Intercomparison of the Extended Reconstructed Sea Surface Temperature v4 and v3b Datasets

WANG Jinping[1], and CHEN Xianyao[2], *

1) *Department of Oceanography, College of Oceanic and Atmospheric Sciences, Ocean University of China,*
  Qingdao 266100, *China*
2) *Physical Oceanography Laboratory/CIMST, Ocean University of China and Qingdao National Laboratory*
  for Marine Science and Technology, Qingdao 266100, *China*

Abstract   Version 4 (v4) of the Extended Reconstructed Sea Surface Temperature (ERSST) dataset is compared with its precedent, the widely used version 3b (v3b). The essential upgrades applied to v4 lead to remarkable differences in the characteristics of the sea surface temperature (SST) anomaly (SSTa) in both the temporal and spatial domains. First, the largest discrepancy of the global mean SSTa values around the 1940s is due to ship-observation corrections made to reconcile observations from buckets and engine intake thermometers. Second, differences in global and regional mean SSTa values between v4 and v3b exhibit a downward trend (around −0.032℃ per decade) before the 1940s, an upward trend (around 0.014℃ per decade) during the period of 1950–2015, interdecadal oscillation with one peak around the 1980s, and two troughs during the 1960s and 2000s, respectively. This does not derive from treatments of the polar or the other data-void regions, since the difference of the SSTa does not share the common features. Third, the spatial pattern of the ENSO-related variability of v4 exhibits a wider but weaker cold tongue in the tropical region of the Pacific Ocean compared with that of v3b, which could be attributed to differences in gap-filling assumptions since the latter features satellite observations whereas the former features *in situ* ones. This intercomparison confirms that the structural uncertainty arising from underlying assumptions on the treatment of diverse SST observations even in the same SST product family is the main source of significant SST differences in the temporal domain. Why this uncertainty introduces artificial decadal oscillations remains unknown.

Key words   ERSST datasets; sea surface temperature; global warming; Arctic; data intercomparison

1 Introduction

As an essential indicator of global climate change and ocean variability, sea surface temperature (SST) has been observed by several means since the 1850s; the findings obtained, however, are often extremely unevenly distributed in either the spatial or temporal domains. To reconstruct real but unknown globally complete SST fields, SST datasets are produced using different statistical techniques, and these sets are frequently upgraded by incorporating more observations and improving assumptions and algorithms (Kaplan *et al.*, 1998; Reynolds *et al.*, 2002; Rayner *et al.*, 2003, 2006; Smith and Reynolds, 2003, 2004; Smith *et al.*, 2008; Kennedy *et al.*, 2011a, 2011b; Hirahara *et al.*, 2014; Huang *et al.*, 2015a; Liu *et al.*, 2015). While these developments enhance our knowledge of how the SST of world oceans varies, careful intercomparison between datasets and estimation of the impacts of each upgrade are also necessary to improve our previous understanding of the global SST variability.

The Extended Reconstructed SST (ERSST) dataset was recently upgraded from version 3b (v3b) (Smith *et al.*, 2008) to version 4 (v4) (Huang *et al.*, 2015a; Liu *et al.*, 2015). The main improvements of this upgrade are based on: 1) the latest *in situ* data set from the International Comprehensive Ocean-Atmosphere Data Set (ICOADS) release 2.5 (R2.5; Woodruff *et al.*, 2011), which provides better duplicate removal and gross quality control, as well as better coverage in previously under-sampled areas; 2) systematical SST bias adjustments since the 1850s in v4 that are not available in v3b; and 3) estimations of parametric uncertainties in the SST reconstructions that were not provided in ERSST v3b (Huang *et al.*, 2015a). Given these improvements, ERSST v4 is expected to exhibit more realistic SST variabilities than v3b across a range of space and time scales (Huang *et al.*, 2015a; Liu *et al.*, 2015). ERSST v4 does not support the notion of a 'slowdown' in the increase in global mean surface temperature (Karl *et al.*, 2015), which intensifies the debate on whether the global warming 'slowdown' is justified by the available data (Lewandowsky *et al.*, 2015; Fyfe *et al.*, 2016). Obviously, while most of our knowledge on the global SST variability is based on ERSST v3b, examining the

---

* Corresponding author. E-mail: chenxy@ouc.edu.cn

differences introduced by the new version is essential.

Intercomparisons between the two ERSST versions are presented in Huang *et al.* (2015a) and Liu *et al.* (2015); these comparisons focus on differences in mean SSTa values in the global scale and different latitudinal zones associated with different parameters. In this short note, we further compared ERSST v3b and ERSST v4 in both the global and regional scales and on areas with and without real observations. Our comparison intends to identify the integrated impacts of the upgrades applied in ERSST v4 on our general understanding of the SST variability established based on ERSST v3b. Unlike pioneering works on cross-comparisons between the SST products from different research branches, the present intercomparison emphasizes SSTs with the same genealogy (ERSST series) but upgraded assumptions and algorithms. Therefore, a comparison with the other SST products is not included in this study.

The paper is structured as follows. In the next section, we describe the data sources and analysis techniques. In Section 3, we present differences between ERSST v3b and ERSST v4 in three aspects: global distributions, global and regional means, and some modes of climate variability, such as ENSO-related variability, Atlantic multidecadal oscillation (AMO), and Indian Ocean dipole (IOD). A discussion and conclusions are given in the final section.

2 Datasets

The ERSST dataset studied in this paper is obtained from the National Centers for Environmental Information (NCEI) of National Oceanic and Atmospheric Administration (NOAA) (http://www.ncdc.noaa.gov/data-access/marineocean-data/extended-reconstructed-sea-surface-temperature-ersst). The monthly-mean ERSST dataset starts from January 1854 and contains anomalies calculated with reference to the 1971–2000 monthly climatology. A starting year of 1900 is selected for better data coverage.

Because infrared-based satellite observations provide 'skin SST' values rather than 'bulk SST' ones and can only be obtained in clear-sky conditions, satellite SST products have not been used to reconstruct the global SST dataset since the release of ERSST v3b. However, in both ERSST v3b and ERSST v4, satellite-based monthly data from OISST v2 of the period 1982–2011 are used to train localized empirical orthogonal teleconnections (EOTs) (van den Dool *et al.*, 2000; Smith *et al.*, 2008). Therefore, OISST v2 (http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html) is used in this analysis.

We also make use of the uninterpolated *in situ* SST dataset from ICOADS R2.5 (http://www.esrl.noaa.gov/psd/data/gridded/data.coads.ltm.html) to locate grid boxes with real *in situ* SST observations.

3 Results

3.1 Differences in Global Structure

Fig.1 shows the correlation coefficients of the two ERSST versions at each grid box. A significant consistency ($r>0.8$ over the period 1900–2015, Fig.1a) between the two datasets is found in the low and middle latitudes (35°S–65°N). In an earlier period (1900–1950), the regions of consistency are generally confined to the North Atlantic and Northeast Pacific Oceans, where a larger number of observations are available compared with that for other oceans (Fig.1b). After 1950, the two versions exhibit high consistency owing to comprehensive observations (Fig.1c). From the 1980s onwards, the correlation coefficients exceed 0.9 at most of the low- and middle-latitude areas (Fig.1d). The correlation coefficients shown in Fig.1 are all significant at the 95% confidence level and have been tested based on Bretherton *et al.* (1999) and Ding *et al.* (2012) by taking into account the auto-correlation inherent in the SST records, which reduces the number of degrees of freedom.
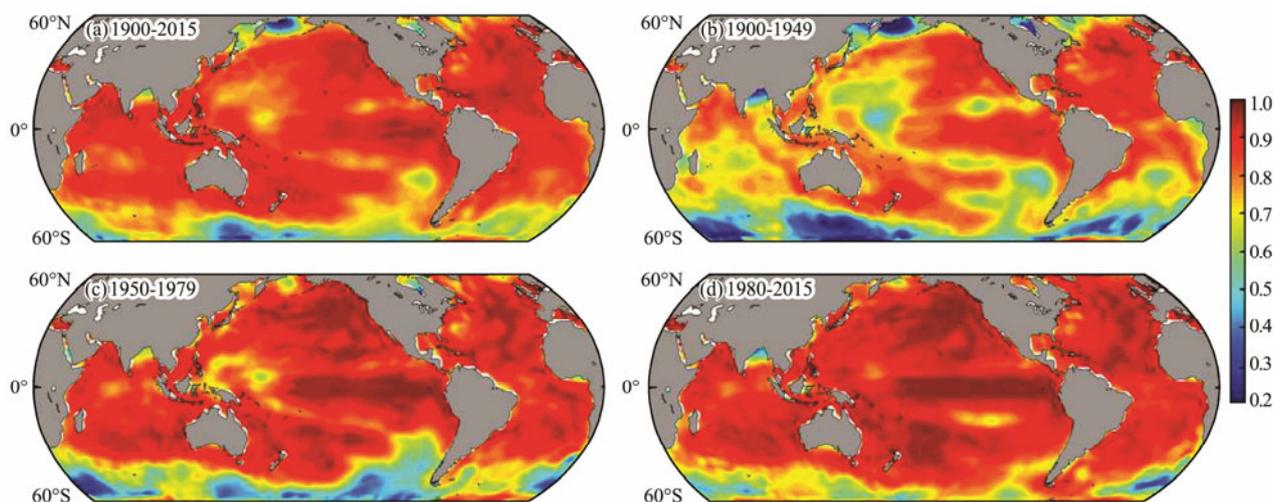
Fig.1 Correlation coefficients of v3b and v4 at each grid box for different periods: (a) 1900–2015, (b) 1900–1949, (c) 1950–1979, and (d) 1980–2015. Only significant correlations above the 95% confidence level are shown in the graph. White regions represent unavailable values.

The correlation map between the two datasets presents the strong impacts of insufficient continuous observations on SST reconstruction. Since the Southern and Arctic Oceans are covered by seasonal or perennial sea ice, which could cause obvious seasonal signals in the observed data, we compare correlation coefficients of these areas during the whole year, summer (for boreal, from May to October), and winter (for boreal, from November to April of the next year) time. In the Southern Ocean, as shown in Fig.2, the correlation coefficients in the South Atlantic and South Indian Ocean sectors generally fall below 0.5 before the 1950s. Afterward, along with in-

creasing observations around the Antarctic Ocean in austral summer, the discrepancy between the two datasets is reduced but remains remarkable. Before the Argo floats were frequently used in the Southern Ocean, at around 2000, the agreement between the two datasets shows strong seasonal characteristics, *i.e.*, higher correlations in summer and lower correlations in winter, due to rare *in situ* observations in the areas affected by the Antarctic Circumpolar Current in winter. After 2000, the widely used Argo floats diminish this seasonal difference by providing better constraints of the observations in winter months.
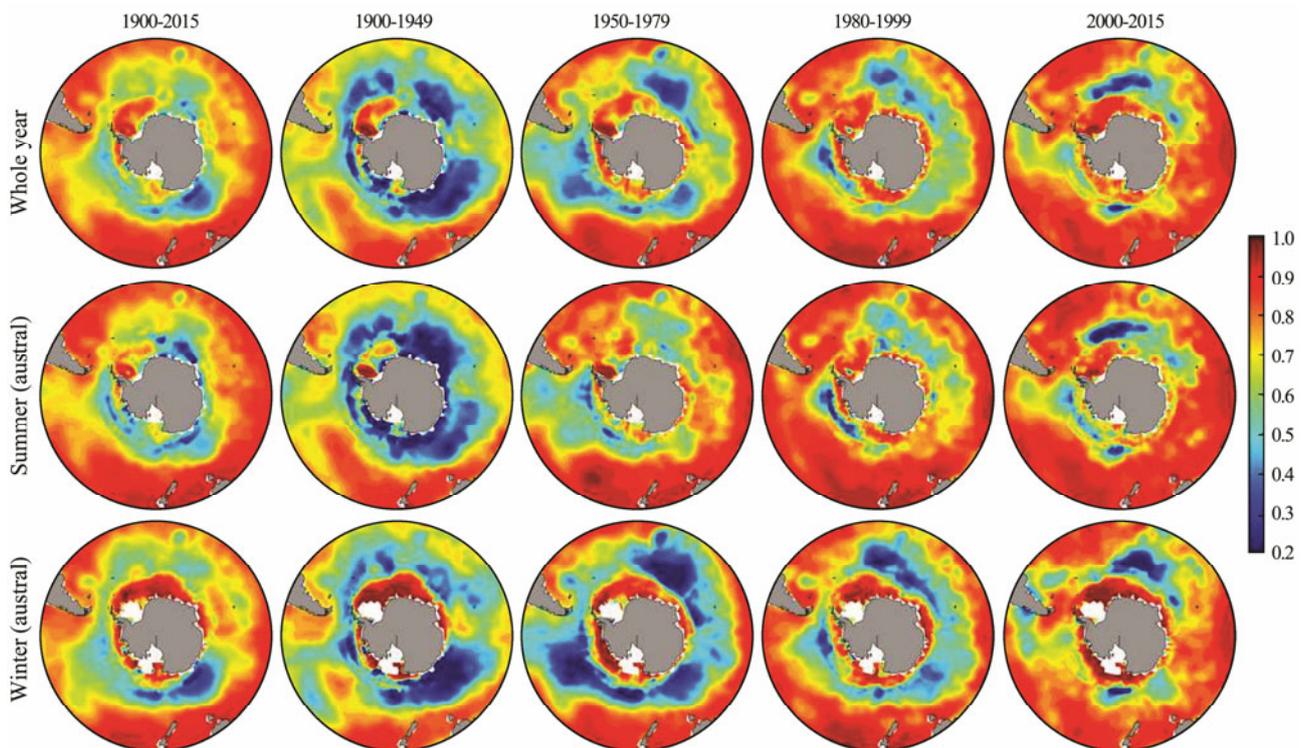
Fig.2 The whole-year monthly mean (top panel), summer (from November to April of the next year) mean (middle panel), and winter (from May to October) mean (bottom panel) SSTa correlation coefficients of the Southern Ocean between v3b and v4 at each grid box for different periods: 1900–2015, 1900–1949, 1950–1979, 1980–1999, and 2000–2015. Only significant correlations above the 95% confidence level are shown in the graph. White regions represent unavailable values.

Unlike those of the Southern Ocean, the correlation coefficients of ERSST v3b and v4 of the Arctic central basin are usually larger than 0.8 until the 1980s (Fig.3); this consistency between the two datasets, however, does not mean that the data are reliable. Before the 1980s, few *in situ* observations of the Arctic central basin were available, and the SSTa of this area is directly assigned to zero in ERSST v3b but the anomaly of the nearest grid in ERSST v4. Since the 1980s, more observations of the Arctic Ocean have been obtained owing to rapid sea ice melting. Unfortunately, these observations are inadequate to constrain the objective interpolation used to fill grid boxes without observations. The same reasoning could explain the decreasing correlation coefficients found in the central basin, as well as in the Greenland–Iceland–Norwegian seas.

3.2 Global and Regional Mean SSTa Values

Fig.4a compares the global and regional mean SSTa values derived from the two datasets, and Fig.4b shows their difference. The most significant discrepancy (over 0.15℃) among global mean SSTa values occurs in the middle of the 1940s, owing to the switching of ship observations from buckets to engine intake thermometers around World War II (Folland and Parker, 1995; Thompson *et al.*, 2008; Kennedy *et al.*, 2011a, 2011b). Before the 1940s, the difference in global mean SSTa values between the two datasets shows a downward trend (around −0.032℃ per decade from 1900 to 1940). After the 1940s, this difference is within ±0.1℃ but two significant features are also found: an upward trend (around 0.014℃ per decade from 1950 to 2015) and a decadal oscillation

with one peak at around the 1980s and one trough each at the end of the 1960s and 2000s. The same pattern from the 1950s is also apparent in regional oceans except the Arctic Ocean, as shown in Fig.4.
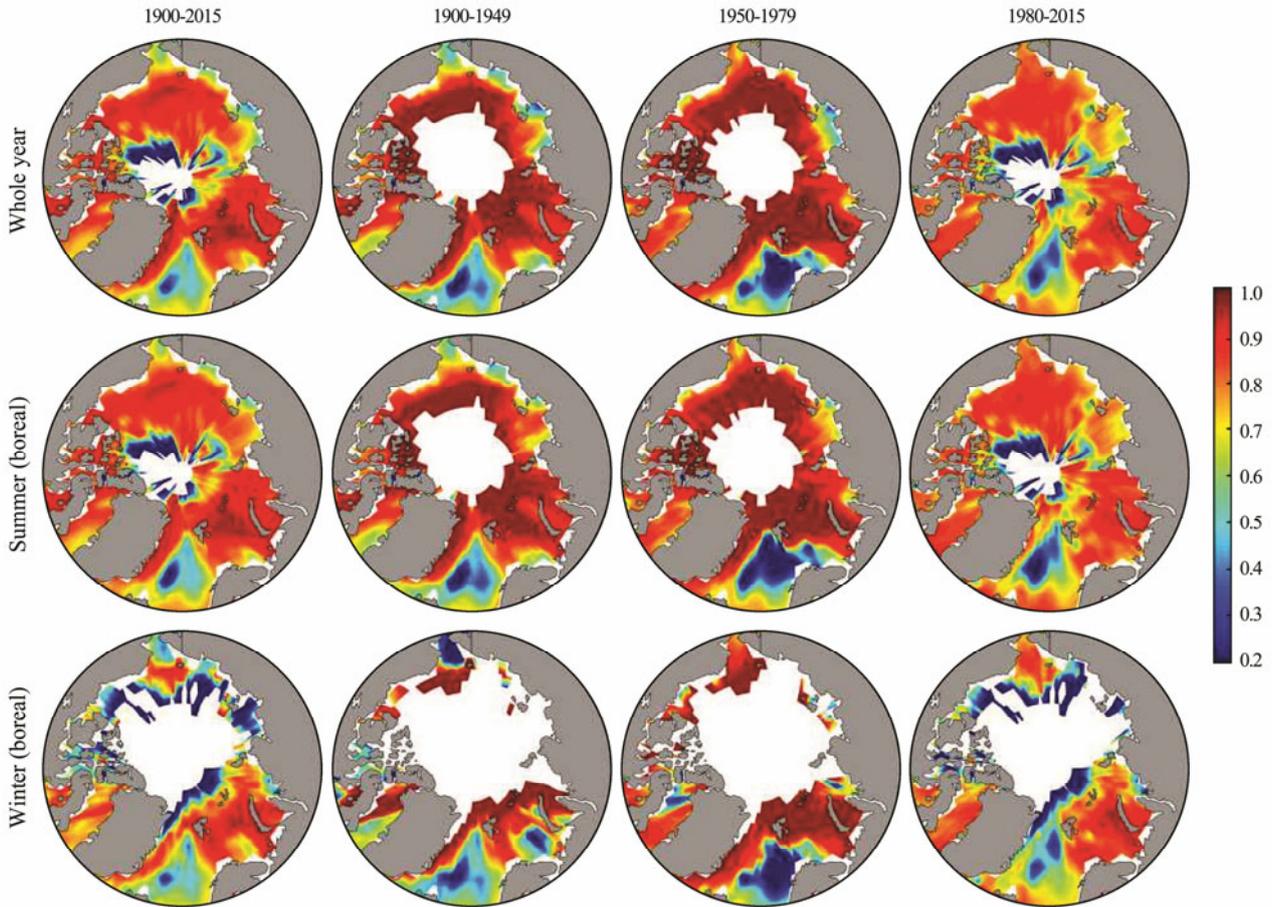
Fig.3 The whole-year monthly mean (top panel), summer (from May to October) mean(middle panel), and winter (from November to April of the next year) mean (bottom panel) SSTa correlation coefficients of the Arctic Ocean between v3b and v4 at each grid box for different periods: 1900–2015, 1900–1949, 1950–1979, and 1980–2015. Only significant correlations above the 95% confidence level are shown in the graph. White regions represent unavailable values.
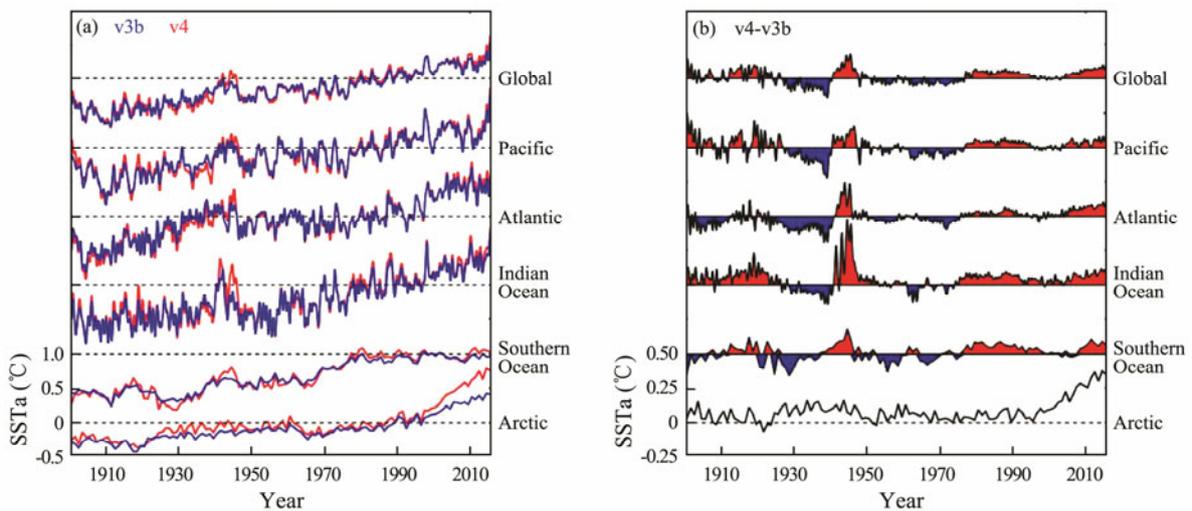
Fig.4 (a) Area averaged monthly SSTa time series over the global, Pacific, Atlantic, Indian, Southern, and Arctic Oceans for v3b (blue) and v4 (red). For the Southern and Arctic Oceans, only the summer SSTa data are chosen to calculate regional mean time series. (b) Differences in time series between v4 and v3b. Eleven-month running means are used to smooth the monthly SSTa time series except those of the Southern and Arctic Oceans. Note that the interval between tick-marks on the vertical axis differs between (a) and (b).

The difference in global and regional mean SSTa values exhibit a downward trend before the 1940s and then an upward trend afterward. This finding suggests that, when comparing v4 with v3b, the latter estimates higher levels of global warming before the 1940s and estimates lower levels after the 1940s, although the overall rate of warming over the full period of 1880–2014 is essentially the same (0.055℃ per decade) between the two datasets (Karl *et al.*, 2015). This difference also explains the debated estimation of the linear global warming trend over the period 2000–2014 between ERSST v3b (0.035℃ per decade) and ERSST v4 (0.099℃ per decade).

Several improvements in ERSST v4 may be attributed to a systematical difference between the two datasets.

Karl *et al.* (2015) proposed that, of the 11 improvements in ERSST v4, the ship-bias correction, which is necessary to achieve consistency between the observations from buckets and engine intake thermometers and was not implemented in ERSST v3b since 1941, exerts the largest impacts on the trends observed for the period 2000–2014, accounting for 0.030℃ of the 0.064℃ trend difference obtained. This proposal implies that the upward trend of the difference in global mean SSTa values after 1940s may also be partly attributable to ship-bias corrections. However, since bucket observations are generally lower than the real SSTs, this correction cannot explain the negative difference observed during the period 1950–1970.

Variations among gap-filling techniques are another main source of uncertainty in SST reconstructions. In ERSST v3b, the low-frequency component of SSTa at the grid boxes without *in situ* observations are filled with zeros; in ERSST v4, filling with the average of neighboring valid proximal SSTa values is performed. The zero-filling technique tends to produce artificially warm (cool) SSTa values in earlier (later) periods in the context of global climate warming. To identify the impact of this difference on the upward trend of the difference in global mean SSTa values shown in Fig.4b, we compare sub-global and regional mean SSTa values in the region with real *in situ* SST observations located by the uninterpolated ICOADS SST dataset; the mean SSTa values of the two datasets in the region without *in situ* observations are also compared. Results demonstrate that despite the larger magnitude of the difference in the region without SST observations compared with that in the region with real observations, the upward trend and oscillation features found are only visible in the latter (Fig.5a). We compare the case of each individual basin and find that the gap-filling technique selected exerts strong impacts on the SST reconstruction in the Atlantic and Arctic Oceans in the recent decade, as well as the Indian Ocean in the period 1970–1990. The main features of an upward trend and oscillations in the Pacific, Atlantic, and Indian Oceans are still only found in the region with real observations. This observation suggests that the gap-filling technique applied presents little contribution to the difference in global mean SSTa values between the two datasets.
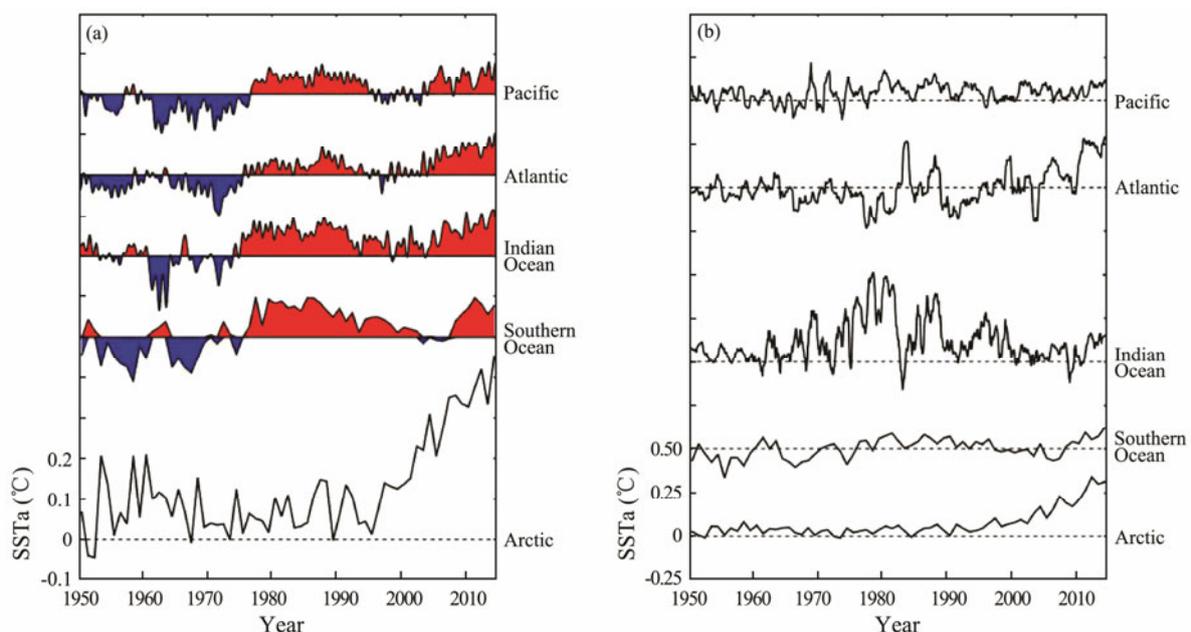
Fig.5 (a) Difference in time series between v4 and v3b over the global, Pacific, Atlantic, Indian, Southern, and Arctic Oceans with real *in situ* SST observations as located by the uninterpolated ICOADS SST dataset. For the Southern and Arctic Oceans, only summer variations are considered. (b) Difference in time series between v4 and v3b over the same oceans in the region without *in situ* observations. Eleven-month running means are used to smooth the monthly SSTa time series except those of the Southern and Arctic Oceans. Note that the interval between tick-marks on the vertical axis differs between (a) and (b).

Another correction is introduced by the difference in ship- and buoy-based SST observations. Before the mid-1990s, the main observations were carried out by ships. Since then, the usage of buoys has dramatically increased.

Because the ship data are systematically warmer than the buoy data (Kent *et al.*, 2010; Reynolds *et al.*, 2002, 2010), a 0.128℃ bias correction is added to the buoy SSTs in ERSST v4 but not in ERRST v3b. This systematic correction partly contributes to the difference in global mean SSTa values observed between datasets after the 1990s.

Except for these three improvements, other upgrades between ERSST v4 and ERSST v3b refer to the general methodologies applied to improve the datasets themselves, data quality control, and empirical orthogonal functions, which may change the temporal variability during reconstruction. Therefore, why a lower SST is observed during the period of 1950–1970 in the upgraded ERSST remains unknown.

3.3 Impacts of Fastest Warming Arctic on the Difference

Incomplete sampling of the fastest warming Arctic Ocean may lead to underestimation of the global trend in instrumental temperature records (Simmons *et al.*, 2010; Folland *et al.*, 2013). ERSST v4, for example, applies more observations of the Arctic Ocean for SST reconstruction than ERSST v3b. To address this issue, we compare mean SSTa values between the global ocean and the global ocean excluding the Arctic. As shown in Fig.6, both cases exhibit similar characteristics, which means the coverage bias in the Arctic Ocean is not adequately strong to change the patterns shown in the Pacific, Atlantic, and Indian Oceans.
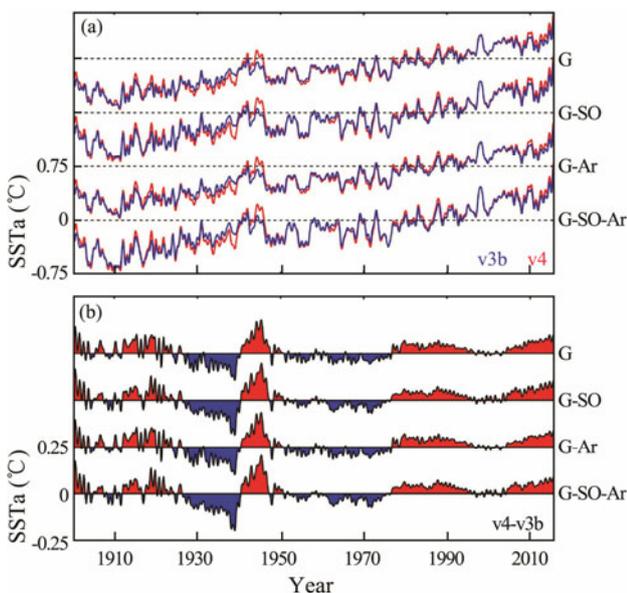
Fig.6 (a) Area averaged monthly SSTa time series over the global ocean (G), without the Southern Ocean (north of 35°S, G-SO), without the Arctic Ocean (south of 65°N, G-Ar), and without both the Southern and Arctic Oceans (middle and low latitude, 35°S–65°N, G-SO-Ar) for v3b (blue) and v4 (red). (b) Differences in time series between v4 and v3b over G, G-SO, G-Ar and G-SO-Ar. Eleven-month running means are used to smooth the time series. Note that the interval between tick-marks on the vertical axis differs between (a) and (b).

For generalizability, we compare mean SSTa values between the global ocean and the global ocean excluding the Southern Ocean, another area without sufficient *in situ* observations. As shown in Fig.6, same conclusions could be reached as in Arctic Ocean.

3.4 ENSO-related Variability, AMO and IOD

3.4.1 ENSO-related variability

In this section, we compare differences between v3b and v4 in terms of three main modes of climate variability: ENSO-like variability (Zhang *et al.*, 1997; Chen and Wallace, 2015), AMO (Schlesinger and Ramankutty, 1994; Trenberth and Shea, 2006) and IOD (Saji *et al.*, 1999).

The ENSO-related variability is defined as the leading principal component (PC1) of the empirical orthogonal functions of the pan-Pacific SST departure from the global mean SSTa (denoted as SST*, which equals SST − [SST], where [∗] denotes the global average of the anomaly field). As shown in Fig.7a, the time series of ENSO-related variability derived from the two datasets are nearly identical, but the associated spatial structures derived by projecting the global SST* upon the ENSO-related index show different characteristics in the tropical region of the Pacific Ocean. The equatorial Pacific cold tongue region derived from ERSST v3b is narrower than that derived from ERSST v4 and presents a stronger amplitude along the equator; the tip of the cold tongue derived from the former is also relatively sharper and extends further westward than that from the latter. We can see these features clearly in Fig.7c, which demonstrates an alternating distribution of positive and negative phases in the domain 10°S–5°S, 5°S–5°N, and 5°N–10°N.

We also compare differences in Pacific decadal oscillation (PDO) and interdecadal Pacific oscillation (IPO) between the two datasets (not shown). The IPO and PDO indices derived from both datasets are quite similar. The spatial patterns of the two datasets are different, but their patterns of difference are similar to those shown in Fig.7c.

To identify the potential reasons behind the difference observed, we project the monthly OISST v2 SST anomaly upon the ENSO-related index for the period 1982–2014. While satellite SST observations are not directly included in the reconstruction of both ERSST v3b and ERSST v4, the EOTs used for reconstruction are trained by OISST (van den Dool *et al.*, 2000; Smith *et al.*, 2008). Therefore, the spatial structure of the global SST from satellite observations is fundamental in both datasets. As shown in Fig.8, the tip of the cold tongue derived from the OISST SST field is similar to that derived from ERSST v3b. The difference noted may be introduced when only 130 EOTs are retained for ERSST v4, resulting in over-smoothing of higher frequency patterns (Huang and Karl, personal communications). By contrast, differences in the ENSO-related spatial structure are mainly associated with the interannual variability of the index.

The difference noted may also be attributed to observations at the tip of the cold tongue being fewer than those in other tropical areas. To confirm this reasoning, we pro-

ject the monthly ICOADS SST upon the ENSO-related index for the period 1982–2014. As the ICOADS SST is an uninterpolated dataset, we only select grid boxes with discontinuous observations of no more than 5 years. As Fig.8b shows, few *in situ* observations of the area (172˚–176˚E, 0˚–4˚N) are available. By contrast, in other areas, the spatial pattern obtained is similar to that derived from OISST dataset.
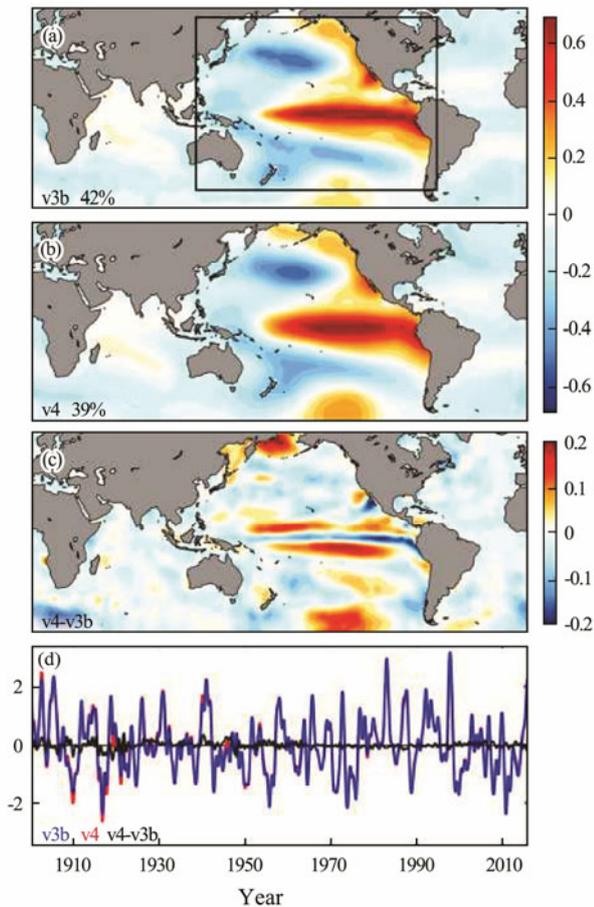
Fig.7 (a, b) Regression coefficients of the global SST* fields based on the PC1 of Pacific SST* fields in v3b and v4, respectively. Regression maps are scaled in units of ℃ per unit amplitude of the corresponding standardized PC time series. Percentages printed in the lower left corners refer to the variance explained by the Pacific SST* field. (c) Difference in regression coefficient patterns between v4 and v3b. (d) Corresponding standardized PCs for v3b (blue) and v4 (red) and the difference in time series between v4 and v3b (black). Eleven-month running means are used to smooth the time series. Rectangles indicate the domains of the regional EOF analyses in the Pacific.

3.4.2 AMO and DMI

We refer to the AMO index in Trenberth's analysis (2006); this index is defined as annual SST anomalies averaged over the North Atlantic (0–60˚N and 0–80˚W). Accordingly, the global mean SST is subtracted to derive a revised AMO index. The AMO index we used here is standardized to the unit variance as a reference time series
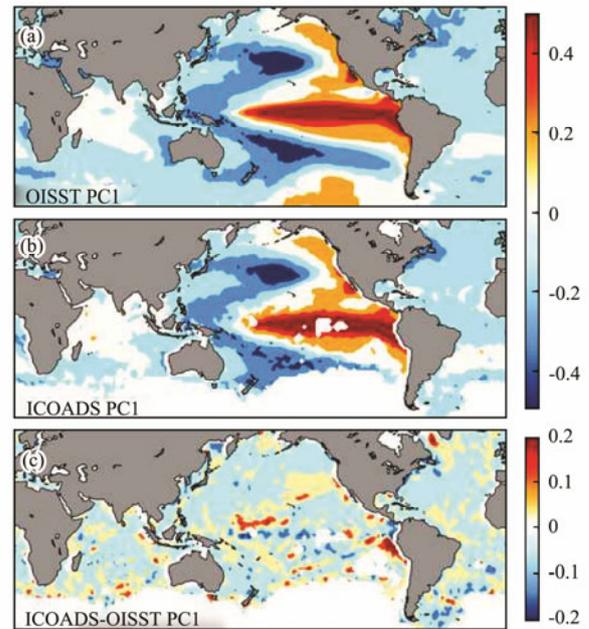
Fig.8 (a, b) Regression coefficients of the global SST* fields based on the PC1 (since the PC1 of v3b is virtually identical to that of v4, we select the PC1 of v4 for this analysis) of the Pacific SST* fields in OISST and ICOADS, respectively. Regression maps are scaled in units of ℃ per unit amplitude of the corresponding standardized PC time series. (c) Difference in regression coefficient patterns between v4 and v3b.

for interrelating AMO variability over the whole global scale. The regression patterns for global SSTa values based on the standardized AMO index are depicted in the top two panels of Fig.9. One of the largest amplitudes observed is located at the northern region of the Atlantic Ocean. The high amplitude in this region is related to the fact that the SSTa varies strongly in this region in a manner seemingly associated with the AMO (Enfield *et al.*, 2001). Another larger amplitude occurs in the eastern tropical Pacific Ocean, and the regression pattern herein exhibits structures highly similar to the ENSO-related pattern based on the PC1 shown in Fig.7, although the amplitude of the AMO regression is much shallower than that of the ENSO-related index. The AMO index of the two datasets shows that their differences are slightly more prominent before the 1940s, which is likely related to a considerable lack of observations (Fig.9d). From the 1940s onwards, the two datasets are fairly consistent with each other except for slight differences in amplitude.

The IOD mode (Saji *et al.*, 1999) is a coupled ocean-and-atmospheric phenomenon in the equatorial Indian Ocean. The intensity of the IOD is represented by anomalous SST gradients between the western equatorial Indian Ocean (50˚–70˚E and 10˚S–10˚N) and the southeastern equatorial Indian Ocean (90˚–110˚E and 10˚S–0˚). This gradient is called the dipole mode index (DMI). We derive the DMI in the two datasets and project the global SSTa values upon their respective normalized time series. The regression patterns and their spatial difference (v4 minus v3b) are closely associated with the ENSO-related

variability but show a much weaker amplitude. The DMIs of the two versions are similar except for a larger dis-crepancy in the 1940s (Fig.9h), which is also attributed to ship-bias corrections.
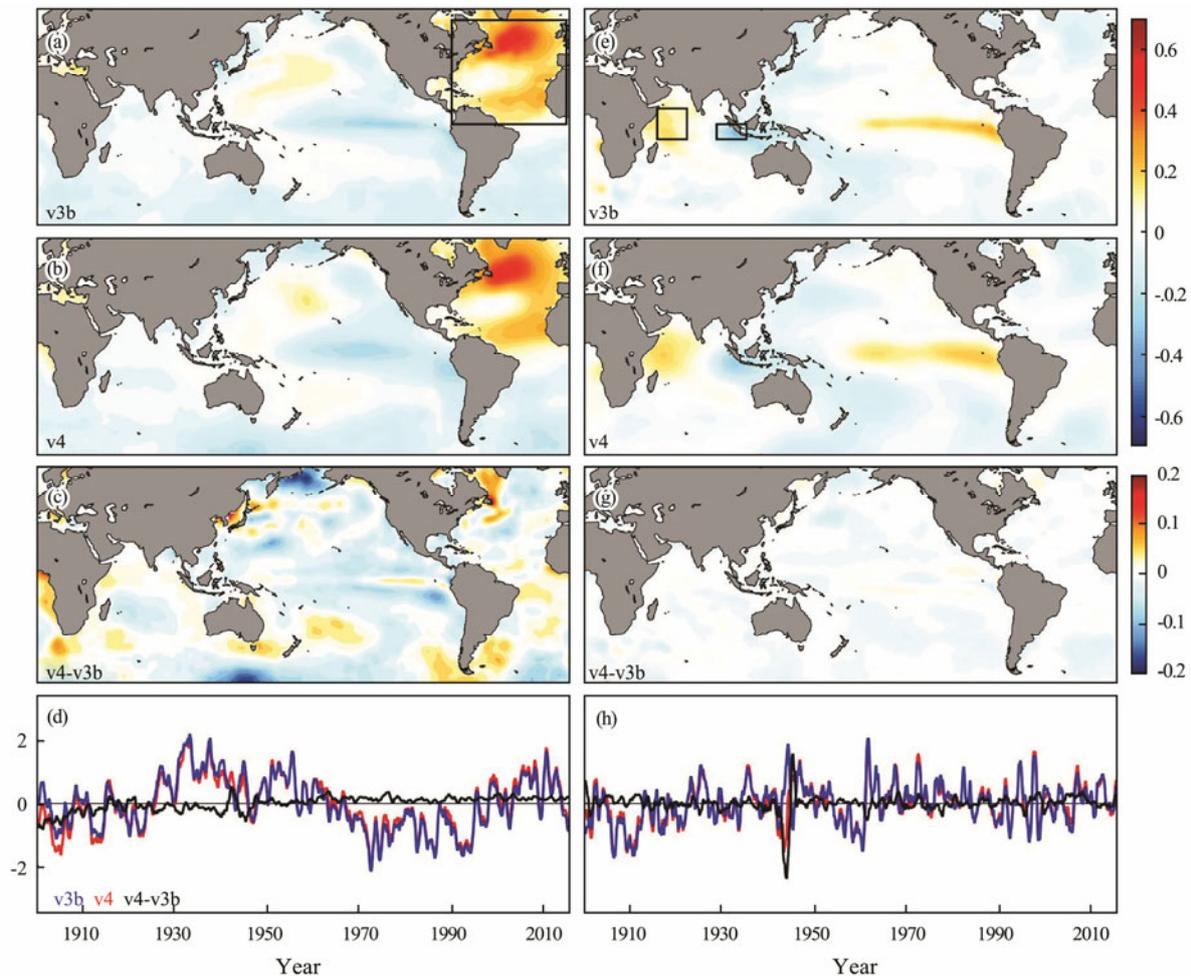
Fig.9 (a, b) Regression coefficients for the global SST* fields based on the AMO index for v3b and v4, respectively. Regression maps are scaled in units of ℃ per unit amplitude of the corresponding standardized AMO time series. (c) Difference in regression coefficient patterns between v4 and v3b. (d) Corresponding standardized AMO indices for v3b (blue) and v4 (red) and the difference in time series between v4 and v3b (black). (e, f) Regression coefficients for the global SST* fields based on DMI for v3b and v4, respectively. Regression maps are scaled in units of ℃ per unit am-plitude of the corresponding standardized DMI time series. (g) Difference in regression coefficient patterns between v4 and v3b. (h) The corresponding standardized DMI indices for v3b (blue) and v4 (red), and the difference in time series between v4 and v3b (black). Eleven-month running means are used to smooth the time series. The rectangles in (a) and (e) indicate the domains used to calculate the AMO index and DMI, respectively.

4 Discussion and Conclusions

ERSST v3b was recently upgraded with 11 improve-ments, as discussed by Huang *et al.* (2015a, b) and Liu *et al.* (2015). These improvements are based on newly available observations, improved analytical methods, and the latest knowledge of global SST variability since the release of the previous version in 2008. Neither the inter-action between two parametric options nor the impacts of multi-parameter combinations are linear. Therefore, the integrated impacts of these improvements on the final SST product are investigated in this work. While no 'real' answer as to which product is better than the other or during which period one is more realistic than the other,

the present intercomparison provides an overview of the changes in our understanding of the SST variability char-acterized by different products but within the same prod-uct family.

Three main differences are identified in this work.

First, the difference in global or regional mean SSTa values between the two datasets has exhibited a signifi-cant upward trend and interdecadal oscillations on the decadal time scale since the 1940s. This difference is par-ticularly important because it slightly shifts our under-standing of the recent global warming slowdown (Held, 2013), which is apparent in v3b but weakly evident in v4. Thus far, we can partly attribute this difference to the ship-buoy bias and ship-observation corrections that have been applied since 1940s. Whether and to what extent

other improvements contribute to the upward trend and oscillations observed have yet to be investigated; why this difference is shared in phases by the regional oceans, including the Pacific, Atlantic, and Indian Oceans, must also be uncovered. Identifying the underlying reasons behind the improvements presented by the systematical upgrade has substantial impacts on our understanding of decadal SST variability.

The second difference is related to the spatial structure of the ENSO-related variability between the two datasets. The slightly wider but weaker cold tongue region found in the new SST product may exert potential impacts on the coupled climate model simulations since the cold tongue region is one of the most active air-sea interaction areas in the world ocean. The difference observed could be due to the gap-filling assumptions and EOT training techniques applied to the two datasets, which means *in situ* observations of SST are insufficient to capture the fine structure of high-frequency global SST variability, even in the tropical Pacific Ocean, which was previously thought to be a well-sampled area.

The third difference refers the obvious local warming trend in the Arctic Ocean in the new SST product, which mainly benefits from recent upgrades of *in situ* observations of this ocean. However, our analysis shows that including Arctic Ocean observations does not change the global mean SSTa significantly. Cowtan and Way (2014) suggested that air temperatures over ice are better reconstructed from land-based air temperatures. Therefore, our analysis does not rule out the possibility of strong impacts of Arctic land warming on the global mean surface temperature, which is not studied in this work.

More general year-to-year differences between the two datasets (not shown), including differences in Arctic derived from updated *in situ* observations and ENSO-related patterns (Fig.7c), are observed. Previous intercomparisons between different SST reconstructions have shown that, in the area and period of rich SST observations, the SST values are more consistent with each other than those in the area and period without sufficient observations (Kennedy, 2014). Our analysis also confirms this understanding by showing that the magnitude of the difference in this area or period is generally larger than that in the well-observed epoch. Furthermore, we show that although the reconstruction of SST includes stronger uncertainties and larger magnitudes of difference in areas poor in *in situ* SST observations, no significant trend or oscillations are observed over the specified time scales. In other words, the difference in SST products in the area without rich SST observations closely resembles random noise and tends to decrease with increasing observations, for example, during the Argo period. However, while the magnitude of the difference in the area and period with rich SST observations is relatively small, the SST reconstructions may be affected by the choice of parameters and assumptions of specific methods, which tends to introduce the un-separable overall trends and/or oscillations to the decadal time scales with the same order of global climate warming and SST variability, as shown in Fig.4.

While this difference improves our understanding of the SST change, the main uncertainties of this difference have yet to be clearly elucidated.

References

Bretherton, C. S., Widmann, M., Dymnidov, V. P., Wallace, J. M., and Blade, I., 1999. The effective number of spatial degrees of freedom of a time-varying field. *Journal of Climate*, **12**: 1990-2009.

Chen, X., and Wallace, J. M., 2015. ENSO-like variability: 1900–2013. *Journal of Climate*, **28**: 9623-9641, DOI: 10.1175/JCLI-D-15-0322.1.

Cowtan, K., and Way, R. G., 2014. Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society*, **140**: 1935-1944, DOI: 10.1002/qj.2297.

Ding, Q., Steig, E. J., Battisti, D. S., and Wallace, J. M., 2012. Influence of the tropics on the Southern Annular Mode. *Journal of Climate*, **25**: 6330-6348, DOI: 10.1175/JCLI-D-11-00523.1.

Enfield, D. B., Mestas-Nuñez, A. M., and Trimble, P. J., 2001. The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S. *Geophysical Research Letters*, **28**: 2077-2080.

Folland, C. K., and Parker, D. E., 1995. Correction of instrumental biases in historical sea surface temperature data. *Quarterly Journal of the Royal Meteorological Society*, **121**: 319-367, DOI: 10.1002/qj.49712152206.

Folland, C. K., Colman, A. W., Smith, D. M., Boucher, O., Parker, D. E., and Vernier, J. P., 2013. High predictive skill of global surface temperature a year ahead. *Geophysical Research Letters*, **40**: 761-767, DOI: 10.1002/grl.50169.

Fyfe, J. C., Meehl, G. A., England, M. H., Mann, M. E., Santer, B. D., Flato, G. M., Hawkins, E. N., Gillett, P., Xie, S., Kosaka, Y., and Swart, N. C., 2016. Making sense of the early-2000s warming slowdown. *Nature Climate Change*, **6**: 224-228, DOI: 10.1038/nclimate2938.

Held, I. M., 2013. The cause of the pause. *Nature*, **501**: 318-319, DOI: 10.1038/501318a.

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

Huang, B., Banzon, V. F., Freeman, E., Lawrimore, J., Liu, W., Peterson, T. C., Smith, T. M., Thorne, P. W., Woodruff, S. D., and Zhang, H. M., 2015a. Extended reconstructed sea surface temperature version 4 (ERSST.v4), Part I. Upgrades and intercomparisons. *Journal of Climate*, **28**: 911-930, DOI: 10.1175/JCLI-D-14-00006.1.

Huang, B., Wang, W., Liu, C., Banzon, V., Zhang, H., and Lawrimore, J., 2015b. Bias adjustment of AVHRR SST and its impacts on two SST analyses. *Journal of Atmospheric & Oceanic Technology*, **32**: 372-387, DOI: 10.1175/JTECH-D-14-00121.1.

Kaplan, A., Cane, M. A., Kushnir, Y., Clement, A. C., Blumen-

thal, M. B., and Rajagopalan, B., 1998. Analyses of global sea surface temperature 1856–1991. *Geophysical Research Letters*, **103**: 18567-18589, DOI: 10.1029/97JC01736.

Karl, T. R., Arguez, A., Huang, B., Lawrimore, J. H., McMahon, J. R., Menne, M. J., Peterson, T. C., Vose, R. S., and Zhang, H. M., 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. *Reviews of Geophysics*, **52**: 1-32, DOI: 10.1002/2013RG000434.

Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E., and Saunby, M., 2011a. Reassessing biases and other uncertainties in sea surface temperature observations measured *in situ* since 1850: 1. Measurement and sampling uncertainties. *Journal of Geophysical Research Atmospheres*, **116**: D14103, DOI: 10.1029/2010JD015218.

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

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

Lewandowsky, S., Oreskes, N., Risbey, J. S., Newell, B. R., and Smithson, M., 2015. Seepage: Climate change denial and its effect on the scientific community. *Global Environmental Change*, **33**: 1-13, DOI: 10.1016/j.gloenvcha.2015.02.013.

Liu, W., Huang, B., Thorne, P. W., Banzon, V. F., Zhang, H.-M., Freeman, E., Lawrimore, J., Peterson, T. C., Smith, T. M., and Woodruff, S. D., 2015. Extended reconstructed sea surface temperature version 4 (ERSST.v4), part II: Parametric and structural uncertainty estimation. *Journal of Climate*, **28**: 931-951, DOI: 10.1175/JCLI-D-14-00007.1.

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

Rayner, N., Brohan, P., Parker, D., Folland, C., Kennedy, J., Vanicek, M., Ansell, T., and Tett, S., 2006. Improved analyses of changes and uncertainties in sea surface temperature measured *in situ* since the mid-nineteenth century: The HadSST2 data set. *Journal of Climate*, **19**: 446-469, DOI: 10.1175/JCLI3637.1.

Reynolds, R. W., Gentemann, C. L., and Corlett, G. K., 2010. Evaluation of AATSR and TMI satellite SST data. *Journal of Climate*, **23**: 152-165, DOI: 10.1175/2009JCLI3252.1.

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., and Wang, W., 2002. An improved *in situ* and satellite SST analysis for climate. *Journal of Climate*, **15**: 1609-1625, DOI: 10.1175/1520-0442(2002)015,1609: AIISAS.2.0.CO;2.

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

Schlesinger, M. E., and Ramankutty, N., 1994. An oscillation in the global climate system of period 65–70 years. *Nature*, **367**: 723-726, DOI: 10.1038/367723a0.

Simmons, A. J., Willett, K. M., Jones, P. D., Thorne, P. W., and Dee, D. P., 2010. Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational data sets. *Journal of Geophysical Research*, **115**: D01110, DOI: 10.1029/2009JD012442.

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

Smith, T. M., and Reynolds, R. W., 2004. Improved extended reconstruction of SST (1854–1997). *Journal of Climate*, **17**: 2466-2477.

Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J., 2008. Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006). *Journal of Climate*, **21**: 2283-2296, DOI: 10.1175/2007JCLI2100.1.

Thompson, D. W., Kennedy, J. J., Wallace, J. M., and Jones, P. D., 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., and Shea, D. J., 2006. Atlantic hurricanes and natural variability in 2006. *Geophysical Research Letters*, **33**: 285-293, DOI: 10.1029/2006GL026894.

van den Dool, H. M., Saha, S., and Johansson, A., 2000. Empirical orthogonal teleconnections. *Journal of Climate*, **13**: 1421-1435.

Woodruff, S. D., Worley, S. J., Lubker, S. J., Ji, Z., Freeman, J. E., Berry, D. I., Brohan, P., Kent, E. C., Reynolds, R. W., Smith, S. R., and Wilkinson, C., 2011. ICOADS Release 2.5: Extensions and enhancements to the surface marine meteorological archive. *International Journal of Climatology*, **31**: 951-967, DOI: 10.1002/joc.2103.

Zhang, Y., Wallace, J. M., and Battisti, D. S., 1997. ENSO-like decade-to-century scale variability. *Journal of Climate*, **10**: 1004-1020.

(Edited by Xie Jun)