Extended Reconstructed Sea Surface Temperature Version 4 (ERSST.v4). Part I: Upgrades and Intercomparisons

BOYIN HUANG,* VIVA F. BANZON,* ERIC FREEMAN,+ JAY LAWRIMORE,* WEI LIU,# THOMAS C. PETERSON,* THOMAS M. SMITH,@ PETER W. THORNE,& SCOTT D. WOODRUFF,** AND HUAI-MIN ZHANG*

* *National Climatic Data Center, Asheville, North Carolina*
+ *National Climatic Data Center, Asheville, North Carolina, and STG, Inc., Reston, Virginia*
# *National Climatic Data Center, Asheville, and Cooperative Institute for Climate and Satellites, Raleigh, North Carolina*
@ *NOAA/STAR/SCSB, and CICS/ESSIC, University of Maryland, College Park, College Park, Maryland*
& *Nansen Environmental and Remote Sensing Center, Bergen, Norway*
** *National Climatic Data Center, Asheville, North Carolina, and Cooperative Institute for Research in Environmental Sciences, Boulder, Colorado*

(Manuscript received 20 December 2013, in final form 3 October 2014)

ABSTRACT

The monthly Extended Reconstructed Sea Surface Temperature (ERSST) dataset, available on global 2° × 2° grids, has been revised herein to version 4 (v4) from v3b. Major revisions include updated and substantially more complete input data from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS) release 2.5; revised empirical orthogonal teleconnections (EOTs) and EOT acceptance criterion; updated sea surface temperature (SST) quality control procedures; revised SST anomaly (SSTA) evaluation methods; updated bias adjustments of ship SSTs using the Hadley Centre Nighttime Marine Air Temperature dataset version 2 (HadNMAT2); and buoy SST bias adjustment not previously made in v3b.

Tests show that the impacts of the revisions to ship SST bias adjustment in ERSST.v4 are dominant among all revisions and updates. The effect is to make SST 0.1°–0.2°C cooler north of 30°S but 0.1°–0.2°C warmer south of 30°S in ERSST.v4 than in ERSST.v3b before 1940. In comparison with the Met Office SST product [the Hadley Centre Sea Surface Temperature dataset, version 3 (HadSST3)], the ship SST bias adjustment in ERSST.v4 is 0.1°–0.2°C cooler in the tropics but 0.1°–0.2°C warmer in the midlatitude oceans both before 1940 and from 1945 to 1970. Comparisons highlight differences in long-term SST trends and SSTA variations at decadal time scales among ERSST.v4, ERSST.v3b, HadSST3, and Centennial Observation-Based Estimates of SST version 2 (COBE-SST2), which is largely associated with the difference of bias adjustments in these SST products. The tests also show that, when compared with v3b, SSTAs in ERSST.v4 can substantially better represent the El Niño/La Niña behavior when observations are sparse before 1940. Comparisons indicate that SSTs in ERSST.v4 are as close to satellite-based observations as other similar SST analyses.

1. Introduction

Sea surface temperature (SST) is one of the most important indicators of climate variability and long-term climate change. SSTs are used to monitor many modes of climate variability such as El Niño–Southern Oscillation (ENSO), the Pacific decadal oscillation (PDO), the Atlantic multidecadal oscillation (AMO), and the Indian Ocean dipole (IOD) (Philander 1990; Latif and Barnett 1994; Saji et al. 1999; Enfield et al. 2001). Historical SST data have played an important role in climate simulation, assessment, and monitoring (Hurrell and Trenberth 1999; Stocker et al. 2014; Gregg and Newlin 2012). Owing to the importance of SST in climate variability and assessment, a variety of global gridded SST datasets have been independently created through historical "reconstruction" techniques, including the Optimum Interpolation SST (OISST), the Hadley Centre SST (HadSST) and Sea Ice and SST datasets (HadISST), Extended Reconstructed SST (ERSST), Kaplan SST, and Centennial Observation-Based Estimates of SSTs (COBE-SST) (Rayner et al. 2003; Reynolds et al. 2002; Parker et al. 1994; Smith et al. 1996; Kaplan et al. 1998; Ishii et al. 2005).

Large-scale multidecadal variations in the SST products are critically dependent on the bias adjustment of historical ship-based SST observations, since buoys and

Corresponding author address: Boyin Huang, 151 Patton Ave., National Climatic Data Center, Asheville, NC 28801.
E-mail: boyin.huang@noaa.gov

DOI: 10.1175/JCLI-D-14-00006.1

© 2015 American Meteorological Society

other automated platforms measuring SST were not introduced widely until the 1970s. The historical ship SST data were measured by a range of methods that have changed through time [see, e.g., the discussion of Hartmann et al. (2014, their Fig. 2.15) based on the earlier study of Kennedy et al. (2011)]. These methodological inhomogeneities are believed to yield, for example, cold biases due to the heat loss by evaporation when SSTs were measured from some (particularly uninsulated) buckets, contrasting with warm biases due to the heat gain from the ship's interior when engine room intake (ERI) samples were measured. To bias adjust for the changing measurement methodologies, quantitative estimates have been made of these various biases by different groups. For example, heat loss estimates have been made for SST measurements from buckets that occur during the time between the hauling of buckets from the ocean surface and the reading of thermometers (Folland and Parker 1995).

For ERSST, in contrast to other SST analyses, ship SSTs are adjusted using Nighttime Marine Air Temperature (NMAT) data. The analysis of the previous version of ERSST, version v3b (ERSST.v3b; Smith and Reynolds 2004; Smith et al. 2008; Banzon et al. 2010) using NMAT from the Comprehensive Ocean–Atmosphere Data Set (COADS; Woodruff et al. 1987), indicated that the NMAT estimates can be used to identify and remove SST biases to construct a climate data record of SSTs (Smith and Reynolds 2002). However, further upgrades of SST holdings and SST bias adjustment understanding means that revisions to the ERSST have now become necessary, specifically given the improved scientific understanding of SST data and their biases during the past decade since the release of ERSST.v3b.

First, ERSST.v3b does not provide SST bias adjustment after 1941 whereas subsequent analyses (e.g., Thompson et al. 2008) have highlighted potential post-1941 data issues and some newer datasets have addressed these issues (Kennedy et al. 2011; Hirahara et al. 2014). The latest release of Hadley NMAT version 2 (HadNMAT2) from 1856 to 2010 (Kent et al. 2013) provided better quality-controlled NMAT, which includes adjustments for increased ship deck height, removal of artifacts, and increased spatial coverage due to added records. These NMAT data are better suited to identifying SST biases in ERSST, and therefore the bias adjustments in ERSST version 4 (ERSST.v4) have been estimated throughout the period of record instead of exclusively to account for pre-1941 biases as in v3b.

Second, the in situ data have been updated from International Comprehensive Ocean–Atmosphere Data Set (ICOADS) release 2.4 (R2.4) [see description of R2.4 in Woodruff et al. (2011)], which is used in ERSST.v3b, to release 2.5 (R2.5) (Woodruff et al. 2011). R2.5 provides better duplicate removal and gross quality control (QC), a larger number of observations, and a better coverage in previously undersampled areas, both spatially and temporally.

Finally, estimates of uncertainty of its SST reconstruction (so-called parametric uncertainty) were not provided in ERSST.v3b, and therefore parametric uncertainty was not included in the total uncertainty of SSTs in ERSST.v3b. Studies have shown that the parametric uncertainty is an important component of the total uncertainty as demonstrated in the latest Hadley Centre dataset, HadSST3 (Kennedy et al. 2011). These have been estimated in this new ERSST.v4 in the accompanying Part II paper (Liu et al. 2015, hereafter Part II).

This paper documents the aforementioned upgrades to and their impacts on ERSST. In ERSST.v4, a total of 11 parameters have been reassessed and revised due to either newly available observations or improved analysis methods (Table 1). Thus, ERSST.v4 is the result of an extensive analysis of the existing algorithm and systematic experimentation on a broad suite of system parameters. Wherever possible these parameter choices are justified in a quantitative and objective manner as discussed herein. The impacts of these choices and uncertainty in the ERSST.v4 product are discussed separately in Part II.

The ERSST methodology is briefly described in section 2. Datasets used in producing and validating ERSST.v4 are described in section 3. Upgrades in ERSST.v4 are described in section 4 except the upgrade for SST bias adjustment using HadNMAT2, which is described in section 5. The SST anomalies (SSTAs) in ERSST.v4 are compared with those in ERSST.v3b, HadSST3, and COBE-SST2 in section 6. The SSTs in ERSST.v4 are compared with independent analyses and satellite-based observations in section 7. A summary is given in section 8.

2. Reconstruction methodology

The methodology of ERSST.v4 reconstruction follows Smith et al. (1996) and Smith and Reynolds (2003). The SST measurements from in situ buoy and ship observations were used to reconstruct monthly 2° × 2° SSTA data in ERSST.v4 from 1875 to present. The reconstruction before 1875 was not accomplished due to sparseness of observations in the Pacific and Indian Oceans in ICOADS R2.5 and the inability to provide sufficient empirical orthogonal teleconnections (EOTs) for construction of a reliable "global" estimate. The SSTs from ships or buoys were accepted (rejected) under a QC

TABLE 1. Major methodological innovations between the current ERSST.v4 and its precursor ERSST.v3b.

Methodological aspect	ERSST.v4 choice	ERSST.v3b choice
SST data	ICOADS R2.5 (1875–2007) NCEP GTS (2008–present)	ICOADS R2.4 (1875–2004) NCEP GTS (2008–present)
Ice data	HadISST (1870–2010) NCEP (2011–present)	UKMO (1870–1980) GFSC (1981–2004) NCEP (2005–present)
130 EOTs	OISST.v2 1982–2011	OISST.v2 1982–2005
EOT criterion	Crit = 0.1	Crit = 0.2
EOT weighting	$\frac{N}{N+\varepsilon^2}w(N,\varepsilon)$	$\cos\phi$
SST STD for QC	OISST.v2 (1982–2011)	COADS (1950–79)
SSTA calculation	At in situ locations	At regular grid boxes
Low-frequency anomaly gap filling	Nearby anomaly filling	Zero-anomaly filling
NMAT in bias adjustment	HadNMAT2 (1875–2010) (adjustments before 1886 are set to be the values of 1886)	R2.4 (1875–1941)
Bias adjustment smoothing	Lowess filter coefficient $f = 0.1$	Linear
Ship-buoy SST adjustment	0.12°C is added to buoy SST	Not applied

criterion that observed SSTs differ from the first-guess SST from ERSST.v3b by less (more) than 4 times standard deviation (STD) of SST (Smith and Reynolds 2003).

The ship and buoy SSTs that have passed QC were then converted into SSTAs by subtracting the SST climatology (1971–2000) at their in situ locations in monthly resolution. The ship SSTA was adjusted based on the NMAT comparators; buoy SSTA was adjusted by a mean difference of 0.12°C between ship and buoy observations (section 5). The ship and buoy SSTAs were merged and bin-averaged into monthly "superobservations" on a 2° × 2° grid. The number of superobservations was defined here as the count of 2° × 2° grid boxes with valid data. The averaging of ship and buoy SSTAs within each 2° × 2° grid box was based on their proportions to the total number of observations. The number of buoy observations was multiplied by a factor of 6.8, which was determined by the ratio of random error variances of ship and buoy observations (Reynolds and Smith 1994), suggesting that buoy observations exhibit much lower random variance than ship observations.

The SSTAs of superobservations were further decomposed into low- and high-frequency components. The low-frequency component was constructed by applying a 26° × 26° spatial running mean using monthly superobservations where the sampling ratio is larger than 3% (five superobservations). An annual mean SSTA was then defined with a minimum requirement of two months of valid data. The annual mean SSTA fields were screened and the missing SSTAs were filled by searching the neighboring SSTAs within 10° in longitude, 6° in latitude, and 3-yr in time. The search areas were tested using ranges of 15°–20° in longitude, 5°–10° in latitude, and 2–5 yr. The final SSTAs did not make much of a difference since the search area is less than the scales of the low-frequency filter. Finally, the annually averaged SSTAs were filtered with a weak three-point binomial filter in longitudinal and latitudinal directions, and further filtered with a 15-yr median filter. These processes were designed to filter out high-frequency noise in time and small scale in space.

The high-frequency component of SSTA, defined as the difference between the original and low-frequency SSTAs, was reconstructed by first applying a 3-month running filter that replaces missing data with an average of valid pre- and postcurrent month data. The filtered SSTAs were then fitted to the 130 leading EOTs (van den Dool et al. 2000; Smith et al. 2008), which are localized empirical orthogonal functions restricted in domain to a spatial scale of 5000 and 3000 km in longitude and latitude, respectively. The EOTs were trained by monthly OISST.v2 from 1982 to 2011:

$$R(x) = \sum_i f_i \Psi_i(x), \tag{1}$$

where $R(x)$ is reconstructed SSTA, $\Psi_i(x)$ is the ith EOT, and f_i is the fitted reconstruction coefficient by minimizing the total error variance:

$$E^2 = \sum_x [O(x) - R(x)]^2 \delta_x w(N,\varepsilon)\cos\phi_x, \tag{2}$$

$$w(N,\varepsilon) = \frac{N_s + 6.8N_b}{N_s + 6.8N_b + \varepsilon^2}, \tag{3}$$

where $O(x)$ represents SST superobservations; δ_x is 1 when a grid box contains observations and 0 otherwise; $\cos\phi_x$ is an area weighting function of latitude; N_s and N_b are the number of observations from ships and buoys, respectively, and $N = N_s + 6.8N_b$; the factor of 6.8 is

determined by the ratio of error variances of ship and buoy observations (Reynolds and Smith 1994); and ε is averaged error of ship (1.3°C) and buoy (0.5°C) SST observations weighted by their observation numbers (Reynolds et al. 2002).

The EOT fitting coefficients f_i were calculated by solving linear equations using the lower upper (LU) decomposition method (Press et al. 1992), and the missing fitting coefficients were filtered out by an average of valid pre- and postcurrent month fitting coefficients weighted with a lag-1 autocorrelation coefficient of EOT fitting coefficients. The autocorrelation coefficients of the fitting functions for 130 EOT modes have been recalculated and updated after the EOTs are revised in ERSST.v4. It should be noted that there is substantial evidence that in the real world there exists correlated uncertainty in the input SST data (Kennedy et al. 2011). However, in ERSST it is necessary to make the simplifying assumption that the errors in Eq. (2) are uncorrelated.

During the SST reconstruction, not all 130 EOTs were actually used in the reconstruction of any given monthly field, depending on whether that mode is supported by actual observations. An EOT mode was accepted only if its variance ratio (r_i) is greater than a criterion (Crit) value of 0.1. The variance ratio r_i was defined as a ratio of accumulated variance, where an EOT mode is covered by superobservations, and the total variance of that EOT mode:

$$r_i = \frac{\sum_x \Psi_i^2(x)\delta_x \cos\phi_x}{\sum_x \Psi_i^2(x) \cos\phi_x}. \tag{4}$$

This ensures against undersampled EOTs being given undue weighting in the reconstruction. The SST data constructed from low- and high-frequency components were then merged, and SSTs at the grid boxes where sea ice concentration is greater than 60% were adjusted toward the freezing point of −1.8°C (Smith and Reynolds 2004).

3. Datasets

Various datasets have been used to create the ERSST.v4 product (section 3a) and independent SST reconstruction datasets have been used for comparisons (section 3b). Necessary details are outlined in this section for the readers.

a. *Input datasets used in ERSST construction*

1) SST OBSERVATIONAL DATA

The in situ SST data used in ERSST.v4 are from ICOADS R2.5 from 1875 to 2007 (Woodruff et al. 2011) and after 2007 from Global Telecommunications System (GTS) receipts from the National Centers for Environmental Prediction (NCEP). The data before 1875 in R2.5 are not used due to sparseness of observations that may result in unreliable EOT modes, most notably in the Pacific and Indian Oceans. R2.5 has substantially more observations than R2.4 (Fig. 1), particularly in the 1880s for ship observations and from 1970 to 1995 for buoy observations. Improvements in data coverage during these periods are indicated by the number of annually accumulated superobservations.

It is important to note that some SSTs from NCEP GTS data and/or ICOADS R2.5 are not utilized for ERSST.v4 due to concerns about their quality, additional biases or uncertainties. These excluded SSTs are from 1) the NOAA National Data Buoy Center (NDBC)'s Coastal-Marine Automated Network (C-MAN), since our focus is primarily on the oceans, and there is the potential for coastal land/topographical influences; and 2) SST estimates derived from the uppermost levels of oceanographic temperature profiles, which were in R2.5 from the National Oceanic and Atmospheric Administration (NOAA) National Oceanographic Data Center (NODC)'s World Ocean Database, owing to the concerns about the possibility of introducing new systematic or time-varying biases as discussed in Woodruff et al. (2008).

2) NIGHT MARINE AIR TEMPERATURES FOR BIAS ADJUSTMENT

Monthly HadNMAT2 data (Kent et al. 2013; 1856–2010 on a 5° × 5° grid) are used to perform the ship SST bias adjustments (section 5). The HadNMAT2 replaces the older COADS NMAT data used for performing SST bias adjustment in ERSST.v3b (Smith and Reynolds 2002). The ship SST bias adjustments are linearly interpolated to the 2° × 2° grid of ERSST.v4.

To validate the assumptions of the SST and NMAT measurands being of sufficient similarity to enable NMAT measurements to be used to adjust SST measurements, monthly SST and surface air temperature (SAT) from the Geophysical Fluid Dynamics Laboratory (GFDL) Coupled Model version 2.1 (CM2.1; Delworth et al. 2006) are partially sampled using monthly observational masks of SST from 1875 to 2000 (section 5). The CM2.1 is a coupled land, atmosphere, and ocean model. The resolution of the land and atmospheric components is 2° in latitude and 2.5° in longitude. The ocean resolution is 1° in longitude, 1° in latitude north/south of 30°N/30°S and ⅓° at the equator, and 10 m in depth above 220 m. The time-varying forcing agents of the CM2.1 are atmospheric CO_2, CH_4, N_2O, halons, tropospheric and stratospheric O_3, anthropogenic tropospheric sulfates, black and organic carbon, volcanic aerosols, solar irradiance, and the distribution of land cover types.

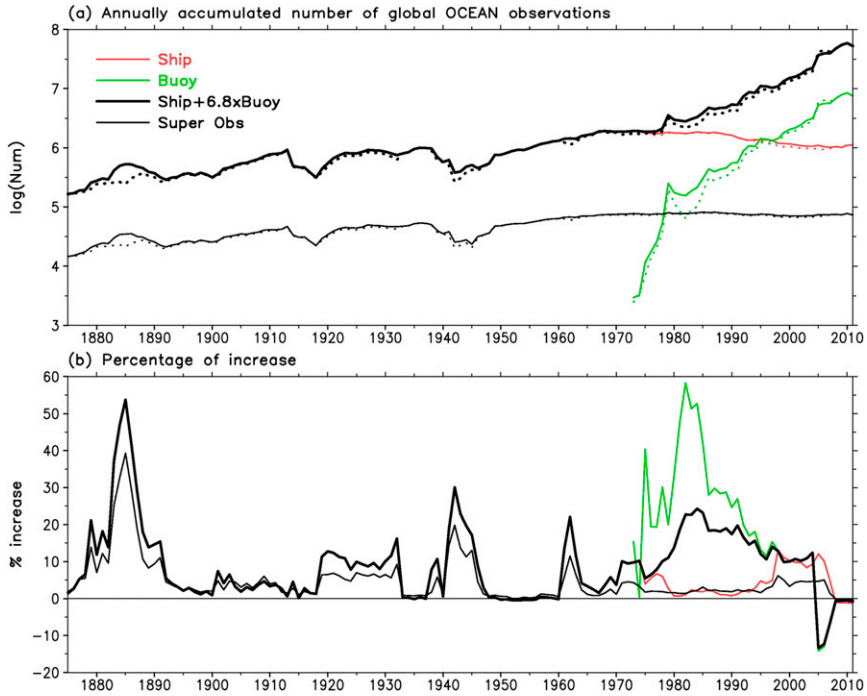

FIG. 1. (a) Annually accumulated number (in log scale) of SST observations by ships (red line) and buoys (green line), equivalent number of combined ship and buoy observations (thick black line), and the number of superobservations on a 2° × 2° grid (thin black line). Solid and dotted lines represent observations selected from ICOADS R2.5 and R2.4, respectively. The factor of 6.8 is determined by the ratio of error variances of ship and buoy observations. (b) As in (a), but for percentage change from R2.4 to R2.5.

3) SEA ICE CONCENTRATION DATA

The sea ice concentrations used to adjust the SSTs over ice-covered areas in ERSST.v4 are from monthly 1° × 1° gridded HadISST data (1870–2010; Rayner et al. 2003) and daily 0.5° × 0.5° gridded NCEP data (2005–present; Grumbine 1996). The NCEP sea ice concentration is adjusted toward HadISST ice concentration by the mean offset during the common period of 2005–10. The ice concentrations are box-averaged to a monthly 2° × 2° grid for ERSST.v4 reconstruction.

4) SPATIALLY COMPLETE DATA TO DERIVE EOT PATTERNS

Monthly SSTs derived from weekly 1° × 1° gridded OISST version 2 (OISST.v2; Reynolds et al. 2002), which is based on in situ and satellite observations, are used between 1982 and 2011 in ERSST.v4 to derive SST STD on a 2° × 2° grid in the QC procedure and to derive EOTs.

b. Datasets used in comparisons to ERSST.v4

Various intercomparisons of ERSST.v4 and the precursor ERSST.v3b are made with other independently derived estimates. SST data, SST bias adjustments, and unadjusted SST data from HadSST3, HadISST, and COBE-SST2 are used to intercompare with ERSST.v4 throughout its record in sections 5 and 6. The HadSST3 data are monthly on a 5° × 5° grid from 1850 to 2012. The HadISST data are monthly on 1° × 1° grid from 1870 to 2012. The SST data of COBE-SST2 are monthly on 1° × 1° grid from 1850 to 2012, and SST bias adjustment data of COBE-SST2 are annually and globally averaged.

The Along-Track Scanning Radiometer (ATSR) satellite SST observations on monthly 1° × 1° grid from 1997 to 2011 (Merchant et al. 2012) are used to evaluate the ERSST.v4 analysis. The ATSR SSTs are adjusted to the water temperature at 20-cm depth (Merchant et al. 2012). All products have been regridded to the common grid of 5° × 5° except where otherwise explicitly noted; and only the data at collocated grids are used in comparisons. The Southern Oscillation index (SOI) using monthly mean sea level pressure anomalies at Tahiti and Darwin (Trenberth 1984) is used to validate the ENSO events in ERSST.v4.

4. Impact assessment of reconstruction upgrades on ERSST.v4 SSTA

The SSTA reconstruction involves many parameter choices within the algorithm used to produce the final

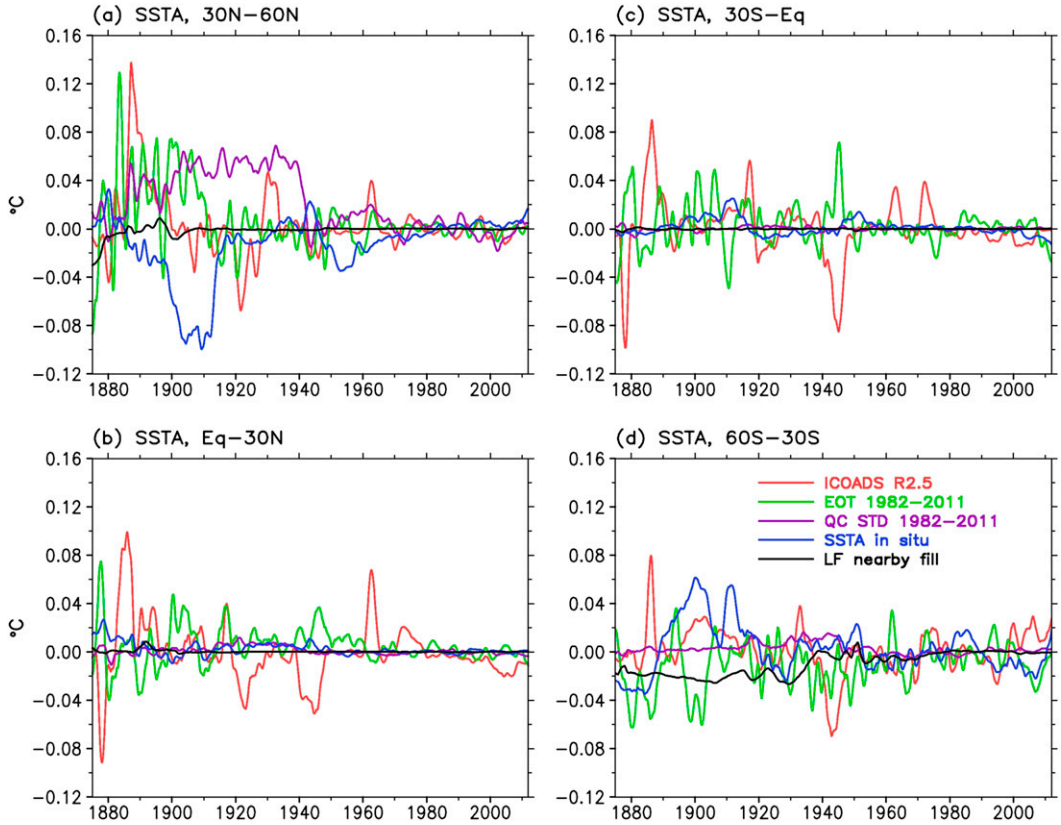

FIG. 2. Areal averaged monthly SSTA difference in (a) 30°–60°N, (b) 0°–30°N, (c) 0°–30°S, and (d) 30°–60°S by changing terms individually from those employed in ERSST.v3b. ICOADS R2.5, EOT 1982–2011, QC STD 1982–2011, SSTA in situ, and low-frequency (LF) nearby fill represent, respectively, the SSTA difference applying R2.5 rather than R2.4, EOTs trained with 1982–2011 SSTs rather than 1982–2005 SSTs, and QC STD from OISST (1982–2011) rather than from COADS (1950–79), SSTA at in situ locations rather than at regular grids, and low-frequency SSTA filled by nearby SSTA observations rather than zero.

SST (Smith and Reynolds 2003; Smith et al. 2008) due to uneven observational data in both space and time. These have been revised wherever deemed necessary in ERSST.v4 using the latest available datasets and improved knowledge and methodologies. Table 1 lists all 11 revisions implemented during data ingest and reconstruction of ERSST.v4. To assess the impacts of each of the individual revisions, test analyses are run progressively by changing one parameter at a time. The mean difference of two or more sets of analyzed SSTAs for one single algorithmic parameter choice are assessed and used as a criterion to select the value of that parameter in the operational version.

a. SST and ice data

As detailed and justified in section 3, the ICOADS R2.5 SST data are used in ERSST.v4, instead of R2.4. The SST data in R2.5 are more complete in early periods, as well as in the recent period due to inclusion of SST observations from delayed-mode sources. Spatial averages of the SSTA differences between the test analyses using R2.5 and R2.4 are small (<0.1°C) most of the time, but they reach up to ±0.1°C in the 1880s (Figs. 2a–d; red lines of "ICOADS R2.5") when data remain sparse (Fig. 1).

The ice concentrations of the latest version from HadISST and NCEP are used in ERSST.v4, whereas previously they were from the Met Office (UKMO; 1870–1980), Goddard Space Flight Center (GFSC; 1981–2004) and NCEP (2005–current) in ERSST.v3b (Smith et al. 2008). Comparisons show that the integrated ice coverage (ice concentration multiplied by grid box area) is approximately 10% lower in HadISST than in the prior UKMO analysis in the Northern Hemisphere oceans, while it is very similar in the Southern Hemisphere oceans. Test analyses show that SSTA changes in the Arctic and Southern Oceans are generally small (<0.1°C) by upgrading the sea ice concentration.

b. Base function EOTs

The high-frequency component of SSTA in ERSST.v4 is reconstructed by projecting the adjusted data fields onto a set of 130 EOTs (localized empirical orthogonal functions) to produce spatially complete estimates. These high-frequency components are key to understanding important modes of variability such as ENSO and how they have changed. The EOTs in ERSST.v4 are trained by OISST.v2 between 1982 and 2011 instead of between 1982 and 2005 as in ERSST.v3b. The spatial structures of the updated EOTs are similar to those used in ERSST.v3b except that the order of EOTs is different because the variance explained by specific EOTs is changed due to the addition of six new years of observations.

Test analyses show that, by revising EOTs, the area averaged SSTA changes are mostly less than 0.1°C between 30° and 60°N in the northern North Pacific and northern North Atlantic before about 1910 when observations are sparse, and they change little afterward when data coverage becomes more complete (Fig. 2a; green line of "EOT 1982–2011"). More importantly, the tests show that the analysis using the EOTs trained using 1982–2011 data resolves the El Niño in 1878 (Fig. 3a; red line) as suggested by the SOI index (Fig. 3a; dotted line), whereas the analyses using the EOTs trained in 1982–2005 and 1988–2011 (Fig. 3a; black and greens lines that mostly overlap) fails to resolve this event.

The criterion (Crit) of variance ratio [Eq. (4)], which is used to accept a specific EOT mode, is set to 0.1 in ERSST.v4, while it was set to 0.2 in ERSST.v3b. Crit is effectively a measurement of data completeness that avoids giving undue weighting to a given EOT due to a grossly inadequate observational constraint. As such, this parameter is only important in the early record or in persistently data-sparse regions such as high-latitude oceans. The number of accepted EOTs is approximately 110 in between 1870s and 1880s, and above 120 after 1900 except for the late 1910s (as low as 110) and between 1940 and 1950 (as low as 100).

The reason for lowering the Crit value is to better represent the El Niño/La Niña events and other variability in the period prior to the early twentieth century when sampling is sparse. This choice is quantified and justified by undertaking test analysis from 1960 to 2012 using historical observational masks (partial sampling) from 1860 to 1912 (e.g., the 1998 ICOADS R2.5 data field is reduced to its data coverage mask of 1898). The test analysis using the actual observational mask (full sampling) from 1960 to 2012 is used as a "truth," since the well-sampled analysis is not sensitive to the slight changes in the EOT training period or Crit selections because the EOTs are fully constrained by the dense observations. The tests show that the analysis with a lower Crit of 0.1 is closer to the truth than that with a higher Crit of 0.2 in the Niño-3.4 region (5°S–5°N, 120°–170°W) (Fig. 3b), with several El Niño/La Niña events better recreated with a lower Crit value than that used in ERSST.v3b. The difference in Niño-3.4 indices between final ERSST.v4 and preceding v3b can be seen clearly before 1970 and particularly prior 1900 (Fig. 3c). The assessment of other regional averaged common indices also indicates (not shown) that a lower Crit of 0.1 better represents the truth. These common indices include the IOD, PDO, North Atlantic Hurricane Main Development Region (HMDR) SST, and global averaged SST. However, the North Atlantic AMO index degrades slightly when a lower Crit is selected. Based on these assessments, the Crit of 0.1 is selected but is not lowered further because the analyzed SSTAs in the midlatitude oceans become noisy when Crit is set to be 0.05.

It should be noted, however, that resolving SST variability in the tropical oceans has a trade-off in some other regions in the high-latitude oceans, which is assessed by root-mean-square-difference (RMSD) between monthly SSTAs of partially and fully sampled experiments from 1960 to 2012. The global averaged RMSD is 0.40°C when Crit is set to 0.1. In contrast, the global averaged RMSD increases to 0.51°C when Crit is set to 0.05. However, the global averaged RMSD decreases slightly to 0.37°C when Crit is set to 0.2. It appears that there is no single correct representation for the value of Crit (see Part II). ERSST.v4 is used for myriad applications, many of which, such as ENSO monitoring by NOAA Climate Prediction Center (CPC), require fidelity in Niño-3.4 more than the global mean. Therefore, a slight increase of global averaged RMSD is deemed an acceptable trade-off and the Crit is lowered from 0.2 of ERSST.v3b to 0.1 in ERSST.v4.

The use of a weighting function $w(N, \varepsilon)$ [Eq. (3)] is necessary to account for the difference in errors from ship and buoy observations and different density of observations in ERSST.v4, whereas it was set to be 1 in ERSST.v3b. Rather than giving each grid box equal weight in determining the ordering and weighting of EOTs, the updated weighting approach gives greater weight to grid boxes containing either a greater data density and/or data of lower random error characteristics. It is well known that buoy-based observations exhibit lower random errors than ship-based observations (Reynolds and Smith 1994), a fact that may be important now and moving forward given the significant shift from mainly ship-based to mainly buoy-based observations over the last two decades. Test analyses show, however, that there is little difference in global or regional average

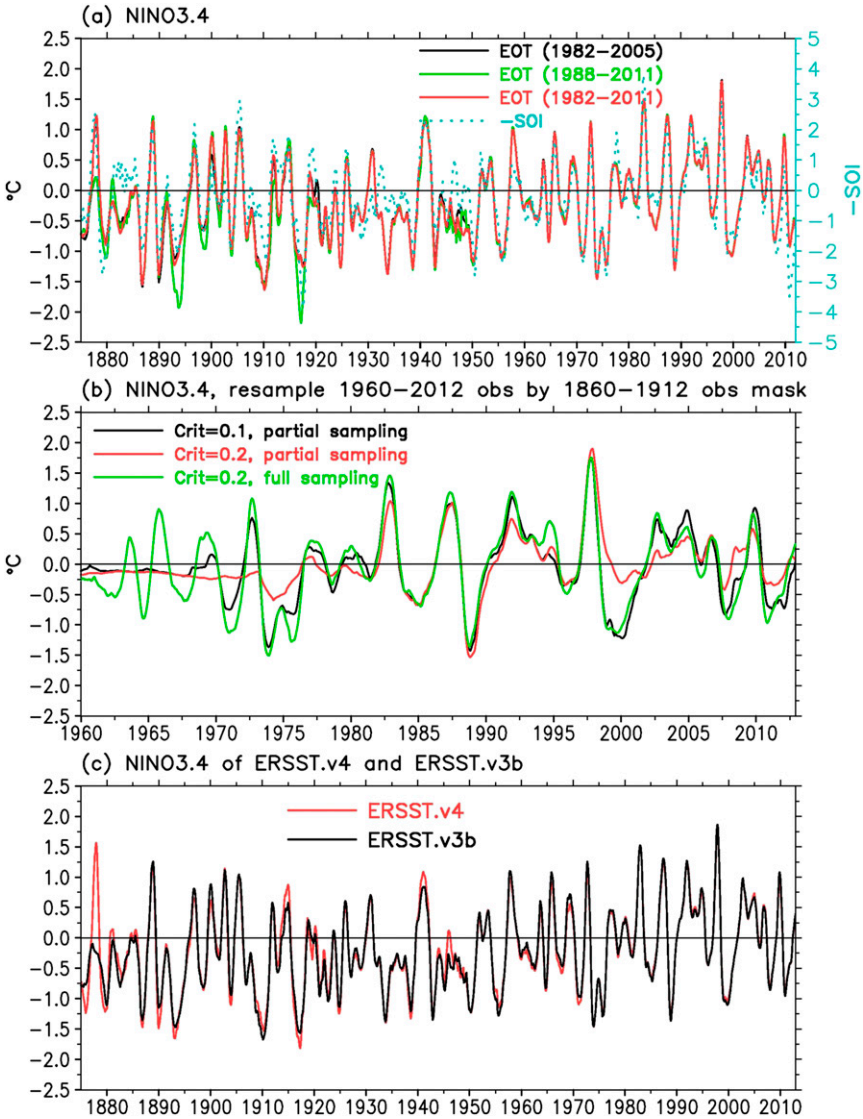

FIG. 3. (a) Niño-3.4 index (left axis) in test analyses using EOTs trained with 1982–2005, 1988–2011, and 1982–2011 data, overlapped with the SOI index (right axis). (b) Niño-3.4 index from 1960 to 2012 in test analyses using Crit of 0.1 and 0.2 and using EOTs trained with 1982–2011 data, when observed data are resampled by observational mask from 1860 to 1912. The Niño-3.4 in full sampled analysis is overlapped. (c) Niño-3.4 index of ERSST.v4 and ERSST.v3b.

behavior in the final SST product due to these changes. This may be because the correlated uncertainty in input SST data (Kennedy et al. 2011) is still not explicitly taken into account. If this correlated uncertainty were to have a significant impact upon EOT ordering or weighting, then it might yield larger changes in the final SST product. However, its inclusion is nontrivial within the ERSST framework and requires substantive further investigation.

c. *SST quality control and SSTA quantification*

The SST data are first screened using a QC procedure checking the differences between observations and first-guess SSTs from ERSST.v3b. Those observations are rejected when they deviate from the first guess by more than 4 times STD. In ERSST.v4, the monthly SST STD is calculated using the weekly OISST.v2 from 1982 to 2011. It was calculated in ERSST.v3b using the original COADS (Woodruff et al. 1987) from 1950 to 1979, but COADS lacks many improvements made subsequently under what is now the ICOADS project. Since the annual averaged STD is 1° to 1.5°C higher in COADS than in OISST.v2 in the western North Pacific and western North Atlantic, fewer cold SST data during the wintertime are accepted in ERSST.v4 by the QC

procedure in those regions before 1940. Therefore, averaged SSTAs between 30° and 60°N before 1940 are approximately 0.05°C warmer in the test analysis using OISST.v2 STD than that using COADS STD (Fig. 2a; purple line of "QC STD 1982–2011"). It is likely that the ERSST.v4 reconstruction system has excluded more extreme cold observations by using OISST.v2 STD, but it may also be a risk to include these extreme cold observations since their reliability is suspect.

In ERSST.v3b, SSTA was calculated by subtracting the monthly climatology between 1971 and 2000 after full SSTs are bin-averaged to the 2° × 2° grid. This can result in an inaccurate SSTA in data-sparse areas in higher-latitude oceans due to coarse latitudinal resolution, since the SSTA may be partially impacted by the climatological SST if SST observations are not representative of the grid box average. Following Reynolds and Smith (1994) and Kennedy et al. (2011), SSTAs are now initially calculated at in situ locations by subtracting SST climatology interpolated to the in situ locations, and then the in situ SSTAs are bin-averaged to the monthly 2° × 2° grid. The test analyses show that the analyzed area averaged monthly SSTAs can differ by 0.1°C. For example, the SSTA decreases by 0.1°C between 30° and 60°N from around 1890 to about 1920 (Fig. 2a; blue line of "SSTA in situ"), and increases by 0.05°C between 30° and 60°S from around 1890 to about 1910 (Fig. 2d).

d. *Low-frequency anomaly filling*

In reconstructing the low-frequency component of SSTA, grid boxes without in situ observations were filled with zeroes in ERSST.v3b. This implicitly makes the SSTs in data-sparse regions and epochs similar to their climatological period average (1971–2000). Under a transient climate change where the climatological subperiod is not necessarily representative of the whole era of record, this may act to artificially warm (cool) the SSTA in the earlier (later) periods, particularly when and where observations are sparse, SST changes have been rapid, or multidecadal variability is marked. In ERSST.v4, instead of zero-filling, the average of neighboring valid proximal SSTAs is used to fill the grid box that originally contains a missing value (Fig. 2d; black line of "LF nearby fill"). The nearby fill cools the Southern Ocean south of 30°S slightly (0.02°C) prior to about 1940 (Fig. 2d). South of 60°S, the SSTAs decrease by 0.2° to 0.4°C before the 1940s (not shown). North of 60°N, the SSTAs increase by 0.2° to 0.6°C after the 1930s. Therefore, the SSTA trend increases by 0.4° to 0.6°C century^{-1} south of 60°S and north of 60°N, although the global averaged SSTA trend is changed little (less than 0.02°C century^{-1}). The nearby fill method used here is not the only way to fill the missing SSTAs. For example, SSTAs are filled using coarse-resolution empirical orthogonal functions in COBE-SST2 (Hirahara et al. 2014).

5. SST bias adjustment

a. *Ship SST bias adjustment*

Historically, SSTs have mostly been observed by commercial, naval, and research ships primarily using various buckets or ship engine room intake (ERI) and hull contact sensors after the World War II (WWII) era (Kennedy et al. 2011). These SST data exhibit marked time-varying systematic biases throughout the record due to changes in observation methods and instruments. The changes in ship deck heights also contribute to the SST biases when SSTs are observed by buckets. The SSTs measured using bucket samples are generally lower than the "true" SSTs due to the heat loss from buckets exposed in air during the hauling and positioning of buckets on the ship deck (Kent and Kaplan 2006). In contrast, the SSTs observed by ERI are mostly higher than true SSTs due to warming from the engine room, although sometimes the ERI measurements are lower than true SST (Kent and Kaplan 2006). Ship SSTs should therefore be adjusted to minimize such artificial variations where they can be identified and quantified.

The bias adjustment for ship SSTs in ERSST.v4 is originally proposed by Smith and Reynolds (2002) and involves using NMAT as a reference. NMAT is selected because the differences from SST are more stable than daytime marine air temperatures, which can have a large range due to solar heating of the ships decks and of the instruments themselves. To formulate the bias adjustment, however, it is necessary to assume that

1) the difference between SST and NMAT is near constant during the climatological period (1971–2000);
2) the climatological difference of SST and NMAT is constant in other periods;
3) the NMAT is less biased (more homogeneous) than the SST data to which it is being compared;
4) the mix of SST measurement methods (bucket or ERI) is invariant across the global oceans, and the spatial pattern of biases follows the climatological difference of SST and NMAT in the modern time (1971–2000); and
5) biases vary relatively slowly and smoothly with time.

To test the first two assumptions, which are assuming broad physical coherence between two highly correlated but physically distinct measurands, the average difference between SST and near-surface air temperature (SAT) of day and night at 2 m is calculated by subsampling monthly

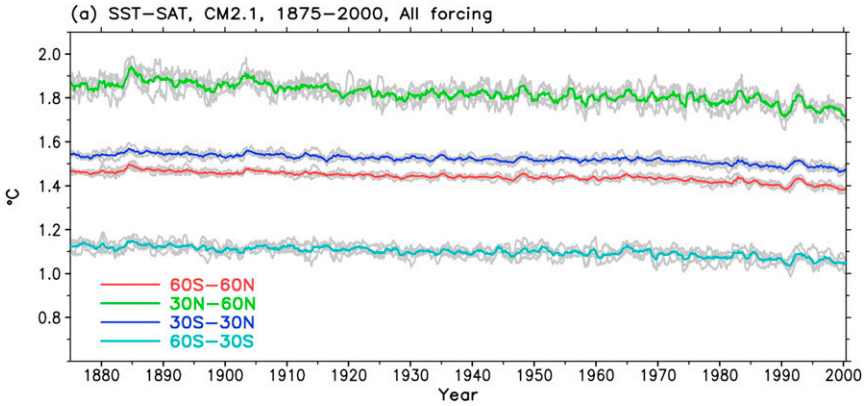

FIG. 4. Ensemble average (colored lines) and five ensemble members (gray lines) of monthly SST and SAT from subsampled simulation of the GFDL coupled model (CM2.1) using monthly historic observation masks from 1875 to 2000 in regions of 60°S–60°N, 30°–60°N, 30°S–30°N, and 60°–30°S. A 12-month running mean filter has been applied. Linear trends are −0.08°, −0.05°, −0.04°, and −0.04°C century^{-1} between 1875 and 2000 for averages over 60°S–60°N, 30°–60°N, 30°S–30°N, and 60°–30°S, respectively.

outputs of the GFDL CM2.1 coupled model with monthly observation masks from 1875 to 2000 (Fig. 4). The model SAT is used since the model bias is assumed to be the same during daytime and nighttime. It is found that the first two assumptions are valid since the model simulations indicate that the difference of SST and SAT is near constant and its linear trend is weak in all four different latitudinal zones (Fig. 4). The slight tendency (less than 0.08°C century^{-1}) of NMAT-SST indicates that NMAT increases faster than SST; bias adjustment may be slightly underestimated in the early period; and therefore the global averaged SST trend may have been slightly overestimated in ERSST.v4, which may partially contribute to the difference of global SST trends between ERSST.v4 and HadSST3 shown in Table 2. However, the potential overestimation of global averaged SST trend (0.08°C century^{-1}) falls within the 95% confidence level (0.11°C century^{-1}; Table 2).

The third assumption regarding NMAT homogeneity being greater than SST homogeneity is tentatively valid since the instruments and methods used to observe NMATs are persistent relative to those used to observe SSTs. However, we note that changes in the instruments observing NMATs were found (Kent et al. 2007), and air temperature sensors may not have been adequately exposed during the latter nineteenth century and the WWII era (when taking a night measurement on deck was considered especially dangerous) (Kent et al. 2013). Even if the data were perfect, the observed NMAT, however, may still be biased mostly due to changes in ship deck height in historic NMAT observations. Observations indicate that ship deck heights have become progressively higher over time as ships themselves have, on average, become larger; and this introduces a sampling artifact that acts to introduce a cooling effect into the record given that atmospheric temperature decreases with height near the ocean surface. This spurious cooling bias relative to the true NMAT at an invariant nominal vertical datum has been adjusted according to individual ship height metadata and shipping fleet characteristics (Kent et al. 2013).

According to the fourth assumption, the SST bias adjustment can be formulated following Smith and Reynolds (2002):

$$d_{x,m,y} \equiv \mathrm{SST} - \mathrm{NMAT} = A_{m,y} C_{x,m}, \tag{5}$$

$$B_{x,m,y} \equiv (\overline{A} - A_y) C_{x,m}, \tag{6}$$

where $d_{x,m,y}$ represents the monthly difference SST − NMAT at location x in month m and year y; $C_{x,m}$ is the monthly climatological difference for SST − NMAT; $A_{m,y}$ is monthly fitting coefficients; and both $C_{x,m}$ and $A_{m,y}$ are updated based on the latest HadNMAT2. Also,

TABLE 2. Ordinary least squares linear trends (in unit of °C century^{-1}) and their uncertainty (95% confidence level) of annually averaged SSTAs from 1901 to 2012 in ERSST.v4, ERSST.v3b, HadSST3, and COBE-SST2. Trend uncertainties have been calculated such as to account for AR(1) effect on the degrees of freedom (von Storch and Zwiers 1999).

	ERSST.v4	ERSST.v3b	HadSST3	COBE-SST2
60°S–60°N	0.73 ± 0.11	0.71 ± 0.10	0.67 ± 0.11	0.73 ± 0.12
30°–60°N	0.57 ± 0.18	0.55 ± 0.18	0.68 ± 0.18	0.71 ± 0.19
30°S–30°N	0.76 ± 0.14	0.73 ± 0.12	0.59 ± 0.13	0.70 ± 0.13
60°–30°S	0.81 ± 0.13	0.88 ± 0.10	0.93 ± 0.13	0.87 ± 0.11

$B_{x,m,y}$ is defined as the SST bias adjustment to be added to the historic SST observations, where A_y is an annually averaged coefficient of $A_{m,y}$; and $\overline{A}$ is the climatological (1971–2000) fitting coefficient of A_y. In ERSST.v3b, NMAT from ICOADS R2.4 was used, while HadNMAT2, which includes deck height corrections and additional QC procedures, is used in ERSST.v4 to calculate the SST bias adjustment on a 5° × 5° grid. Later comparisons will show that the bias adjustment based on a single fitting coefficient over the global oceans [Eqs. (5) and (6)] is consistent with other independent estimates. However, tests show that the bias adjustment may differ when the fitting coefficients are assessed separately in different latitudinal belts in 30°–90°N, 30°S–30°N, and 30°–90°S, although the global averaged bias adjustment does not change. Such regional fitting naturally yields somewhat noisier estimates given the reduced sample sizes of NMAT and SST collocations compared to creating a global-based estimate. Consistent with other aspects of the method we prefer the global fit as it is likely to on average be more robust to sampling effects by averaging over the largest possible sample. However, the local biases may have been overly smoothed by fitting the SST and NMAT differences over the global oceans.

The monthly fitting coefficients (gray lines) are shown in Fig. 5, which overall fits the fifth assumption that the biases vary slowly with time. To filter out potentially spurious high-frequency noise in the fitting coefficients, a linearly fitted coefficient was used in ERSST.v3b (Smith and Reynolds 2002). Subsequent to ERSST.v3b several analyses have highlighted the likely presence of substantive multidecadal bias variability throughout the record (e.g., Kennedy et al. 2011) rather than simply around the transition from mainly buckets to mainly ERI measures around the early 1940s. In ERSST.v4, a Lowess filter (Cleveland 1981) has been applied on A_y (Fig. 5) and allowed to vary the bias adjustments throughout the record. A filter coefficient of 0.1 is applied to the Lowess, which is equivalent to a low-pass filter of 16 years and represents the low-frequency nature of the required bias adjustment. The reason to apply a filter is to make the bias adjustment smoother so that it may be more consistent with the assumption of applying a climatological SST − NMAT pattern of $A_{m,y}$. However, we stress that higher-frequency changes in SST biases are virtually certain to exist as indicated in Thompson et al. (2008), Kennedy et al. (2011), and Hirahara et al. (2014). Shorter windows or use of annually averaged data would be noisier by construction because the estimate at any given point would be based upon a smaller sample and it is not clear at what point there becomes a risk of fitting to random sampling noise

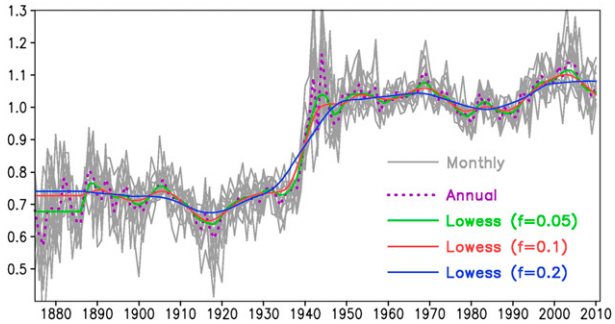

FIG. 5. Monthly fitting coefficients between transient SST − NMAT difference and climatological SST − NMAT difference. Gray lines represent 12 monthly coefficients; dotted lines represent the annual averaged coefficient; colored solid lines indicate filtered coefficients with a Lowess parameter value of 0.05, 0.1, and 0.2, which represent low-frequency filter of approximately 8, 16, and 32 yr, respectively.

rather than systematic bias signal. The preference is for robust estimation of the multidecadal component of the bias adjustments using a coefficient of 0.1 but may come at a cost of accurately portraying biases at times of rapid transition (e.g., the WWII era). The coefficient A_y is set to be the value of 1886 before 1886 and the value of 2010 after 2010 (the final year of the HadNMAT2 dataset at the time of analysis). Kent et al. (2013) cautioned against use of pre-1886 HadNMAT2 for long-term trend analyses. The bias adjustments estimated on the 5° × 5° grid are bilinearly interpolated to our 2° × 2° grid and applied to ERSST.v4.

b. *Comparison of ship SST bias adjustments*

Figure 6 compares the average ship SST bias adjustments in four latitudinal zones for ERSST.v4 and v3b (Smith and Reynolds 2002), HadSST3 (Kennedy et al. 2011), and COBE-SST2 (Hirahara et al. 2014) from 1875 to 2006. The seasonal variation of bias adjustment, which is included in ERSST.v4, v3b, and HadSST3 reconstructions, has been filtered out in Fig. 6 using a 12-month running mean. The bias adjustment between 60°S and 60°N in ERSST.v4 (Fig. 6a) is approximately 0.3°C in the 1870s, increases slightly to 0.4°C in the 1920s, and drops to near 0°C in the mid-1940s. In comparison with ERSST.v3b, the bias adjustments in ERSST.v4 are slightly stronger before about 1920, 0.1°–0.2°C weaker between 60°S and 60°N from around 1920 to 1940 (Fig. 6a), and approximately 0.1°C stronger between 30° and 60°S before about 1940 (Fig. 6d). The bias adjustment was assumed to be zero after 1941 in ERSST.v3b. In contrast, the bias adjustment is explicitly calculated in ERSST.v4 throughout the record, and there is a negative adjustment around 2000 that is consistent with a peak in the fitting coefficient of $A_{m,y}$ (Fig. 5), which may result

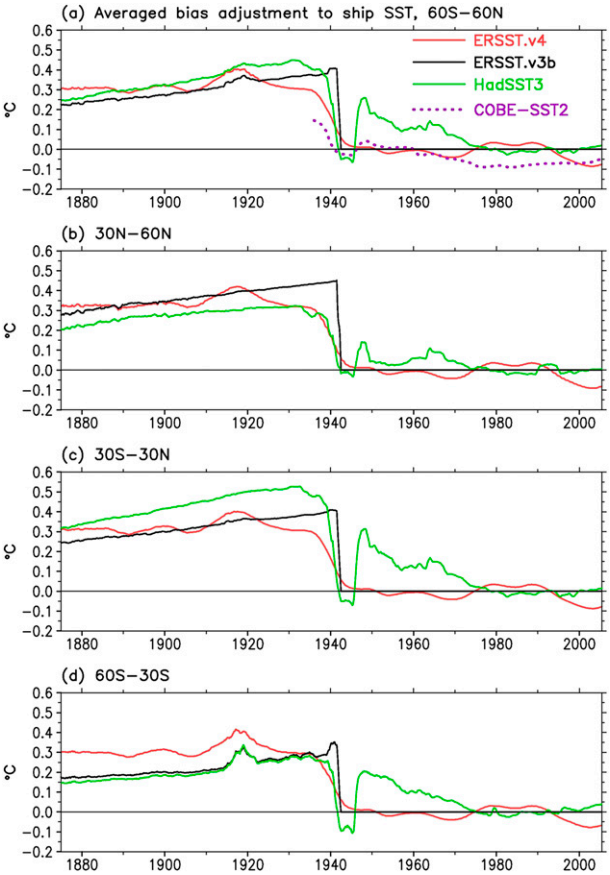

FIG. 6. Collocated monthly bias adjustment to ship SST in ERSST.v4, ERSST.v3b, and HadSST3 in (a) 60°S–60°N, (b) 30°–60°N, (c) 30°S–30°N, and (d) 60°–30°S. A 12-month running mean is applied. Annually and globally averaged bias adjustment of COBE-SST2 from 1936 to 2006 in (a) is adapted from Hirahara et al. (2014).

from a larger difference of SST and NMAT due to an increased number of ship ERI observations (Hirahara et al. 2014).

Spatial differences in bias adjustments between ERSST.v4 and ERSST.v3b are evident from 1880 to 1935 (Figs. 7a,b). The bias adjustment is slightly weaker in ERSST.v4 (0.3° to 0.4°C) than in ERSST.v3b (0.3° to 0.5°C) in most oceans, but is slightly stronger in ERSST.v4 than in ERSST.v3b in the central-eastern equatorial Pacific and Southern Ocean south of 45°S. Despite these differences, a common feature is that the bias adjustment is relatively large in the western North Pacific and western North Atlantic in both ERSST.v4 and ERSST.v3b. The large bias adjustment in those regions is associated with enhanced heat loss due to evaporation of the water from buckets exposed in air during the hauling and positioning of buckets on the ship deck. The bias adjustment is also large in the western tropical Pacific in both ERSST.v4 and ERSST.v3b, which might be associated with higher SSTs that contribute to the larger latent heat loss.

In comparison with HadSST3, which is not globally complete, the collocated bias adjustment in ERSST.v4 is approximately 0.1°C higher in the midlatitude oceans (30°–60°N and 30°–60°S) before around 1930 (Figs. 6b,d), but is 0.1° to 0.2°C lower in the tropics (30°S–30°N) before around 1940 (Fig. 6c) and south of 30°N from the mid-1940s to about 1970 (Figs. 6c,d). The bias adjustment between 60°S and 60°N is approximately 0.1°C weaker in ERSST.v4 than in HadSST3 from approximately 1920 to 1940 and from mid-1940s to around 1970 (Fig. 6a). In contrast to a near-zero bias adjustment in ERSST.v4 in the vicinity of the mid-1940s, a weak negative adjustment is made south of 30°N in HadSST3. The negative adjustment in the 1940s may be associated with a warm bias of SST observations by ship ERI during the WWII era (Kennedy et al. 2011). However, the negative adjustment in the 1940s is not explicitly identified in ERSST.v4, but caution is needed since both SST and NMAT observations are extremely uncertain due to small numbers of observations and SST reconstruction may further be complicated by the ENSO event in the early 1940s (refer to Fig. 3a). Furthermore, as discussed above, the use of a Lowess filter of 0.1 will damp the ability to resolve biases that occur more rapidly than the filter width of 16 years. The stronger bias adjustment between 30°S and 30°N in HadSST3 can be seen in averaged bias adjustments from 1880 to 1935 (Fig. 7c), which is 0.4°C to 0.5°C in most of the tropical oceans, particularly in the western tropical North Pacific, tropical Atlantic, and Indian Ocean. In contrast, the bias adjustment in ERSST.v4 (Fig. 7a) is generally 0.3° to 0.4°C in the tropical oceans.

The reason for the differences in bias adjustments between ERSST.v4 and HadSST3 is not easily discerned, since the algorithms of the bias adjustment in ERSST.v4 and HadSST3 are completely different and independent. In HadSST3, the bias adjustment is assessed based on a data deck dependence and measurement metadata where available, and the bucket corrections pre-1942 are based upon a physical model and climatological atmospheric conditions. In ERSST.v4, the bias adjustment is based on statistical fitting coefficient of $A_{m,y}$ and global climatological difference $C_{x,m}$ of SST and NMAT, which does not explicitly involve individual SST metadata. The estimated heat loss from buckets in HadSST3 is not only associated with the air–sea temperature difference but also with surface wind speed, relative humidity, solar radiation, and ship speed, which is another source of the differences in bias adjustment. It is likely that the difference of the bias adjustments between ERSST.v4 and HadSST3 represents

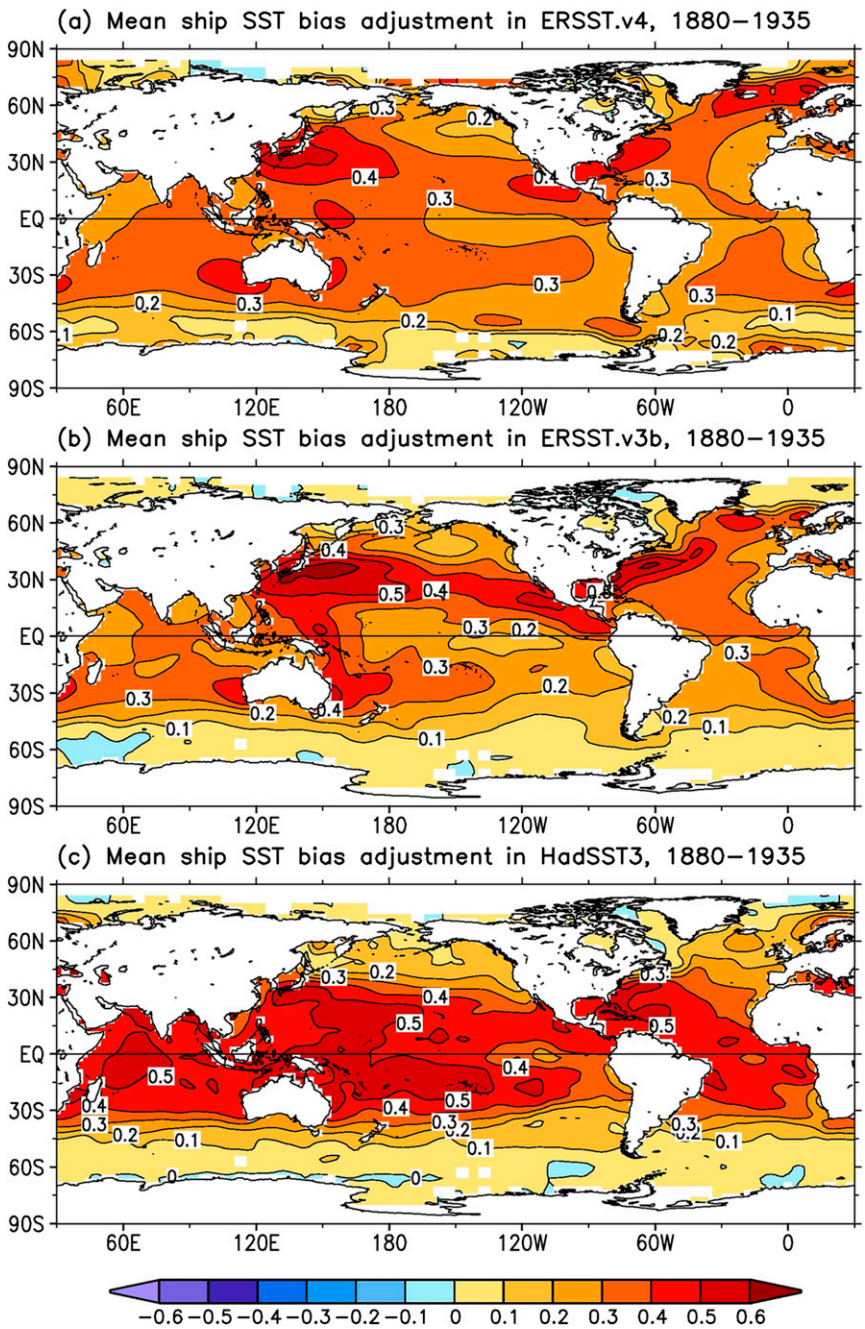

FIG. 7. Collocated average bias adjustment between 1880 and 1935 in (a) ERSST.v4, (b) ERSST.v3b, and (c) HadSST3. Contour intervals are 0.1°C.

some component of the possible uncertainty in SST bias adjustment arising from reasonable methodological choices as suggested by Smith et al. (2008).

The large bias adjustments before the WWII era in both ERSST.v4 and HadSST3 are directly associated with most observations using buckets, particularly uninsulated buckets [refer to Fig. 2 of Kennedy et al. (2011) and Fig. 1 of Hirahara et al. (2014)]. After WWII, the globally averaged bias adjustment in HadSST3 decreases gradually, which appears to be consistent with a gradual decrease of bucket observations. In contrast, the global averaged bias adjustment in ERSST.v4 is near zero after the WWII era (Fig. 6a). The weak bias adjustment in ERSST.v4 after the 1940s may partially be associated with ERI observations that have warming bias cancelling with the cooling bias of bucket observations.

The weaker bias adjustment after the 1940s is also seen in COBE-SST2 (Fig. 6a). The global averaged bias adjustment in COBE-SST2 is weak in the 1950s, which appears to be due to the cancellation of the biases of buckets and ERI observations (Fig. 4b in Hirahara et al. 2014), and is slightly negative after the 1970s. It should be noted that the temporal variations of the bias adjustments in HadSST3 and COBE-SST2 are very consistent, which suggests that the variations of the reconstructed SSTAs would be very consistent after the SSTAs are shifted to have zero mean over the climatological period of 1971–2000.

c. *Ship-buoy SST adjustment*

In addition to the ship SST bias adjustment, the drifting and moored buoy SSTs in ERSST.v4 are adjusted toward ship SSTs, which was not done in ERSST.v3b. Since 1980 the global marine observations have gone from a mix of roughly 10% buoys and 90% ship-based measurements to 90% buoys and 10% ship measurements (Kennedy et al. 2011). Several papers have highlighted, using a variety of methods, differences in the random biases, and a systematic difference between ship-based and buoy-based measurements, with buoy observations systematically cooler than ship observations (Reynolds et al. 2002, 2010; Kent et al. 2010; among others). Here the adjustment is determined by 1) calculating the collocated ship-buoy SST difference over the global ocean from 1982 to 2012, 2) calculating the global areal weighted average of ship-buoy SST difference, 3) applying a 12-month running filter to the global averaged ship-buoy SST difference, and 4) evaluating the mean difference and its STD of ship-buoy SSTs based on the data from 1990 to 2012 (the data are noisy before 1990 due to sparse buoy observations). The mean difference of ship-buoy data between 1990 and 2012 is 0.12°C with a STD of 0.04°C (all rounded to hundredths in precision). The mean difference of 0.12°C is at the lower end of published values of 0.12° to 0.18°C (e.g., Reynolds et al. 2002, 2010; Kent et al. 2010). Although buoy SSTs are generally more homogeneous than ship SSTs, they are adjusted here because otherwise it would be necessary to adjust ship SSTs before 1980 when there were no or very few buoys. As expected, the global averaged SSTA trends between 1901 and 2012 (refer to Table 2) are the same whether buoy SSTs are adjusted to ship SSTs or the reverse. However, the global mean SST is 0.06°C warmer after 1980 in ERSST.v4 because of the buoy adjustments (not shown) and there are therefore impacts on the long-term trends compared to applying no adjustment to account for the change in observational platforms.

6. SSTA comparisons

The SSTAs of ERSST.v4 from 1875 to 2012 are compared with those of ERSST.v3b, HadSST3, and COBE-SST2 to evaluate the consistency of the products. To make the comparisons, SSTAs of COBE-SST2 are derived relative to its own SST climatology of 1971–2000, and box-averaged to the 5° × 5° resolution of HadSST3; SSTAs of HadSST3 relative to their 1961–90 climatology have been adjusted to the 1971–2000 climatology base period used for ERSST.v4 and ERSST.v3b. The SSTAs of 2° × 2° resolution in ERSST.v4 and ERSST.v3b are also box-averaged to 5° × 5° resolution. The regional and temporal averages have been calculated based on collocated data of all four products (HadSST3 has lower coverage as it is uninterpolated) to ensure that mismatches do not arise solely from differences in coverage. Remaining differences could arise from the effects of the different choices in data selection, QC, adjustments, and/or whether, and if so how, to apply spatiotemporal filtering as part of the processing.

Figure 8 shows averaged SSTAs in four latitudinal zones. The consistency among the four products is apparent in the overall variations of SSTAs in different latitudinal zones on interannual to decadal time scales. The consistency is much greater after about 1970. The greater consistency may largely be due to the high density of observations supporting SST bias adjustments and reconstructions using different methods, and may partially be that SSTs are forced to have the same climatology between 1971 and 2000. Overall, SSTAs are higher in HadSST3 (green line) than in ERSST.v4 (red line) in the tropics, while SSTAs are lower in HadSST3 than in ERSST.v4 in the midlatitudes. For example, the SSTAs of HadSST3 are 0.1°C to 0.2°C higher than those of ERSST.v4 between 30°S and 30°N from about 1905 to the mid-1930s and from the mid-1940s to about 1970 (Fig. 8c), and between 30° and 60°S from the mid-1940s to around 1960 (Fig. 8d). In contrast, the SSTAs of ERSST.v4 are approximately 0.1°C higher than those of HadSST3 north of 30°N before around 1930 (Fig. 8b), and south of 30°S before the mid-1910s (Fig. 8d). Further comparisons show that differences between ERSST.v4 and HadSST3 result primarily from the differences in SST bias adjustments rather than differences in the source data. The differences of the SSTAs prior to the bias adjustments (“Unadjusted”; Fig. 9) are an order of magnitude smaller than those in the final adjusted data (“Adjusted”; Fig. 9). The SSTAs prior to the bias adjustments in ERSST.v4 and HadSST3 are derived using the same observational dataset (ICOADS) used in the constructions of ERSST.v4 and HadSST3, respectively.

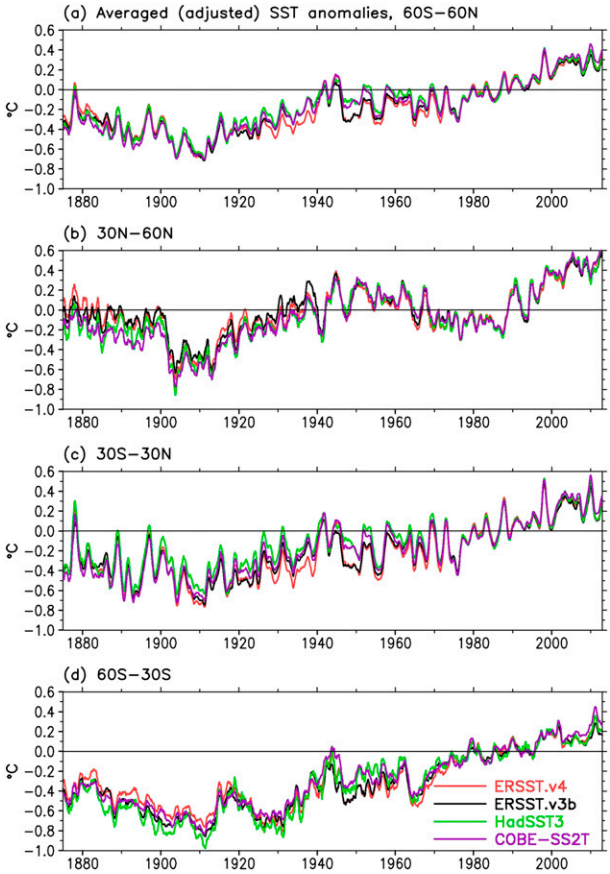

FIG. 8. Collocated monthly SSTA of ERSST.v4, ERSST.v3b, HadSST3, and COBE-SST2 in (a) 60°S–60°N, (b) 30°–60°N, (c) 30°S–30°N, and (d) 60°–30°S. A 12-month running mean has been applied.

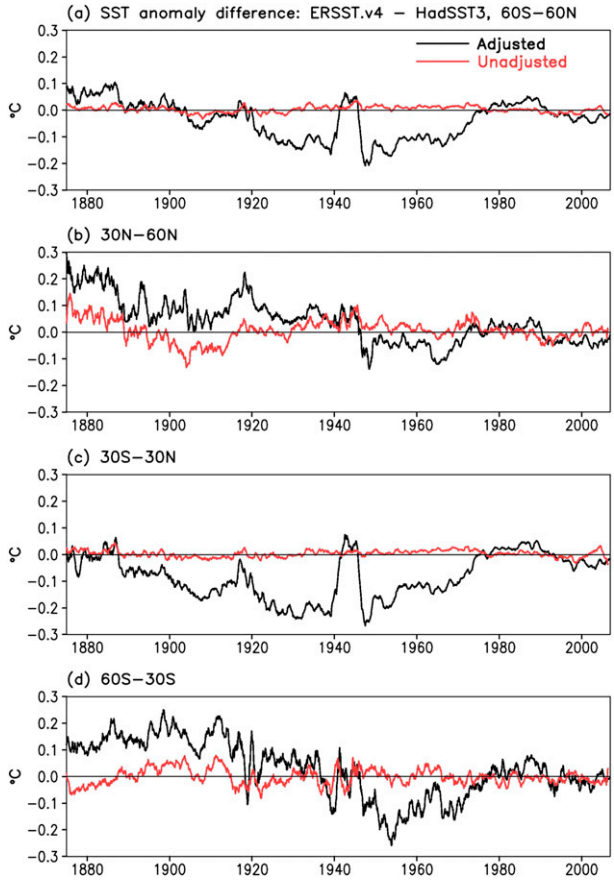

FIG. 9. Differences of collocated monthly "adjusted" and "unadjusted" SSTAs between ERSST.v4 and HadSST3 in (a) 60°S–60°N, (b) 30°–60°N, (c) 30°S–30°N, and (d) 60°–30°S. A 12-month running mean has been applied.

The differences between ERSST.v4 and ERSST.v3b (black lines; Fig. 8) are also noteworthy. The SSTAs are approximately 0.1°C lower in ERSST.v4 than in ERSST.v3b between 30°S and 30°N from the mid-1920s to about 1940 (Fig. 8c), and 0.1°C higher south of 30°S before the mid-1910s (Fig. 8d). The SSTAs of COBE-SST2 are generally closer to HadSST3 than to ERSST.v4 because of the similarity of SST bias adjustment in HadSST3 and COBE-SST2. In contrast, the SSTAs of COBE-SST2 are slightly warmer than HadSST3, and closer to ERSST.v4 in the Southern Ocean south of 30°S before the 1940s (Fig. 8d). It should be noted that the global averaged SSTAs between COBE-SST2 and HadSST3 are very close from the late 1940s to the 1960s (Fig. 8a), while the globally averaged bias adjustment is lower in COBE-SST2 than in HadSST3 (Fig. 6a). The reasons for the apparent inconsistency may be that the adjustments are collocated in ERSST.v4 and HadSST3 but not in COBE-SST2.

The SSTA differences shown in Fig. 8 have an impact on the estimation of long-term SST trends (Table 2). For example, the linear trends of averaged SSTA between 60°S and 60°N from 1901 to 2012 are 0.73°, 0.71°, 0.67°, and 0.73°C century^{-1} in ERSST.v4, ERSST.v3b, HadSST3, and COBE-SST2, respectively. The slightly stronger trend in ERSST.v4 and COBE-SST2 is associated with the lower SSTA from 1925 to 1970 in ERSST.v4 and COBE-SST2 and the higher SSTA after 2000 in COBE-SST2 (Fig. 8a). The lower SSTA from 1925 to 1970 in ERSST.v4 in turn is associated with the weaker SSTA bias adjustment shown in Fig. 6a. The starting year of 1901 is selected because of the greater data coverage after that time. The trend estimates will vary depending upon the chosen start date and end date given that the series does not change linearly through time.

Spatial analysis also aids in understanding the reasons behind differences between ERSST.v4 and the preceding v3b, HadSST3, and COBE-SST2. For this purpose time averaged SSTAs are compared when their differences are relatively large, for example, from the 1910 to 1935 (Fig. 8). Spatial consistencies in time averaged

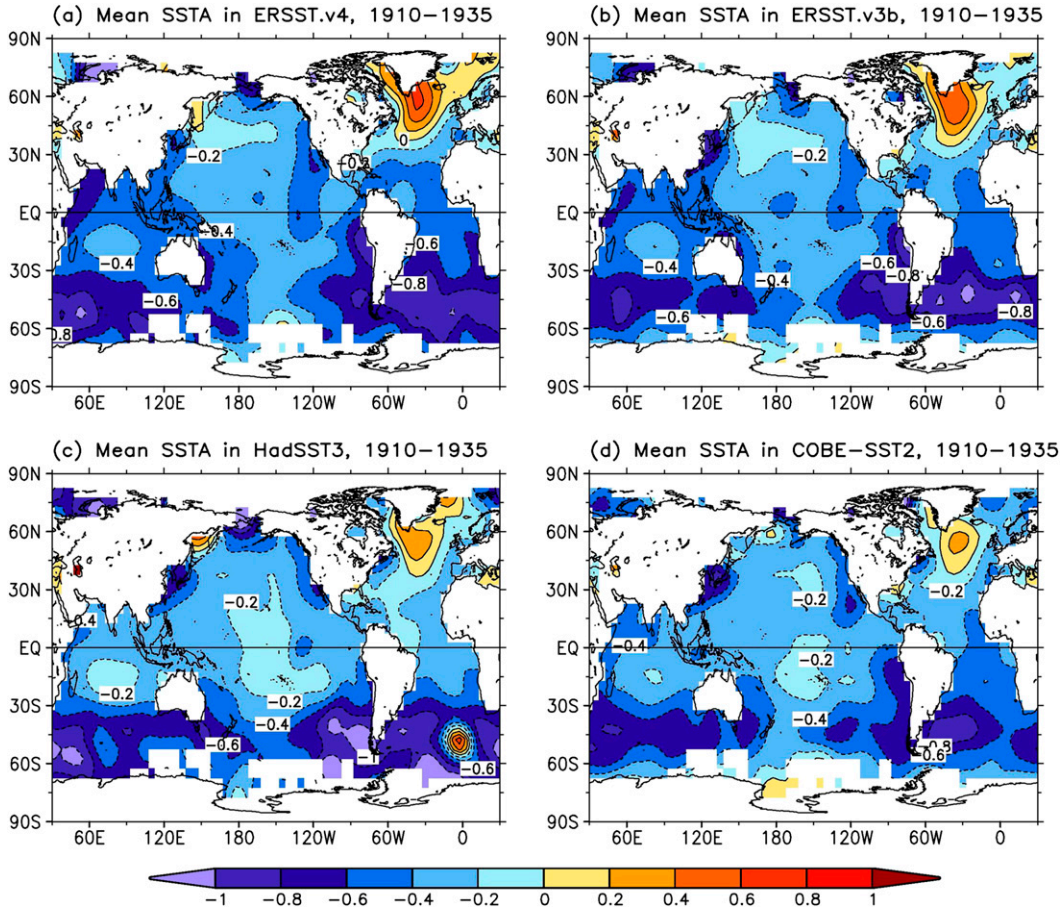

FIG. 10. Collocated SSTA between 1910 and 1935 in (a) ERSST.v4, (b) ERSST.v3b, (c) HadSST3, and (d) COBE-SST2. Contour intervals are 0.2°C.

SST patterns among these four products are found during this (Fig. 10) and other periods (not shown). The averaged (1910–35) SSTAs (Fig. 10) show that SSTAs vary from −0.2° to −0.8°C in most of the world oceans except in the northern North Atlantic where SSTAs are 0.4° to 0.6°C. The negative SSTAs across most of the global oceans between 1910 and 1935 are associated with generally cooler conditions compared to the warm climatological base period of 1971–2000 that is part of an overall warming trend since about 1910 (see Fig. 8a). In contrast, the higher SSTAs in the northern North Atlantic are associated with the fact that in this region SSTA varies strongly in a manner seemingly associated with the AMO (Enfield et al. 2001), which was in peak phase during 1910–35 and a minimum in the climatological period (Fig. 11).

The SSTA differences between datasets are found despite the general similarities in the magnitude and spatial distribution of SSTAs. Over the period of 1910–35, the SSTAs in the tropical oceans are approximately 0.2°C warmer in HadSST3 (Fig. 10c) and COBE-SST2 (Fig. 10d) than in ERSST.v4 (Fig. 10a) and ERSST.v3b (Fig. 10b). The SSTAs south of 30°S are slightly colder and less spatiotemporally consistent in HadSST3 than in ERSST.v4, ERSST.v3b, and COBE-SST2. Overall, SSTA differences are mostly consistent with the differences in bias adjustments in HadSST3, ERSST.v4, and ERSST.v3b (Fig. 6). In the northern North Atlantic south of Greenland, the SSTAs are cooler in HadSST3 and COBE-SST2 than in ERSST.v4 and ERSST.v3b, which can also be seen clearly in Fig. 11 between 1910 and 1935.

7. SST comparisons in the satellite era

The SSTAs have been adjusted to be relative to a 1971 to 2000 climatological base period in the comparisons in section 6. The disadvantage of using SSTAs is that the climatological average has been removed from SST, and the SST climatology may be defined differently among SST products. Many applications require absolute temperatures (i.e., full SSTs, not SSTAs) so it is important to

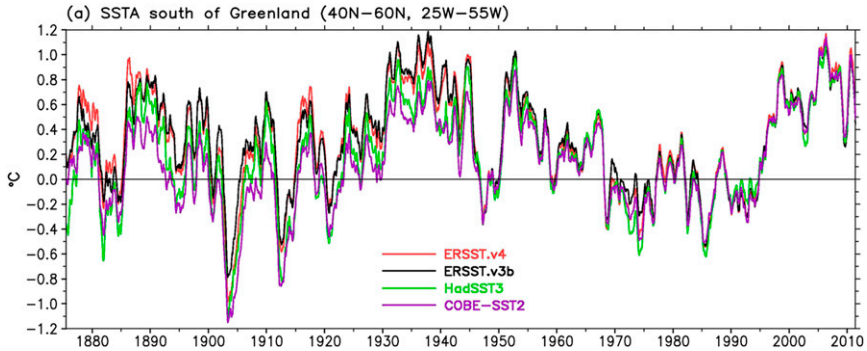

FIG. 11. Averaged SSTA south of Greenland (40°–60°N, 25°–55°W) in ERSST.v4, ERSST.v3b, HadSST3, and COBE-SST2.

understand any differences in full SSTs in addition to SSTAs. Therefore, full SSTs from ERSST.v4, HadISST, and COBE-SST2 are compared to ATSR satellite observations between 1997 and 2011. The HadISST instead of HadSST3 is used in the comparison because HadSST3 fields are not spatially interpolated. ATSR observations have been found to be the most accurate satellite observations of SST (Merchant et al. 2012). The reason for accurate ATSR observations is that the calibration of ATSR observations is ensured by the use of an onboard blackbody and dual view configuration. The use of onboard blackbody makes ATSR SSTs that are almost independent from in situ observations, and the use of a dual-view configuration makes the observations more stable to perturbations such as aerosol loading (Merchant et al. 2012). These comparisons to ATSR SSTs may provide a degree of confidence in the similarity of different SST reconstructions to satellite measurements and the true absolute values.

Comparisons to ATSR show that in the Southern Ocean south of 45°S, SSTs are 0.2° to 0.4°C warmer in ERSST.v4 (Fig. 12a) and HadISST (Fig. 12b), and 0.1° to 0.2°C warmer in COBE-SST2 (Fig. 12c). North of 60°N, SSTs are more than 0.4°C warmer in all three products although SSTs in ERSST.v4 are colder in some very high-latitude regions due to limitations of the EOT decomposition. In the lower latitudes between 45°S and 45°N, SSTs are slightly warmer in ERSST.v4 (Fig. 12a) except in the eastern equatorial Pacific where SSTs are about 0.4°C higher; SSTs are 0.1° to 0.3°C colder in HadISST (Fig. 12b) and SSTs are approximately 0.1°C colder in COBE-SST2 (Fig. 12c). Near the eastern coast of North America, SSTs are about 0.4°C colder in all three products.

Overall, the SST differences relative to ATSR observations are relatively small in COBE-SST2, and larger in ERSST.v4 and HadISST. This can be seen from the RMSD of monthly SSTs relative to ATSR. The RMSD is near 1°C north of 60°N and along the eastern coasts of East Asia and North America in all three products (Figs. 12d–f). South of 30°S, the RMSD is 0.6° to 1°C in ERSST.v4 (Fig. 12d) and HadISST (Fig. 12e), and 0.4° to 0.6°C in COBE-SST2 (Fig. 12f). Between 30°S and 30°N, the RMSD is approximately 0.4°C in ERSST.v4 (Fig. 12d) and HadISST (Fig. 12e), and 0.2° to 0.4°C in COBE-SST2 (Fig. 12f). On the global average, the RMSD is 0.54°, 0.56°, and 0.46°C in ERSST.v4, HadISST, and COBE-SST2, respectively.

8. Summary

The ERSST product has been substantially revised with 11 improvements introduced in version 4. Among the input datasets, the new version utilizes ICOADS R2.5 for a selection of the most complete available historical in situ SSTs, together with HadISST ice concentration datasets. Revisions have been made to many of the algorithmic parameters in the ERSST.v4 by careful selection of parameter values following extensive testing and analyses. These major parameters include the base function EOTs and their acceptance criterion, SST QC procedures, SSTA quantification at in situ locations, and SST bias adjustment using HadNMAT2. The most significant upgrade for long-term trend characterization is the ship SST bias adjustment, which has substantively impacted the SSTA analysis in global and long-term scales, while the impacts of the remaining parameters are predominantly in local and short-term scales which may be as important, if not more so, for many envisaged applications of the product such as monitoring Niño-3.4 temperature variations.

Variations of area averaged SSTA in ERSST.v4 at interannual and decadal time scales are broadly consistent with those in ERSST.v3b, HadSST3, and COBE-SST2 throughout the historic period. However, SSTAs are 0.1°C to 0.2°C lower in ERSST.v4 than in HadSST3

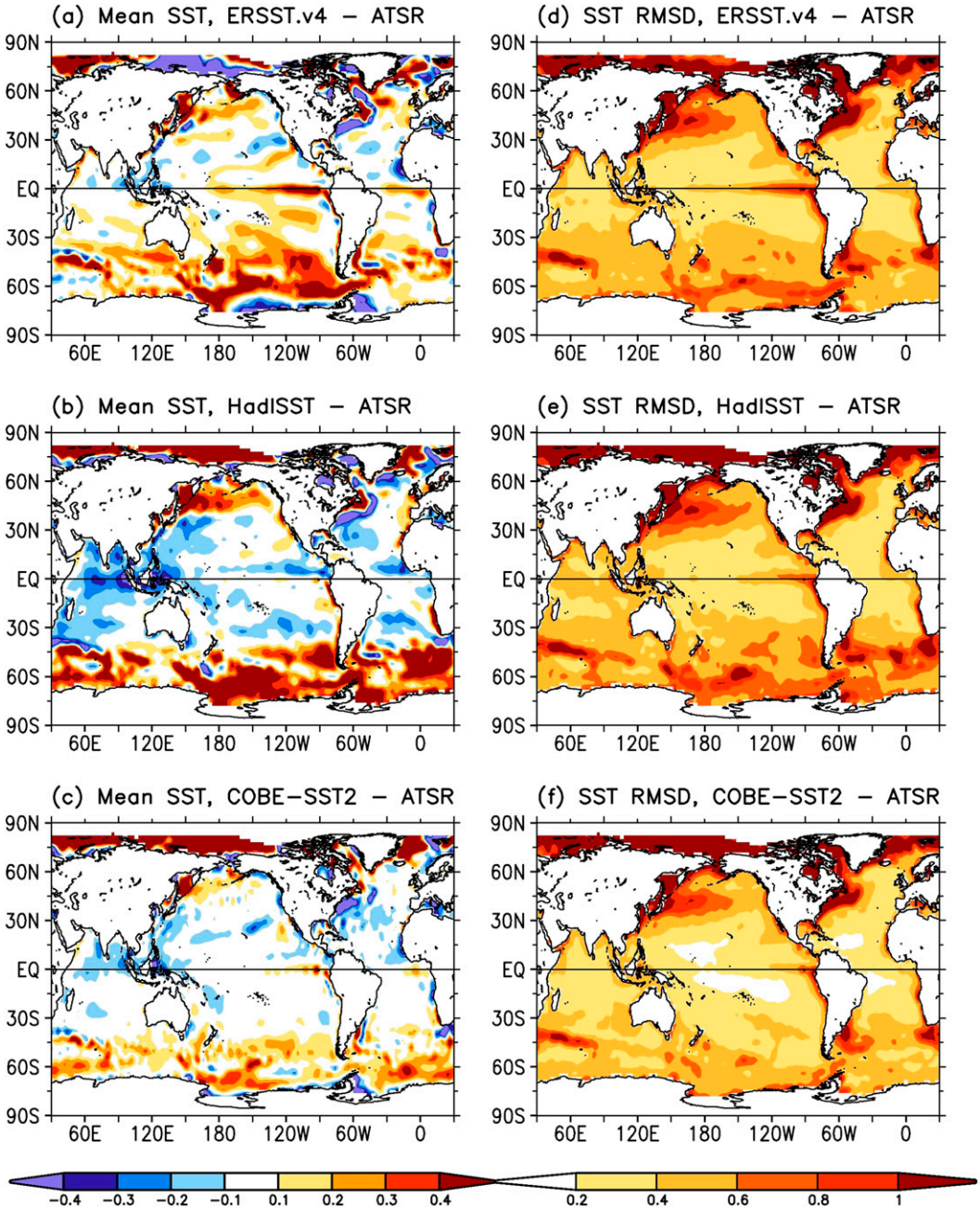

FIG. 12. Collocated mean (1997–2011) difference of SSTs on 2° × 2° grid between (a) ERSST.v4 and ATSR, (b) HadISST and ATSR, and (c) COBE-SST2 and ATSR. (d)–(f) As in (a)–(c), but for RMSD. The difference in the Arctic is blanked due to sparse observations. Contour intervals are 0.1°C in (a)–(c) and 0.2°C in (d)–(f).

and COBE-SST2 between 30°S and 30°N from approximately 1910 to 1970, while they are approximately 0.1°C higher south of 30°S before about 1920 and north of 30°N before around 1935. These differences mostly result from SST bias adjustment differences between the products, and can be attributed in part to the SST parametric uncertainty described in Part II.

Buoy SSTs have been adjusted toward ship SSTs in ERSST.v4 to correct for a systematic difference of 0.12°C between ship and buoy observations. Although buoy SSTs are more homogeneous and reliable than ship observations, buoys were not widely available before around 1980. However, the selection will not affect the evolution of the SSTAs. Further studies are needed

to consider the potential of including C-MAN SSTs and other near-surface ocean temperature measurements not presently incorporated in ERSST.v4 (e.g., from oceanographic profiling instruments).

In conclusion, ERSST.v4 uses the most recent available in situ datasets, includes up-to-date ship and buoy bias adjustments throughout the entire analysis period, and presents uncertainty estimations associated with internal parameters of the analysis (Part II). These innovations permeate the dataset and substantially improve its applicability over ERSST.v3b across a range of space and time scales and end-user applications. The SST in ERSST.v4 exhibits, for example, a substantially more realistic El Niño/La Niña behavior in the early period of the record when data are sparse and therefore a better estimate of long-term variability in this key mode of internal climate variability. The dataset does not change the interdecadal trends significantly at the largest spatial scale and longest time scales over the preceding v3b data, but the dataset provides a more robust estimate due to advances in the application of, in particular, SST bias adjustment and buoy SST adjustment procedures. SSTs in ERSST.v4 are reasonably close to the independent satellite-based ATSR observations. Anomaly series are broadly comparable to the methodologically independent HadSST3, HadISST, and COBE-SST2 reconstructions although some interesting differences remain between these in situ products. Investigators should use several such products to ensure robustness of their analyses to such structural uncertainties.

Acknowledgments. Authors thank three anonymous reviewers for their critiques and comments that have greatly improved the manuscript. Authors appreciate the discussion with Richard Reynolds and David Wuertz in the early stage of the study, and suggestions from two anonymous NOAA internal reviewers are very helpful. Authors thank John Kennedy and Masayoshi Ishii for providing HadSST3 and COBE-SST2 datasets. PWT's early involvement occurred while an employee of CICS-NC, NCSU.

REFERENCES

Banzon, V. F., R. W. Reynolds, and T. M. Smith, 2010: The role of satellite data in extended reconstruction of sea surface temperatures. *Proc. "Oceans from Space,"* Venice, Italy, European Commission, 27–28, doi:10.2788/8394.

Cleveland, W. S., 1981: LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *Amer. Stat.,* **35,** 54–65, doi:10.2307/2683591.

Delworth, T. L., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate,* **19,** 643–674, doi:10.1175/JCLI3629.1.

Enfield, D. B., A. M. Mestas-Nuñez, and P. J. Trimble, 2001: The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S. *Geophys. Res. Lett.,* **28,** 2077–2080, doi:10.1029/2000GL012745.

Folland, C. K., and 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.

Gregg, M. C., and M. L. Newlin, 2012: Global oceans [in "State of the Climate in 2011"]. *Bull. Amer. Meteor. Soc.,* **93,** S57–S92.

Grumbine, R. W., 1996: Automated passive microwave sea ice concentration analysis at NCEP. NOAA Tech. Note 120, 13 pp. [Available from NCEP/NWS/NOAA, 9802 51th Ave. College Park, MD 20740.]

Hartmann, D. L., and Coauthors, 2014: Observations: Atmosphere and surface. *Climate Change 2013: The Physical Science Basis,* T. F. Stocker et al., Eds., Cambridge University Press, 159–254.

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.

Hurrell, J. W., and K. E. Trenberth, 1999: Global sea surface temperature analyses: Multiple problems and their implications for climate analysis, modeling, and reanalysis. *Bull. Amer. Meteor. Soc.,* **80,** 2661–2678, doi:10.1175/1520-0477(1999)080<2661:GSSTAM>2.0.CO;2.

Ishii, M., A. Shouji, S. Sugimoto, and 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.

Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856–1991. *J. Geophys. Res.,* **103,** 18 567–18 589, doi:10.1029/97JC01736.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011: 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.

Kent, E. C., and A. Kaplan, 2006: Toward estimating climatic trends in SST. Part III: Systematic biases. *J. Atmos. Oceanic Technol.,* **23,** 487–500, doi:10.1175/JTECH1845.1.

——, S. D. Woodruff, and D. I. Berry, 2007: Metadata from WMO Publication No. 47 and an assessment of voluntary observing ship observation heights in ICOADS. *J. Atmos. Oceanic Technol.,* **24,** 214–234, doi:10.1175/JTECH1949.1.

——, J. J. Kennedy, D. I. Berry, and R. O. Smith, 2010: Effects of instrumentation changes on sea surface temperature measured in situ. *Climatic Change,* **1,** 718–728, doi:10.1002/wcc.55.

——, N. A. Rayner, D. I. Berry, M. Saunby, B. I. Moat, J. J. Kennedy, and D. E. Parker, 2013: Global analysis of night marine air temperature and its uncertainty since 1880: The HadNMAT2 data set. *J. Geophys. Res. Atmos.,* **118,** 1281–1298, doi:10.1002/jgrd.50152.

Latif, M., and 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.

Liu, W., and Coauthors, 2015: Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4). Part II: Parametric and structural uncertainty estimation. *J. Climate,* **28,** 931–951, doi:10.1175/JCLI-D-14-00007.1.

Merchant, C. J., and Coauthors, 2012: A 20 year independent record of sea surface temperature for climate from Along-Track Scanning Radiometers. *J. Geophys. Res.,* **117,** C12013, doi:10.1029/2012JC008400.

Parker, D. E., P. D. Jones, A. Bevan, and C. K. Folland, 1994: Interdecadal changes of surface temperature since the late 19th century. *J. Geophys. Res.,* **99,** 14 373–14 399, doi:10.1029/94JD00548.

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

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1992: LU decomposition and its applications. *Numerical Recipes in FORTRAN: The Art of Scientific Computing,* 2nd ed., Cambridge University Press, 34–42.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and 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., and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate,* **7,** 929–948, doi:10.1175/1520-0442(1994)007<0929:IGSSTA>2.0.CO;2.

——, N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate,* **15,** 1609–1625, doi:10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2.

——, C. L. Gentemann, and G. K. Corlett, 2010: Evaluation of AATSR and TMI satellite SST data. *J. Climate,* **23,** 152–165, doi:10.1175/2009JCLI3252.1.

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

Smith, T. M., and R. W. Reynolds, 2002: Bias adjustments for historical sea surface temperatures based on marine air temperatures. *J. Climate,* **15,** 73–87, doi:10.1175/1520-0442(2002)015<0073:BCFHSS>2.0.CO;2.

——, and ——, 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.

——, and ——, 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.

——, ——, R. E. Livezey, and 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.

——, ——, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006). *J. Climate,* **21,** 2283–2296, doi:10.1175/2007JCLI2100.1.

Stocker, T. F., and Coauthors, Eds., 2014: *Climate Change 2013: The Physical Science Basis.* Cambridge University Press, 1535 pp.

Thompson, D. W. J., J. J. Kennedy, J. M. Wallace, and P. D. Jones, 2008: A large discontinuity in the mid-20th century in observed global-mean surface temperature. *Nature,* **453,** 646–649, doi:10.1038/nature06982.

Trenberth, K. E., 1984: Signal versus noise in the Southern Oscillation. *Mon. Wea. Rev.,* **112,** 326–332, doi:10.1175/1520-0493(1984)112<0326:SVNITS>2.0.CO;2.

van den Dool, H. M., S. Saha, and A. Johansson, 2000: Empirical orthogonal teleconnections. *J. Climate,* **13,** 1421–1435, doi:10.1175/1520-0442(2000)013<1421:EOT>2.0.CO;2.

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

Woodruff, S. D., R. Slutz, R. Jenne, and P. Steurer, 1987: A comprehensive ocean–atmosphere data set. *Bull. Amer. Meteor. Soc.,* **68,** 1239–1250, doi:10.1175/1520-0477(1987)068<1239:ACOADS>2.0.CO;2.

——, H. F. Diaz, E. C. Kent, R. W. Reynolds, and S. J. Worley, 2008: The evolving SST record from ICOADS. *Climate Variability and Extremes during the Past 100 Years,* S. Brönnimann et al., Eds., Advances in Global Change Research Series, Vol. 33, Springer, 65–83, doi:10.1007/978-1-4020-6766-2_4.

——, and Coauthors, 2011: ICOADS release 2.5: Extensions and enhancements to the surface marine meteorological archive. *Int. J. Climatol.,* **31,** 951–967, doi:10.1002/joc.2103.