AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER

10.1002/2016GL070287

Key Points:

- Error characterization in satellite and in situ SSTs using triple-collocation method
- Analysis of temporal trends and spatial patterns in satellite and in situ SST errors
- Lognormal distribution used to describe random errors in individual in situ SST sensors

Correspondence to:
F. Xu,
fengxu@fudan.edu.cn

Citation:
Xu, F., and A. Ignatov (2016), Error characterization in *i*Quam SSTs using triple collocations with satellite measurements, *Geophys. Res. Lett.*, *43*, 10,826–10,834, doi:10.1002/2016GL070287.

Received 9 JUL 2016
Accepted 13 SEP 2016
Accepted article online 15 SEP 2016
Published online 18 OCT 2016

Error characterization in *i*Quam SSTs using triple collocations with satellite measurements

Feng Xu[1,2,3] and Alexander Ignatov[1]

[1]NOAA STAR, College Park, Maryland, USA, [2]Key Lab for Information Science of Electromagnetic Waves (MoE), Fudan University, Shanghai, China, [3]GST Inc, Greenbelt, Maryland, USA

Abstract Various types of in situ sea surface temperature (SST) measurements have dominated during different periods of the satellite era. Their corresponding errors should be characterized to curtail the nonuniformities in calibration and validation of reprocessed historical satellite SST data. SSTs from several major in situ platform types reported in the NOAA in situ Quality Monitor (*i*Quam) system have been collocated with NOAA-17 Advanced Very High Resolution Radiometer (AVHRR) and Envisat Advanced Along Track Scanning Radiometer (AATSR) satellite SSTs from 2003 to 2009, produced by the European Space Agency (ESA) Climate Change Initiative (CCI) program. The standard deviations of errors in *i*Quam in situ and nighttime satellite CCI SSTs estimated using triple-collocation analyses are 0.75 K for ships, 0.21–0.22 K for drifters and Argo floats, 0.17 K and 0.40 K for tropical and coastal moorings, 0.35–0.38 K for AVHRR, and 0.15–0.30 K for AATSR. The distribution of in situ and satellite errors in space and time is also analyzed, along with their single-sensor error distributions.

1. Introduction

In 2009, NOAA established the in situ sea surface temperature (SST) Quality Monitor (*i*Quam) to support the calibration and validation (Cal/Val) of satellite and blended SST products [*Xu and Ignatov*, 2014]. Uniformly, quality controlled (QCed) *i*Quam SSTs have become a key element of the international Group for High Resolution SST (GHRSST) and are now widely used for a variety of SST analyses. In addition to drifting and moored buoys and ships available in *i*Quam v1, *i*Quam2 has added data from the global Argo array [*Roemmich et al.*, 2009] and several local or experimental programs and extended its temporal coverage to include the full satellite era [*iQuam v2*, 2016].

No single in situ data type covers the full satellite SST era. In particular, ships prevailed in the 1980s but then declined. The number of drifting and moored buoys quickly increased in the 1990s, and now they greatly outnumber ship reports. The Argo floats first deployed in the late 1990s today provide the most globally uniform coverage of the ocean, although the number of their measurements is much smaller compared to drifters due to their 10 day profiling cycle. Uncertainties in each data type should be characterized to facilitate the transition from one in situ standard to another for consistent Cal/Val of reprocessed historical satellite data [*Ignatov et al.*, 2016].

This study employs the triple-collocation method (TCM) to analyze uncertainties in five in situ data types during the 7 year period from January 2003 to December 2009, when they were present in sufficient quantities. The TCM was first tested to estimate the standard deviations (SDs) of errors in three measurements of ocean wind speed by calculating corresponding paired variances and assuming that errors in the three data sets were additive, Gaussian, and independent [*Stoffelen*, 1998]. Since then, the TCM has been widely used to analyze uncertainties in various satellite, model, and in situ geophysical data [e.g., *McColl et al.* 2014, and references therein]. In the SST community, the TCM was first tested using two satellite SSTs (from microwave, MW, Aqua/Advanced Microwave Scanning Radiometer–EOS (AMSR-E) and infrared, IR, Envisat/Advanced Along Track Scanning Radiometer (AATSR)) and drifters [*O'Carroll et al.*, 2008]. *Xu and Ignatov* [2010] employed the TCM to evaluate the spatial and statistical structure of random errors in several in situ data types, including ships, drifters, and moorings, by pairing them with satellite (Pathfinder L3) and analysis (Reynolds L4) SSTs. *Gentemann* [2014] used the TCM to estimate SDs of random errors in the Moderate Resolution Imaging Spectroradiometer (MODIS) IR and AMSR-E MW (both onboard Aqua satellite) SSTs by triple-collocating them with drifters.

In this study, uncertainties in *i*Quam in situ SSTs are estimated by triple-collocating them with two satellite SST products, derived by the European Space Agency (ESA) Climate Change Initiative (CCI) program from NOAA17 Advanced Very High Resolution Radiometer (AVHRR) and Envisat AATSR [*Merchant et al.*, 2014]. The CCI products have been selected for the TCM because they are derived independently from in situ data. Note that both AVHRR and AATSR are IR sensors and their errors may thus be correlated. However, these correlations are expected to be minimized by different measurement principles (cross scan versus dual view) and the corresponding retrieval methodologies. Seven full years of CCI data from January 2003 to December 2009 have been triple collocated with five in situ data types from *i*Quam, with the corresponding platform-specific SDs calculated, first globally, and then stratified in space and time. The in situ single-sensor uncertainties associated with random errors for individual buoy/ship IDs were also analyzed and found to be well represented by lognormal distributions.

Section 2 describes the in situ and satellite data and their triple collocations. Section 3 presents and discusses the results, while section 4 concludes the paper.

2. Data and Triple-Collocation Matchups

2.1. In Situ Data

The in situ SSTs used in this study are from the NOAA *i*Quam2 system. The *i*Quam2 data are uniformly quality controlled and organized into monthly files freely available online [*iQuam v2*, 2016]. The in situ data used in this study are those with quality level = 5. Note that *i*Quam QC is designed to remove outliers (gross errors) while minimizing the disturbances to the intrinsic statistical characteristics of in situ measurement error [*Xu and Ignatov*, 2014].

Five major in situ data types reported in *i*Quam2 are analyzed in this study: ships, drifters, Argo floats, and tropical (T-) and coastal (C-) moorings. From Argo profiles, only one SST closest to the surface is extracted and saved, along with its measurement depth. The remaining three data types—GHRSST High-Resolution (HR) Drifters, Australian Integrated Marine Observing System ships, and NOAA Coral Reef Watch (CRW)—had limited time/space coverage during the 7 year period analyzed in this study and therefore are not considered here.

The time series of the number of individual reporting IDs and corresponding observations from 2003 to 2009 (available on the *i*Quam website) show that the number of ships declined from ~1950 to ~1530 and drifters increased from ~700 to ~1350, while the T- and C-moorings remained fairly steady at ~75–85 and 210–250, respectively. The Argo fleet has grown at the fastest pace, i.e., from ~650 to ~3300. The reporting rate also increased for all platforms, except Argo floats, which provided 12 K observations per month (hereafter, "K" stands for "thousand"), due to their much lower reporting frequency of one profile per 10 day cycle. Despite a declining or steady number of IDs for some data types (e.g., ships and moorings), the number of corresponding reports has increased. The monthly number of observations has changed from 50 K to 90 K for ships, from 25 K to 50 K for T-moorings, and from 100 K to 300 K for C-moorings. Drifters clearly provided the greatest number of observations during these years, with the monthly number of observations having increased from 250 K to 1000 K.

Argo floats and drifters provide the most complete and uniform global coverage. Ships are grouped along the shipping lines. T- and C-moorings are found in the tropics and around the coastal lines of North America, Europe, Hawaii, and India.

2.2. Satellite Data

Satellite SST products have near-global daily coverage and thus are well suited for triple collocations with in situ data. SST retrievals from the NOAA-17 AVHRR and Envisat AATSR, derived by the ESA SST CCI program [*Merchant et al.*, 2014], are selected in this study for the following reasons:

1. CCI employs physical (i.e., radiative transfer model-based) retrieval algorithms, which are largely independent from in situ data.
2. AVHRR and AATSR sensors employ different measurement (cross-track versus along-track) and retrieval (spectral versus dual-view) principles and therefore are expected to be those least correlated in the family of IR sensors and algorithms.
3. The NOAA-17 and Envisat satellites have similar Sun-synchronous orbits, both observing at approximately 10 A.M./P.M. local time and having long overlaps in their periods of operation, which ensures that when an in situ observation is matched with data from one satellite sensor (e.g., AVHRR), it is more likely to also be matched with data from the other satellite sensor (e.g., AATSR).

Seven full years (from January 2003 to December 2009) of AVHRR L2P (i.e., pixel-level data in original swath projection, with resolution ~4 km at nadir and degradation by a factor of 3–4 toward swath edges) v1.0 and

Table 1a. Paired Mean Biases in the Triple Collocated Matchup Data Set

In Situ Anchor	AVHRR-*i*Quam Mean Bias (K)	AATSR- *i*Quam Mean Bias (K)	AVHRR-AATSR Mean Bias (K)	Number of Triplets (% to Total)
Ship	−0.36	−0.30	−0.06	87,807 (9.4%)
Drifter	−0.30	−0.20	−0.10	624,429 (66.9%)
Argo	−0.31	−0.21	−0.10	3,931 (0.4%)
C-Moor	−0.22	−0.21	−0.01	188,368 (20.2%)
T-Moor	−0.34	−0.22	−0.12	28,399 (3.0%)

AATSR L3U (equal-gridded 0.05°, approximately corresponding to ~5 km resolution at the equator and $<$ 5 km toward the high latitudes; "U" stands for "uncollated," meaning that multiple passes from different orbits over a grid are not averaged together and are all preserved in the data set) v1.1 data obtained from the Centre for Environmental Data Analysis (CEDA) website (www.ceda.ac.uk) are used in this study. Only nighttime data are used to minimize the additional noise arising from the uncertain diurnal warming effect. Note that the spatial resolution of the AATSR gridded L3U data at the equator roughly matches the original AVHRR L2P resolution at nadir. Gridding (averaging) AATSR pixel SSTs may smooth out some spatial variability present in the AATSR L2P data and thus reduce the corresponding SDs.

2.3. Triple Collocations and Matchup Window

The triplet matchups are generated by pairing one AVHRR pixel and one AATSR pixel with each in situ data point. In situ data are the main focus of this study and therefore were used as an anchor, with one in situ measurement forming at most one triplet, while the same satellite pixel may contribute to several different triplets. Note that the same satellite pixel being included in multiple triplets may lead to errors being correlated. However, our additional analyses have shown that this effect is negligible. For each in situ data point, the first step of the matchup process is to exclude all satellite pixels falling outside the matchup window and then select the AVHRR and AATSR clear-sky SST pixels closest in space. Based on our sensitivity analyses, the matchup window was selected to be (15 km, 3 h), which minimizes matchup noise while maintaining a sufficient number of triplets.

3. Results

3.1. Global Error Statistics

The number of triplets in the global 7 year triple collocated matchup data set is shown in Table 1a, along with the corresponding paired mean biases.

The AVHRR and AATSR SSTs should be consistent, whereas a cold bias is expected in both CCI SSTs, which are skin SST products, against in situ SSTs, which are measured at depth (e.g., [*Donlon et al.*, 2002]). At night, the average bias is −0.17 K at high wind speeds and is increased at low wind speeds. Furthermore, these average numbers may vary depending upon environmental conditions and may not be distributed normally [*Donlon et al.*, 2002]. However, this variability and its non-Gaussian nature are expected to provide only a small contribution to the uncertainties analyzed in this paper (with typical SDs ~ 0.15 K–0.40 K). Note that the CCI data used here also report a depth SST derived using a diurnal thermocline model [*Merchant et al.*, 2014]. Although such correction is expected to improve the consistency with in situ data, in this study, we choose to use the skin CCI SST, which is not subject to additional errors associated with this conversion, and leave the analysis of CCI depth SST to future work. The mean AATSR-*i*Quam $\Delta T \sim (-0.21 \pm 0.01)$ K is indeed close to the expected −0.17 K value and consistent for all in situ data types, except ships, for which $\Delta T \sim -0.30$ K, likely due to the known warm bias of 0.1–0.2 K in ship SSTs [e.g., *Kennedy*, 2014, and references therein]. The remaining small cold bias of −0.04 K may come from the residual diurnal thermocline in in situ data at low wind speeds, sensor calibration, or residual cloud/aerosol contamination in the AATSR product. AVHRR SSTs also agree well with in situ data (except for C-moorings and ships) but are biased another ~−0.10 K colder, probably signaling more residual cloud contamination in the cruder resolution AVHRR data.

Table 1b lists the individual SDs for each sensor type from the three-way error analysis. The recently proposed extended triple collocation (ETC) method [*McColl et al.*, 2014] was tested, which in addition to the additive errors also allows for estimation of multiplicative biases between the three data sets. As expected for well calibrated satellite and in situ data, all corresponding scaling factors between the three data sets were found to be

Table 1b. SDs of Individual Satellite and In Situ Data Derived Using Three-Way Error Analysis

In situ Anchor	*i*Quam SD (K)	AVHRR SD (K)	AATSR SD (K)
Ship	0.75	0.37	0.22
Drifter	0.21	0.35	0.15
Argo	0.22	0.38	0.15
C-Moor	0.40	0.37	0.30
T-Moor	0.17	0.41	0.04

close to ~1.0 and there was no statistically significant difference between the individual SDs, derived using the ETC and the standard TCM. Hence, only the mean biases and TCM-derived SDs are analyzed below.

If all other conditions are equal, the estimated SDs in satellite SSTs (the last two columns of Table 1b) should be close for all triplets formed against different in situ data. With the exception of "AATSR minus T-moorings," which shows SD ~ 0.04 K, all other triplets provide consistent estimates for the AVHRR (0.35–0.41 K) and AATSR (0.15–0.30 K) SDs. The AATSR SST appears to be more precise (has smaller SDs partially due to the additional spatial smoothing in the L3U data) and therefore is naturally more prone to any residual uncertainties in the TCM and to the (incomplete) representativeness of the domains covered by the different in situ data types (likely explaining why the AATSR SDs estimated from triplets with highly regional T- and C-moorings are out of the family).

The major objective of this study is to estimate random errors in different in situ data types (shown in the second column of Table 1b). Currently, drifters and T-moorings are routinely used in the NOAA Cal/Val, complementing each other with respect to geographical coverage. Their SDs (0.21 and 0.17 K, respectively) are indeed close and suggest that they can be combined together. Comparisons with other published data for the drifters and T-moorings [e.g., *Kennedy*, 2014] suggest that our estimates are comparable—e.g., SD ~ 0.23 K [*O'Carroll et al.*, 2008], 0.26 K [*Xu and Ignatov*, 2010], and 0.20 K [*Gentemann*, 2014].

Argo floats, which were first deployed in the late 1990s, provide measurements with comparable uncertainties to drifters, with SD of ~0.22 K. Recall that for Argo, the closest measurement to the surface is typically from a depth of 4–5 m (cf. 15–20 cm for drifters and ~1 m for the moorings). Nevertheless, the two SDs are close, at least at night, when the diurnal thermocline is minimal. We are not aware of other published estimates of the uncertainties in the Argo SSTs.

The SDs for the C-moorings are clearly elevated over the drifters, T-moorings, and Argo floats. This may be because they operate in a more challenging coastal environment, in conjunction with increased spatial and temporal SST variability, which affects the matchup statistics. This increased spatiotemporal variability, if not accounted for, may propagate from the paired validation SDs into the TCM derived in situ SDs. The SD ~ 0.40 K estimated in this study is in good agreement with other published numbers for the C-moorings—cf., e.g., 0.39 K [*Xu and Ignatov*, 2010] and 0.40 K [*Kennedy*, 2014].

Ships provide the least accurate validation standard of all in situ data, with SD ~ 0.75 K. Nevertheless, they may be used for validation of the satellite product if their noise is known and properly accounted for. In particular, Table 1b suggests that the AVHRR and AATSR SDs estimated from ships' triplets are fairly consistent with the corresponding satellite SDs, estimated using drifters' and Argo floats' triplets. Note that our estimates of ship SDs tend to be on the low side compared with all other published estimates [*Kennedy*, 2014], which all suggest SD > 1 K (including our own earlier estimate of 1.16 K [*Xu and Ignatov*, 2010]). This may be due to the improved QC in *i*Quam2. However, note that QC efficacy may be degraded as one goes back in time due to a greater paucity of reliable background references.

3.2. Temporal Trends

To assess the statistical stationarity of the global statistics shown in Tables 1a and 1b, Figure 1 additionally plots them as a function of time.

Seasonal oscillations are observed in "satellite minus in situ" biases (likely due to the residual diurnal warming at night, resulting from periodic changes in the solar insolation and wind speed) and, to a much smaller degree, in the derived individual SDs, shown in Figures 1d–1f. In terms of long-term trends, the ΔTs are relatively stable in time and consistent with the global statistics shown in Figure 1. One notable exception is the steady increase in the "satellite minus C-moorings" ΔTs, at a rate of ~0.03 K/yr, for both satellite sensors. Similar trends, but reduced by a factor of 3, are also observed against drifters and Argo floats. Within the uncertainties associated with the relatively short-term time series, one may conclude that the C-moorings

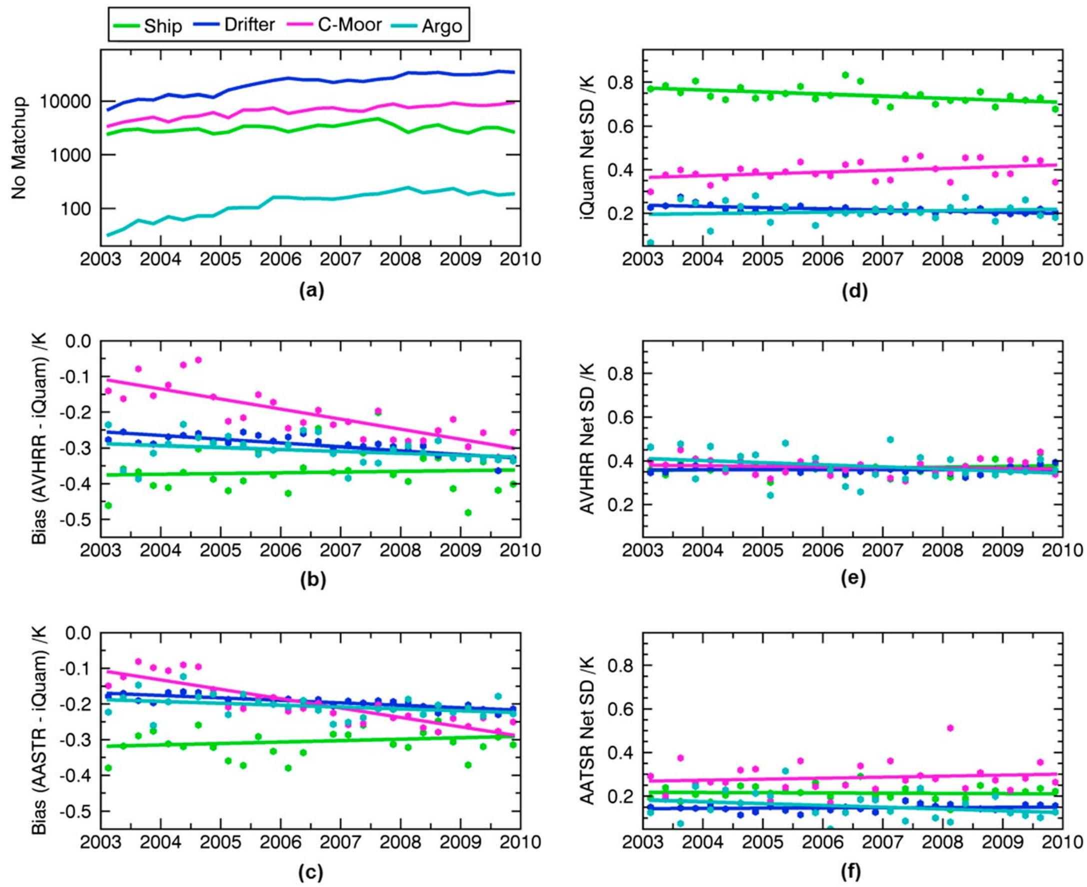

Figure 1. (a) Number of matchups; (b) (AVHRR-*i*Quam) paired biases; (c) (AATSR-*i*Quam) paired biases; (d) TCM-derived individual *i*Quam SDs; (e) AVHRR SDs; and (f) AATSR SDs. Linear fits emphasize prominent trends in the data. T-moorings have been excluded to simplify the plots.

have warmed up by ~0.2 K over the 7 year period analyzed here. The cause of this artifact is not immediately clear. Otherwise, the results in Figure 1 are in good qualitative and quantitative agreement with the global data in Tables 1a and 1b, including excellent consistency between the drifters' and Argo floats' ΔTs for both satellite sensors and the same warm bias in the ship SSTs (resulting in a cold bias in AVHRR and AATSR ΔTs).

The in situ SDs in Figure 1 are in agreement with the global data in Table 1b. The AVHRR SDs are consistent across various in situ SSTs and stable in time. The AATSR SDs are also stable but show more spread from 0.15 to 0.30 K (cf. Table 1b). The most outlying behavior is against the C-moorings, likely due to the challenging and highly regional coastal environment, whereas the AATSR SDs derived from the three near-global matchups (drifters, Argo floats, and ships) are more uniform from 0.15 to 0.22 K.

Turning to in situ data, the SDs of errors in drifters' and Argo floats' SSTs are consistent and stable in time at ~0.21–0.22 K. Over the course of the 7 year period, the ship SDs improved (from ~0.76 K to ~0.70 K) and the C-moorings' SDs slightly degraded (from ~0.36 K to ~0.43 K).

3.3. Spatial Patterns

To test if the satellite minus in situ ΔTs and TCM-derived SDs are spatially uniform, they were aggregated over all 7 years and stratified into 5° × 5° latitude/longitude windows. The TCM analyses were performed in moving 3 × 3 (15° × 15°) arrays (to increase the number of matchups) and assigned to the central 5° × 5° box.

Figures 2a–2c and 3a–3c show global distributions of numbers of matchups and mean satellite minus in situ ΔTs for the two most globally complete in situ data sets, i.e., drifters and Argo floats. Both cover the ocean near fully but nonuniformly. The Argo floats have significantly (2 orders of magnitude) reduced data density. The "AATSR minus in situ" ΔTs are more spatially uniform compared to the AVHRR counterparts, with both satellite SSTs being consistently suppressed in some areas, e.g., influenced by aerosols (off the West

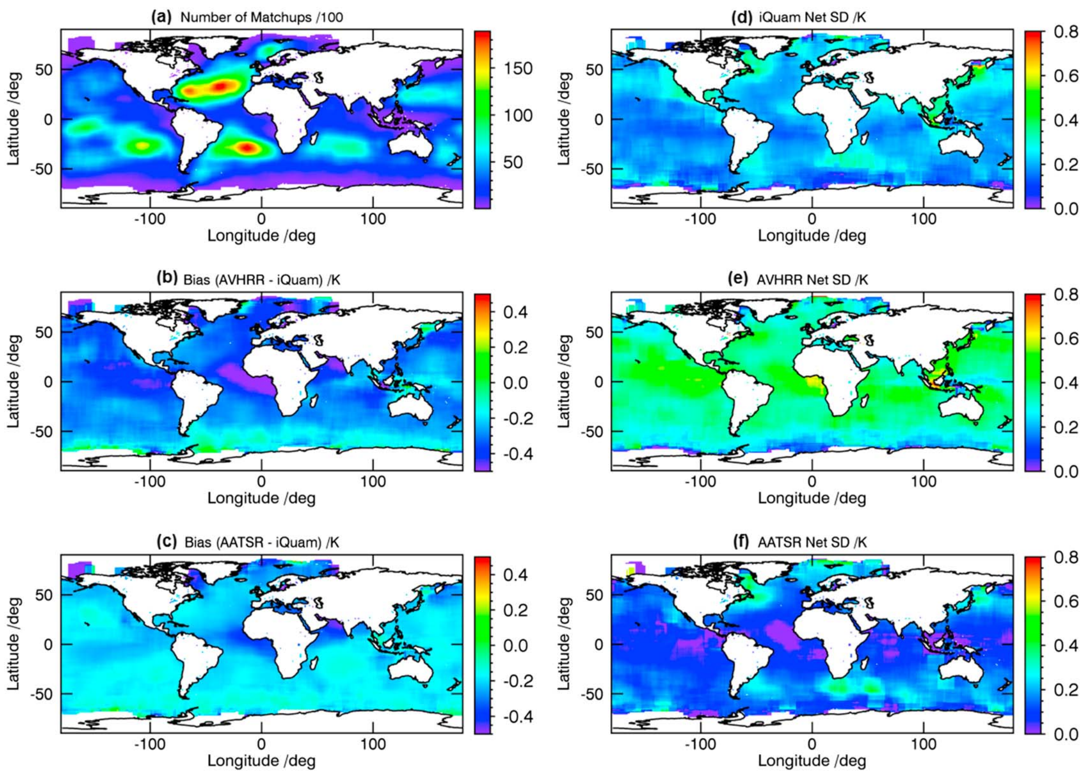

Figure 2. (a–c) Number of observations and satellite minus drifters biases; (d–f) TCM-derived SDs.

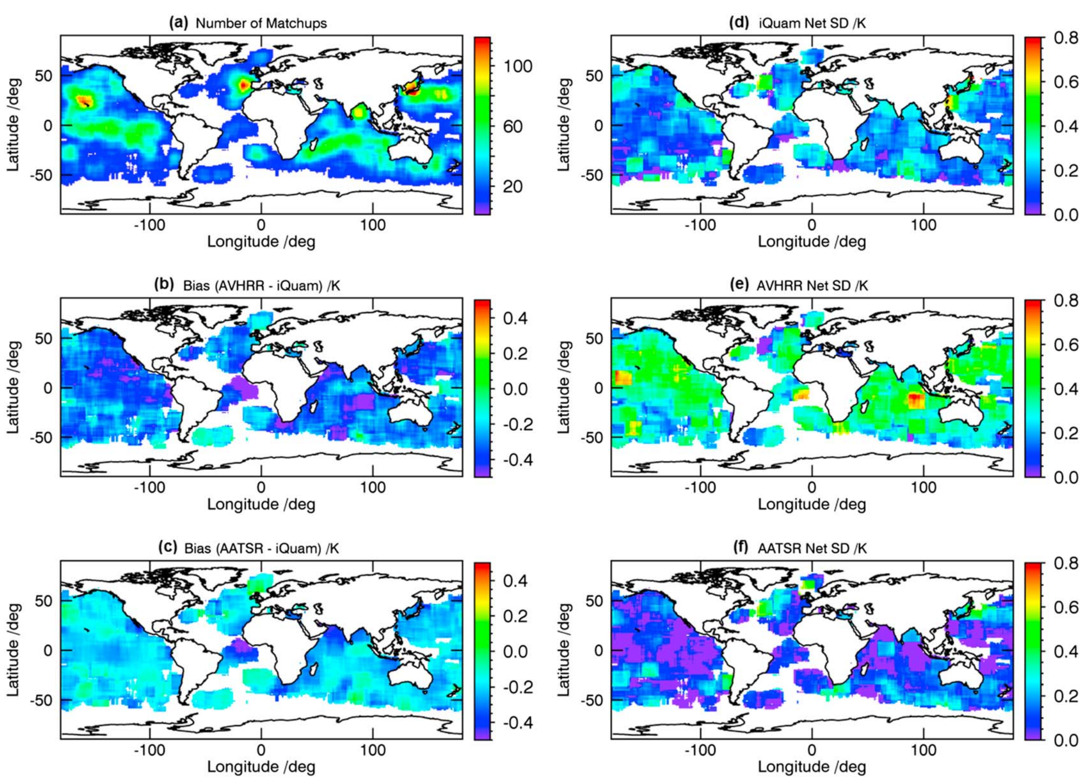

Figure 3. (a–c) Number of observations and satellite minus Argo floats biases; (d–f) TCM-derived SDs.

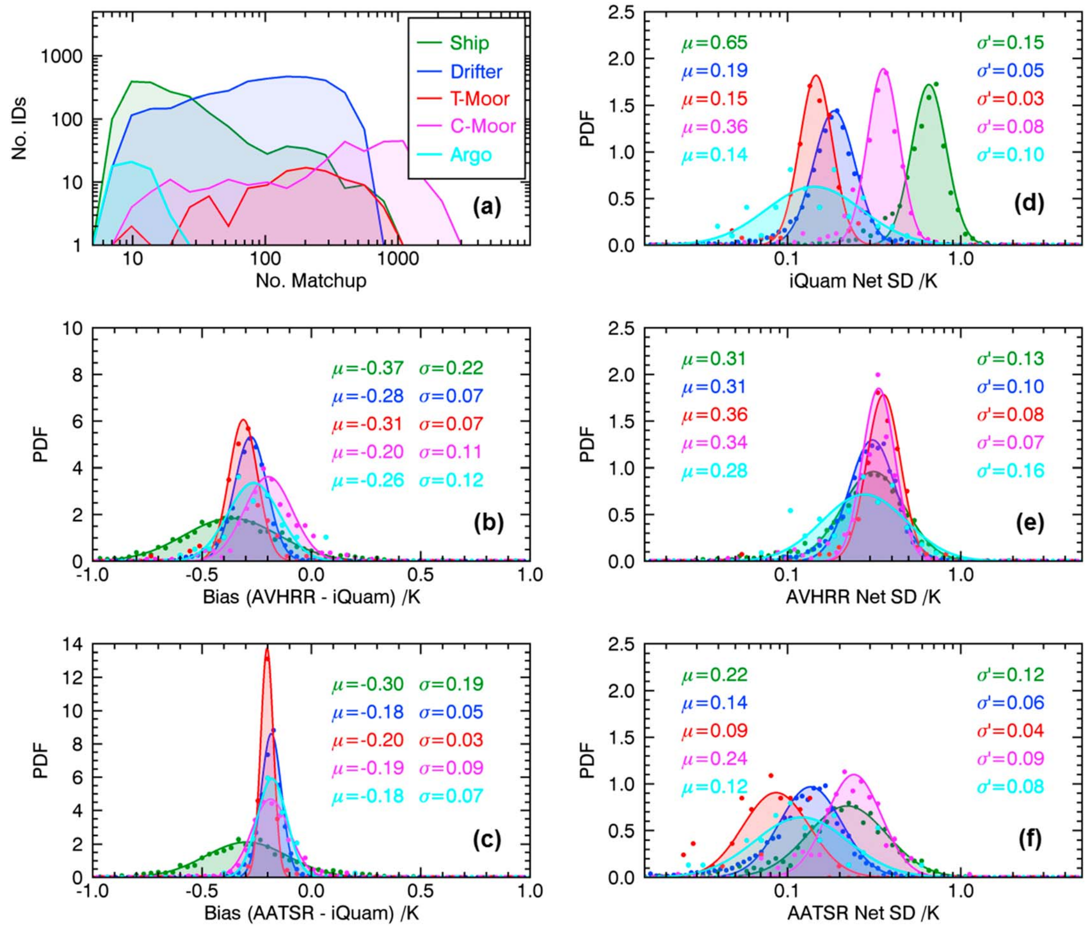

Figure 4. Histograms and PDF fitting of error statistics of individual platforms.

African coast, over the Arabian Sea). Suppressions of smaller magnitudes are also observed in the Pacific (off the Central and North America) and in the high latitudes of the Northern Hemisphere. These areas tend to be sparsely populated with in situ data, and the corresponding estimates may be uncertain. Correlations between the “satellite minus drifters” and “satellite minus Argo” ΔTs were calculated and found to be noisy but statistically significant, with correlation coefficients $R \sim 0.32–0.36$.

The TCM-derived SDs are shown in the right panels of Figures 2d–2f and 3d–3f . The satellite random errors are relatively uniform in space, with SDs being systematically larger for the AVHRR compared to the AATSR SSTs. The AATSR SDs show more structure, being smaller in the tropics and elevated in the “roaring forties” and in some coastal and high-latitude areas (apparently in sync with the drifters’ SDs). Correlations between the satellite minus drifters and satellite minus Argo SDs are noisy but statistically significant, with correlation coefficients $R \sim 0.35–0.41$. One anonymous reviewer of this paper suggested that “there are far more matchups in areas with low standard deviations,” but our special correlation analysis did not support this observation.

The similarity of the spatial patterns of AVHRR and AATSR biases suggests that errors in SSTs derived from these two IR sensors may be correlated, thus violating the major premise of the TCM. One possible way to quantify the sensitivity of the TCM results to the violation of this assumption would be adding an AVHRR-AATSR correlation coefficient as an extra parameter to the TCM equations and analyzing the TCM-derived individual SDs as a function of it. Such analysis may be performed in the future but is outside the scope of this study.

3.4. Single-Sensor Statistics

Uncertainties can also be calculated for each individual ship or buoy ID. Such analyses may give further insight into how uniform the single-sensor uncertainties are across different platforms within the same in situ data type family and provide a statistical model of their performance for use in data assimilation. Figure 4 shows histograms of biases and SDs, with fitted probability density functions (PDF) overlaid.

Figure 4a shows the histogram of the number of IDs versus number of matchups. Most in situ data have more than 10 matchups per ID, except Argo floats, which typically have fewer than 20 matchups per ID, resulting in the Argo statistics being most unstable.

The histograms of satellite minus in situ ΔTs are shown in Figures 4b and 4c and fitted with normal distribution PDFs, i.e.,

$$p(b) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{(b-\mu)^2}{2\sigma^2}\right] \tag{1}$$

Here b denotes the mean ΔT bias; $p(b)$ is the probability density; μ, σ are the two parameters of the Gaussian distribution, also overlaid in the figure. Note that the μ values in Figures 4b–4c, obtained via a two-step averaging (first for each ID and then between different IDs), should be close to the mean biases listed in Table 1a (obtained via a one-step averaging of all measurements from all IDs), but they may not be identical. The μ values for drifters, T- and C-moorings and Argo floats, are all close, whereas for ships, they are off by ~0.1 K due to the warm bias in ship SSTs. The PDF width is deemed to characterize the "sensor-to-sensor consistency" within each in situ data type. To that end, the SSTs measured by the drifters and T-moorings are most reproducible from one sensor to another, followed by the Argo floats and C-moorings. The ships are most nonuniform across different sensors in terms of the spread of estimated SST biases.

The histograms of the TCM-derived SDs are plotted in Figures 4d–4f. They are found to be fitted well with lognormal distributions (one might also choose other similar distributions, such as inverse gamma or inverse chi-square). The lognormal PDF is written as

$$p(d) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{(\ln d - \ln \mu)^2}{2\sigma^2}\right] \tag{2}$$

Here d denotes the SST SD; $p(d)$ denotes the probability density; μ, σ represent the corresponding mean and standard deviation parameters, respectively. Note that μ is defined in the same linear scale as d, while σ is defined in the log scale. An equivalent linear-scale σ' is thus defined using the equivalent 1 sigma extent as

$$2\sigma' = d^+ - d^-, \ln d^\pm = \ln \mu \pm \sigma \tag{3}$$

The fit parameters are overlaid in Figures 4d–4f (cf. SD statistics in Table 1b). Different in situ types have lognormal distributions with similar widths (except Argo floats, which do not have a sufficient number of matchups to produce stable statistics). This finding is potentially useful in modeling the a priori distribution of performance metrics of the ensemble of in situ sensors, which is particularly instrumental for the use in Bayesian-type applications. For example, producers of the L4 analyses blending satellite and in situ measurements could use more accurate error models and a priori parameters [e.g., *Reynolds et al.*, 2007]. As before, AVHRR SDs are less sensitive to the in situ data type used in the TCM than the AATSR SDs.

Note that similar PDF analyses were performed in *Xu and Ignatov* [2010] but using different ensembles (geographical rather than platform specific). The PDF fitting presented in this paper is expected to work better because the single-sensor statistics are more likely to obey (log) normal distributions than the regional/geographical biases and SDs. Moreover, the satellite L2/L3 data used here are more independent from in situ SSTs than the L3 Pathfinder and L4 Reynolds products used in *Xu and Ignatov* [2010].

4. Conclusion

Long-term error characterization in various in situ SST measurements is critical for Cal/Val of reprocessed historical satellite data. Because different in situ types have dominated over different periods, they should be intercalibrated to facilitate seamless transition in the satellite Cal/Val. As a first step toward this objective, this study has intercompared in situ SSTs from *i*Quam with two satellite SST products derived by the ESA CCI project from NOAA-17 AVHRR and Envisat AATSR for a 7 year period from January 2003 to December 2009. Three-way analyses were performed to estimate the relative biases and uncertainties in different in situ and satellite SSTs. The uncertainties measured by the respective SDs were estimated to be ~0.75 K for ships, ~0.21–0.22 K for drifters and Argo floats, and ~0.17 K and ~0.40 K for T- and C-moorings, respectively.

As a by-product of our analyses, relative biases and SDs of the CCI SST products have also been evaluated. The AATSR SST is more accurate, with a range of SD from 0.15 to 0.30 K (cf. SD ~ 0.16 K estimate obtained in *O'Carroll et al.* [2008]), compared with ~0.35–0.41 K for the AVHRR SST (cf. SD ~ 0.38 K for MODIS SST reported in *Gentemann* [2014]). These are very good performance statistics for satellite SST products. However, cold regional biases on the order of several tenths of a degree kelvin are present in both satellite products for the challenging areas affected by Saharan dust outbreaks, Indian aerosol over the Arabian Sea, and over the South China Sea affected by Kosa aerosols. More work is needed to improve the performance of the satellite IR-based SST products in those challenging areas.

The two near-global in situ data types currently best suited for satellite Cal/Val are drifters, in conjunction with geographically complementing tropical moorings, and Argo floats, although their coverage may not be fully uniform and/or globally representative. The Argo SST data are available in significantly smaller (2 orders of magnitude) numbers compared to drifters and may be useful for global and reduced temporal scale (e.g., monthly) validation but challenging for regional and fine timescale (e.g., daily) validation.

Errors for individual IDs show that their mean biases are well described by a Gaussian distribution, whose width is related to the consistency across individual platforms within each in situ data type. The corresponding SDs are well described by lognormal distributions. This result may be useful for specifying a priori probabilistic models of single in situ sensor errors in the data assimilation and quality control methods.

Acknowledgments
This work is supported by the Natural Science Foundation of China under grant 61571134. We thank our STAR colleague, Xinjia Zhou, for help with *i*Quam data. We also acknowledge the constructive feedback from anonymous reviewers, which helped us to improve the presentation of the results. The iQuam data can be obtained freely at ftp://ftp.star.nesdis.noaa.gov/pub/sod/sst/iquam/v2.00/. The CCI data were obtained from the Centre for Environmental Data Analysis (www.ceda.ac.uk). 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

Donlon, C., P. Minnett, C. Gentemann, T. Nightingale, I. Barton, B. Ward, and M. Murray (2002), Toward improved validation of satellite SST measurements for climate research, *J. Clim.*, *15*, 353–369.

Gentemann, C. L. (2014), Three way validation of MODIS and AMSR-E sea surface temperatures, *J. Geophys. Res. Oceans*, *119*, 2583–2598, doi:10.1002/2013JC009716.

Ignatov, A., X. Zhou, B. Petrenko, X. Liang, Y. Kihai, P. Dash, J. Stroup, J. Sapper, and P. DiGiacomo (2016), AVHRR GAC SST Reanalysis version 1 (RAN1), *Remote Sens.*, *8*(4), doi:10.3390/rs8040315.

*i*Quam v2 (2016), The in situ Quality monitor version 2. [Available at www.star.nesdis.noaa.gov/sod/sst/iquam/, Accessed Sep 4, 2016.]

Kennedy, J. (2014), A review of uncertainty in in situ measurements and data sets of sea surface temperature, *Rev. Geophys.*, *52*, 1–32, doi:10.1002/2013RG000434.

McColl, K., J. Vogelzang, A. Konings, D. Entekhabi, M. Piles, and A. Stoffelen (2014), Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target, *Geophys. Res. Lett.*, *41*, 6229–6236, doi:10.1002/2014GL061322.

Merchant, C. J., et al. (2014), Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI), *Geosci. Data J.*, doi:10.1002/gdj3.20.

O'Carroll, A., J. Eyre, and R. Saunders (2008), Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations, *J. Atmos. Oceanic Technol.*, *25*(7), 1197–1207.

Reynolds, R., 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. Clim.*, *20*(22), 5473–5496.

Roemmich, D., G. C. Johnson, S. Riser, R. Davis, J. Gilson, W. B. Owens, S. L. Garzoli, C. Schmid, and M. Ignaszewski (2009), The Argo Program: Observing the global ocean with profiling floats, *Oceanography*, *22*(2), 34–43.

Stoffelen, A. (1998), Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, *J. Geophys. Res.*, *103*, 7755–7766, doi:10.1029/97JC03180.

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.

Xu, F., and A. Ignatov (2014), In situ SST Quality Monitor (*i*Quam), *J. Atmos. Oceanic Technol.*, *31*(1), doi:10.1175/JTECH-D-13-00121.1.