The surface of the World Ocean warmed by ~0.75 K from 1900 to 2016, but the warming did not occur monotonically: temperatures increased during the first half of the twentieth 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 to 1970s are apparent in SSTs averaged over the globe and both the Atlantic and Pacific sectors (Figs. 1k–o, black time series; Figs. 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 to 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 (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 because of the transfer of heat from the superstructure of the ship as water passes through pipes, while bucket measurements are often cold biased because of 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 Met Office Hadley Centre developed SST datasets stratified into bucket and ERI measurements in conjunction with the release of their most recent gridded dataset, the Hadley Centre SST 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° × 5° 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

---

**Fig. 1.** Time series of sea surface temperature anomalies (K; relative to 1961–90) for the datasets and regions indicated. (a)–(e) Area averages of the unadjusted 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. (middle),(bottom) The shaded regions 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 centered 37-month running mean for display purposes. Note that the ERI, bucket, and HadSST3 data are collocated in space (see appendix).
the other measurement type were excluded. The latter step ensures that the bucket-only and ERI-only data are “collocated,” or have the same spatial coverage through time at the gridbox 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 do 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-twentieth-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.

**THE PROBLEM.** The unadjusted bucket and ERI data yield remarkably different renditions of twentieth-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,b). For example, the ERI data exhibit cooling of the Pacific Ocean from 1950 to 1975, whereas the unadjusted bucket data indicate warming (Figs. 1d,e, 2d,e, 3a,b). Likewise, the ERI data exhibit cooling in the global average over the same period, whereas the bucket data indicate warming (Figs. 1a, 2a).

![Fig. 2. Linear least squares trends in sea surface temperature anomalies from 1950 to 1975 for the datasets and regions indicated. (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. (middle), (bottom) Thin error bars 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 the supplemental material. Note that the COBE-SST2, NMAT, and ERSSTv4 data are collocated with the ERI, bucket, and HadSST3 data (see appendix). Note also that the NMAT data are only available in their adjusted form.](image-url)
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 0.1 K decade\(^{-1}\) cooling over the period 1950–75 (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\(^{-1}\) 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. ES1l,m in the supplemental material; https://doi.org/10.1175/BAMS-D-18-0104.2). 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 1950–75 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 1950–75 but the adjusted bucket data exhibit relatively little change in temperature (Figs. 1n,o, 2n,o, 3g,h). In contrast, over the Atlantic sector the adjusted ERI data exhibit significantly weaker cooling than the adjusted bucket data (Figs. 1l,m, 2l,m, 3g,h). These patterns of disagreement are stronger in the North Pacific and North Atlantic during boreal winter (Figs. S2l,m, S3g,h).

**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 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 time steps from 1950 to 1975 were available. Grid boxes are stippled where the (g),(h) trends and (i) trend differences have the same sign across all 100 bias adjustment realizations.
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. ES4g–i).

Importantly, the amplitudes of the trends in the adjusted ERI and bucket data over the 1950–75 period are comparable to the differences between them (cf. 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 (cf. 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 to 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 follows:

\[
\text{ERI}_{\text{unadjusted}} = \text{ERI}_{\text{true}} + \text{ERI}_{\text{true bias}} \quad \text{and} \quad \text{B}_{\text{unadjusted}} = \text{B}_{\text{true}} + \text{B}_{\text{true bias}} \]

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

**Fig. 4.** As in Fig. 3, but for the 1976–2006 period.
In general, over the 1950–75 period, the trends in the
data, which require their own bias adjustments
with nighttime marine air temperature (NMAT)
applied in ERSST.v4 are based only on comparisons
bias correction schemes: The bias
provide a point of comparison with SST data that em
ERSST.v4 and COBE-SST2 datasets are included to
version 4 of the Extended Reconstructed SST dataset
Estimates of SST (COBE-SST2; Hirahara et al. 2014)
archi: the Centennial In Situ Observation-Based
historical SST datasets based on the ICOADS2.5
indicated in Figs. 1–5 also affect the two other major
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 datasets based on the ICOADS2.5
archive: the Centennial In Situ Observation-Based
Estimates of SST (COBE-SST2; Hirahara et al. 2014)
developed by the Japan Meteorological Agency, and
version 4 of the Extended Reconstructed SST dataset
(ERSST.v4; Huang et al. 2015; Liu et al. 2015) released
by the U.S. National Climatic Data Center. The
ERSST.v4 and COBE-SST2 datasets are included to
provide a point of comparison with SST data that em
ploy 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 ERSST.v4 are based only on comparisons
with nighttime marine air temperature (NMAT)
data, which require their own bias adjustments
In general, over the 1950–75 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 ERSST.v4 data are somewhat
weaker than those in the HadSST3 data, particularly
over the Atlantic Ocean and in the global mean
(Figs. 2k–o; see appendix for details of the analysis).

SO WHAT? The results shown here reveal a level
of regional uncertainty in observed SSTs that is not
widely acknowledged in the climate dynamics litera-
ture. 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 ob-
served decadal variability itself. The uncertainty has
important implications for our understanding of the
role of aerosols in twentieth-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-twentieth century
(Kobayashi et al. 2015; Laloyaux et al. 2017; Taylor
et al. 2012; Plato et al. 2013). It also has important
implications for quantifying the amplitudes of pat-
terns of decadal-scale variability, particularly over the
problematic North Pacific sector (Figs. 1, 2) and in as-
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. Com-
plicking 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 are often missing from ICOADS (Carella et al.
2018), and the proportions of recruiting countries
can change substantially over time, especially before
~1970 (Fig. ES5; 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 measure-
ments used in each region, as implied by the differ-
ences in recruiting countries between the two sectors

\[
ERI_{\text{unadjusted}} - B_{\text{unadjusted}} = ERI_{\text{true bias}} - B_{\text{true bias}} \quad (3)
\]
(Fig. ES6). Not surprisingly, despite recent advances (Freeman et al. 2017; Carella et al. 2018; Hausfather et al. 2017; Cowtan et al. 2017; Hirahara et al. 2014), 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 analyzing 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 twentieth-century climate variability.

ACKNOWLEDGMENTS. We thank the three anonymous reviewers for their insightful comments, which greatly improved the paper. I.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 U.K. 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 nighttime 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. 2014) and National Climatic Data Center (ERSST.v4; Huang et al. 2015; Liu et al. 2015) 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 ERSST.v4 data were 1) regridded onto the 5° × 5° resolution HadSST3 grid and 2) had their respective monthly 1961–90 climatologies subtracted (to match the HadSST3 climatology period). The gridded NMAT,

**Fig. 5. Estimating the uncertainty in the bias adjustments for the indicated regions.** The orange lines represent the differences between the bucket and 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).
HadSST3, COBE-SST2, and ERSST.v4 data were matched to the spatial coverage of the collocated 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 1961–90 base period.

### REFERENCES


---

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

<table>
<thead>
<tr>
<th>Region</th>
<th>Boundaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Globe</td>
<td>60ºS–60ºN, 180ºW–180ºE</td>
</tr>
<tr>
<td>Atlantic Ocean</td>
<td>10º–60ºN, 80ºW–0º</td>
</tr>
<tr>
<td></td>
<td>60ºS–10ºN, 70ºW–20ºE</td>
</tr>
<tr>
<td>Pacific Ocean</td>
<td>10º–60ºN, 100ºE–100ºW</td>
</tr>
<tr>
<td></td>
<td>10º–10ºS, 150ºE–70ºW</td>
</tr>
<tr>
<td>North Atlantic</td>
<td>20º–60ºN, 80ºW–0º</td>
</tr>
<tr>
<td>North Pacific</td>
<td>20º–60ºN, 120ºE–100ºW</td>
</tr>
</tbody>
</table>
AMS Members

Give a great gift at a great price

Looking for the perfect present for the weather enthusiast in your life? Want to make a valuable contribution to your local library or community college?

Send a subscription to Weatherwise magazine for just $24.95*—That’s nearly 50% off the list price!

Written for a general audience, Weatherwise offers a colorful and nontechnical look at recent discoveries in meteorology and climatology.

Check out the latest table of contents at www.weatherwise.org.

Want your own?
Then order a personal subscription at the same great price.

Contact Member Services by e-mail at amsmem@ametsoc.org or by phone at 617-227-2425 to place all of your Weatherwise orders today!

*Cost for delivery outside of the U.S. is $40.95. Weatherwise is available to AMS Members through a cooperative agreement with Taylor & Francis Group LLC, the publishers of Weatherwise.