The importance of unresolved biases in 20th century sea-surface temperature observations

Luke L. B. Davis *†, David W. J. Thompson†, John J. Kennedy‡, and Elizabeth C. Kent§

†*Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado USA*

‡*Met Office Hadley Centre, Exeter, UK*

§*National Oceanography Centre, Southampton, UK*

**Corresponding author address:* Luke L. B. Davis, Department of Atmospheric Science, Colorado State University, 3915 W. Laporte Ave., Fort Collins, Colorado USA

E-mail: luke.davis@colostate.edu

ABSTRACT

A new analysis of sea-surface temperature (SST) observations indicates notable uncertainty in observed decadal climate variability in the second half of the 20th century, particularly during the decades following World War II. The uncertainties are revealed by exploring SST data binned separately for the two predominant measurement types: "engine-room intake" (ERI) and "bucket" measurements. ERI measurements indicate large decreases in global-mean SSTs from 1950 to 1975, whereas "bucket" measurements indicate increases in SST over this period before bias adjustments are applied but decreases after they are applied. The trends in the bias adjustments applied to the "bucket" data are larger than the global-mean trends during the period 1950-1975, and thus the global-mean trends during this period derive largely from the adjustments themselves. This is critical, since the adjustments are based on incomplete information about the underlying measurement methods, and are thus subject to considerable uncertainty. The uncertainty in decadal-scale variability is particularly pronounced over the North Pacific, where the sign of low-frequency variability through the 1950s-1970s is different for each measurement type. The uncertainty highlighted here has important – but in our view widely overlooked – implications for the interpretation of observed decadal climate variability over both the Pacific and Atlantic basins during the mid-to-late 20th century.

Capsule Summary

Biases in sea-surface temperature observations lead to larger uncertainties in our understanding of mid-to-late 20th century climate variability than previously thought.

1. Introduction

The world ocean warmed by ∼0.75K from 1900-2016, but the warming did not occur monotonically: temperatures increased during the first half of the 20th century, decreased slightly during the decades following World War II, and increased rapidly after ∼1975 (Hartmann et al. 2013). The decreases in ocean temperatures from the 1950s-1970s are apparent in SSTs averaged over the globe and both the Atlantic and Pacific sectors (Figures 1k-o, black time series; Figures 2k-o, black bars).

The absence of warming during the decades following World War II is important because it coincides with steadily increasing concentrations of carbon dioxide over the same period. Several theories have been proposed to explain the absence of warming during this period, including increases in atmospheric sulfate aerosols (Tett et al. 2002; Lamarque et al. 2010; Booth et al. 2012; Myhre et al. 2013; Folland et al. 2018) and decadal variability in the ocean (Delworth and Mann 2000; Baines and Folland 2007; Knight et al. 2006; Semenov et al. 2010). Here we provide novel analyses of SST data separated into the two primary measurement sources to demonstrate that the uncertainty in decadal variability of SST from the 1950s-1970s is at least as large as the observed decadal variability itself. The results highlight the critical importance of considering uncertainty in SST observations in analyses of observed decadal climate variability.

SST data during the period after 1980 are derived from several *in situ* and remotely-sensed sources (Kent et al. 2010; Kennedy et al. 2011b). But SST data prior to 1980 are derived almost entirely from two *in situ* sources via "ships of opportunity": 1) the temperature of seawater in buckets that have been submerged below the ocean surface and then hauled back onto a ship deck (bucket measurements); and 2) the temperature of the pumped water supply to an engine room (engine-room intake or ERI measurements) (Kent et al. 2010; Kennedy et al. 2011b). A comparatively small number of hull sensor observations are also included in the ERI category, as the biases in both hull sensor and ERI data are thought to be governed by similar factors (Kennedy et al. 2011b).

Bucket and ERI measurements both exhibit substantial measurement biases (Kent and Kaplan 2006; Rayner et al. 2006; Kent et al. 2010; Kennedy et al. 2011b; Kent et al. 2017; Folland and Parker 1995). ERI measurements are often warm-biased due to the transfer of heat from the superstructure of the ship as water passes through pipes, while bucket measurements are often cold-biased due to the exchange of latent and sensible heat with the surrounding air. If the mix of measurement types and their relative biases are well understood, then the biases can be adjusted so that they have little effect on the time evolution of spatially-averaged temperature data. But if the mix of measurement types is poorly documented, large biases can remain after adjustment, even in widely used climate data sources (Folland and Parker 1995; Thompson et al. 2008; Karl et al. 2015).

In principle, SST data stratified by measurement type provide the opportunity to assess the reproducibility of SST variability in subsets of the data not influenced by changes to instrumentation. With this in mind, the UK Met Office Hadley Centre developed SST datasets stratified into bucket and ERI measurements in conjunction with the release of their most recent gridded dataset HadSST3 (Kennedy et al. 2011a,b). The bucket and ERI data are available over the period 1946-2006, and were developed in the same way as the full HadSST3 dataset (Kennedy et al. 2011a,b); that is, by 1) estimating the measurement types of SST observations in the International Comprehensive Ocean-Atmosphere Data Set release 2.5 archive (ICOADS2.5; Woodruff et al. (2011)); 2) consolidating the observations onto monthly $5^\circ \times 5^\circ$ grids; 3) applying bias adjustment schemes unique to each measurement type; and 4) accounting for parametric uncertainty in the bias adjustment schemes by generating 100 plausible realizations of the adjustments. For each of the bucket-only and ERI-only datasets, observations estimated to be from the other measurement type were ignored, and for this analysis, grid boxes without valid data from the other measurement type were excluded. The latter step ensures that the bucket-only and ERI-only data are "co-located," or have the same spatial coverage through time at the grid box level. This is critical, as measurement types are often distributed differently across each ocean basin (Kent and Taylor 2006).

The identification of SST methodology is imperfect; in many cases, the ICOADS2.5 metadata does not provide specific information about the measurement method, and hence the measurement type must be estimated from other information, such as country of origin (Kennedy et al. 2011b). Even if the measurement type is indicated by the metadata, the indication is sometimes incorrect (Kent et al. 2007). In other cases, the general type of measurement is known (e.g., bucket), but specific aspects of the measurement (e.g., the construction and insulation of the bucket) are not. Nevertheless, the bucket-only and ERI-only datasets reflect the best available estimates of mid-20th century SST variability minimally influenced by changes to instrumentation. Together, the two datasets thus provide a unique opportunity to explore uncertainty in observed decadal variability.

2. The Problem

The unadjusted bucket and ERI data yield remarkably different renditions of 20th century SST variability, particularly prior to ~1975 (red and blue time series in Figs 1a-e; red and blue bars in Figs 2a-e; Figs 3a and 3b). For example, the ERI data exhibit cooling of the Pacific ocean from 1950-1975, whereas the unadjusted bucket data indicate warming (Figs 1d-e; Figs 2d-e; Figs

3a and 3b). Likewise, the ERI data exhibit cooling in the global average over the same period, whereas the bucket data indicate warming (Figs 1a, 2a).

The adjustments applied to the ERI data using the HadSST3 bias adjustment scheme are mostly stationary in time, with the exception of the short-term bias adjustments applied to the Atlantic sector during the early 1990s (red time series in Figs 1f-j; see Kent and Kaplan (2006)). Hence they do not notably affect estimates of decadal variability (red bars in Figs 2f-j; Fig. 3d). In contrast, the adjustments applied to the bucket data introduce a substantial 0.1K/decade cooling over the period 1950 to 1975 (blue time series in Figs 1f-j; blue bars in Figs 2f-j; Fig. 3e), due to the assumed transition from canvas to rubber buckets (Kent et al. 2010; Kennedy et al. 2011b). The cooling introduced by the adjustments applied to the bucket data ranges from -0.05 to -0.15 K/decade across the 100 realizations of the HadSST3 bias adjustments (error bars in Figs 2f-j). The adjusted bucket temperature data exhibit robust cooling in Atlantic basin averages but not in the global and Pacific basin averages. The Atlantic cooling is apparent in all 100 realizations of the adjusted bucket data (whiskers on blue bars in Figs 2l-m) and is also statistically significant with respect to the detrended variability in the data (Figs S1l-m, Supplemental Materials). The adjustments for both ERI and bucket data are roughly stationary in time during the 1976-2006 period (Figs 4d-f).

The resulting adjusted ERI and bucket data (red and blue time series in Figs 1k-o; red and blue bars in Figs 2k-o; Figs 3g-h) are in closer agreement with each other than their unadjusted counterparts. However, the trends in the SST field over the period 1950-1975 are still notably different for the two measurement types. In the global-average, the cooling in the adjusted bucket data is roughly half as large as the cooling in the adjusted ERI data (Figs 1k, 2k). The discrepancies are especially notable in the Pacific sector, where the adjusted ERI data exhibit cooling over the period 1950-1975 but the adjusted bucket data exhibit relatively little change in temperature (Figs

1n-o; Figs 2n-o; Figs 3g-h). In contrast, over the Atlantic sector the adjusted ERI data exhibit significantly weaker cooling than the adjusted bucket data (Figs 1l-m; Figs 2l-m; Figs 3g-h). These patterns of disagreement are stronger in the North Pacific and North Atlantic during boreal winter (Figs S2l-m, Figs S3g-h; Supplemental Materials).

After 1975, the agreement between the ERI and bucket data improves as the magnitude of the adjustments decreases and the overall quality and consistency of the observations increases. Nevertheless there remain notable differences in the adjusted ERI and bucket SST trends over the 1976-2006 period, especially over the south-central Pacific sector (Figs 4g-i) and during austral winter (Figs S4g-i, Supplemental Materials).

Importantly, the amplitudes of the trends in the adjusted ERI and bucket data over the 1950-1975 period are comparable to the differences between them (compare the red and blue bars with the pink bars in Figs 2k-o, and Figs 3g-h with Fig. 3i), which points to the scale of the uncertainty in the adjusted data. As indicated by the whiskers on the pink bars in Figs 2k-o and the stippling in Fig. 3i, the 100 bias adjustment realizations cannot account for these differences, and thus do not entirely characterize the bias uncertainties in the trends. When averaged over large spatial domains, the amplitudes of the trends are also comparable to the trends in the bias adjustments themselves (compare the middle and bottom rows of Fig. 2). This is key, as the bias adjustment schemes are subject to considerable uncertainty, particularly prior to 1980 (Kennedy et al. 2011b; Kent et al. 2017).

In the case of HadSST3, the bias adjustment schemes are derived from metadata contained in ICOADS2.5 and historical documentation (Kennedy et al. 2011b). However, the metadata are frequently incomplete, and thus various sources of bias are not known with confidence, including bucket type, the speed of the ship, the depth from which water for the engine-room is drawn, and whether a datum is derived from a bucket or ERI measurement in the first place (Kent and Taylor

2006; Kent et al. 2017). For example, the HadSST3 bias adjustments assume 40-80% of the SST data from 1960-1980 are derived from bucket measurements, whereas a recent reassessment of measurement type suggests the fraction of bucket measurements over this time is consistently closer to 40% (Carella et al. 2018).

The uncertainties in the bias adjustment scheme applied to HadSST3 data can be inferred from the time series in Fig. 1 as follows (see also Kent et al. (2017) and Carella et al. (2018)). The unadjusted ERI and bucket time series can be decomposed as:

$$\mathrm{ERI}_{\mathrm{unadjusted}} = \mathrm{ERI}_{\mathrm{true}} + \mathrm{ERI}_{\mathrm{true\ bias}} \tag{1}$$

$$\mathrm{B}_{\mathrm{unadjusted}} = \mathrm{B}_{\mathrm{true}} + \mathrm{B}_{\mathrm{true\ bias}} \tag{2}$$

where $\mathrm{ERI}_{\mathrm{true}}$ and $\mathrm{B}_{\mathrm{true}}$ are the "true" SST data in the absence of measurement bias, and $\mathrm{ERI}_{\mathrm{true\ bias}}$ and $\mathrm{B}_{\mathrm{true\ bias}}$ are the "ideal" bias adjustments. Since the bucket and ERI data used here are co-located in space, it follows that $\mathrm{ERI}_{\mathrm{true}} = \mathrm{B}_{\mathrm{true}}$ over area averages large enough to suppress sampling and measurement uncertainties (Kennedy et al. 2011a; Carella et al. 2018), and therefore

$$\mathrm{ERI}_{\mathrm{unadjusted}} - \mathrm{B}_{\mathrm{unadjusted}} = \mathrm{ERI}_{\mathrm{true\ bias}} - \mathrm{B}_{\mathrm{true\ bias}} \tag{3}$$

The uncertainty in the bias adjustments applied to the bucket and ERI data (and hence to the HadSST3 dataset) can thus be estimated by comparing (a) the differences between the unadjusted ERI and bucket data with (b) the differences between the ERI and bucket bias estimates (i.e., the negative of the ERI and bucket bias adjustments). If the bias adjustments are ideal, then the time series given by (a) and (b) should be identical. Note that the time series can also be identical if there is a common bias in the ERI and bucket measurements; the series being identical is a necessary but not sufficient criterion for ideal adjustments.

Fig. 5 shows the results of the above calculation for the domains considered in Figs 1-2. The orange lines indicate the differences between the ERI and bucket bias estimates averaged over all

100 pairs of adjustments (the range given by the 100 realizations is indicated by orange shading); the black lines indicate the differences between the unadjusted ERI and bucket data. Again, (1) if the bias adjustments are ideal, then the black and orange lines should overlie each other, and (2) if the 100 realizations of the bias adjustments characterize the uncertainty in the adjustments, then the black lines should lie within the regions of orange shading. Overall, the adjustments required to bring ERI and bucket data into agreement (black lines) are clearly much larger and much more variable than the mean of the actual bias adjustments applied to the HadSST3 data. The inferred uncertainties in the bias adjustments are comparable to the amplitude of the observed decadal variability in the SST field.

The uncertainties in decadal-scale variability indicated in Figs 1-5 also affect the two other major historical SST data sets based on the ICOADS2.5 archive: the Centennial *in situ* Observation-Based Estimates of SST (COBE-SST2; Hirahara et al. (2013)) developed by the Japanese Meteorological Agency, and version 4 of the Extended Reconstructed SST dataset (ERSST4; Huang et al. (2014) and Liu et al. (2014)) released by the US National Climatic Data Center. The ERSSTv4 and COBE-SST2 datasets are included to provide a point of comparison with SST data that employ very different bias correction schemes: The bias adjustments applied in HadSST3 and COBE-SST2 are both based on information about measurement type as inferred from the metadata; the adjustments applied in ERSST4 are based only on comparisons with night-time marine air temperature (NMAT) data, which require their own bias adjustments (Rayner et al. 2003; Kent et al. 2013; Kennedy 2014). In general, over the 1950-1975 period, the trends in the bias-adjusted COBE-SST2 data are similar to those in the HadSST3 data, whereas the trends in the bias-adjusted NMAT and ERSST4 data are somewhat weaker than those in the HadSST3 data, particularly over the Atlantic Ocean and in the global-mean (Figs 2k-o; see Methods for details of the analysis).

3. So What?

The results shown here reveal a level of regional uncertainty in observed SSTs that is not widely acknowledged in the climate dynamics literature. In our view, it should be. Confidence in observed decadal variability derives from confidence in the bias adjustments applied to the SST data. And as shown here, the uncertainty in the bias adjustment schemes is frequently comparable to the amplitude of the observed decadal variability itself. The uncertainty has important implications for our understanding of the role of aerosols in 20th century climate change (Tett et al. 2002; Lamarque et al. 2010; Booth et al. 2012; Myhre et al. 2013; Folland et al. 2018), since aerosols are believed to have contributed to the absence of global warming during the mid 20th century (Kobayashi et al. 2015; Laloyaux et al. 2017; Taylor et al. 2011; Flato et al. 2013). It also has important implications for quantifying the amplitudes of patterns of decadal-scale variability, particularly over the problematic North Pacific sector (Fig. 1 and 2), and in association with Pacific and Atlantic decadal variability (Mantua et al. 1997; Mantua and Hare 2002; Newman et al. 2016; Delworth and Mann 2000; Baines and Folland 2007; Knight et al. 2006; Semenov et al. 2010).

The findings indicate notable shortcomings in our ability to accurately classify SST measurement methods and to quantify the associated biases. Complicating matters is that measurement biases vary not only from one measurement method to the next, but also within the individual methods: ERI biases can vary between individual ships and recruiting countries (Kent et al. 1993); bucket biases depend on the bucket type, and the transition from canvas to rubber buckets for a given recruiting country is highly uncertain (Kennedy et al. 2011b). Additionally, the metadata necessary to identify ships is often missing from ICOADS (Carella et al. 2018), and the proportions of recruiting countries can change substantially over time, especially before $\sim$1970 (Fig. S5, Supplemental Materials; Thompson et al. (2008)). For example, the large differences between trends in the bucket and ERI data over the Pacific sector relative to those over the Atlantic sector (Figs. 2l-o, 3i) are potentially due to differences in the types of bucket and ERI measurements used in each region, as implied by the differences in recruiting countries between the two sectors (Fig. S6, Supplemental Materials). Not surprisingly, despite recent advances (Freeman et al. 2017; Carella et al. 2018; Hausfather et al. 2017; Cowtan et al. 2017; Hirahara et al. 2013), it may take years to resolve the discrepancies between the ERI and bucket time series highlighted here.

What is the best way forward? The recent review of SST biases by Kent et al. (2017) concludes with a series of recommendations for improving the reliability of historical SST bias estimations, especially after World War II. These include improving the metadata and volume of observations in the ICOADS archive, improving the classification of measurement methods from documentation and by analysing of data characteristics, improving the physical and statistical models used to estimate SST bias, and entraining more scientists into the field of SST bias adjustment. Novel analyses could include clustering of observations by individual ship or recruiting country, which may help isolate the bias variations within each measurement method. Our results make clear the critical importance of the recommendations in Kent et al. (2017) for improving our understanding of 20th century climate variability.

Acknowledgments. We thank the three anonymous reviewers for their insightful comments, which greatly improved the paper. L.L.B.D. and D.W.J.T. were supported by the NSF Climate and Large-Scale Dynamics Program (AGS-1547003). J.J.K. was supported by the Joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101). E.C.K. was supported by the NERC (NE/J020788/1).

APPENDIX

Methods

The HadSST3 data (Kennedy et al. 2011b) and night-time marine air temperature data (Kent et al. 2013) were obtained from the Met Office Hadley Centre (`https://metoffice.gov.uk/hadobs`). Subsequent to this study, the ERI-only and bucket-only data are also published on the Hadley Centre website. The unadjusted ICOADS sea-surface temperature observations (Woodruff et al. 2011) were obtained from the National Center for Atmospheric Research (`https://rda.ucar.edu/`). The Japan Meteorological Agency (COBE-SST2; Hirahara et al. (2013)) and National Climatic Data Center (ERSST4; Huang et al. (2014) and Liu et al. (2014)) sea-surface temperature data were both obtained from the Earth System Research Laboratory Physical Sciences Division (`https://esrl.noaa.gov/psd`). To accommodate comparisons with the Hadley Centre data, the COBE-SST2 and ERSST4 data were (1) re-gridded onto the $5^{\circ} \times 5^{\circ}$ resolution HadSST3 grid and (2) had their respective monthly 1961-1990 climatologies subtracted (to match the HadSST3 climatology period). The gridded NMAT, HadSST3, COBE-SST2, and ERSST4 data were matched to the spatial coverage of the co-located ERI-only and bucket-only data. The coordinate boundaries used for each spatial average are shown in Table A1 (grid boxes were weighted by the cosine of the central latitude and the ocean fraction within each box). Note that all data used in this study are in anomaly form with respect to the 1961-1990 base period.

References

Baines, P. G., and C. K. Folland, 2007: Evidence for a rapid global climate shift across the late 1960s. *Journal of Climate*, **20 (12)**, 2721–2744.

Booth, B. B. B., N. J. Dunstone, P. R. Halloran, T. Andrews, and N. Bellouin, 2012: Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, **484 (7393)**, 228–232.

Carella, G., J. J. Kennedy, D. I. Berry, S. Hirahara, C. J. Merchant, S. Morak-Bozzo, and E. C. Kent, 2018: Estimating sea surface temperature measurement methods using characteristic differences in the diurnal cycle. *Geophysical Research Letters*, **45 (1)**, 363–371.

Cowtan, K., R. Rohde, and Z. Hausfather, 2017: Evaluating biases in sea surface temperature records using coastal weather stations. *Quarterly Journal of the Royal Meteorological Society*, doi:10.1002/qj.3235.

Delworth, T. L., and M. E. Mann, 2000: Observed and simulated multidecadal variability in the Northern Hemisphere. *Climate Dynamics*, **16 (9)**, 661–676.

Flato, G., and Coauthors, 2013: *Evaluation of Climate Models*, Book Section 9, 741–866. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Folland, C. K., O. Boucher, A. Colman, and D. E. Parker, 2018: Causes of irregularities in trends of global mean surface temperature since the late 19th century. *Science advances*, **4 (6)**, eaao5297.

Folland, C. K., and D. E. Parker, 1995: Correction of instrumental biases in historical sea surface temperature data. *Quarterly Journal of the Royal Meteorological Society*, **121 (522)**, 319–367.

Freeman, E., and Coauthors, 2017: ICOADS Release 3.0: A major update to the historical marine climate record. *International Journal of Climatology*, **37 (5)**, 2211–2232.

Hartmann, D., and Coauthors, 2013: *Observations: Atmosphere and Surface*, Book Section 2, 159–254. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Hausfather, Z., K. Cowtan, D. C. Clarke, P. Jacobs, M. Richardson, and R. Rohde, 2017: Assessing recent warming using instrumentally homogeneous sea surface temperature records. *Science Advances*, **3 (1)**, e1601 207.

Hirahara, S., M. Ishii, and Y. Fukuda, 2013: Centennial-Scale Sea Surface Temperature Analysis and Its Uncertainty. *Journal of Climate*, **27 (1)**, 57–75.

Huang, B., and Coauthors, 2014: Extended reconstructed sea surface temperature version 4 (ERSST.v4). Part I: Upgrades and intercomparisons. *Journal of Climate*, **28 (3)**, 911–930.

Karl, T. R., and Coauthors, 2015: Possible artifacts of data biases in the recent global surface warming hiatus. *Science*, **348 (6242)**, 1469–1472.

Kennedy, J. J., 2014: A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Reviews of Geophysics*, **52 (1)**, 1–32.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011a: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *Journal of Geophysical Research: Atmospheres*, **116 (D14)**, D14 103.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 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 (D14)**, D14 104.

Kent, E. C., and A. Kaplan, 2006: Toward estimating climatic trends in SST. Part III: Systematic biases. *Journal of Atmospheric and Oceanic Technology*, **23 (3)**, 487–500.

Kent, E. C., J. J. Kennedy, D. I. Berry, and R. O. Smith, 2010: Effects of instrumentation changes on sea surface temperature measured in situ. *Wiley Interdisciplinary Reviews: Climate Change*, **1 (5)**, 718–728.

Kent, E. C., 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. *Journal of Geophysical Research: Atmospheres*, **118 (3)**, 1281–1298.

Kent, E. C., and P. K. Taylor, 2006: Toward estimating climatic trends in SST. Part I: Methods of measurement. *Journal of Atmospheric and Oceanic Technology*, **23 (3)**, 464–475.

Kent, E. C., P. K. Taylor, B. S. Truscott, and J. S. Hopkins, 1993: The Accuracy of Voluntary Observing Ships' Meteorological Observations-Results of the VSOP-NA. *Journal of Atmospheric and Oceanic Technology*, **10 (4)**, 591–608.

Kent, E. C., S. D. Woodruff, and D. I. Berry, 2007: Metadata from WMO Publication No. 47 and an Assessment of Voluntary Observing Ship Observation Heights in ICOADS. *Journal of Atmospheric and Oceanic Technology*, **24 (2)**, 214–234.

Kent, E. C., and Coauthors, 2017: A call for new approaches to quantifying biases in observations of sea surface temperature. *Bulletin of the American Meteorological Society*, **98 (8)**, 1601–1616.

Knight, J. R., C. K. Folland, and A. A. Scaife, 2006: Climate impacts of the Atlantic Multidecadal Oscillation. *Geophysical Research Letters*, **33 (17)**, L17 706.

Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, **93 (1)**, 5–48.

Laloyaux, L., E. De Boisseson, and P. Dahlgren, 2017: CERA-20C: An Earth system approach to climate reanalysis. *J. ECMWF Newsl*, **150**, 25–30.

Lamarque, J.-F., and Coauthors, 2010: Historical (1850-2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: Methodology and application. *Atmospheric Chemistry and Physics*, **10 (15)**, 7017–7039.

Liu, W., and Coauthors, 2014: Extended reconstructed sea surface temperature version 4 (ERSST.v4): Part II. Parametric and structural uncertainty estimations. *Journal of Climate*, **28 (3)**, 931–951.

Mantua, N. J., and S. R. Hare, 2002: The Pacific Decadal Oscillation. *Journal of oceanography*, **58 (1)**, 35–44.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific Interdecadal Climate Oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society*, **78 (6)**, 1069–1079.

Myhre, G., and Coauthors, 2013: *Anthropogenic and Natural Radiative Forcing*, Book Section 8, 659–740. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Newman, M., and Coauthors, 2016: The Pacific Decadal Oscillation, revisited. *Journal of Climate*, **29 (12)**, 4399–4427.

Rayner, N. A., P. Brohan, D. E. Parker, C. K. Folland, J. J. Kennedy, M. Vanicek, T. J. Ansell, and S. F. B. Tett, 2006: Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset. *Journal of Climate*, **19 (3)**, 446–469.

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. *Journal of Geophysical Research: Atmospheres*, **108 (D14)**, 4407.

Semenov, V. A., M. Latif, D. Dommenget, N. S. Keenlyside, A. Strehz, T. Martin, and W. Park, 2010: The impact of North Atlantic-Arctic multidecadal variability on Northern Hemisphere surface air temperature. *Journal of Climate*, **23 (21)**, 5668–5677.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2011: An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, **93 (4)**, 485–498.

Tett, S. F. B., and Coauthors, 2002: Estimation of natural and anthropogenic contributions to twentieth century temperature change. *Journal of Geophysical Research: Atmospheres*, **107 (D16)**, ACL 10–1.

Thompson, D. W. J., J. J. Kennedy, J. M. Wallace, and P. D. Jones, 2008: A large discontinuity in the mid-twentieth century in observed global-mean surface temperature. *Nature*, **453 (7195)**, 646–649.

Woodruff, S. D., and Coauthors, 2011: ICOADS Release 2.5: Extensions and enhancements to the surface marine meteorological archive. *International Journal of Climatology*, **31 (7)**, 951–967.

LIST OF TABLES

Region	Boundaries	
Globe	60°S – 60°N,	180°W – 180°E
Atlantic Ocean	10°N – 60°N, 60°S – 10°N,	80°W – 0°E 70°W – 20°E
Pacific Ocean	10°N – 60°N, 10°S – 10°N, 60°S – 10°S,	100°E – 100°W 100°E – 70°W 150°E – 70°W
North Atlantic	20°N – 60°N,	80°W – 0°E
North Pacific	20°N – 60°N,	120°E – 100°W

Table A1. **Boundaries used for spatial averages of gridded SST data.**

LIST OF FIGURES

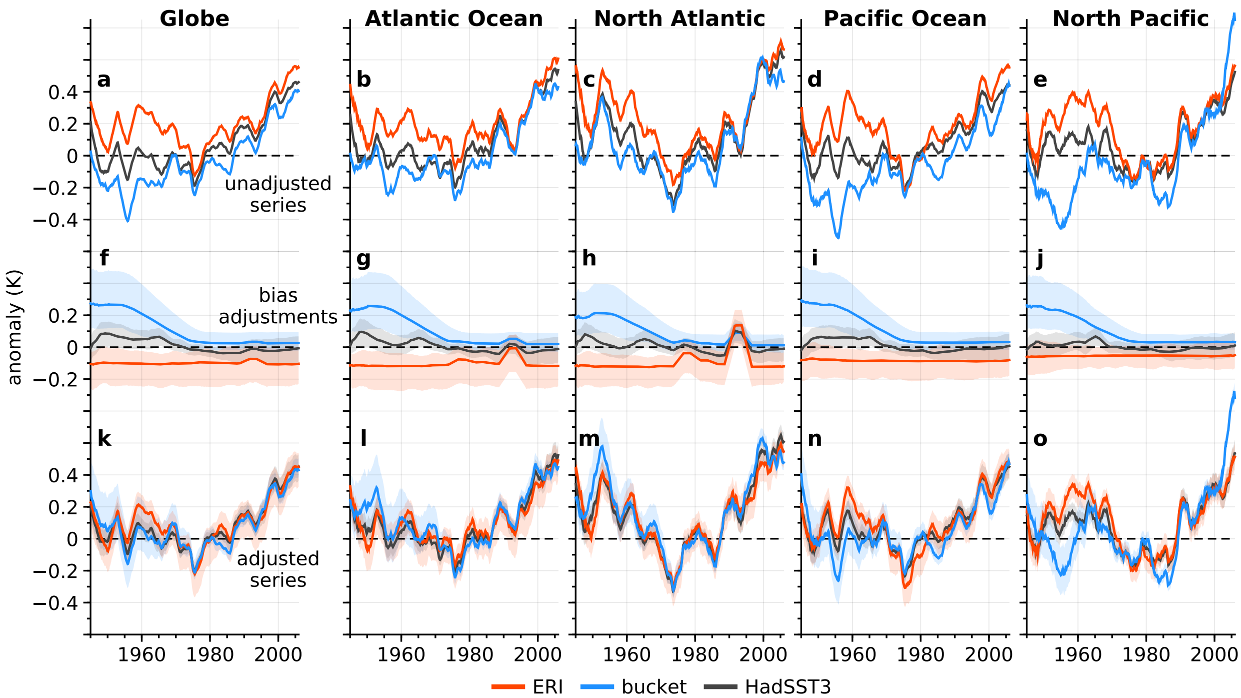


FIG. 1. **Time series of sea-surface temperature anomalies (K, relative to 1961-1990) for the datasets and regions indicated.** From top-to-bottom: (**a-e**) area-averages of the unadjusted engine-room intake (ERI), bucket, and HadSST3 data, (**f-j**) area-averages of the bias adjustments applied to the data, and (**k-o**) area-averages of the adjusted data. The North Atlantic and North Pacific time series indicate averages north of 20°N. The shaded regions in the middle and bottom rows indicate the range of all 100 realizations of the HadSST3 bias adjustments; the thick lines indicate the average over all 100 realizations. The time series are smoothed by a centred 37-month running mean for display purposes. Note that the ERI, bucket, and HadSST3 data are co-located in space (see Methods).

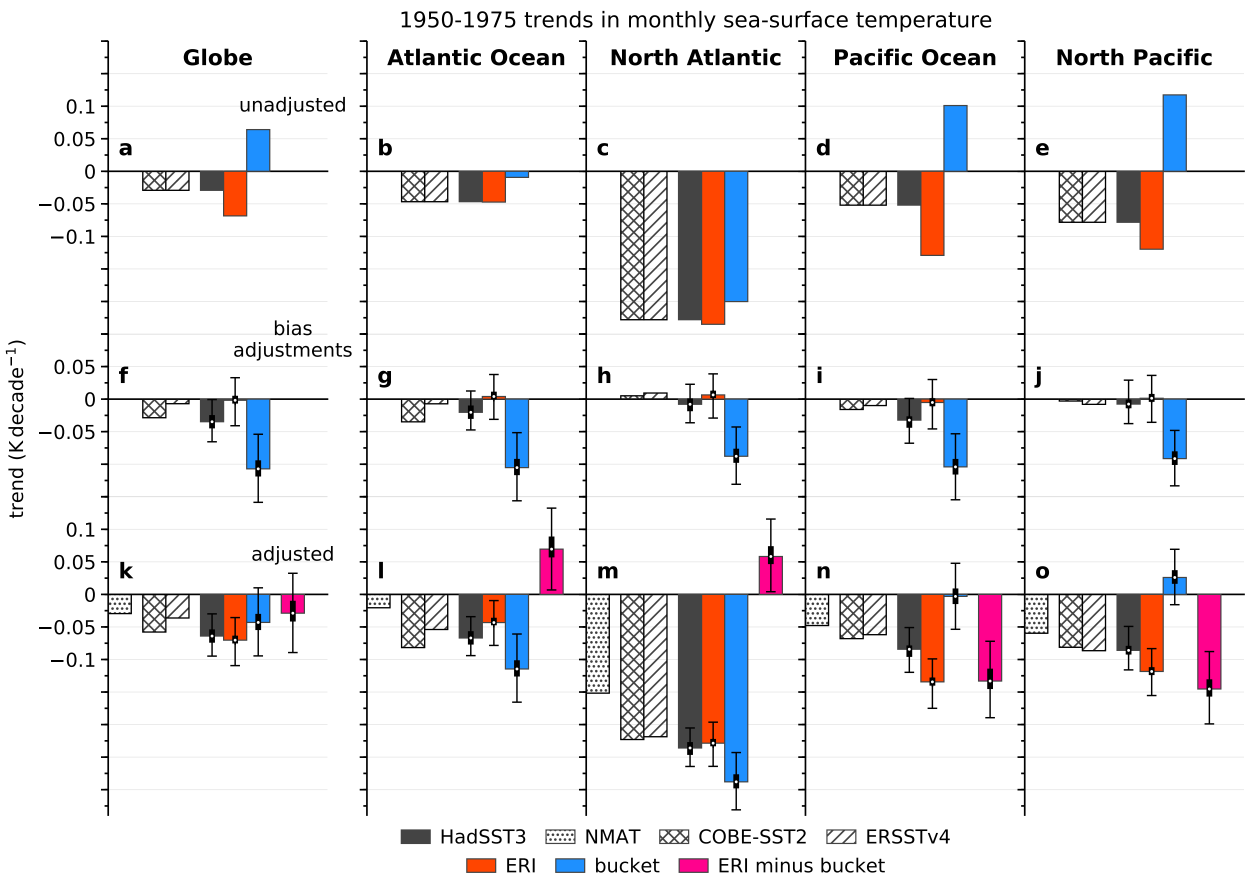
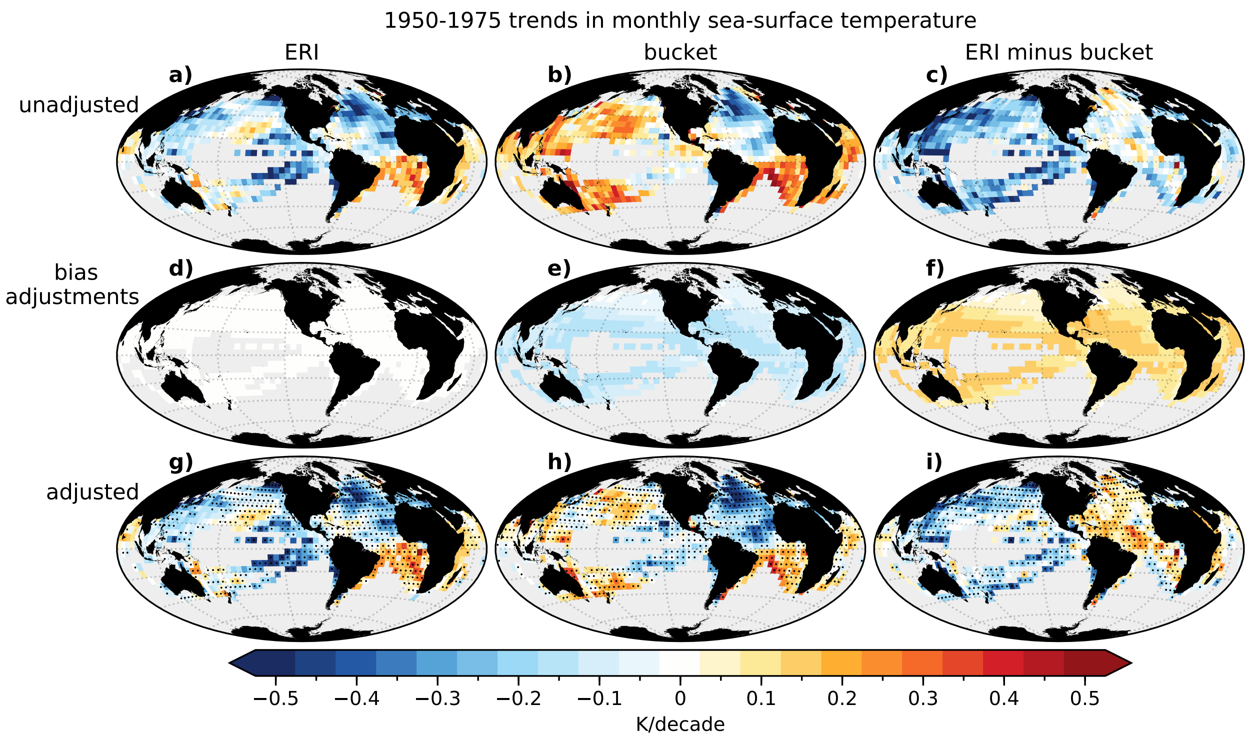


FIG. 2. **Linear least-squares trends in sea-surface temperature anomalies from 1950 to 1975 for the datasets and regions indicated.** From top to bottom: (**a-e**) trends in the area-averaged unadjusted data, (**f-j**) trends in the area-averaged bias adjustments, and (**k-o**) trends in the area-averaged adjusted data. The North Atlantic and North Pacific trends indicate averages north of 20°N. Thin error bars in the middle and bottom rows indicate the range of trends from all 100 realizations of the HadSST3 bias adjustments. The thicker error bars indicate interquartile ranges of the trends. The significance of the mean trends with respect to the internal variability is given in Supplemental Materials. Note that the COBE-SST2, NMAT, and ERSST4 data are co-located with the engine-room intake (ERI), bucket, and HadSST3 data (see Methods). Note also that the NMAT data are only available in their adjusted form.



413 FIG. 3. **Linear least-squares trends in sea-surface temperature anomalies from 1950 to 1975 for the indicated datasets.** From left-to-right: trends in the engine-room intake (ERI) dataset, bucket dataset, and their differences. From top-to-bottom: (**a-c**) trends in the unadjusted data, (**d-f**) trends in the bias adjustments (i.e., the mean of the trends calculated for all 100 realizations of the HadSST3 bias adjustments), and (**g-i**) trends in the adjusted data. Trends were not computed where fewer than 50% of the timesteps from 1950 to 1975 were available. Grid boxes are stippled where the trends (g-h) and trend differences (i) have the same sign across all 100 bias adjustment realizations.

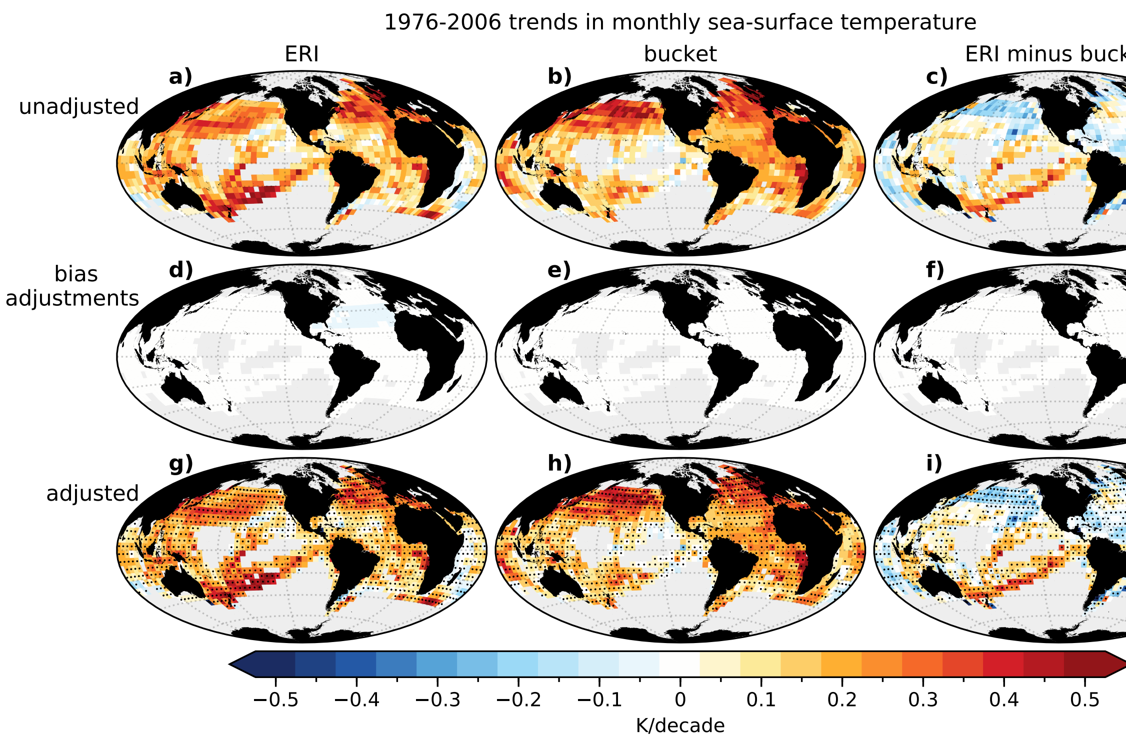


FIG. 4. **As in Figure 3, but for the 1976-2006 period.**

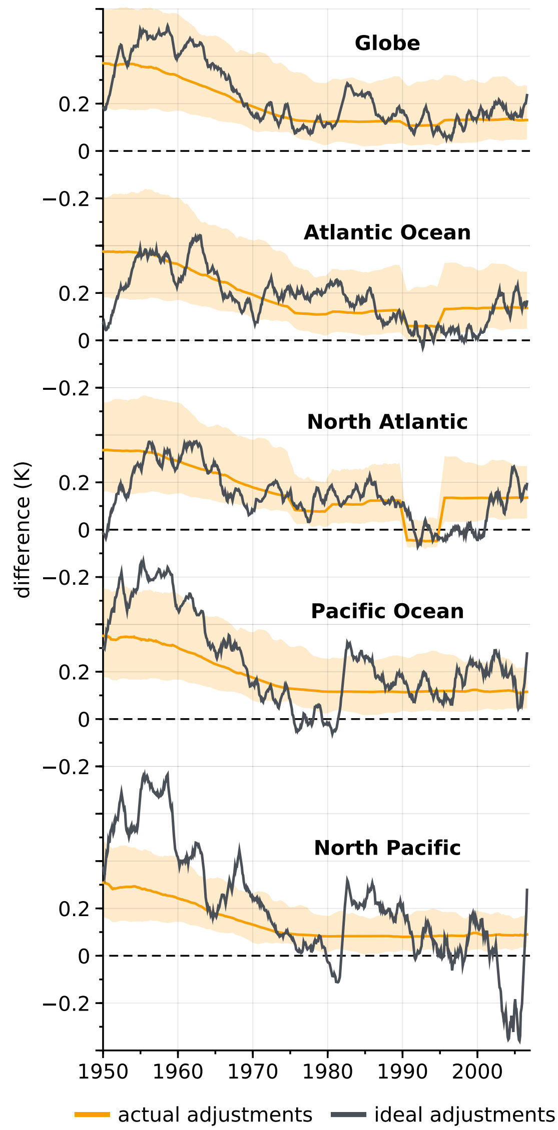


FIG. 5. **Estimating the uncertainty in the bias adjustments for the indicated regions.** The orange lines represent the differences between the bucket and engine-room intake (ERI) bias estimates averaged over all 100 realizations of the bias adjustments, and shading indicates the range of the differences. The black lines represent the differences between the unadjusted ERI and bucket data. The latter correspond to the ideal bias adjustments required to bring the bucket and ERI data into agreement. See Eq. 3 and accompanying discussion in the text. The time series are smoothed with a 12-month running mean to remove the seasonal cycle from the bias adjustments (see Folland and Parker (1995)).