact from the quality control of individual observations in an analysis/forecast system. It is likely, however, that quality control is more important in data monitoring: collection of statistics on the performance of the observing

systems; detection of observing systems that are not performing as expected; and feedback to the data providers so that deficiencies are corrected. An integrated, end-to-end quality control system, therefore, must ensure that results of the quality control procedures are recorded for independent analysis and later use. If the quality control is carried out well, then it can reduce the duplication of effort among the users of ocean data – value added is not lost or misinterpreted. At a minimum, a comprehensive database of raw and processed observed values, independent estimates of the same quantities, and quality control outcomes is needed. The database would be used to look for “unexpected” behavior in observing systems, and allow users and operators of quality control systems to identify systematic problems in order to get errors in the data collection or data transmission corrected. At present, there are few agreed-upon standards for real-time ocean data quality control and very few cases where the procedures and results from the oceanographic centers have been compared. As the GODAE operational oceanographic community continues to develop a range of complex ocean analysis and prediction systems, it is important that procedures be developed for routinely assessing the effectiveness of ocean data quality control and for routinely exchanging statistics from the quality control processes at the operational centers.

**9. Conclusions.** The fully automated ocean data quality control procedures described in this paper are limited to observation data types that are routinely assimilated in ocean forecast models. New ocean observing systems continue to be deployed and new failure modes of existing observing systems continue to be identified. Examples of new observing systems include HF coastal radars that measure surface currents, microwave measurements of sea surface salinity from space, Lagrangian velocity vectors from drifters, and remotely sensed ocean color. Examples of new instrument failure modes are the pressure and salinity sensor issues associated with the long-term, autonomous, deployments of the Argo profiling floats. New observation error models will need to be developed for the automated quality control of new data types, and existing error models will need to be updated to detect, and hopefully correct, new instrument failure modes. The validity of existing and new automated quality control procedures need to continually be confirmed by formal statistical tests and by examining differences between automated and delayed-mode quality control outcomes on the same observation. The automated quality control system can be considered to have performed well if decisions made on observations in real-time are consistent with decisions made to modify or reject the same observation in delayed mode, where more rigorous scientific and expert manual intervention quality control methods are possible. Delayed mode quality control outcomes of the Argo profiling float array are readily available and can be used in this evaluation. It is safe to say real-time ocean data quality control is far from being a static process.

**References:**