Understanding Differences in Sea Surface Temperature Intercomparisons

Boyin Huang[1*], Xungang Yin[1], James A. Carton[2], Ligang Chen[2], Garrett Graham[3], Chunying Liu[4], Thomas Smith[5], Huai-Min Zhang[1]

[1] NOAA National Centers for Environmental Information, Asheville, North Carolina

[2] Department Atmospheric and Oceanic Sciences, University of Maryland, College Park, Maryland.

[3] North Carolina Institute for Climate Studies, North Carolina State University, Asheville, North Carolina

[4] Riverside Technology, Inc., Asheville, North Carolina

[5] NOAA Center for Satellite Applications and Research

[*] Corresponding author email address: boyin.huang@noaa.gov

(Submitted to **JTECH**)

Early Online Release: This preliminary version has been accepted for publication in *Journal of Atmospheric and Oceanic Technology*, may be fully cited, and has been assigned DOI 10.1175/JTECH-D-22-0081.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

Abstract (limit to 260 words)

Our study shows that the intercomparison among sea surface temperature (SST) products is influenced by the choice of SST reference, and the interpolation of SST products. The influence of reference SST depends on whether the reference SST are averaged to a grid or in pointwise in situ locations, including buoy or Argo observations, and filtered by first-guess or climatology quality control (QC) algorithms. The influence of the interpolation depends on whether SST products are in their original grids or pre-processed into common coarse grids.

The impacts of these factors are demonstrated in our assessments of eight widely used SST products (DOISST, MUR25, MGDSST, GAMSSA, OSTIA, GPB, CCI, CMC) relative to buoy observations: (a) when the reference SSTs are averaged onto 0.25º×0.25º grid boxes, the magnitude of biases is lower in DOISST and MGDSST (<0.03ºC), and magnitude of root-mean-square-differences (RMSDs) is lower in DOISST (0.38ºC) and OSTIA (0.43ºC); (b) when the same reference SSTs are evaluated at pointwise in situ locations, the standard deviations (SDs) are smaller in DOISST (0.38ºC) and OSTIA (0.39ºC) on 0.25º×0.25º grids; but the SDs become smaller in OSTIA (0.34ºC) and CMC (0.37ºC) on products' original grids, showing the advantage of those high-resolution analyses for resolving finer scale SSTs; (c) when a loose QC algorithm is applied to the reference buoy observations, SDs increase; and vice versa; however, the relative performance of products remains the same; and (d) when the drifting-buoy or Argo observations are used as the reference, the magnitude of RMSDs and SDs become smaller, potentially due to changes in observing intervals. These results suggest that high-resolution SST analyses may take advantage in intercomparisons.

Significance Statement

Intercomparisons of gridded SST products be affected by how the products are compared with in situ observations: Whether the products are in coarse (0.25º) or original (0.05º−0.10º) grids, whether the in situ SSTs are in their reported locations or gridded and how they are quality-controlled, and whether the biases of satellite SSTs are corrected by localized matchups or large scale patterns. By taking all these factors into account, our analyses indicate that the NOAA

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

DOISST is among the best SST products for the long period (1981−present) and relatively coarse (0.25º) resolution that it was designed for.

1. Introduction

Sea surface temperature (SST) as an important climate indicator has numerous applications at different spatial and temporal scales. For example, SSTs are used in studying short-term extreme weather events (Feudale and Shukla 2011; Hartmann 2015), extreme marine heatwave events (Hobday et al. 2016; Huang et al. 2021a; D'Agata 2022), impacts of El Niño and Southern Oscillation (ENSO) on fisheries and agriculture (Yates et al. 2016; Singles and Bezuidenhout 1999), coastal watch (Miller and DeCampo, 1994; Lima and Wethey 2012; Cole 2000), climatic impacts at decadal and multidecadal timescales (Mohino et al. 2011; Sun et al. 2016; Vibhute et al. 2020), and long-term global warming (Karl et al. 2015; Zhang et al. 2016). Many of these applications require an accurate gridded product for weather and ocean forecasting (O'Carroll et al. 2019, and references therein), climate projections (He and Soden 2016), and coastal watch (Shimada et al. 2015). To meet these applications' requirements, SST products with various spatial (0.01º−5º) and temporal (6-hour, daily, and monthly) resolutions have been developed based on in situ observations from ships, buoys and Argo floats, and satellites (Huang et al. 2017, and references therein).

The quality of gridded SST products is usually assessed by comparing to a reference SST. An ensemble of available SST products may be used as a reference (Dash et al. 2012; Chin et al. 2017; Yang et al. 2021; Huang et al. 2021b), e.g., the Group for High Resolution SST (GHRSST) Multiproduct Ensemble (GMPE; Martin et al. 2012; Fiedler et al. 2019). The in situ SSTs from drifting buoys, moored buoys, and Argo floats were frequently used as an ideal reference due to their accuracy and high spatial and temporal coverages (Huang et al. 2021b; also, in this study). The SSTs derived from Conductivity-temperature-depth (CTD) can be used as a reference, but their accuracy may be impacted by depth errors and near-surface contaminations (Huang et al. 2018; Moteki 2022). The SST measurements from modern saildrones and thermosalinographs are accurate but spatial and temporal coverages are limited, which can be used as references in regional assessment (Vazquez-Cuervo et al. 2022).

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

The reference SST can be processed either to the GMPE grids (0.25º×0.25º; a common grid established for the sake of intercomparison) (Huang et al. 2021a,b) or in its in situ locations (Dash et al. 2012; Fiedler et al. 2019), and may be filtered by different quality-control (QC) algorithms (Reynolds et al. 2007; Dash et al. 2010). Likewise, the gridded SST products can be processed either to the GMPE grids (Martin et al. 2012; Fiedler et al. 2019; Yang et al. 2021; Huang et al. 2021b, 2017) or in their original grids (Dash et al. 2012; Fiedler et al. 2019).

A recent assessment (Huang et al. 2021b) indicated that the NOAA Daily Optimum Interpolation (OI) SST (DOISST) v2.1 has a good performance, while the NOAA SST quality monitor (SQUAM; Fig. S1) showed that UK Met Office operational SST and sea ice analysis (OSTIA; Good et al. 2020) has a better performance. There are questions about whether the intercomparisons are sensitive to these details of the comparisons and how we understand the differences, which are subjects of this paper.

In this study, we address the reasons for the differences among the SST intercomparisons between nine widely used daily gridded SST products (section 2.1). The intercomparison methods are described in section 2.2. In section 3.1, we show why an ensemble SST reference is not preferable. In section 3.2, we demonstrate how intercomparisons are influenced by using in situ SSTs as a reference. In section 3.3, we show how high-resolution products take advantages in intercomparisons. In section 3.2, we demonstrate how QC procedures and moored buoy observations may affect the reference SST and therefore the assessment of products. The results are summarized and discussed in section 4.

2. Datasets and methods

2.1 Nine SST products

(a) DOISST

The NOAA DOISST v2.1 (Table 1) is a daily 0.25º×0.25º product starting September 1981 (Reynolds et al., 2007; Huang et al., 2021a). DOISST includes SST observations from ships, drifting and moored buoys, Argo floats, and Advanced Very High Resolution Radiometer (AVHRR) retrieved from NOAA-series and MetOp-A/B satellites by U. S. Navy (Huang et al. 2021a) before November 2021. After November 2021, DOISST switched to NOAA Advanced Clear Sky Processor for Ocean (ACSPO; Jonasson et al. 2020) satellite SSTs retrieved from

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

AVHRR and the Visible Infrared Imager Radiometer Suite (VIIRS). These observed SSTs are filtered by the QC (see more details in section 2.3) of the first-guess (FG), which is the DOISST analysis in the previous day. The biases of satellite SSTs were quantified by the difference between large scale patterns of satellite and in situ SSTs within 3000 km in latitude, 5000 km in longitude, and 15-day data window, which were determined by the Empirical Orthogonal Teleconnection functions (EOTs; Reynolds et al. 2007).

(b) MUR25

The NASA Multi-scale Ultra-high Resolution (MUR25) v4.1 analysis is a daily 0.25º×0.25º SST product starting from 2002 (Chin et al. 2017). MUR v4.1 includes in situ SSTs from the NOAA iQuam project (Xu and Ignatov 2010), which includes SSTs from ships, drifting and moored buoys, and Argo floats. The in situ SSTs were blended with nighttime SSTs derived from AVHRR, Advanced Microwave Scanning Radiometer-EOS (AMSR-EOS), AMSR2, the Moderate Resolution Imaging Spectroradiometers (MODIS), the US Navy microwave WindSat radiometer. Biases in satellite SSTs are adjusted according to in situ SSTs.

(c) MGDSST

The Japan Meteorological Agency (JMA) Merged satellite and in situ data Global Daily Sea Surface Temperature (MGDSST) is a daily 0.25º×0.25º product starting from 1982 to 2020 (Kurihara et al. 2006). The MGDSST includes in situ SST from buoys and ships, satellite SSTs retrieved from infrared sensors (NOAA/AVHRR, MetOp/AVHRR), microwave sensors (Coriolis/WINDSAT, GCOM-W1/AMSR-2, AQUA/AMSR-E), and ACSPO version 2.60 after December 2018 (Sakurai et al. 2019).

(d) GAMSSA

The Bureau of Meteorology (BoM) Global Australian Multi-Sensor SST Analysis (GAMSSA) v1 is a daily 0.25º×0.25º product starting from 2008 (Zhong and Beggs 2008; Beggs et al. 2011, 2020). GAMSSA uses SSTs derived from AVHRR, the Advanced Along Track Scanning Radiometer (AATSR), the AMSR2, and in situ SSTs from ships, drifting and moored buoys. Biases in AVHRR and AMSR2 SSTs are adjusted using drifting buoy SSTs.

(e) OSTIA

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

The UK Met Office OSTIA v2 is a daily 0.05º×0.05º SST product starting from 2006 (Stark et al. 2007; Donlon et al. 2012; Good et al. 2020). OSTIA includes in situ SSTs from ships, drifting and moored buoys, satellite SSTs derived from AVHRR, AMSR2, VIIRS, the Sea and Land Surface Temperature Radiometer (SLSTR), the Spinning Enhanced Visible and Infrared Imager (SEVIRI). SSTs from drifting and moored buoys and VIIRS nighttime SSTs are used to adjust the biases in other satellite-derived SSTs. Biases in satellite SSTs in a 7º grid are estimated with pairs of in situ SSTs within 25 km.

(f) GPB

The NOAA Geo-Polar Blended (GPB) v1 is a daily 0.05º×0.05º SST product starting from 2014 (Maturi et al. 2017). GPB includes in situ SSTs from ships, drifting and moored buoys, and nighttime SSTs derived from AVHRR, VIIRS, the Geostationary Operational Environmental Satellite (GOES) imager, the Japanese Advanced Meteorological Imager (JAMI) (Xu and Ignatov 2010). Biases in satellite SSTs are corrected by in situ SSTs in a 7º grid based on pairs of in situ and satellite SSTs within 25 km. Additionally, the difference between satellite and GPB analysis of the previous day, and an independent NCEP SST product (Thiébaux et al. 2003).

(g) CCI

The European Space Agency (ESA) Climate Change Initiative (CCI) SST version 2.0/2.1 is a daily 0.05º×0.05º SST product from 1981 (Merchant et al., 2014; 2019). The CCI includes both AVHRR and Along-Track Scanning Radiometer (ATSR) series. The biases in satellite SSTs were adjusted by recalibrating radiances using a reference channel.

(h) CMC

The Canadian Meteorological Centre SST (CMC) v3 is a daily 0.1º×0.1º SST starting from 2016 (Brasnett 1997, 2008; Brasnett and Colan 2016). CMC v3 uses in situ SSTs from ships and drifting buoys, and AVHRR SSTs from satellites NOAA-18 and 19, METOP-A and B, AMSR2. Biases in satellite SSTs in 2.5º grid are estimated with pairs of in situ SSTs within 25 km.

(i) GMPE

The Group High Resolution SST (GHRSST) Multi-Product Ensemble (GMPE) is a daily 0.25º×0.25º product starting from 2009 (Martin et al. 2012; Dash et al. 2012; Fiedler et al. 2019).

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

The GMPE selects the median SST from the GHRSST products. GMPE v2 (2016) and v3 (2017–2020) are used in this study.

Dataset	Version	Resolution	Input	Method	Access
DOISST	v2.0 (1981–2019) v2.1 (2016–)	0.25°	AVHRR/ACSPO + Ship + Buoy + Argo	OI	https://www.ncei.noaa.gov/data/sea-surface-temperature-optimum-interpolation/v2.1/access/avhrr/
MUR25	MUR v4.2 (2002–)	0.25°	AVHRR + Microwave + Ship + Buoy + Argo	Multi-Resolution Variational Analysis (MRVA)	https://podaac-opendap.jpl.nasa.gov/opendap/allData/ghrsst/data/GDS2/L4/GLOB/JPL/MUR25/v4.2
MGDSST	(1982-)	0.25º	AVHRR + Microwave + Ship + Buoy	OI	http://www.data.jma.go.jp/gmd/goos/data/pub/JMA-product/mgd_sst_glb_D
GAMSSA	v1 (2008–)	0.25°	AVHRR + AATSR + AMSRE + Ship + Buoy + ACSPO	OI	https://archive.podaac.earthdata.nasa.gov/podaac-ops-cumulus-protected/GAMSSA_28km-ABOM-L4-GLOB-v01
OSTIA	v2 (2006–)	0.05°	AVHRR + AMSR2 + VIIRS + SEVIRI + SLSTR + Ship + Buoy	OI	https://archive.podaac.earthdata.nasa.gov/podaac-ops-cumulus-protected/OSTIA-UKMO-L4-GLOB-v2.0
GPB	v1 (2014–)	0.05°	Imager + AVHRR + VIIRS + Ship + Buoy	OI	https://archive.podaac.earthdata.nasa.gov/podaac-ops-cumulus-protected/Geo_Polar_Blended_Night-OSPO-L4-GLOB-v1.0
CCI	v2.0 (1981–2019)	0.05°	AVHRR + ATSR + ATSR2 + Adv. ATSR	Variational Assimilation (VA)	https://dap.ceda.ac.uk/neodc/c3s_sst/data/ICDR_v2/Analysis/L4/v2.0;https://dap.ceda.ac.uk/neodc/esacci/sst/data/CDR_v2/Analysis/L4/v2.1
CMC	v3 (2016–)	0.1°	AVHRR + AMSR2 Ship + Buoy	OI	https://archive.podaac.earthdata.nasa.gov/podaac-ops-cumulus-

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

					protected/CMC0.1deg-CMC-L4-GLOB-v3.0
GMPE	v1 (2009-12) v2 (2012-17) v3 (2017–)	0.25°	GHRSST ensemble median SST	N/A	ftp://nrt.cmems-du.eu/Core/SST_GLO_SST_L4_NRT_OBSERVATIONS_010_005/METOFFICE-GLO-SST-L4-NRT-OBS-GMPE-V3

Table 1. Daily SST datasets (from January 2016 to January 2022) used in this study (all data were downloaded on February 15, 2022).

2.2 Reference SSTs from in situ Buoy and Argo observations

In this study, the SSTs from drifting buoys and moored buoys (simply referred as Buoy, hereafter) and Argo observations are used as a reference to assess the nine gridded SST products in section 3. The Buoy SSTs are measured at depth of 0.2–1.0 m (Castro et al. 2012). The temperature measurements of Argo floats and moored buoys above 5 m depth are usually averaged and taken as SST observations (Roemmich et al. 2015; Huang et al. 2017, 2021a). Buoy SSTs are retrieved from the International Comprehensive Ocean-Atmosphere Data Set (ICOADS) Release 3.0.2 (Liu et al. 2022), and Argo SSTs are derived from the Global Data Assembly Centre (GDAC; Argo 2000). It should be noted that observation densities are very distinct in Buoy and Argo due to observing frequency of Buoy (6 minutes to 1 hour in moored buoy and 1 hour in drifting buoy) and Argo (10 days) (Fig. S2), although the collocated difference between Buoy and Argo SSTs are small (0.03º±0.03ºC; Huang et al. 2017). Therefore, their spatial and time coverages are different, which may impact the intercomparison results discussed in sections 3.2 and 3.3.

Buoy SSTs are ingested into eight out of the nine gridded products in section 2.1 except for CCI, and Argo SSTs are used in DOISST and MUR25 (Table 1). Therefore, the impact of independence of Buoy and Argo observations is discussed in section 4. For comparison purposes, the drifting and tropical-moored buoy SSTs from the iQuam project (Xu and Ignatov 2010) are also used as a reference for the purposes of assessments in section 3.6.

2.3 Intercomparison methods

To assess the impacts of spatial resolution on the performance of gridded SST products in reference to Buoy and Argo SSTs, the nine SST products are compared in two resolutions from

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

January 2016 to January 2022: (a) the coarsest resolution (0.25º) for all nine products, which is namely the GMPE convention that degrades the four high resolution SST products (OSTIA, GPB, CCI, CMC) to 0.25º resolution, and (b) the original (Orig) products' resolution (0.05º–0.25º). The reason for degrading these SST products in (a) is to eliminate their potential advantage of high resolution so that the intercomparisons with those in low resolution become fair.

Likewise, the Buoy and Argo SSTs that have passed established QC procedures are used as a reference in two resolutions: (a) the coarsest resolution (0.25º), which is derived using box-average, and (b) the pointwise in situ locations. The reason for using 0.25º resolution is that in situ observations are first processed into superobservations within analysis grid-boxes, and that the reference Buoy SSTs will not be overwhelmed by the moored-buoys that provide high-frequency observations. The reason for using pointwise locations is that observations were actually taken at these in situ locations. When the pointwise locations of observations are used as a reference SST in this study, the gridded SST products are interpolated to the pointwise locations using a bi-linear interpolation method, which linearly in both longitude and latitude interpolates the gridded SSTs surrounding the pointwise observation within the gridbox. Alternatively, we tested an e-fold distance-weighting method, in which the gridded SSTs within 0.25º from the pointwise location are averaged according to their distance to the pointwise location. Our tests indicate that the results using the e-fold method are very close to the bi-linear method.

QC procedures applied to observations, however, may vary among gridded SST products. Therefore, two QC options are used in our assessment: (a) filtering out the outliers deviated from the DOISST FG by more than one SST standard deviation (SD; Reynolds et al. 2007), and (b) filtering out the outliers deviated from climatological SST (CLM) by four times SST SDs (Huang et al. 2017).

Intercomparisons are quantified by globally averaged bias (Bias) and root-mean-square-difference (RMSD), or mean difference (DIFF) and SD:

$$Bias(t) = \frac{1}{W}\sum_{i=1}^{M}\sum_{j=1}^{N}[P(x_i, y_j, t) - O(x_i, y_j, t)] \times \cos(y_j) \quad (1)$$

$$RMSD(t) = \{\frac{1}{W}\sum_{i=1}^{M}\sum_{j=1}^{N}[P(x_i, y_j, t) - O(x_i, y_j, t)]^2 \times \cos(y_j)\}^{0.5} \quad (2)$$

$$DIFF(t) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[P(x_i, y_j, t) - O(x_i, y_j, t)] \quad (3)$$

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

$$SD(t) = \{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[P(x_i, y_j, t) - O(x_i, y_j, t) - DIFF(t)]^2\}^{0.5} \quad (4)$$

where *P* and *O* represent product and observed SSTs at grid *x* and *y*; *x, y*, and *t* represent longitude, latitude, and time, respectively; *W* represents the integrated weighting of $cos(y_j)$. The Bias and RMSD in equations (1)−(2) are weighted by cos(latitude) because they are calculated on 0.25º×0.25º grid-boxes; while DIFF and SD in equations (3)−(4) are not weighted by cos(latitude) because they are calculated on pointwise locations by the bilinear interpolation from *P* to *O*. The uncertainties of Bias, RMSD, DIFF, and SDs at 95% confidence level are quantified by estimating the effective sampling number according to lagged autocorrelation coefficients of time series (Huang et al. 2021b). The Bias and DIFF could be positive or negative when an SST product is warmer or colder than the reference SST. Therefore, Bias and DIFF could be canceled with each other when they are integrated over the global oceans. In contrast, RMSD and SD are always positive and cannot be canceled with each other when they are integrated over the global oceans.

3. Intercomparisons

3.1 GMPE as a reference

The eight gridded SST products were compared with GMPE after OSTIA, CCI, GPB, and CMC were box-averaged to 0.25º×0.25º grids from January 2016 to January 2022 (Fig 1a). GMPE has frequently been used as a reference to assess the performance of SST products (Yang et al. 2021; Huang et al. 2021a; Dash et al. 2012), because it selects the median of various SST products and therefore its biases are relatively small (Fiedler et al. 2019). Comparisons in Figure 1a show that the globally averaged biases are generally within ±0.15ºC varying with time. The biases are mostly positive in MUR25 and MGDSST, mostly negative in GAMSSA, OSTIA, GPB and CCI, and near zero in DOISST and CMC (Table 2). The biases in MGDSST decrease clearly after 2019, whose reasons are not quite clear but may be associated with (a) the selection of the GMPE among numerous SST products since its biases relative to Buoy SSTs are stable as discussed in section 3.2 (Fig. 3a) and (b) the use of ACSPO data after December 2018. The warm biases in MGDSST and MUR25 may be associated with their use of SST observations derived from microwaves (Crewell et al. 1991). The cold biases in GAMSSA, OSTIA, and GPB may partially be associated with the use of nighttime VIIRS SST and rejecting SST measurements during daytime in low wind speed (Martin et al. 2012). The RMSDs are between 0.1ºC and 0.5ºC (Fig. 1b). The RMSDs are

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

low in OSTIA, GPB and CMC (approximately 0.2ºC), and higher in DOISST, MUR25, MGDSST, GAMSSA and CCI (approximately 0.3ºC), which is consistent with the SQUAM analysis at https://www.star.nesdis.noaa.gov/socd/sst/squam/analysis/l4. The RMSDs are generally higher in boreal summer than in boreal winter, which may result from the availability of GMPE that shifts towards the North Pole in boreal summer since the performance of SST products are generally worse in high latitudes. The reason for the data availability of GMPE is not clear, but may be that some SST products are not globally covered or filtered out by sea-ice concentrations.

SST product	**GMPE reference**		**Buoy reference**		**Argo reference**	
	Bias	RMSD	Bias	RMSD	Bias	RMSD
DOISST v2.1	0.002±0.017	0.357±0.014	-0.018±0.013	0.376±0.009	-0.033±0.007	0.346±0.002
MUR25	0.087±0.010	0.281±0.011	0.038±0.010	0.531±0.015	0.036±0.005	0.377±0.004
MGDSST	0.051±0.029	0.391±0.016	0.028±0.009	0.650±0.047	0.006±0.019	0.523±0.011
GAMSSA	-0.024±0.019	0.303±0.022	-0.071±0.011	0.505±0.024	-0.088±0.010	0.480±0.005
OSTIA	-0.020±0.011	0.200±0.016	-0.045±0.011	0.431±0.012	-0.069±0.011	0.370±0.042
GPB	-0.016±0.014	0.203±0.041	-0.051±0.015	0.505±0.016	-0.066±0.011	0.381±0.020
CCI	-0.025±0.013	0.349±0.018	-0.054±0.011	0.608±0.036	-0.068±0.008	0.429±0.017
CMC	-0.001±0.006	0.182±0.012	-0.052±0.009	0.492±0.015	-0.056±0.007	0.380±0.002
GMPE	N/A	N/A	-0.036±0.012	0.454±0.012	-0.055±0.007	0.363±0.011

Table 2. Averaged Biases and RMSDs (ºC) in reference to GMPE, Buoy, and Argo SSTs on 0.25º×0.25º grids from January 1, 2016 to January 31, 2022 in Figures 1, 3, and S1. The ±values represent the uncertainty at 95% confidence level that is determined by the lagged autocorrelation, effective sampling number, and the standard deviation (SD) (Huang et al. 2021b).

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

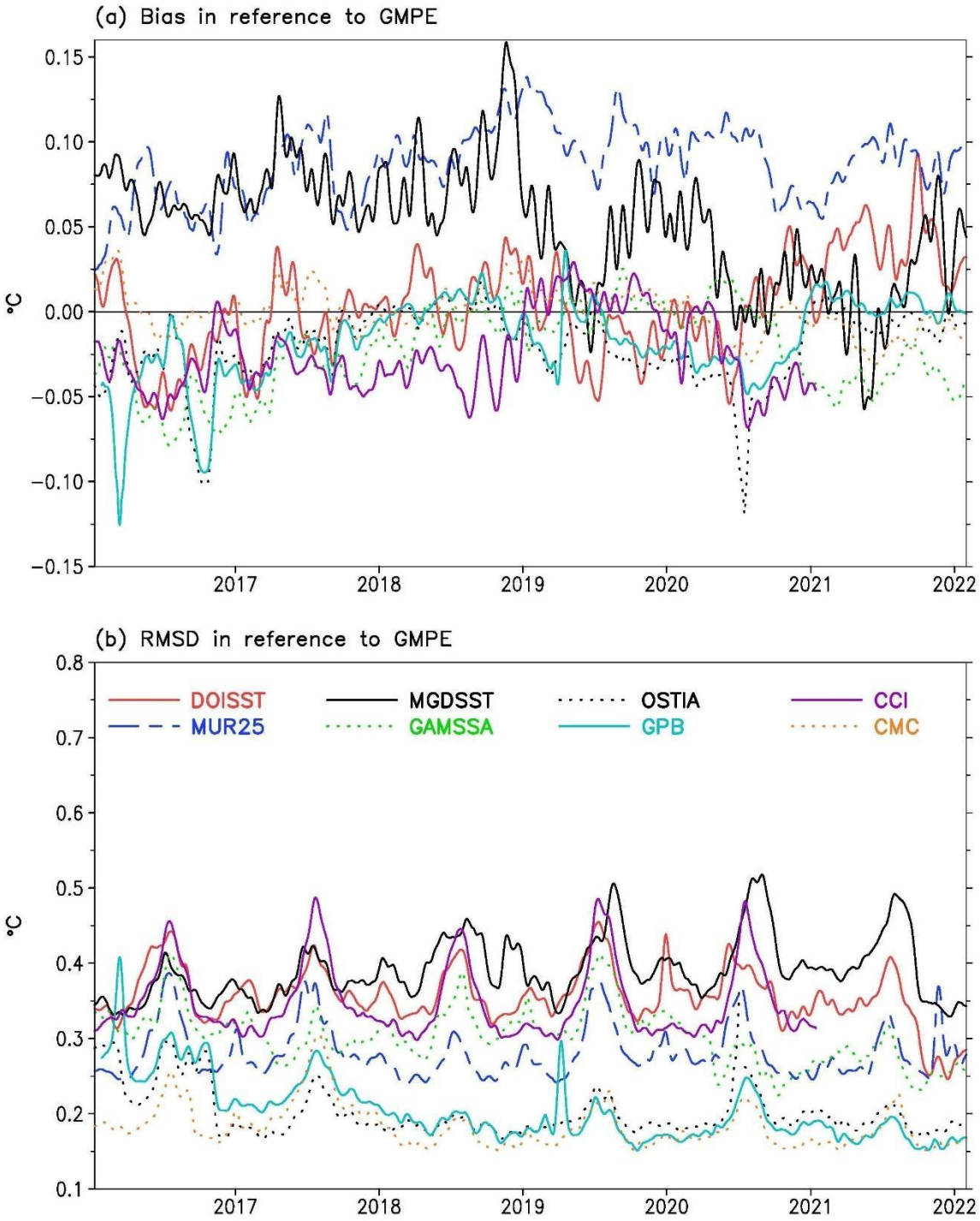

Figure 1. (a) Biases and (b) RMSDs in reference to GMPE in DOISST (solid red), MUR25 (dashed blue), MGDSST (solid black), GAMSSA (dotted green), OSTIA (dotted black), GPB (solid light blue), CCI (solid purple), and CMC (dotted orange). The biases and RMSDs are calculated on 0.25º×0.25º grids. A 15-day running filter is applied in plotting.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

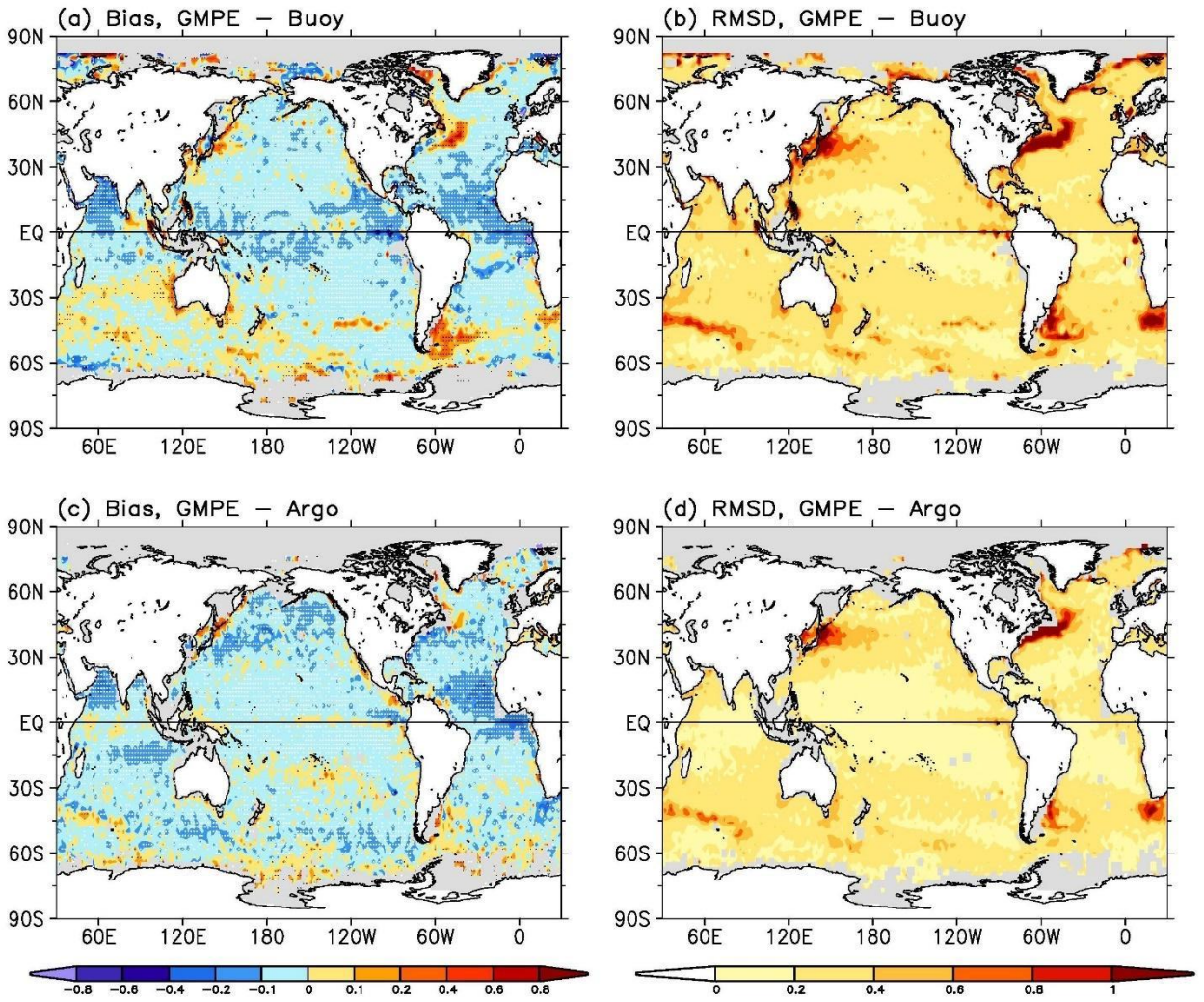

Figure 2. (a) Bias and (b) RMSD of GMPE in reference to Buoy SSTs on 0.25º×0.25º grids from January 1, 2016 to January 31, 2022 in units of ºC. (c) and (d) same as (a) and (b) except for in reference to Argo SST. The biases in (a) and (c) are stippled when they are significant at the 95% confidence level, and the areas without observations are shaded with gray.

However, GMPE itself is biased when it is compared with in situ Buoy and Argo observations during both day- and night-time from January 2016 to January 2022 (Fig. 2). GMPE relative to Buoy SSTs has cold biases (-0.1ºC) in the tropical oceans between 15ºS and 30ºN, warm biases (0.1º to 0.2ºC) south of 15ºS in the Indian Ocean sector and south of 30ºS in the Pacific-Atlantic sectors, and warm biases (0.2º to 0.4ºC) in the regions of the Gulf Stream and Kuroshio. The RMSDs are above 1.0ºC in the regions of the Gulf Stream and Kuroshio, 0.6º to 1.0ºC in the

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

Southern Ocean between 30ºS and 60ºS, and approximately 0.2ºC in the rest of the oceans. The globally averaged bias and RMSD are about -0.04ºC and 0.5ºC, respectively (Table 2).

The spatial distribution of biases and RMSDs of GMPE relative to Argo SSTs are similar to those relative to Buoy SSTs (Figs. 2c and 2d). Exceptions are that the cold biases and RMSDs are slightly lower in the tropical Pacific and Indian Oceans; the warm biases and RMSDs are slightly lower in the regions of the Gulf Stream and Kuroshio and the Southern Ocean. The differences relative to Buoy and Argo SSTs may result from the differences of spatial and time coverages in Buoy and Argo since their observation densities are very distinct (Fig. S2), since the collocated difference between Buoy and Argo SSTs are small (0.03º±0.03ºC; Huang et al. 2017). Nevertheless, the globally averaged bias and RMSD are approximately -0.06ºC and 0.4ºC (Table 2), respectively, which is comparable with the comparisons against Buoy SSTs.

Our analyses indicate that the magnitude of biases and RMSDs in GMPE relative to in situ SSTs are comparable to those in gridded SST products relative to GMPE. Therefore, it is problematic using GMPE as a reference SST to assess the performance of other SST products. A better reference is high-quality in situ observations such as Buoy and Argo SSTs.

3.2 Buoy and Argo as references

The globally averaged biases (Biases) and RMSDs [equations (1)−(2)] relative to ICOADS Buoy (both drifting and moored buoys) observations in the nine gridded SST products including GMPE are calculated on 0.25º×0.25º grids (Fig. 3a). Figure 3a shows that MGDSST and MUR25 are largely warm biased, which may be associated with their use of SST observations derived from microwaves (Crewell et al. 1991). In contrast, DOISST, GAMSSA, OSTIA, GPB, CCI, CMC and GMPE are cold biased. The cold bias in GMPE is clearly seen. The magnitude of the biases is relatively small in DOISST (-0.02ºC; Table 2) and MGDSST (+0.03ºC). The cold biases in GAMSSA, OSTIA, GPB and CMC may partially be associated with the bias correction algorithms using nighttime VIIRS SST and rejecting SST measurements during daytime in low wind speed (Martin et al. 2012). Figure 3b shows that RMSDs are relatively small in DOISST (0.38ºC) and OSTIA (0.43ºC), large in MGDSST and CCI, and in between in MUR25, GAMSSA, GPB, CMC and GMPE, which is consistent with Huang et al. (2021b). The low RMSD in DOISST may result partially from the FG QC procedure applied to the in situ Buoy (and Argo) observations in both DOISST analysis and reference SST, which will be discussed further in section 3.4.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

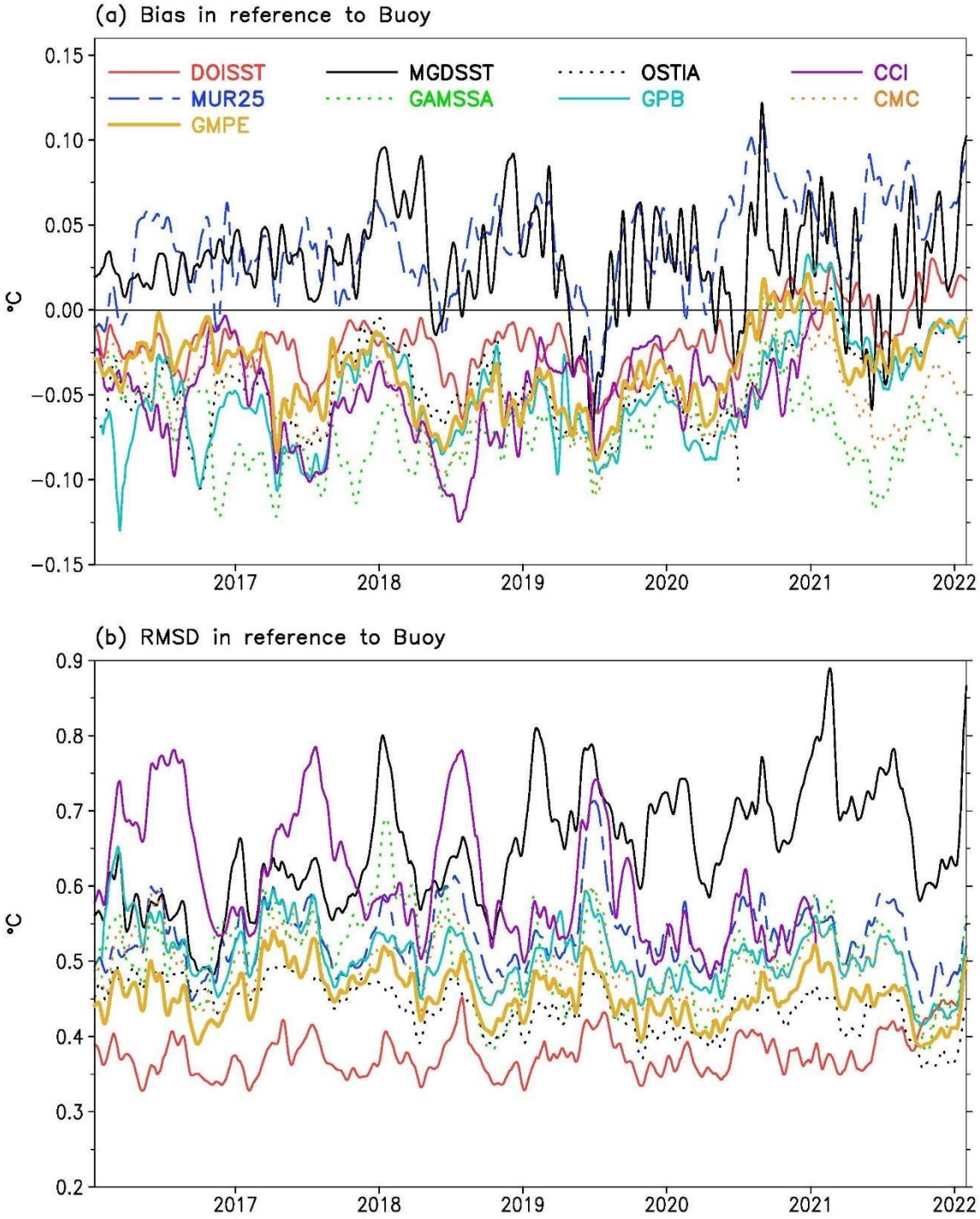

Figure 3. (a) Biases and (b) RMSDs in reference to Buoy SSTs in DOISST (solid red), MUR25 (dashed blue), MGDSST (solid black), GAMSSA (dotted green), OSTIA (dotted black), GPB (solid light blue), CCI (solid purple), CMC (dotted orange), and GMPE (solid orange). The biases and RMSDs are calculated on 0.25º×0.25º grids. A 15-day running filter is applied in plotting.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

In comparison with Argo observations, shown in supplemental materials (Fig. S3a), the biases in the nine gridded products are similar to those in comparison with Buoy observations (Fig. 3a). Exceptions are that the warm biases decrease in MGDSST after 2019 and in MUR25 after 2020. The biases remain small in DOISST (-0.03ºC; Table 2) and MGDSST (+0.01ºC). RMSDs (Fig. S3b) are mostly smaller than those relative to Buoy SSTs (Fig. 3b), which may be associated with the low observing frequency of Argo SSTs and will be discussed further in section 3.3b. The maximum RMSD reduces from about 0.8ºC (Fig. 3b) to 0.5º−0.6ºC (Fig. S3b) in MGDSST and CCI, although the RMSD does not reduce much in DOISST. The RMSDs are relatively low and close with each other in MUR25, OSTIA, GPB, CCI, CMC and GMPE, while the RMSDs are relatively higher in MGSDDT and GAMSSA. Overall, the RMSDs remain relatively small in DOISST (0.35ºC) and OSTIA (0.37ºC), being consistent with the comparison with Buoy observations. However, the low RMSDs in DOISST may in part result from the inclusion of Argo observations (Huang et al. 2021a). It is interesting to note that there is a negative trend of RMSD among most of the SST products except for MGDSST and GAMSSA, which may represent the improvements of the SST analyses, Argo observations, and satellite observations.

The main body of this paper in the following sections focuses on the buoy data as the reference data source, whereas the results using Argo float data as the refence are provided in the supplements. We focus on the buoy data for the following reasons: (a) there are many more (10 times or higher; Huang et al. 2017) Buoy than Argo SSTs as indicated by the observation density shown in the Supplemental Figure S2, and therefore validations against Buoy observations is more reliable; (b) the observing frequency of Buoy (6 minutes to 1 hour in moored buoy and 1 hour in drifting buoy) is much higher than that of Argo (10 days), which can better resolve the high-frequency (1 day) SST variability; (c) DOISST ingests Argo while most of other products do not, and therefore the validations against Argo are unfair to other products. In contrast, the validations against Buoy observations are fair to all products except for CCI; and (d) overall, the conclusions are consistent regardless as to whether Buoy or Argo SSTs are used in the validations. Despite the differences between Buoy and Argo observations, their SST differences at collocated grid-boxes are small (0.03º±0.03ºC) (Huang et al. 2017). Therefore, the figures and tables for validations using Argo as a reference are put to the Supplemental Materials.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

3.3 Impact of resolution of SST products

(a) In reference to Buoy SSTs

The Buoy SSTs from ICOADS used in the intercomparisons in section 3.2 are box-averaged to the daily 0.25º×0.25º grids before comparison to SST products. This is to evaluate analysis accuracy at the grid scale. For highest-resolution validation, it would be necessary to interpolate the analyses to the buoy locations and times before comparison, as in Dash et al. (2012) and Fiedler et al. (2019).

By interpolating gridded SST products to the in situ locations of Buoy SSTs, SDs and DIFFs without cos(latitude) weighting [equations (3)−(4)] are calculated because these are considered to be point values, rather than grid values representative of a region that varies with latitude. SDs are close to the RMSDs without cos(latitude) weighting since DIFFs are relatively small. The use of SDs rather than RMSD is for the intercomparison purpose as presented in SQUM (Dash et al. 2012; Fiedler et al. 2019; https://www.star.nesdis.noaa.gov/socd/sst/squam/analysis/l4). Our analyses showed that SDs in in situ locations are sensitive to the spatial resolution of SST products, and therefore the intercomparisons hereafter are focused on DOISST, OSTIA, GPB, CCI, and CMC with resolutions of 0.25º, 0.05º, 0.05º, 0.05º, and 0.10º, respectively. The results for MUR25, MGDSST, and GAMSSA are similar to DOISST due to their coarse resolution of 0.25º.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

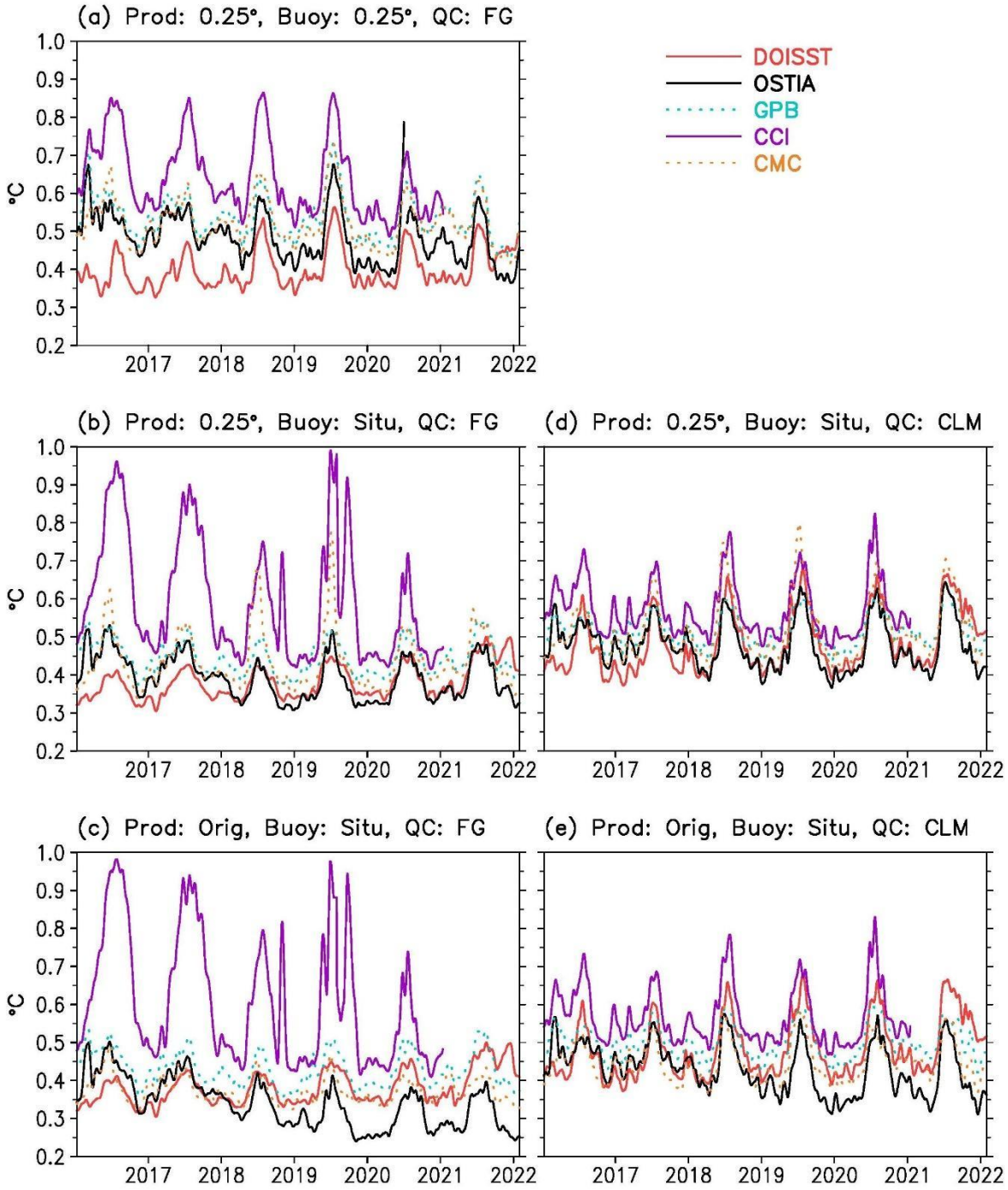

Figure 4. SDs between SST products (Prod) and Buoy SSTs matched with (a) Prod: 0.25º, Buoy: 0.25º, QC: FG, (b) Prod: 0.25º, Buoy: Pointwise in situ (Situ), QC: FG, (c) Prod: Orig, Buoy: Situ, QC: FG (d) Prod: 0.25, Buoy: Situ, QC: CLM, and (e) Prod: Orig, Buoy: Situ, QC: CLM. SDs are calculated without cos(latitude) weighting in DOISST (solid red), OSTIA (solid black), GPB

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

(dotted light blue), CCI (solid purple), and CMC (dotted orange). A 15-day running filter is applied in plotting.

As a reference, SDs are calculated first between gridded SST products and Buoy observations on 0.25º×0.25º grids (Fig. 4a). SDs are small in DOISST (0.40ºC; Table 3) and larger in CCI (0.64ºC), which is consistent with the assessment using RMSDs. The high SD in CCI may result from the fact that CCI is independent from Buoy observations (Table 1). The contrast of SDs in DOISST and CCI is consistent with that of RMSDs in Table 2.

SST product	**Prod:** 0.25º, **Buoy:** 0.25º, **QC:** FG	**Prod:** 0.25º, **Buoy:** Situ, **QC:** FG	**Prod:** Orig, **Buoy:** Situ, **QC:** FG	**Prod:** 0.25º, **Buoy:** Situ, **QC:** CLM	**Prod:** Orig, **Buoy:** Situ, **QC:** CLM
DOISST v2.1	0.399±0.018	0.381±0.021	0.381±0.021	0.483±0.032	0.483±0.032
OSTIA	0.481±0.025	0.391±0.025	0.339±0.055	0.477±0.024	0.429±0.025
GPB	0.536±0.021	0.443±0.017	0.430±0.018	0.504±0.012	0.493±0.013
CCI	0.643±0.041	0.587±0.067	0.593±0.069	0.573±0.026	0.569±0.027
CMC	0.523±0.023	0.435±0.027	0.374±0.012	0.520±0.029	0.447±0.018

Table 3. SDs (ºC) between SST products (Prod) and Buoy SSTs with different matching methods from January 1, 2016 to January 31, 2022 in Figure 4. The ±values represent the uncertainty at 95% confidence level. Globally averaged SDs are calculated without cos(latitude) weighting.

When these SST products on 0.25º×0.25º grids are interpolated to in situ locations of Buoy observations (Fig. 4b), the overall SDs decrease in all five products. Exceptions are that the maximum SDs increase in CCI during the boreal summer of 2016, 2017 and 2019, and increase in CMC in the boreal summer of 2019. The reason for the large seasonal peaks of SDs in CCI are not clear. But it might be possible that CCI is less reliable in the high latitudes, which is detected by more observations during the boreal summer. Particularly, these observations are independent from CCI. The peaks in boreal summer are also visible in other products due to the same reasons,

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

but the peaks are not as strong as those in CCI due to dependence between SST products and Buoy observations.

On average from January 2016 to January 2022, SDs are small in DOISST (0.38ºC; Table 3) and OSTIA (0.39ºC) and larger in CCI (0.59ºC), indicating a good performance of DOISST and OSTIA. However, the amplitude of SD decrease is much smaller in DOISST (about 0.02ºC) than the other four products (about 0.09ºC), which may indicate an advantage of the higher spatial resolution in OSTIA, GPB, CCI and CMC. Therefore, the performance difference among the five products becomes small.

The advantage of high-resolution in OSTIA and CMC is more clear when these gridded SST products are interpolated from their original high-resolution grids to the in situ locations of Buoy observations (Fig. 4c), which is indeed the case in Dash et al. (2012; https://www.star.nesdis.noaa.gov/socd/sst/squam/analysis/l4) and Fiedler et al. (2019). SDs become lower in OSTIA (0.34ºC; Table 3) and CMC (0.37ºC) than DOISST (0.38ºC). In addition, the SDs in OSTIA decrease with time. The improvement of OSTIA performance may result from its unique use of SEVIRI and SLSTR whose quality improves with time.

The advantage of high-resolution in OSTIA and CMC may be understood by the distance between the latitude-longitude locations of gridded SST products and in situ locations of Buoy observations. Assuming the Buoy observations were randomly distributed within a typical grid-box of the SST products, which should be reasonable in the global oceans and within a long time period, the averaged distance from the center of a gridbox to a Buoy observation can statistically be approximated by 0.38δ where δ is the size of the grid-box (https://math.stackexchange.com/questions/15580). Therefore, for the SST products that have a higher resolution as in OSTIA and CMC, the distance from their central grid-point to a Buoy observation is shorter, which may enable the SST products to be closer to the Buoy observation and therefore a smaller SD of the SST differences.

Furthermore, SDs may be associated with the spatial scale of satellite bias correction. In OSTIA and CMC, the biases of satellite SSTs are analyzed on 7º and 2.5º grids, respectively, according to the matchups of in situ and satellite SSTs within 25 km and 1 day. This indicates that the satellite SSTs in these two gridded products were in principle adjusted according to the nearby (25 km) in situ SSTs. In contrast, in DOISST, the biases are calculated within 3000 km in latitude,

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

5000 km in longitude, and 15-day data window. The bias correction in a small scale in OSTIA and CMC may explain why their SDs are lower when SSTs are interpolated from their original grid to in situ locations of Buoy SSTs. In GPB, the bias correction procedure is the same as in OSTIA. However, the SST analysis in GPB is further adjusted by an independent NCEP analysis to prevent a slow drift in its analysis, which may explain why the SD in GPB is not sensitive to how its analysis is interpolated to in situ locations of Buoy SSTs. In CCI, its analysis does not use in situ SSTs at all, and therefore its SD does not change whether CCI is compared with Buoy SSTs in 0.25º or original 0.05º resolution.

In contrast to the reductions of SDs, DIFFs do not change much whether gridded SST products are interpolated from in 0.25º or their original resolutions to in situ Buoy locations (Figs. 5a−5c; Table 4). The robust DIFFs may result from that the biases in different regions have been cancelled with each other. Overall, DOISST has a lower DIFF. The low DIFF in DOISST may result from that the biases in satellite SSTs in DOISST are corrected in very large spatial scales, and therefore globally averaged DIFF is smaller in DOISST. However, the DIFF in DOISST increases after 2020, which may result from a constant bias-correction of ship observations based on 2016−2019 values (Huang et al. 2021a).

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

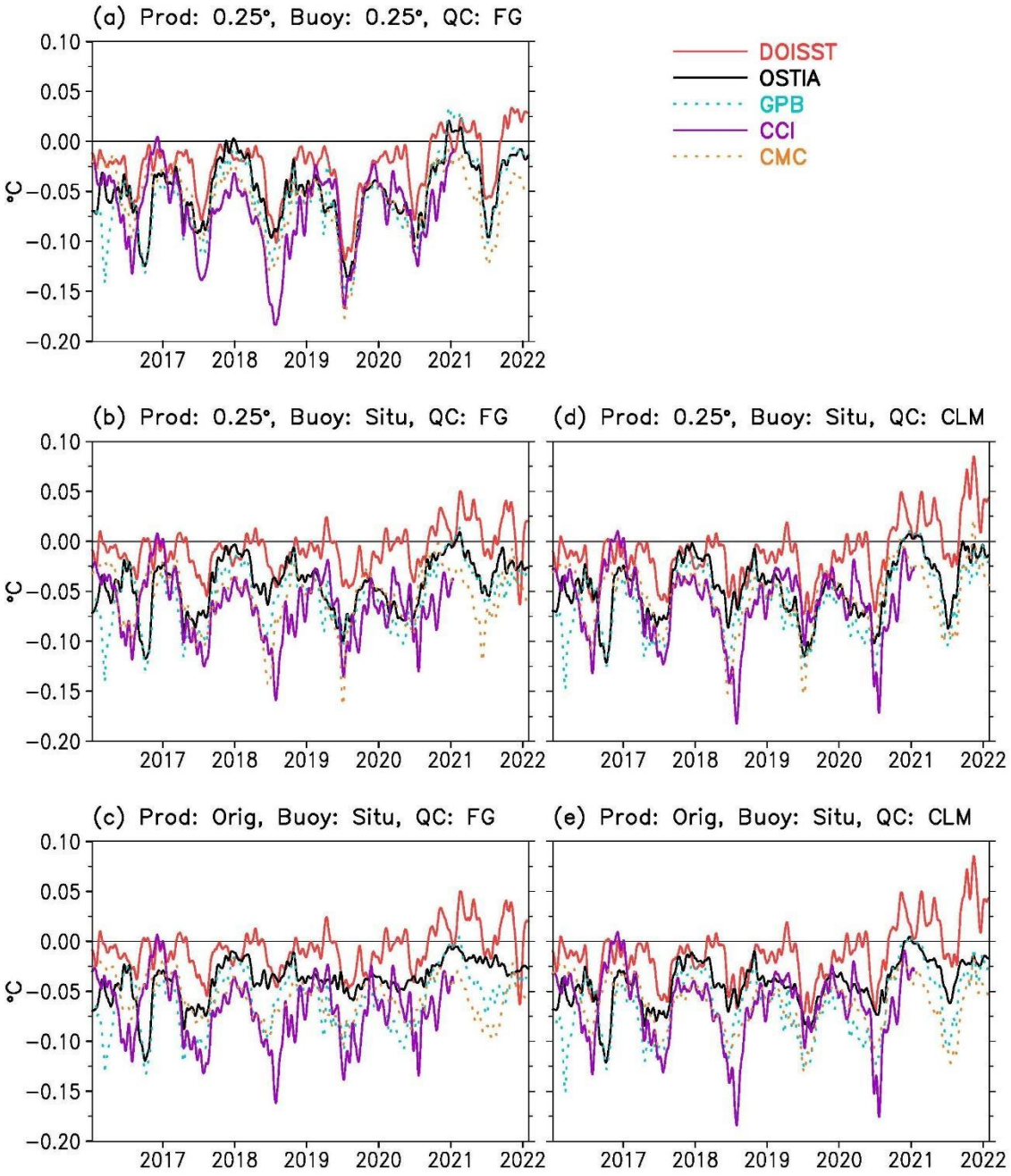

Figure 5. Same as Figure 4 except for DIFFs.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

SST product	**Prod:** 0.25º, **Buoy:** 0.25º, **QC:** FG	**Prod:** 0.25º, **Buoy:** Situ, **QC:** FG	**Prod:** Orig, **Buoy:** Situ, **QC:** FG	**Prod:** 0.25º, **Buoy:** Situ, **QC:** CLM	**Prod:** Orig, **Buoy:** Situ, **QC:** CLM
DOISST v2.1	-0.025±0.012	-0.008±0.007	-0.008±0.007	-0.011±0.011	-0.011±0.011
OSTIA	-0.048±0.012	-0.043±0.010	-0.038±0.008	-0.046±0.012	-0.043±0.009
GPB	-0.057±0.016	-0.055±0.012	-0.060±0.012	-0.06±0.014	-0.064±0.014
CCI	-0.072±0.016	-0.067±0.012	-0.070±0.012	-0.062±0.013	-0.066±0.013
CMC	-0.060±0.014	-0.056±0.011	-0.057±0.008	-0.055±0.012	-0.063±0.010

Table 4. Same as Table 3 except for DIFFs Figure 5.

(b) In reference to Argo SSTs

In comparison with the FG QC applied Argo SSTs, SDs are low in DOISST, slightly higher in CCI, and in between in OSTIA, GPB and CMC (Figs. S4a−S4c; Table S1). The low SD in DOISST may partly result from that DOISST ingests the Argo SSTs so that they are not independent, which will be discussed further in section 4. In addition, there exhibits some interesting features in the SDs: (a) the SDs are overall smaller than those relative to Buoy SSTs, particularly in CCI; the low SDs relative to Argo may partially result from the weaker diurnal variation in Argo than in Buoy SSTs, since the depth of SST measurement is deeper in Argo (above 5m) than in Buoy (0.2−1.0 m); (b) the SDs do not change much whether these products are interpolated from their 0.25º or original resolution to the Argo locations, which may be due to the fact that Argo SSTs are independent from most of these SST products; (c) the differences of the SDs among products are small in comparison with those relative to Buoy SSTs, which may suggest that the performance of these SST products are close at a longer timescale since Argo SSTs are sampled at 10 day cycle; and (d) SDs decrease generally with time, which may indicate a general improvement of these SST products due to improved observations.

In comparison with Argo SSTs, DIFFs are low in DOISST and CMC, higher in OSTIA, GPB, and CCI (Figs. S5a−5c; Table S2). The low DIFF in DOISST may be associated with the bias correction to the satellite SST within a very large spatial scale as described in section 3.3a. The DIFFs do not change much whether these SST products are interpolated from 0.25º or their

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

original resolution to in situ Argo locations. These are consistent with the comparisons in reference to Buoy SSTs.

3.4 Impact of QC for reference SSTs

The intercomparisons in section 3.3 use the reference Buoy and Argo SSTs that have passed the FG QC, which filter out the outliers beyond one SST SD from the FG as described in section 3.2. Since the same FG QC was applied in the DOISST analysis, one may argue that the good performance of DOISST in the intercomparisons may result from using the same in situ data that have passed the FG QC. Indeed, the SST observations filtered out by the FG QC in DOISST analysis may not be filtered out in the other gridded SST analysis systems, and vice versa. To clarify whether the performance of DOISST relies on using FG QC applied reference SSTs, the intercomparisons are repeated in reference to the CLM QC applied Buoy and Argo SSTs. The CLM QC filters out outliers beyond four SST SDs from SST climatology. Overall, the CLM QC is relatively loose and more observations can pass the CLM QC. The CLM QC is also more independent from SST analysis systems, since the SST climatologies are closer than the SST FGs among SST products. However, it should be noted that the selections of one SD in FG QC and four SDs in CLM QC are somehow subjective and can be different among analysis systems.

In comparison with SDs against Buoy SSTs that have passed the FG QC, the SDs in reference to the CLM QC applied Buoy SSTs increase in all gridded products whether these SST products are interpolated from 0.25º or their original resolution (Figs. 4d−4e; Table 3). The exception is that the maximum SDs in CCI decrease clearly during the boreal summers of 2016, 2017, and 2019. The reasons for the reduction of SDs are not immediately clear but may be associated with the independence of CCI from in situ observations, because the increase of observations in CLM QC validation may partially cancel the high SDs in the high latitudes. The time averaged SD against the CLM QC applied Buoy SSTs decreases slightly although the SDs during boreal winters do increase in CCI. Therefore, it should be cautious in assessment of CCI because of its sensitivity to the selection of in situ observations that have passed different QC criteria. The overall increase in SDs in those SST products is intuitive because more observations are included in reference Buoy SSTs due to a loose CLM QC, which is mostly from moored-buoys (by 5−15%) and slightly from drifting-buoy (about 2%). Nevertheless, our conclusions remain the same: SDs are lower in DOISST and OSTIA when those SST products in 0.25º resolution are

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

interpolated to Buoy SSTs; and SDs are lower in OSTIA and CMC when those SST products in their original resolution are interpolated to Buoy SSTs, which results from the advantage of high resolution in OSTIA and CMC as analyzed in section 3.3. The better performance of OSTIA in both FG QC and CLM QC may be associated with using more satellite observations (Table 1). The performance of SST products when they are interpolated from their original resolution is consistent with the SQUAM analyses at https://www.star.nesdis.noaa.gov/socd/sst/squam/analysis/l4. Comparisons also show that DIFFs are lower in DOISST than the other four SST products whether they are interpolated from 0.25º or their original resolution (Figs. 5d−5e; Table 4). The low DIFFs in DOISST are associated with the largescale bias correction to the satellite SSTs, and are consistent with the SQUAM analysis.

In the comparison against the CLM QC applied Argo SSTs (Figs. S4d−S4e; Table S1), the overall features of SDs are similar to those against the FG QC applied Argo SSTs described in section 3.3b. Exception is that there is a clear increase of SDs in all gridded products in comparison with those using tight FG QC, since the CLM QC applied Argo observations are about 1% more than the FG QC applied Argo observations. Likewise, the DIFFs are smaller in DOISST whether these SST products are interpolated from 0.25º or their original resolution and whether FG QC or CLM QC is applied to the reference Argo SSTs (Figs. S5d−S5e; Table S2). These results suggest that the intercomparisons are not sensitive to the selections of QC applied to the reference Buoy and Argo SSTs.

3.5 Impact of moored-buoys

In the comparisons in sections 3.3 and 3.4, it is noticed that the SDs and RMSDs in reference to Buoy SSTs from ICOADS are mostly higher than those in reference to Argo SSTs, particularly in CCI. The difference may result from the higher observing frequency of SSTs from buoys, particularly the moored-buoys. The high-frequency variances may not be well resolved by the gridded SST products at daily resolution, and that the SDs may be overwhelmed by high-frequency moored-buoys due to use of observations at in situ locations and times. Since the locations of the moored-buoys do not change with time, the globally averaged SDs may be biased toward the locations of the moored-buoys.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

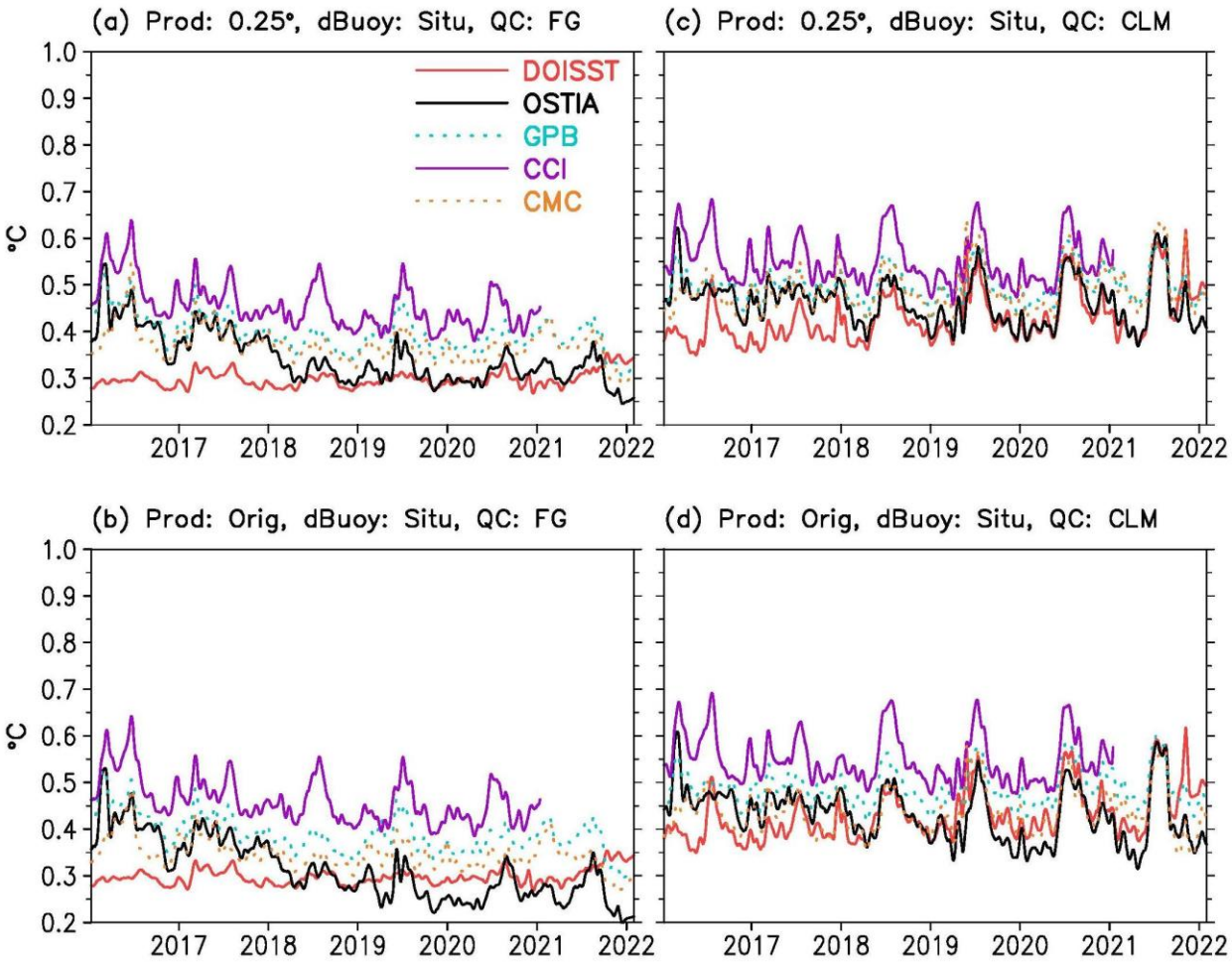

Figure 6. SDs between SST products (Prod) and dBuoy SSTs matched with (a) Prod: 0.25º, dBuoy: Situ, QC: FG, (b) Prod: Orig, dBuoy: Situ, QC: FG, (c) Prod: 0.25º, dBuoy: Situ, QC: CLM, and (d) Prod: Orig, dBuoy: Situ, QC: CLM, SDs are calculated without cos(latitude) weighting in DOISST (solid red), OSTIA (solid black), GPB (dotted light blue), CCI (solid purple), and CMC (dotted orange). A 15-day running filter is applied in plotting.

By removing the moored-buoy from the reference Buoy SST, the comparisons with drifting-buoy (dBuoy) SSTs indeed show an overall decrease of SDs (Fig. 6 and Table 5) from those in the comparisons with Buoy (both drifting and moored buoys) in Figure 5 and Table 3. However, the performance among the five gridded SST products remain similar to that in comparison with combined moored- and drifting-buoy SSTs: SDs are low in DOISST and OSTIA whether FG or CLM QCs are applied to dBuoy SSTs and whether these SST products are

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

interpolated from 0.25º or original resolution to in situ dBuoy locations. The SD is relatively low in CMC when the analysis is interpolated from its original high-resolution and FG QC is applied to the reference dBuoy SSTs. However, the SDs in comparison with dBuoy SSTs (Table 5) remain larger than those in comparison with Argo SSTs (Table S1), which may result from that the observing frequency of dBuoy (1 hour) remain higher than that of Argo (10 days). It is interesting to note that the SDs in OSTIA decrease with time when compared with the Buoy SSTs with FG QC (Fig. 6c). The improvement of OSTIA performance may result from its unique use of SEVIRI and SLSTR whose quality improves with time as noted in section 3.3a.

SST product	**Prod:** 0.25º, **dBuoy:** Situ, **QC:** FG	**Prod:** Orig, **dBuoy:** Situ, **QC:** FG	**Prod:** 0.25º, **dBuoy:** Situ, **QC:** CLM	**Prod:** Orig, **dBuoy:** Situ, **QC:** CLM
DOISST v2.1	0.297±0.007	0.297±0.007	0.430±0.020	0.430±0.020
OSTIA	0.343±0.027	0.312±0.032	0.462±0.017	0.430±0.020
GPB	0.396±0.018	0.388±0.019	0.495±0.009	0.485±0.010
CCI	0.457±0.020	0.462±0.019	0.552±0.016	0.549±0.016
CMC	0.370±0.015	0.347±0.014	0.491±0.014	0.436±0.015

Table 5. SDs (ºC) between SST products (Prod) and dBuoy SSTs with different matching methods from January 1, 2016 to January 31, 2022 in Figure 6. The ±values represent the uncertainty at 95% confidence level. Globally averaged SDs are calculated without cos(latitude) weighting.

SST product	**Prod:** 0.25º, **dBuoy:** Situ, **QC:** FG	**Prod:** Orig, **dBuoy:** Situ, **QC:** FG	**Prod:** 0.25º, **dBuoy:** Situ, **QC:** CLM	**Prod:** Orig, **dBuoy:** Situ, **QC:** CLM
DOISST v2.1	-0.011±0.012	-0.011±0.012	-0.021±0.011	-0.021±0.011
OSTIA	-0.039±0.010	-0.035±0.009	-0.047±0.013	-0.044±0.011
GPB	-0.048±0.015	-0.049±0.015	-0.054±0.015	-0.056±0.015
CCI	-0.044±0.011	-0.047±0.011	-0.052±0.014	-0.055±0.014
CMC	-0.049±0.007	-0.051±0.007	-0.057±0.012	-0.062±0.011

Table 6. Same as Table 5 except for DIFFs in Figure 7.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

Similar to the comparisons with Buoy and Argo SSTs, DIFFs in comparisons with dBuoy SSTs remain low in DOISST whether the SST products are interpolated from 0.25° or their original resolution and whether FG QC or CLM QC is applied to dBuoy SSTs (Fig. 7; Table 6). The low DIFF in DOISST may result from the algorithm of the bias correction to the satellite SST within a large spatial scale at 3000−5000 km.

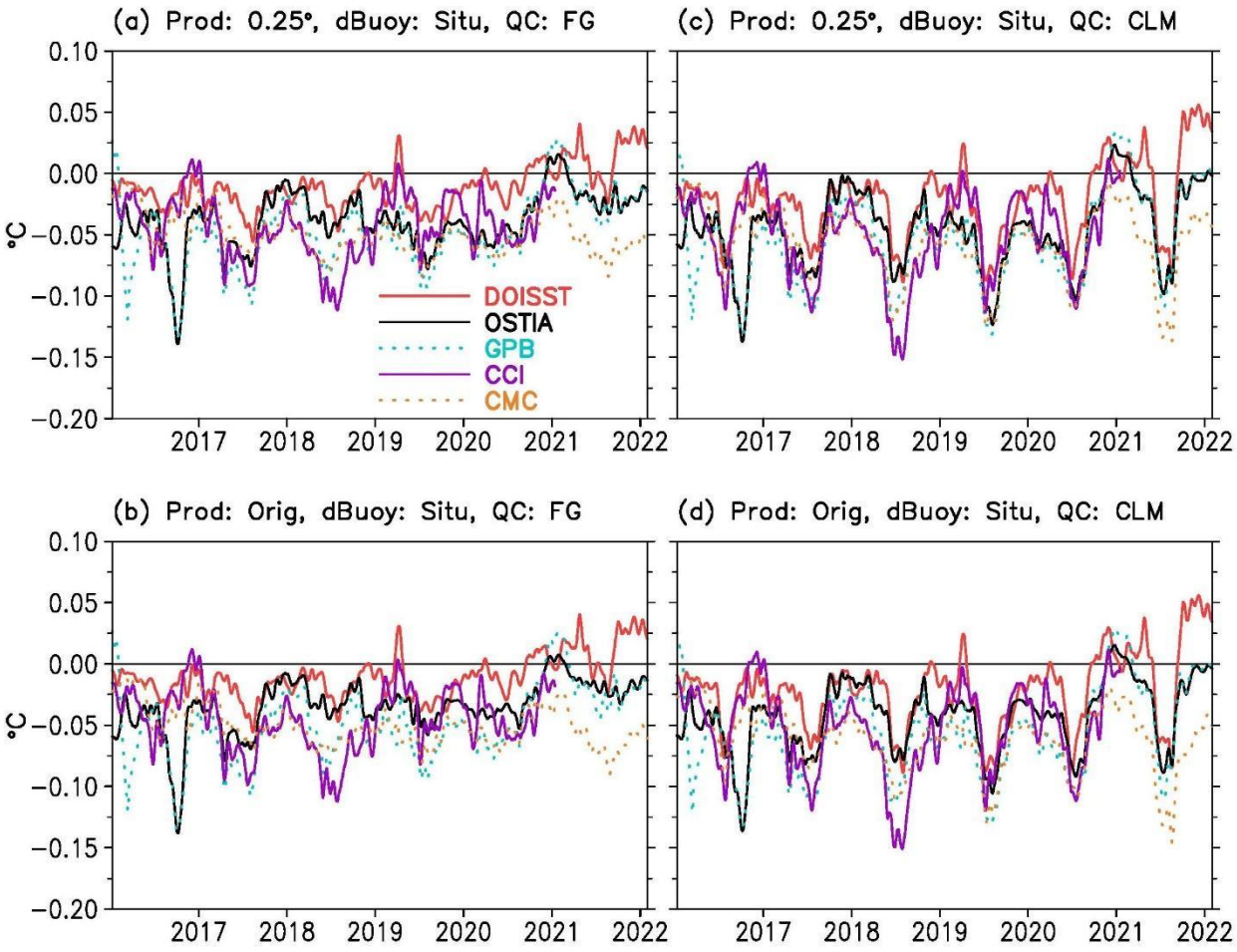

Figure 7. Same as Figure 6 except for DIFFs.

3.6 Impact of iQuam SST

The analysis in section 3.5 (Fig. 4e) shows overall higher SDs in comparison with those on the SQUAM website (Fig. S1; https://www.star.nesdis.noaa.gov/socd/sst/squam/analysis/l4). Our

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

analyses demonstrate that the low SDs in SQUAM result from using the high quality (QC flag 5; Dash et al. 2012) drifting and tropical-moored buoy SSTs from iQuam (Xu and Ignatov 2010).

As an example, Figure 8a shows the SDs of OSTIA relative to iQuam and ICOADS buoy SSTs in January 2020. When drifting and tropical moored buoy SSTs from iQuam with SQUAM QC flag 5 are used as reference, the SD is about 0.20ºC on global average, which is consistent with the SD in SQUAM website. In contrast, when SSTs with SQUAM QC flag 1−5 are used as reference, the SD increases to 0.3º−0.4ºC, which is consistent with that in reference to drifting and moored-buoy SSTs from ICOADS with CLM QC analyzed in section 3.5.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

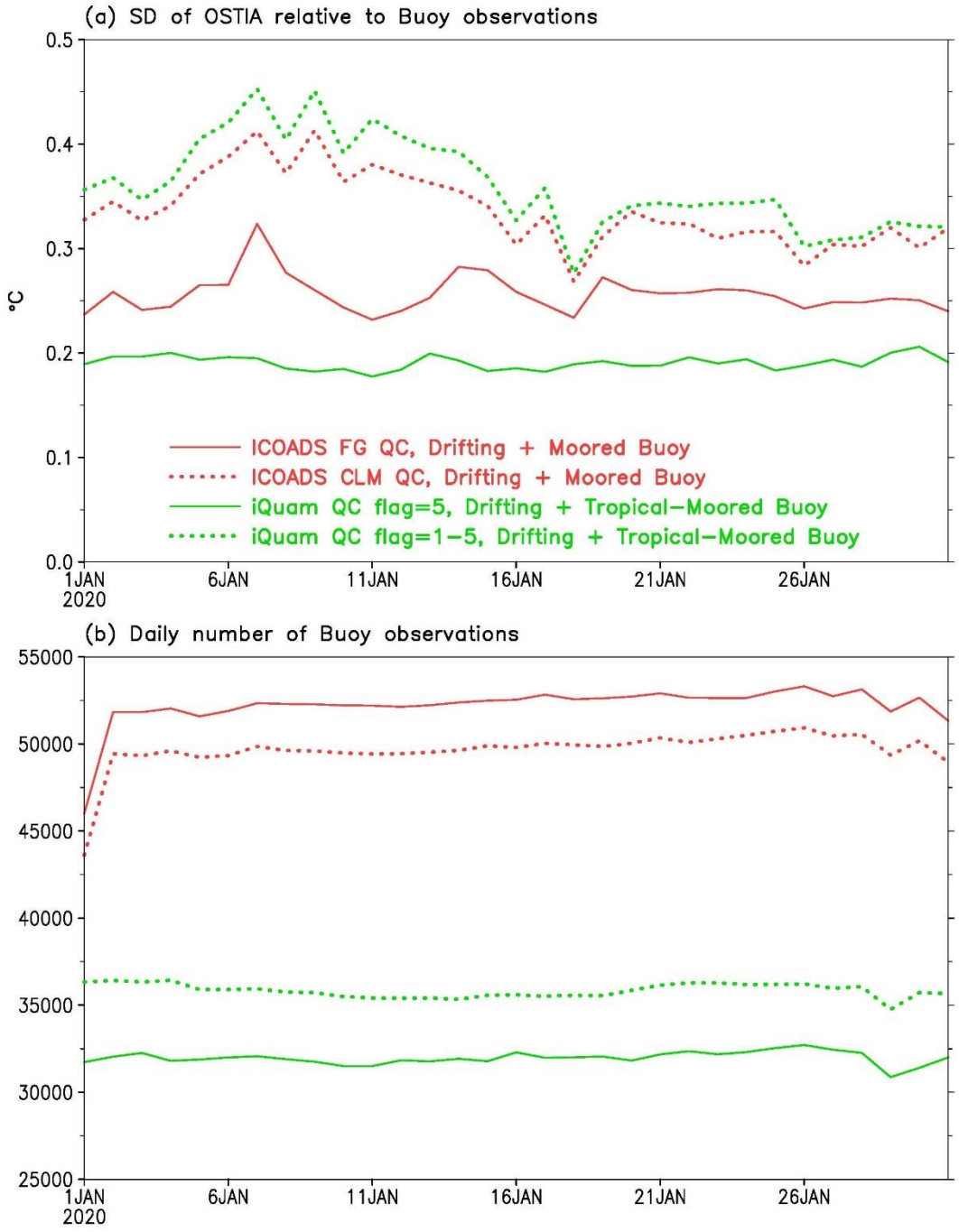

Figure 8. (a) SDs of OSTIA in January 2020 relative to the ICOADS drifting and moored buoy SSTs by FG QC (solid red) and CLM QC (dotted red), and relative to the iQuam drifting and tropical-moored buoy SSTs by SQUAM QC flag 5 (solid green) and flags 1−5 (dotted green); and (b) same as (a) except for daily number of observations.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

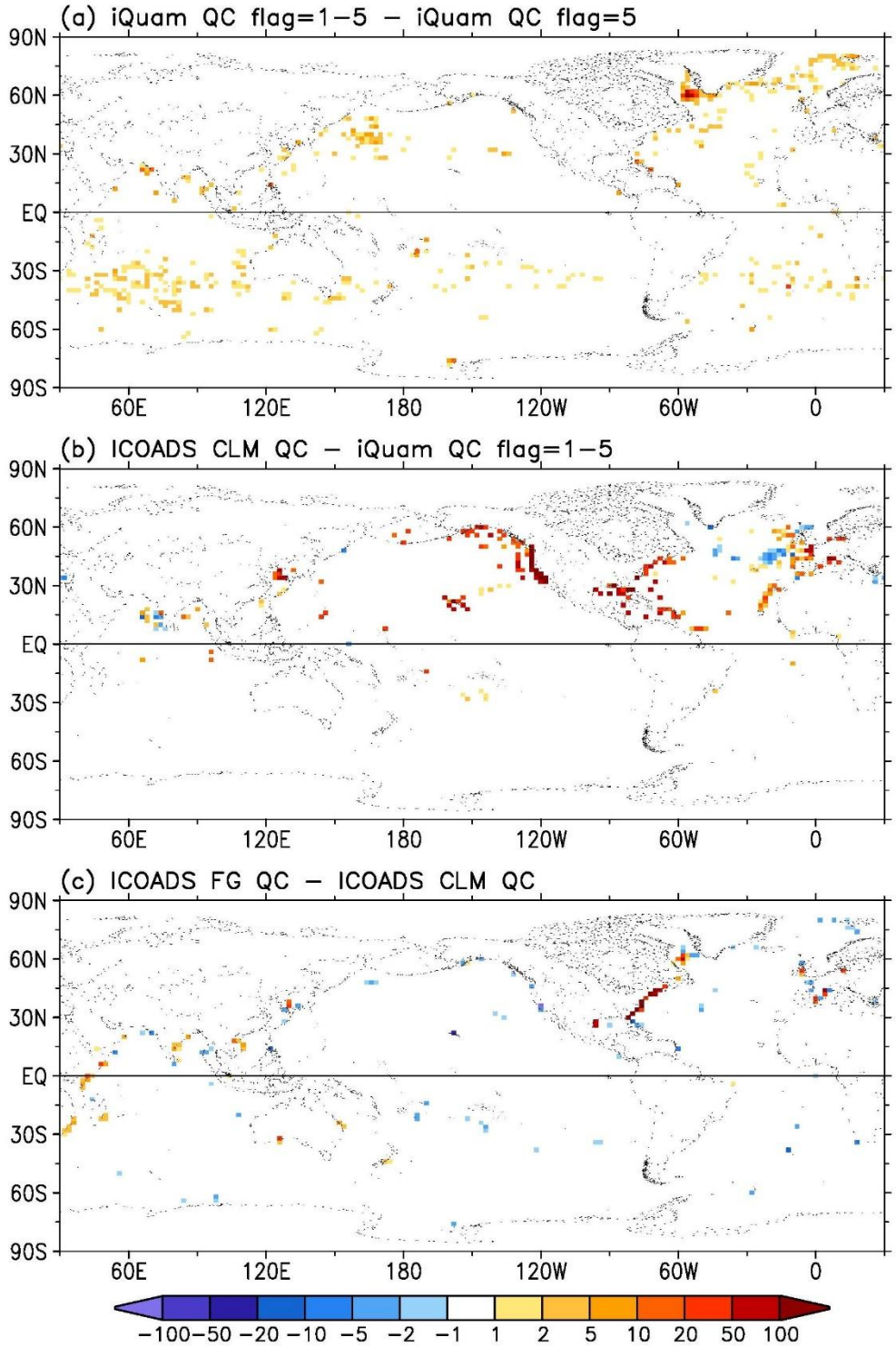

Figure 9. Differences of observation numbers of buoy SSTs between: (a) iQuam QC flag 1−5 and QC flag 5, (b) ICOADS CLM QC and iQuam QC flag 1−5, and (c) ICOADS FG QC and CLM QC. The differences are calculated in 2º×2º grids for the purpose of visualization.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

The increase in SD results from including more low-quality observations from iQuam, which increases from about 32000 to about 37000 per day over the global oceans (Fig. 8b). The increase in observations mostly happened in the northwest North Atlantic, northwest North Pacific, and Southern Hemisphere oceans between 30ºS and 50ºS (Fig. 9a). The inclusion of the observations in these regions results in a high SD on global average, because SST analysis is generally less reliable in those regions (refer to Figs. 2b and 2d).

In comparison with the drifting and tropical-moored buoy from iQuam, there are many more observations from drifting and moored buoy from ICOADS whether CLM or FG QC is applied (Fig. 8b). The higher number of observations results from the inclusion of the moored buoy in the subtropical regions, particularly along the coasts of North America (Fig. 9b). However, the global averaged SD of OSTIA analysis relative to buoy SSTs from ICOADS with CLM QC is close to that relative to buoy SSTs from iQuam with QC flag 1−5 (Fig. 8a). The reason may be that the high SDs along the eastern coasts of North America are compensated by the low SDs along the western coasts. Furthermore, observations from ICOADS increase when FG QC is applied in comparison to that when CLM QC is applied (Fig. 8b). However, the SD of OSTIA decreases from about 0.35ºC to 0.25ºC (Fig. 8a), because the FG QC guarantees that the selected observations are close to the analysis although more observations along the eastern coasts of North America are included (Fig. 9c).

4. Summary and discussions

The performance of nine gridded SST products has been assessed by comparing with an ensemble SST product (GMPE) and in situ observations from buoys and Argo floats during both day- and night-time using metrics of Bias, RMSD, mean difference, and standard deviation. Our analyses indicate that:

(i) Relative to GMPE, RMSDs are low in CMC, OSTIA and GPB, and Biases are low in DOISST and CMC. However, GMPE may not be a preferable reference, since its bias and RMSD relative to in situ Buoy and Argo observations are comparable with the bias and RMSD of individual SST products relative to the same Buoy and Argo observations.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

(ii) On GMPE (0.25º×0.25º) grids, RMSDs are low in DOISST and OSTIA relative to Buoy observations, and Biases are low in DOISST and MGDSST. Therefore, the good performance of DOISST and OSTIA is seen.

(iii) By interpolating to the in situ locations of Buoy observations, the SDs of the four SST products in their original high-resolution are lower than those in GMPE (0.25º) resolution, indicating an advantage of high-resolution in intercomparison. The SDs are low in DOISST and OSTIA. However, the SDs are low in OSTIA and CMC in products' original resolution, which may be associated with their high-resolution in analyses and small spatial scale in the bias correction to satellite SSTs. The DIFFs or Biases are generally low in DOISST, which may be associated with its large spatial scale in the bias correction focused on minimizing the mean bias.

(iv) The relative performance of the SST products remains unchanged when different QC criteria are applied to the reference Buoy and Argo observations, although the SDs are smaller when the tight first-guess QC rather than the loose climatological QC is applied.

(v) The performance of the SST products relative to Argo SSTs is overall similar to that relative to Buoy (combination of drifting and moored buoys) SSTs. However, the SDs and RMSDs are smaller than those relative to Buoy SSTs, which is likely associated with the longer observing interval of SSTs from Argo floats (10 days) than buoys (6 minutes to 1 hour), as well as the deeper depth of SSTs from Argo floats (above 5 m) than buoys (0.2−1.0 m). Similarly, the SDs relative to drifting-buoy SSTs are smaller than those relative Buoy SSTs due to that the observing interval is shorter in moored-buoys (6 minutes to 1 hour) than in drifting buoys (1 hour).

(vi) The magnitude of SDs depends on the selection of reference SSTs. When high-quality SSTs from iQuam with SQUAM QC flag 5 are used as reference, SDs decrease clearly because low-quality observations are excluded in reference in regions where SST analyses are less reliable.

Our study indicates that the performance assessment may depend on whether the gridded products are compared on the GMPE resolution or their original resolutions, whether in situ observations are regridded to the GMPE grids or on their in situ locations, whether comparisons are assessed against drifting- or moored-buoys or Argo floats, whether the bias correction applied to satellite SSTs is based on localized matchups or largescale SST patterns, and whether reference SSTs are independent from SST analyses. The clarification of these questions has helped

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

understand why the intercomparisons in Huang et al. (2021a,b) are different from those in Dash et al. (2012; https://www.star.nesdis.noaa.gov/socd/sst/squam/analysis/l4 and Fiedler et al. 2019): Huang et al. (2021a,b) compared SST analyses in the GMPE grids with in situ observations in the GMPE grids, while Dash et al. (2012) and Fiedler et al. (2019) compared SST analyses in their original grids with the observations in their in situ locations. Particularly, the reference buoy SSTs include the drifting and moored buoys from ICOADS in Huang et al. (2021a,b) but only include drifting and tropical-moored buoys from iQuam with SQUAM QC flag 5 in Dash et al. (2012). The former contains substantially more observations than the latter, which is mostly located in the regions where SST analyses are less reliable.

Our analysis indicates that the performance of DOISST v2.1 is among the best gridded SST products, which is certainly attributed in part to the use of Argo SSTs as demonstrated in Huang et al. (2021a). However, the assessment of DOISST in reference to Argo is no longer independent, which may force us to rethink a different way to assess the performance of SST products. One method is to preserve the 10% of Buoy and/or Argo observations for independent verification without ingesting the analysis system (Huang et al. 2021a, Reynolds et al. 2002). The selection of the 10% can be a one-time random draw or multi-time bootstraps with a data separation between ingest (90%) and validation (10%). Another method is to find other independent observations such as conductivity–temperature–depth (CTD), mechanical bathythermograph (MBT), expendable bathythermograph (XBT) profiles, saildrone measurements, thermosalinograph, and ice buoys (Huang et al. 2018, 2021b; Vazquez-Cuervo et al. 2022; Moteki 2022; Zhang et al. 2022). However, SSTs from CTD, MBT, XBT, ice buoy observations may also be biased due to surface contamination and the retrieval accuracy of measurement-depth (Huang et al. 2018; 2021b), and spatial coverages of the SSTs from saildrones, thermosalinograph, and ice buoy may be low.

Acknowledgements

Authors appreciate two anonymous reviewers' comments that have greatly improve the manuscript. JAC and LC were supported by the NOAA Climate Program Office (NA20OAR4310339).

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

Data availability statements

All gridded SST products are available online and listed in Table 1. The ICOADS data are available at https://www.ncei.noaa.gov/data/marine/nrt and https://www.ncei.noaa.gov/data/marine/nrt. The Argo data were provided by the Global Data Assembly Centre (GDAC; https://doi.org/10.17882/42182; http://www.seanoe.org/data/00311/42182). The iQuam data are available at https://www.star.nesdis.noaa.gov/socd/sst/iquam/data.html. The ACSPO L3S-LEO SST data are provided by NOAA STAR in experimental mode (10.5067/GHLPM-3SS28; 10.5067/GHLAM-3SS28). The access date for all data is February 15, 2022.

References

Argo, 2000: Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE, doi:10.17882/42182, accessed on 1 February 2022.

Beggs, H., L. Qi, P. Govekar, C. Griffin, 2020: Ingesting VIIRS SST into the Bureau of Meteorology's Operational SST Analyses, In: Proceedings of the 21st GHRSST Science Team Meeting, Virtual Meeting hosted by EUMETSAT, 1–4 June 2020. https://imos.org.au/fileadmin/user_upload/shared/SRS/SST/Beggs_GHRSST-XXI_Extended_Abstract_06Sep2020.pdf

Beggs, H., A. Zhong, G. Warren, O. Alves, G. Brassington, T. Pugh, 2011: RAMSSA - An operational, high-resolution, regional Australian multi-sensor sea surface temperature analysis over the Australian region. Australian Meteorological and Oceanographic Journal, 61, 1–22.

Brasnett, B., 1997: A global analysis of sea surface temperature for numerical weather prediction. J. Atmos. Oceanic Technol., 14, 925–937.

Brasnett B., 2008. The impact of satellite retrievals in a global sea-surface-temperature analysis. Q. J. R. Meteorol. Soc., 134, 1745-1760, DOI: 10.1002/qj.319.

Brasnett, B., D. S. Colan, 2016: Assimilating Retrievals of Sea Surface Temperature from VIIRS and AMSR2. J. Atmos. Oceanic Technol., 33, 361–375, DOI: 10.1175/JTECH-D-15-0093.1.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

Castro, S. L., G. A. Wick, W. J. Emery, 2012: Evaluation of the relative performance of sea surface temperature measurements from different types of drifting and moored buoys using satellite-derived reference products. J. Geophys. Res. Oceans, 117, C02029, doi:10.1029/2011JC007472.

Chin, T. M., J. Vazquez-Cuervo, E. M. Armstrong, 2017: A multi-scale high-resolution analysis of global sea surface temperature. Remote Sens. Environ. 200, 154–169.

Cole, J., 2000: Coastal sea surface temperature and coho salmon production off the north- west United States. Fisheries Oceanography, 9, 1-16.

Crewell, S., E. Ruprecht, and C. Simmer, 1992: Latent heat flux over the North Atlantic Ocean—A case study. Journal of Applied Meteorology and Climatology, 30, 1627-1635.

D'Agata, S., 2022: Ecosystems services at risk. Nat. Clim. Chang. 12, 13–14. https://doi.org/10.1038/s41558-021-01256-7.

Dash, P. and coauthors, 2012: Group for High Resolution Sea Surface Temperature (GHRSST) analysis fields inter-comparisons—Part 2: Near real time web-based level 4 SST Quality Monitor (L4-SQUAM). Deep Sea Research Part II: Topical Studies in Oceanography, Volumes 77–80, 31–43, https://doi.org/10.1016/j.dsr2.2012.04.002.

Dash, P., A. Ignatov, Y. Kihai, and J. Sapper, 2010: The SST Quality Monitor (SQUAM). J. Atmos. Oceanic Tech., 27, 1899−1917, DOI: 10.1175/2010JTECHO756.1.

Donlon, C.J., M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, W. Wimmer, 2012: The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. Remote Sens. Environ. 116, 140–158.

Fiedler, E. K., A. McLaren, V. Banzon, B. Brasnett, S. Ishizaki, J. Kennedy, N. Rayner, J. Roberts-Jones, G. Corlett, C. J. Merchant, C. Donlon, 2019: Intercomparison of long-term sea surface temperature analyses using the GHRSST Multi-Product Ensemble (GMPE) system. Remote Sensing of Environment, 222, 18-33, https://doi.org/10.1016/j.rse.2018.12.015.

Feudale, L., and J. Shukla, 2011: Influence of sea surface temperature on the European heat wave of 2003 summer. Part I: an observational study. Clim Dyn 36, 1691–1703, https://doi.org/10.1007/s00382-010-0788-0.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

Freeman, E. and coauthors, 2017: ICOADS release 3.0: A major update to the historical marine climate record. Int. J. Climatol., 37, 2211–2232, doi:10.1002/joc.4775.

Good, S., E. Fiedler, C. Mao, M. J. Martin, A. Maycock, R. Reid, J. Roberts-Jones, T. Searle, J. Waters, J. While, M. Worsfold, 2020: The Current Configuration of the OSTIA System for Operational Production of Foundation Sea Surface Temperature and Ice Concentration Analyses. Remote Sens., 12, 720, https://doi.org/10.3390/rs12040720.

Hartmann, D. L., 2015: Pacific sea surface temperature and the winter of 2014. Geophys. Res. Lett., 42, 1894–1902, doi:10.1002/2015GL063083.

He, J., and B. J. Soden, 2016: The impact of SST biases on projections of anthropogenic climate change: A greater role for atmosphere-only models?, Geophys. Res. Lett., 43, 7745–7750, doi:10.1002/2016GL069803.

Hobday, A. J., and coauthors, 2016: A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227–238. https://doi.org/10.1016/j.pocean.2015.12.014.

Huang, B., W. Angel, T. Boyer, L. Cheng, G. Chepurin, E. Freeman, C. Liu, and H.-M. Zhang, 2018: Evaluating SST analyses with independent ocean profile observations. J. Climate, 31, 5015-5030, doi:10.1175/jcli-d-17-0824.1.

Huang, B., P. W. Thorne, V. F. Banzon, T. Boyer, G. Chepurin, J. H. Lawrimore, M. J. Menne, T. M. Smith, R. S. Vose, H.-M. Zhang, 2017: Extended Reconstructed Sea Surface Temperature version 5 (ERSSTv5), Upgrades, validations, and intercomparisons. J. Climate, 30, 8179–8205, doi:10.1175/JCLI-D-16-0836.1.

Huang, B., C. Liu, V. Banzon, E. Freeman, G. Graham, B. Hankins, T. Smith, H.-M. Zhang, 2021a: Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1, Journal of Climate, 34, 2923-2939, DOI 10.1175/JCLI-D-20-0166.1.

Huang, B., C. Liu, E. Freeman, G., Graham, T. Smith, and H.-M. Zhang, 2021b: Assessment and Intercomparison of NOAA Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1. J. Climate, 34, 7421-7441, https://doi.org/10.1175/JCLI-D-21-0001.s1.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

Huang, B., Z. Wang, X. Yin, A. Arguez, G. Graham, C. Liu, T. Smith, H.-M. Zhang, 2021c: Prolonged Marine Heatwaves in the Arctic: 1982-2020. Geophys. Res. Lett., 48, e2021GL095590, https://doi.org/10.1029/2021GL095590.

Jonasson, O., I. Gladkova, A. Ignatov, Y. Kihai, 2020: Progress With Development of Global Gridded Super-Collated SST Products from Low Earth Orbiting Satellites (L3S-LEO) at NOAA. Proc. SPIE 11420, Ocean Sensing and Monitoring XII, 1142002, doi:10.1117/12.2551819.

Karl, T. R., A. Arguez, B. Huang, J. H. Lawrimore, J. R. McMahon, M. J. Menne, T. C. Peterson, R. S. Vose, and H.-M. Zhang, 2015: Possible artifacts of data biases in the recent global surface warming hiatus. Science, 348, 1469-1472, doi:10.1126/science.aaa5632.

Kurihara, Y., T. Sakurai, T. Kuragano, 2006: Global daily sea surface temperature analysis using data from satellite microwave radiometer, satellite infrared radiometer and in-situ observations (in Japanese). Wea. Service Bull., 73, 1–18.

Lima, F., D. Wethey, 2012: Three decades of high-resolution coastal sea surface temperatures reveal more than warming. Nat Commun., 3, 704, https://doi.org/10.1038/ncomms1713.

Liu, C., E. Freeman, E. C. Kent, D. I. Berry, S. J. Worley, S. R. Smith, B. Huang, H.-M. Zhang, T. Cram, Z. Ji, 2022: Blending BUFR and TAC Marine in Situ Data for ICOADS Near-Real-Time Release 3.0.2. JTECH, in review.

Martin, B. and coauthors, 2012: Group for High Resolution Sea Surface temperature (GHRSST) analysis fields inter-comparisons. Part 1: A GHRSST multi-product ensemble (GMPE). Deep Sea Research Part II: Topical Studies in Oceanography, Volumes 77–80, 21-30, https://doi.org/10.1016/j.dsr2.2012.04.013.

Maturi, E., A. Harris, J. Mittaz, J. Sapper, G. Wick, X. Zhu, P. Dash; P. Koner, 2017: A new high-resolution sea surface temperature blended analysis. Bull. Amer. Meteor. Soc., 98 (5), 1015–1026, https://doi.org/10.1175/BAMS-D-15-00002.1.

Merchant, C. J., O. Embury, J. Roberts-Jones, E. Fiedler, C.E. Bulgin, G.K. Corlett, S. Good, A. McLaren, N. Rayner, S. Morak-Bozzo, C. Donlon, 2014: Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI) starting 1981. Geoscience Data Journal, DOI: 10.1002/gdj3.20.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

Merchant, C. J. and coauthors, 2019: Satellite-based time-series of sea-surface temperature since 1981 for climate applications. Nature Scientific Data, 6:223, https://doi.org/10.1038/s41597-019-0236-x.

Miller, R. L., and J. DeCampo,1994: C Coast: A PC-based program for the analysis of coastal processes using NOAA Coast Watch data, Photogramm. Eng. Remote Sens., 60, 155-160.

Mohino, E., S. Janicot, and J. Bader, 2011: Sahel rainfall and decadal to multi-decadal sea surface temperature variability. Climate dynamics, 373, 419-440, https://doi.org/10.1007/s00382-010-0867-2.

Moteki, Q., 2022: Validation of satellite-based sea surface temperature products against in situ observations off the western coast of Sumatra. Sci Rep, 12.7, https://doi.org/10.1038/s41598-021-04156-0. https://doi.org/10.1038/s41598-021-04156-0

O'Carroll, A. G. and coauthors, 2019: Observational Needs of Sea Surface Temperature. Front. Mar. Sci. 6:420.doi: 10.3389/fmars.2019.00420.

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, M. G. Schlax, 2007: Daily high-resolution blended analyses for sea surface temperature. J. Climate, 20, 5473–5496, DOI:10.1175/2007JCLI1824.1.

Roemmich, D., J. Church, J. Gilson, D. Monsellesan, P. Sutton, S. Wijffels, 2015: Unabated planetary warming and its ocean structure since 2006. Nat. Climate Change, 5, 240–245, doi:10.1038/nclimate2513.

Shimada, S., T. Ohsawa, T. Kogaki, G. Steinfeld, and D. Heinemann, 2015: Effects of sea surface temperature accuracy on offshore wind resource assessment using a mesoscale model. Wind Energy, 18, 1839-1854.

Singels, A., and C. N. Bezuidenhout, 1999: The relationship between ENSO and rainfall and yield in the South African sugar industry. South African Journal of Plant and Soil, 16, 96-101, https://doi.org/10.1080/02571862.1999.10634854.

Stark, J. D., C. J. Donlon, M. J. Martin, M. E. McCulloch, 2007, OSTIA: An operational, high resolution, real time, global sea surface temperature analysis system., Oceans 07 IEEE Aberdeen, conference proceedings. Marine challenges: coastline to deep sea. Aberdeen, Scotland.IEEE.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

Sakurai, T., H. Kobayashi, and A. Yamane, 2019: Report from JMA for GHRSST-XX, https://oa.mg/work/10.5281/zenodo.5121197.

Sun, J., S. Wu, J. Ao, 2016: Role of the North Pacific sea surface temperature in the East Asian winter monsoon decadal variability. Clim Dyn, 46, 3793–3805, https://doi.org/10.1007/s00382-015-2805-9.

Thiébaux, J., E. Rogers, W. Wang, B. Katz, 2003: A new high-resolution blended real-time global sea surface temperature analysis. BAMS, 84, 645-656, DOI: https://doi.org/10.1175/BAMS-84-5-645.

Vazquez-Cuervo, J., S. L. Castro, M. Steele, C. Gentemann, J. Gomez-Valdes, W. Tang, 2022: Comparison of GHRSST SST Analysis in the Arctic Ocean and Alaskan Coastal Waters Using Saildrones. Remote Sens. 14, 692, https://doi.org/10.3390/rs14030692.

Vibhute, A., and coauthors 2020: Decadal variability of tropical Indian Ocean sea surface temperature and its impact on the Indian summer monsoon. Theor Appl Climatol 141, 551–566 (2020). https://doi.org/10.1007/s00704-020-03216-1

Xu, F., A. Ignatov, 2010: Evaluation of in situ SSTs for use in the calibration and validation of satellite retrievals, J. Geophys. Res. Oceans, 115, https://doi.org/10.1029/2010JC006129.

Yang, C., F. E. Leonelli, S. Marullo, V. Artale, H. Beggs, B. B. Nardelli, T. M. Chin, V. D. Toma, S. Good, B. Huang, C. J. Merchant, T. Sakurai11, R. Santoleri, J. Vazquez-Cuervo, H.-M. Zhang, A. Pisano, 2021: Sea Surface Temperature Intercomparison in the Framework of the Copernicus Climate Change Service (C3S). J. Climate, 34, 5257-5283, DOI: 10.1175/JCLI-D-20-0793.1.

Yates, D., R. W. Vervoort, B. Minasny, and A. McBratney, 2016: The history of using rainfall data to improve production in the grain industry in Australia—from Goyder to ENSO. Crop and Pasture Science, 67, 467-479, https://doi.org/10.1071/CP15053.

Zhang, C. and Coauthors, 2022: Evaluation of surface conditions from operational forecasts using in situ saildrone observations in the Pacific Arctic. Mon. Wea. Rev., DOI 10.1175/MWR-D-20-0379.1.

Zhang, L., 2016: The roles of external forcing and natural variability in global warming hiatuses. Clim Dyn 47, 3157–3169, https://doi.org/10.1007/s00382-016-3018-6.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.

Zhong, A., H. Beggs, 2008: Operational Implementation of Global Australian Multi-Sensor Sea Surface Temperature Analysis. Analysis and Prediction Operations Bulletin No. 77. Bureau of Meteorology, Australia, 2 October 2008. http://cawcr.gov.au/projects/SST/GAMSSA_BoM_Operational_Bulletin_77.pdf.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-22-0081.1.