CCI OSTIA as the Standard of Truth: Detailed Error Models for the *in Situ* SST Data From Ships and Other Platforms

Alexey Kaplan
LDEO of Columbia University
Palisades, NY 10964, USA
email: *alexeyk@ldeo.columbia.edu*

GHRSST-XX, Frascatti, Italy
7 June 2019

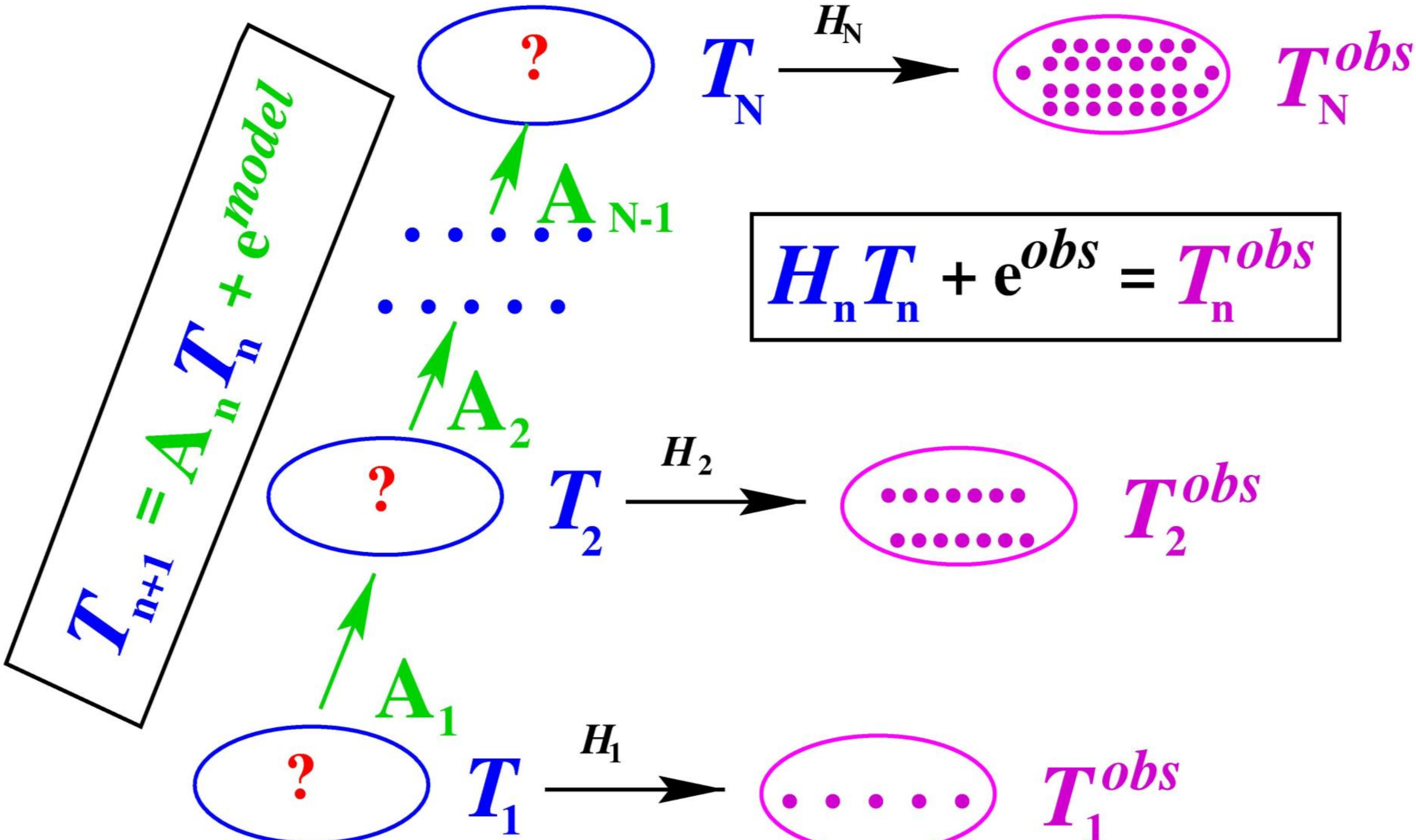
Outline

– Motivation: Why do we need to know the effective error of the in situ SST data

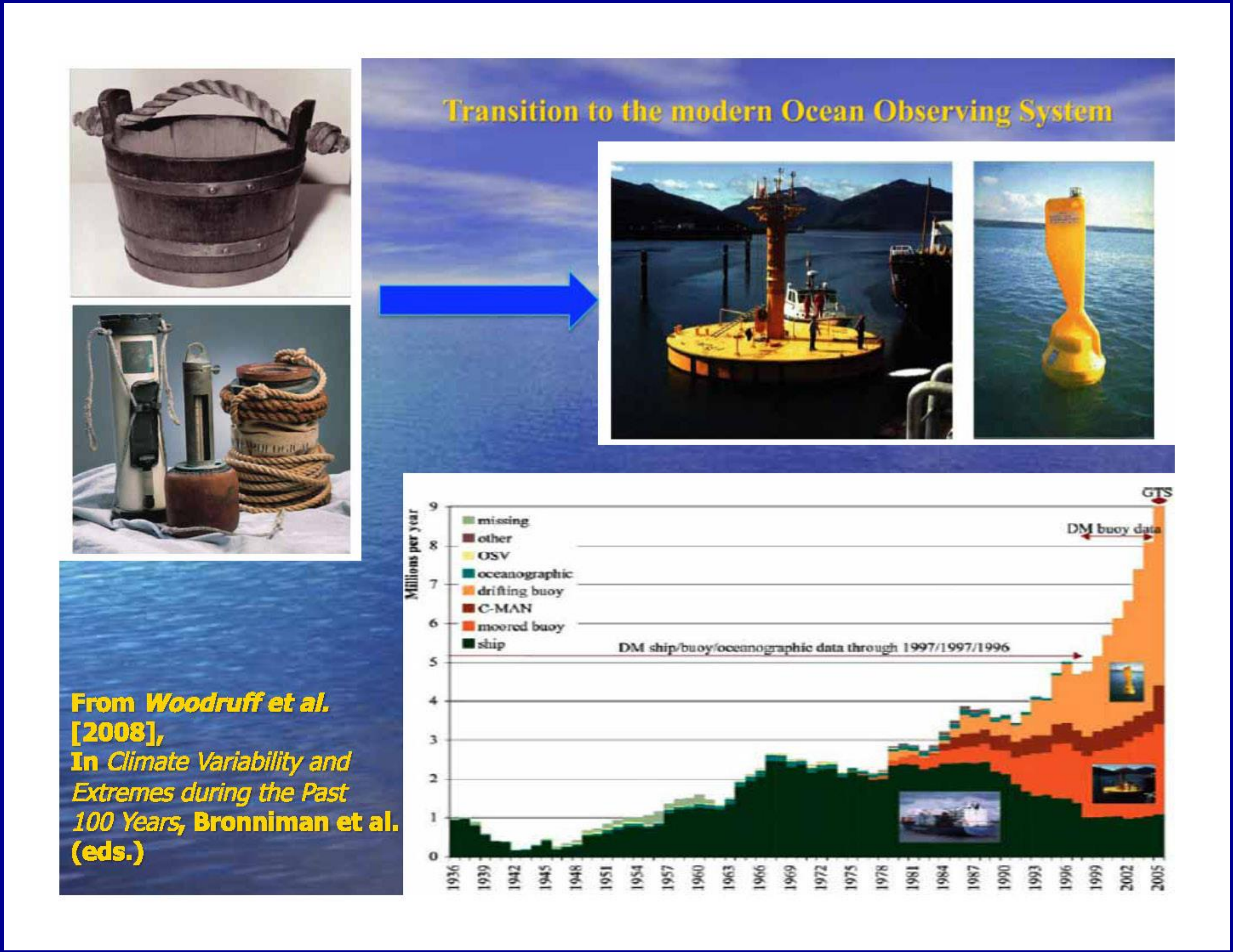
– Simple statistical estimates of variability and error computed for SST data from ships and drifters.

– Contrasting results found; ground-truthing using the CCI OSTIA high-resolution SST data set.

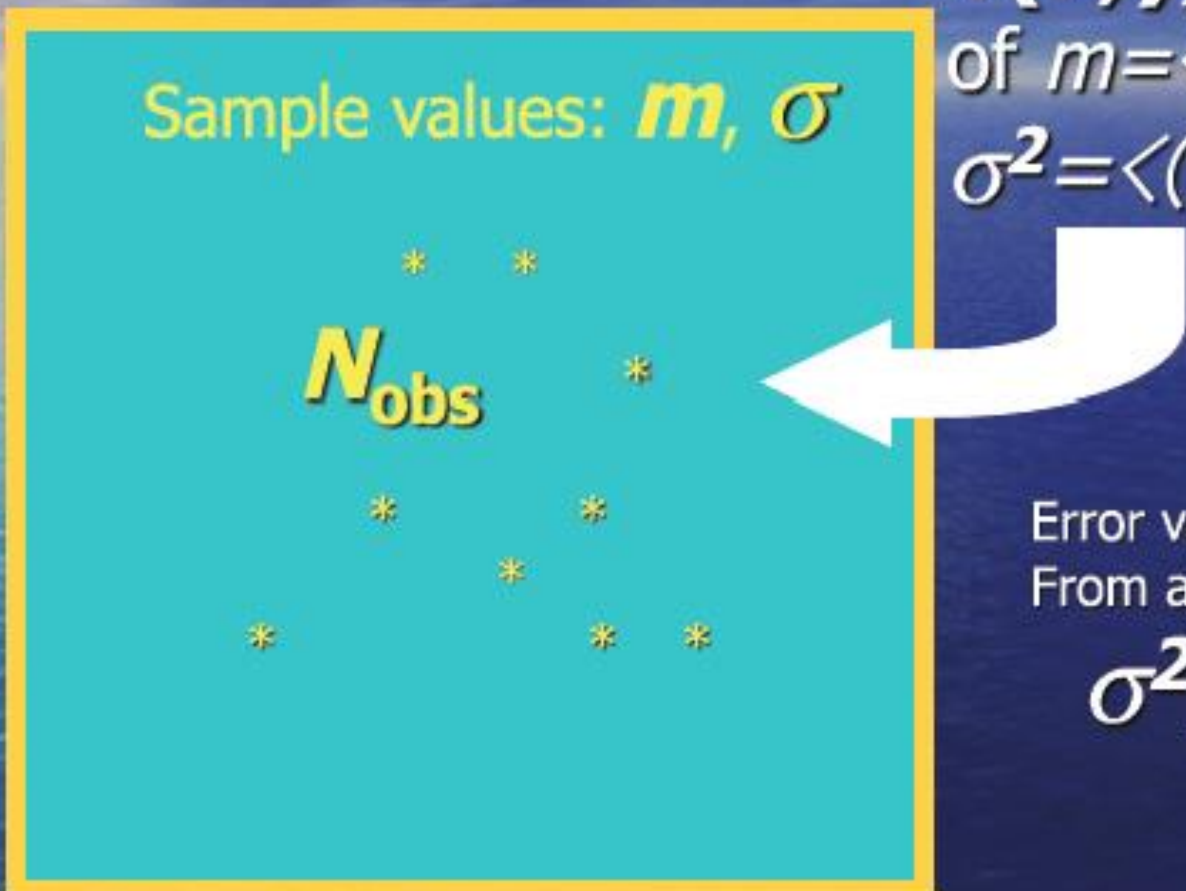
Generic problem of the analysis of time-evolving fields



International Comprehensive Ocean-Atmosphere Data Set (ICOADS, R.2.5, Woodruff et al., 2011)



$T(x,y,t)$: "true" values of $m = \langle T \rangle_{x,y,t}$ and $\sigma^2 = \langle (T-m)^2 \rangle_{x,y,t}$ for a bin

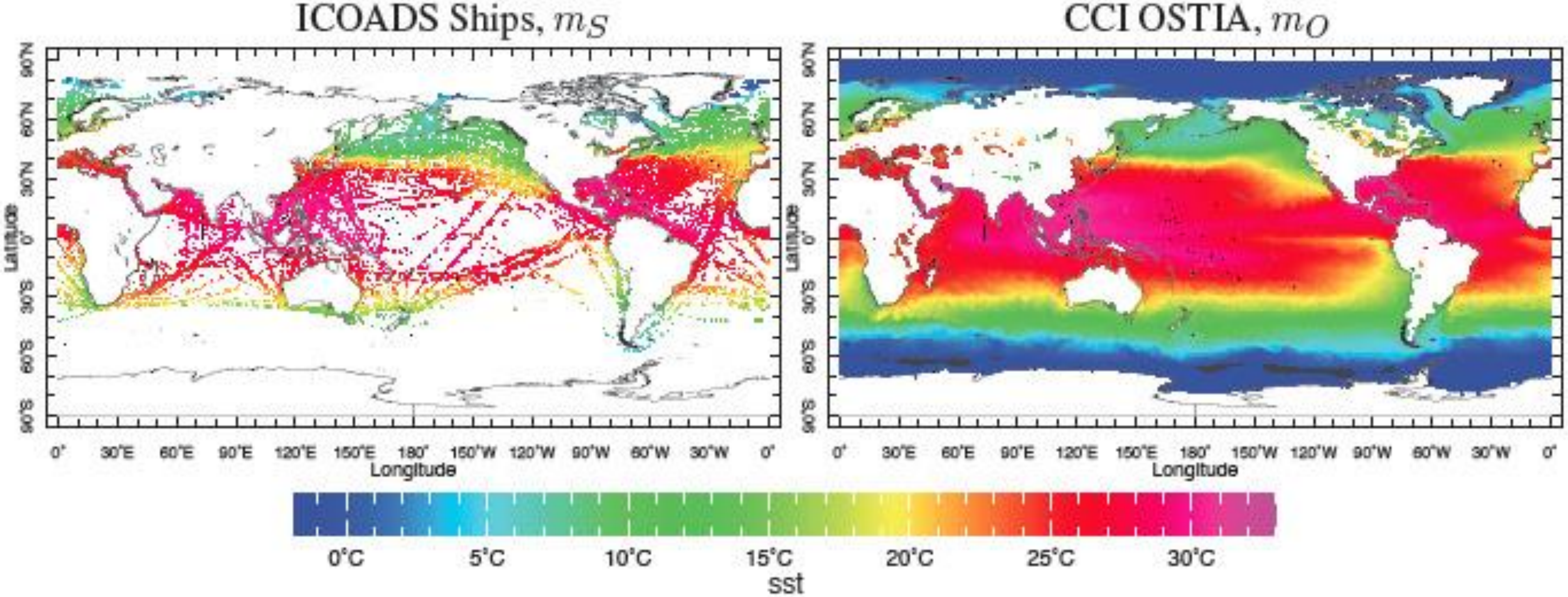


Error variance of the mean m
From averaging N_{obs} observations:

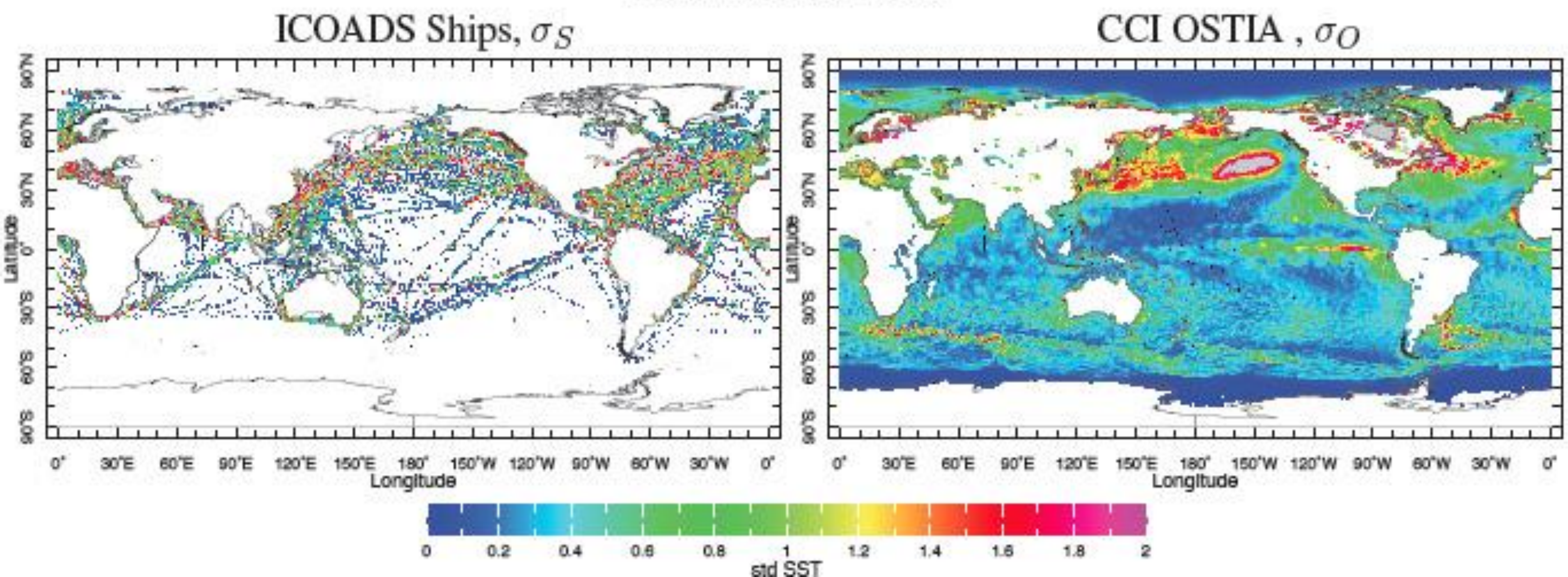
$$\sigma^2 / N_{obs} \qquad (1)$$

SST observations in $1^o \times 1^o$ bins for July 2010

Mean bin SST, m



Intra-bin SST std σ



1. **CCI OSTIA (v.1)** – global daily 6km (0.05 degree) *in situ independent* gapless SST data set (Merchant et al., 2014), with validated error estimates.

2. Separation of ICOADS SST by a platform type.

3. An opportunity to separate sampling and measurement error.

To estimate σ for the use in Equation (1), take the $1^o \times 1^o$ bin value of σ, and average it with square over all months in 1992-2010 (excluding the months with $N_{\mathrm{obs}} < 2$):

$$\sigma_S = \langle \sigma(t)^2 \rangle^{1/2} \tag{2}$$

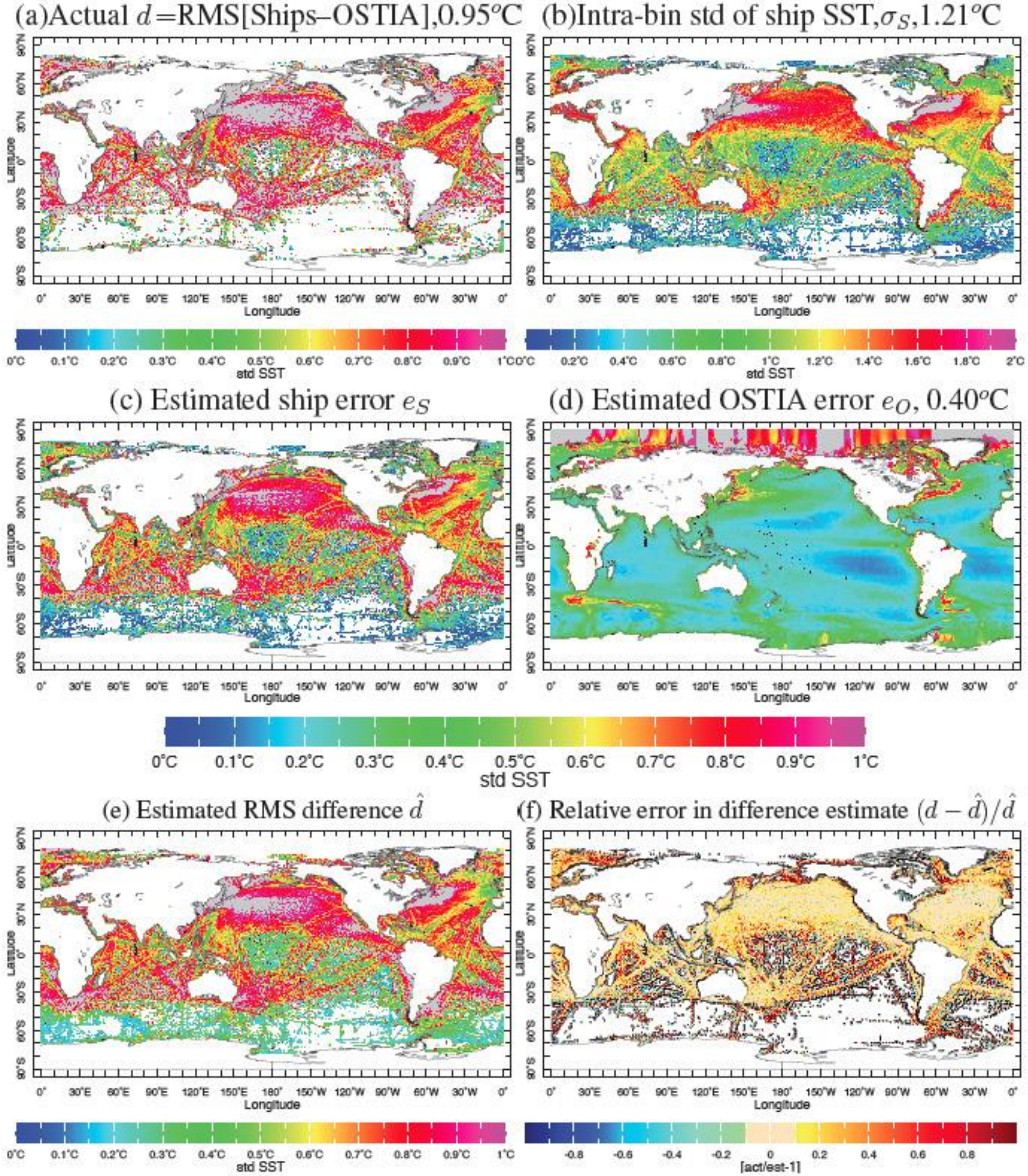
(angle brackets here denote time averaging). The resulting spatial patterns is shown in Figure 3*b* (note a different range of the colorbar). Now compute error of bin averages of ship data for each month t by (1), and average it with square over time:

$$e_S = \langle \sigma_S^2 / N_{\mathrm{obs}}(t) \rangle^{1/2}. \tag{3}$$

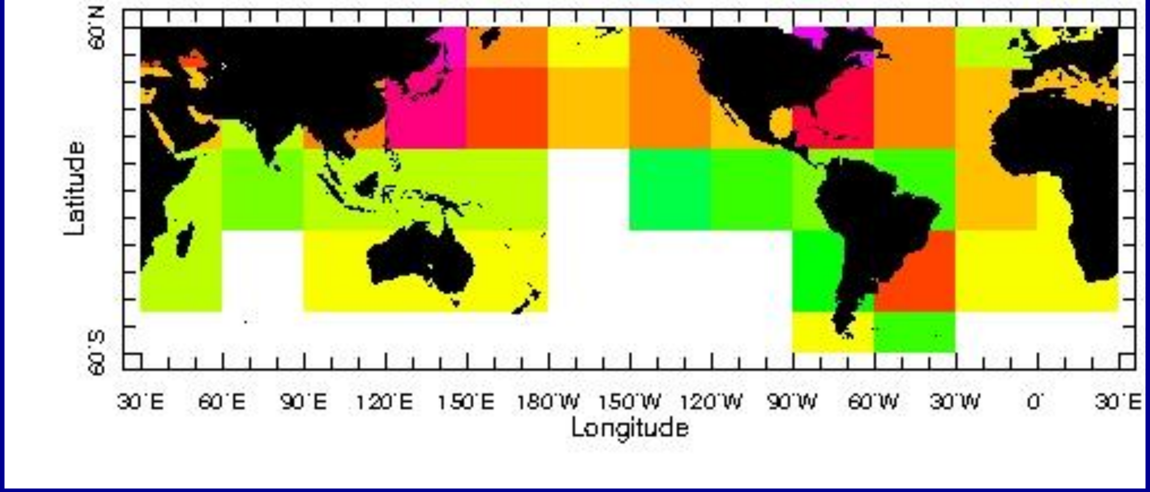
The results, in Figure 3*c*, show a very nice similarity to the RMS difference field d (Figure 5*a*), although e_S is, expectedly, slightly smaller than d, since the OSTIA SST, after all, is not completely free of error either. Error estimates for OSTIA e_O are shown in Figure 3*d*, and the estimate for d is computed as

$$\hat{d} = \left(e_S^2 + e_O^2 \right)^{1/2},$$

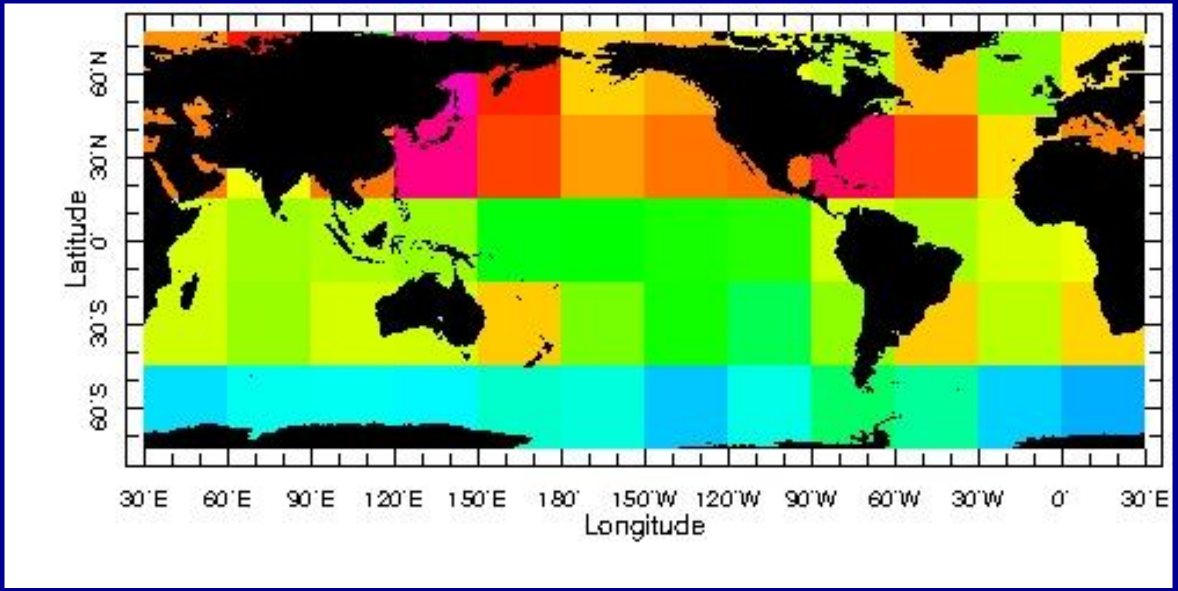
Error model for ICOADS ship SST, monthly $1^o \times 1^o$ bins, 1992-2010



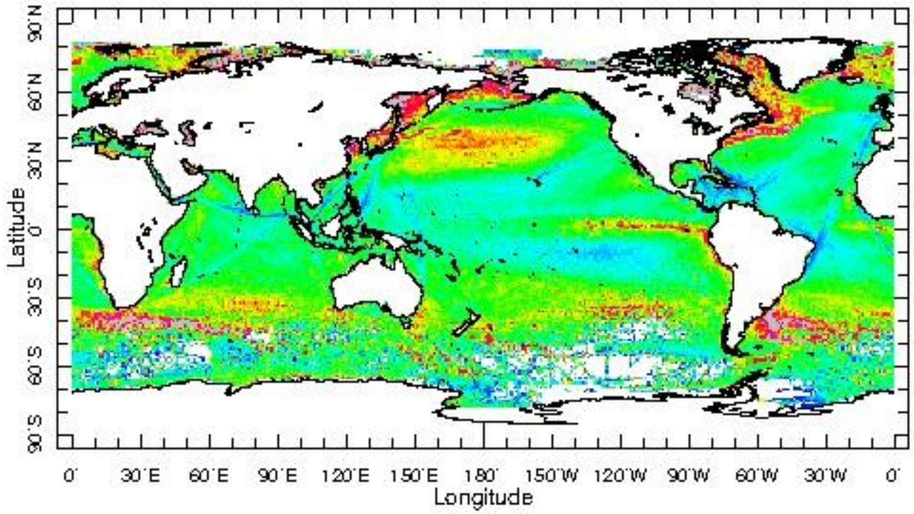
Kent and Challenor (2006)
Variogram method,
1970-1997



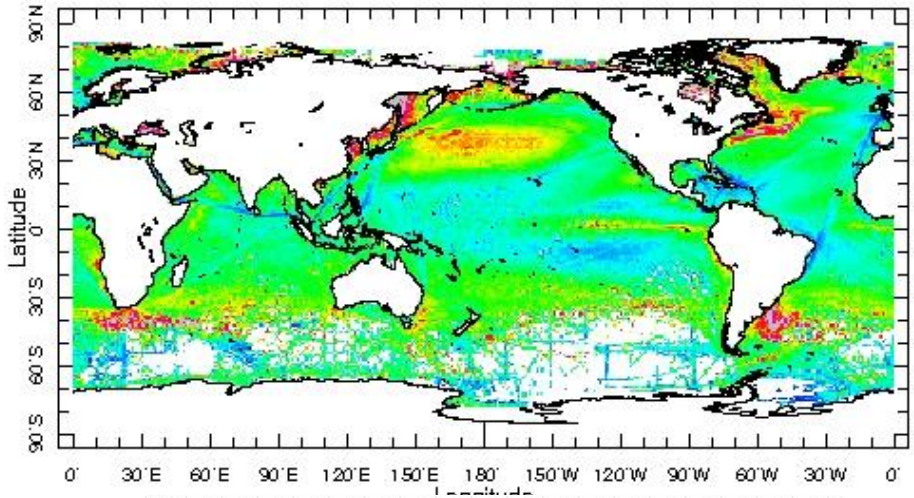
Here

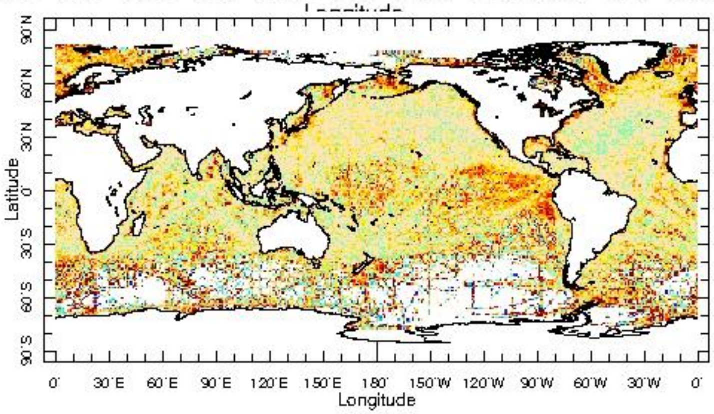


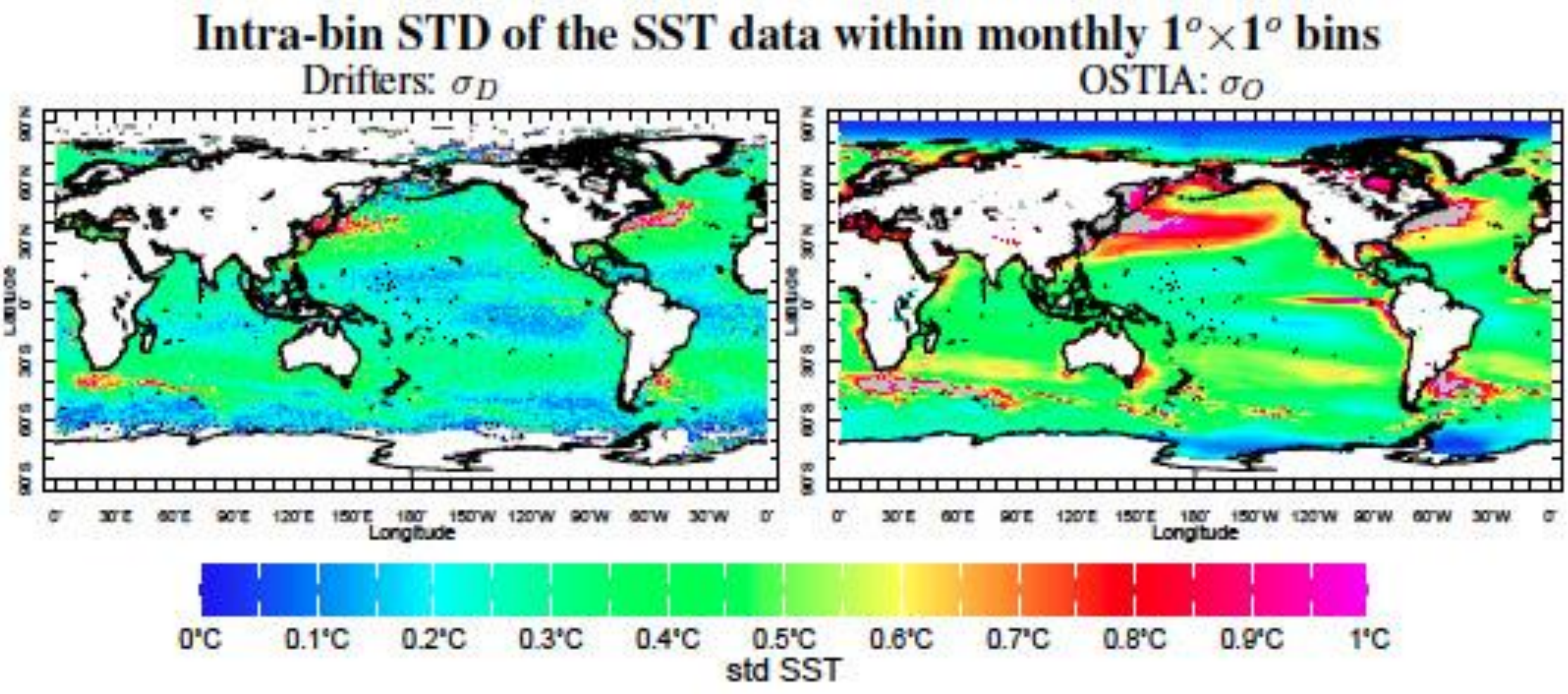
Sampling Error:
Actual



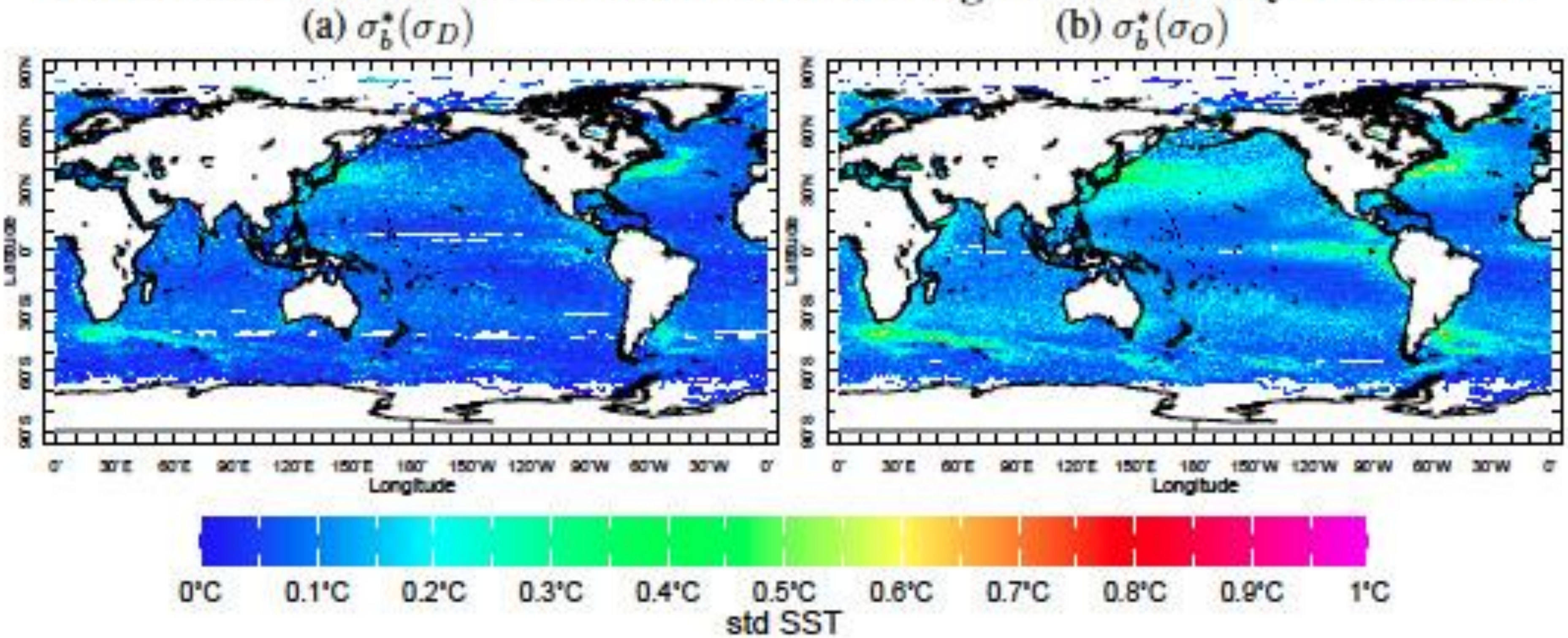
Estimated

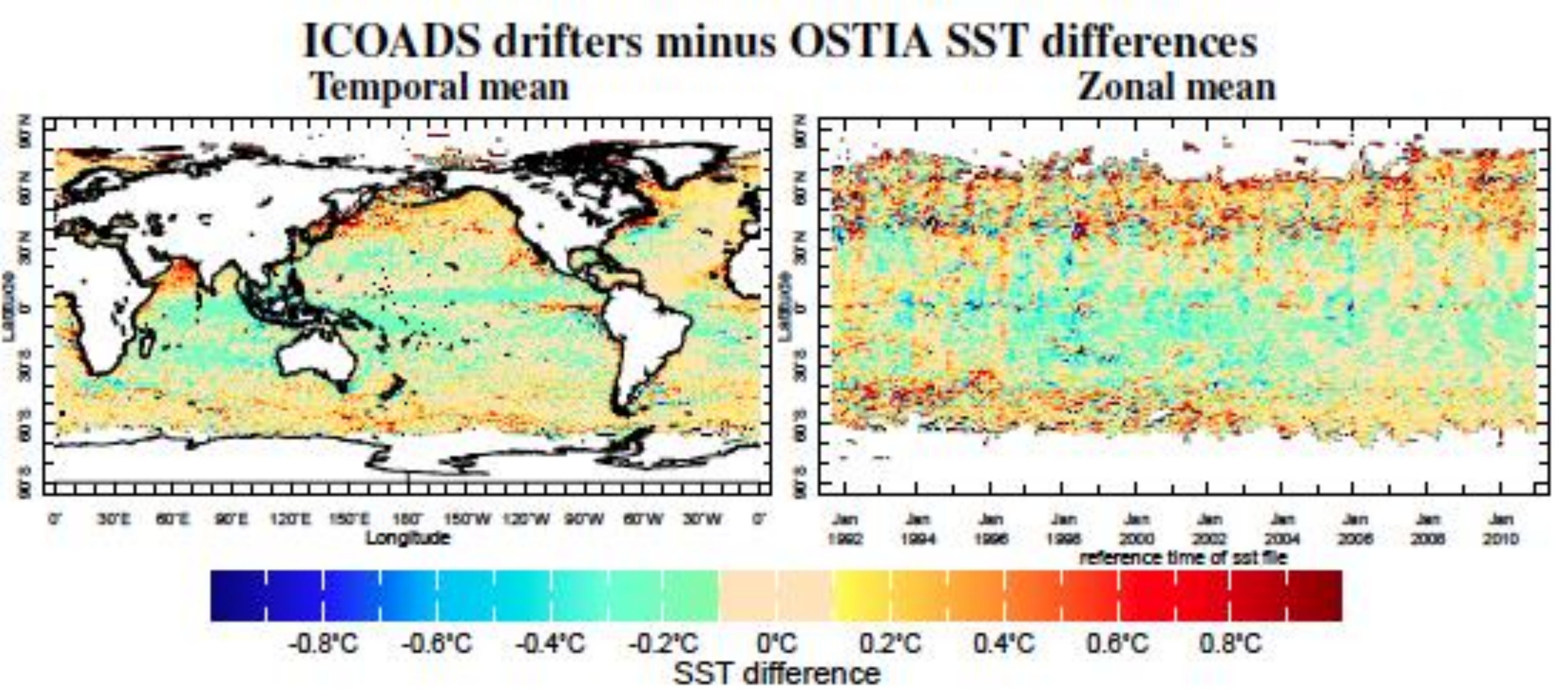





Intra-bin STD of the SST data within monthly 1°×1° bins
Drifters: σD
OSTIA: σO
0° 30°E 60°E 90°E 120°E 150°E 180° 150°W 120°W 90°W 60°W 30°W 0°
Longitude
0° 30°E 60°E 90°E 120°E 150°E 180° 150°W 120°W 90°W 60°W 30°W 0°
Longitude
0°C 0.1°C 0.2°C 0.3°C 0.4°C 0.5°C 0.6°C 0.7°C 0.8°C 0.9°C 1°C
std SST

Estimates for the error std of drifters' average over monthly $1^o \times 1^o$ bins

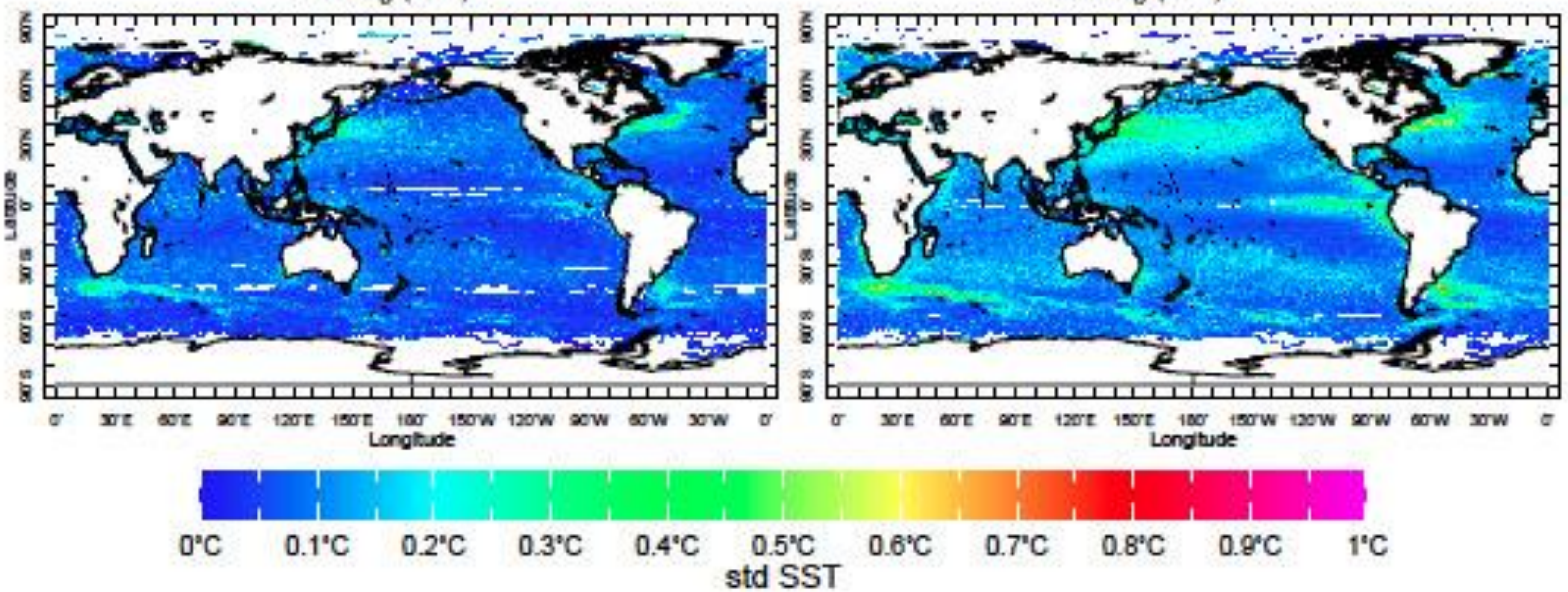




ICOADS drifters minus OSTIA SST differences
Temporal mean
Zonal mean
0° 30°E 60°E 90°E 120°E 150°E 180° 150°W 120°W 90°W 60°W 30°W 0°
Longitude
Jan 1992 Jan 1994 Jan 1996 Jan 1998 Jan 2000 Jan 2002 Jan 2004 Jan 2006 Jan 2008 Jan 2010
reference time of sst file
-0.8°C -0.6°C -0.4°C -0.2°C 0°C 0.2°C 0.4°C 0.6°C 0.8°C
SST difference

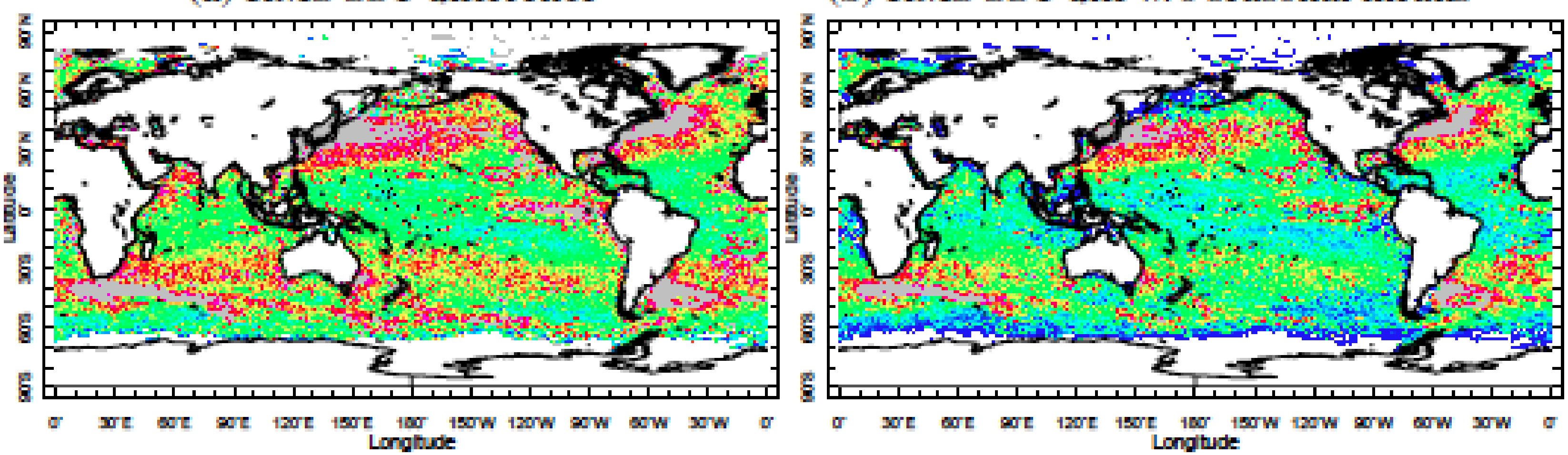
Estimates for the error std of drifters' average over monthly $1^o \times 1^o$ bins

(a) $\sigma_b^*(\sigma_D)$ (b) $\sigma_b^*(\sigma_O)$



RMS[drifters-OSTIA] for monthly $1^o \times 1^o$ bin averages

(a) RMS SST difference (b) RMS SST diff w/o seasonal means



P observational platforms, K observations from each

Example: P observational platforms, with K observations from each.

Measurements from these platforms are characterized by biases b^p, $p = 1, \cdots, P$ and by the random error on top of them, so that K observations from p-th platform will have errors as follows:

$$e_k^p = b^p + \varepsilon_k^p, \quad k = 1, \cdots, K,$$

where ε_k^p for all p and k are independent random numbers with zero mean and standard deviation σ. Since p-th platform's bias value b^p does not depend on k, it is the same for all errors e_k^p, $k = 1, \cdots, K$ of this platform's measurements. But biases of different ptatforms b^p, $p = 1, \cdots, P$ are independent random numbers from a probability distribution with zero mean (assumed this way for simplicity here) and standard deviation β. (With apologies, hereinafter β is **not** the first derivative of the Coriolis parameter over latitude.) Obviously, the mean error of our PK observations is

$$\bar{e} = \frac{1}{PK} \sum_{p=1}^{P} \sum_{k=1}^{K} (b^p + \varepsilon_k^p) = \frac{1}{P} \sum_{p=1}^{P} b^p + \frac{1}{PK} \sum_{p=1}^{P} \sum_{k=1}^{K} \varepsilon_k^p,$$

and the variance of the mean error (the expectation of its squared value) is

$$\mathbb{E}\bar{e}^2 = \frac{\beta^2}{P} + \frac{\sigma^2}{PK}. \tag{1}$$

P observational platforms, K observations from each

Recall that PK is the total number of observations averaged here, and $\mathbb{E}\bar{e}^2$ represents the error variance in this average. Yet, according to Equation (1), when PK becomes so large that

$$\frac{\sigma^2}{PK} \ll \frac{\beta^2}{P}, \tag{2}$$

error variance $\mathbb{E}\bar{e}^2$ only decreases to the value that is essentially independent of the total number of observations:

$$\mathbb{E}\bar{e}^2 \geq \frac{\beta^2}{P}.$$

Instead, it requires an increase in the number of platforms P in order to achieve further reduction.

Platform biases model: Kent and Berry (2008), Kennedy et al. (2011, 2012).

However, platform biases of ~0.3 C for drifting buoys would be inconsistent with, *e.g.*, studies by Reverdin et al. (2010, 2013).

P observational platforms, K observations from each

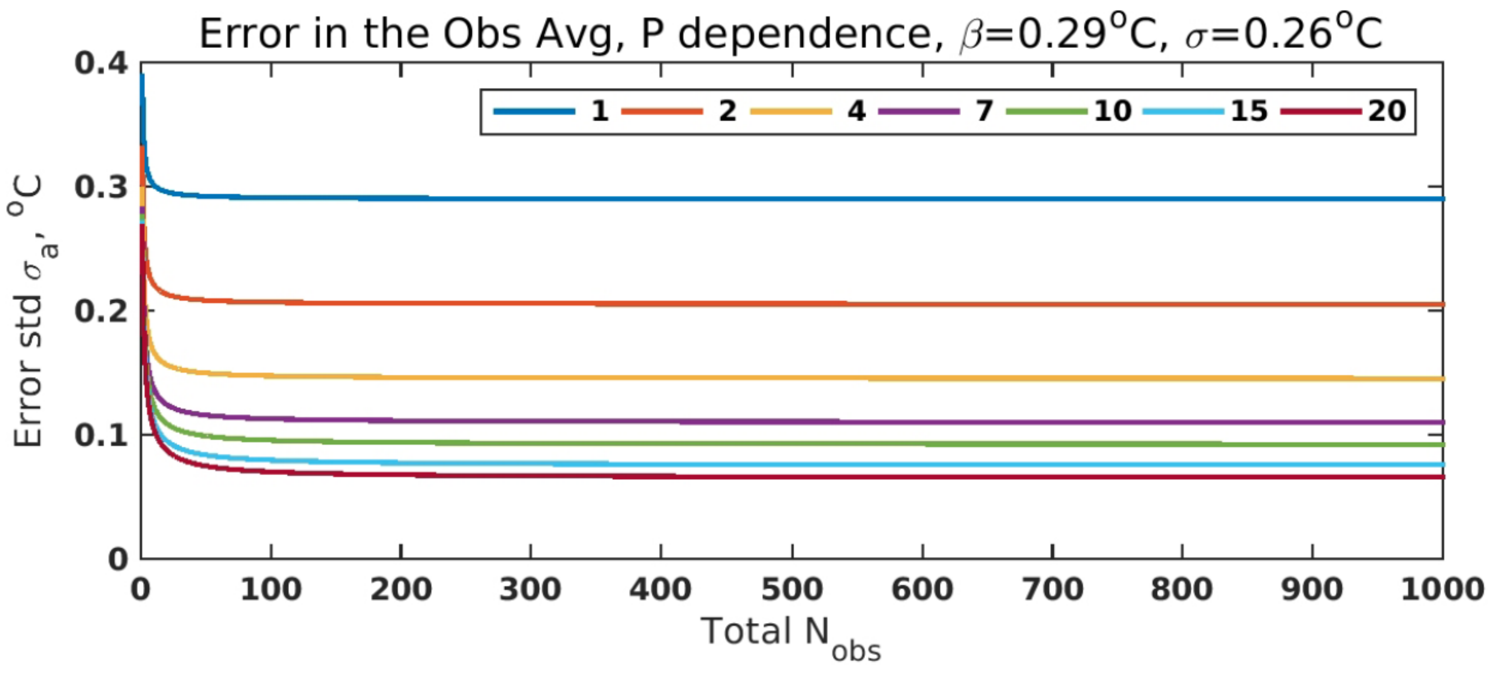


Figure 1: Standard error $\sigma_a = (\mathbb{E}\bar{e}^2)^{1/2}$ in the average of $N_{\rm obs}$ observations that came from P platforms. Calculations are done by formula (1) using values β=0.29^{o}C and σ=0.26^{o}C, as estimated by Kennedy et al. (2012), for ranges P=1–20 (shown by different colors, as indicated in the legend) and $N_{\rm obs} = PK$ =1–1000.

Conclusions & the Outlook

1. Simple statistical formula for the error in the mean works very well for the ICOADS ship SST data in 1x1 degree monthly bins.

2. The same estimate for drifting buoys underestimate the error by the order of magnitude.

3. Comparison with the CCI OSTIA satellite-based SST data set suggests that drifters undersample the SST variability.
4. Lagrangian nature of the platform might be causing these effects, but more research is needed.