Evaluation of the *In Situ* Sea Surface Temperature Data Quality Control in *i*Quam

Haifeng Zhang[1, 2], Alexander Ignatov[1], Dean Hinshaw[1,3]

1. Introduction

- *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. Nonetheless, their quality can sometimes be suboptimal and proper quality control (QC) is required before *in situ* SST data can be used with confidence.
- To support the accurate and consistent Cal/Val of satellite and blended SST products, the *in situ* SST Quality Monitor system was established at NOAA in 2009 (*i*Quam; Xu and Ignatov, 2014; *www.star.nesdis.noaa.gov/socd/sst/iquam*). Objectives of *i*Quam are: (1) pull together a comprehensive set of *in situ* data, from various sources, covering full satellite era 1981-on; (2) perform the advanced, flexible, and unified community consensus QC; (3) monitor QC'ed SST online; and (4) distribute to users, in near-real time.
- This study aims to evaluate the performance of *i*Quam QC by comparing it with QCs from external data sources, such as ICOADS (the International Climate Ocean-Atmosphere Data Set), Argo, and IMOS (Integrated Marine Observing System) ship data.

2. *In situ* QC Schemes & Methods

- ***i*Quam** QC: includes five binary checks (duplicate removal, plausibility check, platform track check, SST spike check, and ID check) and two Bayesian checks (reference check, RC, and buddy check, BC).
- **ICOADS** QC: release 3.0 (R3.0) delayed mode (DM) QC employs duplicate elimination, land lock check, track check, source ID and source exclusion check, trimming flag, external flags inherited from the original sources (NCDC quality flag and World Ocean Database Ocean Station Data flag).
- **Argo** QC: includes up to 22 checks, such as the plausible value check, platform speed check (< 3 m s^{-1}), SST spike check, SST physical range check, and regional temperature range test (such as the Mediterranean Sea and the Red Sea), etc.
- **IMOS** QC: Unlike traditional ship data, IMOS ship SST measurements obtained from hull-contact sensors exhibited overall similar quality with drifting and moored buoys (Beggs et al., 2012). IMOS QC consists of several binary QC checks, including the duplicate removal, plausible value check, and track check, and an RC which flags a value as 'bad' if it exceeds 3°C above or below an operational daily SST analysis.
- ***i*Quam PH**: platform-specific, the 'Performance History' (PH) check in *i*Quam QC is evaluated against the Centre de Meteorologie Spatiale (CMS) drifter blacklist (BL) and Argo Gray List (GL).

QC Comparison Method ("Confusion Matrix") and Data

When comparing any two QC schemes (*i*Quam and external QC, X), we define as 100% the sample where SST has passed at least one of them, and break it into 3 sub-samples as follows:

- IQ*X: SST pass both *i*Quam and external QC ("*i*Quam/X intersection")
- IQ–X: pass *i*Quam QC but fail external QC ("*i*Quam complement")
- X–IQ: fail *i*Quam QC but pass external QC ("X complement")

All analyses are based on these three "Confusion Matrix" categories. If *i*Quam and X QC are consistent, then all 100% fall in the intersection category, with 0% in the two complements. If a complement is non-zero, but its SST "errors" are indistinguishable from the "intersection" – then it's "false alarm". If SST "errors" in a complement are significantly larger, then it's a "leakage". Error is defined as a SD of *in situ* minus reference SST, the reference being a global L4 analysis.

Since ICOADS R3.0 DM data only covers until the end of 2014, three-year (2012-2014) data are adopted in the *i*Quam vs. ICOADS QC comparison. For Argo and IMOS, three different years (2017-2019) are used, to take advantage of the increased data volumes.

3. Results

3.1. Statistics

- The performance of *i*Quam QC is first evaluated by analyzing the distributions of temperature differences ($\Delta T = SST_{insitu} - SST_{ref}$) for the three categories. Two daily L4 analysis fields are used in *i*Quam as a reference: (1) 0.25° Reynolds optimal interpolation (OI); and (2) the 0.10° Canadian Met Centre (CMC).
- The four different platforms from ICOADS, namely ships, drifters, T- (tropical) and C- (coastal) moorings, are investigated separately.

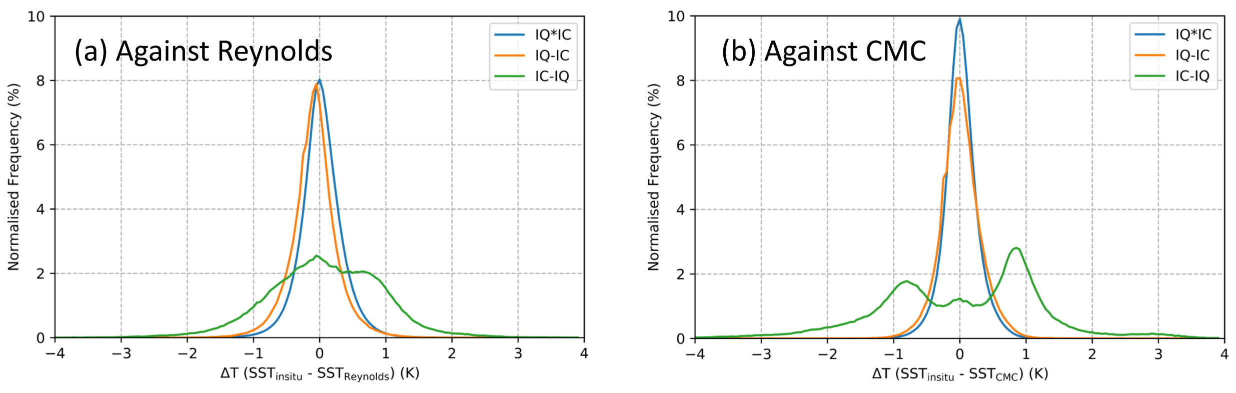

Fig. 1. Normalized frequency of $\Delta T = SST_{insitu} - SST_{ref}$ for ICOADS drifters stratified by iQuam and ICOADS QC flags. Left: Ref=Reynolds; Right: Ref=CMC. Only drifters are shown as an example. Other ICOADS platforms and Argo and IMOS ship data display similar patterns.

Table 1. Statistics, including percentage of number of observations (Nob.), mean (μ) ± SD (σ) / robust SD (RSD) of ΔT against Reynolds and CMC reference fields, are shown for the three categories between iQuam and ICOADS, Argo and IMOS ship.

Platform /Category	Percentage of Nob.	Against Reynolds μ ± σ (RSD) (K)	Against CMC μ ± σ (RSD) (K)
IC. 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)
IC. 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)
IC. 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)
IC. 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	0.26 ± 2.40 (2.35)	0.07 ± 2.75 (2.83)
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)

- For nearly all platforms, the IQ*X category accounts for the majority of all the observations, and has the best statistics, indicating that **a combination of two QC schemes effectively guarantees highest quality data**. (Note that *i*Quam mainly ingests real-time Argo data, which are not defined as 'pass' Argo QC in this study.)
- For nearly all cases, the (IQ-X) complement often has similar shapes and statistics to those of the interception (IQ*X), meaning that the ***i*Quam QC generally picks up reasonably fine quality data that fail the external QC.**
- However, **the X complement (X-IQ) usually shows degraded quality** with a very broad, non-Gaussian distributions with much larger SDs and RSDs.
- **Statistics w.r.t. CMC L4 SST are always superior to those w.r.t. Reynolds.** Only CMC statistics will be used for the rest of the analyses.

3.2. Dependence on Local Solar Time (LST)

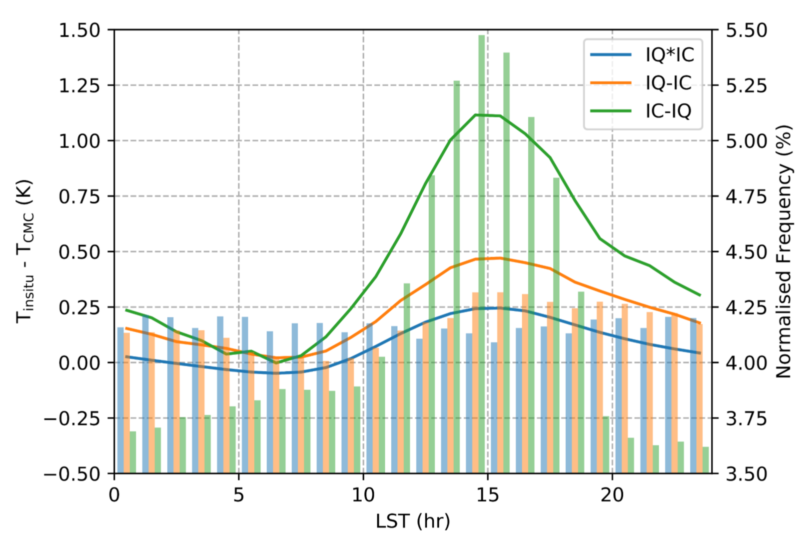

Fig. 2. Dependence of $\Delta T = SST_{insitu} - SST_{CMC}$ (lines; left Y-axis) and normalized frequencies of three categories (bars; right Y-axis) on local solar time (LST) for ICOADS C- moorings, in three "confusion matrix" categories. Colors in legend are for all lines and bars. Only C- mooring is shown here as an example.

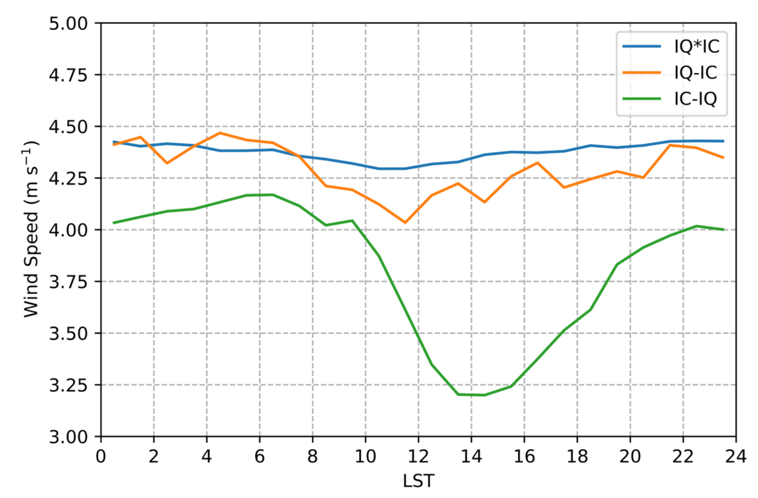

Fig. 3. Wind speed for ICOADS C- mooring platforms as a function of LST, stratified by iQuam and ICOADS QC flags.

- Fig. 2 shows that a large chunk of the C-mooring measurements that fail *i*Quam QC, but pass ICOADS QC **(IC-IQ) are likely to be fine quality data, but with relatively large diurnal variation (DV) signals**.
- ***i*Quam QC may discard good quality SSTs with large DV signals**, especially at calmer winds at 12-18 hr LST as shown in Fig. 3.

3.3. Spatial Distributions of ΔT (*in situ* minus CMC SST)

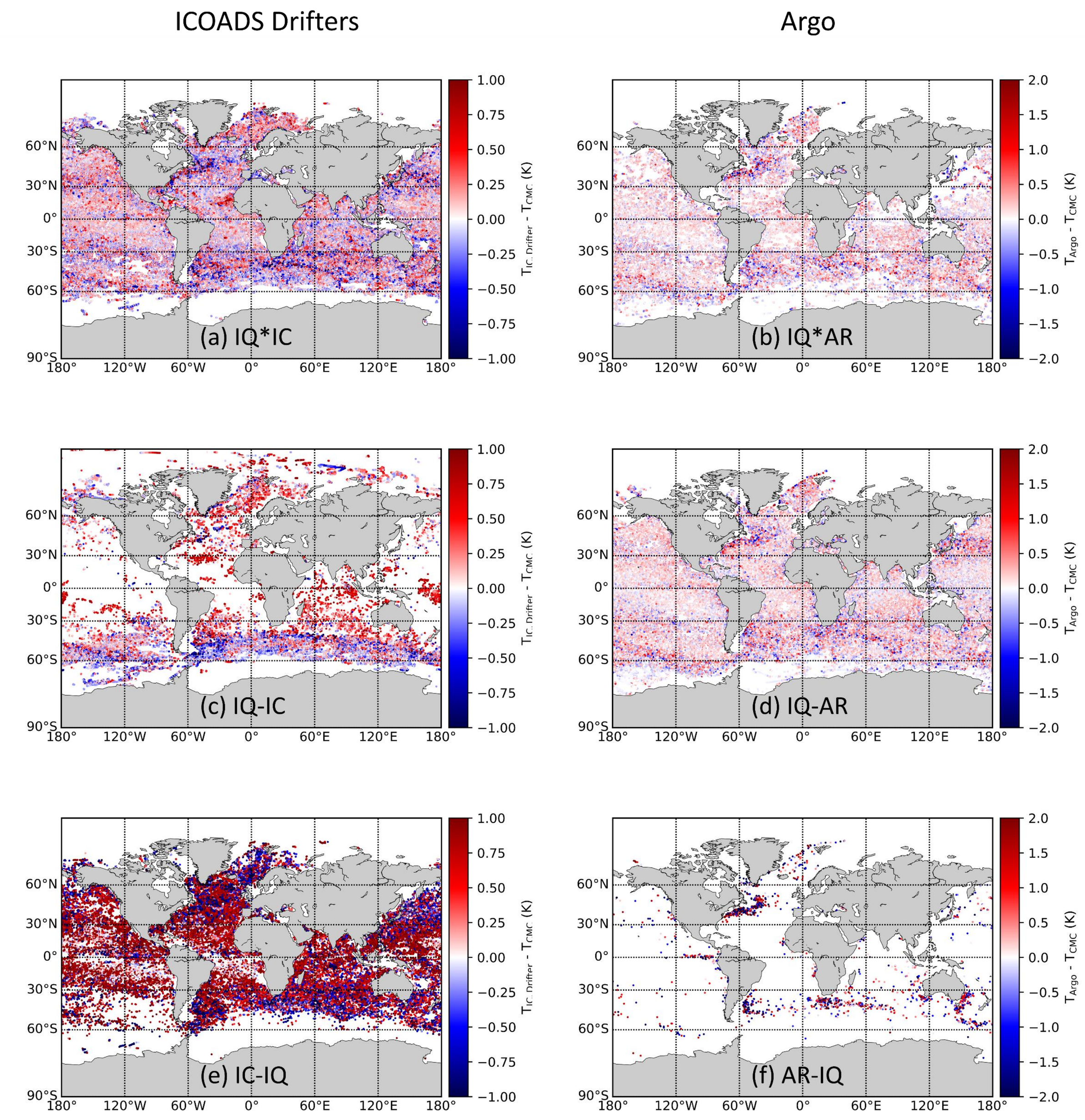

*Fig. 4. Spatial distributions of ΔTs from ICOADS drifters (left) and Argo (right) for each category: IQ*X (top), IQ-X (middle), and X-IQ (bottom).*

- Since only drifters and Argo floats have global coverages, their ΔT distributions in each "confusion matrix" category are plotted in Fig. 4.
- The drifter measurements that pass ICOADS QC but not *i*Quam QC **(IC-IQ)** mostly have **large amplitudes, being either positive or negative**, with an almost global coverage (Fig. 4e).
- The much fewer **AR-IQ** measurements are observed frequently in the **strong currents** (e.g., Golf Stream, Eastern Australian Current, the Brazil Current) regions that are **characterized by sharp horizontal and vertical temperature gradients.**

Table 2. Same as Table 1 but for the comparison between PH&BL (using drifters) and PH&GL (using Argo data).

Platform /Category	Percentage of NOB.	Against CMC μ ± σ (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)

- Table 2 show that *i*Quam PH check compares well with CMS BL and Argo Grey List (GL). **In *i*Quam, PH is applied to all platforms**.

4. Conclusions

- **Overall, the *i*Quam QC shows robust performance for all platforms under various situations.** The IQ-X complements (i.e., where data pass *i*Quam QC but fail external QC) **appear of consistently good quality**.
- Those data that pass external QC but fail *i*Quam QC, frequently show unstable behavior with degraded statistics and appear as large outliers in the corresponding time series. By all characteristics, these measurements usually have questionable quality.
- Several issues in *i*Quam QC are identified as well. **The most prominent one is frequent screening of large diurnal signals.** This is due to the important role the reference check plays in *i*Quam QC, while all reference fields currently available (e.g., CMC, Reynolds) are foundation products, reported only once a day, and do not resolve the diurnal cycle.
- **The *i*Quam QC may also be degraded in certain dynamic regions (strong currents, sharp temperature gradients).** In those areas, the current L4 analyses may underrepresent detailed spatial features. Also, the spatial spike check uses one global threshold for different platforms.
- Future work to improve *i*Quam QC may focus on the **update of the reference field**, **incorporation of a DV model**, and exploring QC enhancements from other advanced QC systems (e.g., Met Office, etc).

Selected References

1) Beggs et al. (2012), J. Oper. Oceanogr. 5, 59–73.
2) Freeman et al. (2017), *Int. J. Climatol. 37, 2211–2232.*
3) Roemmich et al. (2009), Oceanography 22, 46–55.
4) Wong et al. (2020), https://doi.org/10.13155/33951
5) Xu and Ignatov (2014), J. Atmos. Ocean. Technol. 31, 164–180.

Acknowledgement: The *i*Quam is supported by NOAA JPSS, GOES-R, PSDI Programs. The views, opinions, and findings in this report are those of the authors and should not be construed as an official NOAA or U.S. government position or policy.

[1]NOAA/STAR, College Park, MD 20740, USA; [2]CIRA, Colorado State University, Fort Collins, CO, USA; [3]GST, Inc., MD, USA

Emails: haifeng.zhang@noaa.gov; alex.ignatov@noaa.gov; dean.hinshaw@noaa.gov