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. Uncertinty estimates for bin averages obtained under these assumptions translated into the estimated remainder's magnitude that was within 20% of its actual magnitude at 67% of all locations where the temporal coverage for ship data exceeded 50%. Furthermore, the estimates were within 10% of the actual values at 50% of locations with ship coverage exceeding 67%. Uncertainty components due to incomplete sampling and due to the measurement error were estimated as well.

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, 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": T. M. 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' dominance, as an observational platform, in the early part of historical data record.

If binned observations could be viewed as independent and identically distributed (i.i.d.) random variables with mean θ, equal to the true SST average over the full volume of the bin, and variance $\sigma^2_{\mathcal{B}o}$, equal to the full intra-bin variance of SST observations, then, obviously, we would have

$$\mathbb{E}\mathcal{M}_o = \theta, \qquad \mathbb{E}\mathcal{S}^2_o = \sigma^2_{\mathcal{B}o}\,, \tag{1}$$

with the error variance of the bin mean $\mathcal{M}_o$ being

$$e^2_{\mathcal{M}o} \stackrel{\text{def}}{=} \mathbb{E}\,(\mathcal{M}_o - \theta)^2 = \sigma^2_{\mathcal{B}o}/\mathcal{N}_o\,. \tag{2}$$

Hereinafter label "def" above the "=" sign introduces its left-hand side expression as a notation for its right-hand side, and $\mathbb{E}$ denotes mathematical expectation. The intra-bin variance $\sigma^2_{\mathcal{B}o}$ of SST observations is caused both by physical variations of the true SST throughout the bin's volume and by errors in SST measurements; the contributions from both these effects will be quantified in the analysis presented here.

Under the assumption that $\sigma^2_{\mathcal{B}o}$ depends on the bin's location, but is not changing in time, it can be estimated by averaging $\mathcal{S}^2_o$ statistics for that location over some period of relatively good data sampling. Using this approach, error estimates computed by (2) were 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 $\sigma^2_{\mathcal{B}o}$ estimate, by Karspeck et al. (2012). The usefulness of their analyzed fields and uncertainty estimates provides some indirect justification for such uses of formula (2).

However, a direct comparison of error estimates based on (2) with the actual RMS differences between bin means for ICOADS and for satellite SST data, 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). The reasons were many for this lack of detailed agreement: a likely failure of the i.i.d. assumption for bin samples that included SST observations from different platform types, e.g., ships, moorings, drifting buoys; which were obtained by different methods of SST observation; and which were affected by a multitude of systematic biases, thought to be associated with individual methods of observations, with specific types of observing platforms, and even with individual platforms, like persistent thermometer biases on some ships. Furthermore, the interpretation of that comparison was complicated by the dependence of the satellite-based SST data set (due to the commonly used satellite SST calibration procedures) on the in situ observations themselves.

A high-resolution interpolated SST analysis product on a daily 0.05°×0.05° 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). Here it is used in the error analysis of the monthly 1°×1° bin means of the ship-only subset of ICOADS SST observations. The actual RMS differences for the 1992-2010 period between bin-averaged SST from ships and from this satellite-based analysis are compared with estimates based on a version of model (2) for random errors in ship observations, combined with a simple climatological model for their biases, and accounting for the analysis uncertainty, using analysis error estimates supplied by Merchant et al. (2014) with their satellite SST analysis product. While the bias structure of ship SST observations is in reality quite complicated and remains a subject of active research (Kent et al., 2017; Huang et al., 2018; Chan et al., 2019; Kent & Kennedy, 2021), an admittedly simplistic approximation of biases in bin-averaged ship SST data by its seasonally-dependent component is used here. The goal of this study is to show that once the climatological average is removed from the difference between ship and satellite bin means, the residual anomaly can be treated, to a significant extent, as a combination of random errors despite the well-known limitations of the i.i.d. assumption. The multitude and space-time irregularity that characterize, beyond their seasonal dependence, the distributions of measurement method and ship-specific biases, combined with the relatively small (1°) spatial size of bins used here that are unlikely to contain successive measurements from the same moving ship, make the model, based on Eqs. (1),(2) useful for describing the variance of random error in bin means of ship SST observations. An additional advantage of the proposed approach, utilizing a high resolution SST analysis, is the development of separate estimates for sampling and measurement error components of bin-averaged ship SST data.

Section 2 describes the data sets used and their pre-processing for this study. Section 3 presents error models, constructs error estimates, and describes the technique of their comparison with the RMS of the actual difference anomaly between bin-averaged versions of ICOADS ship SST observations and the satellite data analysis product. Section 4 presents the results, which are discussed, together with their caveats, 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 the in situ data (Merchant et al., 2014) that was produced within the Climate Change Initiative (CCI) of the European Space Agency (ESA), is used here. 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 bestfitting 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 and 09/1991–12/2010 period, interpretable as local-time daily averages, with their uncertainties represented by error SD for 09/1991–12/2010. This product, known as ESA SST CCI Analysis, version 1.0, is referred to hereinafter as "CCI Analysis" or simply "CCI." The period of complete 19 years (1992–2010) and 75°S–75°N global ocean domain is 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 the QC flag settings intended for the creation of so called "enhanced Monthly Summary Groups" that trims off all observations outside of the 4.5 SD range from the ICOADS historical climatology (unless they are made in the area with no historical climatology available) and excludes duplicate reports as well as those from the landlocked locations and those whose observation time conflicts with the time range of their ICOADS data source (see S. R. Smith et al., 2016; Freeman et al., 2017, for more information). For each ship SST observation o that passed QC, its local time and date were computed and included into its record for further use in this study (only Coordinated Universal Time and date are in the ICOADS own data format). Then for each ship observation

o the CCI Analysis "match-up" SST value a^o (i.e., the CCI SST for the daily 0.05°×0.05° grid box within whose time-space limits ship observation o was taken) and its estimated error SD e^{ao} were identified and added to the record for o.

For the 1992-2010 period, ICOADS R3.0 contains around 23 million ship SST observations in the latitudinal range 75°S–75°N that pass the QC procedure described above. Among these observations, 3.2% do not have the CCI Analysis match-up values, being made in locations that are too close to the land to be included in the CCI Analysis domain. As Figure S1 illustrates, these are coastal, island, and lake observations. Such observations (lacking CCI match-ups) 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 $\mathcal{S}_o$ in (3) corresponds to the unbiased variance estimate $\mathcal{S}_o^2$ and 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.

Now consider 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:

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

Statistics $\mathcal{M}_a$ and $\mathcal{S}_a$ are computed as follows:

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

Since these represent the spatiotemporal mean and SD of the CCI Analysis SST within the bin $\mathcal{B}$ calculated from the complete set of the CCI Analysis grids covering bin $\mathcal{B}$, formula (4) for $\mathcal{S}_a^2$ has $\mathcal{N}_a$, rather than $\mathcal{N}_a - 1$ in denominator. 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, $\mathcal{N}_a \sim 20\times 20\times 30 \gg \mathcal{N}_o$ for ocean locations.

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

$$\mathcal{B}_{ea} \stackrel{\text{def}}{=} \{e^a_1, e^a_2, \cdots, e^a_{\mathcal{N}_a}\},$$

where each e^a_j is the error 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 $\mathcal{B}$

$$\mathcal{B}_{eao} \stackrel{\text{def}}{=} \{e^{ao}_1, e^{ao}_2, \cdots, e^{ao}_{\mathcal{N}_o}\},$$

where each e^{ao}_i is the error SD for the CCI Analysis SST value $a^o_i \in \mathcal{B}_{ao}$ and compute $\mathcal{M}_{eao}$, $\mathcal{S}_{eao}$ analogously to (3).

Calculations described above were performed to obtain $\mathcal{M}_x$ and $\mathcal{S}_x$ statistics with $x = o, ao, d,$ or eao for all monthly 1°×1° bins with $\mathcal{N}_o \geq 2$, while $\mathcal{M}_x$ and $\mathcal{S}_x$ with $x = a, ea$ were calculated for all monthly 1°×1° bins that contain any CCI Analysis grid points, included into the ocean domain (such bins de facto have $\mathcal{N}_a \geq 28$.) Temporal attribution of bin statistics $\mathcal{M}_x(y,m)$, $\mathcal{S}_x(y,m)$, as well as $\mathcal{N}_o(y,m)$, $\mathcal{N}_a(y,m)$ is done using climatological (calendar) month $m = 1, \cdots, 12$ (January–December) and year $y = 1992, \cdots, 2010$. For any given location, statistics $\mathcal{M}_x(y,m)$, $\mathcal{S}_x(y,m)$ for $x = o$, ao, d, and eao are only available when $\mathcal{N}_o \geq 2$ and are missing when $\mathcal{N}_o \leq 1$. To define rigorously the temporal averaging of available values for such statistics, for a given bin location and a given climatological month m, introduce a subset of years for which such statistical summaries of ship data are available:

$$\Upsilon_m \stackrel{\text{def}}{=} \left\{ y \in \{1992, \cdots, 2010\} \mid \mathcal{N}_o(m,y) \geq 2 \right\},$$

and let

$$Y_m \stackrel{\text{def}}{=} \left| \Upsilon_m \right|$$

be the number of elements in Υ_m, i.e., number of years with available summaries for the given location and climatological month m. Further, let

$$\mathfrak{M} \stackrel{\text{def}}{=} \left\{ m \in \{1, \cdots, 12\} \mid Y_m > 0 \right\}$$

be a set of climatological months for which bin summaries are available at least in one year in this location, with

$$M \stackrel{\text{def}}{=} \left| \mathfrak{M} \right|$$

being a number of such months.

With these definitions, for example, the differences between available bin averages of ship SST $\mathcal{M}_o$ and corresponding bin averages from CCI Analysis $\mathcal{M}_a$

$$d_{\mathcal{M}}(y,m) \stackrel{\text{def}}{=} \mathcal{M}_o(y,m) - \mathcal{M}_a(y,m)\,,\ y \in \Upsilon_m\,,\ m \in \mathfrak{M} \tag{6}$$

constitute a timeseries of length

$$N \stackrel{\text{def}}{=} \sum_{m \in \mathfrak{M}} Y_m\,, \tag{7}$$

and their RMS is calculated as

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

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 (OI) that like other Best Linear Unbiased Estimates (BLUE), e.g., multivariate linear regression (Section 8.4.2 in Von Storch & Zwiers, 2001), produces estimates that are independent of their errors. Specifically for the BLUE produced by kriging (of which OI is a special case called "simple kriging") see section 1.5 in the book by Stein (1999). While variances of analysis errors for different 0.05°×0.05° daily grid boxes can be assumed equal to squares of their SD estimates e_j^a, supplied by Merchant et al. (2014), additional assumptions are needed for their cross-covariances. These do not vanish, since the analysis errors are not mutually independent, especially for grid boxes that 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. Note that because of our definition of true temperature as t^{ao}, its differences $t^o - t^{ao}$ with precise water temperature t^o at the time, location, and depth of ship measurement effectively becomes a part of measurement error ε^o, and will contribute to its statistics, estimated in this study.

Now consider a set of the true SST values for the CCI Analysis grid points within the bin $\mathcal{B}$:

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

Its statistics $\mathcal{M}_{ta}$ and $\mathcal{S}_{ta}^2$, calculated analogously to (4) represent, by construction, the true SST average θ and the space-time variance υ^2 within the bin $\mathcal{B}$:

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

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 $\mathcal{B}_{tao}$calculated analogously to formulas (3), 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, obtain

$$\mathcal{S}^2_{ta} = \mathcal{S}^2_a + \frac{1}{\mathcal{N}_a}\sum_{j=1}^{\mathcal{N}_a}\left(a_j - \mathcal{M}_a\right)\left(\varepsilon_j^a - \mathcal{M}_{\varepsilon a}\right) + \mathcal{S}^2_{\varepsilon a}\,. \tag{19}$$

Due to the assumption of independence between a_j and ε_j^a terms in (9), the cross-terms under summation in the right-hand side of (19) drop out. Therefore, using(14), find for the mathematical expectation of both sides

$$\upsilon^2 = \mathbb{E}\mathcal{S}^2_a + \mathbb{E}\mathcal{S}^2_{\varepsilon a}\,, \tag{20}$$

where $\mathcal{S}^2_{\varepsilon a}$ is the space-time variance of the CCI analysis error within the bin $\mathcal{B}$, defined analogously to (4). Using (10), (11), and (18), derive

$$\mathbb{E}\mathcal{S}^2_{\varepsilon a} = \frac{1}{\mathcal{N}_a}\sum_{i=1}^{\mathcal{N}_a}\mathbb{E}(\varepsilon_i^a)^2 - \mathbb{E}\mathcal{M}^2_{\varepsilon a} \approx \frac{1}{\mathcal{N}_a}\sum_{i=1}^{\mathcal{N}_a}(e_i^a)^2 - \mathcal{M}^2_{ea} = \mathcal{S}^2_{ea}\,, \tag{21}$$

i.e., due to the assumption (11) of the near-perfect correlation of the CCI analysis errors within a bin, their intra-bin variance is approximated by the intra-bin sample variance $\mathcal{S}^2_{ea}$ of the analysis error estimates e_i^a; the latter variance is generally quite small:

it is different from zero only when the analysis error estimates e_i^a vary within the bin. Substituting (21) into (20), find that

$$v^2 \approx \mathbb{E}\mathcal{S}_a^2 + \mathcal{S}_{ea}^2 \,, \tag{22}$$

and therefore

$$\hat{v}^2 \stackrel{\text{def}}{=} \mathcal{S}_a^2 + \mathcal{S}_{ea}^2 \,, \tag{23}$$

is an approximately unbiased estimator of v^2.

Equations (20) and (22) make it clear that under the assumptions made here the expected space-time variance $\mathbb{E}\mathcal{S}_a^2$ of the CCI analysis within the bin $\mathcal{B}$ has to be smaller than such variance v^2 of the true SST field. Moreover, it is the expected variance $\mathbb{E}\mathcal{S}_{\varepsilon a}^2$ of the analysis error that makes up for the variance portion "lost"by the analysis field. To make this especially clear, recall that the outcome of a typical objective analysis, in the Bayesian interpretation, is a claim that the posterior distribution of the target field has its mean and covariance being equal to the analyzed field and to the analysis' error covariance, respectively (Lorenc, 1986; Handcock & Stein, 1993; Stein, 1999, sections 1.2,1.5). Arranging elements of sets $\mathcal{B}_{ta}$, $\mathcal{B}_a$, and $\mathcal{B}_{ea}$ as vector-columns of dimension $\mathcal{N}_a$

$$\mathbf{t}_a = (t_1^a, t_2^a, \cdots, t_{\mathcal{N}_a}^a)^T, \quad \mathbf{a} = (a_1, a_2, \cdots, a_{\mathcal{N}_a})^T, \quad \mathbf{e}_a = (e_1^a, e_2^a, \cdots, e_{\mathcal{N}_a}^a)^T \,,$$

where the superscript T denotes matrix transposition, the results of the CCI analysis, constrained to the bin $\mathcal{B}$, can be stated as

$$\mathbb{E}\mathbf{t}_a = \mathbf{a} \,, \qquad \mathbb{E}(\mathbf{t}_a - \mathbf{a})(\mathbf{t}_a - \mathbf{a})^T = \mathbf{e}_a\mathbf{e}_a^T. \tag{24}$$

If the normality assumption is made as well, the entire posterior distribution for the vector $\mathbf{t}_a$ is known:

$$\mathbf{t}_a \sim \mathbf{N}(\mathbf{a}, \mathbf{e}_a\mathbf{e}_a^T).$$

Here $\mathbf{N}(*,*)$ denotes a multivariate normal distribution with the arguments specifying its mean vector and covariance matrix. But even without the normality assumption, it follows from (24) that

$$\mathbb{E}\mathbf{t}_a\mathbf{t}_a^T = \mathbb{E}\mathbf{a}\mathbf{a}^T + \mathbf{e}_a\mathbf{e}_a^T \,, \tag{25}$$

since

$$\mathbb{E}(\mathbf{t}_a - \mathbf{a})(\mathbf{t}_a - \mathbf{a})^T = \mathbb{E}\mathbf{t}_a\mathbf{t}_a^T - \mathbb{E}\mathbf{a}\mathbf{a}^T \,.$$

Equation (22) can be easily re-derived from (25). By multiplying matrices in both sides of (25) by the $\mathbf{I}-\mathbf{1}\mathbf{1}^T/\mathcal{N}_a$, where $\mathbf{I}$ and $\mathbf{1}$ are respectively $\mathcal{N}_a \times \mathcal{N}_a$ identity matrix and $\mathcal{N}_a \times 1$ vector with all components equal one, obtain

$$\mathbb{E}(\mathbf{t}_a - \mathcal{M}_{ta}\mathbf{1})(\mathbf{t}_a - \mathcal{M}_{ta}\mathbf{1})^T = \mathbb{E}(\mathbf{a} - \mathcal{M}_a\mathbf{1})(\mathbf{a} - \mathcal{M}_a\mathbf{1})^T + (\mathbf{e}_a - \mathcal{M}_{ea}\mathbf{1})(\mathbf{e}_a - \mathcal{M}_{ea}\mathbf{1})^T \,.$$

Averaging diagonal elements of the matrices in both sides of this equation indeed produces equation (22).

Restating equations (9) for $a^o \in \mathcal{B}_{ao}$, i.e., the CCI analysis match-ups to ship observations $o \in \mathcal{B}_o$, obtain

$$t_i^{ao} = a_i^o + \varepsilon_i^{ao}, \quad i = 1, \cdots, \mathcal{N}_o \,, \tag{26}$$

and analogously to the derivation of (23), find, using (15), that

$$\hat{v}_o^2 = \mathcal{S}_{ao}^2 + \mathcal{S}_{eao}^2 \,, \tag{27}$$

is another approximately unbiased estimator of v^2. Unlike the estimate $\hat{v}$ that is given by (23) and is based on the full set $\mathcal{B}_a$ of $\mathcal{N}_a$ CCI Analysis points within the bin $\mathcal{B}$, the estimate $\hat{v}_o$ is based on the much smaller subset $\mathcal{B}_{ao}$ of $\mathcal{N}_o$ CCI Analysis match-ups to ship observationsin $\mathcal{B}_o$.

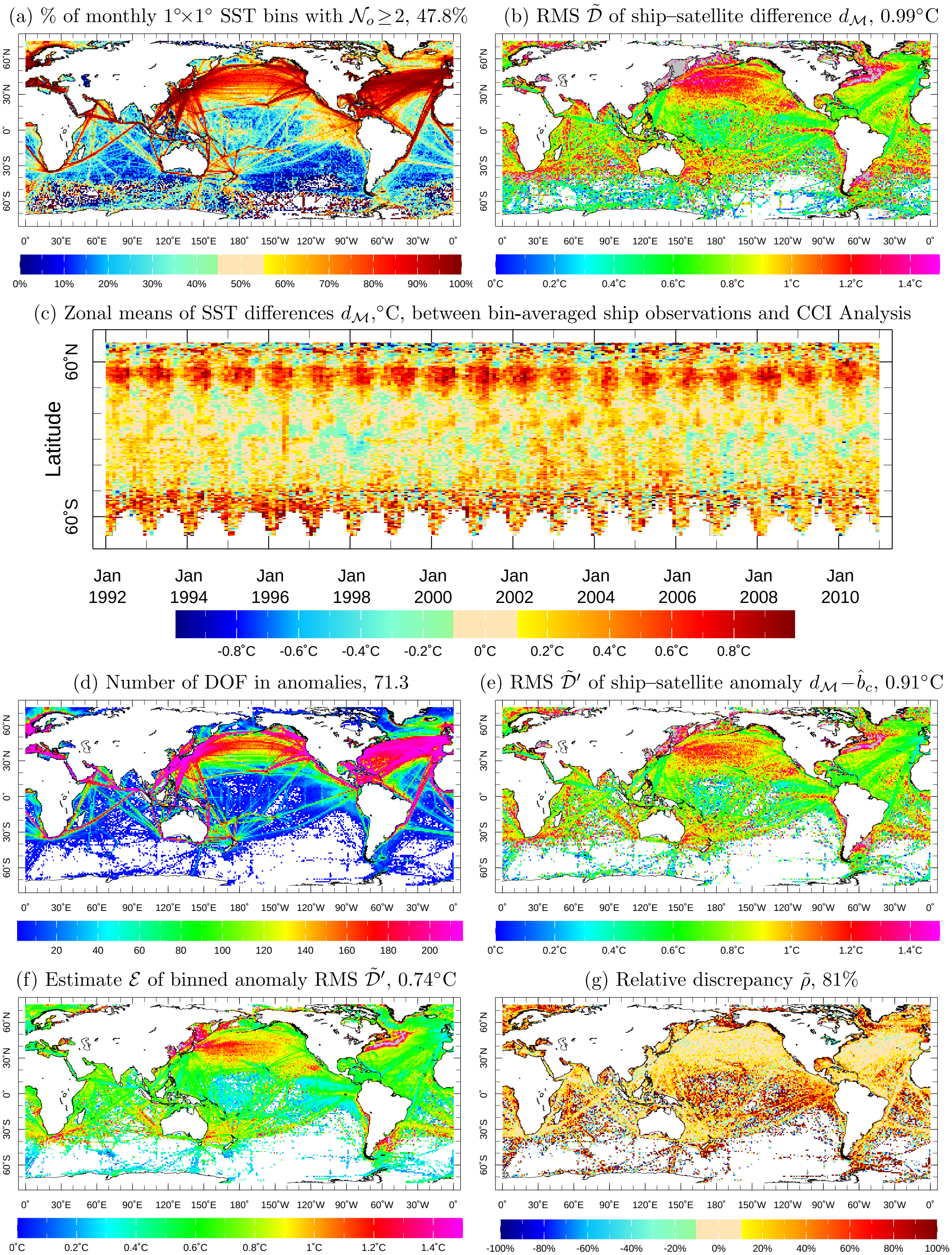
(a) % of monthly 1°×1° SST bins with $\mathcal{N}_o \geq 2$, 47.8%
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
(b) RMS $\tilde{\mathcal{D}}$ of ship–satellite difference $d_{\mathcal{M}}$, 0.99°C
0°C 0.2°C 0.4°C 0.6°C 0.8°C 1°C 1.2°C 1.4°C
(c) Zonal means of SST differences $d_{\mathcal{M}}$,°C, between bin-averaged ship observations and CCI Analysis
Latitude
60°N
60°S
Jan 1992 Jan 1994 Jan 1996 Jan 1998 Jan 2000 Jan 2002 Jan 2004 Jan 2006 Jan 2008 Jan 2010
-0.8°C -0.6°C -0.4°C -0.2°C 0°C 0.2°C 0.4°C 0.6°C 0.8°C
(d) Number of DOF in anomalies, 71.3
20 40 60 80 100 120 140 160 180 200
(e) RMS $\tilde{\mathcal{D}}'$ of ship–satellite anomaly $d_{\mathcal{M}} - \hat{b}_c$, 0.91°C
0°C 0.2°C 0.4°C 0.6°C 0.8°C 1°C 1.2°C 1.4°C
(f) Estimate $\mathcal{E}$ of binned anomaly RMS $\tilde{\mathcal{D}}'$, 0.74°C
0°C 0.2°C 0.4°C 0.6°C 0.8°C 1°C 1.2°C 1.4°C
(g) Relative discrepancy $\tilde{\rho}$, 81%
-100% -80% -60% -40% -20% 0% 20% 40% 60% 80% 100%
0° 30°E 60°E 90°E 120°E 150°E 180° 150°W 120°W 90°W 60°W 30°W 0°
60°N 30°N 0° 30°S 60°S

Figure 1. Comparison of monthly 1°×1° bin-averaged ($\mathcal{N}_o \geq 2$) ICOADS ship SST observations 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 $\tilde{\mathcal{D}}$, °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 $\tilde{\mathcal{D}}'$, °C, of ship–satellite difference anomalies $d_{\mathcal{M}}$-$\hat{b}_c$; (f) Estimate $\mathcal{E}$, °C, of ship–satellite difference anomaly RMS $\tilde{\mathcal{D}}'$; (g) Relative difference $\tilde{\rho}$, %, between $\tilde{\mathcal{D}}'$ 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.

3.2.2 Ship observations sample

Averaging both sides of (12) over i, obtain

$$\mathcal{M}_o = \mathcal{M}_{tao} + b + \mathcal{M}_{\varepsilon o}, \tag{28}$$

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

$$\mathbb{E}\mathcal{M}^2_{\varepsilon o} = \sigma^2_o / \mathcal{N}_o\,. \tag{29}$$

Subtracting (28) from equations (12), then summing up over i squares of both sides of the obtained equations and dividing the results by $\mathcal{N}_o - 1$, find for the mathematical expectation of both sides

$$\sigma^2_{\mathcal{B}o} \stackrel{\text{def}}{=} \mathbb{E}\mathcal{S}^2_o = \mathbb{E}\mathcal{S}^2_{tao} + \mathbb{E}\mathcal{S}^2_{\varepsilon o}, \tag{30}$$

and defining the intra-bin variance of ocean observations $\sigma^2_{\mathcal{B}o}$ as the left-hand side of this equation. Using (13) and (29), derive

$$\mathbb{E}\mathcal{S}^2_{\varepsilon o} = \frac{1}{\mathcal{N}_o - 1}\Big[\sum_{i=1}^{\mathcal{N}_o} \mathbb{E}(\varepsilon^o_i)^2 - \mathcal{N}_o \mathbb{E}\mathcal{M}^2_{\varepsilon o}\Big] = \frac{1}{\mathcal{N}_o - 1}\big[\mathcal{N}_o \sigma^2_o - \mathcal{N}_o \sigma^2_o / \mathcal{N}_o\big] = \sigma^2_o. \tag{31}$$

Inserting (15) and (31) into the right-hand side of (30), obtain

$$\sigma^2_{\mathcal{B}o} = \upsilon^2 + \sigma^2_o. \tag{32}$$

Equation (32) presents the intra-bin variance of ship SST observations as a sum of two terms: the spatiotemporal variance υ^2 of true SST within the bin and the variance σ^2_o of the SST measurement error on ships.

3.2.3 Matched-up differences

Inserting t^{ao}_i from (26) into (12) and recalling (5) definition for matched-up differences d_i, obtain

$$d_i = b + \varepsilon^o_i + \varepsilon^{ao}_i\,, \qquad i = 1, \cdots, \mathcal{N}_o\,. \tag{33}$$

By taking sample variances of both sides of (33) and considering their expectations, find

$$\mathbb{E}\mathcal{S}^2_d = \sigma^2_o + \mathbb{E}\mathcal{S}^2_{\varepsilon ao}\,, \tag{34}$$

therefore presenting the expected variance of matched-up differences as a sum of two terms, namely the variance of SST measurement error on ships σ^2_o and the expected error variance of the CCI analysis match-ups $\mathbb{E}\mathcal{S}^2_{\varepsilon ao}$ in the bin. The latter is approximated as

$$\mathbb{E}\mathcal{S}^2_{\varepsilon ao} \approx \mathcal{S}^2_{eao}\,, \tag{35}$$

based on a derivation similar to (21). From equations (34) and (21), an approximately unbiased estimate of ship SST measurement error σ^2_o is obtained:

$$\hat{\sigma}^2_o = \mathcal{S}^2_d - \mathcal{S}^2_{eao}\,. \tag{36}$$

Note that due to the assumption (11), the analysis error variance term $\mathcal{S}^2_{eao}$ in (36), as well as in (27), is relatively small, being different from zero only when error estimates e_i^{ao} for the CCI analysis match-ups vary within the bin.

3.2.4 Bin mean differences

For differences between bin-averaged ship observations and CCI Analysis, defined by (6):

$$d_{\mathcal{M}} = b + \varepsilon_{d\mathcal{M}}\,, \tag{37}$$

where

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

and based on (16), (18), (29), and (32),

$$\mathbb{E}\varepsilon_{d\mathcal{M}} = 0\,, \qquad e^2_{d\mathcal{M}} \stackrel{\text{def}}{=} \mathbb{E}\varepsilon^2_{d\mathcal{M}} = \sigma^2_{\mathcal{B}o}/\mathcal{N}_o + \mathcal{M}^2_{ea}\,. \tag{38}$$

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 (37), 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 $d_{\mathcal{M}}(y,m)$ over years $y \in \Upsilon_m$ with available bin summary statistics:

$$\hat{b}_c(m) = \frac{1}{Y_m}\sum_{y\in\Upsilon_m} d_{\mathcal{M}}(y,m)\,, \quad m \in \mathfrak{M}\,. \tag{39}$$

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 (7) to

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

becomes

$$\mathcal{D}' = \left[\frac{1}{N-M}\sum_{m\in\mathfrak{M}}\sum_{y\in\Upsilon_m}\left(d_{\mathcal{M}}(y,m) - \hat{b}_c(m)\right)^2\right]^{1/2}. \tag{40}$$

3.3.2 Estimated RMS differences and errors

Based on (37) and (39),

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

where

$$\mu_m \stackrel{\text{def}}{=} (Y_m - 1)/(N-M)\,, \quad m \in \mathfrak{M}\,. \tag{41}$$

is the portion of the total DOF due to each climatological month $m \in \mathfrak{M}$. Note that

$$\sum_{m\in\mathfrak{M}} \mu_m = 1.$$

Based on (38),

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

and

$$\mathbb{E}\mathcal{D}'^{\,2} = \sum_{m=1}^{M} \mu_m \sigma_{\mathcal{B}o}(m)^2/\mathcal{N}_o^h(m) + \sum_{m=1}^{M} \mu_m \mathcal{M}_{ea}^q(m)^2\,, \tag{42}$$

where

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

are harmonic $\mathcal{N}_o^h(m)$ and quadratic $\mathcal{M}_{ea}(m)$ means of $\mathcal{N}_o$ and $\mathcal{M}_{ea}$, respectively, over available years $y \in \Upsilon_m$ for a climatological month $m \in \mathfrak{M}$.

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 $y \in \Upsilon_m$ for each climatological month $m \in \mathfrak{M}$:

$$\hat{\sigma}_{\mathcal{B}o}(m)^2 \stackrel{\text{def}}{=} \sum_{y\in\Upsilon_m} \varphi(y,m)\mathcal{S}_o(y,m)^2, \quad m \in \mathfrak{M}, \tag{43}$$

where weighting coefficients are

$$\varphi(y,m) \stackrel{\text{def}}{=} \left[\mathcal{N}_o(y,m) - 1\right]/\Phi(m), \quad y \in \Upsilon_m, \quad m \in \mathfrak{M}\,, \tag{44}$$

with

$$\Phi(m) \stackrel{\text{def}}{=} \sum_{y\in\Upsilon_m} \left[\mathcal{N}_o(y,m) - 1\right], \qquad m \in \mathfrak{M}\,. \tag{45}$$

being the total number of degrees of freedom used in (43) for the pooled estimate $\hat{\sigma}_{\mathcal{B}o}(m)^2$.

Substituting estimate $\hat{\sigma}_{\mathcal{B}o}(m)^2$ from (43) for the value of $\sigma_{\mathcal{B}o}(m)^2$ in (42), obtain an unbiased estimate for $\mathcal{D}'^{\,2}$:

$$\mathcal{E}^2 \stackrel{\text{def}}{=} \mathcal{E}_{\mathcal{M}o}^2 + \mathcal{E}_{\mathcal{M}a}^2\,, \tag{46}$$

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

$$\mathcal{E}_{\mathcal{M}o}^2 \stackrel{\text{def}}{=} \sum_{m\in\mathfrak{M}} \mu_m \hat{\sigma}_{\mathcal{B}o}(m)^2/\mathcal{N}_o^h(m) \tag{47}$$

and of the CCI Analysis

$$\mathcal{E}_{\mathcal{M}a}^2 \stackrel{\text{def}}{=} \sum_{m\in\mathfrak{M}} \mu_m \mathcal{M}_{ea}^q(m)^2\,. \tag{48}$$

3.3.3 Bias correction of SD and RMS estimates

While $\mathcal{S}_o^2$ given by Equation (3) represents an unbiased estimate of the population variance, its square root $\mathcal{S}_o$ is a biased estimate of the population SD. For an i.i.d. random sample from a normal distribution, its unbiased variance estimate is proportional to a random value from the $\chi^2(f)$ distribution, where f is a number of DOF used in the variance estimate. Based on the properties of the $\chi^2(f)$ distribution, to obtain an unbiased estimate of SD, the square root of the estimated variance has to be multiplied by the correction factor (Holtzman, 1950)

$$c(f) = \sqrt{\frac{f}{2}}\,\Gamma\!\left(\frac{f}{2}\right)\Big/\Gamma\!\left(\frac{f+1}{2}\right), \tag{49}$$

where Γ denotes the gamma function. For example, for the bin sample with the unbiased variance estimate $\mathcal{S}_o^2$ given by Equation (3), the unbiased estimate of SD will be

$$\tilde{\mathcal{S}}_o = c\,(\mathcal{N}_o{-}1)\;\mathcal{S}_o\,. \tag{50}$$

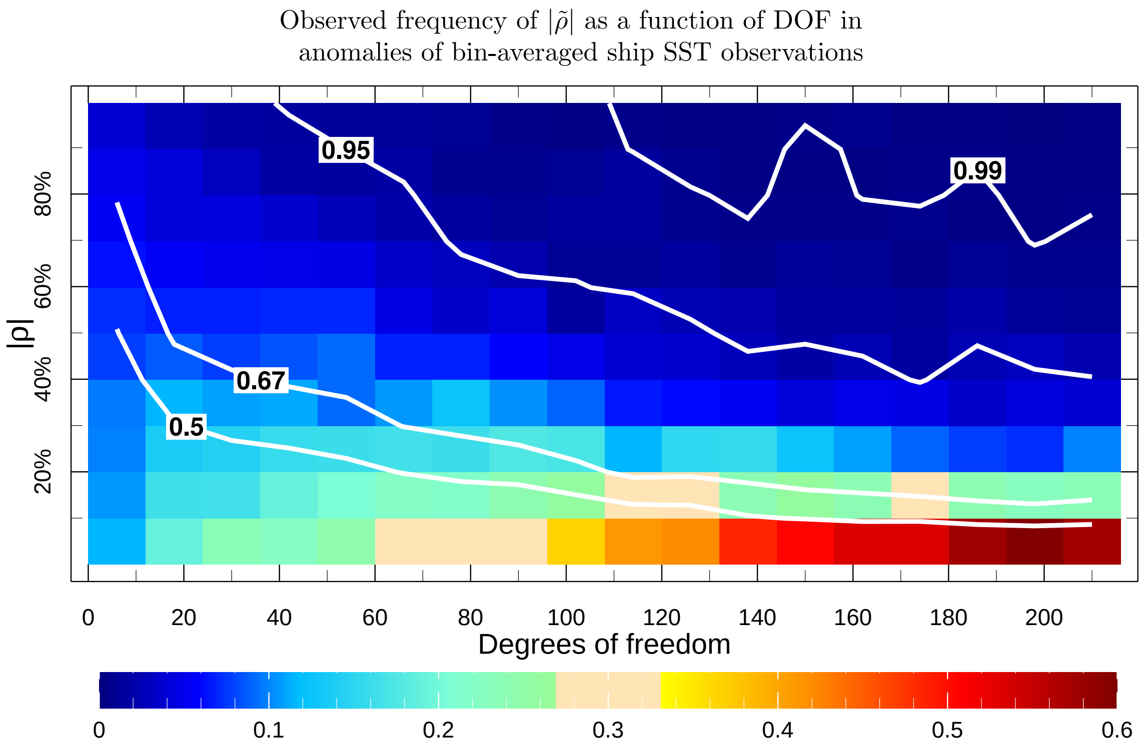

Figure 2. Observed frequency, a.k.a empirical probability, of $|\tilde{\rho}|$ (color) calculated for 10%-wide segments of 0–100% interval (vertical axis) for each of 12-wide sub-ranges of the complete 1–216 range of the possible DOF in the climatological anomaly sample for 1992-2010 (horizontal axis).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

Note that $c(f)$ is a monotonically decreasing function of real $f > 0$, and $c(f) \to 1$ as $f \to \infty$. Naturally, the largest corrections $c(f)$ are required for the smallest DOF numbers f: $c(f) \approx 1.25, 1.13, 1.09, 1.06, 1.05$ when $f = 1, 2, 3, 4, 5$, respectively, and $1 < c(f) < 1.01$ when f exceeds 25. The function $c(f)$ is illustrated by a graph in Figure S2.

When samples for the same calendar month m in different years are pooled together to produce a joint variance estimate $\hat{\sigma}_{\mathcal{B}o}(m)^2$, as given by (43), their DOF numbers add up to $\Phi(m)$, the total number of DOFs in the pooled sample, given by (45). Therefore the unbiased estimate of the intra-bin SD can be obtained by

$$\tilde{\sigma}_{\mathcal{B}o}(m) = c\left(\Phi(m)\right)\, \hat{\sigma}_{\mathcal{B}o}(m). \tag{51}$$

Since $\Phi(m)$ is a sum over all available years $y \in \Upsilon_m$ of the DOF numbers $\mathcal{N}_o(y,m) - 1$ in binned variance estimates for the given location and month m, the argument of function c in Equation (51) is generally much larger than its argument in Equation (50) for the unbiased SD estimate from individual bin samples, thus signifying that a smaller correction is required for producing an unbiased estimate of $\sigma_{\mathcal{B}o}(m)$.

When $\hat{\sigma}_{\mathcal{B}o}(m)^2$ estimates are averaged over all available calendar months $m \in \mathfrak{M}$ as is done in (47) to estimate the error variance in bin averages of ship obsevations, the total DOF number in this calculation becomes

$$\Phi = \sum_{m \in \mathfrak{M}} \Phi(m). \tag{52}$$

However, when $M = |\mathfrak{M}| > 1$, unless the coefficients $\mu_m/\mathcal{N}_o^h(m)$ multiplying individual $\hat{\sigma}_{\mathcal{B}o}(m)^2$ estimates happen to be proportional to $\Phi(m)$, the sum in the right-hand side of (47) will not obey the $\chi^2(\Phi)$ distribution. Therefore the correction factor $c(\Phi)$ is not be applicable in this case, as being too small; a factor $c(f)$, corresponding to a certain DOF number f, lying in the interval

$$\min_{m\in\mathfrak{M}} \Phi(m) < f < \Phi$$

would be necessary for the precise correction.

Analogously to the corrections introduced above, RMS estimates of ship – satellite differences $\mathcal{D}$ and difference anomalies $\mathcal{D}'$, given by formulas (8) and (40), respectively, also need corrections to become unbiased estimates:

$$\tilde{\mathcal{D}} = c(N)\,\mathcal{D}\;, \qquad \tilde{\mathcal{D}}' = c(N{-}M)\,\mathcal{D}'\;. \tag{53}$$

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 ($\mathcal{N}_o$ ≥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 (6) and (8)).

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 (40) 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 (46) and shown in Figure 1f. It matches $\mathcal{D}'$ pattern (Figure 1e) in many details. To aid their visual comparison, their difference

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

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 observed frequency, a.k.a. 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 (46), as a sum of squares of two components: estimated RMS error (ERMSE) $\mathcal{E}_{\mathcal{M}o}$ of bin-averaged ship observations, calculated by (47) and shown in Figure 3a, and ERMSE of bin-averaged CCI Analysis $\mathcal{E}_{\mathcal{M}a}$, calculated using (48) 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).

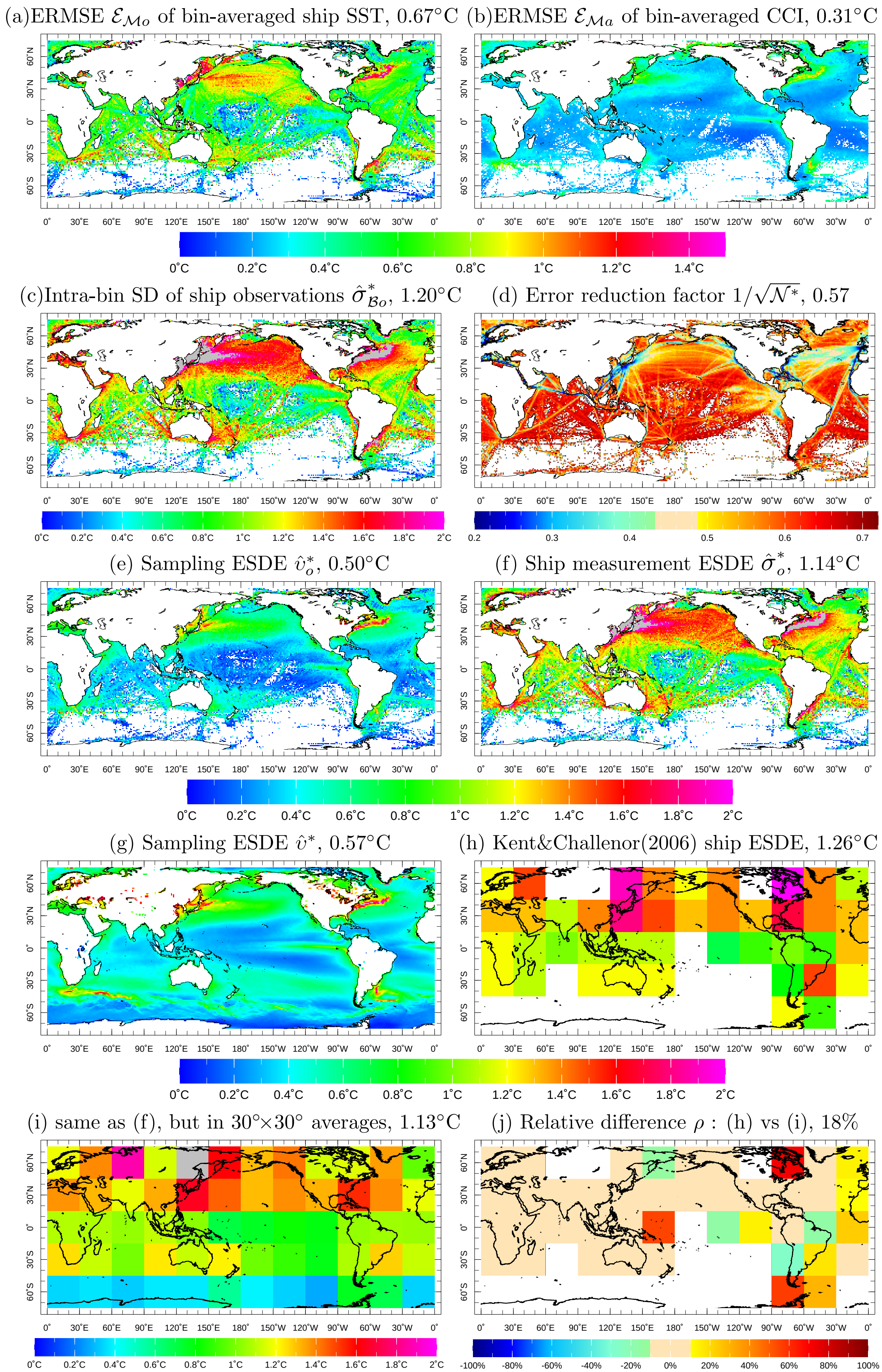
(a)ERMSE $\mathcal{E}_{\mathcal{M}o}$ of bin-averaged ship SST, 0.67°C
(b)ERMSE $\mathcal{E}_{\mathcal{M}a}$ of bin-averaged CCI, 0.31°C
(c)Intra-bin SD of ship observations $\hat{\sigma}^*_{\mathcal{B}o}$, 1.20°C
(d) Error reduction factor $1/\sqrt{\mathcal{N}^*}$, 0.57
(e) Sampling ESDE $\hat{\upsilon}^*_o$, 0.50°C
(f) Ship measurement ESDE $\hat{\sigma}^*_o$, 1.14°C
(g) Sampling ESDE $\hat{\upsilon}^*$, 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 $\hat{\upsilon}^*_o$ from the CCI Analysis match-ups to ship observations, °C; (f) Measurement ESDE $\hat{\sigma}^*_o$ of ship SST, °C; (g) Sampling ESDE $\hat{\upsilon}^*$ 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.

As seen from (47), 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:

$$\hat{\sigma}^*_{\mathcal{B}o} \stackrel{\text{def}}{=} \left[\sum_{m\in\mathfrak{M}} \mu_m \hat{\sigma}^2_{\mathcal{B}o}(m)\right]^{1/2}, \qquad 1/\sqrt{\mathcal{N}^*} \stackrel{\text{def}}{=} \left[\sum_{m\in\mathfrak{M}} \mu_m/\mathcal{N}^h_o(m)\right]^{1/2}, \tag{54}$$

where μ_m are defined by (41).

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 (32), the intra-bin variance $\sigma^2_{\mathcal{B}o}$ of ship observations consists of sampling and measurement error variance components. Using (27), (36), and pooled estimates like (43), these components can be estimated separately; with averaging analogous to (54), obtain

$$\hat{\upsilon}^{*2}_o = \sum_{m\in\mathfrak{M}} \mu_m \sum_{y\in\Upsilon_m} \varphi(y,m)\left[\mathcal{S}_{ao}(y,m)^2 + \mathcal{S}_{eao}(y,m)^2\right], \tag{55}$$

$$\hat{\sigma}^{*2}_o = \sum_{m\in\mathfrak{M}} \mu_m \sum_{y\in\Upsilon_m} \varphi(y,m)\left[\mathcal{S}_{d}(y,m)^2 - \mathcal{S}_{eao}(y,m)^2\right], \tag{56}$$

where φ is defined by (44),(45). The intra-bin sampling $\hat{\upsilon}^*_o$ and measurement $\hat{\sigma}^*_o$ ESDE for ship observations, computed by (55) and (56) 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 $\hat{\upsilon}^*_o$, given by (55) 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 (23) that uses full CCI Analysis and its uncertainty

$$\hat{\upsilon}^{*2} = \frac{1}{228}\sum_{m=1}^{12}\sum_{y=1992}^{2010}\left[\mathcal{S}_a(y,m)^2 + \mathcal{S}_{ea}(y,m)^2\right]$$

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).

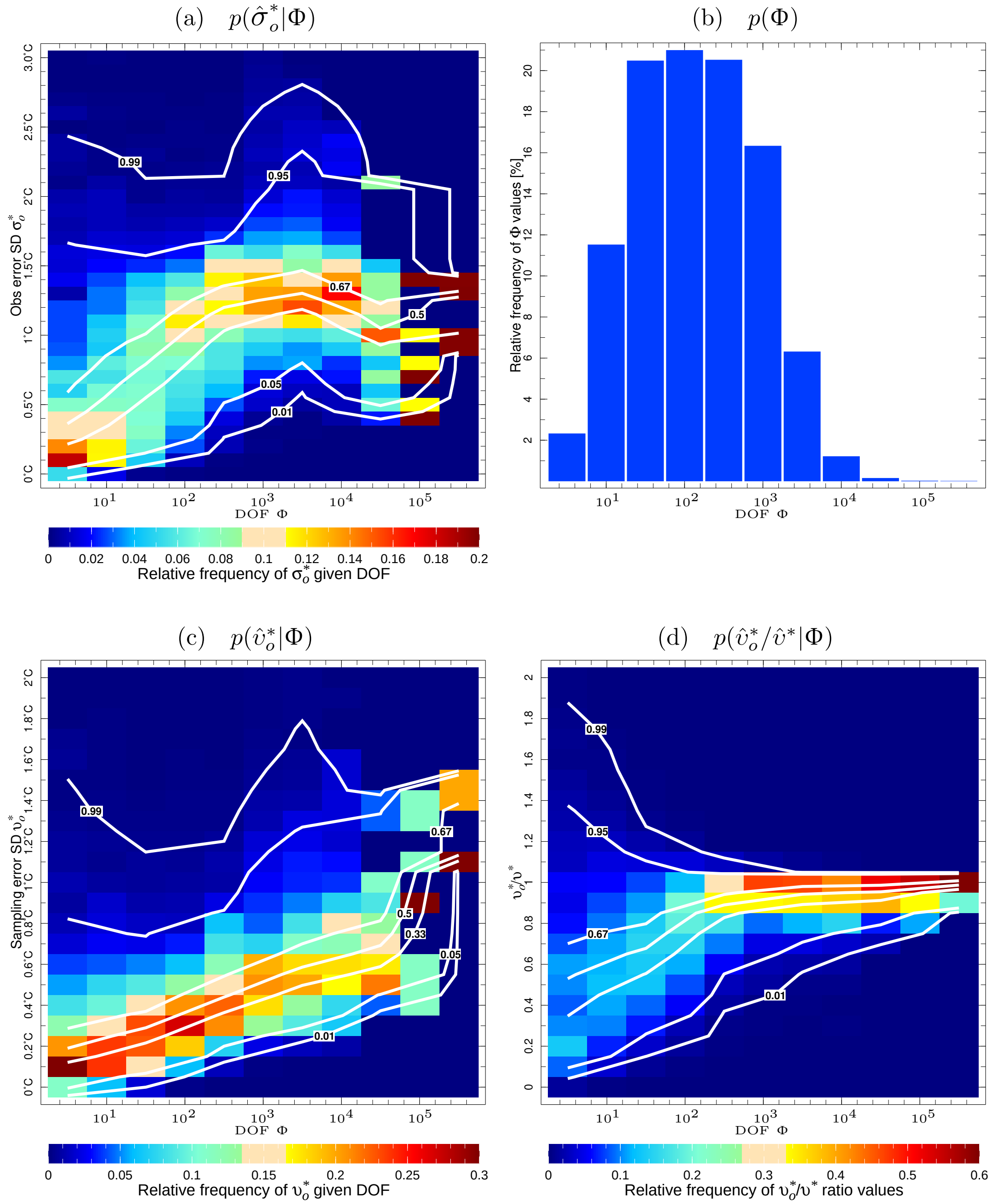

Figure 4. Empirical probability distributions, a.k.a. observed relative frequencies, of estimated SD of measurement $\hat{\sigma}^*_o$ and sampling $\hat{\upsilon}^*_o$ errors, for different ranges of the total number of degrees of freedom (DOF) Φ that are available for error variance estimation at each location in the 1992-2010 ICOADS ship SST data. Shown are (a) the distribution of measurement error SD estimates $\hat{\sigma}^*_o$, given the Φ range; (b) the empirical distribution of Φ values in the whole sample of the data used, for the same Φ bins, as used the calculation of two-dimensional histograms; (c) same as (a), but for the sampling error SD estimates $\hat{\upsilon}^*_o$; (d) same as (a) and (c), but for the ratio of sampling error SDs $\hat{\upsilon}^*_o/\hat{\upsilon}^*$, estimated from the ship data match-ups from the CCI analysis and from the full analysis data. White lines in panels (a), (c), and (d) are contours of cumulative empirical probability, conditional on Φ, for $\hat{\sigma}^*_o$, $\hat{\upsilon}^*_o$, and $\hat{\upsilon}^*_o/\hat{\upsilon}^*$ ratio, respectively, corresponding to 0.01, 0.05, 0.33, 0.5, 0.67. 095, and 0.99 levels, as labels indicate.

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.

Predominant ship tracks are easy to identify in Figure 3(d) as lines of low values. It is somewhat surprising though to see ship tracks characterized by *higher* values of SD estimates $\hat{\sigma}^*_{Bo}$ and $\hat{\sigma}^*_o$ in Figures 3c,f. To investigate this issue, empirical probability distributions, a.k.a. observed relative frequencies, of estimated SD of measurement $\hat{\sigma}^*_o$ and sampling $\hat{\upsilon}^*_o$ errors (whose sum of squares amounts to $\hat{\sigma}^{*\,2}_{Bo}$, for different ranges of the total number of degrees of freedom (DOF) Φ that are available for error variance estimation at each location in the 1992-2010 ICOADS ship SST data. The calculation is based on two-dimensional histograms whose bins are defined as grid boxes of uniform grids for SD σ and $\log_{10}\Phi$ values, with grid steps of 0.1^{o}C and 0.5, respecively. The histogram $h(\sigma, \log_{10}\Phi)$ is then normalized along the σ axis, resulting in the empirical probability function:

$$p(\sigma|\Phi) = h(\sigma, \log_{10}\Phi)\big/H(\Phi),$$

where

$$H(\Phi) = \sum_\sigma h(\sigma, \log_{10}\Phi)$$

is the distribution of absolute frequencies of Φ values. Figures 4a,c show empirical probability distributions, conditional on Φ being within the given range (bin) of values. The overall (marginal) empirical distribution of Φ values is computed as its relative frequencies

$$p(\Phi) = H(\Phi) \Big/ \sum_{\Phi} H(\Phi)$$

and is shown in Figure Figure 4d shows $\hat{v}_o^*$ the empirical probability disribution $p(\hat{v}_o^*/\hat{v}^*|\Phi)$ calculated in a simila way, using for the ratio values histogram bins of the with 0.1.

6 Conclusions

Rigorous formalism is proposed for modeling uncertainties in bin averages of SST observations irregularly sampled by ships, based on a comparison with an independent satellite SST analysis product of higher time-space resolution (compared to the bin's dimensions). The model allows for climatologically-dependent systematic biases in ship observations and assumes i.i.d. random measurement error and randomly distributed times and locations of ship observations within the bin. For the error in the analysis values, product-specified grid point uncertainties are used as given, with a supplementary assumption of high correlation between analysis errors within the bin. The main outcomes of the method are uncertainty estimates for ship SST averages, as well as their sampling and measurement uncertainty components.

The method was applied to the 1992–2010 comparison between ICOADS (Release 3.0) ship SST and ESA SST CCI Analysis (version 1.0). Differences between monthly $1°\times1°$ bin averages (for bins that contain more than one observation) of ICOADS ship SST and of the ESA SST CCI Analysis were presented here as the sum of their climatological bias component and remaining residuals (anomalies), whose magnitudes agreed well in the areas of sufficient data coverage with the estimates based on the proposed random error model. Location-dependent estimates of ship SST measurement and sampling uncertainties were obtained. Estimates of sampling uncertainty were similar in pattern, but larger than those previously published. Ship SST measurement error was consistent with previous estimates in large-scale spatial pattern and global RMS values (found to be 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 paper 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. Insightful remarks and suggestions of two anonymous reviewers resulted in a significant improvement of the manuscript.

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.

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.

Chan, D., Kent, E. C., Berry, D. I., & Huybers, P. (2019). Correcting datasets leads to more homogeneous early-twentieth-century sea surface warming. *Nature*, *571*(7765), 393–397.

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., ... others (2017). ICOADS Release 3.0: a major update to the historical marine climate record. *International Journal of Climatology*, *37*(5), 2211–2232.

Handcock, M. S., & Stein, M. L. (1993). A bayesian analysis of kriging. *Technometrics*, *35*(4), 403–410.

Hartmann, D. L., Tank, A. M. K., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y. A. R., ... others (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.

Holtzman, W. H. (1950). The unbiased estimate of the population variance and standard deviation. *The American Journal of Psychology*, *63*(4), 615–617.

Huang, B., Angel, W., Boyer, T., Cheng, L., Chepurin, G., Freeman, E., ... Zhang, H.-M. (2018). Evaluating SST analyses with independent ocean profile observations. *Journal of Climate*, *31*(13), 5015–5030.

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. doi: {10.1029/97JC01736}

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. doi: {10.1029/97JC01734}

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. doi: {10.1002/qj.900}

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. doi: {10.1175/JTech1844.1}

Kent, E. C., & Kennedy, J. J. (2021). Historical estimates of surface marine temperatures. *Annual Review of Marine Science*, *13*, 283–311.

Kent, E. C., Kennedy, J. J., Smith, T. M., Hirahara, S., Huang, B., Kaplan, A., ... Zhang, H.-M. (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. doi: {10.1175/BAMS-D-15-00251.1}

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., ... Donlon, C. (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. doi: {10.1002/gdj3.20}

Rayner, N. A., Kaplan, A., Kent, E. C., Reynolds, R. W., Brohan, P., Casey, K. S., ... others (2010). Evaluating climate variability and change from modern and historical SST observations. In J. Hall, D. E. Harrison, & D. Stammer (Eds.), *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society* (Vol. 2). ESA Publication WPP-306. (OceanObs' 09, Venice, Italy, September 21-25, 2009) doi: {10.5270/OceanObs09.cwp.71}

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, S. R., Freeman, E., Lubker, S. J., Woodruff, S. D., Worley, S. J., Angel, W. E., ... Kent, E. C. (2016). The international maritime meteorological archive (IMMA) format. *Available from the web page: http://icoads. noaa. gov/e-doc/imma/R3.0-imma1.pdf*.

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.

Stein, M. L. (1999). *Interpolation of spatial data: Some theory for kriging.* Springer Science & Business Media.

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.