Assessment and Intercomparison of NOAA Daily Optimum Interpolation Sea Surface Temperature (DOISST) version 2.1

Boyin Huang[1*], Chunying Liu[1, 2], Eric Freeman[1, 5], Garrett Graham[3], Tom Smith[4], Huai-Min Zhang[1]

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

[2] Riverside Technology, inc., Asheville, North Carolina

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

[4] NOAA Center for Satellite Applications and Research, College Park, Maryland

[5] Cooperative Institute for Satellite Earth System Studies, University of Maryland, Asheville, North Carolina

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

Submitted to J. Climate

Early Online Release: This preliminary version has been accepted for publication in *Journal of Climate*, may be fully cited, and has been assigned DOI 10.1175/JCLI-D-21-0001.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

Abstract

NOAA Daily Optimum Interpolation Sea Surface Temperature (DOISST) has recently been updated to v2.1 (January 2016–present). Its accuracy may impact the climate assessment, monitoring and prediction, and environment-related applications. Its performance, together with those of seven other well-known sea surface temperature (SST) products, is assessed by comparison with buoy and Argo observations in the global oceans on daily 0.25º×0.25º resolution from January 2016 to June 2020. These seven SST products are NASA MUR25, GHRSST GMPE, BoM GAMSSA, UKMO OSTIA, NOAA GPB, ESA CCI, and CMC.

Our assessments indicate that biases and root-mean-square-difference (RMSDs) in reference to all buoys and all Argo floats are low in DOISST. The bias in reference to the independent 10% of buoy SSTs remains low in DOISST, but the RMSD is slightly higher in DOISST than in OSTIA and CMC. The biases in reference to the independent 10% of Argo observations are low in CMC, DOISST, and GMPE; and RMSDs are low in GMPE and CMC. The biases are similar in GAMSSA, OSTIA, GPB, and CCI whether they are compared against all buoys, all Argo, or the 10% of buoy or 10% of Argo observations, while the RMSDs against Argo observations are slightly smaller than those against buoy observations. These features indicate a good performance of DOISST v2.1 among the eight products, which may benefit from ingesting the Argo observations by expanding global and regional spatial coverage of in situ observations for effective bias correction of satellite data.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-21-0001.1.

1. Introduction

The variation of globally averaged sea surface temperature (SST) is one of the most-used indicators of Earth's climate change due to the vast ocean surface area (IPCC 2013, 2019; EPA 2014; Karl et al. 2015; Fyfe et al. 2016). Climate variations over the global oceans are characterized by many SST modes such as the El Niño–Southern Oscillation (ENSO), the Pacific decadal variability (PDV), the Atlantic multidecadal oscillation (AMO), the tropical Atlantic SST mode, and the Indian Ocean dipole (IOD; Philander 1990; Latif and Barnett 1994; Schlesinger and Ramankutty 1994; Mehta 1998; Saji et al. 1999). A reliable SST product is critical to many applications in ocean data assimilation, atmospheric simulation, ocean prediction, climate monitoring and assessment, future climate projection, and calibration of satellite observations (Saravanan 1998; Czaja and Frankignoul 1999; Goddard and Mason 2002; Liu et al. 2006; Schubert et al. 2009; Ashfaq et al. 2011; Liang et al. 2019; Iizuka and Nakamura 2019; Dragaud et al. 2019; Aumann et al. 2020; Ciani et al. 2020).

The reliability of SST products strongly depends on the availability of SST observations, among other factors. In situ SST observations are available as early as 1772 in the International Comprehensive Ocean-Atmosphere Data Set (ICOADS; Freeman et al. 2017). At that time, SST observations were used by commercial sailing ships to locate the Gulf Stream (Franklin et al. 1768; Richardson 1980; Emery 2003). However, SST observations made before the 1950s had two shortcomings: (1) large systematic biases and (2) low spatial coverage over the global oceans (Folland and Parker 1995; Kennedy et al. 2011a, 2011b, 2019; Huang et al. 2015a). Since the 1950s, these shortcomings have been greatly reduced.

Since the 1850s, many in situ SST data products have been developed for climate and weather-related research and applications. Examples of well-known in situ SST products include

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

the National Oceanic and Atmospheric Administration (NOAA) Extended Reconstructed SST (ERSST) in monthly 2°×2° resolution starting from 1854 (Smith et al. 1996; Smith and Reynolds 2003, 2004; Huang et al. 2015a, 2017, 2020a); the UK Met Office Hadley SST (HadSST) in monthly 5°×5° resolution starting from 1850 (Kennedy et al. 2011a, 2011b, 2019); the Hadley Ice and SST (HadISST) in monthly 1º×1º resolution starting from 1870 (Rayner et al. 2003); and the Japan Meteorological Office Centennial Observation-Based Estimates of SSTs (COBE-SST) in daily 1°×1° resolution starting from 1850 (Ishii et al. 2005; Hirahara et al. 2014).

Since the early 1980s, satellite observations have been providing the possibility of global high-resolution SST products in daily 0.25° or finer resolutions. However, satellite-based SST observations may exhibit biases due to instrumental aging and/or contaminations of clouds and atmospheric aerosols (Zhang et al. 2004); biases are generally adjusted using in situ SSTs with various methods (Reynolds et al. 2007; Brasnett 2008; Merchant et al. 2014; Maturi et al. 2017; Good et al. 2020). Examples of well-known satellite-based SST products are the NOAA Daily Optimum Interpolation SST (DOISST) in 0.25° resolution starting from 1981 (Reynolds et al. 2007; Huang et al. 2021), the UK Met Office Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) in 0.05° resolution starting from 1981 (Stark et al. 2007; Donlon et al. 2012; Good et al. 2020), and the European Space Agency (ESA) Climate Change Initiative (CCI) SST in 0.05° resolution (Merchant et al. 2014, 2019). Among these products, CCI uses pure satellite-based observations without explicitly blending in situ observations, while other SST products homogenize the satellite and in situ observations and blend them together.

SST products were assessed commonly by intercomparisons against independent observations such as those from Argo floats or the ensemble median of SST products in regional and global oceans (Barton 2007; Iwasaki et al. 2008; Xie et al. 2008; Martin et al. 2012; Huang et

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

al. 2019; Fiedler et al. 2019; Woo and Park 2020; Yang et al. 2020). The intercomparison system (https://www.star.nesdis.noaa.gov/socd/sst/squam) noted a declined quality of DOISST v2.0 after 2016. DOISST has now been upgraded to v2.1 (Huang et al. 2021) to improve DOISST quality.

This study is to assess the quality of DOISST v2.1 after 2016, particularly its spatial and temporal structures of biases when compared to other similar available SST products. This is important as this topic has been discussed in several GHRSST meetings. Section 2, Data and Methods, describes the eight commonly used daily SST products, SST observations from buoys, Argo floats, and buoys specially designed for the Upper Temperature of the polar Oceans (UpTempO; Steele et al. 2017). Section 3, Intercomparisons, is an assessment of these eight products against buoy and Argo observations and an evaluation of those eight products against independent buoy, Argo, and UpTempO observations over the global oceans from January 2016 to June 2020. Section 4, Discussions, explores the reasons for the resulting differences between the SST products and observations, and the approach needed to provide reliable evaluation of SST products when all or almost all in situ data are ingested. The conclusions of the study are presented in Section 5.

2. Data and Methods

2.1 Data from eight SST products

(a) DOISST

The NOAA DOISST (Table 1) is a global daily product with a resolution of 0.25º starting from 1981 (Reynolds et al. 2007; Huang et al. 2021). DOIST blends in situ measurements and satellite-derived observations from the Advanced Very High Resolution Radiometer (AVHRR). The AVHRR SSTs are adjusted to the buoy SSTs at the nominal depth of 0.2 m (Reynolds et al.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

2007; Huang et al. 2013, 2015b). In ice-covered regions, the SST proxy from ice concentration is blended with SSTs from ships, buoys, and AVHRR, if available.

DOISST has been updated from v2.0 to v2.1 from January 2016 and onward, while data remain unchanged from 1981 to 2015. The updates include (Huang et al. 2021):

(a) Satellite NOAA-19 is replaced by MetOp-B; MetOp-A remains unchanged (MetOp-A and MetOp-B are European, polar-orbiting meteorological satellites);

(b) The SST proxy using the regression between ice concentration and SST is replaced by using the freezing-point temperature in ice-covered oceans (Banzon et al. 2020);

(c) The estimated ship SST bias is reduced from 0.14°C to 0.01°C. The biases of 0.14ºC and 0.01ºC are based on observations in periods 1982–2000 and 2016–2019, respectively;

(d) Ship and buoy observations from Daily ICOADS (ICOADS-D) Release 3.0.2 (R3.0.2; Liu et al. 2020) are used instead of NOAA National Centers for Environmental Prediction (NCEP) Global Telecommunications System (GTS) receipts; and

(e) Argo float observations (Argo 2000; Roemmich et al. 2001) above 5 m depth are included to ensure the best quality of SST production using all available in situ observations.

Note: Argo float observations are first used as independent data to validate the improvements in the updates from steps 1–4 and, in step 5, are included in DOISST in operational production.

To assess the quality of DOISST, additional experiments (DOISST_Buoy90 and DOISST_Argo90; Table 1) were designed following Reynolds et al. (2002). DOISST_Buoy90 is the same as DOISST except that the 90% of buoy drifters (Buoy90) is randomly selected and ingested and the remaining 10% of buoy drifters (Buoy10) is reserved for independent verification

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

and intercomparison. Similarly, DOISST_Argo90 is the same as DOISST except that the 90% of Argo drifters (Argo90) is randomly selected and ingested and the remaining 10% Argo floats (Argo10) is reserved for verification. It is assumed that the reserved Buoy10 and Argo10 drifters/floats are independent from the Buoy90 and Argo90 used in experiments DOISST_Buoy90 and DOISST_Argo90, respectively.

(b) MUR25

The NASA Multi-scale Ultra-high Resolution (MUR) v4.1 analysis is a daily SST product in 0.01° resolution starting from 2002 (Chin et al. 2017). MUR uses wavelets as basis functions in an optimal interpolation approach. MUR v4.1 includes 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, and in situ SST observations from the NOAA iQuam project (Xu and Ignatov 2010). iQuam SSTs include observations from ships, drifting and moored buoys, and Argo floats. Ship and buoy observations in iQuam are from ICOADS (Freeman et al. 2017) and the U.S. Global Ocean Data Assimilation Experiment/Fleet Numerical Meteorology and Oceanography Center (FNMOC). Biases in each satellite sensor are adjusted after the differences between the retrieved and in situ SSTs are assessed. MUR25 v4.1 data are available at https://podaac.jpl.nasa.gov/dataset/MUR-JPL-L4-GLOB-v4.1. For method comparisons, the coarser resolution (0.25°) version MUR v4.2 (MUR25; Table 1) is used in this study.

(c) GMPE

The Group High Resolution SST (GHRSST) Multi-Product Ensemble (GMPE; Table 1) data are a daily near-real-time SST product with a horizontal resolution of 0.25° in latitude and

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

longitude, starting from 2009. The GMPE selects the median SST from the GHRSST products (http://ghrsst-pp.metoffice.gov.uk/ostia-website/gmpe-monitoring.html; Martin et al. 2012; Dash et al. 2012; Fiedler et al. 2019). The current GHRSST products include CCI, OSTIA, Canada Meteorological Center (CMC) SST (Brasnett 1997, 2008; Brasnett and Colan 2016), NOAA DOISST, UK Met Office Hadley Centre Sea Ice and SST (HadISST; Rayner et al. 2003; Titchner and Rayner 2014), and Japan Meteorological Agency Merged Global Daily SST (MGDSST; Kurihara et al. 2006). Because selecting the median produces a less-biased SST product, GMPE has frequently been used as a reference to assess the performance of available SST products (Yang et al. 2020 and references therein). GMPE v2 (2016) and v3 (2017–2020) are used for method comparisons in this study. The use of DOISST in GMPE may result in GMPE's dependence on Argo floats and other in situ observations.

(d) GAMSSA

The Bureau of Meteorology (BoM) Global Australian Multi-Sensor SST Analysis v1 (GAMSSA v1; Table 1) is a daily data product produced by optimum interpolation in 0.25º resolution starting from 2008 (Zhong and Beggs 2008; Beggs et al. 2011, 2020). GAMSSA v1 uses SST data derived from AVHRR, the Advanced Along Track Scanning Radiometer (AATSR), the AMSR2, and in situ SST observations from ships as well as drifting and moored buoys from GTS. Biases in AVHRR and AMSR2 SSTs are adjusted using drifting buoy SSTs. The skin SST data derived from AATSR are first converted to the foundation SST (Donlon et al. 2002) and then merged with other SST data.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

(e) OSTIA

The UK Met Office OSTIA v2 (Table 1) produces global daily SST and ice concentration data using an optimum interpolation method in 0.05º resolution starting from 2006 (Stark et al. 2007; Donlon et al. 2012; Good et al. 2020). OSTIA v2 includes satellite SSTs derived from AVHRR, AMSR2, the Visible Infrared Imager Radiometer Suite (VIIRS), the Sea and Land Surface Temperature Radiometer (SLSTR), the Spinning Enhanced Visible and Infrared Imager (SEVIRI), and in situ SSTs from ships as well as drifting and moored buoys. The ship and buoy SSTs are from the World Meteorological Organization's (WMO) GTS. SSTs from drifting and moored buoys and VIIRS nighttime SSTs are used to adjust the biases in other satellite-derived SSTs using matchups within 25 km and 1 day.

(f) GPB

The NOAA Geo-Polar Blended v1 (GPB; Table 1) is a global daily SST product in 0.05º resolution starting from 2014 (Maturi et al. 2017). GPB v1 includes only nighttime SSTs derived from AVHRR, VIIRS, the Geostationary Operational Environmental Satellite (GOES) imager, the Japanese Advanced Meteorological Imager (JAMI), and in situ SSTs from ships and NOAA iQuam drifting and moored buoys (Xu and Ignatov 2010). The ship and buoy observations in iQuam are from ICOADS (Freeman et al. 2017) and the U.S. FNMOC. GBP v1 employs a rigorous, multiscale, optimum interpolation methodology and a data-adaptive correlation length scale to reduce noises. Biases in satellite-derived SSTs are first corrected by regressing them to in situ SSTs, then by the difference between satellite and GPB analysis of the previous day, and finally adjusted by an independent NCEP SST product of Thiébaux et al. (2003) to avoid long-term drift of GPB. It should be noted that the biases in satellite SSTs in Thiébaux et al. (2003) are

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

adjusted by the SST difference within a seven-day running window between satellite and in situ ship and buoy observations, which is similar to that applied in DOISST.

(g) CCI

The ESA CCI is a daily SST product in 0.05º resolution (Merchant et al. 2014; 2019). CCI applies a variational assimilation scheme to produce a gap-filled estimate of daily mean SST. CCI v2.0 is available from 1981 to 2019, and v2.1 is available from 1981 to 2016 (http://dap.ceda.ac.uk/neodc/esacci/sst/data/CDR_v2/Analysis/L4/v2.1). In this study, v2.0 data from 2016 to 2019 (Table 1) is used for comparison. The CCI SST provides the mean SST at 0.2 m depth, which is close to the nominal depth of drifting buoy measurements. The CCI SST includes both AVHRR (satellites NOAA-7, 9, 11-12, and 14-19) and Along-Track Scanning Radiometer (ATSR) series (ATSR, ATSR2, and AATSR). The biases in satellite observations are adjusted by re-calibrating radiances using a reference channel. Therefore, the CCI SST is not explicitly dependent on in situ observations (Merchant et al. 2014). However, Numerical Weather Prediction (NWP) fields from the European Centre for Medium-range Weather Forecasting (ECMWF) Re-Analysis Interim (ERA-Interim; Dee et al. 2011) are used as auxiliary information for cloud detection and retrieval, which may result in an implicit dependence of CCI SST on in situ observations.

(h) CMC

The CMC v3 (Table 1) is a daily SST in 0.1º resolution starting from 2016 to present (Brasnett 1997, 2008; Brasnett and Colan 2016). The early version CMC v2 from 1991 to 2017 is available at https://podaac.jpl.nasa.gov/dataset/CMC0.2deg-CMC-L4-GLOB-v2.0. CMC v3 merges AVHRR SSTs from satellites NOAA-18 and 19, METOP-A and B, AMSR-EOS, and in

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

situ SSTs from ships and drifting buoys of ICOADS. Biases in satellite observations are estimated from the differences between satellite and in situ pairs within an area of 5º in latitude and 10º in longitude, while the pairs are matched within 25 km. The median difference within the 5°×10º area is selected to adjust the biases in the satellite observations.

2.2 In situ data

(a) Buoy and Argo SSTs

Drifting and moored buoys at the ocean surface measure the SSTs at depth of 0.2–1.0 m (Castro et al. 2012). The temperature measurements of Argo floats above 5 m depth are used as SST observations in SST analyses and/or evaluations (Roemmich et al. 2015; Huang et al. 2017, 2021). Buoy SSTs are ingested into seven out of the eight products except for CCI, and Argo SSTs are used in DOISST v2.1 and MUR25 (Table 1). In this study, both buoy and Argo SSTs are first used to assess the eight SST products we examine. To further evaluate DOISST v2.1, the Buoy90 and Argo90 SSTs are included in experiments DOISST_Buoy90 and DOISST_Argo90, respectively, which are virtually the same as DOISST. The independent data Buoy10 and Argo10 SSTs are reserved to evaluate DOISST and assess the quality of all eight SST products. Buoy and Argo SSTs are first screened with quality control (QC) procedures that filter out the outliers deviated from the first-guess by more than four times standard deviation (STD) as described in Reynolds et al. (2007), and then compared against the eight SST products described in Section 2.1. The first-guess at the current date is the sum of analysis at the previous date and climatological difference between the present and previous dates. The buoy and Argo SSTs are processed into daily 0.25º×0.25º resolutions and compared against those eight SST products. It should be noted that buoy and Argo SSTs passed the same QC procedures whether they were ingested into the DOISST or used for evaluation purpose, which may give some trivial advantage in validating

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

DOISST over other products. It is assumed that the reserved Buoy10 and Argo10 drifters are independent from the Buoy90 and Argo90, respectively.

(b) Arctic buoy SSTs

The buoy SSTs from ICOADS in the Arctic region may be biased because (1) the SST thermistor sensors may be frozen, pushed up, and exposed to the air or (2) ICOADS provides SSTs from the measurements of the topmost thermistor (likely at 0.0 m depth) that may easily be frozen. To assess the SST products in the Arctic region, the SST data from the UpTempO project (Castro et al. 2016; Steele et al. 2017) are used in this study. The UpTempO collects SST measurements from specially designed buoys deployed in the Beaufort Sea and Hudson Bay from January 2016 to January 2019 (Fig. 1). To keep the SSTs from UpTempO observations independent from those of ICOADS and WMO GTS, UpTempO buoy measurements are searched from the second level (mostly at 2.5 m depth) down to 20 m depth and the first measurement from the thermistors submerged completely within water is selected as SST. The maximum depth of 20 m is selected because normal UpTempO observations show that the observed temperature above 20 m in the Arctic is almost uniform. UpTempO SSTs are averaged into daily superobservations on 0.25º×0.25º grids and then compared against the eight SST products described in Section 2.1.

2.3 Assessment methods

The eight SST products are assessed on daily 0.25º×0.25º grids. The products with higher spatial resolution were box-averaged to 0.25º×0.25º. The assessments are quantified by biases and root-mean-square-differences (RMSDs) against observed SSTs (Martin et al. 2012; Yang et al. 2021):

$$B(x,y) = \frac{1}{T}\sum_{t=1}^{T}(P(x,y,t) - O(x,y,t)) \quad (1)$$

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

$$R(x,y) = [\frac{1}{T}\sum_{t=1}^{T}(P(x,y,t) - O(x,y,t))^2]^{0.5} \quad (2)$$

$$b(t) = \frac{1}{W}\sum_{i=1}^{M}\sum_{j=1}^{N}(P(x_i,y_j,t) - O(x_i,y_j,t)) \times \cos(y_j) \quad (3)$$

where *B(x,y)* represents time averaged biases of product *P* relative to observation *O*; *R(x,y)* represents RMSDs between *P* and *O*; *b(t)* represents global average biases between *P* and *O*; *x, y,* and *t* represent longitude, latitude, and time, respectively; *T* represents the total number of time in days; and *W* represents the integrated weighting of $\cos(y_j)$.

The eight products are first assessed by comparing drifting and moored buoys that are dependent on the eight products except for CCI, and then by comparing Argo floats that are independent from most of the eight products except for DOISST and MUR25. To assess the quality of DOISST using independent observations, the 90% of the drifting buoy and Argo floats were ingested into the DOISST experiments while the other 10% were reserved for evaluation purposes as described in Section 2.1a. It is assumed that the residual bias of satellite SST and the analysis bias are larger than in situ SST bias, and therefore the biases and uncertainties of measurements and samplings in O are not taken into account in our assessments in equations (1)–(3).

Our assessments indicate that the spatial patterns of biases and RMSDs are quite similar when different types of observations are used in equation (1)–(3). The similarity of two spatial distribution *f(x,y)* and *g(x,y)* over the global oceans is quantified by a pattern correlation coefficient where *f* and *g* could be biases or RMSDs:

$$c = \frac{cov(f,g)}{\sqrt{var(f) \times var(g)}}, \quad (4)$$

where

$$cov(f,g) = \frac{1}{W}\sum_{i=1}^{M}\sum_{j=1}^{N}(f(x_i,y_j) - \bar{f}) \times (g(x_i,y_j) - \bar{g}) \times \cos(y_j) \quad (5)$$

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

$$var(f) = \frac{1}{W}\sum_{i=1}^{M}\sum_{j=1}^{N}(f(x_i, y_j) - \bar{f})^2 \times \cos(y_j) \quad (6)$$

$$var(g) = \frac{1}{W}\sum_{i=1}^{M}\sum_{j=1}^{N}(g(x_i, y_j) - \bar{g})^2 \times \cos(y_j) \quad (7)$$

Where $\bar{f}$ and $\bar{g}$ represent the global average of *f* and *g*, respectively.

A critical question for local and global averaged biases in equations (1) and (3) is whether these biases are statistically significant. The significance for the SST biases in the eight products is assessed as followed: First, the time series of the biases between analyses and observations at grid (x,y) is calculated, and then the time averaged bias (β) is expressed by

$$\beta(x, y) = \alpha(x, y) \pm u(x, y) \quad (8)$$

where α represents the time averaged bias and u represents its uncertainty. The uncertainty at the 95% confidence level is estimated according to (Walpole et al. 2012)

$$u(x, y) = 1.96\frac{STD(x,y)}{\sqrt{N_e}} \quad (9)$$

where STD represents the standard deviation of the time series of the biases, and N_e represents the equivalent sampling size of the time series that is estimated by (von Storch and Zwiers 1999)

$$N_e = T/[1 + 2\sum_{k=1}^{T-1}\left(1 - \frac{k}{T}\right)C(k)] \quad (10)$$

where $C(k)$ represents the lag-k autocorrelation coefficient of the time series and $C(k)>0.1$ is applied. The averaged bias difference α is statistically significant when $|\alpha| > u$. Similarly, uncertainties can be estimated for the time series of globally averaged biases in equation (3) or bias differences among products.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-21-0001.1.

3. Intercomparisons

3.1 Comparisons against buoy SSTs

The eight daily SST products are compared against buoy SSTs from January 2016 to June 2020. It should be noted the buoy SSTs have been ingested into and are therefore not independent from the eight SST products except for CCI. The SSTs from these eight products are first processed and box-averaged to 0.25º×0.25º resolution if the original resolution is higher than 0.25º (Table 1). The averaged SST differences (or biases) against buoy SSTs according to equation (1) are calculated on 0.25º×0.25º grids and displayed on 2º×2º grids for visualization purposes (Fig. 2). Figure 2 shows that SSTs are dominantly cold-biased in the global oceans in most of the eight products except for MUR25, which is warm-biased. The magnitude of these biases is mostly at 0.1º–0.2ºC, although biases at magnitude of 0.4ºC are found in the region of the Gulf Stream except for DOISST. SSTs are cold-biased in most of the tropical oceans between 20ºS and 20ºN except for MUR25, in most of the Northern Hemisphere mid-latitude (30º–60ºN) oceans except for DOISST and MUR25, in the Southern Ocean south of 45ºS in CCI, and in the South Pacific south of 45ºS in OSTIA and GPB. In the Gulf Stream region, strong warm biases are found except for DOISST, which shows weak warm biases and GAMSSA, which shows strong cold biases. Warm biases dominate over most of the global oceans in MUR25. Warm biases are also found around Australia and the Southern Ocean southeast of Argentina and south of South Africa in all eight products, although the magnitude of these warm biases is relatively small in the Southern Ocean in DOISST. These biases are mostly significant at the 95% confidence level according to equations (8)–(10), indicating a limited capability of these SST analyses in representing observations at local grid scale in those regions.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-21-0001.1.

The globally averaged biases range from −0.08ºC to +0.02ºC (Table 2). The bias in DOISST (-0.04ºC) is in the middle of the range, indicating a good performance of DOISST in the aspect of biases against buoy SSTs due to the recent revision from v2.0 to v2.1 (Huang et al. 2020). The warm bias in MUR25 is unique among the eight products, which might be associated with the unique use of microwave observations from MODIS.

The performance of the eight SST products is stable during the period from January 2016 to June 2020, which is illustrated by the time series of globally averaged biases (Fig. 3a). However, variations in biases are notable. For example, biases vary from −0.06º to 0.00ºC in DOISST, from −0.03º to 0.07ºC in MUR25, and from −0.13º to −0.02ºC in GAMSSA. The biases are large in the Northern Hemisphere summers (May–June–July) of 2017–2020 but smaller in the summer of 2016, which can be seen clearly from the evolution of GMPE (Fig. 3a, solid black). The stronger cold biases during the summers may result from the biases in satellite measurements due to higher cloudiness and dust aerosols in the tropical oceans (Zhang et al. 2004).

The transient variations of biases in Figure 3a can be quantified by their STDs. These STDs are about 0.02º–0.03ºC, which are much smaller (approximately 65% or less) than the mean biases (0.02º–0.07ºC) except for MUR25 (approximately 94%). The smaller STDs suggest that the errors in the eight SST products are mostly attributed to the mean or systematic biases rather than transient or random variability. This indicates that reduction of the satellite biases should be the focus in the future improvements of these eight products. Our tests (Table S1) show that the bias differences relative to DOISST and MUR25 according to equations (3), (8)–(10) are statistically significant at the 95% confidence level. In contrast, the bias differences may not be significant among GMPE, GAMSSA, GPB, CCI, and CMC, which is consistent with the timeseries shown in Figure 3a.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

Despite the varied spatial distributions of biases in the eight SST products shown in Figure 2, the spatial distributions of RMSDs according to equation (2) are rather similar (Fig. 4). RMSDs are less than 0.4ºC in most of the global oceans, particularly the tropical oceans between 20ºS and 20ºN. However, high RMSDs above 1ºC are found along the Gulf Stream, the Kuroshio and their extensions, and in the Southern Ocean southeast of Argentina, south of South Africa, and the sector of the Indian Ocean. The high RMSDs may directly result from mismatches between in situ and satellite observations in these regions. The RMSDs in DOISST are relatively small (< 0.4ºC) in most of the global oceans. The lower RMSDs in DOISST may indicate (1) the role of Argo in increasing SST quality in DOISST, (2) the role of the algorithm in correcting the biases of satellite SSTs in 2º×2º grids and 15-day data window as described later in Section 4.1, and (3) a potential overfitting to the buoy SSTs in DOISST. In contrast, the RMSDs in MUR25, which also ingests Argo SSTs, are higher. The higher RMSDs may result from (1) the quality-control (QC) procedures applied to Argo and buoy SSTs are the same in DOISST, which may differ from the QC procedures in MUR25, and (2) Argo SSTs in DOISST are defined as the temperatures within a 0–5 m depth, while temperatures closest to the surface were used in MUR25 (Xu and Ignatov 2016). The globally averaged RMSDs are 0.28º–0.41ºC (Table 2). The average RMSD in DOISST (0.28ºC) lies in the lower end of the range, indicating that performance of DOISST is good among the eight products.

It should be noted that the RMSDs described above does not include the biases and uncertainties of measurements and samplings in in situ SSTs. The magnitude of RMSDs may change when the biases and uncertainties of the referenced observations are considered, but their impact to the RMSDs would be the same for all products for a given reference.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

3.2 Comparisons against Argo SSTs

Argo SSTs are independent from most of the eight SST products except for DOISST and MUR25 (Table 1). The comparisons of the eight products against Argo SSTs from January 2016 to June 2020 (Fig. S1) show similar biases to those against buoy SSTs in Figure 2. Cold biases are found over most of the global oceans in the eight products except for MUR25, which is warm biased. The RMSDs (Fig. S2) remain higher in the regions of the Gulf Stream, Kuroshio and their extensions, and higher in the Southern Ocean southeast of Argentina, south of South Africa, and the Indian Ocean sector. The time series of the globally averaged biases (Fig. 3b) are overall similar to those in Figure 3a, but the cold biases become stronger in 2018–2020 than in 2016–2017 in DOISST, GMPE, OSTIA, and GPB, with reasons that are not immediately clear. The biases remain stable throughout the entire period of 2016–2020 in MUR25, GAMSSA, CCI, and CMC.

The globally averaged biases and RMSDs against Argo SSTs are overall consistent with those in comparison against buoy SSTs (Table 2), which also show an as good performance of DOISST among the eight products. The differences of averaged biases in reference to DOISST and MUR25 remain significant at the 95% confidence level (Table S2).

One may argue that the good performance of DOISST is associated with using the dependent Argo observations. However, our comparisons show that the spatial distributions of biases and RMSDs are very similar in GAMSSA, OSTIA, GPB, CCI, and CMC when they are compared against the dependent buoy SSTs (Figs. 2 and 4) and the independent Argo SSTs (Figs. S1 and S2). In contrast, the magnitude of biases and RMSDs decreases slightly, which is counterintuitive as one may expect an overall increase of biases and RMSDs against the independent Argo SSTs. The lower biases and RMSDs may suggest that they are largely

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-21-0001.1.

determined by the large-scale features such as non-local bias correction algorithms applied to the satellite observations, and are less determined by whether the reference is dependent.

3.3 Comparison against independent buoy and Argo SSTs

The performance of DOISST is further assessed using experiments DOISST_Buoy90 and DOISST_Argo90. The Buoy90 and Argo90 SSTs from January 2016 to June 2020 were ingested into experiments DOISST_Buoy90 and DOISST_Argo90 (Table 1), and the independent Buoy10 and Argo10 SSTs were reserved for evaluation purposes, respectively. Comparisons indicate that DOISST_Buoy90 and DOISST_Argo90 are almost identical to DOISST (not shown in figures), and therefore we will simply refer to DOISST_Buoy90 and DOISST_Argo90 as “DOISST” for the convenience of description.

The biases against Buoy10 SSTs (Fig. 5) show the similar spatial patterns to those against the full buoy SSTs in Figure 2. Their spatial pattern correlations according to equation (4) are larger than 0.85 in the eight products. The globally averaged biases range from −0.08ºC to 0.04ºC (Table 3). These biases remain close to those against the full buoy SSTs in Table 2, and therefore the differences of biases in reference to DOISST and MUR25 remain significant at 95% confidence level (Table S3).

In addition to the similarity of biases, the RMSDs against the Buoy10 SSTs (Fig. 6) also show a high similarity to those against the full buoy SSTs in Figure 4. Their spatial pattern correlation coefficients are greater than 0.90 in the eight products. The globally averaged RMSDs are 0.28º–0.35ºC (Table 3). These biases and RMSDs are slightly changed in comparison with those against the full buoy SSTs, probably due to the reduced sampling sizes that may not well represent the global oceans. The exception is that the RMSD in DOISST increases slightly from

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-21-0001.1.

0.28º to 0.31ºC when the independent Buoy10 SSTs are used for evaluation. Overall, although the buoy SSTs ingested into DOISST is reduced by 10%, the performance of DOISST remains good among the eight products.

The impact of the sampling size can be seen more clearly when DOISST ingests the Argo90 SSTs in experiment DOISST_Argo90 (Table 1), while the Argo10 SSTs are reserved for independent evaluation. In recent years, the typical number of Argo observations (approximately 1×10^3 per day or 280 surfacing Argo floats) is much less than that of buoy observations (approximately 5×10^4 per day or 1300 drifters) over the global oceans (Huang et al. 2019). Comparisons indicate that the similarity of spatial distributions of biases and RMSDs in reference to Argo10 and the full Argo (Figs. S1 and S3, S2 and S4) is low, 0.44–0.56 for biases and 0.62–0.76 for RMSDs according to equation (4).

The similarity of RMSDs in DOISST is relatively lower (0.62) due to the higher RMSD in the Southern Ocean when the independent Argo10 SSTs are reserved for evaluation (Figs. S2a and S4a). The higher RMSD in DOISST in the Southern Ocean may be due to the fact that the in situ observations are sparse and therefore DOISST is more sensitive to the reservation of the Argo10 SSTs. This is another persuasive reason why DOISST includes all available observations in the operational production to improve the product quality, in particular for data sparse regions. Overall, performance of DOISST in reference Argo10 is good among the eight products. The differences of biases in reference to DOISST and MUR25 remain significant except for GMPE (Table S4).

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

3.4 Comparison against independent UpTempO buoy SST in the Arctic region

The comparisons in Sections 3.1–3.3 do not include the Arctic Ocean because (1) the GMPE SST does not cover the Arctic and (2) buoy observations in ICOADS are from the topmost thermistor and may potentially be biased because the thermistor is exposed to the air by sea ice. Therefore, the independent SSTs from UpTempO project (Steele et al. 2017), collected from specially designed buoys released in the Beaufort Sea from January 2016 to January 2019 (Fig. 1), are used to assess the eight SST products in the Arctic region.

Comparisons of the eight SST products against UpTempO SSTs (Fig. 7) show that the biases are generally small (less than 0.5ºC) during the winter time (from November to May). SSTs in winter are mostly cold biased at the magnitude of −0.2ºC. The biases in DOISST are very small, which may largely be due to the application of freezing-point SST proxy (Banzon et al. 2020). However, biases in summer (from June to August) are as large as 3º–4ºC in all eight products, which may partly result from using nighttime satellite observations in MUR25 and GPB or using nighttime satellite observations to correct other satellites in OSTIA. Variations of biases are large in the eight products. For example, the magnitude of biases during the summer of 2017 is less than 0.2ºC in DOISST but reaches 3º–4ºC in OSTIA and GPB.

The small biases during the boreal winter are associated with the constraint of freezing point in these eight products when ice concentration is high. The large biases during the boreal summer result from large variations of SSTs when ice concentration is low (Banzon et al. 2020), which makes it difficult to constrain the SST proxy from ice concentration. The averaged biases range from −0.22ºC to +0.11ºC in the eight products, and the averaged RMSD are 0.42º–0.69ºC (Table 4). The performance of DOISST in the Arctic region is good among the eight products in

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

the aspect of bias, although its performance in RMSD is relatively worse. It should be noted that the bias and RMSD of GMPE may not be reliable due to its small sampling size.

4. Discussions

4.1 Causes for SST biases in DOISST

To track the source of biases in DOISST described in sections 3.1–3.3, the satellite SSTs (MetOp-A and B of daytime and nighttime) are compared against buoy SSTs from January 2016 to June 2020 (Figs. 8 a–d). The comparisons show that the satellite SSTs exhibit warm biases north of 45ºN and south of 40ºS. The warm biases are larger in MetOp-B (0.4ºC) than MetOp-A (0.2ºC) for both daytime and nighttime. The warm biases in nighttime MetOp-A extend more broadly in the Southern Hemisphere oceans. Between 40ºS and 45ºN, the satellite SSTs are cold biased. The cold biases are larger in MetOp-B (−0.6ºC) than MetOp-A (−0.2ºC) for both the daytime and nighttime. However, the cold biases in nighttime MetOp-A are confined in the northwest Indian Ocean and tropical Atlantic, and are very weak in the tropical Pacific. On a global average, the magnitude of biases is much smaller due to cancellations of cold and warm biases in different regions. The globally averaged biases range from −0.11ºC to +0.02ºC (Table 5).

These biases in AVHRR SSTs are adjusted by in situ observations from ships, buoys, and Argo floats in DOISST via the following four steps (Reynolds et al. 2007; Huang et al. 2015b; Huang et al. 2017):

(a) Daily AVHRR and in situ SSTs are separately bin-averaged to 2º×2º grids, which increases the area coverages of in situ observations over the global oceans.

(b) AVHRR and in situ SSTs are separately filtered by the decomposition of 130 Empirical Orthogonal Teleconnection modes (EOTs). The EOTs are the localized Empirical Orthogonal

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

Functions that are damped to zero 3000/5000 km away from the mode centers in the latitude/longitude direction. The EOT decomposition is critically dependent on the area coverage of data (Huang et al. 2020b, 2021). EOTs decomposition is used to filter out small spatial scale noises.

(c) Differences between the decomposed AVHRR and in situ SSTs are calculated within a 15-day running window, which is to filter out noises in a short-time period.

(d) The differences in step 3 are defined as AVHRR biases, which are interpolated back to DOISST 0.25º×0.25º grids and then subtracted from AVHRR SSTs.

The accuracy of the bias adjustments described above is evaluated by comparing the adjusted AVHRR SSTs against buoy SSTs (Figs. 8 e–h). It is clear that the warm biases decrease from 0.4º to 0.2ºC in MetOp-B for both the daytime and nighttime south of 40ºS and north of 45ºN, and in nighttime MetOp-A in the Southern Hemisphere oceans. The cold biases between 40ºS and 45ºN decrease from −0.6º to −0.2ºC in daytime MetOp-B and from −0.2º to −0.1ºC in nighttime MetOp-B and daytime MetOp-A. However, the warm biases in the North Pacific north of 40ºN increase slightly in MetOp-A for both the daytime and nighttime. The warm biases in the North Atlantic in nighttime MetOp-A are over-adjusted and become cold biases, indicating the limitation of bias correction algorithms in DOISST. The globally averaged biases range from −0.04ºC to −0.02ºC (Table 5). The improvement in globally averaged biases is clear in daytime MetOp-B but very slight in daytime MetOp-A. The globally averaged biases even become slightly larger in nighttime MetOp-A and B.

These remaining residual biases in the adjusted AVHRR SSTs have the following two characteristics: (1) the magnitude of the globally averaged biases is small (−0.02º to −0.04ºC;

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

Table 5), which matches with the final DOISST bias (−0.04ºC; Table 2); (2) the spatial distributions of the AVHRR biases in Figures 8 e–h are similar to that of DOISST in Figure 2a. The spatial correlation coefficients between Figures 8 e–h and 2a range from 0.50 to 0.58. These features suggest that the residual biases from the adjusted AVHRR SSTs are a major source contributing to the final DOISST biases. Therefore, the future development of DOISST should focus on removing these residual biases by improving the bias-adjustment algorithms.

The contribution of the residual biases of the adjusted AVHRR SSTs to the final DOISST biases can also be seen from the RMSDs against buoy SSTs (Fig. 9). Figures 9 a–d show that the RMSDs are large (0.6ºC) in MetOp-A and B for both daytime and nighttime in the regions of the Gulf Stream, the Kuroshio and their extensions, and the Southern Ocean south of 40ºS. The RMSDs are also large (from 0.4º to 0.6ºC) in daytime MetOp-A and B in the tropical oceans. The globally averaged RMSDs range 0.52ºC to 0.62ºC (Table 5).

The RMSDs of AVHRR SSTs mostly remain after the bias adjustment (Figs. 9 e–h), except that the RMSD of the adjusted AVHRR SST from daytime MetOp-B decreases substantially in the tropical and Southern Oceans. The spatial distributions of these RMSDs are very similar to that of the final DOISST shown in Figure 4a. The spatial correlation coefficients between Figures 9 e–h and 4a are 0.68–0.74, which indicate that the residual biases from the adjusted AVHRR SSTs contribute to the final DOISST biases. The globally averaged RMSDs of the adjusted AVHRR SSTs range from 0.54ºC to 0.58ºC, practically showing no improvements over those of the original AVHRRs (Table 5). Nevertheless, these RMSDs are much higher than that of final DOISST (0.28ºC; Table 2). The large contrast between the RMSDs in the adjusted AVHRR SSTs and final DOISST indicates that the noises in the adjusted AVHRR SSTs have been damped by in situ SSTs, which is another reason to include all available in situ SSTs such as Argo SSTs to improve the

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-21-0001.1.

quality of SST products. Although these residual biases of satellite observations may also exist in other SST products, we were unable to assess this as these intermediate data are generally unavailable to the public.

Our analyses indicate that the spatial patterns and magnitude of biases and RMSD do not change much when these products are compared with buoy and Argo, or the 10% of reserved buoys or Argo floats. The difference of biases and RMSD among products are clearly seen. These results suggest that the biases in these products may directly be associated with the algorithms correcting the biases of satellite SSTs as indicated by our earlier studies (Huang et al. 2013, 2015b, 2016).

4.2 Independent observations

One of the challenges in assessing the performance of the eight SST products is the availability of independent in situ observations. It should be noted that the in situ observations were neither perfect in quality nor always consistent among different platforms. The observations must be checked by QC procedures regardless of whether they are to be ingested into or to validate the products. However, the QC procedures may differ among products and impact the number and area coverage of the observations, and their roles may differ among products.

On the one hand, we want to reserve independent observations for evaluations. For example, Argo observations have been reserved to independently evaluate SST productions in GAMSSA, OSTIA, GPB, CCI, and CMC. On the other hand, we want to use as many observations as possible to increase the reliability of SST products. For example, Argo observations are ingested into DOISST and MUR25 to best represent the SST analyses. However, the spatial distributions of biases and RMSDs against Argo observations are similar to those against buoy observations. The magnitude of biases against Argo is similar to that against buoy, while the magnitude of

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

RMSD decreases slightly. These features are exhibited not only in GMPE, DOISST, and MUR25 where Argo observations are dependent, but also in GAMSSA, OSTIA, GPB, CCI, and CMC where Argo observations are independent. In other words, the smaller biases and RMSD in DOISST, MUR25, and GMPE may not necessarily result from comparing against dependent buoy and Argo SSTs.

The similar biases and RMSDs in GAMSSA, OSTIA, GPB, CCI, and CMC suggest that it is not necessary to reserve Argo observations purely for evaluation purposes. Inclusion of all high-quality in situ data (including Argo SSTs) is important to increase the quality of the SST products that utilize both in situ and satellite observations. The addition of Argo sampling can improve the coverage of in situ SSTs in some regions, which is important for satellite bias adjustments.

Our studies indicated that the DOISST biases and RMSDs mainly result from algorithms used for bias adjustment of satellite observations, while impacts are small from methods of blending in situ and satellite observations and from methods of interpolations (Huang et al. 2013, 2015b, 2016). Consistent with previous studies, our study indicates that the residual biases in the adjusted satellite-derived SSTs are the main contributor to the final biases in DOISST, which may also be true for other SST products. The residual biases are caused by imperfect matchups in most products or large-scale differences between in situ and satellite observations in DOISST.

Our analyses show that the residual biases are critically dependent on the coverages of in situ super-posed observations (superobservations) (Fig. 10), which is attributed to the bias correction algorithms using EOTs. The coverages of in situ data (Fig. 10) are defined as a ratio between the counts of days with superobservations and days with or without superobservations from June 1, 2016 to June 31, 2020. The coverages are calculated on 2º×2º grids, since biases of satellite observations are estimated on 2º×2º grids. Figure 10 f shows the difference between the

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

coverages of blended ship+buoy+Argo and ship+buoy observations. The coverage difference highlights the role of Argo observations when they are ingested into SST analysis systems. The figure indicates an increase of coverage by 0.2–0.3 in the Southern Ocean, which is about 100% of ship+buoy coverage (Fig. 10 d). Therefore, we can speculate that the matchups and therefore the overall performance would be notably improved in GAMSSA, OSTIA, GPB, and CMC if Argo observations were ingested, particularly in the Southern Ocean.

In DOISST, biases in satellite SSTs are estimated by the large-scale difference between in situ and satellite observations on 2º×2º grids within the 15-day data window (Reynolds et al. 2007). Large-scale patterns of in situ and satellite SSTs are based on EOTs, which is sensitive to the coverage of in situ observation (Huang et al. 2019). The coverage of buoy SSTs is low in the Southern Ocean (Fig. 10 b), and the total coverage of in situ SSTs (Figs. 10 d–e) is sensitive to the addition of Argo observations. As a result, the estimation of biases in satellite SSTs and therefore the residual biases in the adjusted satellite SSTs are sensitive to the Argo10 SSTs. This may explain why the final biases and RMSDs become larger in the Southern Ocean when the Argo10 SSTs are reserved as evaluation data.

We want to note that the results presented in this study may differ from previous studies due to factors such as using different time periods, validation metric, and validation datasets in assessments. Martin et al. (2012) showed that DOISST has a smaller mean bias but its standard deviation is large, which is consistent with our assessment. Fiedler et al (2019) showed a large bias in DOISST v2.0, which is consistent with Huang et al. (2021); and a smaller bias in CMC, CCI, and GMPE, which is different from our assessment. Yang et al. (2021) showed an overall good performance of CCI and OSTIA and an intermediate performance of DOISST v2.1 and MUR25, which is different from our assessment.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

5. Conclusions

Our assessments of the eight SST products indicate that DOISST v2.1 has a good performance in global-averaged biases and RMSDs in reference to buoy and Argo observations, as well as in reference to the independent Buoy10 and Argo10 SSTs. MUR25 has warm biases, while other seven products have cold biases. The differences of biases in reference to DOISST and MUR25 are statistically significant, while the differences of biases among GMPE, GAMSSA, OSTIA, GPB, CCI, and CMC are less significant. Our comparisons indicate that the quality of SST products may be improved if all in situ observations are included. This is consistent with developments of DOISST from v2.0 to v2.1, as after the inclusion of Argo data the differences relative to Argo data have decreased, and the fitting to regional structures of in situ data resulted in the higher similarity of DOISST spatial structures to those of in situ data. For methodology and product development researches, one may resort to reserving some observations (such as Argo floats) as independent evaluation sets. However, for products that depends on in-situ observation for satellite SST bias corrections (such as in DOISST), the operational production should utilize all good quality data to provide best quality product to users, in particular as there are still data sparse regions as of today (e.g. the Southern Ocean Region). This is akin to the manufacturing industry - manufactures use the best available materials to produce best quality products for customers, not just reserving the best materials for product evaluation purpose (that was done at an earlier experimental stage).

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

Acknowledgments

Authors thank four anonymous reviewers for their constructive comments that greatly improved the manuscript. Authors appreciate comments from Anthony Arguez and Xuepeng Zhao and English proof from Andrea Anderson in improving the manuscript. The sources of the eight SST products are listed in Table 1 and cited in the main text. ICOADS-D R3.0.2 data are provided at NOAA/NCEI (https://doi.org/10.7289/V5CZ3562), access date: August 1, 2020. MetOp-A, and MetOp-B SST were provided by NOAA/NCEI Common Ingest system (ftp://data-internal.ncei.noaa.gov with access permission), access date: August 2020. Argo data were provided by the Global Data Assembly Centre (GDAC; https://doi.org/10.17882/42182; http://www.seanoe.org/data/00311/42182), access date: August 1, 2020. NCEP sea ice concentrations were retrieved from ftp://ftpprd.ncep.noaa.gov/data/nccf/com/omb/prod and ftp://polar.ncep.noaa.gov/cdas/archive, access date October 2019.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

References

Argo, 2000: Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE, accessed 1 August 2020, doi:10.17882/42182.

Ashfaq, M., C. B. Skinner, N. S. Diffenbaugh, 2011: Influence of SST biases on future climate change projections. Clim Dyn 36, 1303–1319, https://doi.org/10.1007/s00382-010-0875-2.

Aumann, H. H., S. E. Broberg, E. M. Manning, T. S. Pagano, R. C. Wilson, 2020: Evaluating the absolute calibration accuracy and stability of AIRS using the CMC SST. Remote Sens. 2020, 12, 2743, https://doi.org/10.3390/rs12172743.

Banzon, V., T. M. Smith, M. Steele, B. Huang, H.-M. Zhang, 2020: Improved Estimation of Proxy Sea Surface Temperature in the Arctic. J. Atmos. Oceanic Technol., doi: 10.1175/JTECH-D-19-0177.1.

Barton, I. J., 2007: Comparison of in situ and satellite-derived sea surface temperatures in the Gulf of Carpentaria. J. Atmos. Oceanic Technol., 24, 1773–1784, https://doi.org/10.1175/JTECH2084.1.

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://www.ghrsst.org/meetings/21st-ghrsst-international-science-team-meeting-g-xxi (in press).

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.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

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.

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.

Castro, S. L., G. A. Wick, M. Steele, 2016: Validation of satellite sea surface temperature analyses in the Beaufort Sea using UpTempO buoys. Remote Sensing of Environment, 187, 458–475, https://doi.org/10.1016/j.rse.2016.10.035.

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.

Ciani, D., M.-H. Rio, B.B. Nardelli, H. Etienne, T. Santoleri, 2020: Improving the altimeter-derived surface currents using sea surface temperature (SST) data: A sensitivity study to SST products. Remote Sens., 12, 1601, https://doi.org/10.3390/rs12101601.

Czaja, A., C. Frankignoul, 1999: Influence of the North Atlantic SST on the atmospheric circulation. Geophysical Research Letters, 26, 2969-2972, https://doi.org/10.1029/1999GL900613.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

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.

Dee, D. P. and coauthor, 2011: The ERA‐ Interim reanalysis: configuration and performance of the data assimilation system. Q. J. Roy. Meteorol. Soc., 137, 553–597.

Donlon, C.J., P. Minnett, C. Gentemann, T. J. Nightingale, I. J. Barton, B. Ward, J. Murray, 2002: Towards improved validation of satellite sea surface skin temperature measurements for Climate Research. J. Climate, 15, 353–69.

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.

Dragaud, I. C. D. V. and coauthors: 2019: The impact of SST on the wind and air temperature simulations: a case study for the coastal region of the Rio de Janeiro state. Meteorol. Atmos. Phys., 131, 1083–1097, https://doi.org/10.1007/s00703-018-0622-5.

Emery, W. J., 2003. AIR-SEA INTERACTION | Sea Surface Temperature. In R. H. James, ed. Encyclopedia of Atmospheric Sciences. Oxford: Academic Press, pp. 100–109, DOI: 10.1016/B0-12-227090-8/00065-8.

EPA, 2014: Climate change indicators in the United States, 2014. 3rd ed. EPA Rep. EPA 430-R-14-004, 112 pp. [Available online at https://www.epa.gov/sites/production/files/2016-07/documents/climateindicators-full-2014.pdf.]

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

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.

Folland, C. K., D. E. Parker, 1995: Correction of instrumental biases in historical sea surface temperature data. Quart. J. Roy. Meteor. Soc., 121, 319–367, doi:10.1002/qj.49712152206.

Franklin, B., T. Folger, E. Wright, E. Halley, H. Moll, J. Mount, T. Page, 1768: Franklin-Folger chart of the Gulf Stream. [London: Sold by Jno. Mount and Tho. Page] [Map] Retrieved from the Library of Congress, https://www.loc.gov/item/88696412/.

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.

Fyfe, J. C. and coauthors, 2016: Making sense of the early-2000s warming slowdown, Nat. Clim. Change, 6, 224–228, doi:10.1038/nclimate2938.

Goddard, L., S. Mason, 2002: Sensitivity of seasonal climate forecasts to persisted SST anomalies. Climate Dynamics, 19, 619–631, DOI 10.1007/s00382-002-0251-y.

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.

Hirahara, S., M. Ishii, and Y. Fukuda, 2014: Centennial-scale sea surface temperature analysis and its uncertainty. J. Climate, 27, 57–75, doi:10.1175/JCLI-D-12-00837.1.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

Huang, B., V. F. Banzon, E. Freeman, J. Lawrimore, W. Liu, T. C. Peterson, T. M. Smith, P. W. Thorne, S. D. Woodruff, H.-M. Zhang, 2015a: Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4), Part I. Upgrades and intercomparisons. J. Climate, 28, 911-930, doi:10.1175/JCLI-D-14-00006.1.

Huang, B., M. L'Heureux, J. Lawrimore, C. Liu, V. Banzon, Z.-Z. Hu, A. Kumar, 2013: Why did large differences arise in the sea-surface temperature datasets across the tropical Pacific during 2012? J. Atmos. Oceanic Technol., 30, 2944–2953, DOI:10.1175/ JTECH-D- 13-00034.1.

Huang, B., M. L'Heureux, Z.-Z. Hu, X. Yin, H.-M. Zhang, 2020b: How significant was the 1877–78 El Nino? J. Climate, 33, 4853-4869, doi: 10.1175/JCLI-D-19-0650.1.

Huang, B., C. Liu, V. Banzon, E. Freeman, G. Graham, B. Hankins, T. Smith, H.-M. Zhang, 2021: 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, V. F. Banzon, H.-M. Zhang, T. R. Karl, J. H. Lawrimore, R. S. Vose, 2016: Assessing the impact of satellite-based observations in sea surface temperature trends. Geophys. Res. Lett., 43, DOI:10.1002/2016GL068757.

Huang, B., C. Liu, G. Ren, H.-M. Zhang, L. Zhang, 2019: The role of buoy and Argo observations in two SST analyses in the global and tropical Pacific oceans. J. Climate, 32, 2517–2535, doi:10.1175/JCLI-D-18-0368.1.

Huang, B., M. J. Menne, T. Boyer, E. Freeman, B. E. Gleason, J. H. Lawrimore, C. Liu, J. J. Rennie, C. Schreck, F. Sun, R. Vose, C. N. Williams, X. Yin, H.-M. Zhang, 2020a: Uncertainty estimates for sea surface temperature and land surface air temperature in NOAAGlobalTemp version 5. J. Climate, 33, 1351–1379, DOI: 10.1175/JCLI-D-19-0395.1.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

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., W. Wang, C. Liu, V. F. Banzon, J. Lawrimore, H.-M. Zhang, 2015b: Bias adjustment of AVHRR SST and its impacts on two SST analyses. J. Atmos. Oceanic Technol., 32, 372-387, DOI:10.1175/JTECH-D-14-00121.1.

Iizuka, S. and H. Nakamura, 2019: Sensitivity of mid-latitude heavy precipitation to SST: A case study in the area of Japan area on 9 August 2013. Journal of Geophysical Research: Atmospheres, 124, 4365–4381, https://doi.org/10.1029/2018JD029503.

IPCC, 2013: Climate Change 2013: The Physical Science Basis. T. F. Stocker et al., Eds., Cambridge University Press, 1535 pp., https://doi.org/10.1017/CBO9781107415324.

IPCC, 2019: Technical summary. In: Pörtner H-O, Roberts DC, Masson-Delmotte V, Zhai P, Tignor M, Poloczanska E, Mintenbeck K, Alegría A, Nicolai M, Okem A, Petzold J, Rama B, Weyer NM (eds): IPCC special report on the ocean and cryosphere in a changing climate (In press).

Ishii, M., A. Shouji, S. Sugimoto, T. Matsumoto, 2005: Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the Kobe Collection. Int. J. Climatol., 25, 865–879, doi:10.1002/joc.1169.

Iwasaki, S., M. Kubota, H. Tomita, 2008: Inter-comparison and evaluation of global sea surface temperature products, International Journal of Remote Sensing, 29, 21, 6263-6280, DOI: 10.1080/01431160802175363.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

Karl, T. R., A. Arguez, B. Huang, J. H. Lawrimore, J. R. McMahon, M. J. Menne, T. C. Peterson, R. S. Vose, 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.

Latif, M., T. P. Barnett, 1994: Causes of Decadal Climate Variability over the North Pacific and North America. Science, 266, 634-637, DOI: 10.1126/science.266.5185.634.

Liang, X., M. Losch, L. Nerger, L. Mu, Q. Yang, C. Liu, 2019: Using sea surface temperature observations to constrain upper ocean properties in an Arctic sea ice- ocean data assimilation system. Journal of Geophysical Research: Oceans, 124, 4727–4743, https://doi.org/10.1029/2019JC015073.

Liu, C. et al. (2020): Blending BUFR and TAC Marine in Situ Data for ICOADS Near-Real-Time Release 3.0.2. [Unpublished raw data].

Liu, Q., N. Wen, Z. Liu, 2006: An observational study of the impact of the North Pacific SST on the atmosphere. Geophys. Res. Lett., 33, L18611, DOI:10.1029/2006GL026082.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011a: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling errors. J. Geophys. Res., 116, D14103, doi:10.1029/2010JD015218

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, M. Saunby, 2011b: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. J. Geophys. Res., 116, D14104, doi:10.1029/2010JD015220.

Kennedy, J. J., N. A. Rayner, C. P. Atkinson, and R. E. Killick, 2019: An ensemble data set of sea surface temperature change from 1850: The Met Office Hadley Centre HadSST.4.0.0.0 data set. J. Geophys. Res., 124, 7719–7763, https://doi.org/10.1029/2018JD029867.

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.

Mehta, V. M., 1998: Variability of the Tropical Ocean Surface Temperatures at Decadal–Multidecadal Timescales. Part I: The Atlantic Ocean. J. Climate, 11, 2351–2375, https://doi.org/10.1175/1520-0442(1998)011<2351:VOTTOS>2.0.CO;2.

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 Climate*. DOI 10.1175/JCLI-D-21-0001.1.

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.

Philander, S. G., 1990: El Niño, La Niña, and the Southern Oscillation. Academic Press, 293 pp.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J. Geophys. Res., 108, 4407, doi:10.1029/2002JD002670.

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.

Richardson P.L., 1980: The Benjamin Franklin and Timothy Folger Charts of the Gulf Stream. In: Sears M., Merriman D. (eds) Oceanography: The Past. Springer, New York, NY, https://doi.org/10.1007/978-1-4613-8090-0_64.

Roemmich, D. and coauthors, 2001: The global array of profiling floats. Observing the Ocean in the 21st Century, C. J. Koblinsky and N. R. Smith, Eds., Australian Bureau of Meteorology, 248–258.

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.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

Saji, N. H., B. N. Goswami, P. N. Vinayachandran, T. Yamagata, 1999: A dipole mode in the tropical Indian Ocean. Nature, 401, 360–363.

Saravanan, R., 1998: Atmospheric low-frequency variability and its relationship to mid-latitude SST variability: Studies using the NCAR climate system model. J. Climate, 11, 1386–1404, https://doi.org/10.1175/1520-0442(1998)011<1386:ALFVAI>2.0.CO;2.

Schlesinger, M. E., and N. Ramankutty, 1994: An oscillation in the global climate system of period 65–70 yr. Nature, 367, 723–726, https://doi.org/10.1038/367723a0.

Schubert, S. and coauthors, 2009: A US CLIVAR project to assess and compare the responses of global climate models to drought-related SST forcing patterns: Overview and results. Journal of Climate, 22, 5251-5272.

Smith, T., R.W. Reynolds, 2003: Extended reconstruction of global sea surface temperature based on COADS data (1854–1997). J. Climate, 16, 1495–1510, doi:10.1175/1520-0442-16.10.1495.

Smith, T., R.W. Reynolds, 2004: Improved extended reconstruction of SST (1854–1997). J. Climate, 17, 2466–2477, doi:10.1175/1520-0442(2004)017,2466:IEROS.2.0.CO;2.

Smith, T., R.W. Reynolds, R. E. Livezey, D. C. Stokes, 1996: Reconstruction of historical sea surface temperatures using empirical orthogonal functions. J. Climate, 9, 1403–1420, doi:10.1175/1520-0442(1996)009,1403:ROHSST.2.0.CO;2.

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 Climate*. DOI 10.1175/JCLI-D-21-0001.1.

Steele, M., W. Ermold, I. Rigor, 2017: UpTempO buoys deployed in the Arctic Ocean in 2017. Arctic Data Center, accessed 1 August 2020, https://doi.org/10.18739/A2GB1XG6P.

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.

Titchner, H. A., N. A. Rayner, 2014: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. J. Geophys. Res. Atmos., 119, 2864–2889, doi:10.1002/2013JD020316.

von Storch, H., and F. W. Zwiers, 1999: Statistical Analysis in Climate Research. Cambridge University Press, 484 pp.

Walpole, R. E., R. H. Myers, and S. L. Myers, and K. Ye: 2012: Probability and Statistics for Engineers and Scientists. Prentice Hall, 9th ed., 791pp.

Woo, H.-J., K.-A. Park, 2020: Inter-Comparisons of Daily Sea Surface Temperatures and In-Situ Temperatures in the Coastal Regions. Remote Sensing, 12, 1592; doi:10.3390/rs12101592.

Xie, J., J. Zhu, Y. Lib, 2008: Assessment and inter-comparison of five high-resolution sea surface temperature products in the shelf and coastal seas around China. Continental Shelf Research, 28, 1286–1293, https://doi.org/10.1016/j.csr.2008.02.020.

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.

Xu, F., A. Ignatov, 2016: Error characterization in iQuam SSTs using triple collocations with satellite measurements, Geophys. Res. Lett., 43, 10,826–10,834, doi:10.1002/2016GL070287.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

Yang, C. and coauthors, 2020: Sea Surface Temperature intercomparison in the framework of the Copernicus Climate Change Service (C3S), J. Climate, 34, 5257–5283, https://doi.org/10.1175/JCLI-D-20-0793.1.

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.

Zhang, H.-M., R. W. Reynolds, T. M. Smith (2004), Bias characteristics in the AVHRR sea surface temperature, Geophys. Res. Lett., 31, L01307, doi:10.1029/2003GL018804.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

Table Captions

Table 1. Daily SST datasets (from January 2016 to July 2020) used in this study (all data were downloaded on August 15, 2020).

Table 2. Globally averaged biases and RMSDs (°C) of SST datasets against buoy and Argo observations. The lowest biases and RMSDs are in bold text.

Table 3. Globally averaged biases and RMSDs (°C) of SST datasets against Argo10 SSTs. The lower biases and RMSDs are in bold text, but the low values in MUR25 are not highlighted due to its dependence on Argo10 SSTs.

Table 4. Average biases and RMSDs (°C) of SST datasets against UpTempO Level-2 data in the Arctic (Fig. 7). The lower biases and RMSDs are in bold text. The pair numbers are the counts of collocated data pairs between SST products and UpTempO.

Table 5. Globally averaged biases and RMSDs (°C) of original and bias-adjusted daytime and nighttime satellite observations in comparison with buoy SSTs.

Figure Captions

Figure 1. Geographic locations of Level 2 UpTempO buoy observations (green dots). Each dot represents a daily average.

Figure 2. Average SST biases (ºC, January 2016 – June 2020) against buoy observations. The biases are stippled when they are significant at the 95% confidence level. (a) DOISST, (b) MUR25, (c) GMPE, (d) GAMSSA, (e) OSTIA, (f) GPB, (g) CCI, and (h) CMC.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-21-0001.1.

Figure 3. (a) Collocated and globally averaged SST biases of DOISST (solid red), MUR25 (dashed blue), GMPE (solid black), GAMSSA (dotted green), OSTIA (dotted black), GPB (solid light green), CCI (solid purple), and CMC (dotted orange) against buoy observations. (b) Same as (a) except for against Argo observations. A 15-day filter is applied to all curves for readability.

Figure 4. RMSDs of SSTs (ºC, January 2016 – June 2020) against buoy observations. (a) DOISST, (b) MUR25, (c) GMPE, (d) GAMSSA, (e) OSTIA, (f) GPB, (g) CCI, and (h) CMC.

Figure 5. Average biases of SSTs (°C) against the independent Argo10 SSTs. The biases are stippled when they are significant at the 95% confidence level. (a) DOISST_Argo90, (b) MUR25, (c) GMPE, (d) GAMSSA, (e) OSTIA, (f) GPB, (g) CCI, and (h) CMC.

Figure 6. RMSDs of SSTs (°C) against the independent Argo10 SSTs. (a) DOISST_Argo90, (b) MUR25, (c) GMPE, (d) GAMSSA, (e) OSTIA, (f) GPB, (g) CCI, and (h) CMC.

Figure 7. Collocate SST biases (°C) against independent Level 2 UpTempO buoy observations. (a) DOISST, (b) MUR25, (c) GMPE, (d) GAMSSA, (e) OSTIA, (f) GPB, (g) CCI, and (h) CMC. Each red dot represents one pair of daily averaged SSTs. The total number of pairs are shown in Table 5.

Figure 8. Average biases of satellite SSTs (°C) against buoy observations. (a) Daytime MetOp-A, (b) Nighttime MetOp-A, (c) Daytime MetOp-B, (d) Nighttime MetOp-B, (e)–(h) The same as (a)–(d) except for the bias-adjusted satellite SSTs.

Figure 9. RMSDs of satellite SSTs (°C) against buoy observations. (a) Daytime MetOp-A, (b) Nighttime MetOp-A, (c) Daytime MetOp-B, (d) Nighttime MetOp-B, (e)–(h) The same as (a)–(d) except for the bias-adjusted satellite SSTs.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

Figure 10. Coverage by observations of (a) ship, (b) buoy, (c) Argo, (d) ship+buoy, and (e) ship+buoy+Argo. Use of 2º×2º spatial resolution and 15-day data windows, January 2016 – June 2020. (f) Coverage difference between (e) and (d).

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

Table 1. Daily SST datasets (from January 2016 to July 2020) used in this study (all data were downloaded on August 15, 2020).

Dataset	Version	Resolution	Input	Access
DOISST	v2.0 (1981–2019) v2.1 (2016–present)	0.25°	AVHRR + Ship + Buoy + Argo	https://www.ncei.noaa.gov/data/sea-surface-temperature-optimum-interpolation/v2.1/access/avhrr/
	DOISST_Argo90	0.25º	AVHRR + Ship + Buoy + 90% of Argo	This study
	DOISST_Buoy90	0.25º	AVHRR + Ship + 90% of Buoy + Argo	This study
MUR25	MUR v4.2 (2002–)	0.25°	AVHRR + Microwave + Ship + Buoy + Argo	https://podaac-opendap.jpl.nasa.gov/opendap/allData/ghrsst/data/GDS2/L4/GLOB/JPL/MUR25/v4.2
GMPE	v1 (2009-12) v2 (2012-17) v3 (2017–)	0.25°	GHRSST ensemble SSTs	https://resources.marine.copernicus.eu/?option=com_csw&view=details&product_id=SST_GLO_SST_L4_NRT_OBSERVATIONS_010_005
GAMSSA	v1 (2008–)	0.25°	AVHRR + AATSR + AMSRE + Ship + Buoy	https://podaac-opendap.jpl.nasa.gov/opendap/allData/ghrsst/data/L4/GLOB/ABOM/GAMSSA_28km
OSTIA	v2 (2006–)	0.05°	AVHRR + AMSR2 + VIIRS + SEVIRI + SLSTR + Ship + Buoy	ftp://ftp.nodc.noaa.gov/pub/data.nodc/ghrsst/GDS2/L4/GLOB/UKMO/OSTIA/v2
GPB	v1 (2014–)	0.05°	Imager + AVHRR + VIIRS + Ship + Buoy	ftp://ftp.nodc.noaa.gov/pub/data.nodc/ghrsst/GDS2/L4/GLOB/OSPO/Geo_Polar_Blended_Night/v1/
CCI	v2.0 (1981–2019)	0.05°	AVHRR + ATSR + ATSR2 + Adv. ATSR	http://dap.ceda.ac.uk/neodc/c3s_sst/data/ICDR_v2/Analysis/L4/v2.0
CMC	v3 (2016–)	0.1°	AVHRR + AMSR2 Ship + Buoy	ftp://ftp.nodc.noaa.gov/pub/data.nodc/ghrsst/GDS2/L4/GLOB/CMC/CMC0.1deg/v3

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

Table 2. Globally averaged biases and RMSDs (°C) of SST datasets against buoy and Argo observations. The lowest biases and RMSDs are in bold text.

Data set	**Buoy reference**		**Argo reference**	
	Bias	RMSD	Bias	RMSD
DOISST	−0.035	0.279	−0.039	0.243
MUR25	+0.018	0.386	+0.034	0.261
GMPE	−0.052	0.327	−0.053	0.245
GAMSSA	−0.083	0.364	−0.072*	0.338*
OSTIA	–0.068	0.358	−0.070*	0.287*
GPB	−0.076	0.373	−0.070*	0.269*
CCI	−0.084*	0.411*	−0.070*	0.284*
CMC	−0.059	0.338	−0.052*	0.258*

Table 3. Globally averaged biases and RMSDs (°C) of SST datasets against Argo10 SSTs. The lower biases and RMSDs are in bold text, but the low values in MUR25 are not highlighted due to its dependence on Argo10 SSTs.

Dataset	**Buoy10 reference**		**Argo10 reference**	
	Bias	RMSD	Bias	RMSD
DOISST_Buoy90	–0.039*	0.310*	N/A	N/A
DOISST_Argo90	N/A	N/A	–0.061*	0.322*
MUR25	+0.026	0.332	+0.031	0.245
GMPE	−0.047	0.327	–0.058	0.231
GAMSSA	−0.077	0.306	–0.080*	0.317*
OSTIA	–0.059	0.281	–0.072*	0.256*
GPB	−0.070	0.318	–0.076*	0.254*
CCI	−0.075	0.351	–0.070*	0.268*
CMC	−0.052	0.286	–0.055*	0.244*

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

Table 4. Average biases and RMSDs (°C) of SST datasets against UpTempO Level-2 data in the Arctic (Fig. 7). The lower biases and RMSDs are in bold text. The pair numbers are the counts of collocated data pairs between SST products and UpTempO.

Dataset	**Bias**	**RMSD**	**Pair number**
DOISST	+0.057	0.473	1706
MUR25	−0.206	0.444	1706
GMPE	+0.115	0.686	113
GAMSSA	−0.232	0.469	1706
OSTIA	+0.082	0.664	1706
GPB	+0.086	0.674	1668
CCI	−0.218	0.416	1706
CMC	−0.139	0.487	1706

Table 5. Globally averaged biases and RMSDs (°C) of original and bias-adjusted daytime and nighttime satellite observations in comparison with buoy SSTs.

Dataset	**Bias**		**RMSD**	
	Original	Adjusted	Original	Adjusted
MetOp-A, daytime	−0.04	−0.02	0.57	0.58
MetOp-A, nighttime	+0.02	−0.04	0.52	0.54
MetOp-B, daytime	−0.11	−0.02	0.62	0.58
MetOp-B, nighttime	−0.02	−0.04	0.55	0.54

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

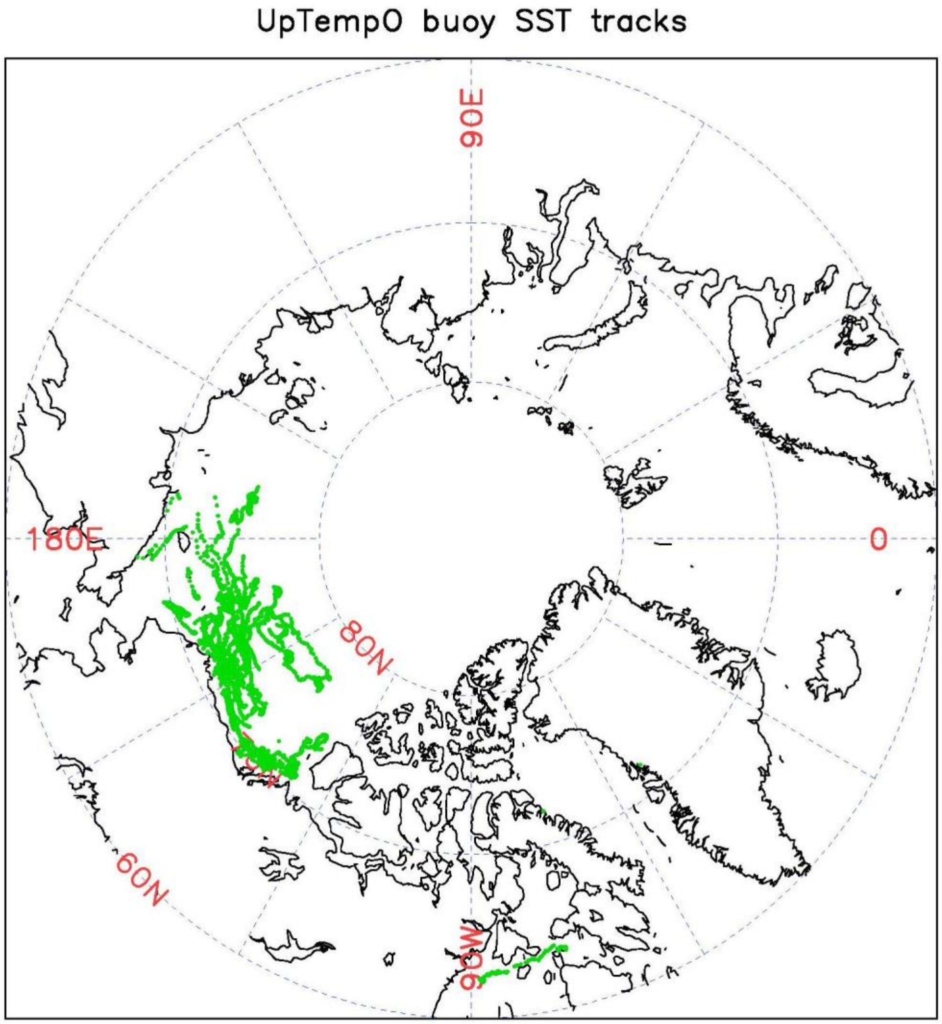

Figure 1. Geographic locations of Level 2 UpTempO buoy observations (green dots). Each dot represents a daily average.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

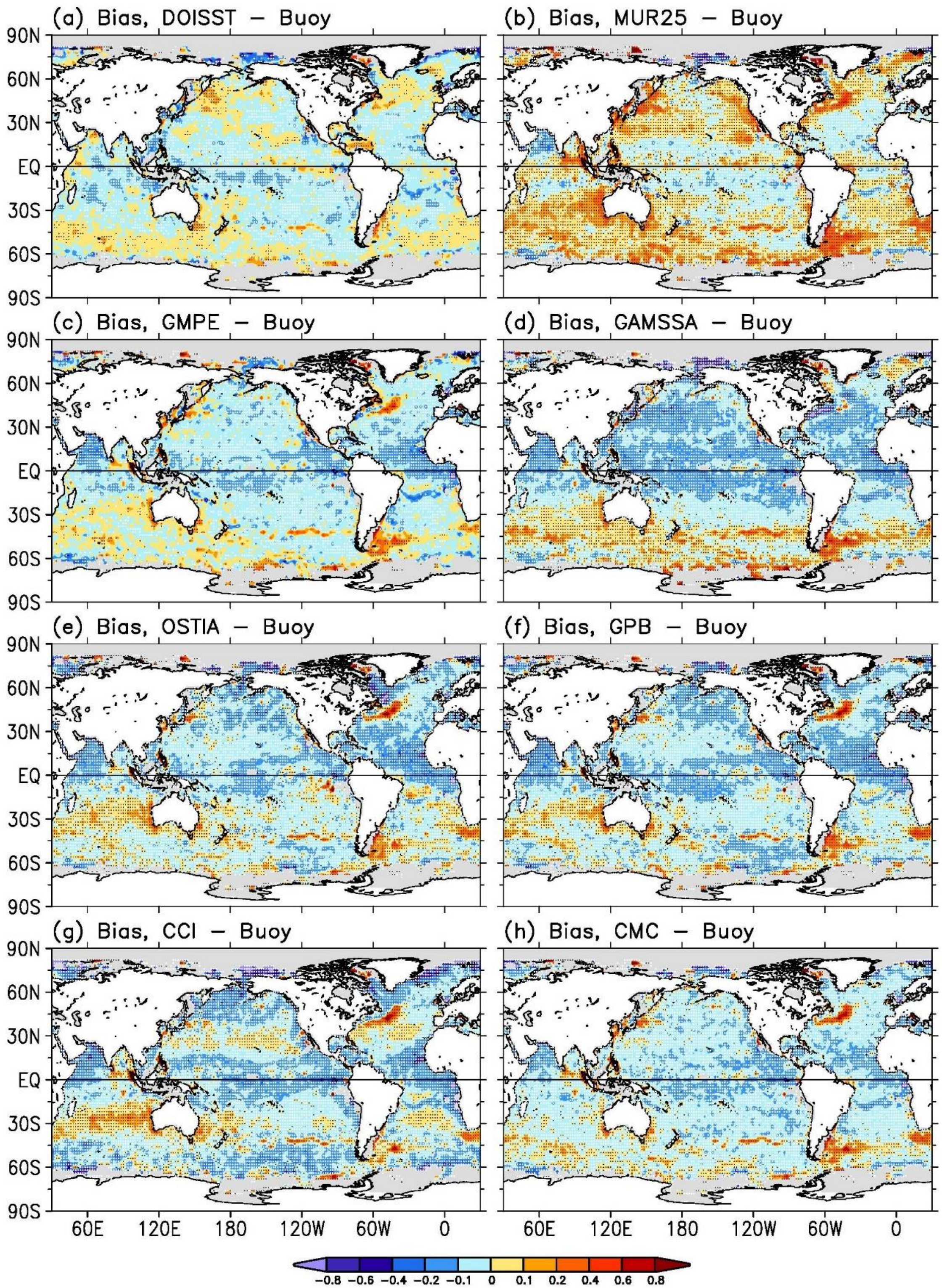

Figure 2. Average SST biases (ºC, January 2016 – June 2020) against buoy observations. The biases are stippled when they are significant at the 95% confidence level. (a) DOISST, (b) MUR25, (c) GMPE, (d) GAMSSA, (e) OSTIA, (f) GPB, (g) CCI, and (h) CMC.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-21-0001.1.

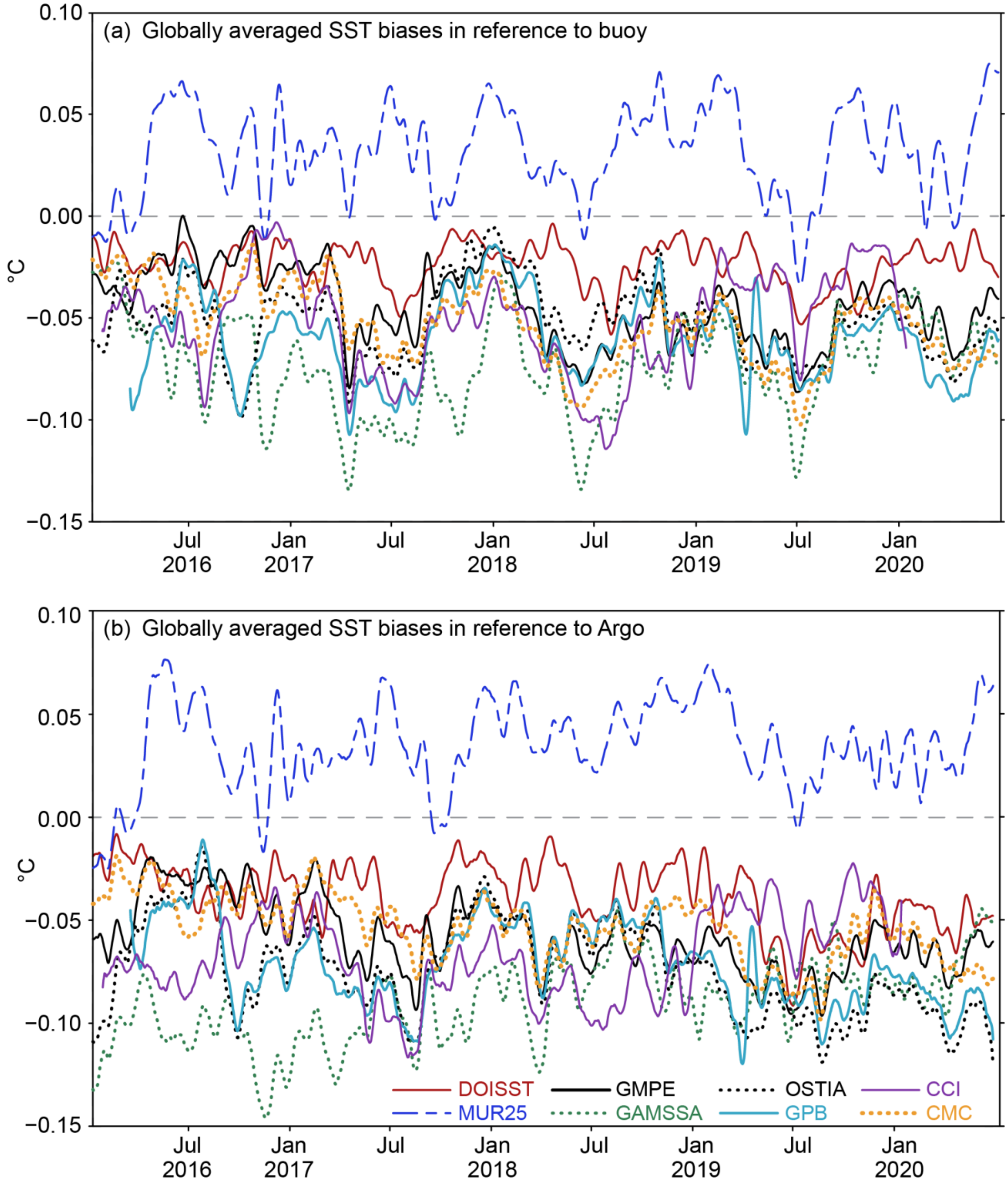

Figure 3. (a) Collocated and globally averaged SST biases of DOISST (solid red), MUR25 (dashed blue), GMPE (solid black), GAMSSA (dotted green), OSTIA (dotted black), GPB (solid light green), CCI (solid purple), and CMC (dotted orange) against buoy observations. (b) Same as (a) except for against Argo observations. A 15-day filter is applied to all curves for readability.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

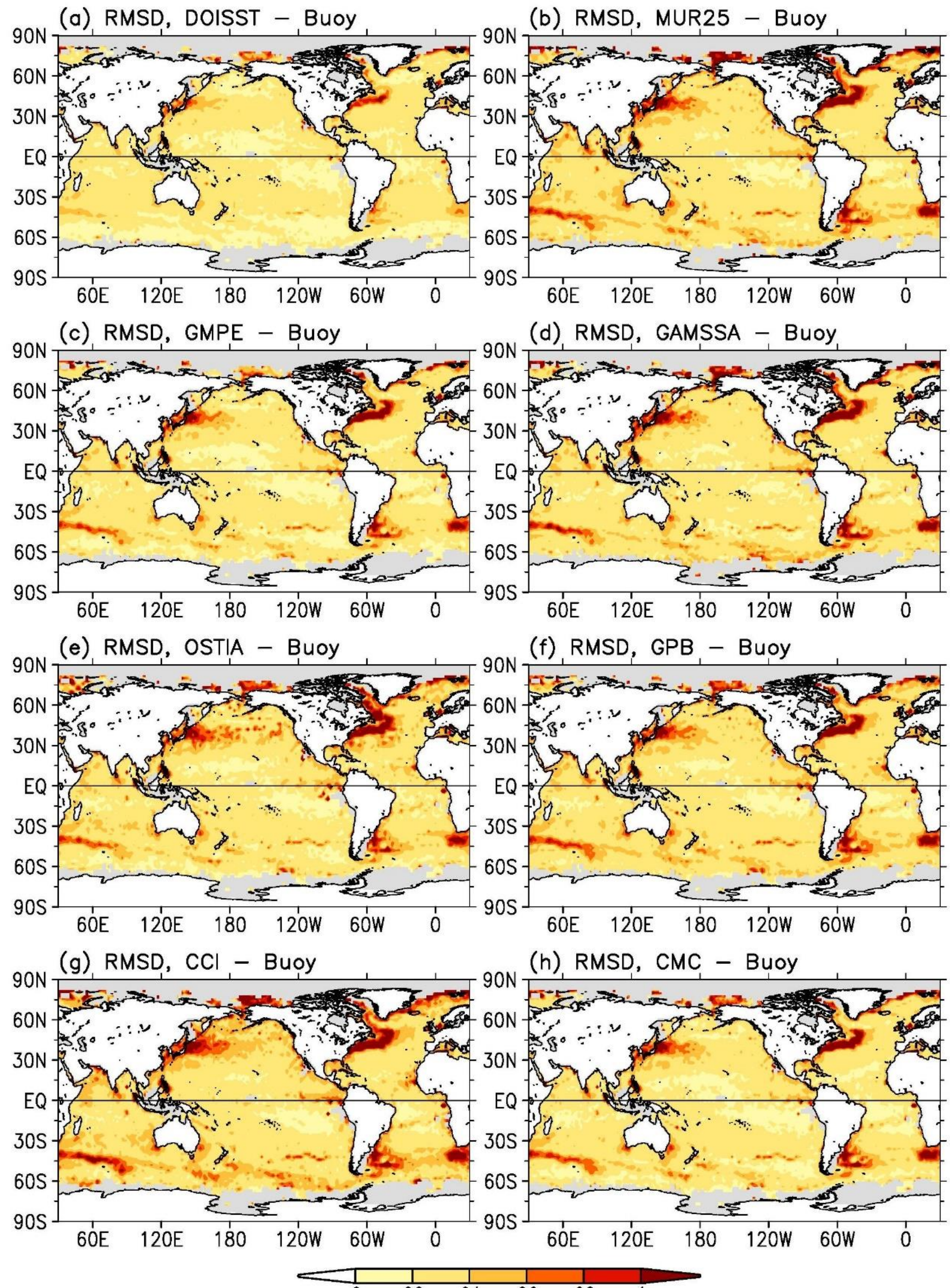

Figure 4. RMSDs of SSTs (ºC, January 2016 – June 2020) against buoy observations. (a) DOISST, (b) MUR25, (c) GMPE, (d) GAMSSA, (e) OSTIA, (f) GPB, (g) CCI, and (h) CMC.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-21-0001.1.

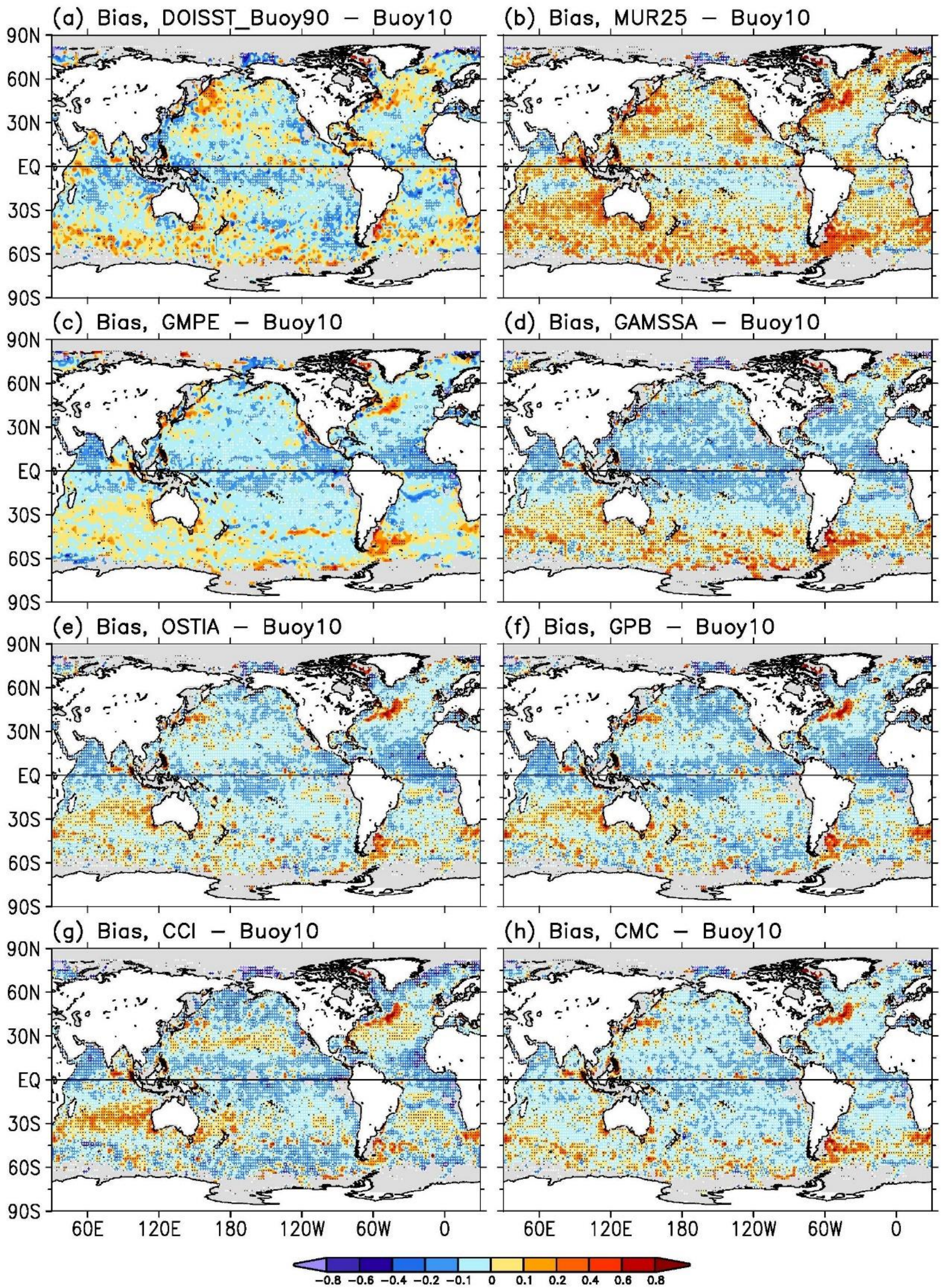

Figure 5. Average biases of SSTs (°C) against the independent Argo10 SSTs. The biases are stippled when they are significant at the 95% confidence level. (a) DOISST_Argo90, (b) MUR25, (c) GMPE, (d) GAMSSA, (e) OSTIA, (f) GPB, (g) CCI, and (h) CMC.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-21-0001.1.

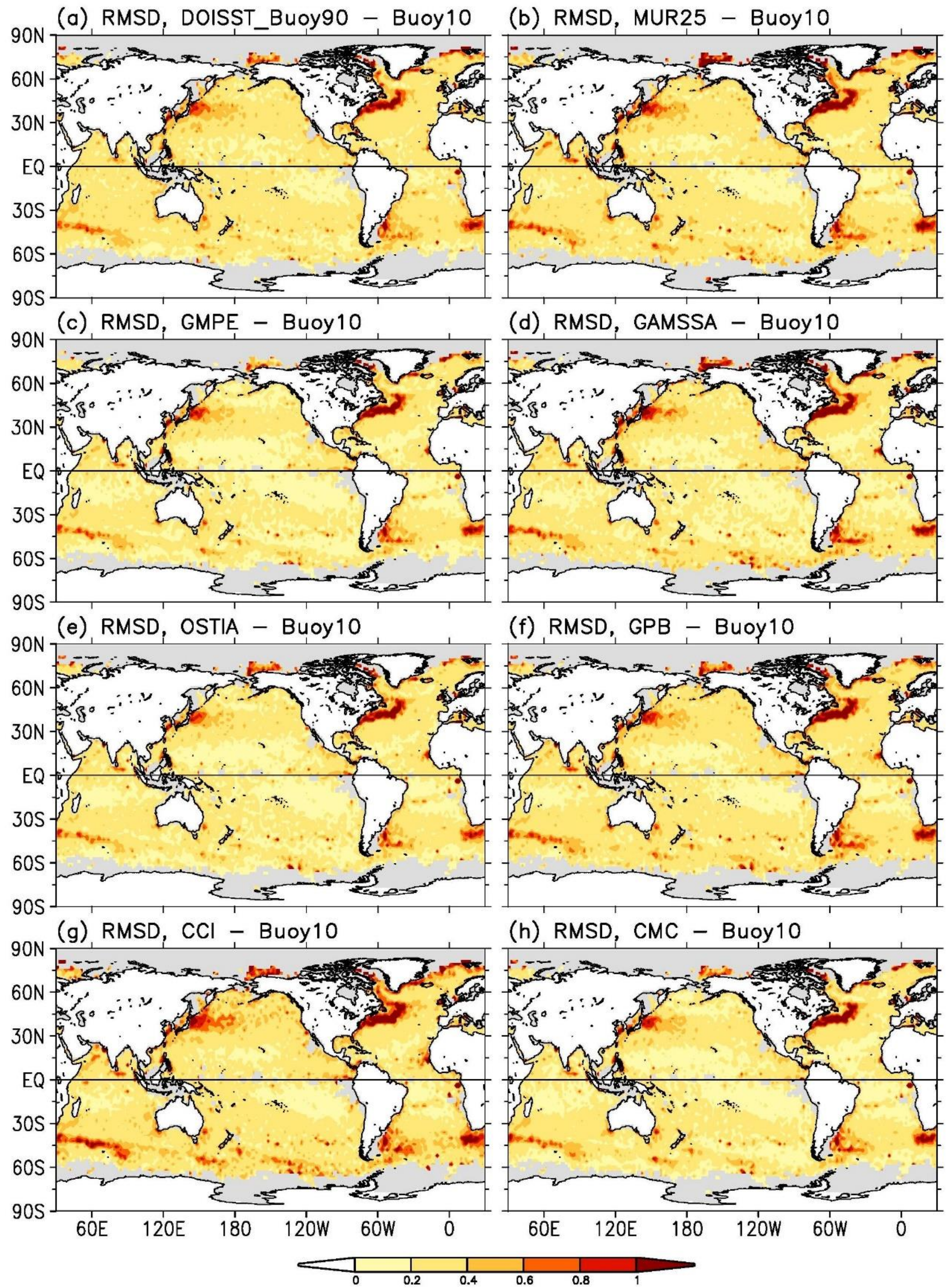

Figure 6. RMSDs of SSTs (°C) against the independent Argo10 SSTs. (a) DOISST_Argo90, (b) MUR25, (c) GMPE, (d) GAMSSA, (e) OSTIA, (f) GPB, (g) CCI, and (h) CMC.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-21-0001.1.

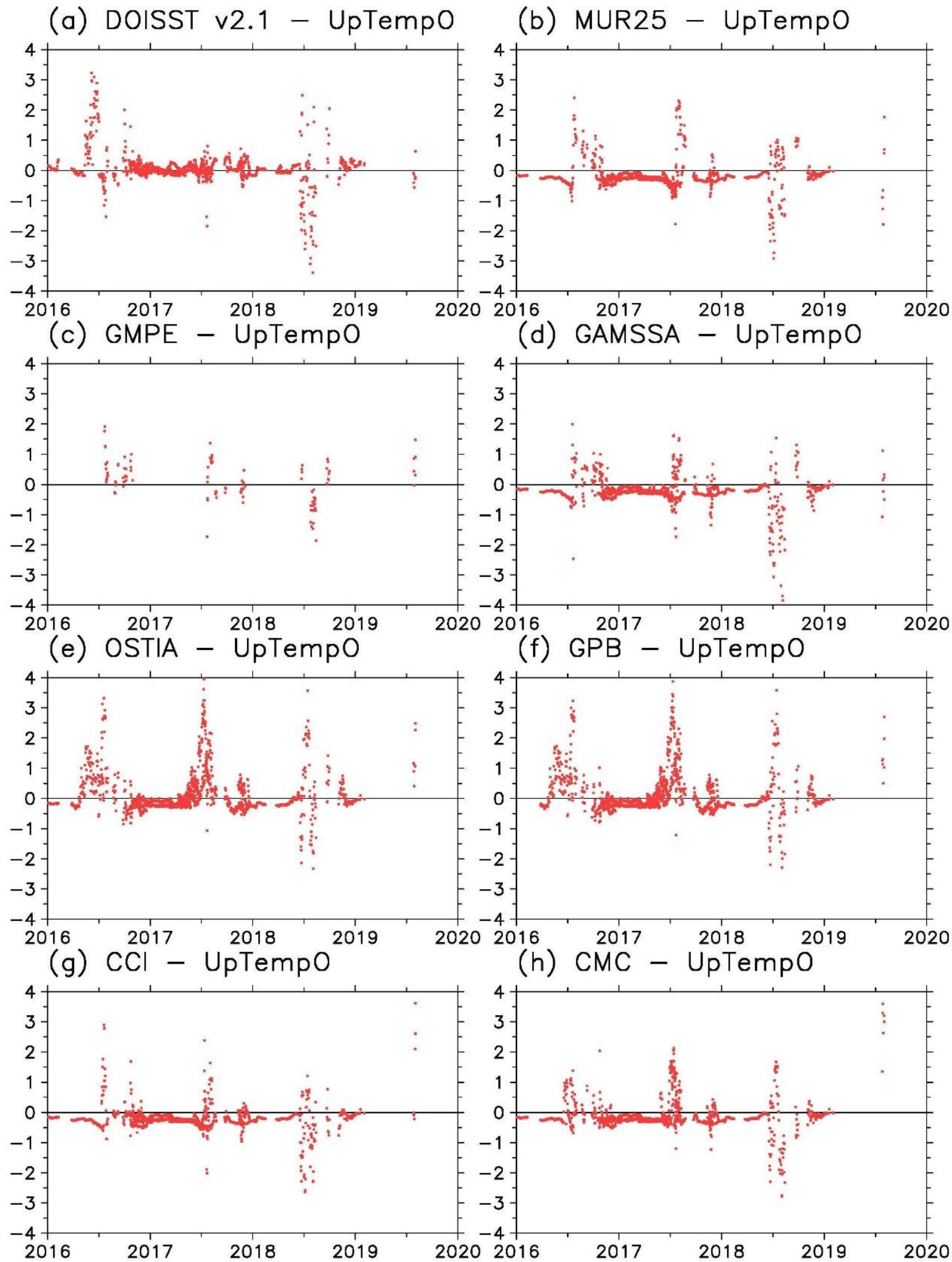

Figure 7. Collocate SST biases (°C) against independent Level 2 UpTempO buoy observations. (a) DOISST, (b) MUR25, (c) GMPE, (d) GAMSSA, (e) OSTIA, (f) GPB, (g) CCI, and (h) CMC. Each red dot represents one pair of daily averaged SSTs. The total number of pairs are shown in Table 5.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-21-0001.1.

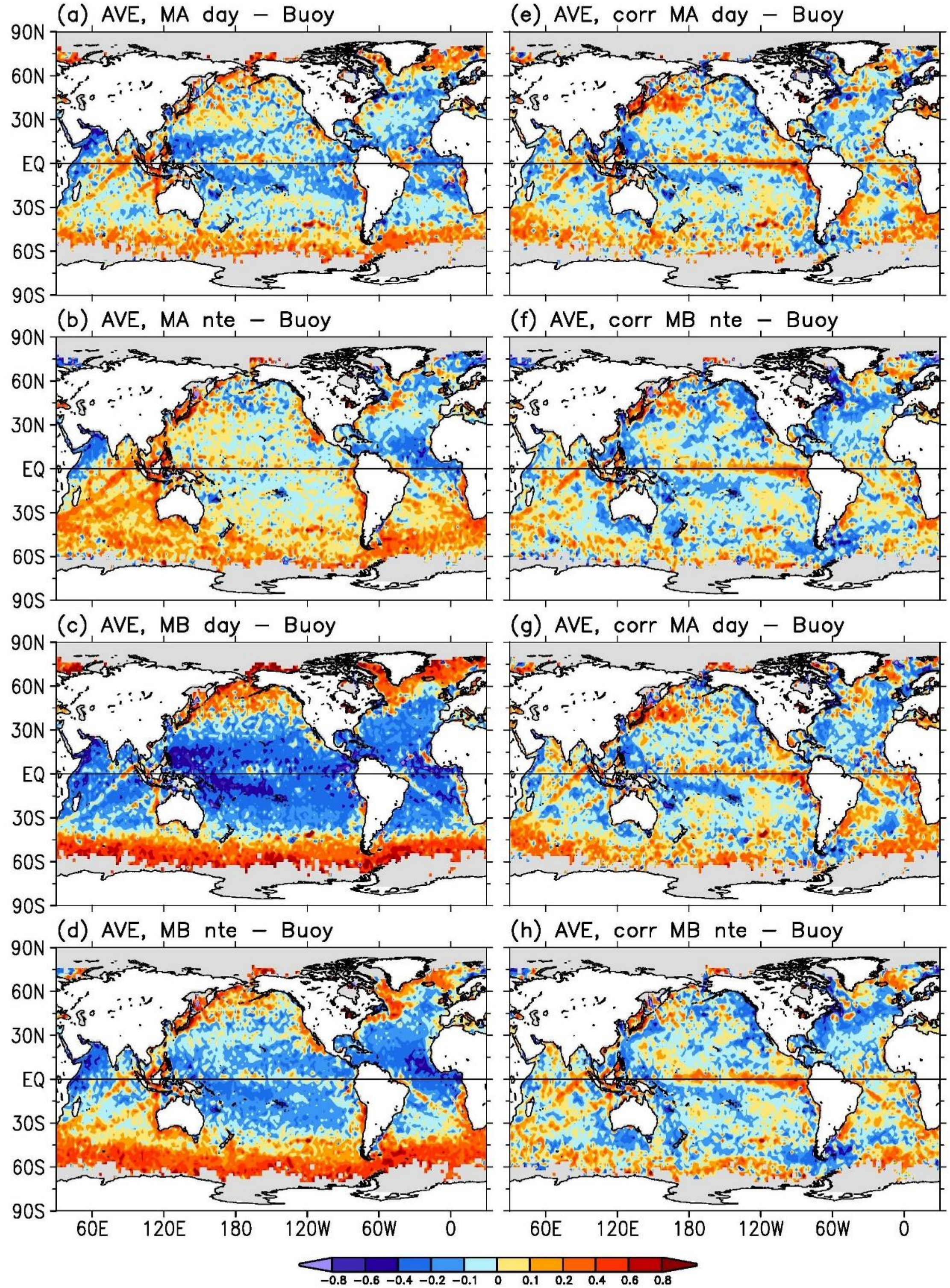

Figure 8. Average biases of satellite SSTs (°C) against buoy observations. (a) Daytime MetOp-A, (b) Nighttime MetOp-A, (c) Daytime MetOp-B, (d) Nighttime MetOp-B, (e)–(h) The same as (a)–(d) except for the bias-adjusted satellite SSTs.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-21-0001.1.

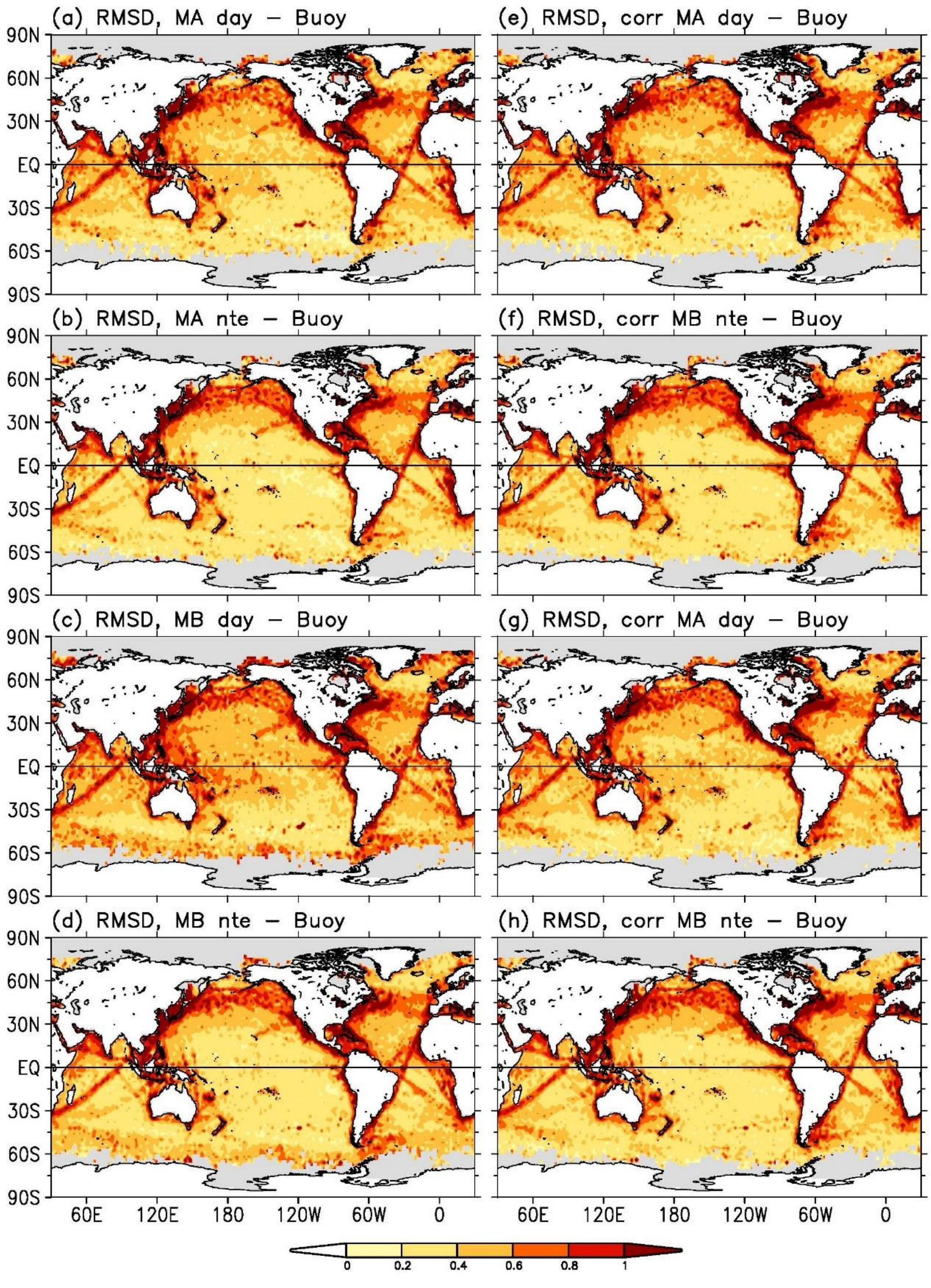

Figure 9. RMSDs of satellite SSTs (°C) against buoy observations. (a) Daytime MetOp-A, (b) Nighttime MetOp-A, (c) Daytime MetOp-B, (d) Nighttime MetOp-B, (e)–(h) The same as (a)–(d) except for the bias-adjusted satellite SSTs.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-21-0001.1.

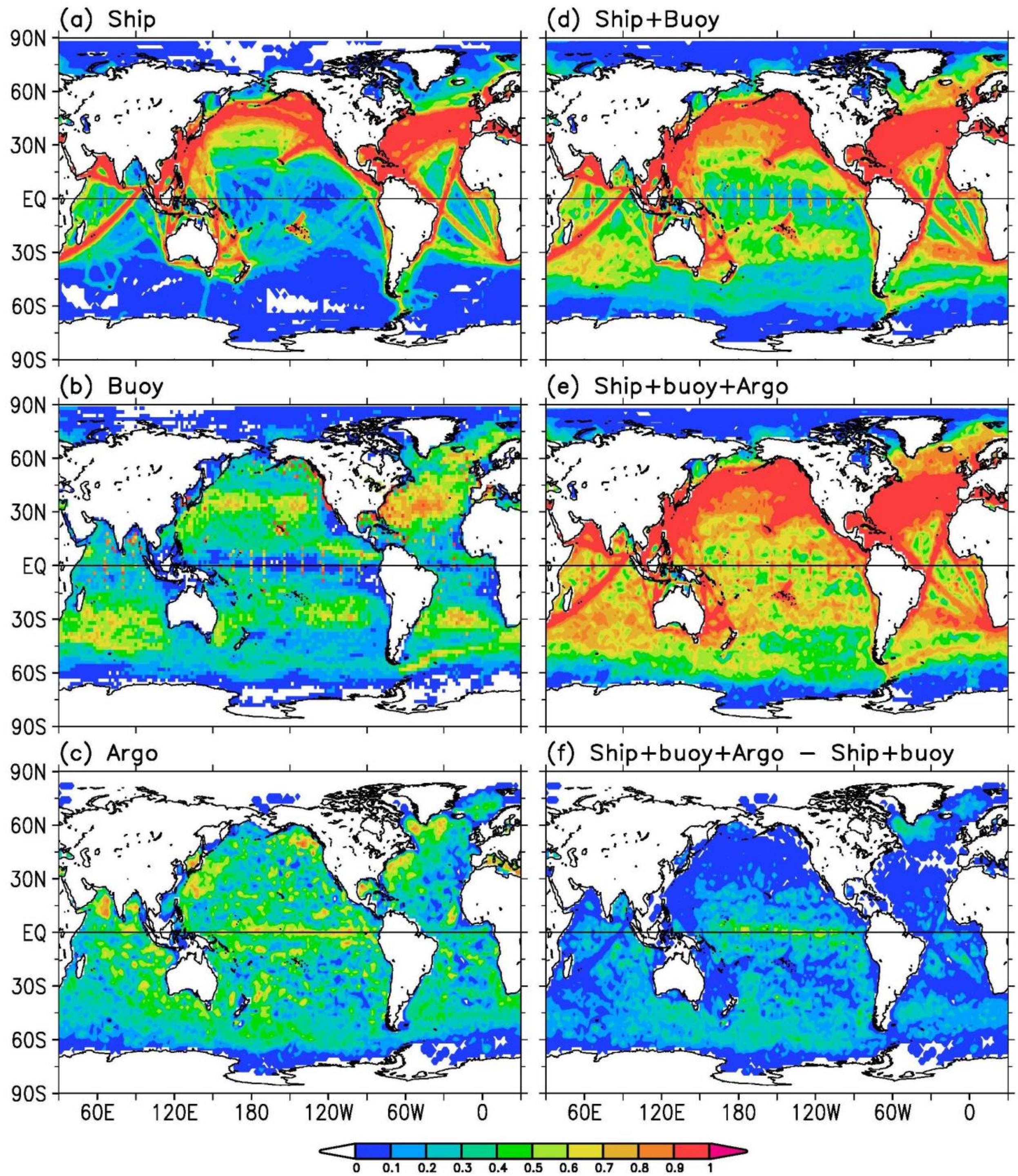

Figure 10. Coverage by observations of (a) ship, (b) buoy, (c) Argo, (d) ship+buoy, and (e) ship+buoy+Argo. Use of 2°×2° spatial resolution and 15-day data windows, January 2016 – June 2020. (f) Coverage difference between (e) and (d).

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-21-0001.1.