In situ SST Quality Monitor (*i*Quam)

FENG XU

NOAA/Center for Satellite Applications and Research (STAR), and Global Science and Technology, Inc., College Park, Maryland

ALEXANDER IGNATOV

NOAA/Center for Satellite Applications and Research (STAR), College Park, Maryland

(Manuscript received 24 May 2013, in final form 23 July 2013)

ABSTRACT

The quality of in situ sea surface temperatures (SSTs) is critical for calibration and validation of satellite SSTs. In situ SSTs come from different countries, agencies, and platforms. As a result, their quality is often suboptimal, nonuniform, and measurement-type specific. This paper describes a system developed at the National Oceanic and Atmospheric Administration (NOAA), the in situ SST Quality Monitor (*i*Quam; www.star.nesdis.noaa.gov/sod/sst/iquam/). It performs three major functions with the Global Telecommunication System (GTS) data: 1) quality controls (QC) in situ SSTs, using Bayesian reference and buddy checks similar to those adopted in the Met Office, in addition to providing basic screenings, such as duplicate removal, plausibility, platform track, and SST spike checks; 2) monitors quality-controlled SSTs online, in near–real time; and 3) serves reformatted GTS SST data to NOAA and external users with quality flags appended. Currently, *i*Quam's web page displays global monthly maps of measurement locations stratified by four in situ platform types (drifters, ships, and tropical and coastal moorings) as well as their corresponding "in situ minus reference" SST statistics. Time series of all corresponding SST and QC statistics are also trended. The web page user can also monitor individual in situ platforms. The current status of *i*Quam and ongoing improvements are discussed.

1. Introduction

In situ observations of sea surface temperature (SST) are critical for calibration and validation (Cal/Val) of satellite retrievals. These applications require a highly accurate standard. However, the quality of in situ data is often suboptimal. These data vary in space and time, and across different countries, agencies, platforms, sensors, and manufacturers (e.g., Bitterman and Hansen 1993; Hansen and Poulain 1996; Brasnett 1997, 2008; Emery et al. 2001a,b; Kent and Berry 2005; Rayner et al. 2003, 2006; Kent and Challenor 2006; Kent and Kaplan 2006; Kent and Taylor 2006; Gronell and Wijffels 2008; Kent et al. 2010; Ingleby 2010; Kent and Ingleby 2010; Reverdin et al. 2010; Kennedy et al. 2011b, 2012; Castro et al. 2012). At the same time, even if a small fraction of outliers is included in Cal/Val, it may render its results unusable (e.g., Xu and Ignatov 2010, and references therein). On the other hand, rejecting some unexplained but correct data could miss important climate and diurnal warming signals, and leave voids in some geographic areas (Lorenc and Hammon 1988).

At the National Oceanic and Atmospheric Administration (NOAA) and other satellite SST-producing centers, including the U.S. Naval Oceanographic Office (NAVOCEANO), the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Ocean and Sea Ice Satellite Application Facility (OSI SAF), and the National Aeronautics and Space Administration (NASA)–University of Miami SST Team, in situ data provided by the National Centers for Environmental Prediction (NCEP) Global Telecommunication System (GTS) are employed for near–real time (NRT) Cal/Val applications. GTS data available from NCEP in NRT from January 1991 to present are not quality controlled (QC), and an efficient QC is needed before they can be used in satellite Cal/Val (e.g., Xu and Ignatov 2010, and references therein). This need has long been recognized, and QC of in situ data is

Corresponding author address: Alexander Ignatov, NOAA/STAR, NCWCP, 5830 University Research Court, Room 3750, College Park, MD 20740.
E-mail: alex.ignatov@noaa.gov

DOI: 10.1175/JTECH-D-13-00121.1

always performed in satellite Cal/Val efforts. However, the practices adopted in the remote sensing community remain largely ad hoc, overly simplistic, and nonuniform. For instance, outlier data points are often identified by merely applying a constant threshold to the deviation of the in situ SST from a reference (climatological or analysis) SST field (e.g. Kilpatrick et al. 2001; Francois et al. 2002; Brisson et al. 2002). Some authors specify the global thresholds from the data using ± 3 standard deviation (SD) of "in situ minus reference" SST without removing the corresponding global mean (e.g. O'Carroll et al. 2006; Merchant et al. 2008). In any case, these QC methods remain far inferior to the more sophisticated, systematic, and well-developed procedures employed in the meteorological and oceanographic communities (e.g. Slutz et al. 1985; Lorenc and Hammon 1988; Woodruff et al. 1998; Rayner et al. 2003, 2006; Worley et al. 2005; Kent and Taylor 2006; Ingleby and Huddleston 2007; Thomas et al. 2008).

At the same time, satellite Cal/Val is very demanding on the quality of in situ data and requires a flexible and scalable QC depending on the specific Cal/Val task. Presently, NOAA is responsible for the maintenance and development of SST products from the current operational polar [from NOAA and Meteorological Operation (METOP) Advanced Very High Resolution Radiometers (AVHRRs)] and geostationary [from Geostationary Operational Environmental Satellite (GOES), Meteosat, and Multifunctional Transport Satellite (MTSAT)] as well as the new generation Joint Polar Satellite System (JPSS) and Geostationary Operational Environmental Satellite R-Series (GOES-R) satellites. A NRT in situ SST Quality Monitor (*i*Quam; www.star.nesdis.noaa.gov/sod/sst/iquam/) was developed to support these products and applications in a consolidated and cohesive way, and as a NOAA contribution toward a community effort coordinated by the international Group for High Resolution SST (GHRSST; Donlon et al. 2007).

The following are three major functionalities of the *i*Quam:

- Implementation of advanced, flexible, and unified community consensus QC for in situ SSTs, maximally consistent with the procedures that are adopted in wider meteorological and oceanographic communities;
- Web-based NRT quality monitoring (QM) of quality-controlled in situ SSTs relative to reference SST (currently, the daily Optimal Interpolation version 2 (OI v2) product; Reynolds et al. 2007) stratified by platform types (drifters, tropical and coastal moored buoys, and ships) and/or by platform identification (ID) numbers;
- Serving quality-controlled in situ SST data with quality flags (QFs) appended (but not applied) to NOAA and wider external SST users, in support of various tasks and applications (primarily, satellite Cal/Val).

The QC algorithm in *i*Quam includes, in addition to basic screenings (such as the duplicate removal and plausibility, platform tracking, and SST spike checks), more sophisticated reference and cross-platform checks. The two latter checks follow the Bayesian approaches proposed by Lorenc and Hammon (1988) and Ingleby and Huddleston (2007), and adopted for QC of in situ data in the Met Office. In *i*Quam, these approaches are applied with only minor modifications.

The QM component of *i*Quam picks up quality-controlled in situ data, calculates their monthly statistical summaries, which are stratified by platform types and individual ID numbers, and displays them on the web (at www.star.nesdis.noaa.gov/sod/sst/iquam/). Global maps and histograms are available along with their summary Gaussian statistics (both conventional and robust) and fractions of in situ data that failed various QC checks. Long-term time series of monthly statistics that include number of platforms and observations, all Gaussian parameters, and QC error rates can be viewed. A sortable table of all individual platforms is also provided with one-click-of-a-button access to precalculated graphs showing the platform track, SST time series, and performance history.

Finally, quality-controlled in situ SST data, are served online in Hierarchical Data Format (HDF). Historical data are organized into monthly files. The current month file is updated every 12 h, with a 2-h latency following GTS data availability, and is finalized on the fifth day of the following month. For each observation, all individual QFs are provided. A summary QF is also set, using the recommended *i*Quam logic. Users always have freedom to define their own summary QF using a different logic with individual QFs.

QC algorithms and configurations are described in section 2. Web-based QM and statistics are introduced in section 3. Section 4 describes the *i*Quam data and defines the QFs. Section 5 concludes the paper and discusses ongoing work toward *i*Quam version 2.

2. Quality control algorithm

a. Principles

The basic principle of the QC is to check the in situ data for self-consistency and for cross consistency with other data. Commonly used QC checks were summarized by Woodruff (2008), and are based on the condition and the method. Those checks can be categorized into five major groups based on the physical principles they rely on:

TABLE 1. The QC checks employed in *i*Quam.

Category	Check	Type of error handled	Physical basis
Preprocessing	DR	Duplicates arise from multiple transmission or dataset merging	Identical spatial and temporal coordinates and ID numbers
Plausibility/geolocation	GC	Unreasonable field values including geolocation	Range of single fields and relationships among them
Internal consistency	TC	Points falling out of track	Travel speed exceeds limit
	SC	Discontinuities in SST time series along track	SST gradient exceeds limit
External consistency	RC	Measurements deviating from reference	Bayesian approach (reference SST: daily OI SST v2)
Mutual consistency	XC	Mutual verification with nearby measurements (buddies check)	Bayesian approach based on space/time correlation of SST field (correlation model: two-scale SOAR; Martin et al. 2002)

- Prescreening—Resolves data-specific problems (e.g., duplicate removal, and data cleaning and/or reorganizing).
- Plausibility/geolocation—Assures that each individual field and relationships between different fields are realistic (e.g., field range, geolocation, ID number versus platform-type checks).
- Internal consistency—Checks different measurements from the same platform for internal consistency (e.g., platform track and SST spike checks).
- External consistency—Checks individual measurements for consistency with the reference (first guess) SST field. Termed the reference check in this paper, it is also sometimes referred to as background check (e.g., Lorenc and Hammon 1988).
- Mutual consistency—Checks for consistency between nearby measurements from different platforms. This check, termed cross-platform check in this paper, is also often referred to as the buddy check (e.g., Lorenc and Hammon 1988).

A summary of *i*Quam QC is presented in Table 1. All checks are performed independently, meaning that no check relies on the results of other checks. The two exceptions are the cross-platform check, which only uses data points that pass all other checks as ''buddies'' (cf. section 2c), and the duplicate removal, which uses the result of the reference check. No data are excluded in *i*Quam based on QC, but rather all data are retained and QFs are appended.

b. *Binary checks*

1) DUPLICATE REMOVAL

Duplicates arise from multiple receptions of the same report via different paths, or from merging different datasets. The algorithm checks the differences between any two neighboring records originating from the same platform. Only latitude, longitude, and time are checked. Tolerances are set as the corresponding digitization precision of each field—for example, 0.01° for latitude and longitude, and 1 min for time.

For a group of duplicates, the one with the best quality will be kept. If quality information is not available and all the duplicates have SSTs within 0.1°C tolerance, then the first in the sequence is kept and the rest are dropped; otherwise, all are dropped.

In *i*Quam, duplicate removal is preceded by the reference check described in section 2c below, which compares each individual record with a reference field and is set for all duplicates. Quality information from the reference check is then used in the duplicate removal to select the record with the best quality.

2) PLAUSIBILITY/GEOLOCATION CHECK

Geolocation check evaluates whether the location of a platform is plausible. For instance, SST measurements should not be reported over land, and buoys are supposed to be located in the regions indicated by their corresponding area codes, which are embedded in their World Meteorological Organization (WMO) ID numbers. This check may also remove those reports found too close to coastlines, depending upon the resolution and the accuracy of the water mask employed. Currently in *i*Quam, the University of Maryland's (UMD) 1-km land cover classification is used (Hansen et al. 2000). Note that near-coastal in situ SSTs are highly variable in space and time because of shallow waters and high dynamics, and should be avoided in satellite Cal/Val in any case. Plausibility checks also include valid range checks for each fields, for example, latitude within [−90°, 90°], longitude within [−180°, 180°], and SST within [−2°, 35°C].

3) Platform track check

This check verifies that consecutive locations of a platform (identified by its ID number) are consistent with the respective time stamps, assuming that the platform cannot move faster than a predefined maximum moving speed. Significant errors in time and latitude–longitude will cause deviations from this expected pattern. At first, a least-required speed is estimated, assuming that the platform had traveled between the locations of any two reports through the shortest path (great-circle distance). Next, the report with the most speed violations is identified and excluded. The operation is iterated until no violation is detected.

The maximum speed is chosen as $60\,\mathrm{km\,h^{-1}}$ for ships and $15\,\mathrm{km\,h^{-1}}$ for drifters. These values have been estimated from the global histograms of the least-required speed traveling between a pair of locations. It should be noted that digitization error of time, latitude, or longitude may raise false alarms when the time and location differences are very small. Therefore, the condition of track check is written as

$$\frac{\max(\Delta_d - \delta_d, 0)}{\Delta_t + \delta_t} > \upsilon_{\max}. \tag{1}$$

Here, Δ_d and Δ_t denote distance and time differences, respectively; and δ_d and δ_t correspond to their errors caused by digitization; $\upsilon_{\max}$ is the maximum travel speed. If the condition is met, the record is labeled as erroneous. For moored buoys, the procedure can be simplified. If a report is located far away from the majority of reports of the same mooring, then it is regarded as erroneous. The maximum allowed distance is chosen as 100 km to tolerate reasonable drifting and latitude–longitude error. Note that platforms with invalid or group ID numbers [cf. ID check in section 2b(5)] are not subject to track check.

Figure 1 shows several examples of abnormal reports identified by track check. In Fig. 1a, one observation apparently falls off the ship track because of an error caused by a swapped sign in the latitude field. In this case, it would be difficult to detect such an error merely by comparing to the reference SST, which could be close for the similar latitude zones in the north and in the south. Another example for a drifting buoy is shown in Fig. 1b. Such an error may be even more difficult to detect by comparing to the reference SST, which may not change significantly within 2° latitude or longitude. Figure 1c shows an example of a mooring buoy, which reported two observations located far off the main body of the cluster.

4) SST spike check

For a continuously reporting platform, an erroneous report may appear as an SST spike (or step) along its track or in the time series because of sensor malfunction or occasional maintenance operation. Spike check employs the same logic and algorithm as the track check, except the maximum SST gradient in space and time is checked, instead of travel speed. The maximum SST gradient is chosen as $g_d = 0.5\,\mathrm{K\,km^{-1}}$ in space and $g_t = 1.0\,\mathrm{K\,h^{-1}}$ in time. To accommodate normal fluctuation between successive records caused by, for example, instrument noise, an exempt threshold δ_T is set so that SST differences $<\delta_T$ are exempt from spike check. The condition for spike check is written as

$$\Delta_T > \max(\delta_T, \Delta_d g_d, \Delta_t g_t). \tag{2}$$

Here, Δ_T, Δ_d, and Δ_t are SST, and space and time differences, respectively; and g_d, g_t are the corresponding maximum SST gradients. If the condition is met, then the record is labeled as erroneous. Note that the exempt threshold, δ_T, is set specifically for each type of platform based on its noise level. Currently in *i*Quam, $\delta_T = 2.0\,\mathrm{K}$ for ships, 1.0 K for tropical moored and drifting buoys, and 1.6 K for coastal moored buoys.

5) ID check

A valid platform ID is critical because several QC checks are applied on an individual platform basis, for example, track check and spike check. Hence, an ID check is performed to determine whether the ID field of a measurement is valid. If not, it will not be subject to individual platform (track and spike) checks and labeled as such in the final quality flags.

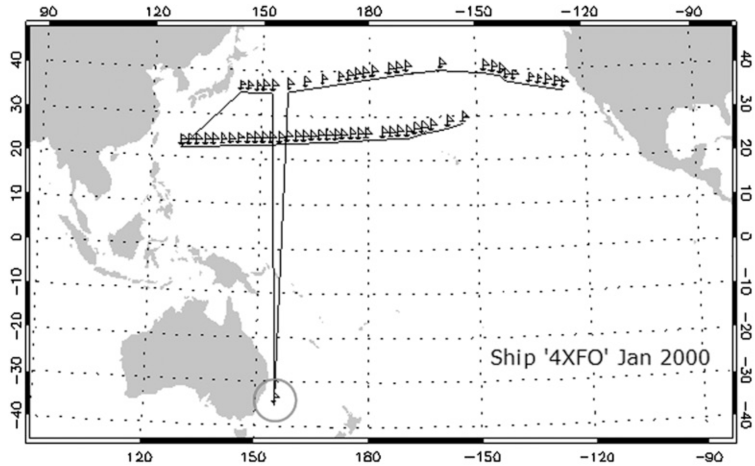
The most common invalid IDs are group IDs (several platforms that share the same ID, for example, call sign SHIP representing all anonymous ships) and ‘‘single reporter’’ IDs (IDs with fewer than three reports per month).

Other invalid IDs are those containing illegal characters, that is, not numbers or letters. IDs are also checked for consistency with corresponding platform types according to the WMO’s call sign allocation rules.

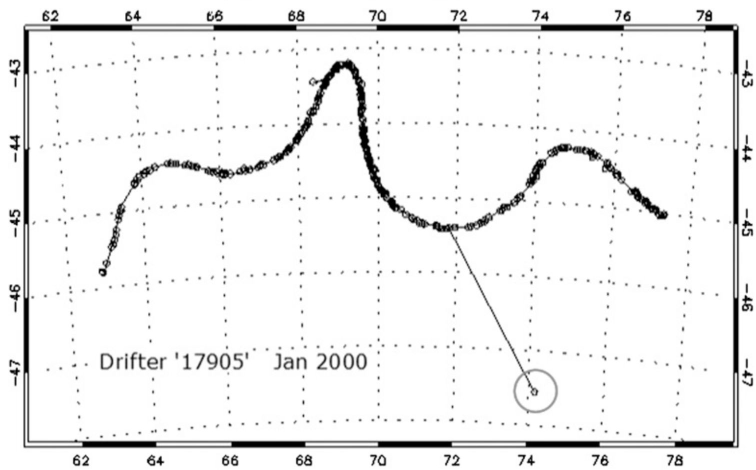
c. *Bayesian checks*

1) Reference (background) check

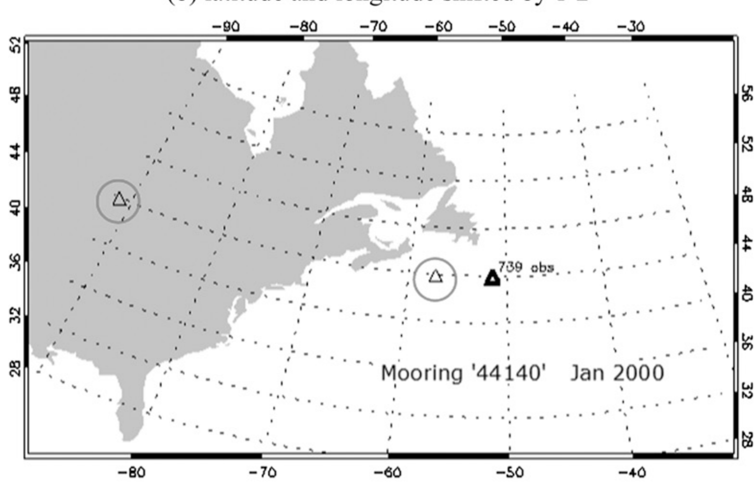
Reference check (RC) is the major check of many QC methods, which identifies most outliers. The Bayesian-based approach by Lorenc and Hammon (1988) was adopted in the *i*Quam QC algorithm. Compared to conventional outlier detection methods, it employs the Bayesian probability theory to take into better account factors such as the accuracy of the reference field itself,



(a) latitude sign swapped



(b) latitude and longitude shifted by 1-2°



(c) located off from the moored position

FIG. 1. Erroneous records of (a) ship, (b) drifter, and (c) mooring buoy detected by tracking check.

error due to the difference in the locations of observation and reference grid point, and the instrumental noise of in situ data. A brief description is given below. For details and theoretical derivation, the reader is referred to Lorenc and Hammon (1988).

According to Bayes' theorem, the *posterior* probability of gross error is calculated as (Lorenc and Hammon 1988)

$$P(E\,|\,O) = \frac{kP(E)}{kP(E) + P(O\,|\,\overline{E})[1 - P(E)]}. \qquad (3)$$

Here, events E and $\overline{E}$ denote gross error and normal situations, respectively. The O denotes the event of getting an observation, T_o, and k is the density of probability distribution of an observation value when a gross error occurs. Assuming a uniform distribution within a range of 10 K, k is set to 0.1 (Lorenc and Hammon 1988). The $P(E)$ is the a priori probability of gross error event, which is empirically chosen according to the percentage of outliers in each platform type.

The quantity $P(O\,|\,\overline{E})$ is the probability distribution of an observation without a gross error. Assuming that both observation T_o and reference T_r obey normal distributions around the true SST value, it is written as

$$P(O\,|\,\overline{E}) = N(T_o - T_r, \sigma_o^2 + \sigma_r^2), \qquad (4)$$

where σ_o and σ_r are the a priori noise (SDs) of the observation and reference, respectively.

In our implementation, the *prior* for σ_o and $P(E)$ is set differently for different types of platforms. These numbers are chosen empirically based on statistical analyses described by Xu and Ignatov (2010). Specifically, the a priori noise is chosen as 1.0 K for ships, 0.3 K for tropical moored and drifting buoys, and 0.6 K for coastal moored buoys. The a priori $P(E)$ is selected as 0.06 for ships, 0.05 for drifters, 0.02 for tropical moorings, and 0.04 for coastal moorings.

Reynolds optimal interpolation (OI) global 0.25° daily analysis SST (AVHRR only) was selected as reference (Reynolds et al. 2007). Recall that Reynolds SST is a blended product of AVHRR satellite retrievals and quality-controlled International Comprehensive Ocean–Atmosphere Data Set (ICOADS) in situ SSTs (or NCEP GTS in situ SSTs, for NRT applications), and it is available from September 1981 onward. Gridded 0.25° resolution data are bilinearly interpolated in space, to each in situ observation. No interpolation in time is attempted, as it would require a reference field with resolved diurnal cycle, which is currently unavailable in *i*Quam. Note that the previous-day Reynolds SST is used in current-day QC, in an attempt to improve *i*Quam latency and minimize the cross dependence of reference and in situ data.

The SD of reference SST, σ_r, should also include the matching errors rising from the space and time difference between the reference field and the actual measurement point. Therefore, an empirical reference SD is calculated based on local statistics as follows:

$$\sigma_r = \sqrt{\sigma_{\text{r-local}}^2/4 + \sigma_{\text{r-base}}^2}. \qquad (5)$$

Here, the base SD, $\sigma_{\text{r-base}}$, is set to 0.2 K for the Reynolds daily product, and the local SD, $\sigma_{\text{r-local}}$, is calculated from the reference 0.25° SST field within a 1° × 1° × 3 days running window (i.e., SD of 4 × 4 × 3 grid points), and further scaled by 1/4, based on empirical analyses and sensitivity studies. Equation (5) was verified by comparing the estimated $\sigma_o^2 + \sigma_r^2$ to the statistics of "in situ–reference" SST given in (Xu and Ignatov 2010). Note that the diurnal warming present in in situ measurements (e.g., Kennedy et al. 2007) is not accounted for in the Reynolds SST. In the future, using a diurnally resolving reference SST, or an empirical bias and/or SD correction adaptive to the local hour, may be considered. Alternatively, the *i*Quam QC reference check may only be applied at night, and the derived QF may be extended to daytime data.

The Bayesian reference check is more flexible than a conventional approach that is based on setting fixed thresholds. The relationship between the posterior probability of gross error $P(E\,|\,O)$ and the departure from the reference SST is not a simple and global one, as it varies in space and time and differs for different sensors (Kent and Berry 2005; Kennedy et al. 2011a).

2) Cross-platform (buddy) check

Cross-platform check (XC) is a critical complement to the reference check, which may compensate for some RC deficiencies, resulting from possible inaccuracies in the reference field, for example. The Bayesian XC is performed on the top of the RC—that is, it updates the *posteriori* probability of gross error by incorporating information from nearby measurements (a.k.a buddies) (Lorenc and Hammon 1988; Ingleby and Huddleston 2007).

The simplest case of cross checking two nearby observations, O_1 and O_2, and adjusting their probabilities of gross error—that is, $P(E_1\,|\,O_1), P(E_1\,|\,O_2)$—is derived as (Lorenc and Hammon 1988)

$$P(E_1\,|\,O_1 \cap O_2) = \frac{P(E_1\,|\,O_1)P(O_1)P(O_2)}{P(O_1 \cap O_2)}. \qquad (6)$$

When simultaneously checking multiple nearby observations, computation may become prohibitively expensive (Ingleby and Lorenc 1993). The iterative approximation, initially suggested by Lorenc and Hammon (1988), proved efficient and accurate for QC purposes (Ingleby and Lorenc 1993). This approximation sequentially adjusts the probability of gross error as checks with nearby observations are performed, one by one. Assuming N nearby observations (buddies), $O_i, i = 2, 3, \ldots, N+1$; the approximation is expressed as

$$P(E_1 \mid O_1 \cap \bigcap_i O_i) \approx P(E_1 \mid O_1) \prod_i \frac{P(O_1)P(O_i)}{P(O_1 \cap O_i)}. \qquad (7)$$

The iterative approximation in Eq. (7) could make the probability overly adjusted, when too many related buddies are included. For example, a significant number of nearby observations from the same problematic platform may amplify the adjustment and wrongly reject good data. One of the three anonymous reviewers of this paper rightly pointed out that rejection of good data is a problem in any QC process that uses data from neighboring platforms. One technique that has been shown to reduce the rejection rate of good data is to introduce a second pass of the platform cross check, this time checking only stations rejected in the first pass and omitting flagged observations when doing the calculations. This strategy has been used elsewhere in the context of an OI-based buddy check with quite good results. To alleviate this problem, Ingleby and Lorenc (1993) proposed a damping factor of 0.5 to buddy check.

In *i*Quam, an adaptive damping factor is used instead, that is,

$$P(E_1 \mid O_1 \cap \bigcap_i O_i) \approx P(E_1 \mid O_1) \left[\prod_i \frac{P(O_1)P(O_i)}{P(O_1 \cap O_i)} \right]^{N_0/N}, \qquad (8)$$

where N is the number of buddies being checked and N_0 is an average number of buddies to which N is normalized. The N_0 is empirically set to 6, meaning that up to six independent nearby measurements are usually expected. Note that it also results in amplification of adjustments in cases of five or fewer buddies. However, our analyses suggest that the effect of this amplification is negligible.

Following Ingleby and Huddleston (2007) and Martin et al. (2002), the correlation coefficient between nearby (in either space or time) observations is modeled by two, second-order autoregressive (SOAR) functions. The correlation lengths of two different scales—that is, mesoscale and synoptic scales—are chosen as $\ell_{\text{mes}} = 100$ km and $\ell_{\text{syn}} = 400$ km, respectively; and $t_c = 5$ days is selected as the *e*-folding time.

For NRT applications, the algorithm has to be implemented efficiently in order to reduce data latency, and to optimize computing resources. In *i*Quam implementation, the upper limit is set to 300 km in space and 4 days in time to exclude those buddies from the XC that are too far away. In addition, a space partitioning technique (Moore 1991) is employed to accelerate the buddy search process. As a result, processing time is significantly reduced.

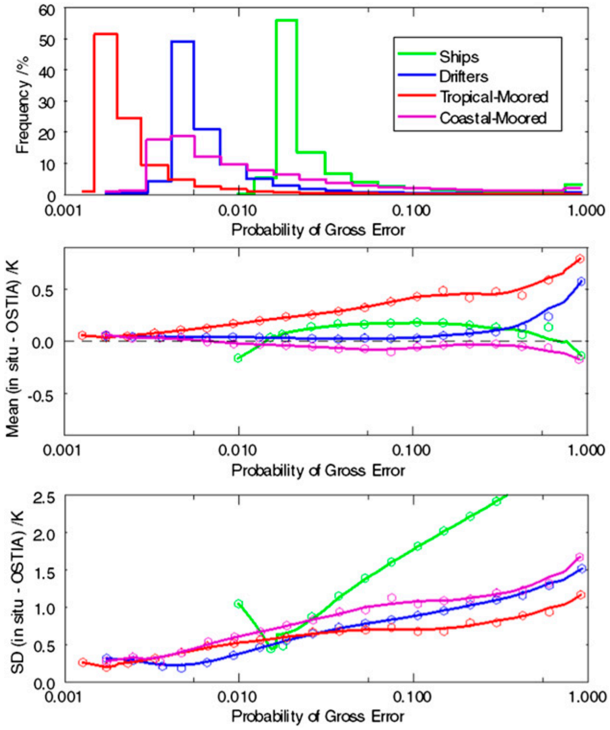


FIG. 2. (top to bottom) Histograms, and in situ minus OSTIA SST mean biases and SDs as functions of probability of gross error.

Note that both RC and XC produce continuous quality indicators that serve as the probabilities of gross error, and are saved on *i*Quam output. In *i*Quam, the threshold of $P > 0.5$ was selected to set up the default overall *i*Quam QF. Other thresholds can be applied to these probabilities by the user if different data quality is desired. See analyses in the following subsection and Fig. 2 for more details.

d. *Efficacy of iQuam QC*

To quickly evaluate the efficacy of *i*Quam QC, one year of NCEP GTS data in 2009 was used in the following analyses. Percentages of detected bad reports and the mean bias and SD of both "bad" and "good" data are calculated for different checks independently. Three binary checks and the two Bayesian checks are analyzed. Note that the XC is applied on the top of the RC, and it adjusts the results of the RC. In an attempt to minimize the effect of using the same Reynolds SST in QC on the below-mentioned diagnostics, statistics in this subsection were all calculated with respect to Operational

TABLE 2. Summary of *i*Quam QC statistics in 2009 NCEP GTS in situ data, including percentage of data identified by each QC check (using Reynolds SST as reference) and the corresponding mean and SD (calculated against independent L4 field and OSTIA SST to more objectively quantify performance of *i*Quam QC based on Reynolds SST). The (mean ±SD) statistics for each check are shown for the corresponding detected outliers, and for remaining points (next row, shown in bold). The all checks row is for the data for which at least one check has failed (i.e., the Boolean sum of all checks: DR and TC and SC and RC and XC). Note that the all checks row percentage will not exactly equal to the arithmetic sum of percentages for individual checks. (Note that GC is not included in this table because reference SST is missing for data points identified by this check.)

% (Mean ±SD)	Ships	Drifters	Tropical moorings	Coastal moorings
Before QC	100% (−0.01 ±2.19)	100% (0.03 ±1.01)	100% (0.06 ±0.76)	100% (−0.02 ±0.88)
DR	0.10% (0.19 ±0.84)	0.44% (−0.04 ±1.09)	0.33% (0.05 ±0.33)	0.003% (−0.535 ±3.67)
	(−0.01 ±2.19)	**(0.03 ±1.01)**	**(0.06 ±0.76)**	**(−0.02 ±0.88)**
TC	0.51% (−0.81 ±6.54)	0.04% (0.40 ±2.06)	0.004% (0.86 ±1.46)	0.001% (1.04 ±7.52)
	(−0.00 ±2.15)	**(0.03 ±1.01)**	**(0.06 ±0.76)**	**(−0.02 ±0.88)**
SC	0.14% (−0.89 ±14.74)	0.19% (−1.16 ±8.87)	0.30% (−0.11 ±12.26)	0.32% (−0.92 ±6.87)
	(−0.01 ±2.12)	**(0.03 ±0.94)**	**(0.06 ±0.37)**	**(−0.02 ±0.78)**
RC	5.75% (−1.67 ±8.08)	1.74% (−0.57 ±7.20)	1.00% (0.22 ±6.98)	3.13% (−0.42 ±2.94)
	(0.10 ±0.97)	**(0.04 ±0.35)**	**(0.06 ±0.30)**	**(−0.01 ±0.71)**
XC (increment from RC)	0.74% (−1.53 ±7.67)	0.39% (−0.51 ±6.55)	0.14% (0.11 ±6.57)	3.87% (−0.21 ±2.26)
	(0.10 ±0.94)	**(0.04 ±0.33)**	**(0.06 ±0.30)**	**(−0.01 ±0.66)**
All checks	6.96% (−1.41 ±7.43)	2.63% (−0.41 ±5.90)	1.50% (0.09 ±5.74)	7.17% (−0.21 ±2.26)
After QC	**(0.10 ±0.94)**	**(0.04 ±0.33)**	**(0.06 ±0.30)**	**(−0.01 ±0.66)**

Sea Surface Temperature and Sea Ice Analysis (OSTIA) SST (Donlon et al. 2012).

Table 2 summarizes the percentage of in situ data, identified by each individual QC check, and the corresponding statistics of ''in situ minus OSTIA'' SST for the points excluded by QC checks and for those retained. Aside from the duplicate removal, where a smaller standard deviation is expected in the excluded sample, because of many identical points, all other checks show significantly degraded statistics in the excluded sample and incremental improvement in the remaining data, with each subsequent check.

Track check detects ~0.5% ship and <0.04% buoy reports with erroneous latitude–longitude–time information. Spike check detects ~(0.1–0.3)% reports with significant SST discontinuities. Although the number of these two types of bad reports is quite low, the data are large in error and must be excluded, even if they only minimally affect the overall statistics of the remaining sample. Moreover, time series in the *i*Quam web interface show that these two checks contributed more in pre-2007 years, and the percentage of bad reports changes greatly from year to year probably because of changes in the procedures of handling the source GTS data. Thus, it is absolutely necessary to have these checks.

The RC is the major check that removes most bad reports (1%–7%) and improves the statistics most significantly.

Table 2 further suggests that the XC additionally removes up to 0.75% more outliers (4% for coastal moorings) on top of the RC. The SST statistics continue to improve following the application of the XC, clearly indicating its valuable contribution to the QC. The much higher XC rate in the case of coastal moorings is likely due to the overestimated SST correlations in coastal areas. These areas are usually shallow and dynamic, and the actual space–time correlation is much weaker here than specified by the global set of parameters adopted in Eq. (6). Consistent with this explanation, the degradation of the statistics is smallest here, although still significant (also likely due to high variability in the SST field that is not captured in the OSTIA SST analysis).

Comparing the error rates in all individual rows with the all checks row, contributions from all checks are improvements and are all significant, suggesting that they are all complementary and an indispensable part of *i*Quam QC.

Note that the XC not only identifies more outliers but it may also rescue good measurements that were wrongly removed by the RC (e.g., because of a biased reference SST). Contribution from the XC is further analyzed in Table 3. The first column is the percentage of reports with one or more buddies available for XC. The second column is for six or more buddies. Note that it is easier to find buddies for ships and coastal moorings than for more sparsely distributed drifters and tropical moorings. The last two columns are percentages of reports (relative to all reports), which were originally identified as good by the RC and then subsequently reclassified as bad by the XC, and vice versa. Apparently, a substantial number of good RC results are additionally

TABLE 3. Contribution of cross-platform check. Note that statistics (mean ±SD) are of the changed portion.

% (mean ±SD)	1+ buddies (%)	6+ buddies (%)	Good by RC → bad by XC	Bad by RC → good by XC
Ships	96	93	0.98% (−0.16 ±2.83)	0.24% (0.68 ±2.76)
Drifters	91	87	0.53% (−0.13 ±1.60)	0.14% (0.17 ±1.09)
Tropical moorings	87	76	0.21% (−0.22 ±1.48)	0.07% (0.65 ±0.66)
Coastal moorings	99	98	4.34% (−0.05 ±1.44)	0.46% (−0.10 ±1.08)

screened out by the XC (from 0.5% to 4%) as well as bad RC results reversed by the XC check (0.2%–0.5%). Comparing the statistics of these two portions of data whose QC is flipped by the XC, the smaller SD in the second group indicates that it has a better quality. The larger bias in the second group could potentially be an indicator that these data actually carry abnormal climate and/or diurnal warming signals not captured in the reference field and therefore are wrongly rejected by the RC. If this is true, then one concludes that the XC is an effective and essential part of *i*Quam QC.

Figure 2 shows the histograms and statistics of in situ–minus–OSTIA SST as a function of P. The histograms look very different for four types of platforms because of different platform-specific a priori settings in the RC. The biases and SDs tend to increase with P, except for some instabilities on the left-hand side, likely due to small samples there. In *i*Quam, the default recommended setting for P is 0.5. Figure 2 should be consulted by those *i*Quam data users who want to utilize the continuous probability of gross error, P (also available in *i*Quam), rather than the default setting of 0.5 adopted in the overall *i*Quam QF.

3. QM and web interface

a. NRT QM

The QC algorithm was implemented at the NOAA Center for Satellite Applications and Research (STAR) with NRT GTS data and routine QM commenced in September 2009. All available GTS data from January 1991 onward have been reprocessed and backfilled. This section describes the QM with a particular emphasis on its web interface (at www.star.nesdis.noaa.gov/sod/sst/iquam/).

The flowchart of the *i*Quam system is shown in Fig. 3. Raw GTS in situ data are automatically accessed twice daily, and then reformatted and appended to the intermediate file, which is subsequently used as input into QC processing. In addition to GTS, QC also uses two ancillary datasets, the land–sea mask and reference SST. Results of QC processing are output into another intermediate file with QFs appended, and the current month file is refreshed on *i*Quam web page. On the fifth day or so of the following month, the current monthly file ceases to be updated. At this point in time, a monthly report is generated and graphics on the web are updated. The primary goal of the QM is to provide *i*Quam users a quick snapshot of the quality-controlled in situ data in NRT. Summary statistics are also useful for *i*Quam developers to monitor the performance of the QC, and to adjust configurations as needed.

Analyses and plots are available in *i*Quam QM from January 1991 to the present and are organized into four sections:

1) monthly maps, stratified by four platform types;
2) corresponding monthly QC statistics, and histograms and corresponding statistics of quality-controlled ΔSST = in situ minus reference;
3) time series of these statistics; and
4) summary tables and visualization plots for individual platforms.

Note that because of the current NOAA web server security settings, following a user's query, *i*Quam data are downloaded to a user's computer, and are processed and displayed there. The current *i*Quam web interface is partially implemented based on the Yahoo! User Interface (YUI), version 2.0, library, which relies on Flash Player for plotting. Thus, users should have Flash Player installed on their web browsers to be able to view some of the *i*Quam QM results.

b. Web interface and maps

The *i*Quam web interface is shown in Fig. 4. The buttons on the left correspond to the four *i*Quam sections. The top menu facilitates a user's navigation

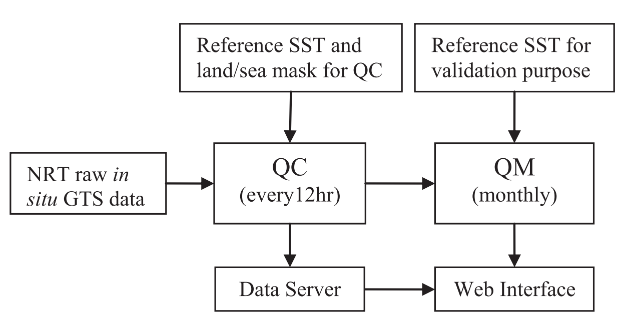
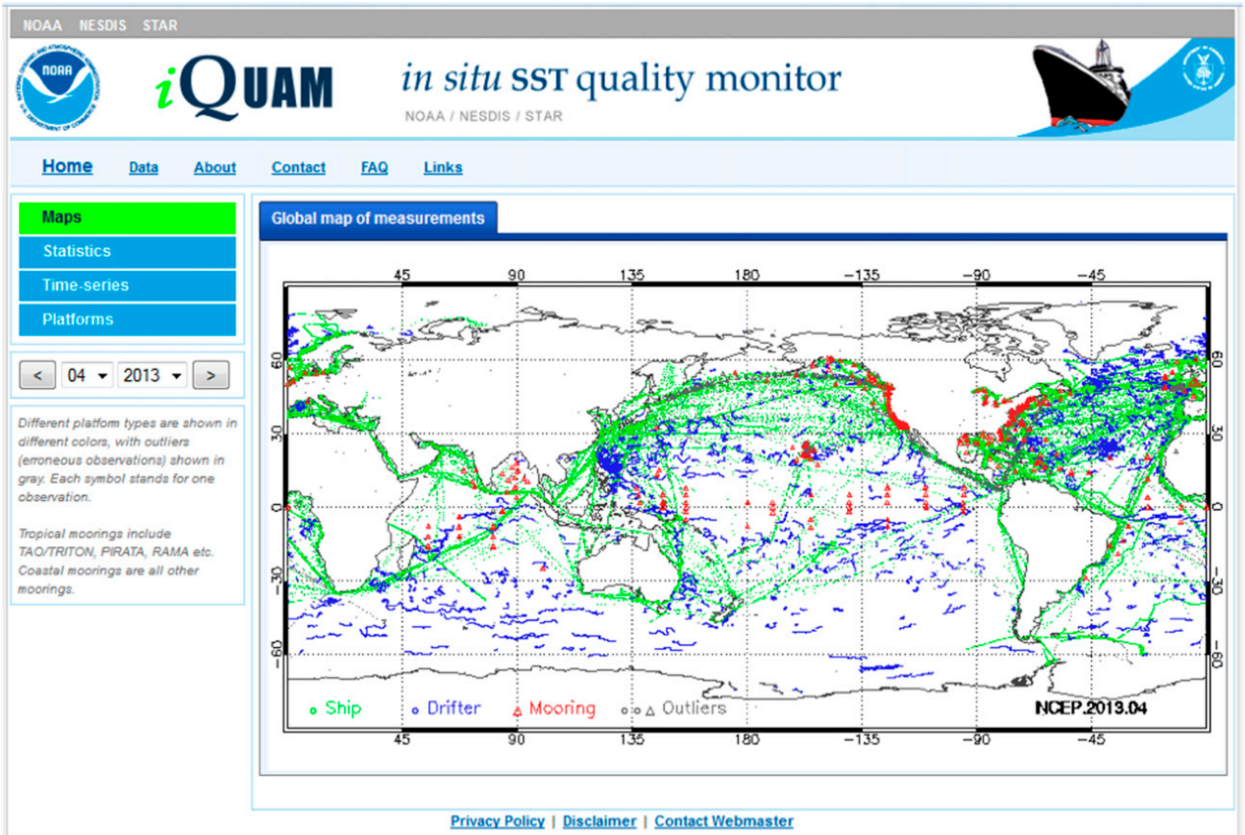
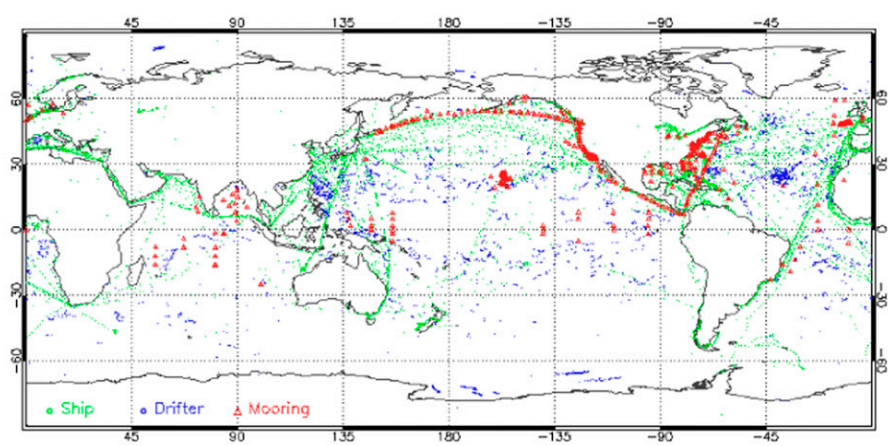


FIG. 3. *i*Quam in situ SST QC and monitoring system. [Currently, Reynolds daily v2 SST (Reynolds et al. 2007) is used for both QC and QM purposes.]



(a)



(b)

FIG. 4. (a) *i*Quam web interface and global map of in situ measurements for April 2013. (b) Global map of detected outliers for April 2013.

through the *i*Quam page, and provides access to the data button (described in section 4).

The default home page is set to display the latest monthly global map (for instance, from 6 May to 5 June, the April map will be displayed). The user can select the year and month, using the drop-down menus, or arrow functions. Four types of in situ measurements are rendered in different colors, whereas outliers detected by

QC are shown in gray. Comparisons of later (e.g., 2013) with earlier (e.g., 1991) maps suggest that the number of ship data has declined, whereas measurements from buoys (both drifters and moorings) have significantly increased. Large areas of the ocean are now covered with in situ data, and geographical biases and voids are significantly reduced, but they are still observed in the data.

To emphasize the number and geographical distribution of outliers, a separate map is shown in Fig. 4b, using the color codes adopted for individual platforms. Outliers are found in all types of platforms, although to a different degree. Consistent tracks for some ships or drifting buoys suggest that all (or at least majority of) data are consistently excluded as outliers, prompting the need for more in-depth analyses of those platforms. One ship was misclassified as a moored buoy, and likely wrongly rejected by the track check. Analyses of outliers provide a useful feedback to producers of in situ data, or to producers of reference SST fields. If some areas of the ocean consistently show an anomalously high data rejection rate, then this might indicate a problem with the reference field.

c. *Statistics and histograms*

The second section of *i*Quam reports statistics of the QC and quality-controlled ΔSSTs, and corresponding histograms of ΔSSTs (Fig. 5). The QC statistics are summarized in a table that shows the total number of observations (N_Obs), the number of observations that passed QC (N_QC), and the number of outliers detected by individual QC tests: duplicate removal (DR), geolocation (GC), track (TC), spike (SC), RC, and XC checks. (Note that the XC column shows cumulative numbers of detections by both RC and XC.) Another table summarizes statistics of quality-controlled ΔSSTs, including the mean bias, SD, skewness, kurtosis, median, robust SD (RSD), and number of matchups (N_Mtchup). N_Mtchup may be different from N_QC, since not all in situ data have matching reference SST—for example, observations from lakes are not defined in the OI v2 land–sea mask. Finally, histograms of ΔSSTs are also plotted. Their shape is near Gaussian (cf. relatively small values of skewness and kurtosis reported in Fig. 5b).

d. *Time series*

The time series section plots time series of the number of platforms (N_ID) and observations (N_Obs), from January 1991 to present (Figs. 6a,b). The number of ships has gradually declined, whereas the number of buoys increased. Mean biases and SDs of ΔSSTs are also plotted (Figs. 6c,d). Drifters and tropical moorings, customarily used in satellite Cal/Val, show comparable SDs—historically, ~0.4 K and closer to ~0.3 K in recent years (cf. Xu and Ignatov 2010, and references therein). Note that ship SSTs are known to be biased warm because of the use of engine intake and specifics of the thermometers on the Voluntary Observing Ships (VOS; e.g. Kent et al. 1993; Emery et al. 2001a; Kent and Taylor 2006). Time series of the number of outliers detected by different QC checks are also shown.

e. *Platform-specific statistics*

The last *i*Quam section reports statistics for individual platforms (Fig. 7). First, a sortable list of all platforms is displayed, with QC and ΔSST statistics similar to those described in section 3c, but they now calculated for each individual platform. Clicking on the platform ID brings up a platform monitor window that shows either a monthly trajectory, a time series of ΔSST, or a complete history of outlier rate for this platform.

4. *i*Quam data: Formats, quality flags, and users

a. *Quality-controlled in situ data*

Quality-controlled data generated by *i*Quam are served online in self-documented HDF format (cf. www.star.nesdis.noaa.gov/sod/sst/iquam/data.html). Although GTS data are processed in NRT with ~12-h latency, data from the 4 previous days are continuously reprocessed as new "buddies in time" become available for the XC. Hence, the *i*Quam QFs are continuously updated and not finalized until 5 days later. Preliminary analyses suggest that the updates to QFs are minimal, and the value of such reprocessing may be reexamined in future *i*Quam versions, to improve the latency of data.

In situ data with QFs appended are aggregated into monthly files. The latest month file is continuously updated and available with a 12-h lag. The naming convention for the *i*Quam quality-controlled monthly data files is IQUAM.NCEP.YYYY.MM.HDF, where NCEP denotes the data source, YYYY denotes the four digit year, and MM is the two-digit month. The general information regarding each HDF file is found in the global attributes section. Definitions of the common global attributes are listed in Table 4.

The *i*Quam data files preserve all information from the original GTS reports, including SST as well as other in situ measurements. However, only SST is quality controlled and corresponding QFs are only set for SST. Table 5 summarizes data layers contained in *i*Quam HDF files. The first several layers are time and location information. The ID layer is the particular buoy ID or ship call sign that is reported in the GTS system. The Type layer is used to distinguish different in situ platforms: 0—unknown; 1—ship; 2—drifting buoy; 3—open-sea (tropical) moored buoy; and 4—coastal moored buoy. The last layer Quality_Flag is a 16-bit field packed with both individual QC results, Bayesian quantitative QC results, QC Statistics

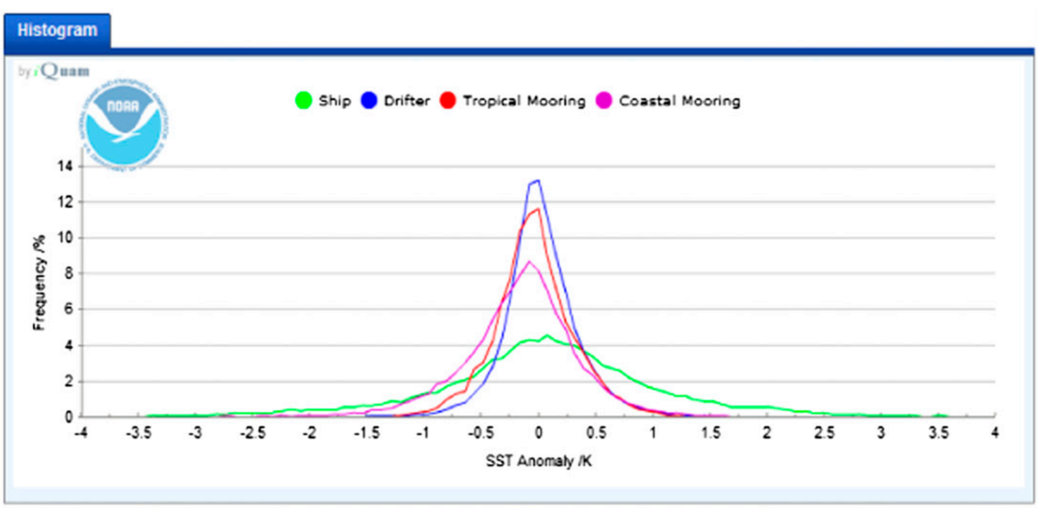
Platform	N_Obs	N_QC	DR	GC	TC	SC	RC	XC
Ship	87,442	74,745	0	6,652	321	245	5,045	5,479
Drifter	628,818	606,887	3,414	1,292	263	1,578	12,782	15,384
Tropical Mooring	32,743	30,410	2,189	2	0	1	94	141
Coastal Mooring	178,957	167,050	0	2,147	123	375	5,617	9,262

(a)

SST Statistics

Platform	BIAS	SD	SKEW	KURT	MED	RSD	N_Mtchp
Ship	0.11	1.01	-0.14	1.47	0.12	0.78	67,506
Drifter	0.06	0.33	0.03	4.41	0.05	0.26	605,417
Tropical Mooring	0.01	0.35	0.09	1.23	-0.01	0.29	29,947
Coastal Mooring	-0.12	0.51	-0.57	2.98	-0.08	0.41	135,697

(b)



(c)

FIG. 5. Monthly statistics stratified by platform types for April 2013: (a) Number of in situ data; (b) statistics of ΔSST = in situ minus reference (bias, SD, skewness, kurtosis, median, RSD, and number of matchups of QC in situ with reference SST); and (c) frequency of QC ΔSSTs. Note that the frequency curves are widest for ships (lower data quality, larger noise) and narrowest for drifters and tropical moorings (higher data quality, smaller noise). For additional discussion, see Xu and Ignatov (2010) and references therein.

and the overall QC result. Definitions and recommended usage are described in the next subsection.

b. *Quality_Flag layer*

All layers in the HDF file listed in Table 5 are passed along from the GTS data unaltered, except the Quality_Flag layer, which is produced by *i*Quam and appended to the data. From the lowest bit of 0 to the highest bit of 15, explanations are given in Table 6.

The lowest two bits are reserved for the overall quality flag, which is derived from individual flags and indicators as explained in Table 7. This summary flag is intended

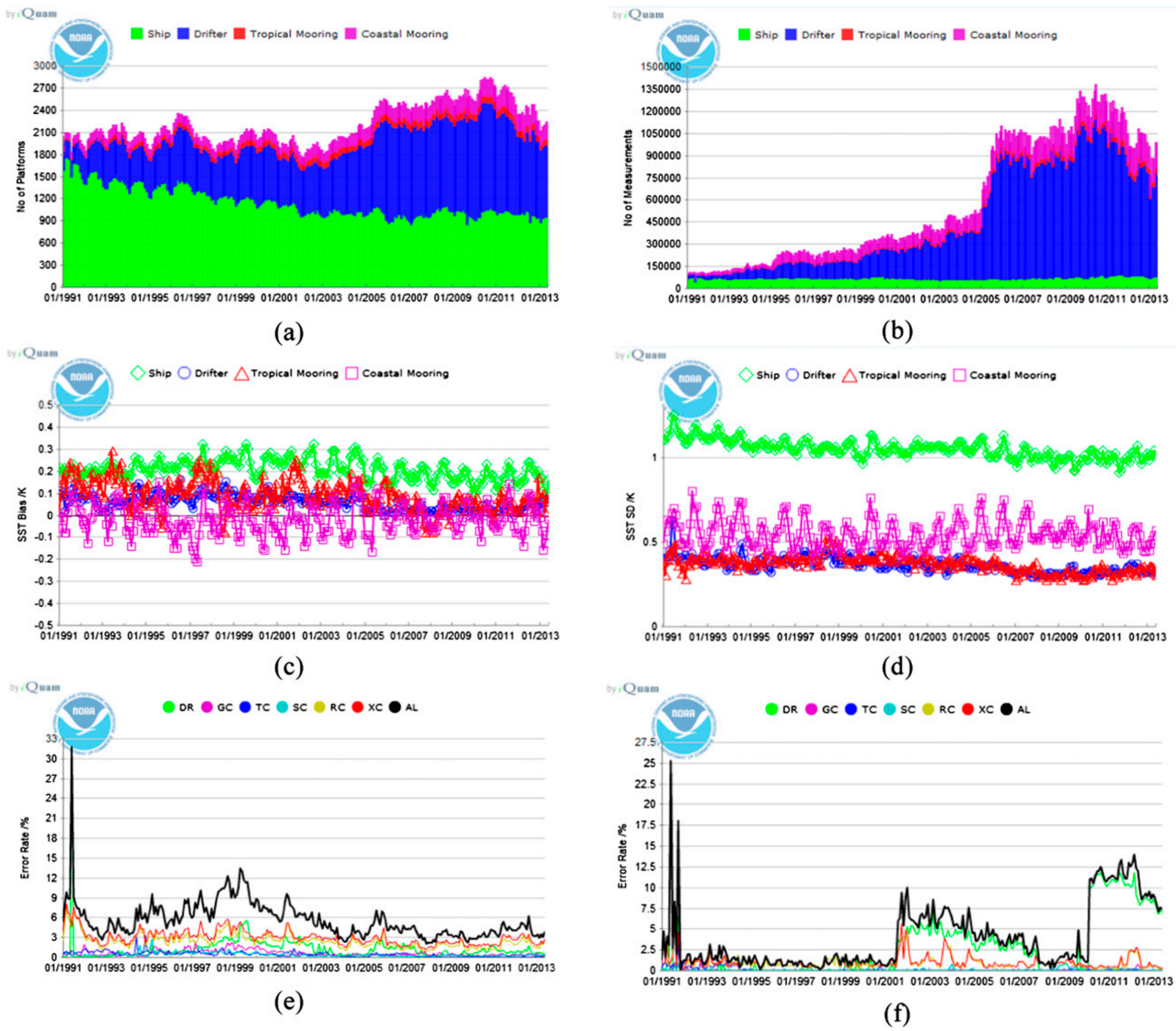


FIG. 6. Time series of monthly statistics stratified by platform types: (a) number of platforms; (b) number of observations; (c) mean biases of SST anomalies after QC; (d) SD of SST anomalies after QC; (e) error rates (percentage of detected erroneous measurements) of each QC check for drifters; and (f) error rates of each check for tropical moorings. The reason for two sharp increases in the fraction of duplicate records in tropical moored buoys around 2002 and 2010 is not immediately clear.

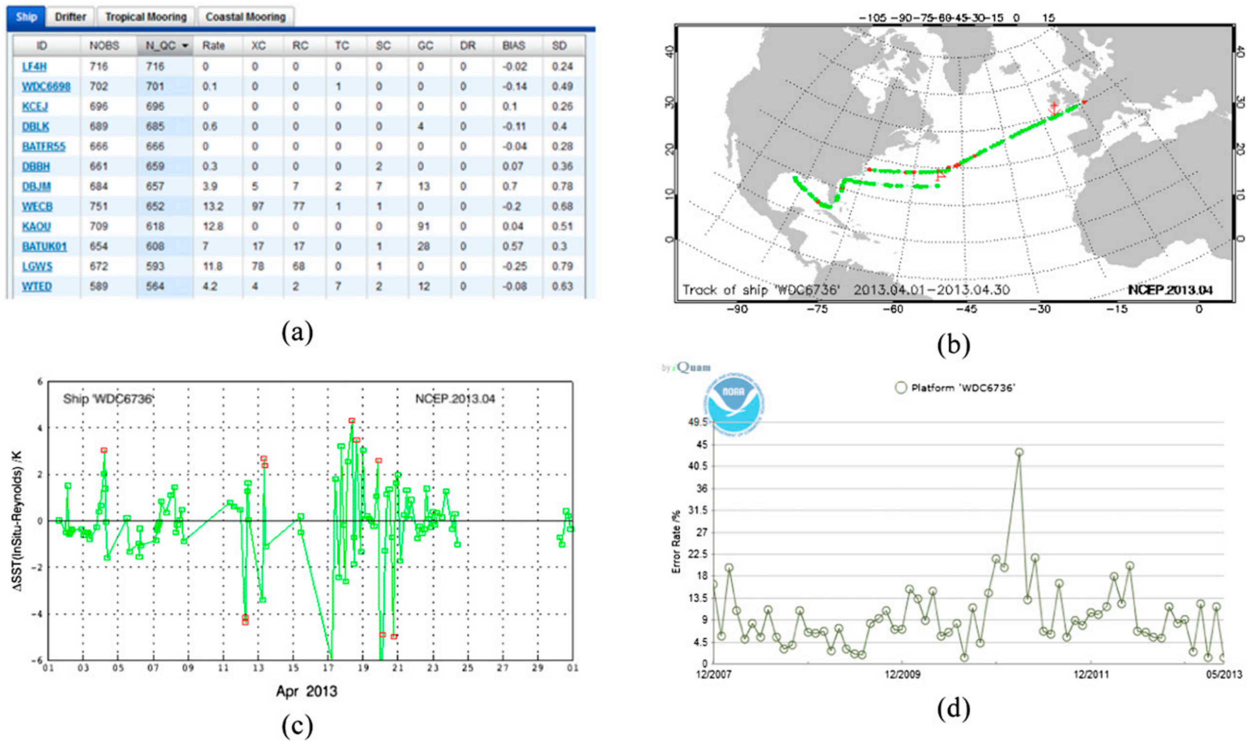
for general users, who need "good in situ data" but are not interested in digging into the individual QFs. It is recommended that

- for high-accuracy applications, use data with the lowest two bits cleared (QF AND 0x0003 == 0), that is, Normal only;
- for general application, use data with the lowest bit cleared (QF AND 0x0001 == 0), that is, Normal and Noisy.

All individual checks are also reported and available for more advanced applications. Bits 2 to 6 report results of individual binary checks described in section 2b. Bit 7 reports the number of buddies checked [cf. section 2c (2)]. This layer is not used in setting the overall quality flag, but it may be useful for a more advanced user. The second byte is a continuous probability of gross error (P) ranging from 0 (byte value 0x00) to 1 (byte value 0xFF), which is produced cumulatively by the reference and cross-platform checks (for its interpretation and characteristics, see Fig. 2). Users can customize the threshold of P for their own application requirements.

c. *i*Quam data users

The *i*Quam data have been used at NOAA and several external organizations for satellite Cal/Val applications. In particular, the *i*Quam system was identified as



ID	NOBS	N_QC	Rate	XC	RC	TC	SC	GC	DR	BIAS	SD
LF4H	716	716	0	0	0	0	0	0	0	-0.02	0.24
WDC6698	702	701	0.1	0	0	1	0	0	0	-0.14	0.49
KCEJ	696	696	0	0	0	0	0	0	0	0.1	0.26
DBLK	689	685	0.6	0	0	0	0	4	0	-0.11	0.4
BATFR55	666	666	0	0	0	0	0	0	0	-0.04	0.28
DBBH	661	659	0.3	0	0	0	2	0	0	0.07	0.36
DBJM	684	657	3.9	5	7	2	7	13	0	0.7	0.78
WECB	751	652	13.2	97	77	1	1	0	0	-0.2	0.68
KAOU	709	618	12.8	0	0	0	0	91	0	0.04	0.51
BATUK01	654	608	7	17	17	0	1	28	0	0.57	0.3
LGWS	672	593	11.8	78	68	0	1	0	0	-0.25	0.79
WTED	589	564	4.2	4	2	7	2	12	0	-0.08	0.63

FIG. 7. Individual platform statistics for ship WDC6736, April 2013: (a) list of platforms and their statistics of QC results and SST deviations from reference; (b) monthly track map of individual platform; (c) monthly time series of individual platform SST anomalies [with erroneous points in (b) and (c) labeled in red]; and (d) error rate history of individual platform.

the in situ data source for the current heritage polar and geostationary SST products, as well as advanced JPSS and GOES-R programs. The *i*Quam also serves as the NOAA contribution to the international Group for High Resolution SST (Donlon et al. 2007).

The major use of the *i*Quam data is the routine generation of match-up datasets with various level 2 (L2) and L3 satellite SST products from AVHRR, Moderate Resolution Imaging Spectroradiometer (MODIS), and Visible Infrared Imager Radiometer Suite (VIIRS) produced by various data centers, including NOAA, NAVOCEANO, NASA, and OSI SAF. Work is also underway to generate consistent matchups with geostationary SSTs. These match-up data are used to calculate coefficients of the regression equations, and to perform algorithm development and comparisons (e.g., Petrenko et al. 2011; Petrenko et al. 2013, manuscript submitted to *J. Geophys. Res.*). Also, match-up data are routinely input into the NOAA SST Quality Monitor (SQUAM; www.star.nesdis.noaa.gov/sod/sst/squam/; Dash et al. 2010), which, among other functions, performs routine validation of all products and reports their statistical summaries online in NRT. More recently, the L4 SST module was added to SQUAM, and all L4 products are also consistently validated against *i*Quam data (Dash et al. 2012).

Importantly, all products are validated against uniformly quality-controlled in situ data, using QC checks consistent with the larger meteorological and oceanographic communities. Using a uniform and community consensus validation standard rules out product performance differences caused by deficiencies or differences in in situ data, and allows a fair and consistent cross evaluation of various products.

TABLE 4. Global attributes of *i*Quam HDF files.

Name	Description
FILE_NAME	File name
FIRST_CREATED	Time when the file was first created
LAST_UPDATED	Time when the file was last updated
RAW_DATA_SOURCE	Data source
IQUAM_VERSION	Version of *i*Quam
HDF_VERSION	Version of HDF
START_TIME	Time of the first in situ data record included in this file
END_TIME	Time of the last in situ data record included in this file
SOURCE	Data distributor
CONTACT	Contact points

TABLE 5. Data layers in *i*Quam HDF files. The variable *n* is the number of records in the file, the letter "b" stands for "bit," and the abbreviation NaN means "not a number."

Name	Dimension	Data type	Unit	Range	Fill value
Year	n	16b INT	—	—	—
Month	n	8b UNIT	—	1–12	—
Day	n	8b UNIT	—	1–31	—
Hour	n	8b UNIT	—	0–23	—
Minute	n	8b UNIT	—	0–59	—
Latitude	n	32b Float	Degrees	−90–90	—
Longitude	n	32b Float	Degrees	0–360	—
ID	$n \times 8$	8b UNIT	—	—	—
Type	n	8b UNIT	—	1–4	0
Sea_Surface_Temperature	n	32b Float	Degrees Celsius	—	NaN
Sea_Surface_Pressure	n	32b Float	Pascal	—	NaN
Wind_Direction	n	32b Float	Degrees	0–360	NaN
Wind_Speed	n	32b Float	$m\,s^{-1}$	—	NaN
Air_Temperature	n	32b Float	Degrees Celsius	—	NaN
Dew_Point	n	32b Float	Degrees Celsius	—	NaN
Cloud_Coverage	n	32b Float	%	—	NaN
Quality_Flag	n	16b UNIT	—	—	0xFFFF

One anonymous reviewer of this paper also suggested that platform-specific monitoring information in *i*Quam can be used to identify problematic instruments and take remedial actions. Thus, *i*Quam could potentially contribute to improvements in the quality of in situ data.

5. Conclusions and future work

The NRT in situ SST Quality Monitor (*i*Quam; www.star.nesdis.noaa.gov/sod/sst/iquam/) has been developed with the primary goal to support satellite Cal/Val at NOAA, including heritage polar and geostationary, as well as the newer JPSS and GOES-R, SST products. The following are three major *i*Quam functions: 1) performing advanced and uniform QC of GTS data that are consistent with best practices adopted in wider meteorological and oceanographic communities; 2) monitoring quality-controlled SSTs online; and 3) serving data to NOAA and external users.

QC checks implemented in *i*Quam include several binary (duplicate removal, geolocation, tracking, and spike) and Bayesian (reference against Reynolds L4 analysis and cross platform) checks, the latter two being the major checks. Processing time ranges from $\sim$0.5 h yr^{-1} of data for early years to $\sim$6 h yr^{-1} of data after 2005, on an average NOAA PC. All checks are necessary and unique and improve SST performance statistics measured against an independent L4 SST field—OSTIA.

The online quality monitoring system provides four types of diagnostics: 1) monthly global maps of in situ platforms; 2) corresponding monthly statistics of the number of platforms and measurements, QC results, and SST deviations from Reynolds SST, stratified by four

TABLE 6. Definition of 16-bit Quality_Flag layer. QI = quality indicator.

Bit	Name	Description
0–1	Overall quality	0—Normal, 2—Noisy, 1—Erroneous, 3—QC unavailable
2–3	DR	0—No duplicate, 1—Duplicate kept, 2—Duplicate removed
4	TC and GC	0—Pass, 1—Fail
5	SST SC	0—Pass, 1—Fail
6	ID check	0—Valid, 1—Invalid
7	Number of buddies	0—Checked with at least six buddies, 1—Otherwise
8–15	Scaled probability of gross error	BYTE (Probability_of_Gross_Error × 255), i.e., QI

TABLE 7. Definition of 2-bit overall quality flag. Note that in *i*Quam, a combination of Normal and Noisy data are monitored, whereas in SQUAM validation, only Normal data are used. (For example, in April 2013, 73.1% ships, 94.8% drifters, 91.7 tropical moorings, and 86.7% coastal moorings are Normal; 14.5%, 3.4%, 7.1%, and 6.7% are Noisy; 6.8%, 1.7%, 1.2%, and 5.2% are Erroneous; and 5.7%, 0%, 0%, and 1.5% are QC not performed.)

Bit	Overall quality	Definition
00	Normal	(DR, TC, SC, GC all pass) and ($P_{GE} < 0.1$) and (ID = valid)
10	Noisy	[(DR, TC, SC, GC all pass) and ($0.1 \le P_{GE} < 0.5$)] or (ID = invalid)
01	Erroneous	($P_{GE} \ge 0.5$) or (any of DR, TC, SC, GC fail)
11	QC unavailable	QC not performed

platform types monitored in *i*Quam—ships, drifters, and tropical and coastal moorings; 3) time series of all those statistics; and 4) statistics stratified by individual platforms.

Quality-controlled data are served online via the *i*Quam website in HDF format, and include all the layers originally included in GTS as well as an additional layer of SST Quality_Flag. The 16-bit Quality_Flag includes a 2-bit overall QF, which is recommended for an average user, as well as all individual QFs, so that the user has a flexibility to derive a different overall QF. The *i*Quam data are routinely used as input to another NOAA online NRT system, the SST Quality Monitor (SQUAM; www.star.nesdis.noaa.gov/sod/sst/squam/; Dash et al. 2010, 2012), where they are used for a consistent and uniform validation of various L2, L3, and L4 SST products. The *i*Quam system serves as the official source of in situ data for all NOAA heritage as well as newer JPSS and GOES-R SST products.

Ongoing work toward *i*Quam version 2 includes adding Argo profilers; extending *i*Quam time series back to the start of satellite era (i.e., early 1980s; currently, they only cover the period from 1991 to present); using ICOADS data instead of GTS, whenever available; using diurnally resolving reference SST, or stratifying QC by day and night to account for SST diurnal warming; adding QFs from the external "black lists" developed by the Met Office and OSI SAF; and testing more accurate L4 analysis fields as a reference SST (e.g., Saha et al. 2012). Future work will also include more accurate estimation of an error budget in each type of in situ SST through three-way (or multiway) joint error estimation (e.g., O'Carroll et al. 2008; Xu and Ignatov 2010).

Acknowledgments. The *i*Quam development is supported by the JPSS and GOES-R Programs and by the Polar Product System Development and Implementation, NOAA Data Exploitation, and Ocean Remote Sensing Programs. We thank our colleagues at NOAA (J. Sapper, D. Stokes, S. Woodruff, P. Dash, Y. Kihai, X. Liang, and B. Petrenko), JPSS SST (P. LeBorgne, P. Minnett, B. Evans), and GHRSST (N. Rayner, J. Kennedy, E. Kent, C. Merchant, H. Beggs, M. Chin) for helpful discussions and feedback on the use of *i*Quam data. Thanks also go to three anonymous reviewers of this manuscript and to JTECH Editor Prof. William J. Emery for the valuable recommendations. The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official NOAA or U.S. government position, policy, or decision.

REFERENCES

Bitterman, D. S., and D. V. Hansen, 1993: Evaluation of SST measurements from drifting buoys. *J. Atmos. Oceanic Technol.,* **10,** 88–96.

Brasnett, B., 1997: A global analysis of SST for numerical weather prediction. *J. Atmos. Oceanic Technol.,* **14,** 925–937.

——, 2008: The impact of satellite retrievals in a global SST analysis. *Quart. J. Roy. Meteor. Soc.,* **134,** 1745–1760.

Brisson, A., P. Le Borgne, and A. Marsouin, 2002: Results of one year of preoperational production of sea surface temperatures from *GOES-8*. *J. Atmos. Oceanic Technol.,* **19,** 1638–1652.

Castro, S. L., G. A. Wick, and W. J. Emery, 2012: Evaluation of the relative performance of SST measurements from different types of drifting and moored buoys using satellite-derived reference products. *J. Geophys. Res.,* **117,** C02029, doi:10.1029/2011JC007472.

Dash, P., A. Ignatov, Y. Kihai, and J. Sapper, 2010: The SST Quality Monitor (SQUAM). *J. Atmos. Oceanic Technol.,* **27,** 1899–1917.

——, and Coauthors, 2012: Group for High Resolution Sea Surface Temerature (GHRSST) analysis fields inter-comparisons—Part 2: Near real time web-based level 4 SST Quality Monitor (L4-SQUAM). *Deep-Sea Res. II,* **77–80,** 31–43, doi:10.1016/j.dsr2.2012.04.002.

Donlon, C. J., and Coauthors, 2007: The Global Ocean Data Assimilation Experiment High-Resolution Sea Surface Temperature Pilot Project. *Bull. Amer. Meteor. Soc.,* **88,** 1197–1213.

——, M. Martin, J. D. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer, 2012: The Operational Sea Surface Temperature and Sea Ice analysis (OSTIA). *Remote Sens. Environ.,* **116,** 140–158, doi:10.1016/j.rse.2010.10.017.

Emery, W. J., D. J. Baldwin, P. Schluessel, and R. W. Reynolds, 2001a: Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements. *J. Geophys. Res.,* **106** (C2), 2387–2405.

——, S. Castro, G. A. Wick, P. Schluessel, and C. Donlon, 2001b: Estimating sea surface temperature from infrared satellite and in situ temperature data. *Bull. Amer. Meteor. Soc.,* **82,** 2773–2785.

Francois, C., A. Brisson, P. Le Borgne, and A. Marsouin, 2002: Definition of a radiosounding database for sea surface brightness temperature simulations: Application to sea surface temperature retrieval algorithm determination. *Remote Sens. Environ.,* **81,** 309–326.

Gronell, A., and S. E. Wijffels, 2008: A semiautomated approach for quality controlling large historical ocean temperature archives. *J. Atmos. Oceanic Technol.,* **25,** 990–1003.

Hansen, D. V., and P.-M. Poulain, 1996: Quality control and interpolations of WOCE-TOGA drifter data. *J. Atmos. Oceanic Technol.,* **13,** 900–909.

Hansen, M., R. DeFries, J. R. G. Townshend, and R. Sohlberg, 2000: Global land cover classification at 1 km resolution using a decision tree classifier. *Int. J. Remote Sens.,* **21,** 1331–1365.

Ingleby, B., 2010: Factors affecting ship and buoy data quality: A data assimilation perspective. *J. Atmos. Oceanic Technol.,* **27,** 1476–1489.

——, and A. C. Lorenc, 1993: Bayesian quality control using multivariate normal distributions. *Quart. J. Roy. Meteor. Soc.,* **119,** 1195–1225.

——, and M. Huddleston, 2007: Quality control of ocean temperature and salinity profiles—Historical and real-time data. *J. Mar. Syst.,* **65,** 158–175.

Kennedy, J. J., P. Brohan, and S. F. B. Tett, 2007: A global climatology of the diurnal variations in sea-surface temperature and implications for MSU temperature trends. *Geophys. Res. Lett.,* **34,** L05712, doi:10.1029/2006GL028920.

——, N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011a: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res.,* **116,** D14103, doi:10.1029/2010JD015218.

——, ——, ——, ——, and ——, 2011b: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *J. Geophys. Res.,* **116,** D14104, doi:10.1029/2010JD015220.

——, R. O. Smith, and N. A. Rayner, 2012: Using AATSR data to assess the quality of in situ sea-surface temperature observations for climate studies. *Remote Sens. Environ.,* **116,** 79–92.

Kent, E. C., and D. I. Berry, 2005: Quantifying random measurement errors in Voluntary Observing Ships' meteorological observations. *Int. J. Climatol.,* **25,** 843–856, doi:10.1002/joc.1165.

——, and P. K. Taylor, 2006: Toward estimating climatic trends in SST. Part I: Methods of measurement. *J. Atmos. Oceanic Technol.,* **23,** 464–475.

——, and P. G. Challenor, 2006: Toward estimating climatic trends in SST. Part II: Random errors. *J. Atmos. Oceanic Technol.,* **23,** 476–486.

——, and A. Kaplan, 2006: Toward estimating climatic trends in SST. Part III: Systematic biases. *J. Atmos. Oceanic Technol.,* **23,** 487–500.

——, and B. Ingleby, 2010: From observations to forecast—Part 6. Marine meteorological observations. *Weather,* **65,** 231–238.

——, P. K. Taylor, B. S. Truscott, and J. S. Hopkins, 1993: The accuracy of Voluntary Observing Ship's meteorological observations: Results of the VSOP-NA. *J. Atmos. Oceanic Technol.,* **10,** 591–608.

——, J. J. Kennedy, D. I. Berry, and R. O. Smith, 2010: Effects of instrumentation changes on sea surface temperature measured *in situ*. *Wiley Interdiscip. Rev.: Climate Change,* **1,** 718–728.

Kilpatrick, K. A., G. P. Podestá, and R. Evans, 2001: Overview of the NOAA/NASA Advanced Very High Resolution Radiometer Pathfinder algorithm for sea surface temperature and associated matchup database. *J. Geophys. Res.,* **106** (C5), 9179–9197.

Lorenc, A. C., and O. Hammon, 1988: Objective quality control of observations using Bayesian methods: Theory, and a practical implementation. *Quart. J. Roy. Meteor. Soc.,* **114,** 515–543.

Martin, M. J., M. J. Bell, and A. Hines, 2002: Estimation of three-dimensional error covariance statistics for an ocean assimilation system. Ocean Applications Tech. Note 30, Met Office, 23 pp.

Merchant, C. J., P. Le Borgne, A. Marsouin, and H. Roquet, 2008: Optimal estimation of sea surface temperature from split-window observations. *Remote Sens. Environ.,* **112,** 2469–2484.

Moore, A. W., 1991: Efficient memory-based learning for robot control. Tech. Rep. 209, Robotics Institute, Carnegie Mellon University, 82 pp.

O'Carroll, A. G., J. G. Watts, L. A. Horrocks, R. W. Saunders, and N. A. Rayner, 2006: Validation of the AATSR Meteo product sea surface temperature. *J. Atmos. Oceanic Technol.,* **23,** 711–726.

——, J. R. Eyre, and R. W. Saunders, 2008: Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. *J. Atmos. Oceanic Technol.,* **25,** 1197–1207.

Petrenko, B., A. Ignatov, N. Shabanov, and Y. Kihai, 2011: Development and evaluation of SST algorithms for GOES-R ABI using MSG SEVIRI as a proxy. *Remote Sens. Environ.,* **115,** 3647–3658.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.,* **108,** 4407, doi:10:1029/2002JD002670.

——, P. Brohan, D. E. Parker, C. K. Folland, J. J. Kennedy, M. Vanicek, T. J. Ansell, and S. F. Tett, 2006: Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset. *J. Climate,* **19,** 446–469.

Reverdin, G., and Coauthors, 2010: Temperature measurements from surface drifters. *J. Atmos. Oceanic Technol.,* **27,** 1403–1409.

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily high-resolution-blended analyses for sea surface temperatures. *J. Climate,* **20,** 5473–5496.

Saha, K., A. Ignatov, X. Liang, and P. Dash, 2012: Selecting a first-guess sea surface temperature field as input to forward radiative transfer models. *J. Geophys. Res.,* **117,** C12001, doi:10.1029/2012JC008384.

Slutz, R. J., S. J. Lubker, J. D. Hiscox, S. D. Woodruff, R. L. Jenne, D. H. Joseph, P. M. Steurer, and J. D. Elms, 1985: Comprehensive Ocean-Atmosphere Data Set, release 1. NOAA Environmental Research Laboratories Tech. Rep., 263 pp. [NTIS PB86-105723.]

Thomas, B. R., E. C. Kent, V. R. Swail, and D. I. Berry, 2008: Trends in ship wind speeds adjusted for observation method and height. *Int. J. Climatol.,* **28,** 747–763.

Woodruff, S. D., 2008: Marine QC scoping document. JCOMM Data Management Coordination Group, WMO DMCG-III/Doc. 5.3 Rev. 1, 17 pp.

——, H. F. Diaz, J. D. Elms, and S. J. Worley, 1998: COADS release 2 data and metadata enhancements for improvements of marine surface flux fields. *Phys. Chem. Earth,* **23,** 517–527.

Worley, S. J., S. D. Woodruff, R. W. Reynolds, S. J. Lubker, and N. Lott, 2005: ICOADS release 2.1 data and products. *Int. J. Climatol.,* **25,** 823–842.

Xu, F., and A. Ignatov, 2010: Evaluation of in situ sea surface temperatures for use in the calibration and validation of satellite retrievals. *J. Geophys. Res.,* **115,** C09022, doi:10.1029/2010JC006129.