Random Error in Space-Time Bin Averages of Sea Surface Temperature Observations from Ships

Alexey Kaplan[1]

[1]Lamont–Doherty Earth Observatory of Columbia University

Key Points:

- Relative biases between SST data sets from ships and satellites, averaged to one degree monthly bins, are estimated as climatological means
- Magnitudes of difference anomalies between one degree monthly averages of SST from ships and satellites agree with the random error model
- Separate estimates are obtained for sampling and measurement error components of the total error in bin averages of ship SST observations

Corresponding author: Alexey Kaplan, LDEO of Columbia University, P.O. Box 1000, 61 Route 9W, Palisades, NY 10964, USA; alexeyk@ldeo.columbia.edu

Abstract

Sea surface temperature (SST) observations made at ships are distributed irregularly in space and time and are affected by systematic biases and random errors. Such observations are often "binned": split into samples, contained within "bins" – grid boxes of a space-time grid (1°×1° monthly bins are used here), and their statistics are computed. Bin averages often serve as gridded representations of such data, thus requiring reliable uncertainty estimates, which for ship observations are particularly important because of their domination in the early observational records. Here ship SST observations for 1992–2010 are compared with an independent high-resolution satellite-based SST data set. To remove systematic biases, seasonal means were subtracted from the difference between bin-averaged data sets. In more than 66%(50%) of locations with binned temporal coverage exceeding 50%(66%), the magnitude of remaining anomalies agreed within 20%(10%) with random error model estimates. Separate estimates for sampling and measurement error components were obtained.

Plain Language Summary

Sea surface temperature (SST) is an important climate variable. SST observations made at ships are distributed irregularly in space and time and are affected both by systematic biases and randomly-varying measurement errors. To make them easier to use, such data sets are often "binned", i.e., split into samples contained within "bins", which usually are grid boxes of some space-time grid (monthly 1° longitude by 1° latitude bins are used here), and the statistics of these binned samples are computed. Bin averages often serve as gridded representations for data sets of ship observations; hence their uncertainty estimates have to be reliable. This is especially important since ship observations dominate early on in the historical observational record. Ship SST observations for 1992–2010 are compared here with an independent high-resolution satellite-based SST data set. To remove systematic biases, seasonal means were subtracted from the difference between bin-averaged versions of these data sets, and the remainder was interpreted as a sum of random errors. In more than 66%(50%) of locations where binned temporal coverage exceeded 50%(66%), the remainder's magnitude agreed within 20%(10%) with the random error model estimates. Error components due to incomplete sampling and due to measurement error were estimated.

1 Introduction

Sea surface temperature (SST) is one of the "essential" climate variables (Bojinski et al., 2014), particularly well-suited for monitoring changes in the Earth's mean surface temperature and very visible in the climate change debate (Hartmann et al., 2013). More than two centuries of SST observations together with other in situ data for surface ocean are assembled in the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, Woodruff et al., 1987; Freeman et al., 2017). These observations are irregularly distributed in space and time. A typical preparatory step for their use in climate studies is "binning," i.e., splitting them into subsamples, contained in non-overlapping spatiotemporal "bins", usually grid boxes of a regular space-time grid (hereinafter, "bins" are monthly 1°×1° grid boxes), and reporting statistical summaries of each bin's sample, e.g., number of observations $\mathcal{N}_o$ in the bin, their sample mean $\mathcal{M}_o$, standard deviation (SD) $\mathcal{S}_o$, etc. By construction, each of these statistics forms a gridded field, albeit usually incomplete. In lieu of averages over the complete bin's volume, which are generally unavailable, bin means $\mathcal{M}_o$ (a.k.a. "super-observations": Smith & Reynolds, 2005; Kennedy, 2014) are often used as input data for objective analyses or data assimilation; hence having reliable uncertainty estimates for binned data averages is important. For ship data this importance is especially high because of ships' domination, as an observational platform, in the early part of historical data record.

If binned observations were independent and identically distributed (i.i.d.) with variance

$$\sigma^2_{\mathcal{B}o} \stackrel{\text{def}}{=} \mathbb{E}\mathcal{S}^2_o, \tag{1}$$

the error SD estimate (ESDE) for $\mathcal{M}_o$ would be

$$e_{\mathcal{M}o} \stackrel{\text{def}}{=} \sigma_{\mathcal{B}o}/\sqrt{\mathcal{N}_o}\,. \tag{2}$$

Hereinafter label "def" above the "=" sign introduces an expression on the left of this sign as denoting the expression on its right, and $\mathbb{E}$ denotes mathematical expectation. Error estimates computed by (2), but with $\sigma_{\mathcal{B}o}$ calculated as the root-mean-square (RMS) over the best data coverage period of all bins' SDs in that location, was introduced by Kaplan et al. (1997) and used for objective analyses of historical SST observations by Kaplan et al. (1998), Ilin and Kaplan (2009), and, with further modifications to $\sigma_{\mathcal{B}o}$ estimate, by Karspeck et al. (2012). Their analyzed fields and uncertainty estimates provide some indirect validation for such uses of formula (2).

However, a direct comparison of the RMS difference between ICOADS bin means and satellite SST data with error estimates based on (2), while showing general large-scale agreement between global patterns of error magnitude, had many regional and smaller-scale differences (Rayner et al., 2010, cf. their Figures 1e vs. 1f). There were many possible reasons for this lack of detailed agreement: likely failure of the i.i.d. assumption, since the binned samples included observations from different platform types (ships, moorings, drifting buoys), obtained by different observational methods and affected by biases associated with particular methods of observations, platform types, or individual platforms (e.g., persistent thermometer biases on individual ships). Furthermore, the interpretation of that comparison was complicated by the dependence of the satellite-based SST data set on the in situ observations themselves.

A high-resolution interpolated SST analysis product (available on a daily $0.05°\times0.05°$ grid), based on the satellite data, independent of the concurrent in situ SST observations, and accompanied by verified uncertainty estimates, had become available several years ago (Merchant et al., 2014). The actual RMS differences for 1992-2010 between bin-averaged SST from ship observations and from this independent satellite-based analysis are compared here with their estimates based on the random error model and the analysis uncertainty estimates. While the bias structure of ship SST observations is complicated and remains a subject of active research (Kent et al., 2017), it is hypothesized here that a large part of biases in bin-averaged ship SST data can be approximated by its seasonally-dependent component. Once the climatological average is removed from the difference between ship and satellite bin means, the residual anomaly is treated as a combination of random errors. Separate estimates are obtained for sampling and measurement error components in bin-averaged ship SST data.

Section 2 describes data sets used and their pre-processing for this study. Section 3 presents error models, their estimates, and the technique of their comparison with the RMS of difference anomaly between bin-averaged versions of ICOADS ship SST and the satellite analysis product. Section 4 presents the results, which are discussed in section 5. Conclusions are given in section 6.

2 Data

2.1 High-resolution satellite SST analysis product

High-resolution globally-complete satellite SST data set, independent of in situ data (Merchant et al., 2014), produced within the Climate Change Initiative (CCI) of the European Space Agency (ESA), is used. It is based on the consistent re-processing of major global streams of the infrared satellite SST data, namely, the data from (Advanced)

Along-Track Scanning Radiometer and from the Advanced Very High Resolution Radiometer missions, with the deliberate avoidance of product dependency on the concurrent in situ SST observations (coefficients in SST retrievals were computed by optimal estimation, based on the atmospheric radiative transfer simulations, rather than by best-fitting in situ SST observations). In addition to the more traditional "skin" SST, the time-adjusted temperature at 20 cm depth was also produced, by modeling the near-surface thermally-stratified ocean layer. These temperature values with their uncertainty estimates were fed into the optimal interpolation system for the U.K. Met Office Ocean Sea Surface Temperature and Sea Ice Analysis (OSTIA, Donlon et al., 2012; Roberts-Jones et al., 2012, 2016), producing globally-complete ocean temperature fields at 20 cm depth, 0.05°≈6 km spatial resolution, interpretable as local-time daily averages, with uncertainty estimates, for 09/1991–12/2010. This product, known as ESA SST CCI Analysis, version 1.0, hereinafter will be referred to as "CCI Analysis" or simply "CCI." The period of complete 19 years (1992–2010) and 75°S–75°N global ocean domain will be used.

2.2 Ship Observations of SST

Ship observations of SST in ICOADS (Release 3.0; Freeman et al., 2017) were identified by the "Platform Type" indicator value (PT=5), corresponding to the "ship" observational platform type, and put through the ICOADS own quality control (QC) system with settings for the "enhanced Monthly Summary Trimmed Group". For each ship SST observation o that passed QC, its local time and date were computed and included into its record (only Coordinated Universal Time and date are in the ICOADS own data format). Then "match-up" SST value a^o and its uncertainty SD e^{ao} from the CCI Analysis to this ship observation o were extracted and also added to the record for o. Match-up values from the CCI Analysis to the given ship observation are the analyzed SST value and its uncertainty SD for the daily 0.05°×0.05° grid box within whose time-space limits this ship observation was made. (A small number of ICOADS ship SST observations that passed QC but did not have CCI Analysis match-up values were excluded from this study.)

2.3 Data Preparation

Consider bin $\mathcal{B}$, representing a grid box of a regular monthly 1°×1° grid, and a sample $\mathcal{B}_o$ of $\mathcal{N}_o$ SST observations from ships that were taken within its space and time limits and successfully passed ICOADS QC:

$$\mathcal{B}_o \stackrel{\text{def}}{=} \{o_1, o_2, \cdots, o_{\mathcal{N}_o}\}\,.$$

This "binned" sample is characterized by its mean $\mathcal{M}_o$ and SD $\mathcal{S}_o$, as follows:

$$\mathcal{M}_o \stackrel{\text{def}}{=} \frac{1}{\mathcal{N}_o}\sum_{i=1}^{\mathcal{N}_o} o_i\,, \qquad \mathcal{S}_o^2 \stackrel{\text{def}}{=} \frac{1}{\mathcal{N}_o-1}\sum_{i=1}^{\mathcal{N}_o}(o_i-\mathcal{M}_o)^2\,. \tag{3}$$

Note that $\mathcal{S}_o$ can only be computed if $\mathcal{N}_o>1$. Therefore bins with only one ship observation ($\mathcal{N}_o=1$) form a special class of data samples: their means, but not variability can be estimated directly from their data. Dealing with this more complicated subset is left for further investigaton, and only bins with $\mathcal{N}_o\geq 2$ are considered in this study.

Consider also a set $\mathcal{B}_a$ of $\mathcal{N}_a$ SST values from the CCI Analysis for all daily 0.05°×0.05° grid boxes contained within that same bin $\mathcal{B}$ as above:

$$\mathcal{B}_a \stackrel{\text{def}}{=} \{a_1, a_2, \cdots, a_{\mathcal{N}_a}\}\,.$$

Its statistics $\mathcal{M}_a$ and $\mathcal{S}_a$ are computed similarly to (3):

$$\mathcal{M}_a \stackrel{\text{def}}{=} \frac{1}{\mathcal{N}_a}\sum_{j=1}^{\mathcal{N}_a} a_j\,, \qquad \mathcal{S}_a^2 \stackrel{\text{def}}{=} \frac{1}{\mathcal{N}_a-1}\sum_{j=1}^{\mathcal{N}_a}(a_j-\mathcal{M}_a)^2\,. \tag{4}$$

Unless the land or ice cover are present within the bin $\mathcal{B}$, the number of data points in $\mathcal{B}_a$ is quite large: typically, $\mathcal{N}_a \sim 20\times20\times30 \gg \mathcal{N}_o$ for ocean locations. In fact, constraint $\mathcal{N}_a \geq 2$ is satisfied automatically for all non-empty binned samples $\mathcal{B}_a$ from the CCI Analysis.

Recall that for each $o_i \in \mathcal{B}_o$, its CCI SST match-up a_i^o has been identified and stored in the record for o_i (Section 2.2). Therefore it is easy to assemble a sample of CCI Analysis match-ups to ship observations in $\mathcal{B}_o$:

$$\mathcal{B}_{ao} \stackrel{\text{def}}{=} \{a_1^o, a_2^o, \cdots, a_{\mathcal{N}_o}^o\}$$

and to compute its statistics $\mathcal{M}_{ao}$ and $\mathcal{S}_{ao}$ analogously to (3). Additionally, differences between ship observations and their CCI Analysis SST match-ups

$$d_i \stackrel{\text{def}}{=} o_i - a_i^o, \quad i = 1, \cdots, \mathcal{N}_o \tag{5}$$

are binned as well, resulting in the sample

$$\mathcal{B}_d \stackrel{\text{def}}{=} \{d_1, d_2, \cdots, d_{\mathcal{N}_o}\}$$

and its bin statistics $\mathcal{M}_d$ and $\mathcal{S}_d$.

It will prove useful to have bin statistics for CCI Analysis uncertainties pre-computed as well. These are calculated in exactly the same way as was done above for corresponding SST values. Specifically, let

$$\mathcal{B}_{ea} \stackrel{\text{def}}{=} \{e_1^a, e_2^a, \cdots, e_{\mathcal{N}_a}^a\},$$

where each e_j^a is the uncertainty SD for the CCI Analysis SST value $a_j \in \mathcal{B}_a$ and compute $\mathcal{M}_{ea}$, $\mathcal{S}_{ea}$ analogously to (4). For the sample of the CCI Analysis uncertainty value match-ups to the ship observations in $\mathcal{B}$

$$\mathcal{B}_{eao} \stackrel{\text{def}}{=} \{e_1^{ao}, e_2^{ao}, \cdots, e_{\mathcal{N}_o}^{ao}\},$$

where each e_i^{ao} is the uncertainty SD for the CCI Analysis SST value $a_i^o \in \mathcal{B}_{ao}$ and compute $\mathcal{M}_{eao}$, $\mathcal{S}_{eao}$ analogously to (3).

Described above calculations of $\mathcal{M}_x$ and $\mathcal{S}_x$ are done for all monthly 1°×1° bins with sample sizes larger than one, i.e., for bins with $\mathcal{N}_o \geq 2$, when $x = o$, ao, d, or eao, and for all bins with non-empty samples when $x = a$ or ea. Temporal attribution of bin statistics $\mathcal{M}_x(y,m)$, $\mathcal{S}_x(y,m)$ is done using climatological (calendar) month $m = 1, \cdots, 12$ (January–December) and year $y = 1, \cdots, 19$ (corresponding to 1992–2010). Depending on location, statistics $\mathcal{M}_x(y,m)$, $\mathcal{S}_x(y,m)$ for $x = o$, ao, d, and eao might not be availabe for all (y,m) combinations. To simplify the formal treatment of their temporal averaging, for each location if there are $Y_m > 0$ years for month m with $\mathcal{N}_o \geq 2$, other years are skipped, while available years are renumbered as $y = 1, \cdots, Y_m$ for that month. If only $M < 12$ climatological months are available, these are renumbered as well from 1 to M. This change in temporal arguments is applied to statistics $\mathcal{N}_o$ and $\mathcal{M}_x$, $\mathcal{S}_x$ for $x = o$, ao, d, and eao, as well as for $\mathcal{M}_a$ and $\mathcal{M}_{ea}$, even though the latter two are available for all (y,m) for bins in ocean locations; statistics $\mathcal{N}_a$, $\mathcal{S}_a$, $\mathcal{S}_{ea}$ remain attributed to the full set of months and years in 1992-2010 period.

With these definitions, available bin averages of SST from ship observations $\mathcal{M}_o$ and corresponding bin averages from CCI Analysis $\mathcal{M}_a$ have the same temporal arguments, with the timeseries length

$$N \stackrel{\text{def}}{=} \sum_{m=1}^{M} Y_m \,. \tag{6}$$

Hence for their differences

$$d_{\mathcal{M}} \stackrel{\text{def}}{=} \mathcal{M}_o - \mathcal{M}_a \,, \tag{7}$$

the full period RMS is computed as

$$\mathcal{D} \stackrel{\text{def}}{=} \left[\frac{1}{N} \sum_{m=1}^{M} \sum_{y=1}^{Y_m} d_{\mathcal{M}}(y,m)^2 \right]^{1/2} . \tag{8}$$

3 Methods

3.1 Models and assumptions

CCI Analysis values a_j are estimates of water temperature at 20 cm depth, averaged over daily 0.05°×0.05° grid boxes. Corresponding "true" values t_j^a are averages of true water temperature t at 20 cm depth over such grid boxes, so for values within bin $\mathcal{B}$

$$t_j^a = a_j + \varepsilon_j^a, \qquad j = 1, \cdots, \mathcal{N}_a \,; \tag{9}$$

$$\mathbb{E}\varepsilon_j^a = 0\,, \;\; \mathbb{E}\left(\varepsilon_j^a\right)^2 = \left(e_j^a\right)^2, \quad j = 1, \cdots, \mathcal{N}_a\,, \tag{10}$$

where ε_j^a are the CCI analysis errors. These are assumed uncorrelated with the analyzed values a_j, since the CCI Analysis is a form of optimal interpolation (Lorenc, 1986). Analysis errors for different grid boxes, however, are not mutually independent, especially when these are not greatly separated in time and space. CCI Analysis uses the increased range (20–350 km) of spatial decorrelation scales of background error that resulted in improved feature resolution (Roberts-Jones et al., 2016), hence the analysis error is likely dominated by spatial scales larger than 1°×1° . Since the OSTIA background solution uses day-to-day persistence and relaxes to reference climatology with the 30 day decorrelation time scale (Donlon et al., 2012), a near-perfect correlation of the analysis error within monthly 1°×1° bins is assumed here:

$$\mathbb{E}\left(\varepsilon_j^a \varepsilon_k^a\right) \approx e_j^a e_k^a\,, \qquad j,k = 1, \cdots, \mathcal{N}_a\,. \tag{11}$$

For conceptual simplicity, the same "truth" definition, as for the CCI Analysis (9), is used for ship observations as well:

$$o_i = t_i^{ao} + b + \varepsilon_i^o\,, \quad i = 1, \cdots, \mathcal{N}_o\,, \tag{12}$$

where t_i^{ao} is true 20 cm depth temperature averaged over the daily 0.05°×0.05° grid box containing ship observation o_i, bias b is assumed constant within each 1°×1° monthly bin, thus it does not depend on i in (12). Measurement errors ε_i^o are assumed independent of true temperature variations t_i^{ao} and i.i.d. within each bin, with

$$\mathbb{E}\varepsilon_i^o = 0, \quad \mathbb{E}\left(\varepsilon_i^o\right)^2 = \sigma_o^2, \quad i = 1, \cdots, \mathcal{N}_o\,, \tag{13}$$

where σ_o^2 is an (unknown) measurement error variance.

Introduce bin average and intra-bin variance of true SST

$$\theta \stackrel{\text{def}}{=} \mathcal{M}_{ta}\,, \qquad \upsilon^2 \stackrel{\text{def}}{=} \frac{\mathcal{N}_a - 1}{\mathcal{N}_a} \mathcal{S}_{ta}^2\,, \tag{14}$$

where $\mathcal{M}_{ta}$ and $\mathcal{S}_{ta}$ are the mean and variance of the set of true SST values in the bin $\mathcal{B}$:

$$\mathcal{B}_{ta} \stackrel{\text{def}}{=} \{t_1^a, t_2^a, \cdots, t_{\mathcal{N}_a}^a\}.$$

Note that υ definition above changes the denominator of $\mathcal{S}_{ta}^2$ from $\mathcal{N}_a$–1 to $\mathcal{N}_a$, since (14) uses discrete analogues of the true value integrals over the bin, rather than statistical estimates.

Another important assumption is that times and locations of ship observations are random and uniformly distributed over the bin's volume. Hence the true SST match-ups to them form a set of $\mathcal{N}_o$ equiprobable draws

$$\mathcal{B}_{tao} \stackrel{\text{def}}{=} \{t_1^{ao}, t_2^{ao}, \cdots, t_{\mathcal{N}o}^{ao}\}$$

from the full set $\mathcal{B}_{ta}$ of the true SST values in the bin. Based on statistical theorems that lay the foundation of the classical Monte Carlo method for evaluating definite integrals (e.g., Section 3.2 of Robert & Casella, 2004), sample mean $\mathcal{M}_{tao}$ and variance $\mathcal{S}^2_{tao}$ of this set of random draws $\mathcal{B}_{tao}$ are unbiased estimates of the true mean and variance of the bin

$$\mathbb{E}\mathcal{M}_{tao} = \theta\,, \quad \mathbb{E}\mathcal{S}^2_{tao} = \upsilon^2\,, \tag{15}$$

and the error of sample mean, a.k.a. sampling error,

$$\varepsilon_s \stackrel{\text{def}}{=} \mathcal{M}_{tao} - \theta$$

has variance

$$\mathbb{E}\varepsilon_s^2 = \upsilon^2/\mathcal{N}_o \tag{16}$$

(for detailed derivation see Section 2.10 of Cochran, 1997).

3.2 Single bin statistics

3.2.1 CCI Analysis samples

Averaging equations (9) over j and using (14), obtain

$$\theta = \mathcal{M}_a + \mathcal{M}_{\varepsilon a}, \tag{17}$$

where $\mathcal{M}_{\varepsilon a}$ is the CCI Analysis error, averaged over the bin. Based on (10) and (11),

$$\mathbb{E}\mathcal{M}_{\varepsilon a} = 0,$$

$$e^2_{\mathcal{M}a} \stackrel{\text{def}}{=} \mathbb{E}\mathcal{M}^2_{\varepsilon a} = \frac{1}{\mathcal{N}_a^2}\sum_{j,k=1}^{\mathcal{N}_a}\mathbb{E}\big(\varepsilon_j^a\varepsilon_k^a\big) \approx \frac{1}{\mathcal{N}_a^2}\sum_{j,k=1}^{\mathcal{N}_a} e_j^a e_k^a = \frac{1}{\mathcal{N}_a^2}\left(\sum_{j=1}^{\mathcal{N}_a} e_j^a\right)^2 = \mathcal{M}^2_{ea}. \tag{18}$$

Subtracting (17) from (9), averaging squares of both sides over j, find, using (14), for mathematical expectation of both sides

$$\frac{\mathcal{N}_a}{\mathcal{N}_a - 1}\upsilon^2 = \mathbb{E}\mathcal{S}_a^2 + \mathbb{E}\mathcal{S}^2_{\varepsilon a}\,, \tag{19}$$

where $\mathcal{S}^2_{\varepsilon a}$ is the sample variance of the CCI analysis error in the bin. Using (10), (11), and (18), derive

$$\mathbb{E}\mathcal{S}^2_{\varepsilon a} = \frac{1}{\mathcal{N}_a - 1}\sum_{i=1}^{\mathcal{N}_o}\mathbb{E}(\varepsilon_i^a)^2 - \frac{\mathcal{N}_a}{\mathcal{N}_a - 1}\mathbb{E}\mathcal{M}^2_{\varepsilon a} = \mathcal{S}^2_{ea}\,. \tag{20}$$

Substituting (20) into (19), find that

$$\hat{\upsilon}^2 = \frac{\mathcal{N}_a - 1}{\mathcal{N}_a}\left(\mathcal{S}_a^2 + \mathcal{S}^2_{ea}\right)\,, \tag{21}$$

is an unbiased estimator of υ^2.

Subselecting from equations (9) those for the CCI analysis match-ups a^o to ship observations from $\mathcal{B}_o$,

$$t_i^{ao} = a_i^o + \varepsilon_i^{ao}, \quad i = 1, \cdots, \mathcal{N}_o\,, \tag{22}$$

analogously to the derivation of (21), find that

$$\hat{\upsilon}_o^2 = \mathcal{S}^2_{ao} + \mathcal{S}^2_{eao}\,, \tag{23}$$

is another unbiased estimator of υ^2, but based on a small subset of CCI Analysis grid points ($\mathcal{N}_o$ match-ups to ship observations) than the full set of $\mathcal{N}_a$ CCI Analysis points in $\mathcal{B}$, used in (21).

3.2.2 Ship observations sample

Averaging both sides of (12) over i, obtain

$$\mathcal{M}_o = \mathcal{M}_{tao} + b + \mathcal{M}_{\varepsilon o}, \tag{24}$$

where $\mathcal{M}_{\varepsilon o}$ is the bin mean of measurement errors with, based on (13),

$$\mathbb{E}\mathcal{M}^2_{\varepsilon o} = \sigma^2_o/\mathcal{N}_o\,. \tag{25}$$

Subtracting (24) from (12), averaging squares of both sides over i, find for mathematical expectation of both sides, using (13), (15), and recalling (1),

$$\sigma^2_{\mathcal{B}o} = \mathbb{E}\mathcal{S}^2_{tao} + \mathbb{E}\mathcal{S}^2_{\varepsilon o} = \upsilon^2 + \sigma^2_o\,. \tag{26}$$

Inserting t^{ao}_i from (22) into (12), obtain, recalling (5),

$$d_i = b + \varepsilon^o_i + \varepsilon^{ao}_i\,, \qquad i = 1, \cdots, \mathcal{N}_o\,. \tag{27}$$

By taking sample variances of both sides of (27) and considering their expectations, find an unbiased estimate of ship SST measurement error σ^2_o:

$$\hat{\sigma}^2_o = \mathcal{S}^2_d - \mathcal{S}^2_{eao}\,. \tag{28}$$

3.2.3 Bin mean differences

For differences between bin-averaged ship observations and CCI Analysis, defined by (5):

$$d_{\mathcal{M}} = b + \varepsilon_{d\mathcal{M}}\,, \tag{29}$$

where

$$\varepsilon_{d\mathcal{M}} \stackrel{\text{def}}{=} \varepsilon_s + \mathcal{M}_{\varepsilon o} + \mathcal{M}_{\varepsilon a}\,,$$

and based on (16), (18), (25), and (26),

$$\mathbb{E}\varepsilon_{d\mathcal{M}} = 0\,, \qquad e^2_{d\mathcal{M}} \stackrel{\text{def}}{=} \mathbb{E}\varepsilon^2_{d\mathcal{M}} = \sigma^2_{\mathcal{B}o}/\mathcal{N}_o + \mathcal{M}^2_{ea}\,. \tag{30}$$

3.3 Statistics for a temporal sample of bins

3.3.1 Actual RMS differences

Consider a temporal sample of bin statistics for a certain location of the bin. Due to (29), straight RMS $\mathcal{D}$ of differences $d_{\mathcal{M}}(y,m)$, calculated by (8), is affected by bias b. Bias estimate $\hat{b}_c(m)$ is obtained by climatological averaging of $d_{\mathcal{M}}(y,m)$:

$$\hat{b}_c(m) = \frac{1}{Y_m}\sum_{y=1}^{Y_m} d_{\mathcal{M}}(y,m)\,, \quad m = 1, \cdots, M\,. \tag{31}$$

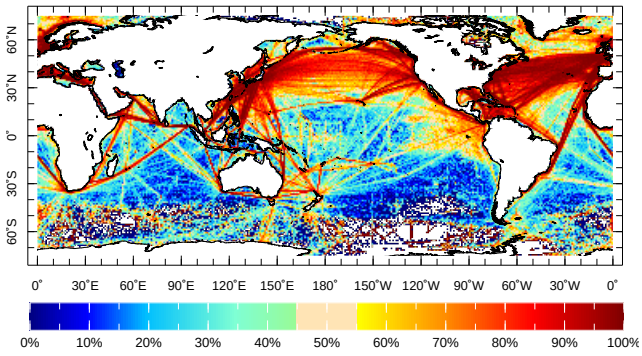
The RMS of the differences $d_{\mathcal{M}}$ with the estimated bias removed, taking into account the reduction in the number of degrees of freedom (DOF) from (6) to

$$\sum_{m=1}^{M}(Y_m - 1) = N - M\,,$$

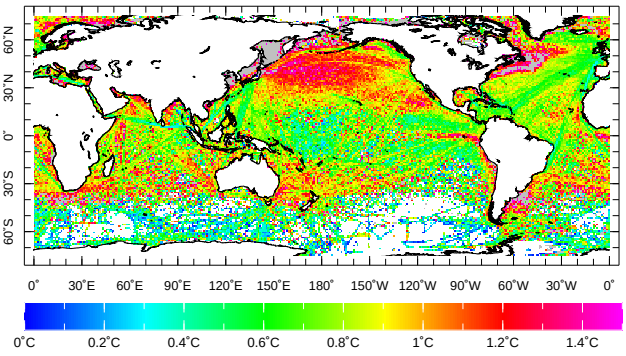
becomes

$$\mathcal{D}' = \left[\frac{1}{N-M}\sum_{m=1}^{M}\sum_{y=1}^{Y_m}\Big(d_{\mathcal{M}}(y,m) - \hat{b}_c(m)\Big)^2\right]^{1/2}. \tag{32}$$

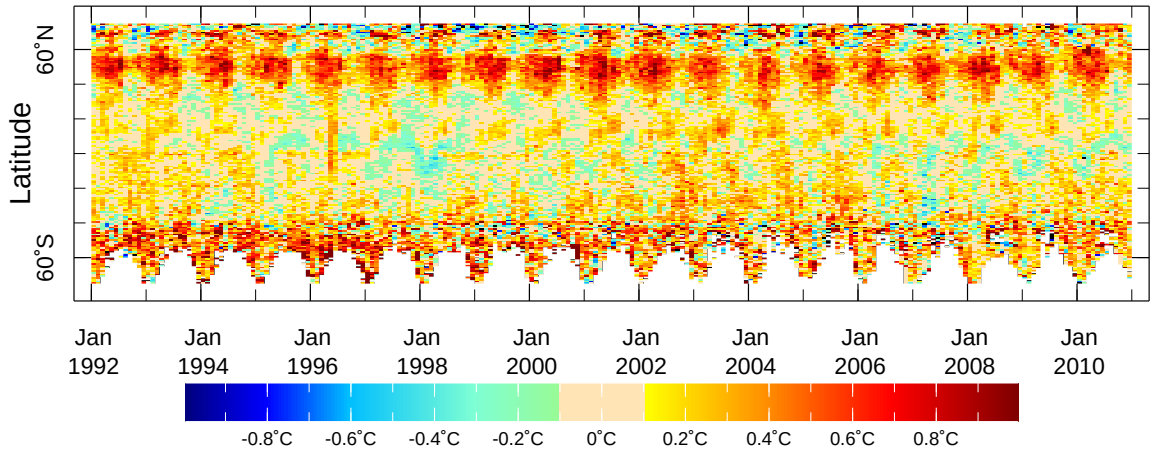
(a) % of monthly 1°×1° SST bins with $\mathcal{N}_o \geq 2$, 47.8%



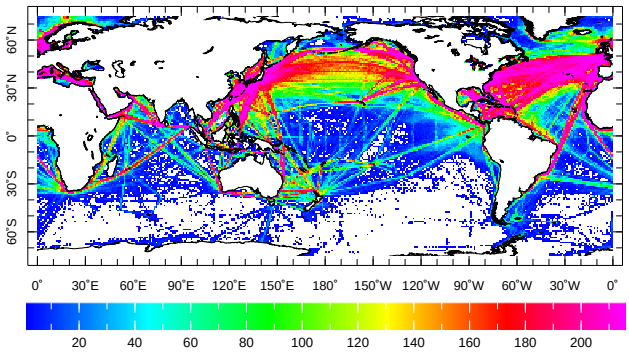
(b) RMS $\mathcal{D}$ of ship–satellite difference $d_{\mathcal{M}}$, 0.99°C



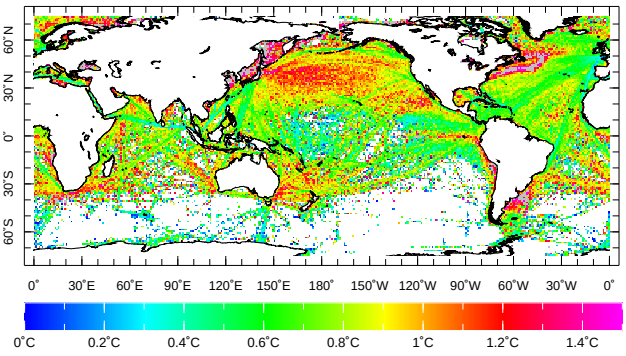
(c) Zonal means of SST differences $d_{\mathcal{M}}$,°C, between bin-averaged ship observations and CCI Analysis



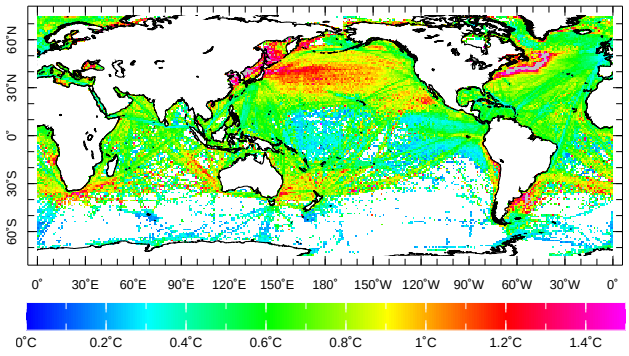
(d) Number of DOF in anomalies, 71.3



(e) RMS $\mathcal{D}'$ of ship–satellite anomaly $d_{\mathcal{M}} - \hat{b}_c$, 0.91°C



(f) Estimate $\mathcal{E}$ of binned anomaly RMS $\mathcal{D}'$, 0.74°C



(g) Relative discrepancy ρ, 81%

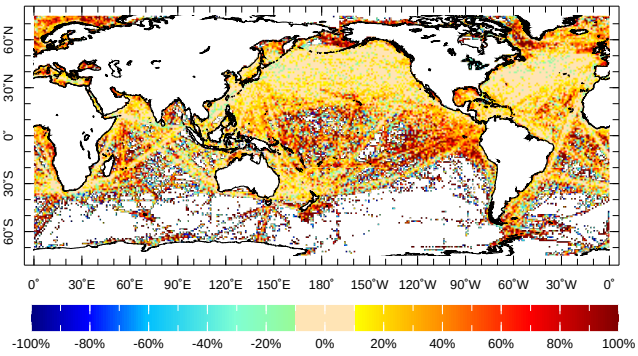


Figure 1. Comparison of monthly 1°×1° bin-averaged ($\mathcal{N}_o \geq 2$) ICOADS ship SST obserations with the ESA CCI Analysis for 1992-2010: (a) Percentage of ICOADS ship SST bins with $\mathcal{N}_o \geq 2$ among all bins with data ($\mathcal{N}_o \geq 1$); (b) RMS $\mathcal{D}$, °C, of difference $d_{\mathcal{M}}$ between bin-averaged ship and satellite data; (c) Zonal averages of differences $d_{\mathcal{M}}$, °C; (d) DOF in anomalies of bin-averaged ship data (zero DOF grids are shown as missing data, in white); (e) RMS $\mathcal{D}'$, °C, of ship–satellite difference anomalies $d_{\mathcal{M}}$-$\hat{b}_c$; (f) Estimate $\mathcal{E}$, °C, of ship–satellite difference anomaly RMS $\mathcal{D}'$; (g) Relative difference ρ, %, between $\mathcal{D}'$ and $\mathcal{E}$. Numbers at the end of panel labels are: for (a),(d) – global averages of displayed fields; for (b),(e),(f),(g) – global RMS of displayed fields.

3.3.2 Estimated RMS differences and errors

Based on (29) and (31),

$$\begin{aligned}
\mathbb{E}\mathcal{D}'^2 &= \frac{1}{N-M}\mathbb{E}\left[\sum_{m=1}^{M}\sum_{y=1}^{Y_m}\left(\varepsilon_{d\mathcal{M}}(y,m) - \frac{1}{Y_m}\sum_{q=1}^{Y_m}\varepsilon_{d\mathcal{M}}(y,m)\right)^2\right] = \\
&= \frac{1}{N-M}\sum_{m=1}^{M}\left[\mathbb{E}\sum_{y=1}^{Y_m}\varepsilon_{d\mathcal{M}}(y,m)^2 - \frac{1}{Y_m}\mathbb{E}\left(\sum_{q=1}^{Y}\varepsilon_{d\mathcal{M}}(y,m)\right)^2\right] = \\
&= \sum_{m=1}^{M}\frac{\mu_m}{Y_m}\sum_{y=1}^{Y_m} e_{d\mathcal{M}}(y,m)^2\,,
\end{aligned}$$

where

$$\mu_m \stackrel{\text{def}}{=} (Y_m - 1)/(N-M)\,, \quad m = 1, \cdots, M\,. \tag{33}$$

is the portion of the total DOF due to each climatological month m (note that $\sum_{m=1}^{M}\mu_m = 1$). Based on (30),

$$e_{d\mathcal{M}}(y,m)^2 = \sigma_{\mathcal{B}o}(m)^2/\mathcal{N}_o(y,m) + \mathcal{M}_{ea}(y,m)^2$$

and

$$\mathbb{E}\mathcal{D}'^2 = \sum_{m=1}^{M}\mu_m\sigma_{\mathcal{B}o}(m)^2/\mathcal{N}_o^h(m) + \sum_{m=1}^{M}\mu_m\mathcal{M}_{ea}^q(m)^2\,, \tag{34}$$

where

$$\mathcal{N}_o^h(m) \stackrel{\text{def}}{=} \left[\frac{1}{Y_m}\sum_{y=1}^{Y_m}\mathcal{N}_o(y,m)^{-1}\right]^{-1}, \qquad \mathcal{M}_{ea}^q(m) \stackrel{\text{def}}{=} \left[\frac{1}{Y_m}\sum_{y=1}^{Y_m}\mathcal{M}_{ea}(y,m)^2\right]^{1/2}.$$

are harmonic $\mathcal{N}_o(y,m)$ and quadratic $\mathcal{M}_{ea}(m)$ means of $\mathcal{N}_o$ and $\mathcal{M}_{ea}$, respectively, over available years $y = 1, \cdots, Y_m$.

An estimate of $\sigma_{\mathcal{B}o}(m)^2$ is computed as pooled variance (Section 9.2.16 in Von Storch & Zwiers, 2001) of binned samples over all available years $y = 1, \cdots, Y_m$:

$$\hat{\sigma}^2_{\mathcal{B}o}(m) \stackrel{\text{def}}{=} \sum_{y=1}^{Y_m}\gamma(y,m)\mathcal{S}_o(y,m)^2, \quad m = 1, \cdots, M, \tag{35}$$

where weighting coefficients are

$$\gamma(y,m) \stackrel{\text{def}}{=} [\mathcal{N}_o(y,m) - 1]\,/\Gamma(m), \quad y = 1, \cdots, Y_m, \quad m = 1, \cdots, M, \tag{36}$$

$$\Gamma(m) \stackrel{\text{def}}{=} \sum_{y=1}^{Y_m}[\mathcal{N}_o(y,m) - 1]\,, \qquad m = 1, \cdots, M\,. \tag{37}$$

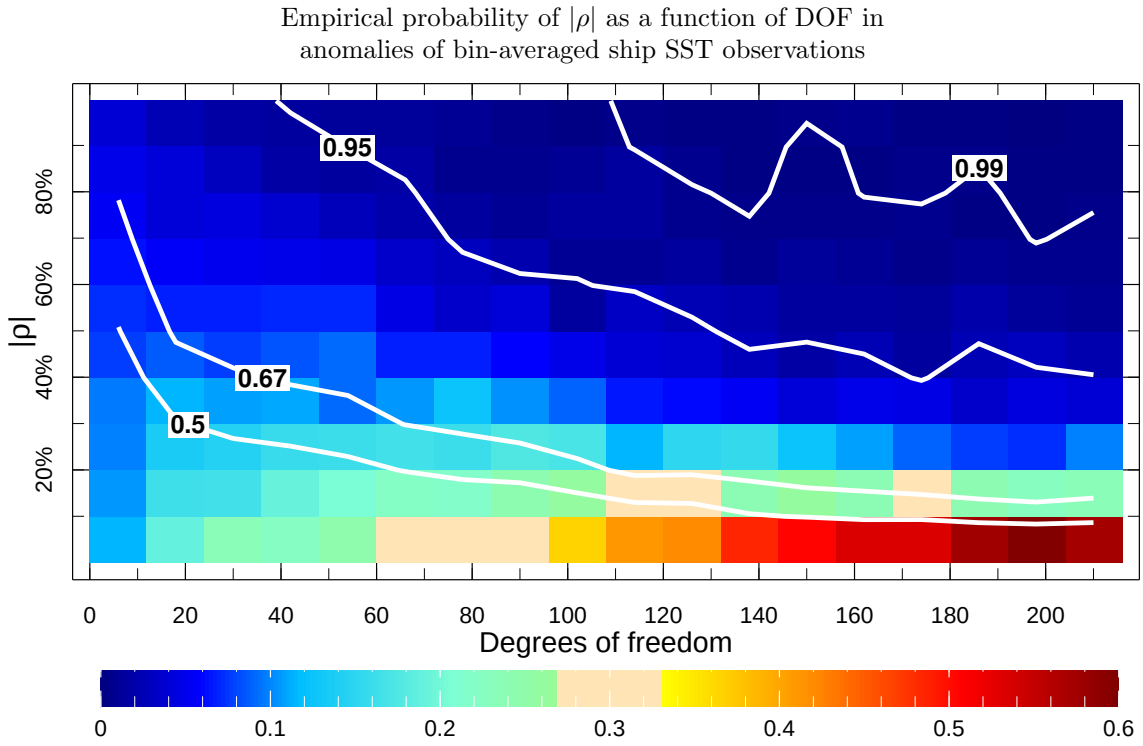


Figure 2. Empirical probability (color) of $|\rho|$ calculated for 10%-wide segments of 0–100% interval (vertical axis) for each of 12-wide sub-ranges of the complete 1–216 range of the possible DOF in the climatological anomaly sample for 1992-2010 (horizontal axis). White lines are contours of cumulative empirical probability of $|\rho|$, conditional on the given DOF range, corresponding to the values of 0.5, 0.67, 0.95, and 0.99, as labels indicate.

Substituting estimate $\hat{\sigma}_{\mathcal{B}o}(m)^2$ from (35) for the value of $\sigma_{\mathcal{B}o}(m)^2$ in (34), obtain an unbiased estimate for $\mathcal{D}'^2$:

$$\mathcal{E}^2 \stackrel{\text{def}}{=} \mathcal{E}^2_{\mathcal{M}o} + \mathcal{E}^2_{\mathcal{M}a}\,, \tag{38}$$

where the terms in the right-hand side are estimates of error variances in bin averages of ship observations

$$\mathcal{E}^2_{\mathcal{M}o} \stackrel{\text{def}}{=} \sum_{m=1}^{M} \mu_m \hat{\sigma}_{\mathcal{B}o}(m)^2 / \mathcal{N}^h_o(m) \tag{39}$$

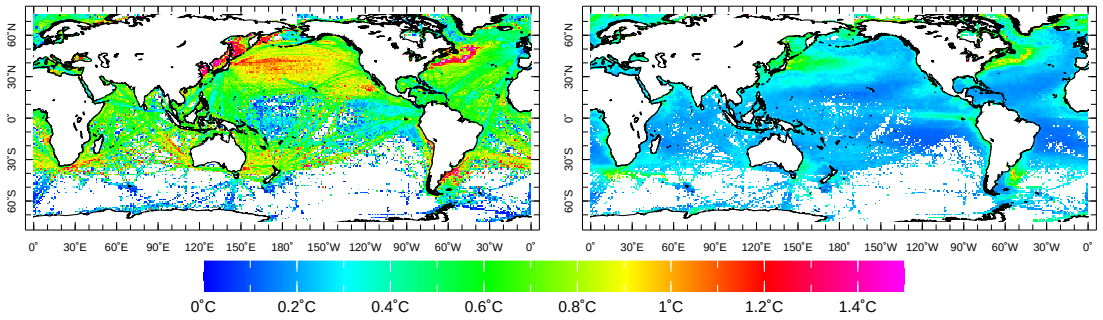
and of the CCI Analysis

$$\mathcal{E}^2_{\mathcal{M}a} \stackrel{\text{def}}{=} \sum_{m=1}^{M} \mu_m \mathcal{M}^q_{ea}(m)^2\,. \tag{40}$$

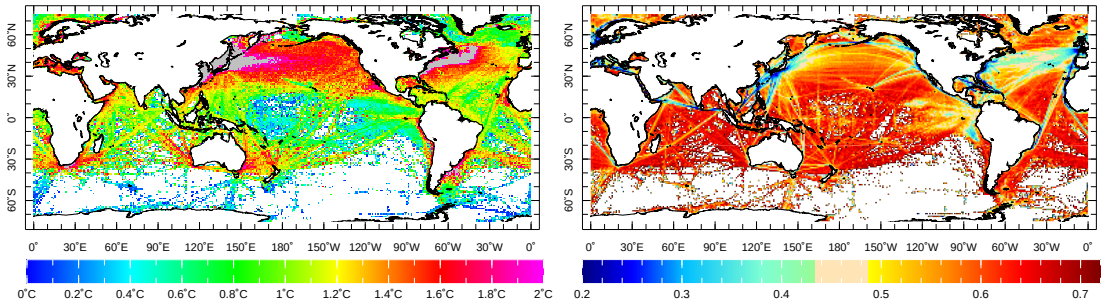
4 Results

Excluded from this study monthly 1°×1° bins with a single ship SST observation constitute a surprisingly large percentage (31.8%) of all ICOADS 1992-2010 monthly 1°×1° bins with ship SST observations (with any $\mathcal{N}_o > 0$) Figure 1a shows local percentages of bins included in this study ($\mathcal{N}_o \geq 2$) among all bins with ship SST data ($\mathcal{N}_o > 0$), identifying better-sampled areas in North Atlantic and North Pacific Oceans and along ship tracks. Figure 1b shows RMS $\mathcal{D}$ of differences $d_{\mathcal{M}}$ between bin averages of ship SST observations and CCI Analysis for 1992-2010 (see equations (7) and (8)).

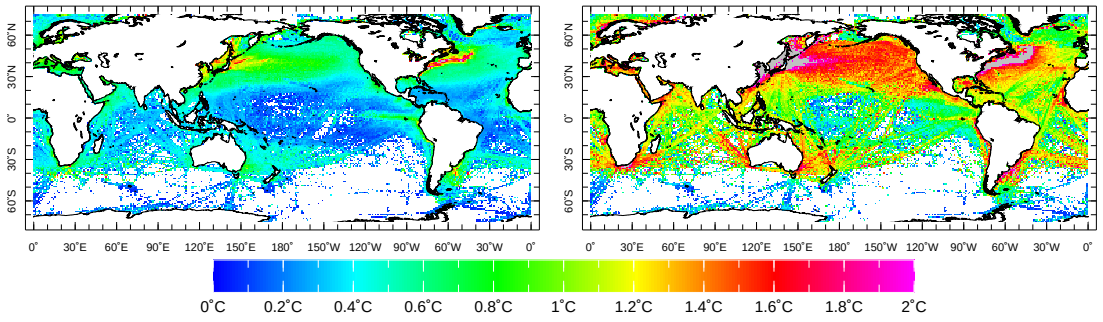
(a)ERMSE $\mathcal{E}_{\mathcal{M}o}$ of bin-averaged ship SST, 0.67°C (b)ERMSE $\mathcal{E}_{\mathcal{M}a}$ of bin-averaged CCI, 0.31°C



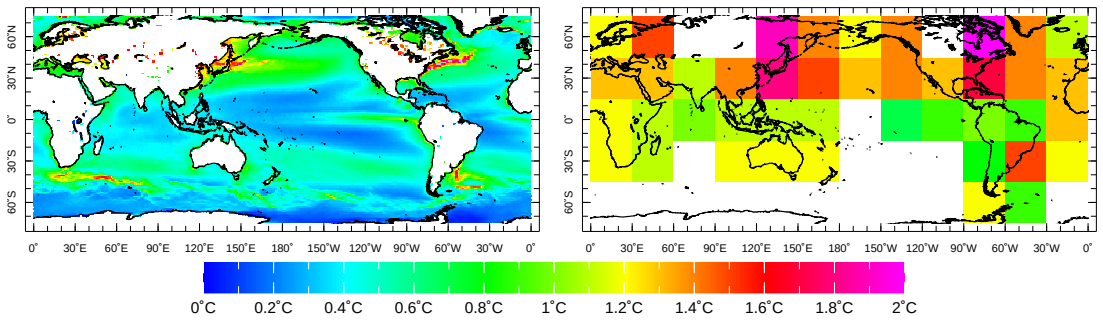
(c)Intra-bin SD of ship observations $\hat{\sigma}^*_{\mathcal{B}o}$, 1.20°C (d) Error reduction factor $1/\sqrt{\mathcal{N}^*}$, 0.57



(e) Sampling ESDE $\hat{v}^*_o$, 0.50°C (f) Ship measurement ESDE $\hat{\sigma}^*_o$, 1.14°C



(g) Sampling ESDE $\hat{v}^*$, 0.57°C (h) Kent&Challenor(2006) ship ESDE, 1.26°C



(i) same as (f), but in 30°×30° averages, 1.13°C (j) Relative difference ρ : (h) vs (i), 18%

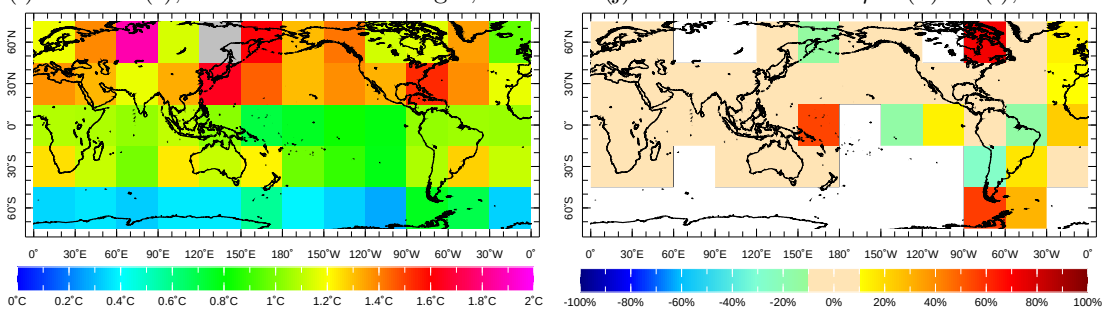


Figure 3. Components of estimated RMS difference anomaly between bin-averaged ship SST observations and CCI Analysis: (a) ERMSE $\mathcal{E}_{\mathcal{M}o}$ of bin-averaged ship SST, °C; (b) ERMSE $\mathcal{E}_{\mathcal{M}a}$ of bin-averaged CCI Analysis SST, °C; (c) Intra-bin SD $\hat{\sigma}^*_{\mathcal{B}o}$ of ship observations, °C; (d) Average error reduction factor $1/\sqrt{\mathcal{N}^*}$; (e) Sampling ESDE $\hat{v}^*_o$ from the CCI Analysis match-ups to ship observations, °C; (f) Measurement ESDE $\hat{\sigma}^*_o$ of ship SST, °C; (g) Sampling ESDE $\hat{v}^*$ from the full CCI Analysis, °C; (h) ship SST random ESDE from Kent and Challenor (2006, their Figure 2), °C; (i) same as (f), but in 30°×30° averages, °C; (j) Relative difference ρ between (h) and (i), %. Numbers at the end of panel labels indicate displayed fields' global RMS.

These differences have substantial mean and seasonal components, as seen in the time-latitude plot of zonally-averaged $d_{\mathcal{M}}$ (Figure 1c). Subtracting from $d_{\mathcal{M}}$ their climatological mean reduces the DOF by one for each climatological month, represented in the data (Figure 1d), but even accounting for the reduced DOF, the RMS $\mathcal{D}'$ of $d_{\mathcal{M}}$ anomaly, calculated by (32) and shown in Figure 1e, is appreciably smaller than $\mathcal{D}$ (8% global RMS reduction).

The difference $d_{\mathcal{M}}$ anomaly is interpreted here as the sum of random errors in bin averages of ship observations and of CCI analysis; the estimate $\mathcal{E}$ of its RMS $\mathcal{D}'$, based on this model, is computed by equation (38) and shown in Figure 1f. It matches $\mathcal{D}'$ pattern (Figure 1e) in many details. To aid their visual comparison, their difference

$$\rho = (\mathcal{D}' - \mathcal{E})/\mathcal{E}$$

is expressed as the percentage of the estimate $\mathcal{E}$ and is shown in Figure 1g, where large areas of the actual and estimated RMS agreeing within 10% or so are clearly seen.

The areas of poor agreement in Figure 1g appear to colocate with areas of smaller DOF in Figure 1d. To quantify this relationship, the empirical probability of $|\rho|$ in 10% intervals is shown in Figure 2 for different 12-wide DOF ranges of $d_{\mathcal{M}}$ anomalies (1-12, 13-24,..., 205-216). As DOF increases, $|\rho|$ concentrates more in its interval of smallest values. For more than 67% of points where DOF exceeds 50% of its maximum value (108= 0.5×12×(19-1)), $|\rho|$<20%; for more than half of the points, where for DOF exceeds 144 (2/3 of its maximum), $|\rho|$<10%.

The variance of difference anomaly between bin-averaged ship SST and CCI Analysis is modeled by (38), as a sum of squares of two components: estimated RMS error (ERMSE) $\mathcal{E}_{\mathcal{M}o}$ of bin-averaged ship observations, calculated by (39) and shown in Figure 3a, and ERMSE of bin-averaged CCI Analysis $\mathcal{E}_{\mathcal{M}a}$, calculated using (40) and shown in Figure 3b. The former clearly dominates: the CCI analysis error represents only 17.6% of the global variance in the total ERMSE $\mathcal{E}$ (Figure 1f).

As seen from (39), ERMSE for bin-averaged ship observations averages over the climatological month m products of intra-bin variance estimates $\hat{\sigma}_{\mathcal{B}o}(m)^2$ with inverse harmonic means $1/\mathcal{N}^h_o(m)$ of observational counts. Figures 3c,d show square roots of these quantities averaged over available climatological months:

$$\hat{\sigma}^*_{\mathcal{B}o} \stackrel{\text{def}}{=} \left[\sum_{m=1}^{M} \mu_m \hat{\sigma}^2_{\mathcal{B}o}(m)\right]^{1/2}, \qquad 1/\sqrt{\mathcal{N}^*} \stackrel{\text{def}}{=} \left[\sum_{m=1}^{M} \mu_m / \mathcal{N}^h_o(m)\right]^{1/2}, \tag{41}$$

where μ_m are defined by (33).

Figure 3c shows, in effect, the ESDE for the bin-averaged ship SST, if all monthly bins in the given location only had single observations in them, while 3d shows the ESDE reduction factor due to the multiple observations. Because of $\mathcal{N}_o \geq 2$ constraint, all values shown in 3d do not exceed $\sqrt{1/2} \approx 0.71$; their global RMS is 0.57, and the reductions to much smaller factors are relatively rare: the interquartile range is 0.51–0.64, and only 3.1% of shown grid boxes have a reduction factor below 0.3.

As seen from (26), the intra-bin variance $\sigma^2_{\mathcal{B}o}$ of ship observations consists of sampling and measurement error variance components. Using (23), (28), and pooled estimates like (35), these components can be estimated separately; with averaging analogous to (41), obtain

$$\hat{v}_o^{*2} = \sum_{m=1}^{M} \mu_m \sum_{y=1}^{Y_m} \gamma(y,m) \left[\mathcal{S}_{ao}(y,m)^2 + \mathcal{S}_{eao}(y,m)^2\right], \tag{42}$$

$$\hat{\sigma}_o^{*2} = \sum_{m=1}^{M} \mu_m \sum_{y=1}^{Y_m} \gamma(y,m) \left[\mathcal{S}_{d}(y,m)^2 - \mathcal{S}_{eao}(y,m)^2\right], \tag{43}$$

where γ is defined by (36),(37). The intra-bin sampling $\hat{v}_o^*$ and measurement $\hat{\sigma}_o^*$ ESDE for ship observations, computed by (42) and (43) are shown in Figures 3e,f. As with $\hat{\sigma}^*_{\mathcal{B}o}$, these are essentially ESDE components for a single observation, which are reduced by the factor $1/\sqrt{\mathcal{N}^*}$ (Figure 3d), when more observations are available.

5 Discussion

5.1 Sampling error

Estimate $\hat{v}_o^*$, given by (42) is based on the match-ups of the CCI Analysis SST and its uncertainty to the ship SST observations, a relatively small data sample. An estimate, based on the equation (21) that uses full CCI Analysis and its uncertainty

$$\hat{v}^{*2} = \frac{1}{228} \sum_{m=1}^{12} \sum_{y=1}^{19} \frac{\mathcal{N}_a(y,m) - 1}{\mathcal{N}_a(y,m)} \left[\mathcal{S}_a(y,m)^2 + \mathcal{S}_{ea}(y,m)^2\right]$$

is shown in Figure 3g. Expectedly, this estimate is larger (by about 10% in areas of high DOF numbers) and smoother than the one based on the incomplete data (Figure 3e). It has the uncanny similarity in pattern, but generally is larger than the estimate presented by Kennedy et al. (2011, their Figure 1d).

5.2 Measurement error

Kent and Challenor (2006) used the semivariogram method to estimate SST measurement error in 1970–1997 ICOADS data from ships. They identified pairs of ship SST observations made at the same hour and within 300 km of each other; squared differences between paired observations were binned by distance to construct the semivariogram; a linear fit to its points was extended towards zero distance separation to obtain the measurement error variance as the semivariogram's nugget. Ship measurement ESDE in 30°×30° averages from Kent and Challenor (2006, their Figure 2) is compared here with the measurement error estimates $\hat{\sigma}_o^*$, averaged to the same 30°×30° grid (Figure 3h,i). Two estimates have a great deal of similarity (their pattern correlation is 0.75), despite the differences in the study period and estimation method. Relative difference ρ, shown in Figure 3j has global RMS of 18.0%, with $|\rho| \leq 10\%$ in most of grid boxes. Grid boxes with $|\rho| > 10\%$ are generallyly in the areas of poor data coverage (cf. Figure 1d).

Kent and Berry (2008) introduced the measurement error model for marine observations that combines random error with a "platform-dependent" bias or "micro-bias", with the randomly distributed value over the platforms (ships). For this kind of error structure, if a bin contains many observations from a relatively small number of platforms, the error variance of its mean decreases inversely-proportionally to the number of platforms, rather than to the total number of observations. However, since moving ships, even at 14 knots (a relatively slow speed for modern ships), would cross the equatorial 1°×1° bin in less than six hours (a typical time interval between ship observations), multiple observations from the same ship would not typically appear in the same bin, thus making equation (2) usable in this study.

Kennedy (2014, Table 1) listed published in 1965-2011 ship SST measurement ESDE that did not separate micro-biases from the purely random error parts. There are 19 estimates there, ranging from 0.11°C to 3.5°C, with the median of 1.2°C, and 1–1.3°C interquartile range. Depending on the way of averaging measurement error estimates and especially on the averaging domain, global estimates can change appreciably. (Kent and Challenor (2006) report their global ESDE for ship SST random error as 1.2°C, if weighted by ocean area, and 1.3°C, if weighted by number of observations.) Estimates $\hat{\sigma}_o^*$ here can average to the global RMS of 1.14°C (Figure 3f), 1.13°C (Figure 3i), or 1.21°C, if the latter is constrained to the exact domain, where estimates in Figure 3h (global RMS of 1.26°C) are available.

6 Conclusions

Differences for 1992–2010 between monthly 1°×1° bin averages (for bins that contain more than one observation) of ICOADS ship SST and of the ESA SST CCI Analysis are presented here as the sum of their climatological bias component and remaining residuals (anomalies) with magnitudes that agree with the random error model in the areas of sufficient data coverage. The model assumes for ship observations the i.i.d measurement and sampling errors within bins and high intra-bin correlation for the CCI Analysis uncertainty. Location-dependent estimates of ship SST measurement and sampling error were obtained. Estimates of sampling error are similar in pattern, but larger than those previously published. Ship SST measurement error is consistent with previous estimates in spatial pattern and global RMS (1.13-1.21°C, depending on the averaging domain and procedure).

Acknowledgments

To the memory of M. Benno Blumenthal (1959-2018), creator of Data Library and Ingrid software, with which all figures in this Letter were calculated and plotted. Discussions with P.C.Cornillon, E.C.Kent, C.J.Merchant, J.J.Kennedy, and N.A.Rayner are gratefully acknowledged. ICOADS, Release 3.0, is obtained at

https://rda.ucar.edu/datasets/ds548.0/

ESA SST CCI Analysis for 9/1991–12/2010, version 1.0, is obtained at

https://catalogue.ceda.ac.uk/uuid/916986a220e6bad55411d9407ade347c

Supported by grants OCE-1853717 from NSF and NA17OAR4310156 from NOAA. LDEO contribution number xxxx.

References

Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., & Zemp, M. (2014). The concept of essential climate variables in support of climate research, applications, and policy. *Bulletin of the American Meteorological Society*, *95*(9), 1431–1443.

Cochran, W. G. (1997). *Sampling Techniques* (3rd ed.). New York: John Wiley & Sons.

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., & Wimmer, W. (2012). The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sensing of Environment*, *116*, 140–158.

Freeman, E., Woodruff, S. D., Worley, S. J., Lubker, S. J., Kent, E. C., Angel, W. E., et al. (2017). ICOADS Release 3.0: a major update to the historical marine climate record. *International Journal of Climatology*, *37*(5), 2211–2232.

Hartmann, D. L., Tank, A. M. K., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y. A. R., et al. (2013). Observations: Atmosphere and surface. In *Climate change 2013 the physical science basis: Working group i contribution to the fifth assessment report of the intergovernmental panel on climate change* (pp. 159–254). Cambridge University Press.

Ilin, A., & Kaplan, A. (2009). Bayesian PCA for Reconstruction of Historical Sea Surface Temperatures. In *IJCNN: 2009 INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, VOLS 1- 6* (p. 1138+). Int Neural Network Soc; IEEE Computat Intelligence Soc. (International Joint Conference on Neural Networks, Atlanta, GA, JUN 14-19, 2009)

Kaplan, A., Cane, M. A., Kushnir, Y., Clement, A. C., Blumenthal, M. B., & Rajagopalan, B. (1998). Analyses of global sea surface temperature 1856-1991. *Journal of Geophysical Research*, *103*(C9), 18567-18589.

Kaplan, A., Kushnir, Y., Cane, M. A., & Blumenthal, M. B. (1997). Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. *Journal of Geophysical Research*, *102*(C13), 27835-27860.

Karspeck, A. R., Kaplan, A., & Sain, S. R. (2012). Bayesian modelling and ensemble reconstruction of mid-scale spatial variability in North Atlantic sea-surface temperatures for 1850-2008. *Quarterly Journal of the Royal Meteorological Society*, *138*(662, A), 234-248.

Kennedy, J. J. (2014). A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Reviews of Geophysics*, *52*(1), 1–32.

Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E., & Saunby, M. (2011). Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *Journal of Geophysical Research: Atmospheres*, *116*(D14), D14103.

Kent, E. C., & Berry, D. I. (2008). Assessment of the marine observing system (asmos).

Kent, E. C., & Challenor, P. G. (2006). Toward estimating climatic trends in SST. Part II: Random errors. *Journal of Atmospheric and Oceanic Technology*, *23*(3), 476-486.

Kent, E. C., Kennedy, J. J., Smith, T. M., Hirahara, S., Huang, B., Kaplan, A., et al. (2017). A CALL FOR NEW APPROACHES TO QUANTIFYING BIASES IN OBSERVATIONS OF SEA SURFACE TEMPERATURE. *Bulletin of the American Meteorological Society*, *98*(8), 1601-1616.

Lorenc, A. C. (1986). Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, *112*(474), 1177–1194.

Merchant, C. J., Embury, O., Roberts-Jones, J., Fiedler, E., Bulgin, C. E., Corlett, G. K., et al. (2014). Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI). *Geoscience Data Journal*, *1*(2), 179-191.

Rayner, N. A., Kaplan, A., Kent, E. C., Reynolds, R. W., Brohan, P., Casey, K. S., et al. (2010). Evaluating climate variability and change from modern and historical SST observations. In J. Hall, D. E. Harrison, & D. Stammer (Eds.), *Proceedings of OceanObs09: Sustained Ocean Observations and Information for Society* (Vol. 2). ESA Publication WPP-306. (OceanObs' 09, Venice, Italy, September 21-25, 2009)

Robert, C., & Casella, G. (2004). *Monte Carlo Statistical Methods* (2nd ed.). New York: Springer–Verlag.

Roberts-Jones, J., Bovis, K., Martin, M. J., & McLaren, A. (2016). Estimating background error covariance parameters and assessing their impact in the OSTIA system. *Remote Sensing of Environment*, *176*, 117–138.

Roberts-Jones, J., Fiedler, E. K., & Martin, M. J. (2012). Daily, global, high-resolution SST and sea ice reanalysis for 1985–2007 using the OSTIA system. *Journal of Climate*, *25*(18), 6215–6232.

Smith, T. M., & Reynolds, R. W. (2005). A global merged land–air–sea surface temperature reconstruction based on historical observations (1880–1997). *Journal of climate*, *18*(12), 2021–2036.

Von Storch, H., & Zwiers, F. W. (2001). *Statistical analysis in climate research*. Cambridge University Press.

Woodruff, S. D., Slutz, R. J., Jenne, R. L., & Steurer, P. M. (1987). A comprehensive ocean-atmosphere data set. *Bulletin of the American meteorological society*, *68*(10), 1239–1250.