Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons

Boyin Huang,[a] Peter W. Thorne,[b] Viva F. Banzon,[a] Tim Boyer,[a]
Gennady Chepurin,[a,c] Jay H. Lawrimore,[a] Matthew J. Menne,[a]
Thomas M. Smith,[d,e] Russell S. Vose,[a] and Huai-Min Zhang[a]

[a] *NOAA/NCEI, Asheville, North Carolina*
[b] *Irish Climate Analysis and Research Units, Department of Geography, Maynooth University, Maynooth, Ireland*
[c] *Department of Atmospheric and Ocean Science, University of Maryland, College Park, College Park, Maryland*
[d] *NOAA/STAR/SCSB, College Park, Maryland*
[e] *CICS/ESSIC, University of Maryland, College Park, College Park, Maryland*

(Manuscript received 22 November 2016, in final form 26 June 2017)

ABSTRACT

The monthly global $2° \times 2°$ Extended Reconstructed Sea Surface Temperature (ERSST) has been revised and updated from version 4 to version 5. This update incorporates a new release of ICOADS release 3.0 (R3.0), a decade of near-surface data from Argo floats, and a new estimate of centennial sea ice from HadISST2. A number of choices in aspects of quality control, bias adjustment, and interpolation have been substantively revised. The resulting ERSST estimates have more realistic spatiotemporal variations, better representation of high-latitude SSTs, and ship SST biases are now calculated relative to more accurate buoy measurements, while the global long-term trend remains about the same. Progressive experiments have been undertaken to highlight the effects of each change in data source and analysis technique upon the final product. The reconstructed SST is systematically decreased by 0.077°C, as the reference data source is switched from ship SST in ERSSTv4 to modern buoy SST in ERSSTv5. Furthermore, high-latitude SSTs are decreased by 0.1°–0.2°C by using sea ice concentration from HadISST2 over HadISST1. Changes arising from remaining innovations are mostly important at small space and time scales, primarily having an impact where and when input observations are sparse. Cross validations and verifications with independent modern observations show that the updates incorporated in ERSSTv5 have improved the representation of spatial variability over the global oceans, the magnitude of El Niño and La Niña events, and the decadal nature of SST changes over 1930s–40s when observation instruments changed rapidly. Both long- (1900–2015) and short-term (2000–15) SST trends in ERSSTv5 remain significant as in ERSSTv4.

1. Introduction

Sea surface temperature (SST) is an essential climate variable (Bojinski et al. 2014) and one of the most important indicators of Earth's climate (EPA 2014). Historic SST data are widely used in climate simulations, assessments, and monitoring activities (IPCC 2013; Xue et al. 2016). Several SST datasets have been developed by independent groups and are available to the public, with several of these updated monthly or more frequently. Some analyses only use in situ observations, prominent examples being the Extended Reconstructed SST (ERSST; Smith et al. 1996; Huang et al. 2015a), Hadley Centre SST, version 3 (HadSST3; Kennedy et al. 2011a,b), and Japan Meteorological Agency Centennial Observation-Based Estimates of SSTs (COBE-SST; Ishii et al. 2005) and COBE-SST, version 2 (COBE-SST2; Hirahara et al. 2014). Others use both in situ and satellite observations, examples including the National Centers for Environmental Prediction (NCEP) Weekly Optimum Interpolation SST (WOISST) (Reynolds et al. 2002), National Centers for Environmental Information (NCEI) Daily Optimum Interpolation SST (DOISST; Reynolds et al. 2007), Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST; Rayner et al. 2003), and Kaplan SST (Kaplan et al. 1998). Reliable SST retrievals from satellites start in the early 1980s while

---

in situ observations are available much earlier but have changed substantively through time in both their methods of measurement and where the measurements are taken. Therefore, the problem of creating a dependable estimate of the true historical variability and change is a substantial challenge (Kent et al. 2017).

Several intercomparison studies and assessments have indicated that the global-scale features and long-term trends are broadly consistent among SST products (e.g., Yasunaka and Hanawa 2011; IPCC 2013; Kennedy 2014). Differences among the available products can largely be reconciled by the quantified uncertainties associated with those SST analyses (e.g., Huang et al. 2016a; Kennedy et al. 2011b). However, the differences are often larger than the recognized single-dataset uncertainties in some regions (e.g., the tropical Pacific) and over shorter time scales (e.g., the past two decades) (Huang et al. 2013, 2016b). These interdataset differences mostly result from how the in situ SST data biases are corrected (Huang et al. 2015a,b; Kent et al. 2017) and may also result from quality control and gap-filling choices when and where observations are sparse, particularly in early record periods.

Recent studies based on ERSST, version 4 (ERSSTv4; Huang et al. 2015a), indicated that the global SST has been warming in the most recent decade as fast as in the past 50 years (Huang et al. 2016b), which called into question the existence of a recent global warming "hiatus" reported in the IPCC Fifth Assessment Report (IPCC 2013; Karl et al. 2015). These changes from the preceding ERSSTv3b were mostly associated with the bias correction of ship SST observations using the most recent Hadley Centre Nighttime Marine Air Temperature, version 2 (HadNMAT2; Kent et al. 2013), and how buoy SSTs were handled. These conclusions were further confirmed by satellite observations from the Advanced Very High Resolution Radiometer (AVHRR; Huang et al. 2016b) and, more recently, independently using other independent high-quality observations (e.g., Argo floats; Hausfather et al. 2017).

Dataset construction is not a one-time operation. Any given dataset version represents a snapshot of the then-current best knowledge of the data issues and technical and methodological capabilities available to address them. Knowledge, data availability, and technical capabilities evolve with time, and hence periodic reassessments and updates are warranted. Several suggestions regarding new data sources, current limitations, and so forth have been forthcoming from users. In addition, the work of Huang et al. (2016a) on uncertainty quantification led to potential innovations and improvements that warranted further investigation.

A number of analyses have been undertaken since the ERSSTv4 release in early 2014 to constantly question the assumptions underlying the algorithm and seek improved estimates of the true SST state through time globally, regionally, and locally. The innovations pursued in ERSSTv5 and their rationale can be considered to fall into the following four broad classes.

a. Choice of reference observation

A global averaged offset (0.12°C; 1990–2010) was found between ship and buoy observations (Rayner et al. 2003; Kennedy et al. 2011b; Huang et al. 2015a; Hirahara et al. 2014). This offset could apply to either ship or buoy observations as far as the SST anomalies (SSTAs) are concerned. In ERSSTv4, the offset was applied to the buoy observations after the 1980s; the buoy offsetting avoided applying a bias correction to all pre-1980s data that are dominated by ship observations (Huang et al. 2015a). On the other hand, it is now realized that applying an offset to ship observations (instead of buoys) has the advantage of making timely updates simpler as the present-day data (at least by volume) primarily are from buoy measurements, and in doing so we are also no longer dependent upon third-party-provided NMAT data (from Met Office in ERSSTv4). In addition, buoy-observing techniques have been shown to be more homogenous and the buoy data are more accurate than ship data (Reynolds and Smith 1994; Reynolds et al. 2002).

b. Spatial variability of SSTA

Since the public release of ERSSTv4, both internal and external users have suggested that SSTAs appear to be too smooth over much of the global oceans. This can seriously affect its utility for regional and local scale studies. For example, it tends to damp the magnitude of El Niño and La Niña events relative to similar SST products. This damping results from the strong spatial filters applied to the training data when the base functions [empirical orthogonal teleconnections (EOTs); van den Dool et al. 2000] were calculated. In addition, the EOTs in ERSSTv4 were damped in the high latitudes, and there were no EOTs in the Arctic at all. These made it difficult to reconstruct the SSTs in partially ice-covered oceans, particularly in recent decades, when more SST observations have become available because of decreased sea ice coverage.

c. Quality screening of in situ data

The bias-corrected first-guess (FG) SSTs from ERSSTv3b were used in the quality control (QC) procedures in ERSSTv4. SST observations were discarded if they deviated from the FG by more than 4.5 times the SST standard deviation (STD). Since the raw SST observations are approximately normally distributed near the average of bias-uncorrected SSTs, the selection of

the bias-corrected FG resulted in the inclusion of more warm SSTs, since the SST observations were generally cold biased before the 1940s (Kennedy et al. 2011b; Huang et al. 2015a; Hirahara et al. 2014). Huang et al. (2016a) found a strong effect of this choice on the final analysis within the ensembles. This shortcoming is addressed in this ERSSTv5 improvement.

d. New input datasets

The SSTs from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS), release 2.5 (R2.5; Woodruff et al. 2011), and sea ice concentration from HadISST were used in ERSSTv4. The ICOADS SSTs have now been updated to R3.0 (Freeman et al. 2017; https://doi.org/10.7289/V5CZ3562). The HadISST sea ice concentration dataset has been updated to version 2 (HadISST2; Titchner and Rayner 2014). The Argo program of autonomous ocean subsurface profilers, with increased numbers starting from 1999 (Argo 2000; Roemmich et al. 2001), reached global coverage around 2005 (except marginal seas, ice-covered areas, and continental shelves). The number of observations from Argo floats between 0- and 5-m depth (Argo5obs) has rapidly expanded over 2000–06 (Fig. 1a) and has maintained near-global coverage since 2006 (Fig. 1b). Therefore, Argo observations may improve coverage, making up for the reduction in ship numbers, and reporting timeliness. However, it is necessary that the Argo5obs data are corrected toward the SSTs from drifting buoys at a typical depth of 0.2 m, since the observing depths of the two different types of instruments differ. As a result, ERSSTv5 is representative of SST measured at a nominal depth of 0.2 m.

These four innovations have been implemented in ERSSTv5 using an eight-step process, one step at a time (Table 1). The improvements were evaluated using a combination of independent observations and cross-validation testing.

The remainder of the paper is arranged as follows: The validation datasets and the source datasets used in ERSSTv5 are described in section 2. The ERSST reconstruction method is briefly described in section 3. The upgrades and their impacts are assessed in section 4. Intercomparisons are presented in section 5. A summary is given in section 6.

2. Datasets

a. Data sources used in ERSSTv5 processing

1) ICOADS R3.0

The objective of ICOADS is to support climate assessment and monitoring, reanalyses, and near-real-time (NRT) applications, among others. In comparison

to R2.5 (Woodruff et al. 2011), R3.0 (Freeman et al. 2017) includes additional metadata such as assignment of a unique identifier (UID) to each marine report, new near-surface oceanographic data elements, and cloud parameters. Many new input data sources have been acquired and updated, and improvements were made to existing data sources. Other improvements include removal of erroneous data. R3.0 is available from both the National Center for Atmosphere Research (NCAR) and NCEI.

The in situ observations of R3.0 archived at NCEI from 1854 to 2015 are used in ERSSTv5, while R2.5 was used in ERSSTv4. The Global Teleconnections System (GTS) receipts from NCEP after January 2016 are used in operational ERSSTv5. NCEP GTS may slightly differ from NRT ICOADS R3.0 (Freeman et al. 2017), but the difference should not affect the discussion here that considered the period of 1854–2015. In comparison with R2.5, ICOADS R3.0 includes substantially more ship observations in the 1850s–60s, 1950s–60s, 1990s, and 2000s–10s and includes more buoy observations in the 1980s–2000s (Fig. 1a). Spatial coverage is slightly higher in ICOADS R3.0 than R2.5 in the 1850s–60s and in the later 2000s (Figs. 1b and 1c).

2) ARGO SST ABOVE 5 M

Argo observations were not included in ERSSTv4 but are included in ERSSTv5. The Argo data used in ERSSTv5 are from the Global Data Assembly Centre, France (https://doi.org/10.17882/42182). The Argo program's main purpose was to provide as complete a picture as possible of the oceans' subsurface temperature and salinity structure in the upper 2000 m in order to track changes in ocean heat and freshwater content (Riser et al. 2016). Floats are deployed oceanwide. They normally drift at depths of near 1000 m, and then on usually 10-day cycles, they typically descend to near 2000-m depth; they ascend from that depth to the surface, measuring temperature and salinity along the way. Data are transmitted to satellite before the floats descend to drift for another 10 days. In this manner, most of the ice-free ocean above 2000-m depth outside of marginal seas has been observed. Moreover the international program has a coordinated quality control and dissemination system ensuring the highest quality and availability of the Argo observations.

For ERSSTv5, Argo5obs are retrieved from Argo floats and used as SST observations. Roemmich et al. (2015) showed that the global mean Argo temperature anomaly above 5 m tracks closely with SST change. The number of Argo5obs data receipts has been expanded over 2000–06 and is nearly equivalent to the number of ship observations by the end of 2015 (Fig. 1a). The
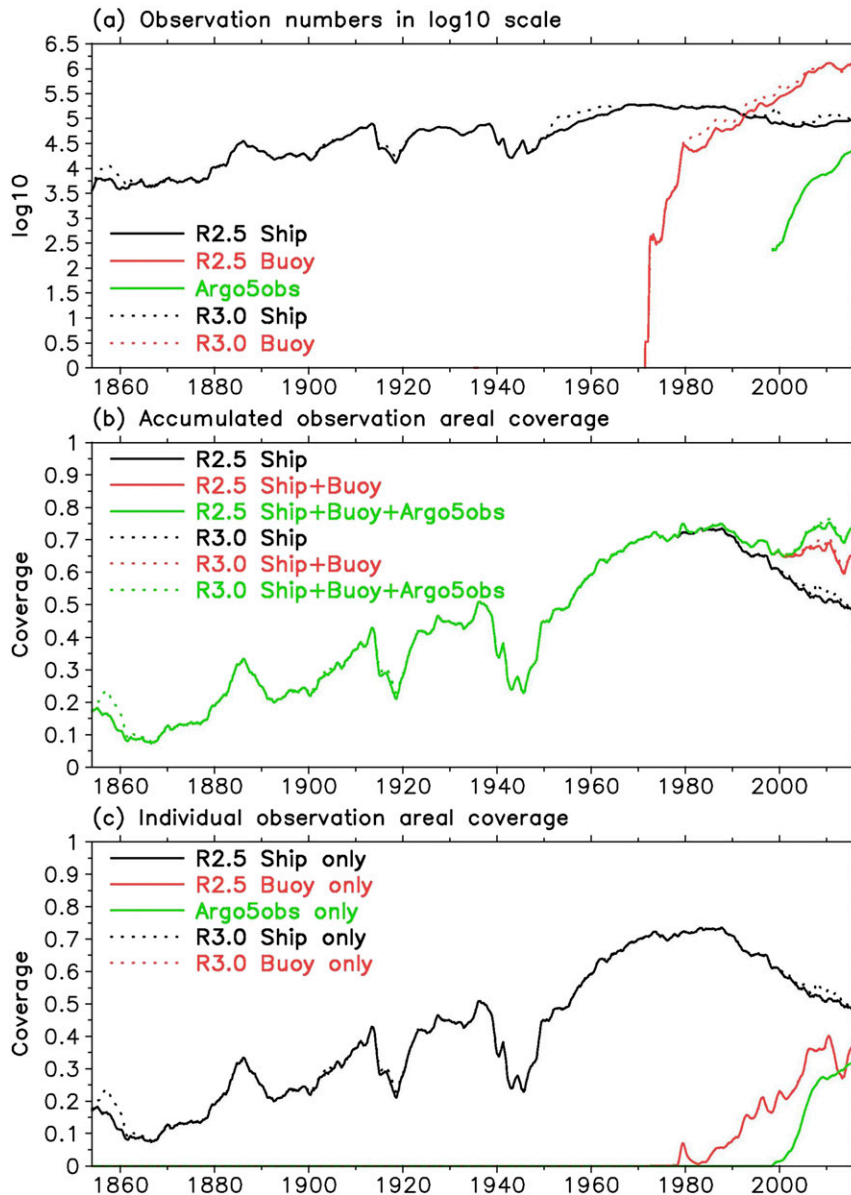
FIG. 1. (a) Number (log10 scale), (b) accumulated areal coverage ratio, and (c) individual area coverage ratio of monthly ship (black), buoy (red), and Argo5obs (green) observations. The spatial coverage is a ratio between the area of 2° × 2° boxes containing observations and the total ocean area. The ICOADS R2.5 (R3.0) is represented by solid (dotted) lines. A 12-month running filter is applied in plotting.

global areal coverage of Argo5obs increases to 30% by the end of 2015, which is as high as that of buoy observations (Fig. 1c). Argo observations provide approximately 5%–10% extra area coverage in addition to ship and buoy observations after 2000 (Fig. 1b).

Our analysis in section 3 shows that Argo5obs data are close to buoy observations with an averaged difference and root-mean-square difference (RMSD) of 0.03°C. Thus they are first used as a validation dataset to assess

the improvements in the progressive experiments listed in Table 1 and only then included in the ERSSTv5 at the final stage as described in section 4d(2) and for operational monthly updates.

3) HADISST2 SEA ICE CONCENTRATION

Sea ice concentrations from HadISST2 are used to relax reconstructed SSTs in partial ice-covered areas toward the freezing point (−1.8°C) at the very final stage

TABLE 1. The eight progressive experiments (Exp) toward ERSSTv5 starting from ERSSTv4.

Exp	Progressive revisions
ERSSTv4	
UnadjFG	Same as ERSSTv4 except for using unadjusted FG.
NDP	Same as UnadjFG except for no high-latitude damping in EOT training.
SMT	Same as NDP except for using a smoothing of one-time 6-deg instead of three-time 14-deg running filter in EOT training.
EOT140	Same as SMT except for including additional 10 EOTs in the Arctic.
ShipBias	Same as EOT140 except for using combined ship bias corrections referenced to buoy SSTs and a linear fit estimation prior to Lowess filtering to better account for the 1940s transition in ship observing methods.
ICOADS3	Same as ShipBias except for using ICOADS R3.0.
Argo5m	Same as ICOADS except for including Argo observations above 5 m (Argo5obs).
ERSSTv5	Same as Argo5m except replacing the HadISST sea ice concentration with the HadISST2 sea ice concentration.

in section 4d(3). During the development of ERSSTv5, we attempted to replace the SST relaxation by a method that used sea ice proxy SST as described in Reynolds et al. (2002). However, we found that the proxy SSTA from sea ice concentration is 0.2°–0.5°C lower than available in situ SSTA in the Southern Ocean before the 1970s, and therefore the proxy SST is not applied in ERSSTv5. Further study is needed on sea ice proxy SST and its dependence on historic sea ice reconstruction prior to potential implementation.

Monthly 1° × 1° sea ice concentration from HadISST2 over 1870–2015 (Titchner and Rayner 2014) is averaged by area weighting to 2° × 2° and used in ERSSTv5. The monthly sea ice concentration before 1870 is set to periodic monthly values of 1870, since the concentration in 1870–1900s (1870–1940s) was set to a monthly climatology in HadISST2 in the Northern (Southern) Hemisphere oceans. In comparison with the sea ice concentration from HadISST (Rayner et al. 2003) that was used in ERSSTv4, the sea ice concentration in HadISST2 is approximately 25% higher before the 1980s and 5% higher after the 1980s in the Southern Hemisphere (Fig. 2b) and is slightly higher in the Northern Hemisphere in the 1940s–50s and 1970s–90s (Fig. 2a). The interannual variability after the 1960s is larger in HadISST2 than in HadISST. These differences indicate that the sea ice concentration in HadISST2 and HadISST may have large uncertainty in the Arctic before the 1950s and in the Southern Ocean before the 1970s.

Starting from January 2016, the daily 0.5° × 0.5° sea ice concentration from NCEP (Grumbine 2014) is averaged by area weighting to monthly 2° × 2° ERSST ocean grids and used for monthly operational ERSSTv5 production. Then, the NCEP sea ice concentration is adjusted toward the HadISST2 sea ice concentration, when the sea ice concentration in NCEP is higher than 0.3. The adjustment uses the monthly varying averaged offsets between the two products over 2006–15. The adjustment is estimated

by first calculating monthly differences between HadISST2 and NCEP ice concentration over 2006–15 as a function of month from January to December and ice concentration from 0.0 to 1.0 with an increment of 0.1. The difference is then applied to NCEP ice concentration on 2° × 2° grids. The procedure is done separately for Northern and Southern Hemisphere oceans.

The adjusted NCEP sea ice concentration is close to the HadISST2 sea ice concentration (Fig. 2). However, the adjustment is not perfect because the concentration below 0.3 is not adjusted to avoid negative sea ice concentrations. This will not impact the SST analysis since SSTs are only adjusted in regions where ice concentrations are above 0.6 as described in section 4d(3).

4) HADNMAT2

The monthly 5° × 5° HadNMAT2 data in 1880–2010 (Kent et al. 2013) were used in ERSSTv4 to perform the ship SST bias correction. The bias correction was calculated at 5° × 5° grid, regridded to 2° × 2°, and applied to ERSSTv4. The bias correction to the ship SSTs in ERSSTv5 is similar to that in ERSSTv4 except that the bias correction is adjusted by subtracting 0.077°C to match the average ship-buoy offset over 1990–2010. The offset is derived as the difference between buoy-based correction (−0.118°C) and NMAT-based correction (−0.041°C) over 1990–2010, which will be discussed further in section 4c(1). The reasons for using NMAT to correct ship SST observations are 1) NMAT excludes the potential bias due to diurnal changes in solar radiation during daytime, 2) NMAT has a good relationship with SST (Huang et al. 2015a), 3) biases in NMAT observations can be corrected relatively easily using available metadata (Kent et al. 2013), and 4) NMAT is largely independent of SST observations. It should be noted that the bias correction for the transition from buckets to engine room intakes (ERI) starting in the 1940s using NMAT is different from using the individual bucket model applied in HadSST3
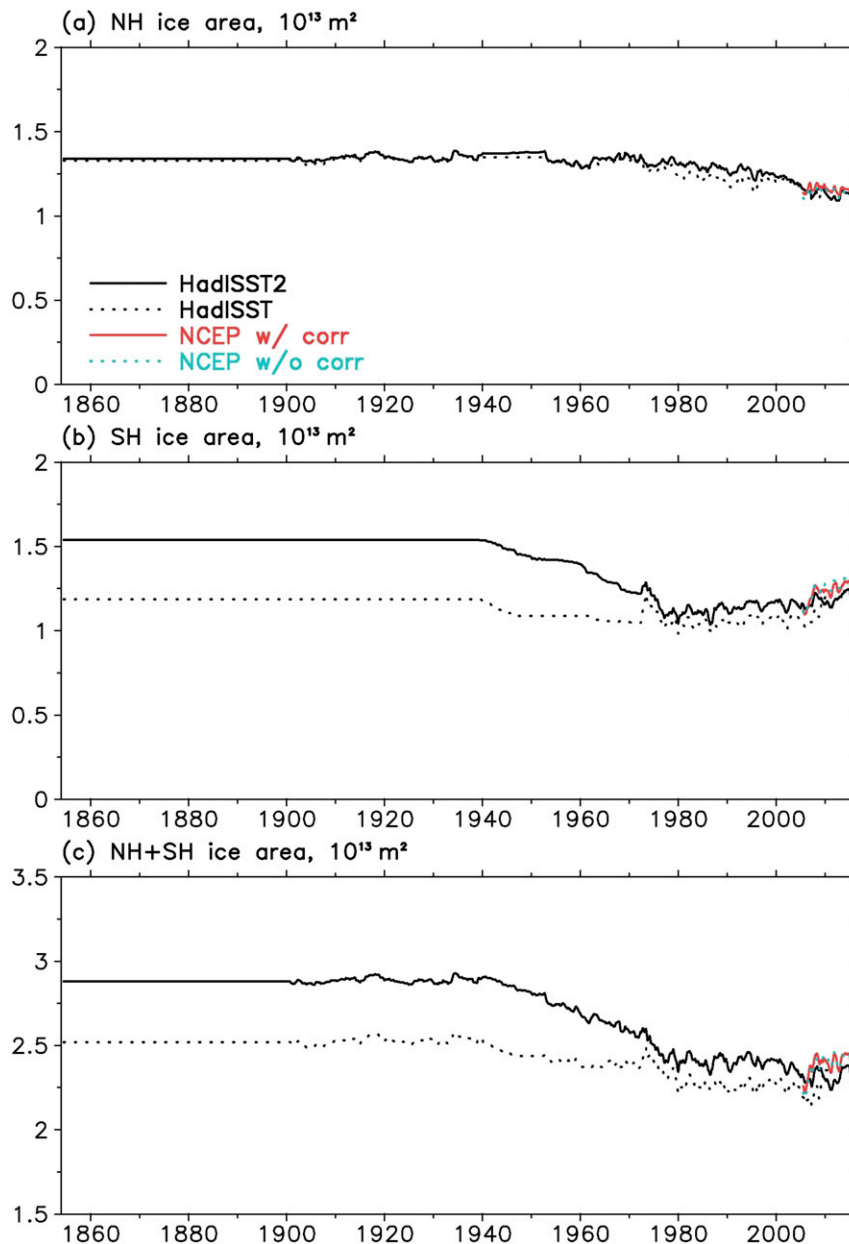
FIG. 2. Integrated ice-covered area (10^{13} m^2) in (a) Northern Hemisphere, (b) Southern Hemisphere, and (c) global oceans from HadISST2 (solid black line), HadISST (dotted black line), NCEP (solid red line), and corrected NCEP (dotted green line). A 12-month running filter is applied in plotting.

(Kennedy et al. 2011a,b). This represents a major source of uncertainty in the reconstruction of SSTs (Kent et al. 2017; Huang et al. 2015a).

5) WOISST

Monthly SSTs were derived by area-weight averaging to a 2° × 2° grid from 1° × 1° WOISST over 1982–2011 (Reynolds et al. 2002). The derived SSTs were filtered by the 2° × 2° ERSST ocean mask. The WOISST includes both in situ and satellite observations. The WOISST is used in ERSSTv5 1) to derive SST STD in the QC procedure, 2) to derive 140 EOTs, and 3) to cross-validate the improvements in the progressive experiments listed in Table 1.

6) UNADJUSTED SST

Monthly 2° × 2° unadjusted SSTs are derived by subtracting the bias correction from ERSSTv4 SST. The

unadjusted SSTs are more comparable to raw observational SSTs and thus are used as FG in the QC procedure to remove outliers in the SST observations when the deviation between observations and FG is larger than 4.5 times SST STD.

b. Validation datasets

The previous version, ERSSTv4, is the starting point of the progressive experiments in Table 1. ERSSTv5, ERSSTv4 (Huang et al. 2015a), HadISST (Rayner et al. 2003), and COBE-SST2 (Hirahara et al. 2014) are compared with the European Space Agency (ESA) Climate Change Initiative (CCI) level 4 version 1.1 SST from September 1991 to December 2010 (Merchant et al. 2014) to evaluate the improvements and identify limitations in ERSSTv5 (for potential future development).

- The CCI SST includes both AVHRR and Along-Track Scanning Radiometer (ATSR) observations on a monthly $1° × 1°$ grid. The CCI SST provides the mean SST at 20-cm depth that is close to the nominal depth of drifting buoy measurements. The CCI SST is retrieved by a reduced-state-vector optimal estimation algorithm and therefore is largely independent from in situ observations (Merchant et al. 2014).
- The SST data from HadISST are monthly $1° × 1°$ from 1870 to 2015, including both in situ and AVHRR observations.
- The COBE-SST2 data are monthly $1° × 1°$ from 1850 to 2012 (no update has been made after 2012), including only in situ observations as in ERSSTv5.

All datasets have been averaged by area weighting to $2° × 2°$ for comparisons. The Southern Oscillation index (SOI) using monthly mean sea level pressure anomalies at Tahiti and Darwin (Trenberth and Stepaniak 2001) over 1866–2015 is used to validate the El Niño and La Niña events.

3. Reconstruction methods

The overarching reconstruction methodology of ERSSTv5 is the same as ERSSTv4 (Huang et al. 2015a) and its predecessors (Smith et al. 1996; Smith and Reynolds 2003). The upgrades (Table 1) for ERSSTv5 are 1) the unadjusted FG fields (that are more comparable to observed raw SSTs) are used in QC (section 4a), 2) EOT functions are revised (section 4b), 3) the biases of ship SSTs are corrected relative to buoy observations (section 4c), and 4) the latest and additional datasets (including ICOADS R3.0 and Argo5obs) are used (section 4d).

SST measurements from ship, buoy, and Argo observations from 1854 to present are used to reconstruct

monthly $2° × 2°$ SSTA data. SST observations are accepted (rejected) under a QC criterion that observed SSTs differ from the unadjusted FG from ERSSTv4 by less (more) than 4.5 times STD of WOISST over 1981–2011 (Smith and Reynolds 2003). The QC test is applied relative to the time-varying mean state in the reconstruction over 1854–2015 in order to ensure that outliers are rejected more or less symmetrically around the period mean and so that long-term changes do not skew the resulting distributions. SST observations that pass QC are then converted into SSTAs by subtracting the SST climatology (1971–2000) at their in situ locations at monthly resolution. Ship SSTAs are then adjusted based on the NMAT before 2010 and by the globally averaged offset between ship and buoy SSTs after 2010 (section 4c). The globally averaged corrections are continuous at 2010, and the discontinuity on a local grid scale is small relative to local SST variability. Argo SSTs are similarly adjusted according to the globally averaged offset between Argo and buoy SSTs over 2000–14 (section 4d).

Ship, buoy, and Argo SSTAs are then merged and bin-averaged into monthly superobservations on a $2° × 2°$ grid. The averaging of ship, buoy, and Argo SSTAs is based on their proportional contribution to the grid at a given time step. In forming the monthly 2° averages, the buoy and Argo observations are weighted by a factor of 6.8 more than ship observations. That factor is based on the ratio of random-error variances of ship and buoy observations (Reynolds and Smith 1994).

Our additional analysis (not shown in figure) indicates that the random error variance of Argo SSTA is close to that of buoy SSTA over 2000–14, so the same factor is employed. The globally averaged mean difference and STD between buoy and Argo SSTAs over 2000–14 period are both, coincidentally, 0.03°C and are derived as follows. The difference between buoy and Argo SSTs is first calculated over the global ocean where both Argo and buoy SSTs are available within a $2° × 2°$ grid box. The global average is then calculated from 1998 to 2014, and finally the mean difference and STD are calculated based on global average series over 2000–14 since there were few Argo floats over 1999–2000. The buoy–Argo SST difference may be associated with the two types of different instruments and/or two different observing depths. Specifically, the depth of Argo SST (above 5-m depth) may not match with the depth of buoy floats (nominally 0.2 m). The mean difference of 0.03°C, albeit small, might not be trivial to the short-term SST trend since the number of Argo observations is increasing. Therefore, Argo SSTs are adjusted accordingly in ERSSTv5.

To filter out high-frequency noise due to measurement and sampling errors in time and small scale in

space, the monthly superobservations are separated into low- and high-frequency components as described in ERSSTv4 (Huang et al. 2015a). The low-frequency component is retrieved by applying a filter of 26° in both latitude and longitude and 15 years in time. The high-frequency component is retrieved first from the superobservations by subtracting the low-frequency component. The high-frequency component of SSTA is then reconstructed by a 3-month running filter and the 140 leading EOTs (van den Dool et al. 2000; Smith et al. 2008) that are computed using monthly WOISST from 1982 to 2011. However, the training processes are revised in ERSSTv5 over ERSSTv4 as follows:

- The number of EOTs has been increased to 140 in ERSSTv5 from 130 in ERSSTv4 by adding 10 EOTs in the Arctic region.
- The damping to the EOTs in the high latitudes (Smith et al. 2008) is removed in ERSSTv5.
- The spatial smoothing applied to the training data is reduced to a 6-degree running filter in ERSSTv5. In contrast, a 14-degree running filter was applied three times in EOT training data in ERSSTv4.

Tests (section 4b) indicated that these revisions in EOT computations can improve the spatial variability of the reconstruction.

For the high-frequency SSTA reconstruction, the sampling of each of the EOTs is evaluated for each month, as in earlier versions of ERSST (Huang et al. 2015a; Smith and Reynolds 2003; Smith et al. 1996). Here an EOT mode is accepted for reconstruction if the fraction of variance sampled is at least 0.1. This eliminates EOTs without enough sampling to adequately filter out data noise. The SSTA is finally formed by combining the low- and high-frequency components. Where sea ice concentration is greater than 60%, SSTs are adjusted toward the freezing point of $-1.8°C$ (Smith and Reynolds 2004).

4. Upgrades and their impacts on ERSSTv5

Eight progressive experiments (Table 1) were undertaken to demonstrate the efficacy of the proposed improvements to ERSSTv5. Each experiment quantifies the impact of a new update of a specific parameter or input data source in ERSSTv5, while all other parameters remain unchanged. The impacts of each consecutive change are thus assessed by the difference between two consecutive experiments. The differences were calculated in the latitudinal belts of 60°–90°N, 40°–60°N, 20°–40°N, 20°S–20°N, 20°–40°S, and 40°–90°S over the 1854–2015 time period (Fig. 3) and as the averaged differences (Fig. 4) and RMSD (Fig. 5) to assess

the spatial distribution and magnitude in each consecutive change. The averaged differences are mostly shown over 1860–1920 when the effects of consecutive changes are large owing to input data sparsity. The exceptions are ShipBias and Argo5m, both of which only affect modern period portions and whose averaged differences are thus evaluated over 2000–15 instead. The long-term (defined as the period 1900–2015) and short-term (defined as the period 2000–15) SST trends over different regions are also evaluated and compared. The impacts of these upgrades to ERSSTv5 are first described and then applied to the final ERSSTv5 based on their improvements over ERSSTv4 according to cross validations and validations against independent datasets whenever possible.

a. Use of unadjusted first guess

Assuming that random errors in raw SST observations are normally distributed, the outliers can be screened reasonably by comparison to an expectation of the mean state plus the normal distribution about that mean. The first-guess expectation should be an unbiased estimate of the mean of the input data being evaluated. Thus it should vary over time to account for both real-world trends and systematic biases that affect the data. In ERSSTv4 the first guess was based upon the adjusted ERSSTv3b data. This is not a major issue when the raw and bias-adjusted data are reasonably close. However, before the 1940s the adjusted FG is warmer than the unadjusted SST since the SST observations from bucket measurements are generally cold biased (see Fig. 8 and associated discussion below) (Huang et al. 2015a; Kennedy et al. 2011b). When the adjusted FG is used in the QC procedure as in ERSSTv4, more warm-tail observations may be included than cold-tail observations because the distributional mean against which tails are trimmed shall be warmer than the true distributional mean. Such an effect, all else being equal, would result in an artificially warm SST estimate and potentially skew the SST variability in the pre-1940 period reconstruction.

The study of SST uncertainty in ERSSTv4 confirmed that the selection of unadjusted FG in quality controlling raw SST observations can affect the SST long-term trend (Huang et al. 2016a). Therefore, the experiment UnadjFG, using the unadjusted mean as the first guess, was chosen as the first necessary change and compared against ERSSTv4 (Table 1). As expected, when unadjusted FG is used in ERSSTv5, the SSTA decreases by approximately 0.02°C over 40°–60°N (Fig. 3b, black dotted line) before the 1940s as a result of including more cold-tail SST input data. Over 1860–1920, the averaged difference is approximately $-0.1°C$ in the
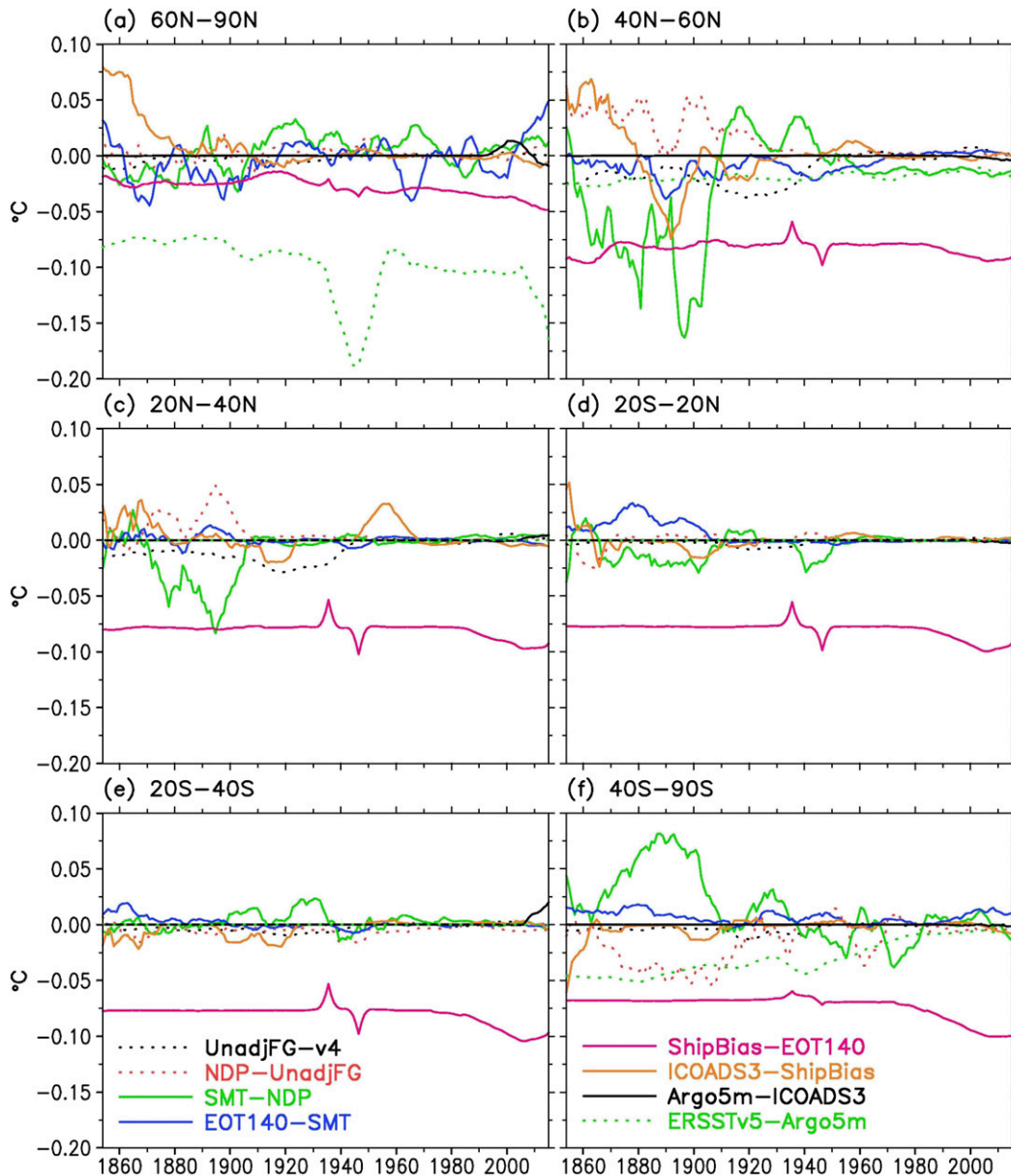
FIG. 3. SSTA differences averaged in (a) 60°–90°N, (b) 40°–60°N, (c) 20°–40°N, (d) 20°S–20°N, (e) 20°–40°S, and (f) 40°–90°S. The differences are calculated from progressive experiments: UnadjFG − v4 (dotted black line), NDP − UnadjFG (dotted red line), SMT − NDP (solid green line), EOT140 − SMT (solid blue line), ShipBias − EOT140 (solid pink line), ICOADS3 − ShipBias (solid orange line), Argo5m − ICOADS3 (solid black line), and ERSSTv5 − Argo5m (dotted green line). An 11-yr running filter is applied in plotting.

northwestern North Pacific and northwestern North Atlantic (Fig. 4a), and the RMSD is 0.1°–0.2°C in these regions (Fig. 5a). Overall, the long-term trend of the globally and regionally averaged SST increases by 0.01°C century^{-1} (Table 2, rows 2 and 3). In contrast, the short-term trend decreases by 0.02°C century^{-1} (Table 3, rows 2 and 3), which is associated with including less cold SST data in the 2000s when unadjusted FG is slightly warmer than the bias-corrected

FG used in ERSSTv4 (refer to the bias correction shown later in Fig. 8).

b. EOT revisions

1) USE OF UNDAMPED EOTs IN HIGH LATITUDES

The EOT modes in ERSSTv4 were damped in the high latitudes beyond 70°N and 60°S, to damp the reconstructed SSTAs where observations are very sparse
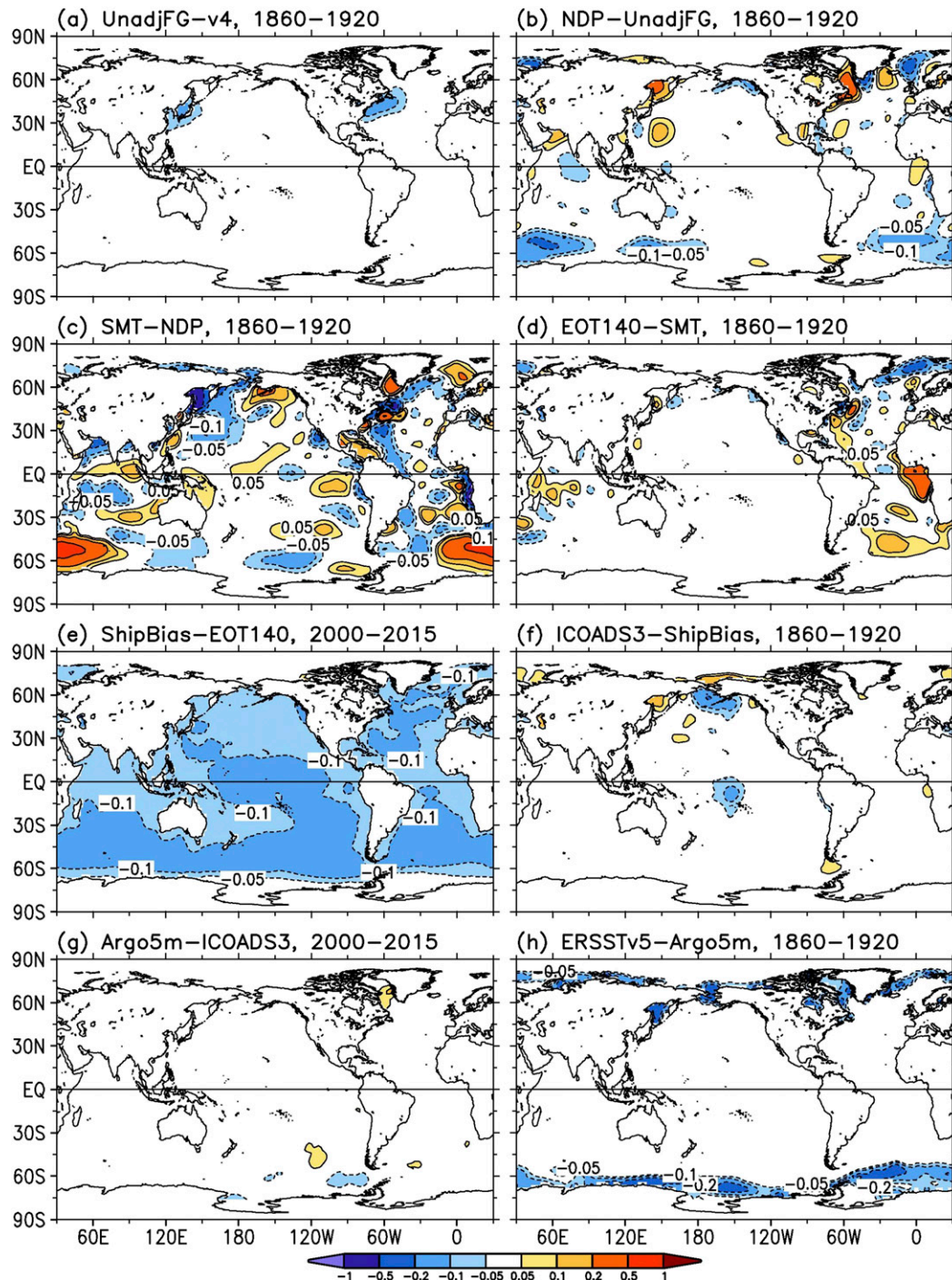
FIG. 4. Averaged differences calculated from progressive experiments: (a) UnadjFG − v4, (b) NDP − UnadjFG, (c) SMT − NDP, (d) EOT140 − SMT, (e) ShipBias − EOT140, (f) ICOADS3 − ShipBias, (g) Argo5m − ICOADS3, and (h) ERSSTv5 − Argo5m. The averages are calculated in periods of 1860–1920 in (a)–(d) and (f), and 2000–15 in (e) and (g).

(Smith et al. 2008). However, the number of observations has increased in these regions in the modern period and ICOADS R3.0 also has a greater number of early high-latitude observations. For example, the ship SST

observations increased by 40% in observation count and 2%–10% in areal coverage north of 60°N and south of 60°S after 2005. It is therefore possible that ERSSTv5 would benefit from removing the high-latitude damping
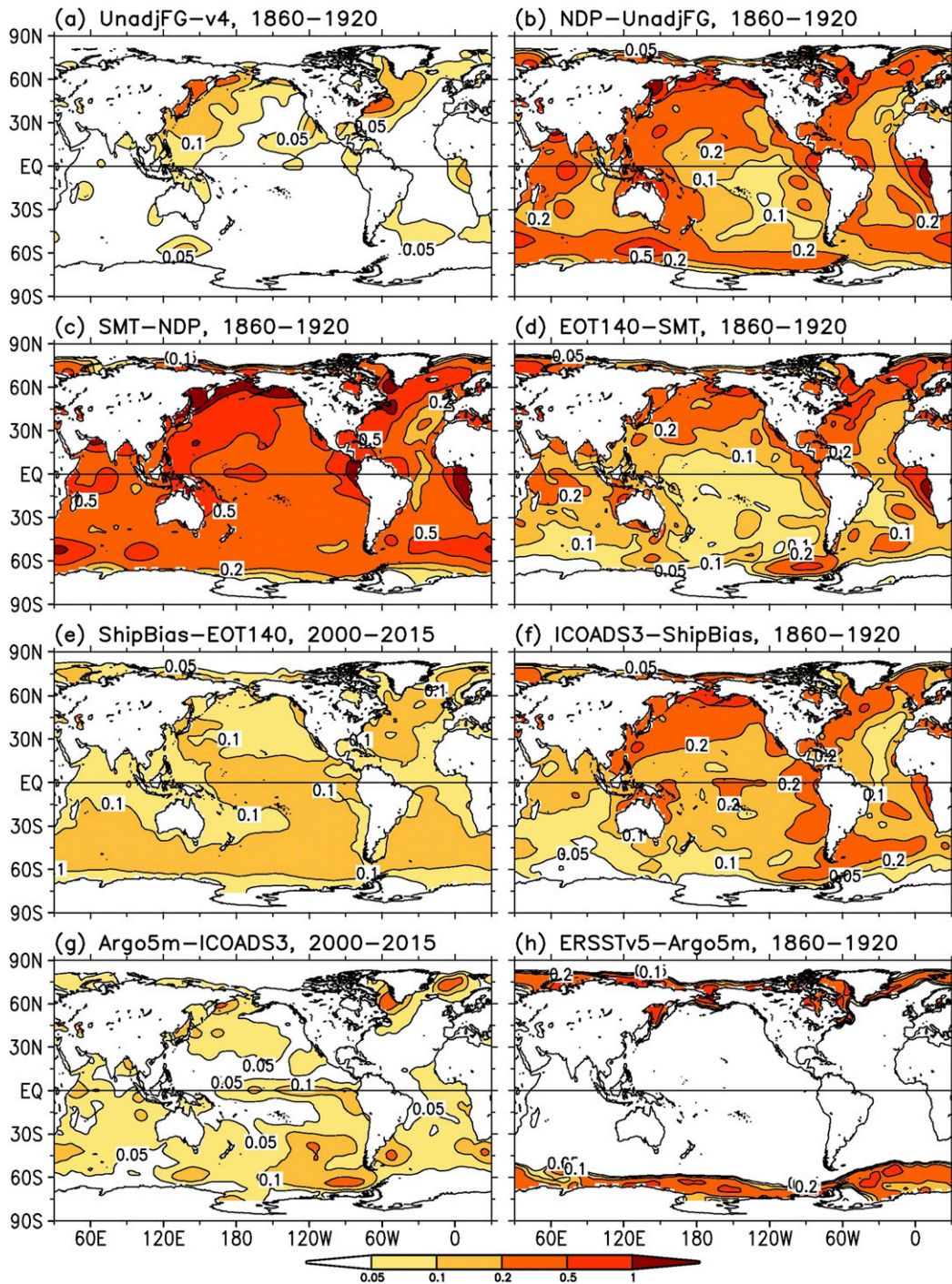
FIG. 5. As in Fig. 4, but for RMSD.

of EOTs, permitting more realistic expressions of variability and change in those areas. Because the EOTs have large spatial scales, removal of the damping can affect lower latitudes in each hemisphere through changes in ordering and weighting of EOTs that ensues. This step is denoted as NDP (Table 1).

In removing the high-latitude damping, averaged SSTA increases as much as 0.05°C in 40°–60°N before the 1930s (Fig. 3b, pink dotted line) and in 20°–40°N before the 1900s (Fig. 3c), and decreases as much as −0.05°C in 40°–90°S (Fig. 3f). The averaged difference has a magnitude of 0.1°–0.2°C in the high-latitude

TABLE 2. Trends (°C century^{-1}) over 1900–2015 and their uncertainty at 95% confidence level of annually and area-weighted averaged SSTA. The confidence level has taken into account the effective sampling number quantified by lag-1 auto correlation.

Exp	90°S–90°N	60°–90°N	40°–60°N	20°–40°N	20°S–20°N	40°–20°S	90°–40°S
ERSSTv4	0.69 ± 0.08	0.46 ± 0.11	0.56 ± 0.15	0.67 ± 0.11	0.75 ± 0.12	0.79 ± 0.11	0.63 ± 0.09
UnadjFG	0.70 ± 0.08	0.47 ± 0.11	0.60 ± 0.15	0.70 ± 0.11	0.76 ± 0.12	0.80 ± 0.11	0.64 ± 0.09
NDP	0.71 ± 0.08	0.47 ± 0.11	0.58 ± 0.15	0.69 ± 0.11	0.75 ± 0.13	0.80 ± 0.10	0.67 ± 0.09
SMT	0.71 ± 0.07	0.46 ± 0.11	0.57 ± 0.15	0.71 ± 0.11	0.76 ± 0.12	0.79 ± 0.10	0.65 ± 0.08
EOT140	0.71 ± 0.07	0.48 ± 0.12	0.59 ± 0.15	0.71 ± 0.11	0.76 ± 0.12	0.80 ± 0.10	0.65 ± 0.08
ShipBias	0.69 ± 0.07	0.46 ± 0.12	0.58 ± 0.15	0.69 ± 0.10	0.74 ± 0.12	0.78 ± 0.10	0.62 ± 0.08
ICOADS3	0.70 ± 0.07	0.46 ± 0.12	0.58 ± 0.15	0.70 ± 0.11	0.75 ± 0.12	0.79 ± 0.09	0.63 ± 0.08
Argo5m	0.70 ± 0.07	0.46 ± 0.12	0.58 ± 0.15	0.70 ± 0.11	0.75 ± 0.12	0.79 ± 0.09	0.62 ± 0.08
ERSSTv5	0.70 ± 0.07	0.44 ± 0.11	0.59 ± 0.15	0.70 ± 0.11	0.75 ± 0.12	0.79 ± 0.09	0.66 ± 0.08

oceans (Fig. 4b), and the RMSD is as high as 0.5°–1.0°C in the high latitudes (Fig. 5b). Both long-term and short-term trends of the globally averaged SST increase slightly, primarily due to a change in the representation of warming in the Southern Ocean (Tables 2 and 3, rows 3 and 4).

The removal of the damping has a minor influence on SSTA behavior globally in the data-sparse period prior to the 1940s. Otherwise the influence is largely limited to higher latitudes. Cross validations [section 4b(4)] show a slight improvement by removing the damping. The original rationale for applying the damping was to avoid the situation where a small number of high-latitude observations could have a large influence on the global mean (Smith and Reynolds 2003, 2004). Given the rather minor effect on global mean series, we undo this damping in ERSSTv5.

2) REDUCTION IN SMOOTHING OF EOTs

In ERSSTv4, the training data for the EOT calculation were filtered three times by a 14-degree running spatial average. This strong filtering resulted in a set of smooth EOTs, which resulted in a smooth reconstructed SSTA field over the global oceans and weaker El Niño and La Niña magnitudes than in ERSSTv3b. On the flip side, in periods when the data constraint is sparse these choices avoid potential undue weight being given to single sparse observations and yield an appropriately conservative estimate of our ability to reconstruct details of the SST fields. There is a continuum of choices that could be made in regard to how aggressively to smooth. A range of user feedback on ERSSTv4 suggested that the filtering was too aggressive for a range of user needs.

We have therefore tested and reduced the strength of the smoothing to a 6-degree running average in experiment SMT (Table 1). The SSTA changes in comparison with experiment NDP are 0.05°–0.15°C in 40°–60°N in the 1880s–1900s (Figs. 3b,c,f, solid green lines). The averaged (1860–1920) differences are 0.1°–0.5°C in the Southern Ocean and the coastal regions and 0.05°–0.1°C in other regions of the global oceans (Fig. 4c). The RMSDs are 0.5°–1.0°C in the Southern Ocean and the coastal regions (Fig. 5c). The long-term trend of globally averaged SSTA does not change much (Table 2, rows 4 and 5) owing to cancellations in the regions of 20°S–40°N, south of 20°S, and north of 40°N.

The reduction in smoothing has some effects regionally and over short periods but has limited impact upon the global mean behavior. The short-term trend of globally averaged SSTA decreases by 0.03°C century^{-1}, which is largely associated with a decreased (by 0.22°C century^{-1}) SSTA trend over 90°–40°S (Table 3, rows 4 and 5). The subsequent cross validation [section 4b(4)] shows that the reduction in smoothing clearly reduces the reconstruction error, particularly in the tropical

TABLE 3. As in Table 2, but for trends and uncertainties (°C century^{-1}) over 2000–15.

Exp	90°S–90°N	60°–90°N	40°–60°N	20°–40°N	20°S–20°N	40°–20°S	90°–40°S
ERSSTv4	1.34 ± 0.75	2.65 ± 0.63	2.61 ± 1.24	1.72 ± 1.00	1.45 ± 1.56	0.97 ± 0.55	0.46 ± 0.70
UnadjFG	1.32 ± 0.75	2.62 ± 0.63	2.54 ± 1.24	1.67 ± 1.06	1.44 ± 1.56	0.95 ± 0.54	0.46 ± 0.70
NDP	1.33 ± 0.76	2.75 ± 0.61	2.50 ± 1.22	1.67 ± 1.01	1.44 ± 1.57	0.95 ± 0.54	0.50 ± 0.75
SMT	1.30 ± 0.77	2.88 ± 0.54	2.49 ± 1.20	1.68 ± 1.06	1.46 ± 1.57	0.94 ± 0.51	0.28 ± 0.63
EOT140	1.29 ± 0.75	3.11 ± 0.70	2.44 ± 1.19	1.64 ± 1.00	1.47 ± 1.57	0.92 ± 0.54	0.27 ± 0.56
ShipBias	1.29 ± 0.77	3.00 ± 0.71	2.44 ± 1.27	1.62 ± 1.09	1.50 ± 1.58	0.93 ± 0.53	0.23 ± 0.57
ICOADS3	1.27 ± 0.77	2.90 ± 0.74	2.51 ± 1.20	1.60 ± 1.03	1.51 ± 1.56	0.90 ± 0.52	0.15 ± 0.56
Argo5m	1.29 ± 0.77	2.65 ± 0.68	2.48 ± 1.27	1.66 ± 1.02	1.49 ± 1.55	1.09 ± 0.50	0.16 ± 0.54
ERSSTv5	1.25 ± 0.77	2.23 ± 0.66	2.41 ± 1.28	1.66 ± 1.02	1.49 ± 1.55	1.09 ± 0.50	0.13 ± 0.53

TABLE 4. Averaged (60°–90°N) RMSD (°C) between the ERSST analyses using historical sampling grids and WOISST. The 10-yr validation data are independent from the other 24-yr EOT training data. The 24-yr training data for EOTs are over 1992–2015 in the analysis periods in columns 2–4 and over 1982–2005 in the analysis period in columns 5–7.

Exp	1882–91	1932–41	1982–91	1906–15	1956–65	2006–15
UnadjFG	0.57	0.75	0.63	0.65	0.68	0.66
NDP	0.56	0.76	0.62	0.65	0.67	0.66
SMT	0.57	0.78	0.63	0.71	0.68	0.68
EOT140	0.51	0.60	0.56	0.52	0.62	0.62

Pacific Niño-3.4 region. Hence we apply the reduction in smoothing in ERSSTv5.

3) MORE EOT MODES

An additional 10 Arctic EOTs were computed for experiment EOT140 (Table 1) to better represent SST observations in the Northern Hemisphere high latitudes. As expected, the SSTA changes due to the increased number of EOT modes in comparison with experiment SMT are mostly in 60°–90°N (Fig. 3a, solid blue line) and 40°–60°N (Fig. 3b). There is also a change in the tropics (20°S–20°N) before the 1910s (Fig. 3d). The change in the tropics is associated with an increased SSTA in the eastern equatorial Atlantic near the coast of Africa in 1860–1920 (Figs. 4d and 5d). Analyses indicate that there are few observations in the eastern equatorial Atlantic in 1860–1920, and therefore the reconstruction is sensitive to the selection of EOT modes supported by nearby observations. Similar features happen in experiments SMT (Fig. 5c) and NDP (Fig. 5b) when EOTs are revised. Despite changes in local SSTAs, both long-term and short-term SSTA trends do not change much (Tables 2 and 3, rows 5 and 6).

Given the small impact on global time series and the clear improvement in representing the Arctic changes as shown in the subsequent cross validations [section 4b (4)], we adopt the additional 10 Arctic EOTs in ERSSTv5.

4) VALIDATION OF EFFICACY OF EOT CHANGES IN ERSSTv5

As described in sections 4b(1)–4b(3), the upgrades in EOTs can result in changes in local SSTAs, although changes in the global trends are slight. Cross validations indicate that overall these local SSTA changes represent improvements in reconstructing available observations in ERSSTv5. For cross-validation testing, the monthly averaged WOISST (1982–2015) is used. The WOISST data are separated into two parts: one for calculating EOTs and the other independent period for analysis. The EOTs are calculated in the same way as in experiments UnadjFG, NDP, SMT, and EOT140, and therefore the same experiment names are used (but note that

they are for different periods of training data). Two sets of 24-yr training data are selected, 1992–2015 and 1982–95, leaving two 10-yr independent periods, 1982–91 and 1996–2015 respectively. The 34 years of WOISST data are subsampled, using the superobservation masks for historic observations in 1882–1915, 1932–65, and 1982–2015. The subsampled WOISST is reconstructed by ERSST method and is cross-validated in the independent periods against original WOISST, which gives estimates of errors in the historical analyses.

The comparisons for the different independent periods are shown in Tables 4–7. The pointwise RMSDs between reconstructions and perfect data are assessed and averaged in the high latitudes of 60°–90°N (Table 4) and in the tropical Pacific Niño-3.4 region (5°S–5°N, 170°–120°W; Table 5). In comparison with experiment UnadjFG, the averaged RMSDs in experiment EOT140 decrease by 0.06°–0.15°C in the high latitudes of 60°–90°N (Table 4) and by 0.02°–0.07°C in the Pacific Niño-3.4 region (Table 5) in all validation periods. Similar features are found in other regions of the global oceans. The experiment NDP also indicates reduced RMSDs over 60°–90°N, although not as much as in experiment EOT140. However, the RMSDs in experiment SMT increase the RMSDs slightly over 60°–90°N in comparison with experiment NDP.

Similarly, RMSDs of averaged SSTAs over 60°–90°N (Table 6) and Niño-3.4 (Table 7) regions are calculated. Comparisons indicate that the RMSDs in EOT140 in comparison to UnadjFG decrease by 0.01°–0.07°C over 60°–90°N and by 0.02°–0.12°C in Niño-3.4 regions in all validation periods. The reduction in RMSDs in the Niño-3.4 region shows that the net effect of the EOT changes improves the fidelity of the reconstructed

TABLE 5. As in Table 4, but for the Niño-3.4 region (5°S–5°N, 120°–170°W).

Exp	1882–91	1932–41	1982–91	1906–15	1956–65	2006–15
UnadjFG	0.30	0.31	0.29	0.26	0.24	0.26
NDP	0.31	0.31	0.29	0.26	0.24	0.25
SMT	0.25	0.23	0.22	0.24	0.21	0.23
EOT140	0.25	0.24	0.23	0.24	0.21	0.23

TABLE 6. RMSD (°C) between averaged (60°–90°N) SSTA series in the ERSST analyses using historical sampling grids and using WOISST. The 24-yr training data for EOTs are over 1992–2015 in columns 2–4 and over 1982–2005 in columns 5–7.

Exp	1882–1915	1932–65	1982–2015	1882–1915	1932–65	1982–2015
UnadjFG	0.18	0.08	0.11	0.20	0.10	0.10
NDP	0.21	0.08	0.10	0.20	0.10	0.08
SMT	0.21	0.10	0.12	0.19	0.09	0.10
EOT140	0.17	0.07	0.04	0.14	0.05	0.04

Niño-3.4 index. Figure 6 shows an example of the Niño-3.4 indices in three time periods (1882–1915, 1932–65, and 1982–2015) in UnadjFG, EOT140, and WOISST when EOTs are trained by 1992–2015 data. The magnitudes of major El Niño and La Niña events in EOT140 are much closer to the original WOISST than those in UnadjFG, particularly when observation data become dense after 1932 (Figs. 6b,c). The magnitudes of major El Niño events are also improved in EOT140 in comparison with UnadjFG before 1915 (Fig. 6a), but they remain weaker than the WOISST, indicating the difficulty in reconstructing El Niño events when data are sparse.

Not only is the magnitude of El Niño and La Niña improved, the spatial distributions of SSTAs become more faithful reconstructions of the underlying SSTA fields. For example, the WOISST SSTA in December 1982 (Fig. 7a) shows one of the strongest El Niño events (Huang et al. 2016c). How does the reconstruction method cope when given this El Niño event but with the sampling coverage constraint of a century prior? To answer this, the data in December 1982 is filtered using a December 1882 observation mask and reconstructed using the EOTs trained with 1992–2015 data (hence the training period is independent). The reconstructed SSTAs differ from WOISST in all of the tests: UnadjFG (Fig. 7b), NDP (Fig. 7c), SMT (Fig. 7d), and EOT140 (Fig. 7e). But the differences are considerably reduced in the consecutively applied experiments. The positive SSTA in the central-eastern tropical Pacific is much smoother (and hence unrealistic), and the area of 3°C contour is much smaller in UnadjFG and NDP than in WOISST. In contrast, these issues are greatly reduced in SMT and EOT140 due to the reduced filtering of EOTs. In addition, the positive SSTA in the South Pacific near 140°W is damped south of 65°S in UnadjFG but stretches reasonably toward the Antarctic in NDP,

SMT, and EOT140 due to the removal of damping in EOT training.

To test whether the improvements shown in experiments NDP, SMT, and EOT140 (Tables 4–7) are robust, parallel experiments are designed in which a set of random errors is added to the validation data when subsampled with historic observation masks. The random error has a mean of zero and an STD of 1.0°C. The magnitude of the random errors is selected to be between those of ship (1.3°C) and buoy (0.5°C) observations (Reynolds et al. 2002). The tests confirm that RMSDs decrease very similarly as shown in Tables 4–7, except that the magnitude of RMSDs in the parallel experiments increase by 0.02°–0.05°C owing to the inclusion of random errors.

Overall, the cross validation using modern data but historical sampling masks suggests that the changes in EOTs produce a more realistic set of SSTAs if the input data are unbiased. As with all processing choices the uncertainty increases in the early, data-sparse, period. We therefore believe that, independent of questions about efficacy of remaining steps in the method, the changed treatment of EOTs improves the representation of SSTAs and provides more realistic reconstructions of important aspects such as Niño-3.4 temperature series used in downstream applications.

c. Bias correction of ship SSTs

1) REVISION OF SHIP SST BIAS CORRECTION BASIS

In ERSSTv4, biases of ship SSTs were first corrected using HadNMAT2 (Kent et al. 2013) before 2010 (Fig. 8, dotted black line), and buoy SSTs were then adjusted to agree with the ship observations by subtracting a globally averaged buoy–ship offset (−0.12°C) over 1990–2010 (Fig. 8, dotted red line) (Huang et al. 2015a). This

TABLE 7. As in Table 6, but for Niño-3.4 region.

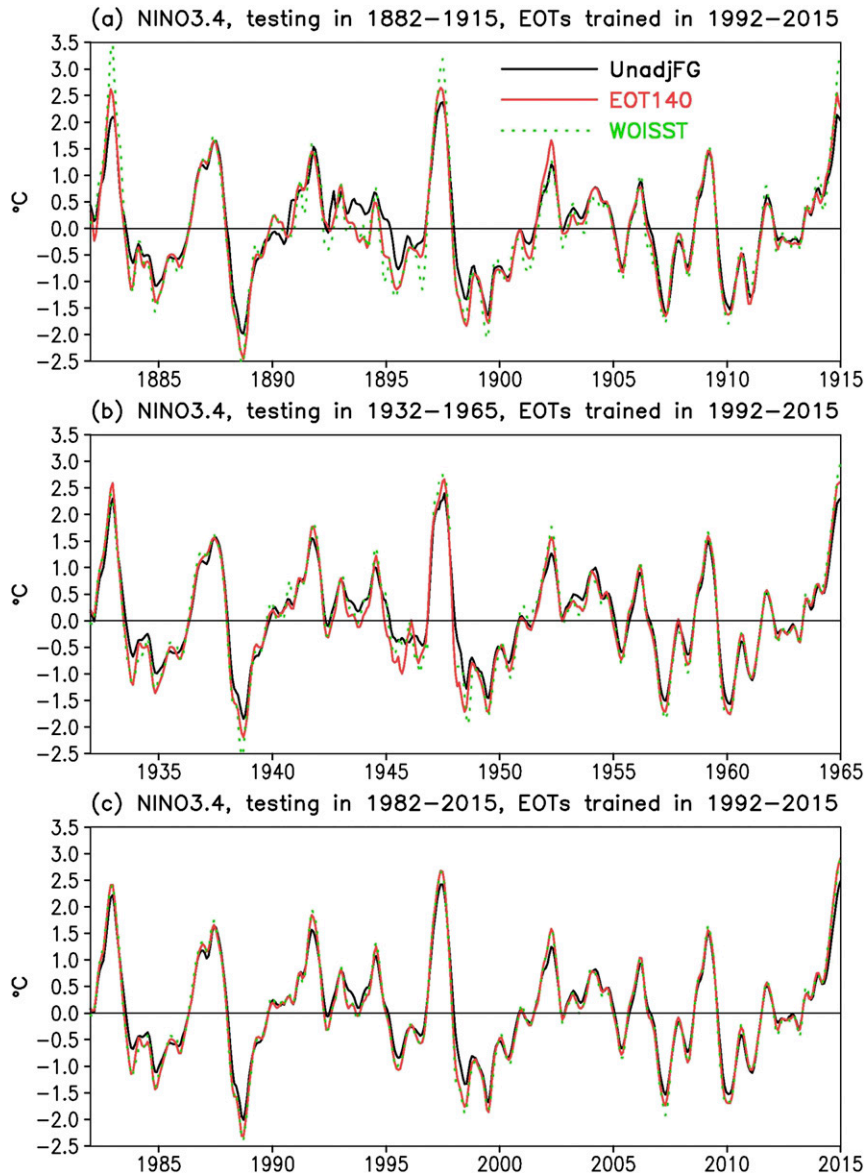| Exp	1882–1915	1932–65	1982–2015	1882–1915	1932–65	1982–2015
UnadjFG	0.34	0.23	0.19	0.30	0.18	0.13
NDP	0.35	0.25	0.20	0.30	0.19	0.13
SMT	0.22	0.17	0.06	0.23	0.16	0.06
EOT140	0.22	0.17	0.07	0.23	0.16	0.06

FIG. 6. Niño-3.4 indices of UnadjFG (solid black line), EOT140 experiment (solid red line), and WOISST (dotted green line) in (a) 1882–1915, (b) 1932–65, and (c) 1982–2015 testing periods. The validation data are from WOISST over 1982–2015 and are subsampled and reconstructed with observation masks in periods in (a)–(c). The RMSDs are shown in Table 4 (columns 1–3). EOTs are trained with WOISST 1992–2015 data. A 3-month running filter is applied in plotting.

approach has several drawbacks: 1) studies have shown that buoy SSTs are more well behaved, with lower spread and greater interplatform (buoy to buoy vs ship to ship) consistency than ship SSTs (Reynolds and Smith 1994; Reynolds et al. 2002); 2) the offset between buoy and ship SSTs has never been a constant in time (Fig. 8, dotted red line); and 3) the biases of ship SSTs after 2010 were corrected using the periodic monthly corrections in 2010 since HadNMAT2 was not updated after 2010

(now updated to 2014), while the tendency of the biases over 2000–10 is very clear albeit small in magnitude (Fig. 8, dotted black line).

These issues have been addressed in ERSSTv5: The SSTs before 2010 are corrected by NMAT and readjusted by subtracting 0.077°C [i.e., the difference between the NMAT-based estimate (−0.041°C) and the buoy-based offset estimate (−0.118°C) over 1990–2010]. The SSTs after 2010 are corrected using only buoy SSTs.
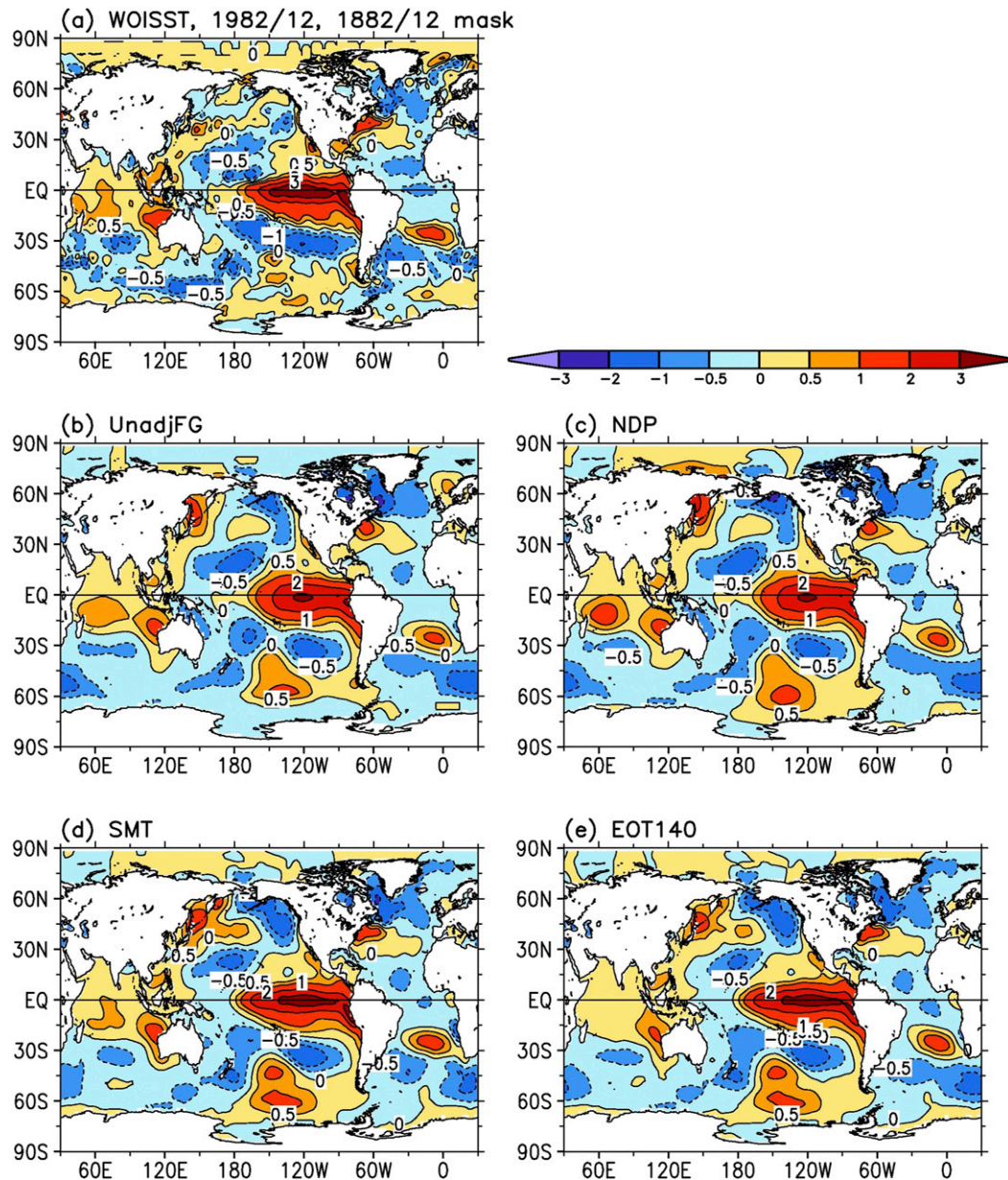
FIG. 7. SSTAs from (a) WOISST, (b) UnadjFG with EOT130, (c) NDP, (d) SMT, and (e) EOT140 experiments. WOISST in December 1982 in (a) is used as validation data and filtered with observation masks in December 1882 and reconstructed in (b)–(e). The RMSDs are 0.50°, 0.52°, 0.48°, and 0.48°C in (b)–(e), respectively, relative to (a). EOTs are trained with WOISST 1992–2015 data.

Importantly, the use of buoy SSTs enables the month-to-month updates of ship SSTs to be bias corrected without any dependency upon a third party data source (for corrected NMAT data). It also addresses the issue that the assumption of a constant SST–NMAT difference may increasingly not hold in a warming climate. In Fig. 4 of Huang et al. (2015a) it was shown that there was a slight discrepancy between MAT and SST in climate model historical runs but contended this effect to be very slight

relative to the substantive data bias issues in the raw SST observations through time. Subsequent analyses indicate that the NMAT and SST although highly correlated shall diverge somewhat in their trend behavior under transient change with NMAT showing more warming than the SSTs (Cowtan et al. 2015). Most importantly, the offset between buoy and ship SSTs is no longer set to a constant in time, which is more realistic according to ship and buoy observations over 1990–2015 (Fig. 8).
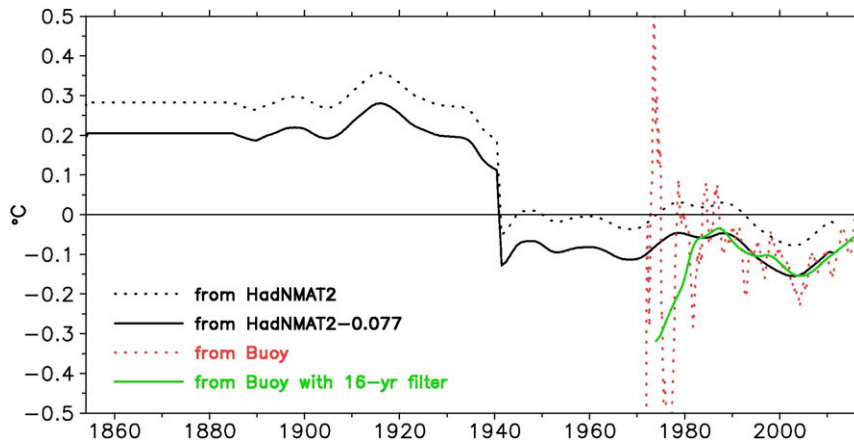
FIG. 8. Globally averaged ship SST bias corrections (added to observations) derived from NMAT (dotted black line) (Huang et al. 2015a), buoy SST (dotted red line), and its low-frequency (16 yr) component (solid green line). The combined bias corrections derived from HadNMAT2 with an offset of 0.077°C (solid black line) before 2010 and derived from buoy SST with 16-yr filtering (solid green line) after 2010 are used in ERSSTv5. The offset is the difference between buoy-based correction (−0.118°C) and NMAT-based correction (−0.041°C) over 1990–2010.

To achieve this change, first the adjustment applied to buoy SSTs in ERSSTv4 has been removed. Second, the globally averaged buoy–ship differences are filtered by a Lowess filter equivalent to 16-yr low-pass filter (Fig. 8, solid green line), an approach identical to that applied in NMAT-derived corrections in ERSSTv4. The averaged (1990–2010) difference (0.077°C) is calculated between the ship SST bias corrections derived from HadNMAT2 (−0.041°C; Fig. 8, dotted black line) and the buoy–ship offset (−0.118°C; Fig. 8, solid green line), noting the very close congruence (with a correlation coefficient of 0.87 over 1990–2010) in behavior of these two curves over this period. Third, the NMAT-based ship SST

bias corrections are readjusted by subtracting 0.077°C (Fig. 8; solid black line) so that the corrections between 1985 and 2010 match with the equivalent buoy–ship offset. These procedures readjust the ship SST bias correction with minimal changes in temporal variations of ship SSTs.

The reconstructed SSTAs in experiment ShipBias (Table 1) are consistent with that in EOT140 except for a systematic cooling offset of approximately 0.077°C (relative to a common climatology, and prior to rebasing; Figs. 9, 3, 4e, and 5e), as expected. However, the long-term trend (Table 2, rows 6 and 7) of globally averaged SSTAs is reduced slightly by 0.02°C century^{-1}.
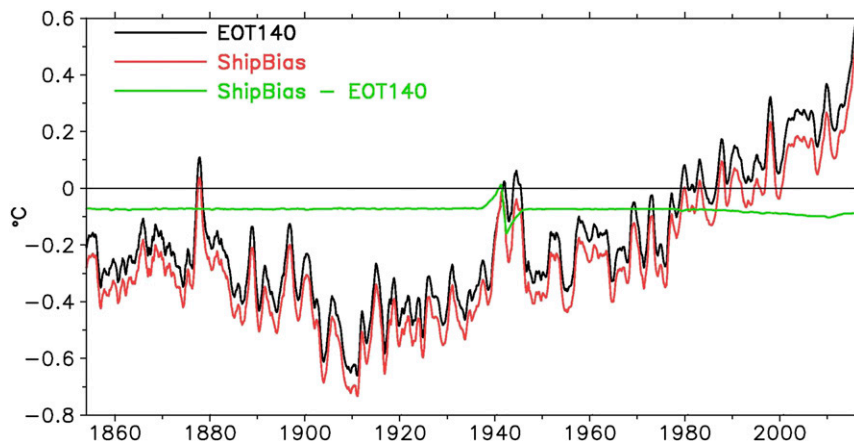

FIG. 9. Globally averaged SSTA in experiments EOT140 (solid black line), ShipBias (solid red line), and their difference of ShipBias − EOT140 (solid green line). A 12-month running filter is applied in plotting.

The reduction relates to changes between adding 0.12°C to buoy observations in ERSSTv4 and subtracting 0.077°C from ship observations in ERSSTv5 in post-1990 data. Because the number of ship observations is much smaller than that of buoy observations, as well as buoy SSTs that are weighted almost 7 times larger than ship SSTs, the net effect is to slightly reduce estimates of warming in the period that affects trends for both periods being considered. In contrast, the short-term trend does not change (Table 3, rows 6 and 7) since similar adjustments are made throughout the recent decades.

The application of buoy-based adjustments allows us to account for time-varying biases in ship–buoy differences going forward, removes an operational update dependency upon third-party products, and avoids issues around potential divergence between NMAT and SST measurands moving forward under a changing climate system. The change therefore makes the operational basis for ERSSTv5 updates easier and scientifically more rigorous. The change also improves overall reconstruction error in comparison with independent Argo5obs over 2004–15 as shown in the subsequent discussion in section 4c(3). Given close congruence of ship–buoy and ship–NMAT series over the period of dense buoy deployment, which permits a consistent transition, we adopt this approach in ERSSTv5. However, potential small spatial variabilities of the offsets being applied are not considered in the first-order approximation for the following reasons: 1) the NMAT-based correction is relatively uniformly distributed over the global oceans after the 1940s, which is consistent with the globally averaged approximation of buoy-based correction; and 2) the magnitude of both NMAT- and buoy-based correction is small (0.1°–0.2°C) after the 1940s while the magnitude of SSTA locally in any given month is generally large (1°–2°C), and therefore the potential local discontinuity between NMAT-based correction before 2010 and buoy-based correction after 2010 is negligible.

2) APPLICATION OF AN A PRIORI ESTIMATE OF THE 1940S BUCKET CORRECTION

User feedback has highlighted concerns about the behavior of SST series from ERSSTv4 over the period when the measurement method rapidly changed from predominantly buckets to predominantly engine room intakes (i.e., the Second World War). In ERSSTv4, the biases of ship SSTs were calculated using the Lowess filter applied to the annual fitting coefficient of SST-NMAT over 1875–2010 (Huang et al. 2015a). The purpose of the Lowess filter is to account for long-term variations of ship SST biases, but the filter cannot necessarily fully account for the sudden change in ship SST biases due to the abrupt change that we know occurred.

To attempt to better account for the effect, in ERSSTv5, the fitting coefficient is first fitted linearly over the periods of 1875–1941 and 1942–2010. The residual between the original and linearly fitted coefficients in 1875–2010 is then filtered by the Lowess filter. The combination of Lowess filtered residual and linearly fitted coefficients is used to recalculate the ship SST bias correction, which results in a steeper change in ship SST bias correction in the 1930s–40s (Fig. 8) while the correction in other periods remains the same.

Conceptually, such an approach has precedence in the U.S. Historic Climatology Network (USHCN; Menne et al. 2009) whereby the known impacts of observation bias are quantified by a homogenization algorithm. In employing this revised approach in ERSSTv5 we are effectively performing an initial adjustment estimate for the 1940s transition, which is the largest artifact in the entire record, and then letting the Lowess filter nuance this estimate and find and adjust for additional data issues. On a methodological basis this is preferable, even if the effects are relatively small. The verification using the independent SOI index [section 4c(3)] shows a clear improvement in the evolution of global averaged SST. This change is therefore adopted in ERSSTv5.

3) VALIDATION OF CHANGES IN SHIP BIAS CORRECTIONS

Since Argo5obs are not included until a later experiment (Argo5m), Argo5obs are used here to validate the improvement in ShipBias over the period 2004–15 since Argo5obs are very sparse before 2004. Comparisons show that the averaged (2004–15) difference is higher in EOT140 (Fig. 10a) than in ShipBias (Fig. 10b). The globally averaged difference is approximately 0.13°C in EOT14 and 0.02°C in ShipBias. The error of globally averaged SSTA is near 0.1°C in ERSSTv4 (Fig. 11a, solid black line) and EOT140 (Fig. 11a, dotted green line) but decreases to nearly 0.0°C in ShipBias (Fig. 11a, dotted red line). The error in the Niño-3.4 region is also reduced by approximately 0.1°C in ShipBias in comparison with ERSSTv4 and EOT140 (Fig. 11b). However, the RMSDs of EOT140 and ShipBias relative to Argo5obs over 2004–15 are very close: they are approximately 0.62°C in EOT140 (Fig. 10d) and 0.61°C in ShipBias (Fig. 10e). The RMSDs of averaged SSTs in global and Niño-3.4 regions decrease, albeit by a small amount, in ShipBias in comparison with ERSSTv4 and EOT140 (Figs. 11c and 11d).

The global averaged SSTA is correlated with El Niño and La Niña, which are directly associated with the SOI derived from atmospheric sea level pressure (Trenberth
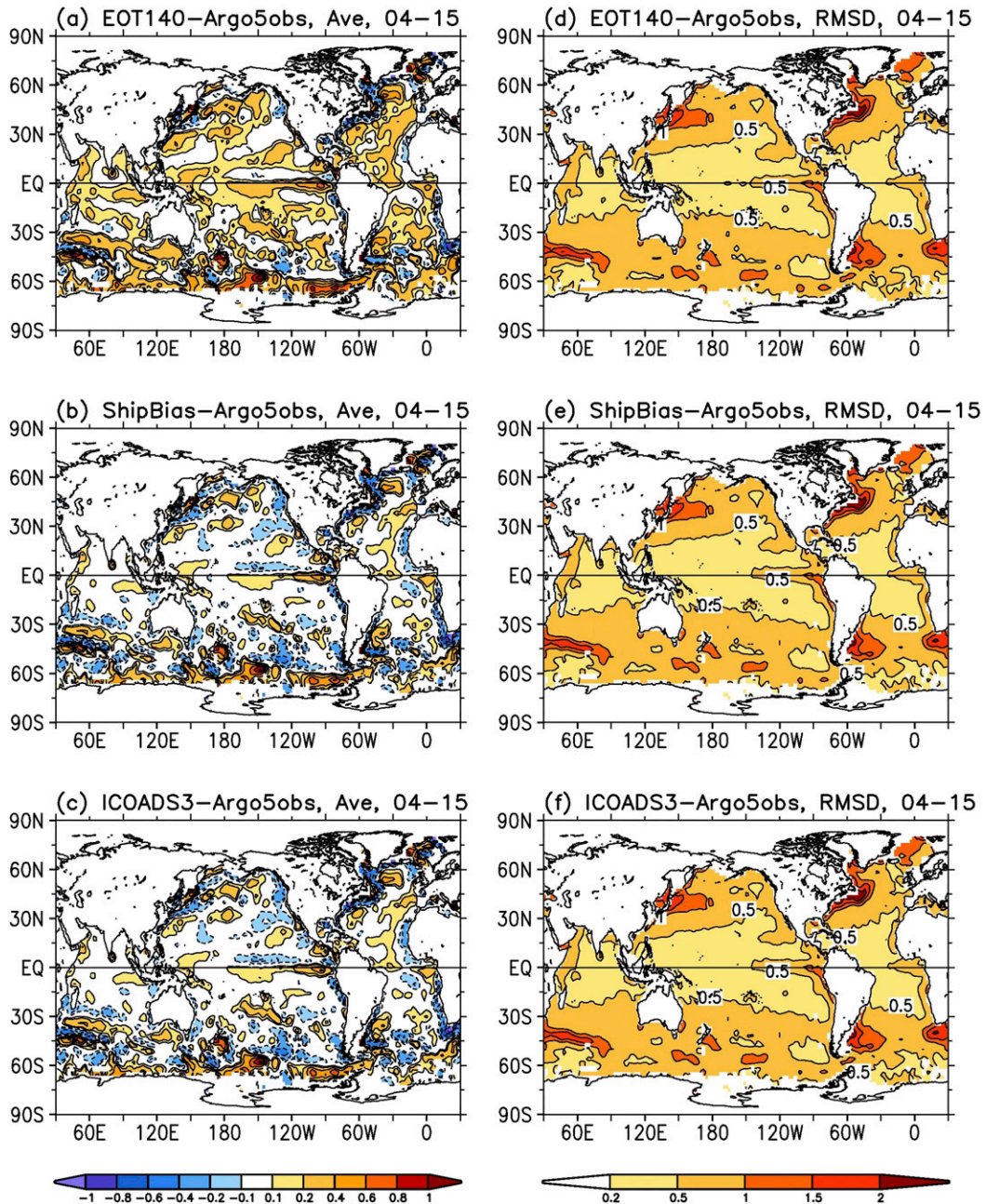
FIG. 10. Averaged (2004–15) difference in (a) EOT140, (b) ShipBias, and (c) ICOADS3 relative to Argo5obs. (d)–(f) As in (a)–(c), but for RMSD. The globally averaged values are 0.13°, 0.02°, 0.02°, 0.62°, 0.61°, and 0.60°C in (a)–(f), respectively.

and Stepaniak 2001). The correlation between global annually averaged SSTA and SOI is approximately 0.4 over 1880–2015. Therefore, the SOI is used to validate the revision in ship SST bias correction in the 1930s–40s. The validations show that the correlation between globally averaged SSTAs and SOI increases from 0.37 in EOT140 to 0.65 in ShipBias in the short period of 1937–45, which suggests a potentially better representation

in global SSTA in ShipBias, which includes the change in adjustment approach.

To the extent that validation is possible, both changes to the ship bias correction deployed in ERSSTv5 lead to improved performance. In the modern era the high-quality Argo data provide very strong constraints and show greatly improved performance. In the 1940s we have no such luxury, but the only
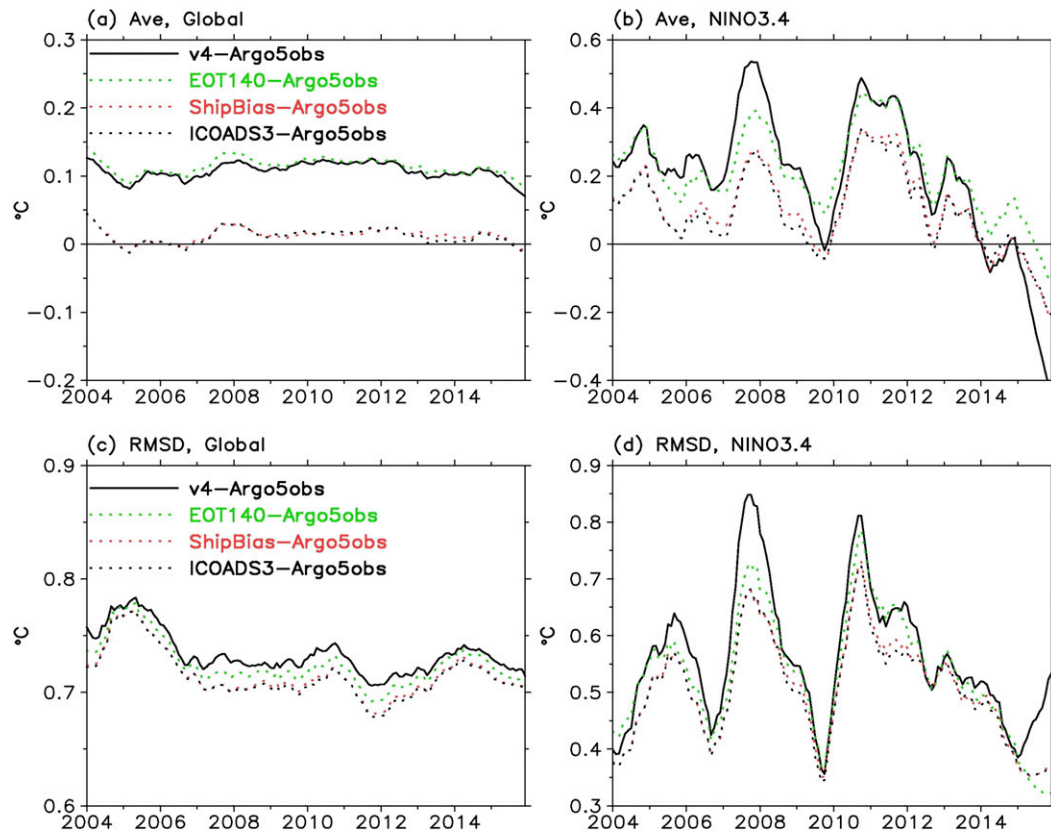
FIG. 11. Averaged difference in (a) global and (b) Niño-3.4 (5°S–5°N, 120°–170°W) regions and averaged RMSD in (c) global and (d) Niño-3.4 regions. The difference and RMSD are relative to Argo5obs and quantified in ERSSTv4 (solid black line), EOT140 (dotted green line), ShipBias (dotted red line), and ICOADS3 (dotted black line). A 12-month running mean is applied in plotting.

available metric, albeit a weak constraint, suggests improved performance.

d. New datasets

1) ICOADS R3.0

Experiment ICOADS3 uses the latest in situ SST observations from ICOADS R3.0 (Freeman et al. 2017) as the in situ data source, while ICOADS R2.5 was used in ERSSTv4. In comparison with R2.5, R3.0 contains more observations throughout, but particularly in the 1850s, 1950s–60s, 1990s–2000s (Fig. 1a), and its areal coverage increases in the 1850s and 2000s–10s. The reconstructed SSTA difference between ICOADS3 and ShipBias is approximately 0.05°C over 60°–90°N before the 1880s (Fig. 3a, solid orange line) and over 40°–60°N before the 1900s (Fig. 3b). The long-term trend of globally averaged SSTA increases slightly (Table 2, rows 7 and 8). However, the short-term trend (Table 3, rows 7 and 8) decreases slightly as a result of reduced SST trends in the high latitudes: 60°–90°N (0.10°C century^{-1}) and 40°–90°S (0.08°C century^{-1}). Analyses suggest (not

shown in figure) that the reduction in the short-term trend south of 40°S is associated with the increased ship observations south of 40°S, where the SST trend is generally lower owing to sea ice melting and strong vertical mixing (Huang et al. 2016a). However, the validations against Argo5obs over 1999–2015 show that the averaged difference and RMSD are comparable in ICOADS3 (Figs. 10c and 10f) and ShipBias (Figs. 10b and 10e), which is also the case for the averaged SSTAs in global and Niño-3.4 regions (Fig. 11, dotted red line vs dotted black line).

Given the substantial improvements in the raw data holdings encapsulated in ICOADS R3.0, ERSSTv5 transitions to this updated data source. Other SST product providers shall also likely transition in the future to ICOADS R3.0 or its successors and in future ERSST version releases we intend to do likewise.

2) ARGO SST ABOVE 5 M

A criticism leveled at ERSSTv4 was that it did not utilize the available data from Argo profilers. In going from ERSSTv3b to ERSSTv4 substantial upgrades and

updates were incorporated, and our primary interest was the centennial time scale changes. As such, we did not consider Argo measurements at that time because Argo data only exist since the late 1990s, and their role in near-surface monitoring was less well advanced. Following Karl et al. (2015) there is interest in the recent period data that, together with an ever-lengthening record and the high measurement system quality, led to revisiting the question of inclusion of Argo data within ERSSTv5. The question is vexed. On the one hand retaining Argo as an independent estimate would permit strong independent validation. On the other hand these data are some of the best data available, and, unlike drifting buoys, they tend not to coalesce around ocean surface current convergence zones and hence sample many regions infrequently visited by drifting buoys. Noting that there is no right answer on the issue; we eventually decided to include these data on the grounds of improved resilience and data constraint availability for informing monitoring activities.

The Argo5obs have first been derived [section 2a(2)] and then compared against collocated buoy SSTs (section 3). The small difference (0.03°C) between Argo5obs and buoy SSTs indicates good agreement between buoy and Argo SST observations, but this difference needs to be corrected for given the rapid increase in Argo observations. Therefore, the averaged offset of 0.03°C is added to the Argo5obs and merged with ship and buoy observations with a weighting coefficient of 6.8 (identical to that applied to buoys) in experiment Argo5m (Table 1).

The differences between Argo5m and ICOADS3 indicate that the changes due to including Argo5obs are small over the global oceans after 1999, except for a slight SSTA change in 60°–90°N (Fig. 3a) and 20°–40°S (Fig. 3e). The averaged (1999–2015) SSTA difference is small (Fig. 4g), but the RMSD is 0.05°–0.10°C in the eastern equatorial Pacific, northern North Atlantic, northern North Pacific, and Southern Ocean (Fig. 5g). The long-term trend of globally averaged SSTAs does not change (Table 2, rows 8 and 9), while the short-term trend increases slightly (Table 3, rows 8 and 9). The increase is most evident in 40°–20°S, where the SST trend over 2000–15 increases by 0.19°C century^{-1}. In contrast, the SST trend over 2000–15 decreases by 0.25°C century^{-1} in 60°–90°N owing to slight decreased coverage of Argo floats in the northern North Atlantic. Argo serves to improve coverage and therefore constraints on the EOTs in both regions.

Argo will undoubtedly be a key component of the ocean-observing system going forward. Its inclusion in ERSST would help ensure long-term monitoring capabilities were the buoy or ship fleet to suffer any data reductions. Therefore, even though the effect of inclusion is relatively small we include Argo data in ERSSTv5.

3) HADISST2 SEA ICE CONCENTRATION

In ERSSTv5 and previous versions, the sea ice concentration is used to adjust the analyzed SSTs toward the freezing temperature of seawater (-1.8°C). When sea ice concentration is greater than 90% in a grid box, the SST is set to -1.8°C; when sea ice concentration is between 60% and 90%, the SST is linearly interpolated between -1.8°C and the reconstructed SST. The reconstructed SST is not changed when sea ice concentration is less than 60% (Smith et al. 2008). The reason for using criterion 0.6 is that the relationship between ice and SST becomes noisy when sea ice concentration is less than 60%.

In a final innovation, ERSSTv5 (Table 1) is analyzed by replacing HadISST (1870–2010; Rayner et al. 2003) with the latest HadISST2 sea ice concentration (1870–2015; Titchner and Rayner 2014). The reconstructed SSTA in ERSSTv5 in comparison with experiment Argo5m decreases systematically by 0.05°–0.10°C over 60°–90°N (Fig. 3a) and 40°–90°S (Fig. 3f), which can clearly be seen in the averaged difference (Fig. 4h) and RMSD (Fig. 5h) for 1860–1920. The decrease in SSTAs in the Southern Hemisphere is clearly associated with the increased sea ice concentration in HadISST2 over HadISST (Fig. 2b). Analyses show that the decrease in SSTAs in the Northern Hemisphere oceans is associated with the changes in the distribution of sea ice concentration, although the integrated ice-covered area increases slightly in HadISST2 (Fig. 2a). In particular, the area with high sea ice concentration (greater than 90% or equal to 100%) in HadISST2 increases in comparison with HadISST in the 1940s, and therefore the SSTA decreases as much as 0.15°C (Fig. 3a).

5. Intercomparisons

The globally averaged SSTA (relative to their respective climatologies over 1971–2000) in ERSSTv5 is first compared with that in ERSSTv4, HadISST, and COBE-SST2 (Fig. 12). The SSTAs are higher in HadISST (solid black line) and COBE-SST2 (dotted black line) than ERSSTv5 and ERSSTv4 because of higher ship SST bias correction in the 1880s–1940s and 1950s–1960s as indicated in Huang et al. (2015a). Therefore, the long-term trend (1900–2015) of globally averaged SSTA is higher in ERSSTv5 (0.70° ± 0.07°C century^{-1}; Table 2, row 10) and ERSSTv4 (0.69° ± 0.08°C century^{-1}; Table 2, row 2) than HadISST (0.53° ± 0.05°C century^{-1}). The short-term trend (2000–15) is also higher in
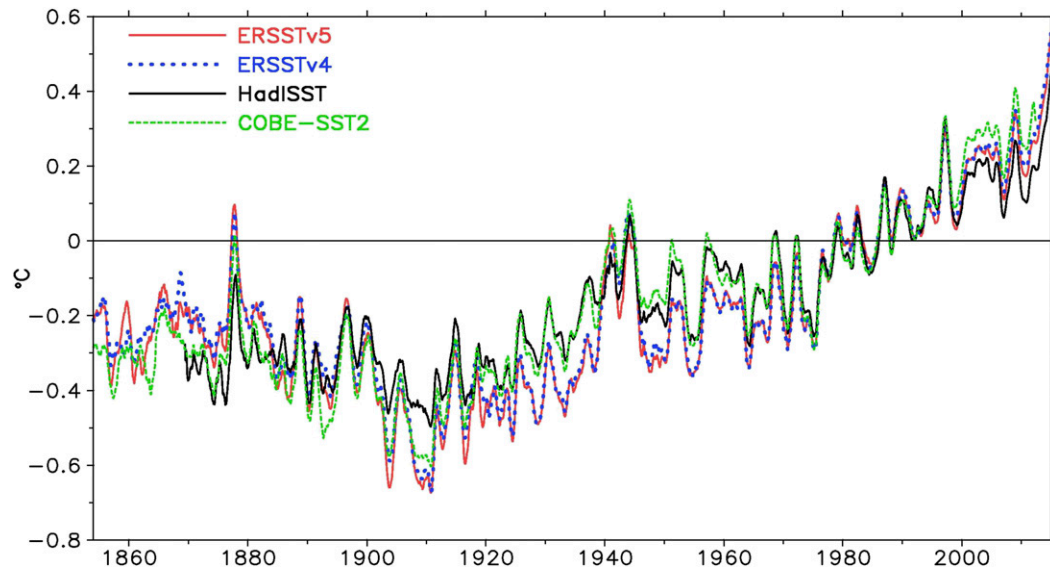
FIG. 12. Globally averaged SSTAs from ERSSTv5 (solid red line), ERSSTv4 (dotted blue line), HadISST (solid black line), and COBE-SST2 (dashed green line). Note that the SSTAs have been set to be relative to their own climatologies over 1971–2000.

ERSSTv5 ($1.25° \pm 0.77°C$ century^{-1}; Table 3, row 10) and ERSSTv4 ($1.34° \pm 0.75°C$ century^{-1}; Table 3, row 2) than HadISST ($0.80° \pm 0.70°C$ century^{-1}). The higher SST trends in ERSSTv5 and ERSSTv4 are associated with an upward trend in ship SST bias correction after 2000 (refer to Fig. 8) while the bias correction is near zero in HadISST after the 1970s (see Fig. 6 in Huang et al. 2015a).

The SSTs from ERSSTv5, ERSSTv4, HadISST, and COBE-SST2 are further compared with CCI SST product for September 1991–December 2010. The ESA CCI SST product is built from ATSR and AVHRR and is largely independent from in situ observations (Merchant et al. 2014). Therefore, CCI SST is used for validation of ERSST and comparison with other SST products.

Comparisons indicate that, on average over 1992–2010, the reconstructed SSTs are warmer than CCI in the high latitudes (40°–80°N and south of 40°S) in all products (Figs. 13a–d), and colder in the lower latitudes (40°S–40°N) in ERSSTv5 (Fig. 13a), HadISST (Fig. 13c), and COBE-SST2 (Fig. 13d) but not in ERSSTv4 (Fig. 13b). The globally averaged differences are 0.03°, 0.13°, 0.07°, and 0.05°C in ERSSTv5, ERSSTv4, HadISST, and COBE-SST2, respectively. Similar to the averaged difference, the RMSDs between the reconstructed SSTs and CCI are large in the high latitudes and small in the lower latitudes (Figs. 13e–h). The globally averaged RMSDs are 0.44°, 0.47°, 0.48°, and 0.39°C in ERSSTv5, ERSSTv4, HadISST, and COBE-SST2, respectively.

Those differences can be seen more clearly in the SSTAs (Fig. 14a) relative to the same ERSSTv4

climatology over 1971–2000. Overall, the differences between reconstructed SSTAs and CCI are approximately 0.1°–0.2°C higher in ERSSTv4 (Fig. 14b, dotted red line) and HadISST (dotted blue line) than in ERSSTv5 (solid red line) and COBE-SST2 (solid green line) over 1992–2000. The difference in ERSSTv4 is nearly 0.2°C over 1992–96 and 0.1°C over 1998–2010. In contrast, the differences become smaller in ERSSTv5, HadISST, and COBE-SST2 over 2000–06 and further decrease in HadISST and ERSSTv5 after 2006, although the difference in HadISST is as high as that in ERSSTv4 over 1992–98. Comparisons indicate that the SST trends are lower than CCI in all reconstructions. The SST trends over 1992–2010 are 1.25°, 1.29°, 0.69°, and 1.56°C century^{-1}, respectively, in ERSSTv5, ERSSTv4, HadISST, and COBE-SST2, while the SST trend in CCI is 1.61°C century^{-1}.

Overall, these comparisons indicate a potential improvement in ERSSTv5 over ERSSTv4, but the discrepancy between ERSSTv5 and CCI is clear before 1997. Further analyses (not shown) imply that the improvement in ERSSTv5 is mostly associated with the change in ship SST bias correction to using buoy SST as a reference (refer to Fig. 9) rather than NMAT over the most recent period.

6. Summary and discussion

The ERSST has been upgraded from ERSSTv4 to ERSSTv5 and in a range of aspects, and the impact of each change has been progressively assessed. The following aspects have been revised: the first-guess, base
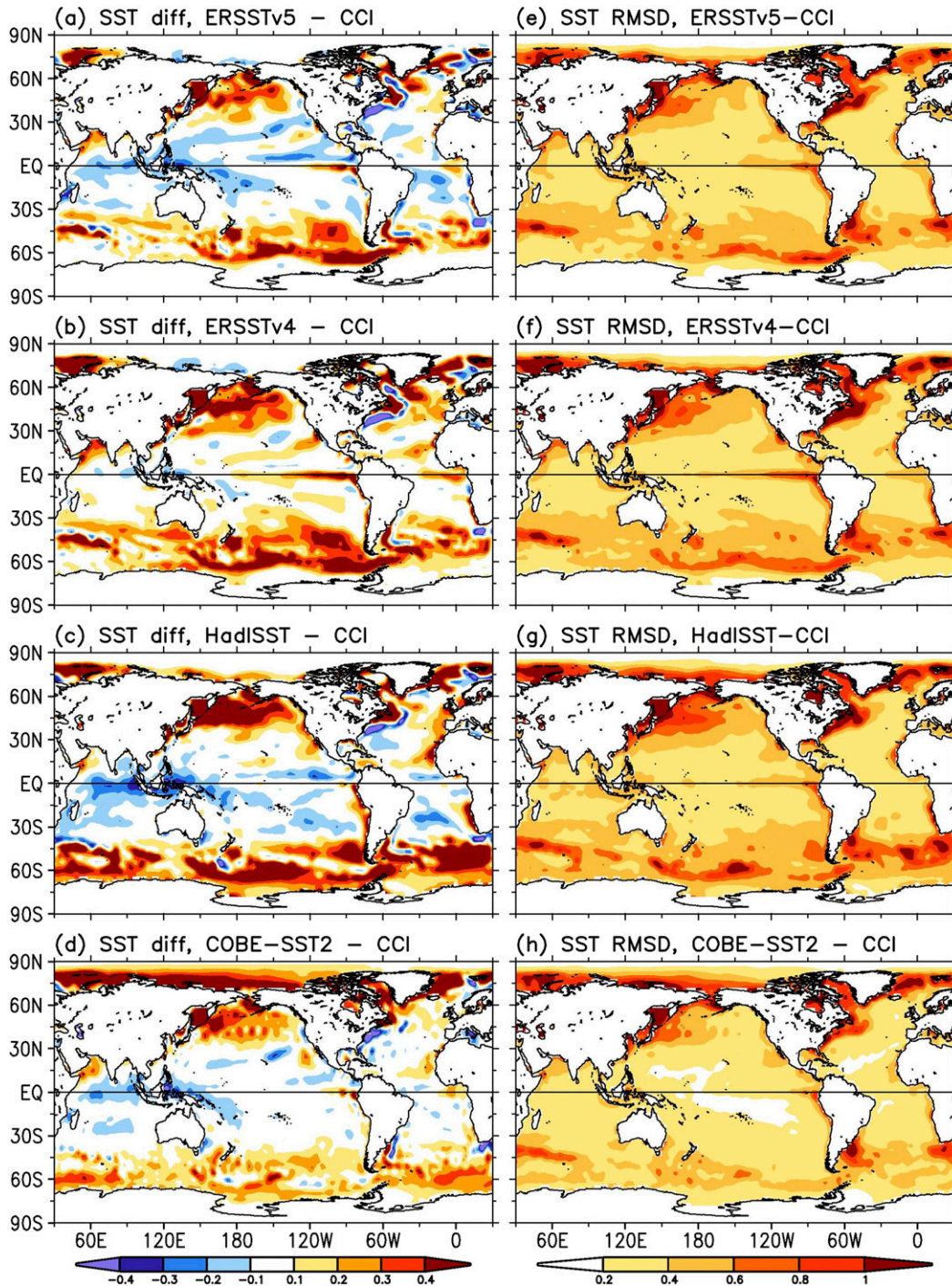
FIG. 13. Averaged (1992–2010) SST difference between (a) ERSSTv5 and CCI, (b) ERSSTv4 and CCI, (c) HadISST and CCI, and (d) COBE-SST2 and CCI. (e)–(h) As in (a)–(d), but for RMSD. The globally averaged values are 0.03°, 0.13°, 0.07°, and 0.05°C in (a)–(d) and 0.44°, 0.47°, 0.48°, and 0.39°C in (e)–(h), respectively.

function of EOT, ship SST bias correction, and updated and newly available source and ancillary datasets. The EOTs are updated by removing the damping in high latitudes, reducing spatial filters applied to the training

data, and adding 10 more modes in the Arctic. The ship SST bias is corrected using NMAT before 2010 and applying a separate linear fit before and after 1941 to better account for the transition in observation techniques in the
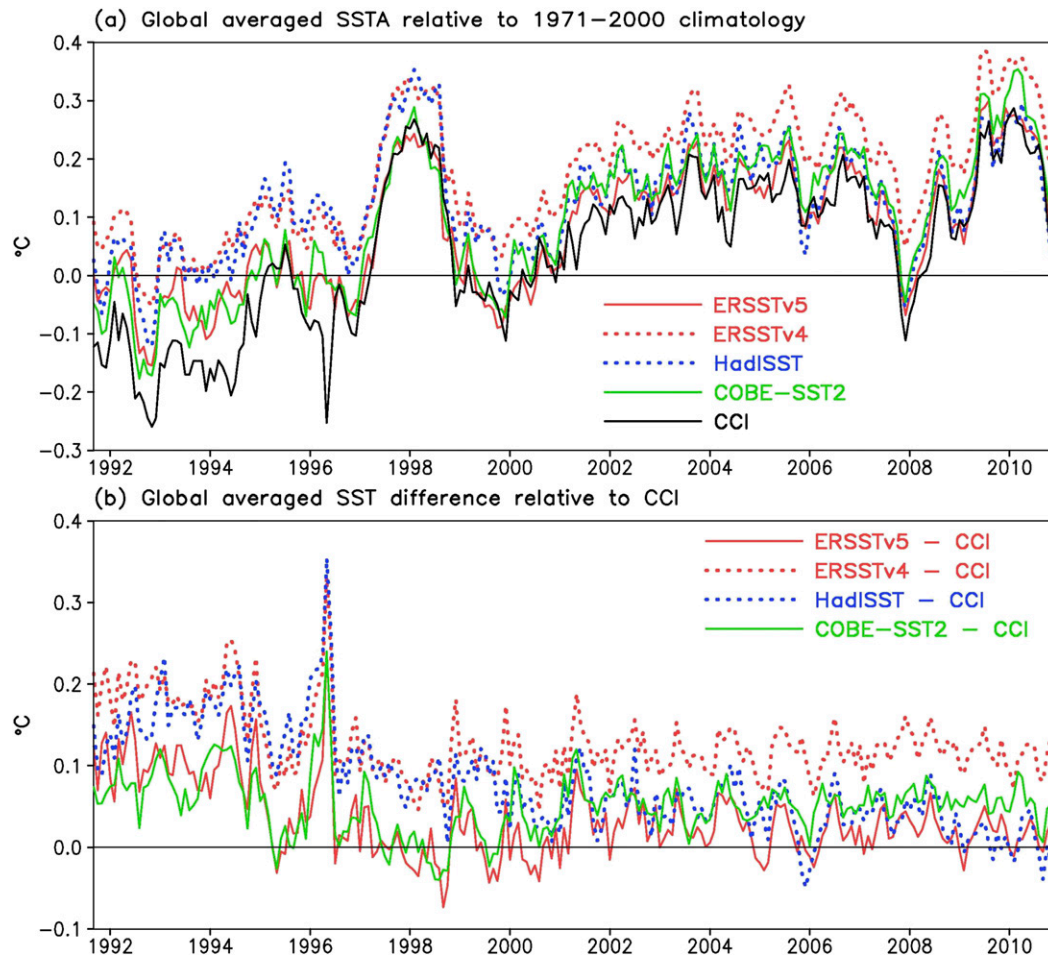
FIG. 14. (a) Globally averaged SSTA from ERSSTv5 (solid red line), ERSSTv4 (dotted red line), HadISST (dotted blue line), COBE-SST2 (solid green line), and CCI (solid black). The same 1971–2000 climatology of ERSSTv4 is used. (b) Globally averaged SST difference relative to CCI from ERSSTv5, ERSSTv4, HadISST, and COBE-SST2. The averaged (September 1991–December 2010) differences are 0.04°, 0.13°, 0.08°, and 0.05°C, respectively, in ERSSTv5, ERSSTv4, HadISST, and COBE-SST2. SST trends over 1992–2010 are 1.25°, 1.29°, 0.69°, 1.56°, and 1.61°C century^{-1}, respectively, in ERSSTv5, ERSSTv4, HadISST, COBE-SST2, and CCI.

1940s and is corrected using buoy SST after 2010. The updated SST observations from ICOADS R3.0 (1854–2015) and sea ice concentration from HadISST2 (1870–2015) are used to replace ICOADS R2.5 and HadISST. The SST observations from Argo drifters above 5 m (1999–2015) are merged into ERSSTv5 for the first time, which increases coverage and long-term viability of monitoring updates.

The impacts of some upgrades can be large scale in space and long-term in time. First, SSTs are systematically lowered by approximately 0.077°C over the global oceans when the biases of ship SSTs are corrected by reference to buoy observations in ERSSTv5. Validations from independent (at that step) Argo observations above 5 m show that the reconstructed SSTs have been improved when buoy SSTs are used as a reference in correcting ship SST biases.

Second, the SSTs decrease by 0.1°–0.2°C in the high latitudes oceans in ERSSTv5 because the newer HadISST2 is used, which has generally higher sea ice concentration. Since these changes occur primarily throughout the entire period of the reconstruction, the changes to the trends of the globally averaged SSTAs are small in both the long term (by 0.01°–0.02°C century^{-1}; Table 2, first column) and short term (by 0.01°–0.04°C century^{-1}; Table 3, first column).

In contrast, the impacts of some upgrades from EOTs and new datasets of Argo SST and ICOADS R3.0 can be local (i.e., small scale in space and short term in time). First, cross validations show that the EOT upgrades to remove the damping in the high latitudes reduce spatial filtering in training data and add more modes in the Arctic that have clearly improved the local SSTA distributions and the magnitude of El Niño and La Niña events. As they relate to

high-frequency components, the changes in local SSTAs due to these upgrades do not substantively impact the long-term trend of globally averaged SSTA. However, the short-term SST trend in 90°–40°S decreases when EOTs are trained with reduced spatial filtering, while the trend decreases slightly in the global oceans. Second, the impact of using ICOADS R3.0 is mostly in the high latitudes before the 1900s, and hence the long-term trend of globally averaged SSTA over 1900–2015 does not change much. However, the short-term SST trend in 90°–40°S decreases owing to better coverage in parts of the Southern Ocean that have exhibited muted or no warming. Finally, the inclusion of Argo SST increases the spatial coverage of observations by 5%–10%, but its impact on the SST reconstruction is limited, since the ship and buoy SSTs have already covered much of the global oceans in the most recent period. However, the impact of Argo observations is evident in 60°–90°N and 20°–40°S where observations taken by ships and buoys are less densely covered.

Comparisons show that the differences between several independently produced SST products remain. The differences have been attributed mostly to the ship SST bias correction (Huang et al. 2015a; Kent et al. 2017). Our results would confirm this. The differences in SST bias correction result in different long-term and short-term SST trends, particularly the short-term SST trend. Overall, the difference of long-term (1900–2015) trends of globally averaged SST is small ($0.01°\text{C century}^{-1}$) between ERSSTv5 ($0.70°\text{C century}^{-1}$) and ERSSTv4 ($0.69°\text{C century}^{-1}$). The difference of short-term (2000–15) trend is somewhat larger ($-0.09°\text{C century}^{-1}$) between ERSSTv5 ($1.25°\text{C century}^{-1}$) and ERSSTv4 ($1.34°\text{C century}^{-1}$). The short-term SST trends suggest a continued warming in the global oceans, which contributes to the warming of the global surface temperature as indicated in Karl et al. (2015). These trend changes between ERSSTv4 and ERSSTv5 fall well within quantified uncertainties in the ERSSTv4 product (Huang et al. 2016a).

Some restrictions of ERSSTv5, however, should be noted. First, the data prior to around the 1880s may not be very reliable owing to sparseness of observations in the Pacific and Indian Oceans in ICOADS R3.0 and the inability to provide sufficient valid EOTs for construction of a reliable global estimate. Second, the time evolution of ship SST bias correction remains similar to that in ERSSTv4, which will be a focus for future SST data development as suggested by Kent et al. (2017). Third, each SST data production requires an estimation of uncertainty because of the limitations in data availability and reconstruction methodology (Huang et al. 2016a). Therefore, a new estimation of ERSSTv5 uncertainty is under development according to our present understanding of uncertainty quantification approaches. In particular, the ERSSTv5 uncertainty in the high latitudes may be large due to large uncertainty in sea ice coverage as indicated by the difference of sea ice concentrations between HadISST and HadISST2.

The purpose to include Argo-derived SST in ERSSTv5 is to provide the best-possible estimate of SST. However, to encourage an independent validation against ERSSTv5 using Argo-derived SST observations, a version of ERSSTv5 with no Argo observations (ERSSTv5nargo) is constructed in parallel with ERSSTv5 after 2000 and available to researchers upon demand. This is consistent with the approach taken in the wider SST community and with the recommendations of the recent community paper (Kent et al. 2017).

In conclusion, ERSSTv5 represents an improvement upon the previous version ERSSTv4 in source datasets used and in key aspects of quality control, homogenization, and interpolation. The global SST warming trend in the past century (1900–2015) remains essentially unchanged, while the warming trend since 2000 decreases slightly but remains highly significant. The spatial and temporal variabilities of local SSTs are more realistic. The SSTAs of El Niño and La Niña events are more latitudinally bound in the equatorial Pacific, and therefore their magnitude is enhanced and closer to observations. Use of improved observations holdings and the Argo floats increases the resilience of the product to provide operational monitoring capabilities into the future.

REFERENCES

Argo, 2000: Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE, accessed 1 June 2016, doi:10.17882/42182.

Bojinski, S., M. Verstraete, T. C. Peterson, C. Richter, A. Simmons, and M. Zemp, 2014: The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Amer. Meteor. Soc.*, **95**, 1431–1443, doi:10.1175/BAMS-D-13-00047.1.

Cowtan, K., and Coauthors, 2015: Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures. *Geophys. Res. Lett.*, **42**, 6526–6534, doi:10.1002/2015GL064888.

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.]

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.

Grumbine, R. W., 2014: Automated sea ice concentration analysis history at NCEP 1996–2012. Marine Modeling and Analysis Branch Tech. Note 321, 39 pp.

Hausfather, Z., and Coauthors, 2017: Assessing recent warming using instrumentally homogeneous sea surface temperature records. *Sci. Adv.*, **3**, e1601207, doi:10.1126/sciadv.1601207.

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.

Huang, B., M. L'Heureux, J. Lawrimore, C. Liu, V. Banzon, Z.-Z. Hu, and 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.

——, and Coauthors, 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.

——, W. Wang, C. Liu, V. F. Banzon, J. Lawrimore, and 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.

——, and Coauthors, 2016a: Further exploring and quantifying uncertainties for Extended Reconstructed Sea Surface Temperature (ERSST) version 4 (v4). *J. Climate*, **29**, 3119–3142, doi:10.1175/JCLI-D-15-0430.1.

——, C. Liu, V. F. Banzon, H.-M. Zhang, T. R. Karl, J. H. Lawrimore, and R. S. Vose, 2016b: Assessing the impact of satellite-based observations in sea surface temperature trends. *Geophys. Res. Lett.*, **43**, 3431–3437, doi:10.1002/2016GL068757.

——, M. L'Heureux, Z.-Z. Hu, and H.-M. Zhang, 2016c: Ranking the strongest ENSO events while incorporating SST uncertainty. *Geophys. Res. Lett.*, **43**, 9165–9172, doi:10.1002/2016GL070888.

IPCC, 2013: *Climate Change 2013: The Physical Science Basis*. Cambridge University Press, 1535 pp., doi:10.1017/CBO9781107415324.

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.

Karl, T. R., and Coauthors, 2015: Possible artifacts of data biases in the recent global surface warming hiatus. *Science*, **348**, 1469–1472, doi:10.1126/science.aaa5632.

Kennedy, J. J., 2014: A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Rev. Geophys.*, **52**, 1–32, doi:10.1002/2013RG000434.

——, 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.

——, ——, ——, ——, and ——, 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.

Kent, E., 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.

——, and Coauthors, 2017: A call for new approaches to quantifying biases in observations of sea-surface temperature. *Bull. Amer. Meteor. Soc.*, doi:10.1175/BAMS-D-15-00251.1, in press.

Menne, M. J., C. N. Williams Jr., and R. S. Vose, 2009: The U.S. Historical Climatology Network monthly temperature data, version 2. *Bull. Amer. Meteor. Soc.*, **90**, 993–1007, doi:10.1175/2008BAMS2613.1.

Merchant, C. J., and Coauthors, 2014: Sea surface temperature datasets for climate applications from phase 1 of the European Space Agency Climate Change Initiative (SST CCI). *Geosci. Data J.*, **1**, 179–191, doi:10.1002/gdj3.20.

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.

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

Riser, S. C., and Coauthors, 2016: Fifteen years of ocean observations with the global Argo array. *Nat. Climate Change*, **6**, 145–153, doi:10.1038/nclimate2872.

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.

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

Smith, T., and 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.

——, 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.

Titchner, H. A., and 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.

Trenberth, K. E., and D. P. Stepaniak, 2001: Indices of El Niño evolution. *J. Climate*, **14**, 1697–1701, doi:10.1175/1520-0442(2001)014<1697:LIOENO>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.

Woodruff, S. D., 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.

Xue, Y., Z.-Z. Hu, A. Kumar, V. Banzon, B. Huang, and J. Kennedy, 2016: Global oceans [in "State of the Climate in 2015"]. *Bull. Amer. Meteor. Soc.*, **97** (8), S63–S92.

Yasunaka, S., and K. Hanawa, 2011: Intercomparison of historical sea surface temperature datasets. *Int. J. Climatol.*, **31**, 1056–1073, doi:10.1002/joc.2104.