Evaluation of the In Situ Sea Surface Temperature Quality Control in the NOAA In Situ SST Quality Monitor (*i*Quam) System

HAIFENG ZHANG,[a,b] ALEXANDER IGNATOV,[a] AND DEAN HINSHAW[a,c]

[a] *NOAA/Center for Satellite Applications and Research, College Park, Maryland*
[b] *Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado*
[c] *Global Science and Technology, Inc., College Park, Maryland*

(Manuscript received 18 December 2020, in final form 29 March 2021)

ABSTRACT: In situ sea surface temperature (SST) measurements play a critical role in the calibration/validation (Cal/Val) of satellite SST retrievals and ocean data assimilation. However, their quality is not always optimal, and proper quality control (QC) is required before they can be used with confidence. The in situ SST Quality Monitor (*i*Quam) system was established at the National Oceanic and Atmospheric Administration (NOAA) in 2009, initially to support the Cal/Val of NOAA satellite SST products. It collects in situ SST data from multiple sources, performs uniform QC, monitors the QCed data online, and distributes them to users. In this study, the *i*Quam QC is compared with other QC methods available in some of the in situ data ingested in *i*Quam. Overall, the *i*Quam QC performs well on daily to monthly time scales over most global oceans and under a wide variety of environmental conditions. However, it may be less accurate in the daytime, when a pronounced diurnal cycle is present, and in dynamic regions, because of the strong reliance on the ''reference SST check,'' which employs daily low-resolution level-4 analyses with no diurnal cycle resolved. The *i*Quam ''performance history check,'' applied to all in situ platforms, is an effective alternative to the customary ''black/gray'' lists, available only for some platforms (e.g., drifters and Argo floats). In the future, *i*Quam QC will be upgraded [e.g., using improved reference field(s), with enhanced temporal and spatial resolutions]. More comparisons with external QC methods will be performed to learn and employ the best QC practices.

KEYWORDS: Surface temperature; In situ oceanic observations; Quality assurance/control

1. Introduction

In situ sea surface temperatures (SST) serve as ''ground truth'' for the calibration and validation (Cal/Val) of satellite SST retrievals (e.g., Saunders 1967; Strong and McClain 1984; Kilpatrick et al. 2001; Brisson et al. 2002; Donlon et al. 2002; Merchant et al. 2006; Castro et al. 2008; O'Carroll et al. 2008; Petrenko et al. 2014). They are also assimilated in ocean analysis and forecast models and are often used to ''bias correct'' various satellite inputs (e.g., Reynolds et al. 2007; Brasnett and Surcel-Colan 2016; Akella et al. 2017; Chin et al. 2017; Storto and Oddo 2019). However, their quality, geolocation, and time attribution may not be known or optimal. Many drifting and moored buoys often remain unattended for years in a hostile environment, and ship records have higher odds of being subject to human errors (e.g., Xu and Ignatov 2010, and references therein). Oftentimes, biases and noise in the data are platform-type and individual-sensor specific (e.g., Ingleby 2010; Xu and Ignatov 2016). Additional errors may occur when data are transmitted from the in situ platform to the satellite and back to the ground, and during their distribution via the Global Telecommunications System (GTS) (Xu and Ignatov 2010). A small fraction of erroneous or improperly geolocated/time-stamped in situ measurements may significantly affect the results of satellite Cal/Val. Because of their relatively sparse distribution, if erroneous values are assimilated, they can cause immediate spurious overtuning and error propagation (Ingleby and Huddleston 2007). Therefore, proper quality control (QC) is required before in situ SST data can be used with confidence.

The need for QCing in situ data was recognized in the early years of satellite SST (e.g., Strong and McClain 1984). However, QC practices have been largely ad hoc, undocumented, nonuniform and overly simplistic. For example, a very common approach to detecting outliers used to be setting a constant [or a multiple of standard deviation (SD)] threshold on the deviation of the in situ SST from a climatological or analysis reference field (e.g., Llewellyn-Jones et al. 1984; Kilpatrick et al. 2001; Brisson et al. 2002). As in situ SST QC methods evolved toward being more sophisticated and systematic, more-advanced QC algorithms were established (e.g., Slutz et al. 1985; Lorenc and Hammon 1988; Ingleby and Lorenc 1993; Rayner et al. 2006; Ingleby and Huddleston 2007; Woodruff et al. 2008; Beggs et al. 2012; Atkinson et al. 2013; Xu and Ignatov 2014; Kent et al. 2019). Details of such QC methods will be introduced in section 2.

The National Oceanic and Atmospheric Administration (NOAA) is responsible for a wide range of operational and reprocessed satellite and blended SST products (Xu and Ignatov 2010, 2014). To support their accurate and consistent Cal/Val, the in situ SST Quality Monitor system (*i*Quam) was established in 2009 (Xu and Ignatov 2014; www.star.nesdis.noaa.gov/socd/sst/iquam). The *i*Quam collects in situ SSTs from as many platforms and data producers as possible, and performs an advanced, flexible, and unified community consensus QC on all data. Some input data come with their own QC. These external quality flags (QFs) are not applied in *i*Quam, but retained in the output files, for completeness

Corresponding author: Haifeng Zhang, Haifeng.zhang@noaa.gov

DOI: 10.1175/JTECH-D-20-0203.1

and further use by the interested users. Last, these QCed data are monitored online, and distributed to SST data producers and users.

Overall, the *i*Quam QC proved efficient and accurate and well suited for many Cal/Val and data assimilation applications (e.g., Xu and Ignatov 2010, 2014, 2016). However, some issues have been raised, such as the discarding of good quality SSTs when large diurnal warming events occur (e.g., Zhang et al. 2019). As a first step toward improving the *i*Quam QC, this study compares the performance of the *i*Quam QC with other QCs, available from some of the external data sources. This paper is organized as follows: section 2 introduces the datasets and different QC methods, section 3 presents the results of comparisons between *i*Quam and other QC schemes, and section 4 concludes the paper by discussing the results and outlines future work.

2. Datasets, QC schemes, and method

a. Datasets

Since its establishment in 2009, *i*Quam has been regularly updated. For each newer version, more in situ measurements from different platforms are being incorporated into the system, among other improvements. The current *i*Quam version v2.1 incorporates in situ SSTs from the following platforms: conventional ships (e.g., via bottles or engine room intake waters), drifting buoys, tropical (T-) and coastal (C-) moorings, Argo floats (Roemmich et al. 2009), high-resolution drifters (i.e., typically featured by a 0.01-K report resolution; e.g., Le Menn et al. 2019; Poli et al. 2019), Integrated Marine Observing System (IMOS) ships (which employ high quality hull-sensor measurements; Beggs et al. 2012), and Coral Reef Watch (CRW) coastal buoys (M. Eakin and E. Geiger 2015, personal communication). Multiple data for the same platform may be extracted from different data sources produced at different agencies, to ensure data completeness and best quality, at the expense of some redundancy. For example, commercial ships, drifting buoys, and T- and C-moorings in *i*Quam v2.1 are obtained from at least three sources during the satellite SST era [which started in 1981, when the first Advanced Very High Resolution Radiometer second-generation instrument (AVHRR/2), with a split-window capability, was launched on board the *NOAA-7* satellite]: the International Climate Ocean–Atmosphere Dataset (ICOADS; Freeman et al. 2017), the NOAA National Centers for Environmental Prediction (NCEP) GTS near-real-time stream, and the U.S. Global Ocean Data Assimilation Experiment (GODAE) Fleet Numerical Meteorology and Oceanography Center (FNMOC). For more details, refer to Xu and Ignatov (2014) and the *i*Quam website (https://www.star.nesdis.noaa.gov/socd/sst/iquam/about.html).

In *i*Quam, only some input datasets come with relatively complete and well-documented QC, including ICOADS, Argo, and IMOS ships. The *i*Quam QC is evaluated in this study against these three external QC schemes, using measurements from the respective platforms. Therefore, only these datasets and their QC schemes are introduced here.

1) ICOADS DATA

The ICOADS is a global ocean surface and marine meteorological dataset, formed by merging many U.S. and international data sources, obtained from ships, moored and drifting buoys, coastal stations, and other marine platforms (e.g., Slutz et al. 1985; Woodruff et al. 2011; Freeman et al. 2017). Variables in ICOADS include SST, air temperature, wind direction and speed, pressure, humidity, and cloudiness. ICOADS release 3 (R3.0) is the largest collection of surface marine observations spanning from 1662 to the present day (Freeman et al. 2017, 2019). ICOADS R3.0 comprises two archives: delayed-mode (DM; prior to 31 December 2014) and near–real time (NRT; from 1 January 2015 to the present). All ICOADS R3.0 data ingested in *i*Quam are downloaded from the National Center for Atmospheric Research (NCAR) without using any filters to exclude data.

2) ARGO DATA

Argo is a global array of free-drifting and profiling floats for observing temperature, salinity and currents in the upper 2000 m of the ocean (Roemmich et al. 2009). Between 3500 and 4000 Argo floats are active at the time of this writing, covering the global oceans from 60°N to 60°S (with the exception of marginal seas and coastal areas) near uniformly with a roughly 3° latitude/longitude (300 km) spacing, on average. Typically, Argo floats do not sample above the ~4-m depth, where the pump shuts off to protect the salinity sensor. Because of a strong interest in the satellite SST and salinity communities, a limited number of new experimental Argo floats have been deployed since 2008 that carry also an auxiliary surface temperature and salinity (STS) conductivity–temperature–depth (CTD) sensor to sample the top layer of the ocean up to approximately 20-cm depth at steps of ~10 cm (Anderson and Riser 2014). In *i*Quam, Argo data are downloaded once daily from three different sources: the NOAA National Centers for Environmental Information (NCEI; https://data.nodc.noaa.gov/argo/gadr/data/); the Institut français de recherche pour l'exploitation de la mer (IFREMER; Brest, France); and the FNMOC, mainly for fallback and redundancy. Our experience suggests that the data holdings in these three centers are similar but not identical.

3) IMOS SHIP DATA

From 2008, the Australian IMOS has enabled SST data to be supplied in real time (RT; within 24 h) from hull-contact sensors on commercial vessels and water injection sensors on research vessels in the Australian region. Six vessels use hull-contact temperature sensors, and nine other vessels use thermistors located in water intakes. The IMOS ships are intended to provide in situ SST observations in the sparsely sampled regions, such as the coastal areas of Australia and the Southern Ocean. In contrast to traditional ship SST observations, which are made through engine water intakes and often suffer from the heating in the engine room, the IMOS ship SST measurements obtained from hull-contact sensors have exhibited overall quality that is similar to that of drifting and moored buoys (Beggs et al. 2012).

b. Quality control methods

1) *i*QUAM QC

The *i*Quam QC represents one of the most inclusive and systematic in situ SST QC systems (Xu and Ignatov 2014), as briefly summarized here. The QC includes five binary checks: 1) duplicate removal, which deletes duplicates that arise from multiple transmissions or merging multiple datasets; 2) plausibility check, which evaluates whether a value is at a reasonable geolocation (e.g., not over land), as well as within a physical SST range (i.e., from −2° to +35°C); 3) platform track check, which makes sure 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 speed (i.e., $<17\ m\ s^{-1}$ for ships and $<3.75\ m\ s^{-1}$ for drifters; Xu and Ignatov 2014); 4) SST spike check, which employs the same logic and algorithm as the track check except that the maximum SST gradient in space and time is checked, instead of travel speed; and 5) identifier (ID) check, performed to determine whether the ID field of a measurement is valid. More advanced QC includes two Bayesian checks, namely, a reference check (RC) and a buddy check (BC). Rather than setting a simple threshold as in some previous RC checks, the approach by Lorenc and Hammon (1988) is employed in *i*Quam. The RC proves to be the most critical of all *i*Quam checks, which screens out the most in situ data (Xu and Ignatov 2014). In *i*Quam, two gap-free reference SST fields are used: the Reynolds version-2 optimal interpolation (OI) global 0.25° daily level-4 (L4) analysis (AVHRR only), which dates back to September 1981 (Reynolds et al. 2007), and the Canadian Meteorolgical Centre (CMC) L4 product, comprising two versions: 0.2°-resolution v1 (1 September 1991–31 December 2015) and 0.1° v2 (1 January 2016–present) (Brasnett and Surcel-Colan 2016). Last, a BC is performed on top of the RC. The BC updates the posteriori probability of gross error by incorporating information from nearby measurements (aka buddies). It may compensate for some RC deficiencies (e.g., resulting from possible inaccuracies in the reference field). Its effectiveness also depends on the density of the neighbors and their data quality (Xu and Ignatov 2014; Kent et al. 2019).

Note that these reference fields, which blend satellite and in situ SST measurements, are intended to be foundation SSTs, that is, at a depth where no diurnal warming signal is present, which is often deeper than the depths where many in situ platforms take their measurements. Also, in situ data are point measurements, as opposed to L4 space-average measurements. Although perfect agreement between an in situ measurement and the reference field is therefore not realistic, the latter typically is a good first guess for the former. The two L4 products employed in *i*Quam were selected because of their long record and quality, respectively, and their mutual complementarity and reasonable redundancy.

The quality level (QL) in *i*Quam is defined using both reference fields. Note that both RC and BC produce continuous quality indicators that serve as the probabilities of gross error (PGE; or P). Two PGE indices are used here: 1) PGE without BC, $P(\mathrm{RC})_n$, and 2) PGE $P(\mathrm{QC})_n$, where n represents the reference field ($n = 1$ for Reynolds, and $n = 2$ for CMC). The definition of QL 4 and 5 is as follows:

- if $P(\mathrm{RC})_1 \le 0.5$ and $P(\mathrm{QC})_1 \le 0.1$ and $P(\mathrm{RC})_2 \le 0.5$ and $P(\mathrm{QC})_2 \not\le 0.1$ then QL = 5;
- else if $P(\mathrm{RC})_1 \le 0.5$ and $P(\mathrm{QC})_1 \le 0.5$ and $P(\mathrm{RC})_2 \le 0.5$ and $P(\mathrm{QC})_2 \le 0.5$ then QL = 4.

Note that all providers' QFs (e.g., ICOADS, Argo, and IMOS) are reported in *i*Quam files, along with the *i*Quam QLs, so that users have a choice of using one of the two, or both.

2) ICOADS QC

Similar to *i*Quam, ICOADS R3.0 delayed-mode (DM) QC includes several prescreening checks, such as the duplicate elimination, land lock, and track checks. In addition, there are three major QF groups, as follows:

1) Two flags, "source ID" and "source exclusion," are set in ICOADS based on legacy rules to remove data from known questionable sources during particular time periods.
2) Trimming flags are the result of ICOADS' own QC, defined as normalized deviations of in situ SST from its background average (calculated as a running median within monthly 2° × 2° boxes). The deviations are normalized against the running lower and upper median deviations ("sigmas") and digitized to several discrete levels.
3) External flags inherited from the original sources include 1) the National Climatic Data Center (NCDC) QF (produced by the former NOAA NCDC, currently a part of the newly formed NOAA NCEI), and 2) the World Ocean Database (WOD) Ocean Station Data (OSD) flag (https://icoads.noaa.gov/e-doc/imma/R3.0-imma1_short.pdf).

ICOADS NRT QC is less comprehensive than DM and is not used in this study.

3) ARGO QC

There are up to 22 checks included in Argo QC (Argo Data Management Team 2019; Wong et al. 2020). Basic QC comprises several checks, including plausibility, platform speed ($<3\ m\ s^{-1}$), SST spikes, and SST physical range (from −2.5° to 40.0°C) (Wong et al. 2020). In several regions, such as the Mediterranean Sea and the Red Sea, Argo data will go through a further regional temperature range test. There are also other checks unique to Argo floats, such as the deepest-pressure and near-surface unpumped CTD salinity tests (Argo Data Management Team 2019). The overall Argo QL may take any of 10 values (0–9): 0 = no QC performed; 1 = good data; 2 = probably good data; 3 = probably bad data but may be correctable; 4 = bad data; 5 = value changed; 6 and 7 = not used; 8 = estimated value; and 9 = missing value. Of the 10 values, QC flags of 1, 2, 5, or 8 are claimed to be good data (Argo Data Management Team 2019; Wong et al. 2020). The Argo DM data are generated up to several months after the NRT data and overwrite those. The current *i*Quam processing does not update NRT with DM data.

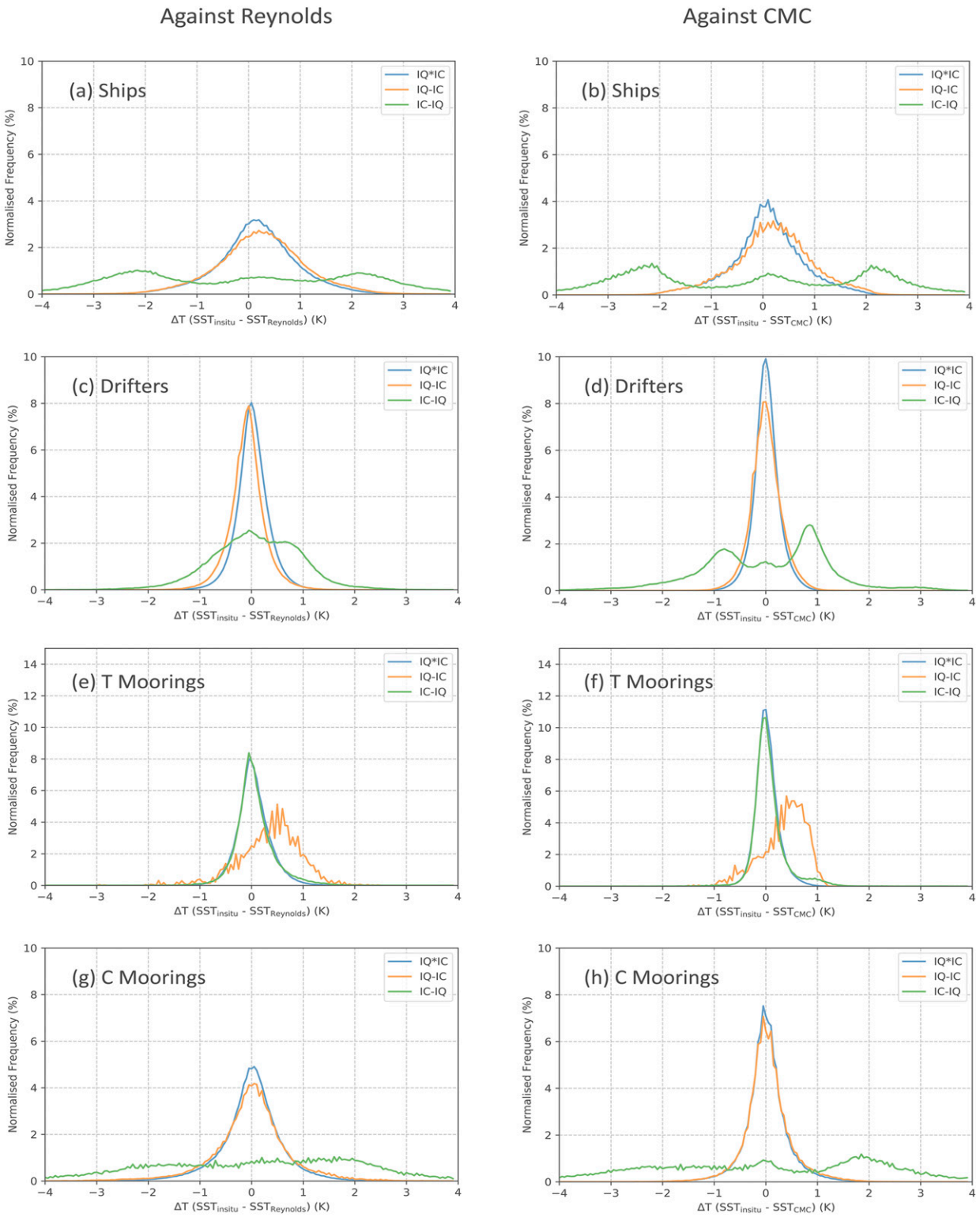

FIG. 1. Normalized frequencies of $\Delta T = T_{\text{in\_situ}} - T_{\text{ref}}$ for four ICOADS platform types from 2012 to 2014, stratified by *i*Quam/ICOADS QFs: (left) reference (ref) = Reynolds and (right) reference = CMC.

TABLE 1. Statistics from Fig. 1, including percentage of number of observations (NOB), mean (μ) ± SD (σ)and robust SD (RSD) of ΔT against Reynolds and CMC reference fields, are shown for the three categories between *i*Quam and ICOADS QC for each platform.

Platform/category	Percentage of NOB	Against Reynolds $\mu \pm \sigma$ (RSD) (K)	Against CMC $\mu \pm \sigma$ (RSD) (K)
		Ships (100% = 3 409 948)	
IQ*IC	71.2	0.19 ± 0.79 (0.66)	0.11 ± 0.70 (0.57)
IQ-IC	19.2	0.26 ± 0.85 (0.76)	0.22 ± 0.77 (0.68)
IC-IQ	9.6	−0.08 ± 2.42 (2.93)	−0.23 ± 2.48 (3.47)
		Drifters (100% = 25 603 597)	
IQ*IC	89.9	0.05 ± 0.32 (0.26)	0.04 ± 0.24 (0.21)
IQ-IC	6.1	−0.06 ± 0.35 (0.27)	0.03 ± 0.30 (0.26)
IC-IQ	4.0	0.11 ± 0.88 (0.85)	0.04 ± 1.26 (1.24)
		T-moorings (100% = 1 728 170)	
IQ*IC	87.7	0.06 ± 0.32 (0.27)	0.05 ± 0.22 (0.18)
IQ-IC	0.3	0.39 ± 0.59 (0.52)	0.37 ± 0.43 (0.40)
IC-IQ	12.0	0.09 ± 0.38 (0.27)	0.09 ± 0.33 (0.20)
		C-moorings (100% = 6 167 468)	
IQ*IC	90.8	0.03 ± 0.60 (0.44)	0.04 ± 0.42 (0.29)
IQ-IC	7.2	0.02 ± 0.74 (0.52)	0.06 ± 0.45 (0.32)
IC-IQ	2.0	0.26 ± 2.40 (2.35)	0.07 ± 2.75 (2.83)

The current study incorporates both NRT and DM Argo data as they are reported in *i*Quam v2.1.

4) IMOS QC

IMOS QC consists of several binary QC checks, including duplicate removal, plausibility, and track checks, and an RC that flags an SST value as ''bad'' if it exceeds 3°C above or below the most recent Australian Bureau of Meteorology operational daily SST analysis, produced by blending satellite and in situ SSTs. Beggs et al. (2012) indicate that this RC is simply a warning flag and is not used to reject data.

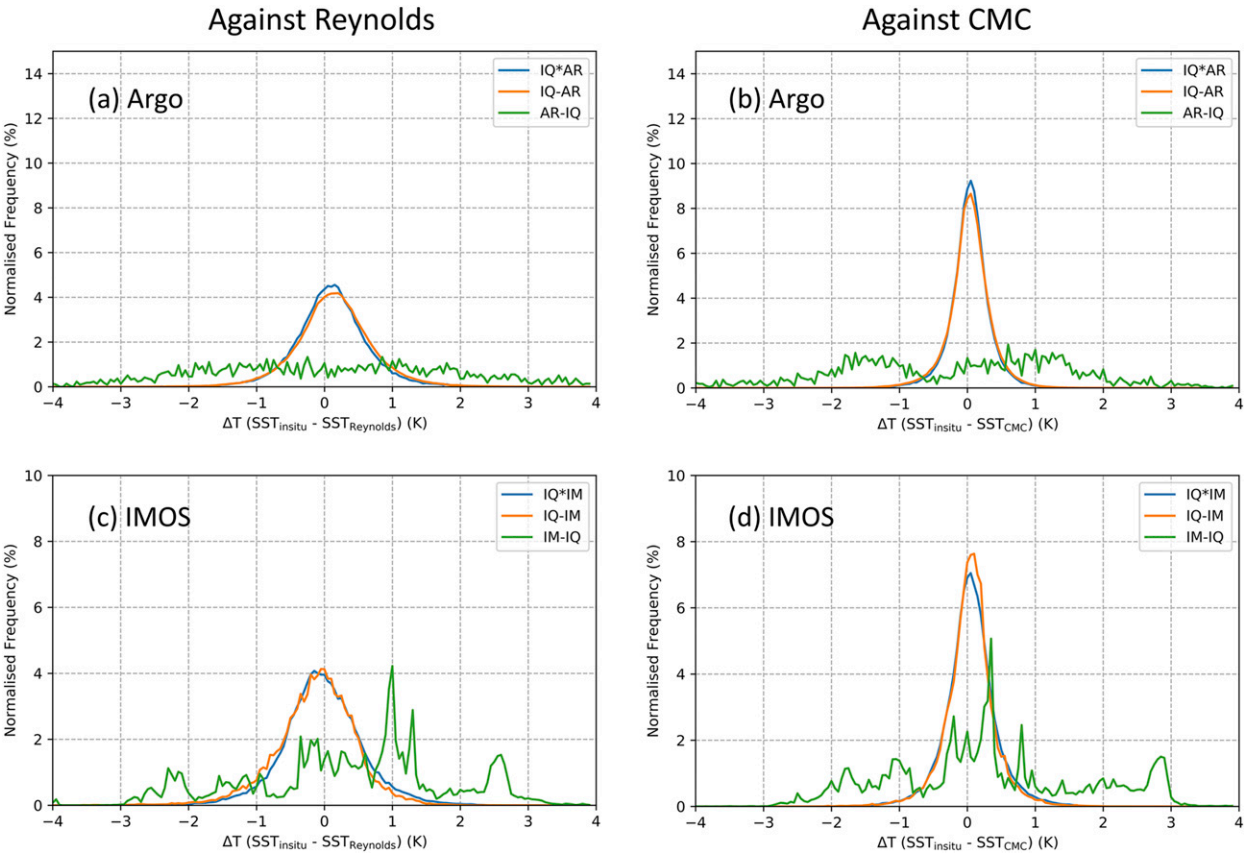

FIG. 2. As in Fig. 1, but for Argo floats and IMOS ships.

TABLE 2. As in Table 1, but for Argo floats and IMOS ships.

Platform/category	Percentage of NOB	Against Reynolds $\mu \pm \sigma$ (RSD) (K)	Against CMC $\mu \pm \sigma$ (RSD) (K)
		Argo (100% = 643 666)	
IQ*AR	20.2	0.13 ± 0.55 (0.45)	0.06 ± 0.30 (0.23)
IQ-AR	79.4	0.17 ± 0.58 (0.48)	0.06 ± 0.32 (0.24)
AR-IQ	0.4	0.03 ± 2.58 (2.20)	−0.33 ± 2.24 (1.94)
		IMOS (100% = 3 675 244)	
IQ*IM	78.1	0.00 ± 0.59 (0.51)	0.08 ± 0.38 (0.30)
IQ-IM	21.2	−0.10 ± 0.59 (0.52)	0.06 ± 0.37 (0.27)
IM-IQ	0.7	−0.37 ± 1.46 (1.20)	0.32 ± 1.42 (1.44)

5) CMS DRIFTER BLACKLIST, ARGO GRAY LIST, AND *i*QUAM PERFORMANCE HISTORY CHECK

Le Centre de Météorologie Spatiale (CMS) in France maintains an operational blacklist (BL) of drifting buoys, which is essentially a platform-specific QC detecting abnormal platforms, rather than individual measurements. The CMS BL is generated from comparison of drifter SSTs with satellite SSTs from available MetOp and NOAA satellites processed in Lannion, France, and is automatically updated every 10 days. The main criterion for acceptance in the CMS BL is a bias larger than 1.5 K relative to nighttime satellite SSTs from one satellite (for two consecutive 10-day periods) or from two satellites (for one 10-day period; e.g., Marsouin et al. 2015).

Similarly, if an Argo float is deemed to report unrealistic data, it is added to the "gray list" (GL) maintained by the two global data assembly centers (GDACs; GODAE in Monterey, California, and Coriolis in Brest, France). The decision to add a float to the GL comes from the principal investigator or from the DM operator (Wong et al. 2020). A float is added to the GL when its sensor drifts too strongly to be adequately corrected for in the RT processing, or when the sensor is deemed to be malfunctioning. Note that the Argo GL is only set in RT files. When an anomalous float is dead or the offending data have been "adjusted" in DM, it is removed from the GL (Wong et al. 2020). See more details about the "adjustment" in Wong et al. (2020).

Similar to the BL/GL, an additional "performance history" (PH) QC check was introduced in *i*Quam v2.1. Its major premise is that if a platform has a track record of poor performance in the near past, it is likely that its sensor is subject to long-term malfunction or failure, so that all measurements following the period of such poor performance are flagged as suspicious. The PH check is performed daily based on the results of regular *i*Quam QC checks. If the *i*Quam error rate (i.e., percentage of measurements flagged as bad by the regular *i*Quam QC, excluding the PH) of each individual platform ID for a prior 10-day window is larger than the preset threshold (currently, 50%), then the corresponding ID and all its measurements for that day are flagged by the PH check as "suspicious." Note that, unlike the drifter BL or Argo GL, the *i*Quam PH is applied to all platforms reported in *i*Quam and is recalculated daily.

c. *Comparison of* i*Quam QC with data providers' QCs: "Confusion matrix"*

Note that flagging a measurement as "good" (="pass QC") is done differently in each QC scheme. In *i*Quam, in addition to the positive binary checks, a good value should have $P(\mathrm{QC}) \leq 0.5$

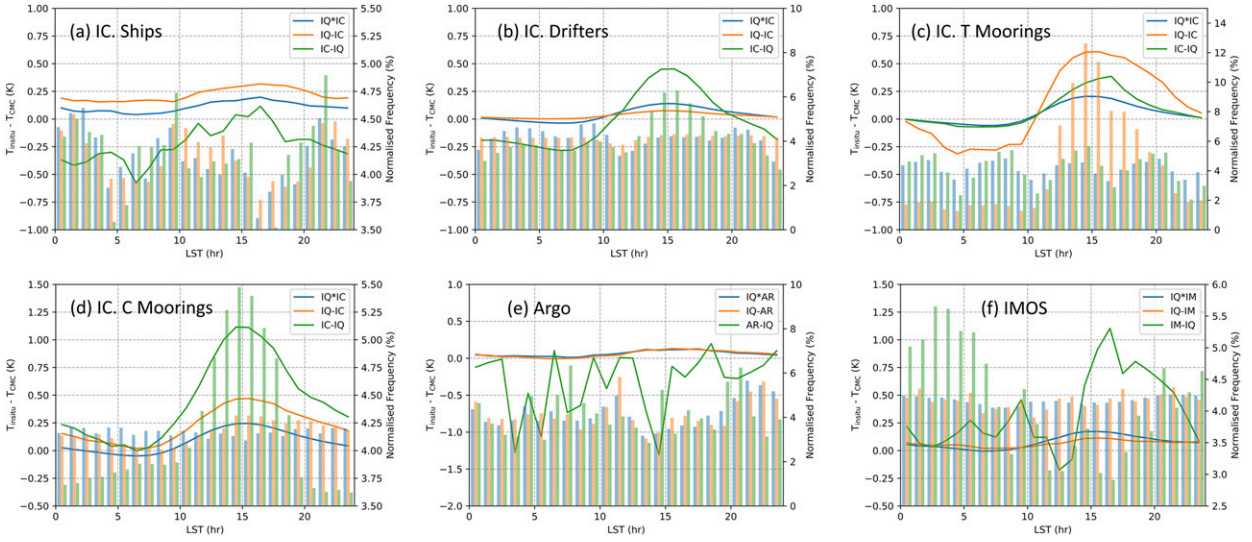

FIG. 3. Dependence of $\Delta T = T_{\mathrm{in\_situ}} - T_{\mathrm{CMC}}$ on LST (lines; left Y axis) and corresponding normalized frequencies (bars; right Y axis) for different platform types, in three QC subsamples: IQ*X, IQ-X, and X-IQ.

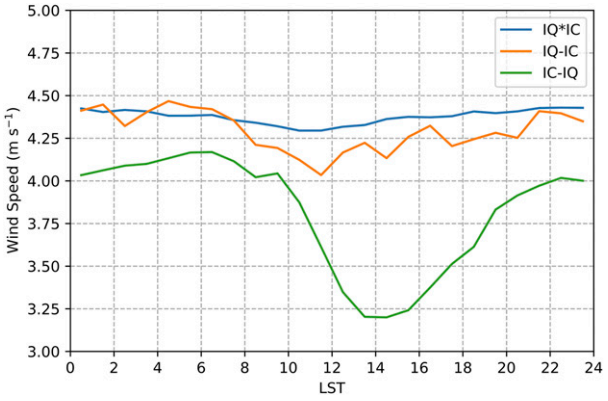

FIG. 4. Wind speed for ICOADS C-moorings as a function of LST, stratified by three *i*Quam/ICOADS QC subsamples.

or 0.1 when compared with a Reynolds or CMC reference field, respectively (Xu and Ignatov 2014). In ICOADS, a measurement is considered to have ''passed'' when binary flags are positive and the trimming flag is within the mean $\pm$ 2.8σ range. In Argo QC, a ''pass'' is assigned when the value passes all checks and has been properly adjusted, with its vertical sampling scheme being either primary (within 10-m depth) or near-surface. For IMOS ships, a value passes QC when all tests have positive outcomes.

In this study, *i*Quam QLs are compared with the external QFs, using the ''confusion matrix'' method. When comparing *i*Quam (IQ) and external (X) QC, the sample is defined as the set where SST passes at least one of them, and broken into three subsamples as follows:

1) IQ*X: SST passes *both* *i*Quam *and* external (X) QCs (''*i*Quam/X intersection''),
2) IQ-X: SST passes *i*Quam QC but fails external (X) QC (''*i*Quam complement''), and
3) X-IQ: SST fails *i*Quam QC but passes external (X) QC (''X complement'').

All analyses are based on these three ''confusion matrix'' categories. If *i*Quam and X QCs are consistent, then 100% should fall in the intersection category, with 0% in the two complements. If a complement is nonzero, but its SST ''errors'' are indistinguishable from the intersection, then it is a ''false alarm'' (i.e., a good measurement flagged as bad). Error is defined as SD or robust SD (RSD; defined as 1.5 times the median of the absolute differences from the median biases) of in situ minus reference SST (from either of the two global L4 analyses). If SST ''errors'' in a complement are significantly larger than in the intersection, then it is a ''leakage'' (i.e., bad measurement flagged as good). For example, if the values in the *i*Quam complement have larger ''errors,'' it means the *i*Quam QC method failed to pick out those bad measurements and wrongfully flag them as good, hence a ''leakage'' in the *i*Quam QC method. (Note that one cannot rule out that both QC methods may improperly flag a valid observation as ''bad,'' which is then not included in this study. However, such occurrences, if they occur, are deemed to be infrequent and are expected to have only an insignificant effect on the results and overall conclusions of this study).

Considering that ICOADS R3.0 DM data are only available through the end of 2014, a 3-yr period from 2012 to 2014 was used in the *i*Quam/ICOADS QC comparison. For Argo and IMOS, three different years from 2017 to 2019 were used to take advantage of increased data volumes.

3. Results

a. Statistics

The performance of *i*Quam QC is first evaluated by analyzing the frequency distributions of $\Delta T = T_{\text{in\_situ}} - T_{\text{ref}}$, in the

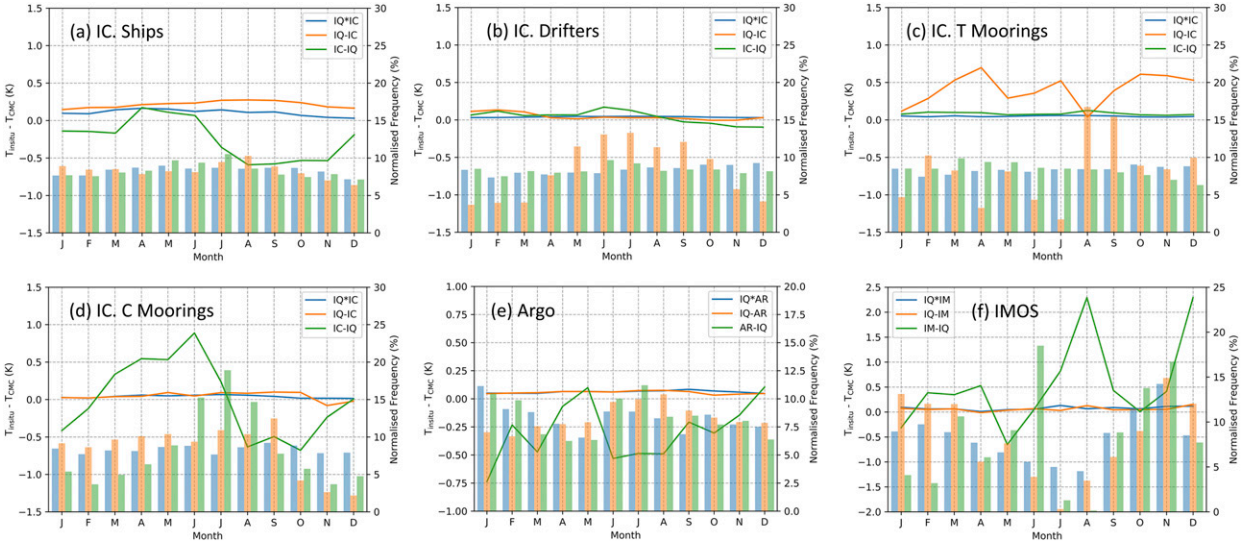

FIG. 5. As in Fig. 3, but as a function of month.

three *i*Quam/ICOADS QC subsamples: IQ*IC, IQ-IC, and IC-IQ. The four platform types in ICOADS, ships, drifters, and T- and C- moorings, are investigated separately. The normalized frequency distributions of ΔT (against both Reynolds and CMC L4 SSTs) are shown in Fig. 1, and the corresponding statistics are summarized in Table 1.

Several observations emerge from Fig. 1 in conjunction with Table 1. First, traditional ship SST measurements have the least favorable statistics, out of the four platforms. This observation holds against both Reynolds and CMC SSTs. Their SD/RSDs are larger, as attested to by wider distributions in Figs. 1a and 1b, and their less Gaussian shapes. Drifters (which have the largest number of observations) and T-moorings have comparable, and the smallest, SD/RSDs, of all four platform types. The IQ*IC category accounts for the majority of all the observations (ranging from ~71% to ~91% for different platforms) and has the best statistics for all platforms. For nearly all cases, the *i*Quam complements IQ-IC (i.e., when *i*Quam QC passes and ICOADS QC rejects the data) have shapes (in Fig. 1; orange lines) and statistics (in Table 1) similar to the intersection IQ*IC (Fig. 1; blue lines), suggesting that these values are of reasonably fine quality, despite the fact that they fail the ICOADS QC. However, the ICOADS complement IC-IQ (i.e., data passing ICOADS QC but failing *i*Quam QC; Fig. 1, green line) usually shows broader non-Gaussian distributions, with larger SD/RSDs. The only exception is the T-moorings, for which the shapes of IQ*IC and IC-IQ histograms are similar, just with a wider and positively shifted IQ-IC peak (Figs. 1e,f). The likely cause of the positive bias observed in Figs. 1e and 1f may be diurnal warming, as discussed later. The small IQ-IC data volume (only 0.3% of $>1.7 \times 10^6$ data points) may also be a contributing factor to the unstable/unreliable statistics in this subsample. In all cases, the CMC reference appears a better proxy of in situ SSTs, consistent with similar observations made in previous studies (Dash et al. 2012; Martin et al. 2012; Saha et al. 2012).

Similar QC comparisons have been performed for *i*Quam with Argo floats and *i*Quam with IMOS ships (Fig. 2 and Table 2). For both platforms, the *i*Quam complements (IQ-AR and IQ-IM) have nearly identical distribution shapes (Fig. 2; orange lines) and statistics to those of IQ*AR and IQ*IM (Fig. 2; blue lines), respectively, suggesting that some good quality data may have been discarded by the Argo or IMOS QCs. In contrast, the Argo and IMOS complements have wider distributions (with SD/RSDs that are 3–4 times as large) with non-Gaussian shapes (Fig. 2; green lines). Note that unlike all other platforms, the IQ-AR (not IQ*AR) accounts for the majority of observations (79.4%). This is because the current implementation of *i*Quam, in which the RT data (which come with no QC) are ingested first and are never replaced with the DM data, leads to a much larger portion of RT data in the output files. Work is under way to replace RT with DM data as soon as the latter become available. If only DM Argo data were considered in the comparison, then the IQ*AR intersection would have constituted ~90%. A closer agreement of the CMC SSTs with in situ data is consistently observed. For the remainder of this paper, only the CMC reference is used.

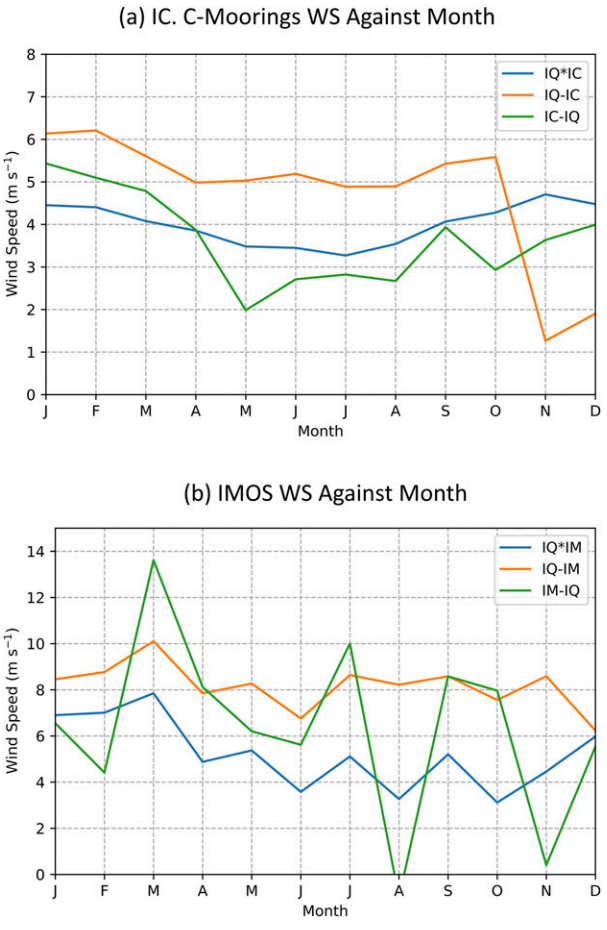

FIG. 6. Wind speed from (a) ICOADS C-moorings and (b) IMOS ships as a function of month, stratified by *i*Quam/ICOADS and *i*Quam/IMOS QC subsamples.

b. *Dependence on environmental conditions*

1) DIURNAL CYCLE

To gain further insight into the results of section 3a, the $\Delta T = T_{\text{in\_situ}} - T_{\text{CMC}}$ was plotted as a function of local solar time (LST) in Fig. 3 for all platforms and QC categories. As expected, for most platforms, the ΔT trends for *i*Quam complements (orange lines) follow closely those of IQ*X (blue lines). A mild diurnal variation (DV) pattern is observed, with the ΔTs reaching maximum at ~1400–1600 LST, for both subsamples. The only exception is the T-moorings, for which the *i*Quam complement (IQ-IC; orange line in Fig. 3c) shows a stronger DV shape (cf. slightly positive IQ-IC peaks in Figs. 1e,f). Also, note that a relatively large portion of data in *i*Quam complements (orange bars) falls between 1200 and 1800 LST (Fig. 3c). Some measurements here that fail ICOADS QC but pass *i*Quam QC are likely good data with larger DV amplitudes (up to ~0.7 K).

However, for other platforms the DV signal is not always picked up by *i*Quam QC. For instance, the ICOADS complements for drifters (Fig. 3b) and C-moorings (Fig. 3d) (IC-IQ; green lines) have a typical DV pattern with amplitudes of

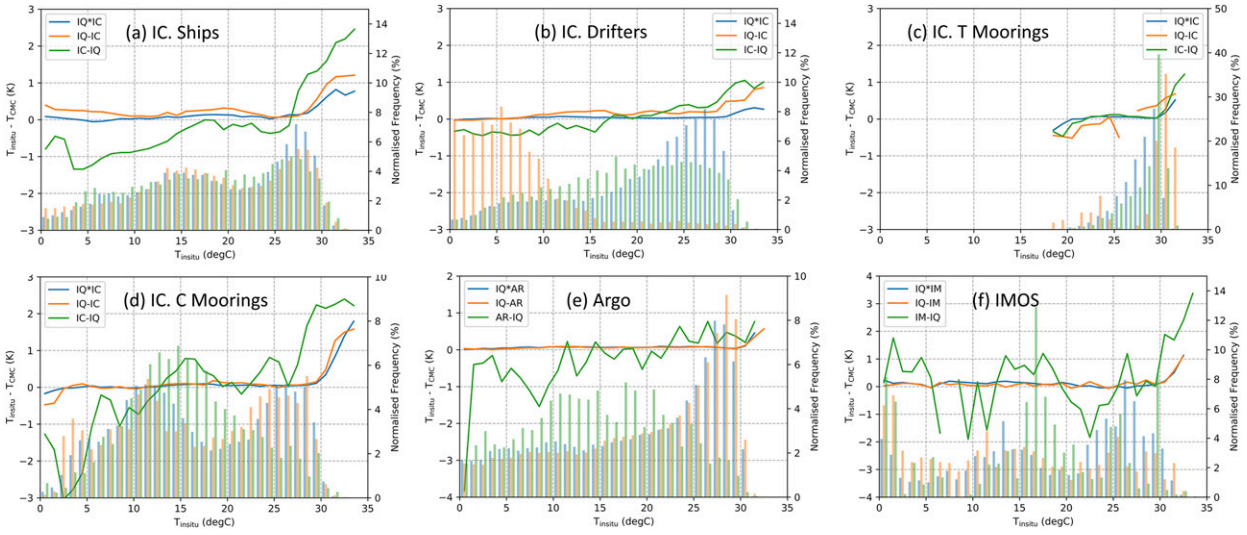

FIG. 7. As in Fig. 3, but as a function of $T_{\text{in\_situ}}$.

~0.45 and 1.10 K, respectively. A large portion of IC complements for C-moorings is found between 1200 and 1800 LST (Fig. 3d; green bars). However, these values, which appear to be good quality data with larger DV, do not pass the *i*Quam QC.

To further confirm that these are diurnal warming signals, the wind speed dependence for the C-moorings is plotted as a function of LST (Fig. 4). The wind speed is chosen since it is the most important factor in propagating insolation through the surface and controlling the DV amplitude (e.g., Fairall et al. 1996; Kawai and Wada 2007; Zhang et al. 2016). Clearly, in comparison with the other two categories, the winds for the IC complement (green line) are calmer between 1200 and 1800 LST, explaining the larger-amplitude DV in this subsample.

2) SEASONAL CYCLE

Seasonal evolution of ΔT is shown in Fig. 5. For all platforms, except for T-moorings (Fig. 5c), the IQ*X and *i*Quam complement (IQ-X) behave similarly, with both lines following each other closely. The very small data volume of IQ-IC for T-moorings may be partially responsible for the unexpected drift of the IQ-IC line (orange). The external QC complements (X-IQ; green lines) show strong seasonal fluctuation, indicating that their quality may be less stable in time.

A further analysis of wind speed as a function of month is conducted to see if the seasonal SST DV is a contributor to the behaviors of external QC complements (Fig. 6). Only ICOADS C-mooring and IMOS results are shown because of their larger X-IQ ΔT monthly variations (Figs. 5d,f; green lines). Figure 6a shows that the wind speed from C-moorings keeps decreasing from January to May for IC-IQ (green line), which correlates with the increasing trend of ΔT (green line) in Fig. 5d. Similarly, both the maximum ΔT and the minimum wind speeds for the IMOS complement (green lines in Fig. 5f and Fig. 6b, respectively) are seen in August. Although the numbers of valid wind speed observations may be small for some months, this trend still suggests that seasonal SST DV could be partially responsible for the fact that the measurements pass external QC but fail *i*Quam QC.

3) SEA SURFACE TEMPERATURE TRENDS

Next, the performance of *i*Quam QC under different temperature conditions is studied (Fig. 7). For all platforms, the IQ-X measurements show fine quality under all temperature conditions, with the lines (orange) following closely those of IQ*X (blue). The external QC complement measurements, however, display larger ΔT variations. Another pattern that stands out is that for $T_{\text{in\_situ}} > 28$°C, all lines start to have positive ΔTs, which are very likely to be diurnal variation signals at stratified waters. Note that the amplitudes of ΔTs for Argo are the smallest, given that the majority of Argo measurements are taken at a depth (~5 m) more representative of the reference fields. Very warm waters (>28°C) are more often

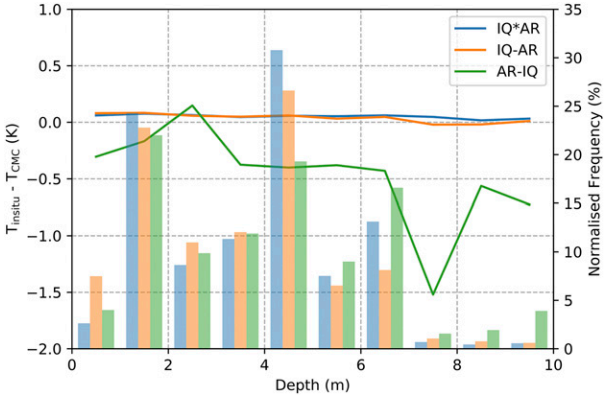

FIG. 8. Dependence of $\Delta T = T_{\text{in\_situ}} - T_{\text{CMC}}$ on depth for Argo data (left *Y* axis), and corresponding normalized frequencies (right *Y* axis), by three QC categories.

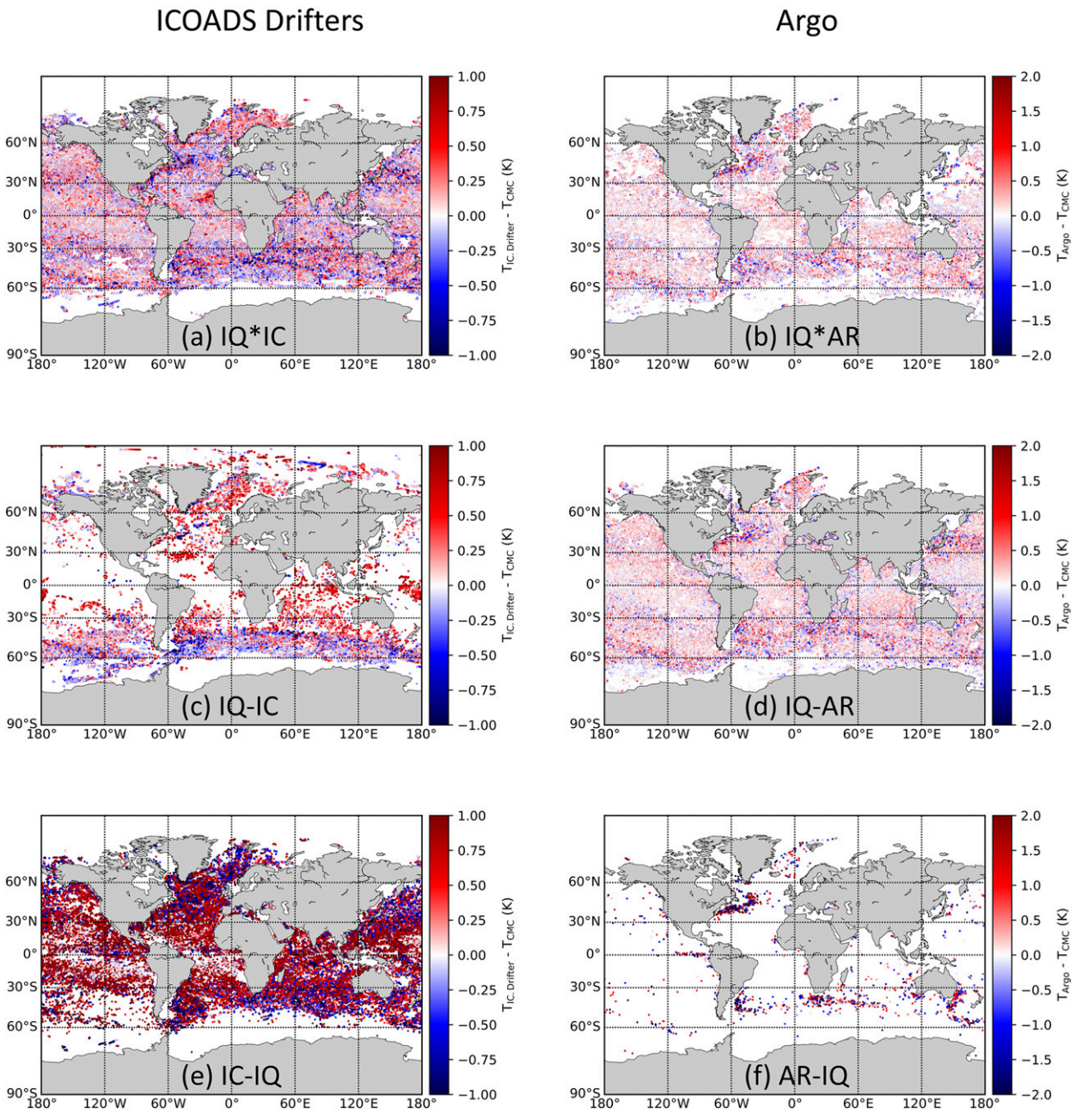

FIG. 9. Spatial distribution of ΔT from (a),(c),(e) ICOADS drifters and (b),(d),(f) Argo floats, by QC categories: (top) IQ*X, (middle) IQ-X, and (bottom) X-IQ.

observed in the tropical regions, especially in the tropical warm pool (western Pacific and eastern Indian Oceans), where DV events may occur frequently due to a combination of calm winds and strong insolation (Zhang et al. 2016). The external QC complement values typically have much larger ΔTs for $T_{in\_situ} > 28°C$ (Figs. 7a,d,f; green lines). Again, both less stable data quality and DV are contributors. Finally, it is noticed that a larger portion of IQ complement values are measured at lower temperature conditions ($T_{in\_situ} < 10°C$) for ICOADS drifters (orange bar in Fig. 7b). A spatial distribution indicates that these drifters are seen more in the mid- to high latitudes (shown later).

4) ARGO DEPTH

Since Argo measurements within the top 10-m ocean layer are all included in *i*Quam, the behavior of *i*Quam QC on different depths is investigated by comparing with Argo QC (Fig. 8). Most Argo observations are made at 4–5-m depth, which is the uppermost layer at which the primary sensor measures temperature. Above ~4-m depth, SSTs are usually

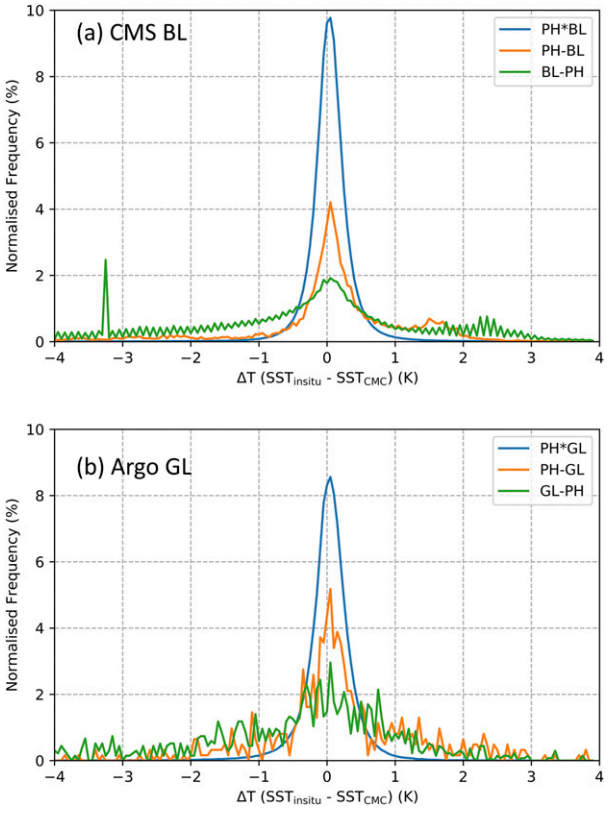

FIG. 10. Normalized frequencies of ΔT for (a) drifters and (b) Argo floats, stratified by PH/BL and PH/GL quality flags.

measured by the auxiliary sensor with a second observation number peak at 1–2-m depth. Many fewer data are collected below 7 m. The *i*Quam QC performs very well at all depths, with the average IQ-AR values being around ~0.1 K. The jumps in the Argo QC complement ΔTs (green line) observed, especially toward deeper layers (>7-m depth), may, at least in part, be due to the reduced data volume.

c. *Spatial distribution*

The performance of *i*Quam QC is now assessed spatially. Since only drifters and Argo floats have near-global coverage, their ΔT distributions for each category are plotted and shown in Fig. 9. For drifters, IQ*IC has relatively small amplitude ΔTs globally (Fig. 9a). The IQ complement measurements are mostly found in the North Atlantic Ocean, the Indian Ocean, and the Southern Ocean (Fig. 9c). While those large amplitude ΔTs found in the North Atlantic and the Indian Ocean are positive, most ΔTs over the Southern Ocean are negative. The values that pass ICOADS QC but not *i*Quam QC mostly have large amplitudes, either positive or negative, with an almost global coverage (Fig. 9e).

The ΔT distributions and amplitudes of IQ*AR (Fig. 9b) and IQ-AR (Fig. 9d) are very similar (recall that IQ-AR accounts for 79.4% of all observations). The much fewer AR-IQ measurements are observed frequently in the Gulf Stream, the Eastern Australian Current, the Brazil Current, and part of the Antarctic Circumpolar Current. All these regions are characterized by sharp horizontal and vertical temperature gradients. The amplitudes of AR-IQ ΔTs appear large (Fig. 9f). It cannot be ruled out that in such areas, the CMC reference field may underrepresent some finer temperature details, possibly affecting and degrading performance of *i*Quam QC.

d. *PH against CMS BL and Argo GL*

Since the *i*Quam PH, CMS BL, and Argo GL are all platform-specific QC checks (i.e., detecting abnormal performance of individual platforms rather than individual measurements), they are the only QC checks applied for this part with all other QC filters ignored. Drifter data are used for the comparison between PH and BL because the CMS BL is available only for drifters. Only RT Argo measurements are included for the PH-versus-GL comparison because GL does not apply for DM data. Note that if a platform passes the BL test, it means it is of good quality.

The ΔT distributions are shown in Fig. 10, and statistics are in Table 3. It is observed that *i*Quam PH is largely consistent with the CMS BL and Argo GL, with PH*BL and PH*GL accounting for 96.5% and 99.7% of all observations, respectively. Since no other QC checks are employed, the SDs are much larger when compared with those QCed drifters and Argo data, yet the RSDs are similar (Table 2). This is clearly due to the many outliers included in this analysis. The PH test appears rather strict with both PH-BL and PH-GL values making up only 0.1% of all data. Although the PH-BL (orange line in Fig. 10a) and PH-GL (orange line in Fig. 10b) ΔT distribution shapes tend to resemble those of PH*BL (blue line in Fig. 10a) and PH*GL (blue line in Fig. 10b), respectively,

TABLE 3. Statistics of ΔT for the three categories between PH/BL and PH/GL for drifters and Argo floats, respectively.

Platform/category	Percentage of NOB	Against CMC $\mu \pm \sigma$ (RSD) (K)
	CMS BL (100% = 40 245 749)	
PH*BL	96.5	0.06 ± 0.98 (0.21)
PH-BL	0.1	−3.27 ± 10.48 (1.89)
BL-PH	3.3	−1.46 ± 6.81 (2.12)
	Argo GL (100% = 660 168)	
PH*GL	99.7	−0.01 ± 1.18 (0.24)
PH-GL	0.1	−0.18 ± 3.23 (0.69)
GL-PH	0.2	−1.37 ± 4.76 (1.37)

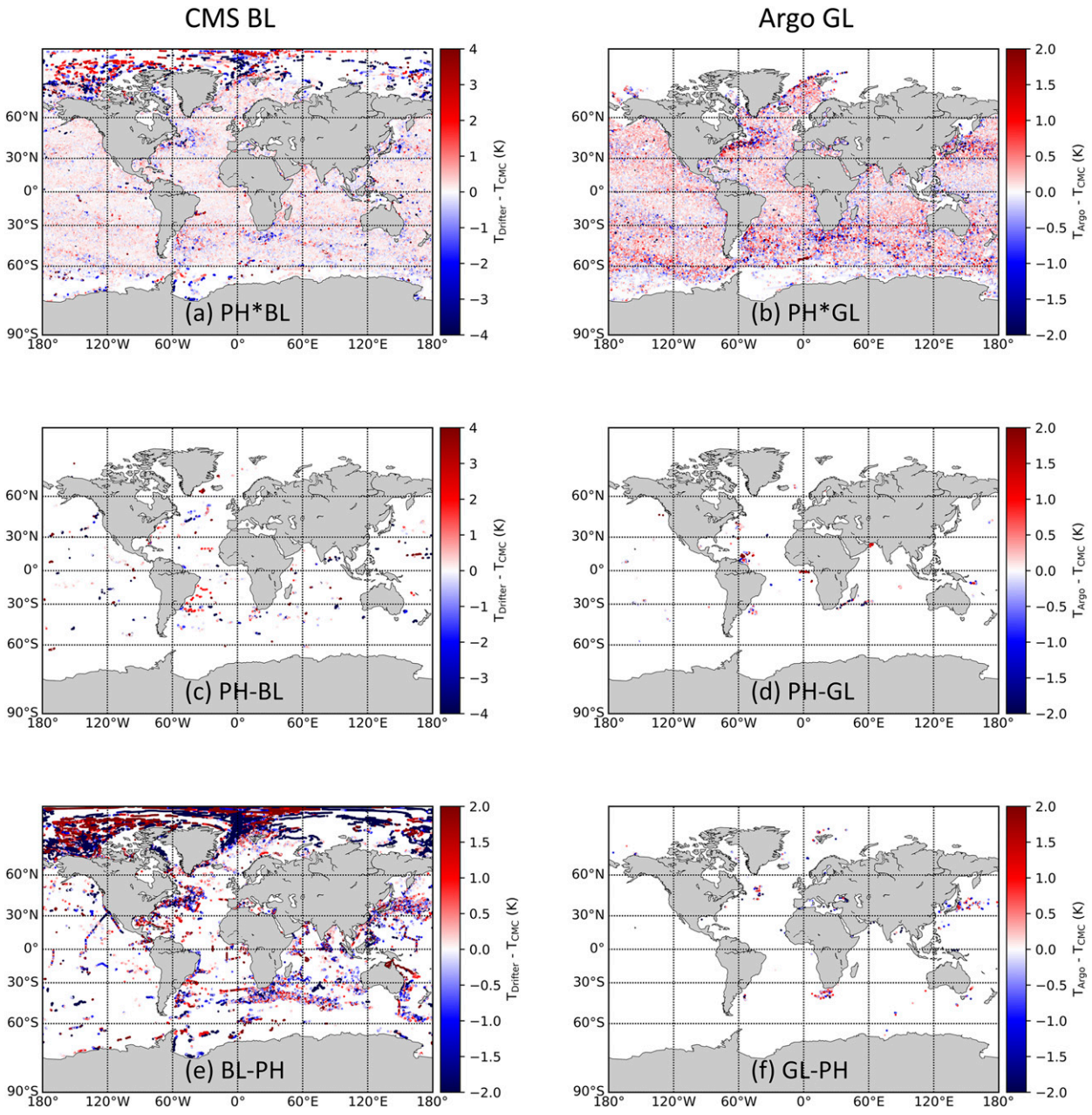

FIG. 11. Geographical distribution of ΔT for (a),(c),(e) drifters blacklist (X = BL); and (b),(d),(f) Argo floats gray list (X = GL), in three *i*Quam performance history (PH) QC subsamples: (top) PH*X, (middle) PH-X, and (bottom) X-PH.

the PH complement data may not be sufficiently representative to draw far-reaching conclusions. Although the data in the BL-PH and GL-PH complements are somewhat larger in numbers, their statistics are even less favorable.

The spatial distributions of ΔTs from all three categories for two platforms are shown in Fig. 11. The ΔTs from PH*BL (Fig. 11a) and PH*GL (Fig. 11b) are relatively uniformly distributed across global oceans with mild amplitudes. Drifters that pass BL yet fail PH tend to cluster in higher latitudes as well as in the areas with strong currents, such as the Gulf Stream, the Kuroshio, and the Eastern Australian Current (Fig. 11e). Because of the small data volumes, the platforms in the PH complements (Figs. 11c,d) and GL complements (Fig. 11f) are too sparsely distributed to display any statistically significant pattern.

4. Discussion and conclusions

Given the importance of the in situ SST measurements in satellite SST Cal/Val and data assimilation, the *i*Quam system was established at NOAA in 2009 to support such activities by collecting, quality controlling, and monitoring different in situ

observations. In this study, the performance of *i*Quam QC is compared with some of the other quality flags supplied with the in situ SST inputs ingested into the *i*Quam system, including the ICOADS R3.0, Argo, and IMOS ship measurements. Different platforms (e.g., ships, drifters, and moorings) in ICOADS are investigated separately. The QC comparisons are conducted under different environmental conditions (e.g., local times, SST conditions, and depth) and spatially across global oceans.

Overall, the *i*Quam QC has illustrated good performance for all platform types, under various observational conditions. As expected, when a value passes both QC schemes, it has the best quality. The IQ complement measurements (i.e., the data that pass *i*Quam QC but fail external QC) continue to consistently display good quality. In many cases, they are comparable to those that pass both QC schemes. Conversely, data that pass external QC but fail *i*Quam QC often show less stable behavior, with worse performance statistics and large discontinuities on daily and monthly time scales. These measurements may have questionable quality and more outliers. The need for improved ICOADS QC was recognized as a major priority in previous ICOADS versions (Wolter 1997; Woodruff et al. 2011). Although some updates were added in R3.0, ICOADS QC is deemed to still have room for improvement. The relatively robust performance of *i*Quam QC, in the context of Argo and IMOS QCs, is largely due to a combination of the well-behaved CMC L4 reference field, employed in conjunction with the Bayesian reference and buddy checks. More outliers are therefore discarded, yet many good measurements retained. The period from 1981 to 1991 not covered by CMC data is more problematic, as the Reynolds L4 product is less accurate relative to the CMC L4 product, and likely even more degraded in the earlier satellite years. The PH check in *i*Quam turns out to be as effective as the CMS BL and Argo GL, with the advantage of being applied to all platforms, unlike the BL and GL, which are only applied to drifters and Argo floats, respectively. In the future, another blacklist produced by the Met Office, may be explored.

Several issues with *i*Quam QC have been identified. The most prominent one is the overscreening of large DV signals. The major reason is that both the CMC and Reynolds reference fields are designed to be foundation temperatures (typically corresponding to ~5–10-m depth, with, supposedly, no DV signals). However, the *i*Quam QC applies the reference checks not only during the nighttime, but during the daytime as well. The effect is most evident in coastal regions (e.g., when *i*Quam QC is compared with ICOADS QC for C-moorings), where DV can have large amplitudes, yet CMC and Reynolds SSTs may be of degraded quality. Another issue noticed in this study is that spatially, *i*Quam QC may also degrade in certain regions with strong currents and sharp vertical and horizontal temperature gradients, such as the Gulf Stream and the Eastern Australian Current, where the CMC and Reynolds fields miss some detailed spatial features. The globally uniform spatial spike check thresholds may not be applicable in certain areas, for some platforms.

Future work toward improved *i*Quam QC may include a possible replacement of the CMC field with a more accurate dataset, with more realistic and detailed spatial–temporal resolution, resolving the DV signals and finer spatial features. If such a reference is not available in the short term, another option is to superimpose modeled DV signals onto the CMC data (e.g., While et al. 2017). The globally uniform spike check threshold may be tweaked regionally, to meet the more stringent requirements in dynamic areas, such as the Gulf Stream, Kuroshio, Agulhas or Brazil currents. New approaches to QC that are based on reduced weight on the reference fields, and more analyses of continuity of the time series, will be also explored.

Acknowledgments. The *i*Quam development is supported by the NOAA JPSS, GOES-R and Ocean Remote Sensing Programs. We thank Eric Freeman and Scott Woodruff (NOAA) and Steve Worley (NCAR) for help with the ICOADS data; Pierre Le Borgne, Anne Marsouin, and Sonia Péré (CMS/OSI-SAF in France) for help with the CMS blacklist data; and Helen Beggs (Bureau of Meteorology in Australia) for help with the IMOS ship data. Argo data used in *i*Quam were collected and made freely available by the International Argo Program and the national programs that contribute to it (https://www.argo.ucsd.edu and https://argo.jcommops.org). The Argo Program is a part of the Global Ocean Observing System. We also thank the two anonymous reviewers, whose comments have helped to make this paper clearer. The views, opinions, and findings contained in this paper are those of the authors and should not be construed as an official NOAA or U.S. government position, policy, or decision.

REFERENCES

Akella, S., R. Todling, and M. Suarez, 2017: Assimilation for skin SST in the NASA GEOS Atmospheric Data Assimilation System. *Quart. J. Roy. Meteor. Soc.*, **143**, 1032–1046, https://doi.org/10.1002/qj.2988.

Anderson, J. E., and S. C. Riser, 2014: Near-surface variability of temperature and salinity in the near-tropical ocean: Observations from profiling floats. *J. Geophys. Res. Oceans*, **119**, 7433–7448, https://doi.org/10.1002/2014JC010112.

Argo Data Management Team, 2019: Argo user's manual V3.3. IFREMER Rep., 113 pp., https://doi.org/10.13155/29825.

Atkinson, C. P., N. A. Rayner, J. Roberts-Jones, and R. O. Smith, 2013: Assessing the quality of sea surface temperature observations from drifting buoys and ships on a platform-by-platform basis. *J. Geophys. Res. Oceans*, **118**, 3507–3529, https://doi.org/10.1002/jgrc.20257.

Beggs, H. M., R. Verein, G. Paltoglou, H. Kippo, and M. Underwood, 2012: Enhancing ship of opportunity sea surface temperature observations in the Australian region. *J. Oper. Oceanogr.*, **5**, 59–73, https://doi.org/10.1080/1755876X.2012.11020132.

Brasnett, B., and D. Surcel Colan, 2016: Assimilating retrievals of sea surface temperature from VIIRS and AMSR2. *J. Atmos. Oceanic Technol.*, **33**, 361–375, https://doi.org/10.1175/JTECH-D-15-0093.1.

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, https://doi.org/10.1175/1520-0426(2002)019<1638:ROOYOP>2.0.CO;2.

Castro, S., G. Wick, D. Jackson, and W. Emery, 2008: Error characterization of infrared and microwave satellite sea

surface temperature products for merging and analysis. *J. Geophys. Res.*, **113**, C03010, https://doi.org/10.1029/2006JC003829.

Chin, M., J. Vazquez-Cuervo, and E. Armstrong, 2017: A multi-scale high-resolution analysis of global sea surface temperature. *Remote Sens. Environ.*, **200**, 154–169, https://doi.org/10.1016/j.rse.2017.07.029.

Dash, P., and Coauthors, 2012: Group for High Resolution Sea Surface Temperature (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, https://doi.org/10.1016/j.dsr2.2012.04.002.

Donlon, C. J., P. J. Minnett, C. Gentemann, T. J. Nightingale, I. J. Barton, B. Ward, and M. J. Murray, 2002: Toward improved validation of satellite sea surface skin temperature measurements for climate research. *J. Climate*, **15**, 353–369, https://doi.org/10.1175/1520-0442(2002)015<0353:TIVOSS>2.0.CO;2.

Fairall, C. W., E. F. Bradley, J. S. Godfrey, G. A. Wick, J. B. Edson, and G. S. Young, 1996: Cool-skin and warm-layer effects on sea surface temperature. *J. Geophys. Res.*, **101**, 1295–1308, https://doi.org/10.1029/95JC03190.

Freeman, E., and Coauthors, 2017: ICOADS release 3.0: A major update to the historical marine climate record. *Int. J. Climatol.*, **37**, 2211–2232, https://doi.org/10.1002/joc.4775.

——, and Coauthors, 2019: The International Comprehensive Ocean-Atmosphere Data Set—Meeting users needs and future priorities. *Front. Mar. Sci.*, **6**, 435, https://doi.org/10.3389/fmars.2019.00435.

Ingleby, B., 2010: Factors affecting ship and buoy data quality: A data assimilation perspective. *J. Atmos. Oceanic Technol.*, **27**, 1476–1489, https://doi.org/10.1175/2010JTECHA1421.1.

——, and A. C. Lorenc, 1993: Bayesian quality control using multivariate normal distributions. *Quart. J. Roy. Meteor. Soc.*, **119**, 1195–1225, https://doi.org/10.1002/qj.49711951316.

——, and M. Huddleston, 2007: Quality control of ocean temperature and salinity profiles—Historical and real-time data. *J. Mar. Syst.*, **65**, 158–175, https://doi.org/10.1016/j.jmarsys.2005.11.019.

Kawai, Y., and A. Wada, 2007: Diurnal sea surface temperature variation and its impact on the atmosphere and ocean: A review. *J. Oceanogr.*, **63**, 721–744, https://doi.org/10.1007/s10872-007-0063-0.

Kent, E. C., and Coauthors, 2019: Observing requirements for long-term climate records at the ocean surface. *Front. Mar. Sci.*, **6**, 441, https://doi.org/10.3389/fmars.2019.00441.

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**, 9179–9197, https://doi.org/10.1029/1999JC000065.

Le Menn, M., P. Poli, A. David, J. Sagot, M. Lucas, A. O'Carroll, M. Belbeoch, and K. Herklotz, 2019: Development of surface drifting buoys for fiducial reference measurements of sea-surface temperature. *Front. Mar. Sci.*, **6**, 578, https://doi.org/10.3389/fmars.2019.00578.

Llewellyn-Jones, D. T., P. J. Minnett, R. W. Saunders, and A. M. Zavody, 1984: Satellite multichannel infrared measurements of sea surface temperature of the N.E. Atlantic Ocean using AVHRR/2. *Quart. J. Roy. Meteor. Soc.*, **110**, 613–631, https://doi.org/10.1002/qj.49711046504.

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, https://doi.org/10.1002/qj.49711448012.

Marsouin, A., P. Le Borgne, G. Legendre, S. Péré, and H. Roquet, 2015: Six years of OSI-SAF MetOp-A AVHRR sea surface temperature. *Remote Sens. Environ.*, **159**, 288–306, https://doi.org/10.1016/j.rse.2014.12.018.

Martin, M., and Coauthors, 2012: Group for High Resolution Sea Surface Temperature (GHRSST) analysis fields inter-comparisons. Part 1: A GHRSST multi-product ensemble (GMPE). *Deep-Sea Res. II*, **77–80**, 21–30, https://doi.org/10.1016/j.dsr2.2012.04.013.

Merchant, C. J., L. Horrocks, J. Eyre, and A. O'Carroll, 2006: Retrievals of sea surface temperature from infrared imagery: Origin and form of systematic errors. *Quart. J. Roy. Meteor. Soc.*, **132**, 1205–1223, https://doi.org/10.1256/qj.05.143.

O'Carroll, A. G., 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, https://doi.org/10.1175/2007JTECHO542.1.

Petrenko, B., A. Ignatov, Y. Kihai, J. Stroup, and P. Dash, 2014: Evaluation and selection of SST regression algorithms for JPSS VIIRS. *J. Geophys. Res. Atmos.*, **119**, 4580–4599, https://doi.org/10.1002/2013JD020637.

Poli, P., and Coauthors, 2019: The Copernicus Surface Velocity Platform drifter with Barometer and Reference Sensor for Temperature (SVP-BRST): Genesis, design, and initial results. *Ocean Sci.*, **15**, 199–214, https://doi.org/10.5194/os-15-199-2019.

Rayner, N. A., P. Brohan, D. E. Parker, C. K. Folland, J. J. Kennedy, M. Vanicek, T. J. Ansell, and S. F. B. 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, https://doi.org/10.1175/JCLI3637.1.

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 temperature. *J. Climate*, **20**, 5473–5496, https://doi.org/10.1175/2007JCLI1824.1.

Roemmich, D., and Coauthors, 2009: Argo: The challenge of continuing 10 years of progress. *Oceanography*, **22** (3), 46–55, https://doi.org/10.5670/oceanog.2009.65.

Saha, K., A. Ignatov, X. M. 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, https://doi.org/10.1029/2012JC008384.

Saunders, P., 1967: Aerial measurements of sea surface temperature in the infrared. *J. Geophys. Res.*, **72**, 4109–4117, https://doi.org/10.1029/JZ072i016p04109.

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 Climate Research Program Rep., 268 pp.

Storto, A., and P. Oddo, 2019: Optimal assimilation of daytime SST retrievals from SEVIRI in a regional ocean prediction system. *Remote Sens.*, **11**, 2776, https://doi.org/10.3390/rs11232776.

Strong, A., and P. McClain, 1984: Improved ocean surface temperatures from space—Comparison with drifting buoys. *Bull. Amer. Meteor. Soc.*, **65**, 138–142, https://doi.org/10.1175/1520-0477(1984)065<0138:IOSTFS>2.0.CO;2.

While, J., C. Mao, M. J. Martin, J. Roberts-Jones, P. A. Sykes, S. A. Good, and A. J. McLaren, 2017: An operational analysis system for the global diurnal cycle of sea surface temperature: Implementation and validation. *Quart. J. Roy. Meteor. Soc.*, **143**, 1787–1803, https://doi.org/10.1002/qj.3036.

Wolter, K., 1997: Trimming problems and remedies in COADS. *J. Climate*, **10**, 1980–1997, https://doi.org/10.1175/1520-0442(1997)010<1980:TPARIC>2.0.CO;2.

Wong, A., R. Keeley, and T. Carval, 2020: Argo quality control manual for CTD and trajectory data. IFREMER Rep., 63 pp., https://doi.org/10.13155/33951.

Woodruff, S. D., H. F. Diaz, E. C. Kent, R. W. Reynolds, and S. J. Worley, 2008: The evolving SST record from ICOADS. *Advances in Global Change Research*, S. Brönnimann et al., Eds., Springer International Publishing, 65–83, https://doi.org/10.1007/978-1-4020-6766-2_4.

——, and Coauthors, 2011: ICOADS release 2.5: Extensions and enhancements to the surface marine meteorological archive. *Int. J. Climatol.*, **31**, 951–967, https://doi.org/10.1002/joc.2103.

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, https://doi.org/10.1029/2010JC006129.

——, and ——, 2014: In situ SST quality monitor (*i*Quam). *J. Atmos. Oceanic Technol.*, **31**, 164–180, https://doi.org/10.1175/JTECH-D-13-00121.1.

——, and ——, 2016: Error characterization in *i*Quam SSTs using triple collocations with satellite measurements. *Geophys. Res. Lett.*, **43**, 10 826–10 834, https://doi.org/10.1002/2016GL070287.

Zhang, H., H. Beggs, L. Majewski, X. H. Wang, and A. Kiss, 2016: Investigating sea surface temperature diurnal variation over the tropical warm pool using MTSAT-1R data. *Remote Sens. Environ.*, **183**, 1–12, https://doi.org/10.1016/j.rse.2016.05.002.

——, A. V. Babanin, Q. Liu, and A. Ignatov, 2019: Cool skin signals observed from Advanced Along-Track Scanning Radiometer (AATSR) and in situ SST measurements. *Remote Sens. Environ.*, **226**, 38–50, https://doi.org/10.1016/j.rse.2019.03.035.