BAMS Article

# Choosing the Right Sea Surface Temperature Datasets: Benchmarking and Guidance for Climate Applications

Duo Chan,[a] Elizabeth C. Kent,[b] Nathan Lenssen,[c,d] Clara Deser,[d] Christopher J. Merchant,[e,f] Masayoshi Ishii,[g,h] Caroline Sandford,[i] Boyin Huang,[j] Xungang Yin,[j] John J. Kennedy,[k] Richard C. Cornes,[b] Peter Huybers,[l] and Geoffrey Gebbie[m]

**KEYWORDS:** Sea surface temperature; Climate change; Data processing/distribution; Databases

**ABSTRACT:** Sea surface temperature (SST) datasets underpin many climate applications, including climate monitoring, the study of historical climate variability and change, attribution, model evaluation, ecosystem assessment, paleoproxy calibration, and use as boundary conditions for atmospheric model simulations. Many different SST products are available. This paper addresses why SST products differ, what these differences mean for climate analyses, and which products are best suited for various purposes. Differences among SST products are first reviewed with respect to improvements in bias adjustments, gridding and infilling techniques, and uncertainty quantification. The implications of these advances are then assessed through historical case studies, evaluation of spatial patterns, and comparison of global means and key regional indices. Substantial discrepancies in trends are found during the satellite era using older SST products, but recently released datasets are much more consistent. Recent datasets also show a more-consistent SST evolution during World War II and in trends associated with tropical Pacific zonal gradients. Disagreements persist, however, with respect to early-twentieth-century warming and in data-sparse regions such as the Southern Ocean and Arctic. To assist users across disciplines, we articulate principles for dataset selection based on application needs and highlight the NCAR Climate Data Guide and an accompanying web-based data selector tool that provides updated benchmarking and access to SST products.

**SIGNIFICANCE STATEMENT:** Sea surface temperature is an "essential climate variable" used for tracking climate change, evaluating models, calibrating paleorecords, and understanding events such as marine heat waves and El Niño. Many different datasets exist, produced by various scientific groups. In addition, there are multiple versions of many of these datasets, yet older versions remain in use long after improved versions have superseded them. This article explains how SST datasets have developed and improved, shows how differences between them can influence scientific results, and highlights where recent versions agree and where important uncertainties persist. Alongside a general encouragement to use up-to-date SST products, we offer practical, application-focused guidance as well as an online tool that helps researchers identify, understand, and access SST datasets well suited to their needs, promoting proper, consistent use of sea surface temperature information.

DOI: 10.1175/BAMS-D-25-0210.1
*Corresponding author*: Duo Chan, duo.chan@soton.ac.uk
Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/BAMS-D-25-0210.s1.

Manuscript received 5 August 2025, in final form 27 July 2026, accepted 30 July 2026

© 2026 American Meteorological Society. This published article is licensed under the terms of the default AMS reuse license. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

**AFFILIATIONS:** [a] School of Ocean and Earth Science, University of Southampton, Southampton, United Kingdom; [b] National Oceanography Centre, Southampton, United Kingdom; [c] Applied Mathematics and Statistics, Colorado School of Mines, Golden, Colorado; [d] National Center for Atmospheric Research, Boulder, Colorado; [e] National Centre for Earth Observation, Reading, United Kingdom; [f] Department of Meteorology, University of Reading, Reading, United Kingdom; [g] Department of Climate and Geochemistry Research, Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan; [h] Geoenvironmental Sciences, University of Tsukuba, Tsukuba, Japan; [i] Met Office, Exeter, United Kingdom; [j] NOAA/National Centers for Environmental Information, Asheville, North Carolina; [k] Verdun, France; [l] Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts; [m] Department of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts

## 1. Introduction

Sea surface temperature (SST) is a critical variable in climate science, providing the primary measure of ocean surface warming and a key indicator for monitoring climate change and variability. It informs analyses of marine heat waves (Oliver et al. 2021), estimates of climate sensitivity (Sherwood et al. 2020), and attribution of observed changes to anthropogenic forcing (Eyring et al. 2023). SST also provides boundary conditions for atmospheric reanalyses (e.g., Hersbach et al. 2020; Kosaka et al. 2024) and atmospheric model simulations, e.g., the Atmospheric Model Intercomparison Project (AMIP; Eyring et al. 2016), and represents key modes of climate variability such as El Niño–Southern Oscillation (ENSO; McPhaden et al. 2006) and Atlantic multidecadal variability (AMV; Knight et al. 2006).

The SST datasets considered here are listed in Table 1 with acronyms defined and citation and access information. Each of these SST datasets generally targets one of three main user requirements (Fig. 1):

1) Long historical in situ records beginning before 1900, e.g., Hadley Centre SST dataset (HadSST), ERSST, and Centennial In Situ Observation-Based Estimates of SST (COBE-SST), intended for decadal-to-centennial-scale climate applications. These products often provide the oceanic component for global surface temperature datasets, and new major versions are released approximately every 6 years aligned with the Intergovernmental Panel on Climate Change (IPCC) assessment cycle.
2) High-resolution SST analyses, e.g., ESA Climate Change Initiative (CCI) SST and NOAA OISST, over the era of sustained satellite observations since 1980. These datasets utilize remotely sensed SST observations, may blend with in situ measurements, and provide data at high spatial and temporal resolutions. Several satellite-era analyses are also updated in near–real time for weather and climate prediction applications.
3) Centennial and multidecadal records that combine long-term in situ observations with satellite data, where available, to achieve extended temporal coverage and greater spatial completeness whenever possible, often for use in atmospheric reanalyses or as boundary conditions for atmosphere-only dynamical models. An example is HadISST.

The creation of an SST product generally requires four elements: 1) data selection (and, in the case of satellite-based products, retrieval of SST from radiances), 2) bias corrections to


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

**Table 1. SST datasets used in this paper.**

| Dataset | Dataset name | Citation | Available from |
|---|---|---|---|
| DCENT SST | Dynamically Consistent Ensemble of Temperature (-infilled) SST | Chan et al. (2024a) | https://doi.org/10.7910/DVN/NU4UGW |
| DCENT-I SST | | Chan et al. (2026) | https://doi.org/10.7910/DVN/ROG38Q |
| HadSST4.2 | Met Office Hadley Centre SST | Sandford and Rayner (2026) | https://www.metoffice.gov.uk/hadobs/hadsst4 |
| HadSST4.0 | | Kennedy et al. (2019) | https://www.metoffice.gov.uk/hadobs/hadsst4/previous_versions.html |
| HadSST3 | | Kennedy et al. (2011a,b) | https://www.metoffice.gov.uk/hadobs/hadsst3 |
| HadSST2 | | Rayner et al. (2006) | https://www.metoffice.gov.uk/hadobs/hadsst2 |
| ERSSTv6 | Extended Reconstructed SST | Huang et al. (2025) | https://www.ncei.noaa.gov/data/sea-surface-temperature-extended-reconstructed/v6/access/ |
| ERSSTv5 | | Huang et al. (2017) | https://www.ncei.noaa.gov/data/sea-surface-temperature-extended-reconstructed/v5/netcdf |
| ERSSTv4 | | Huang et al. (2015) | https://www.ncei.noaa.gov/data/sea-surface-temperature-extended-reconstructed/v4/netcdf |
| ERSSTv3b | | Smith et al. (2008) | https://www.ncei.noaa.gov/data/sea-surface-temperature-extended-reconstructed/v3b/netcdf |
| COBE-SST3 | Centennial In Situ Observation-Based Estimates | Ishii et al. (2025) | https://climate.mri-jma.go.jp/pub/archives/Ishii-et-al_COBE-SST3/cobe-sst3 |
| COBE-SST2 | | Hirahara et al. (2014) | https://climate.mri-jma.go.jp/pub/archives/Hirahara-et-al_COBE-SST2/ |
| COBE-SST1 | | Ishii et al. (2005) | https://ds.data.jma.go.jp/tcc/tcc/products/elnino/cobesst_doc.html |
| HadISST1 | Hadley Centre Sea Ice and SST dataset | Rayner et al. (2003) | https://www.metoffice.gov.uk/hadobs/hadisst/ |
| COBE-SST3H | Centennial Observation-Based Estimates | Ishii et al. (2025) | https://climate.mri-jma.go.jp/pub/archives/Ishii-et-al_COBE-SST3/cobe-sst3h |
| OISSTv2.1 | Optimum Interpolation SST | Huang et al. (2021) | https://www.ncei.noaa.gov/products/optimum-interpolation-sst |
| OISSTv2 | | Reynolds et al. (2007, 2002) | https://www.ncei.noaa.gov/data/sea-surface-temperature-optimum-interpolation/v2 |
| ESA CCI SST3.0 | European Space Agency Climate Change Initiative SST | Embury et al. (2024) | Easy access: https://surftemp.net; Full global resolution: https://data.ceda.ac.uk/neodc/eocis/data/global_and_regional/sea_surface_temperature/CDR_v3/Analysis |
| CMIP6 ensemble | Coupled Model Intercomparison Project phase 6 | Eyring et al. (2016), Abernathey et al. (2021) | https://console.cloud.google.com/storage/browser/cmip6 |

remove artifacts in measurements, 3) gridding and infilling to provide estimates in regions without direct measurements, and 4) derivation of an estimate of uncertainty for each value in the final product. Each element has improved over the years, leading to updated versions of the long-standing product families as well as newly developed datasets, e.g., Dynamically Consistent Ensemble of Temperature (DCENT) SST.

Despite these advances, uptake of newer SST datasets by the research community can be slow. As a result, some datasets that do not contain any bias adjustments, e.g., the gridded


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

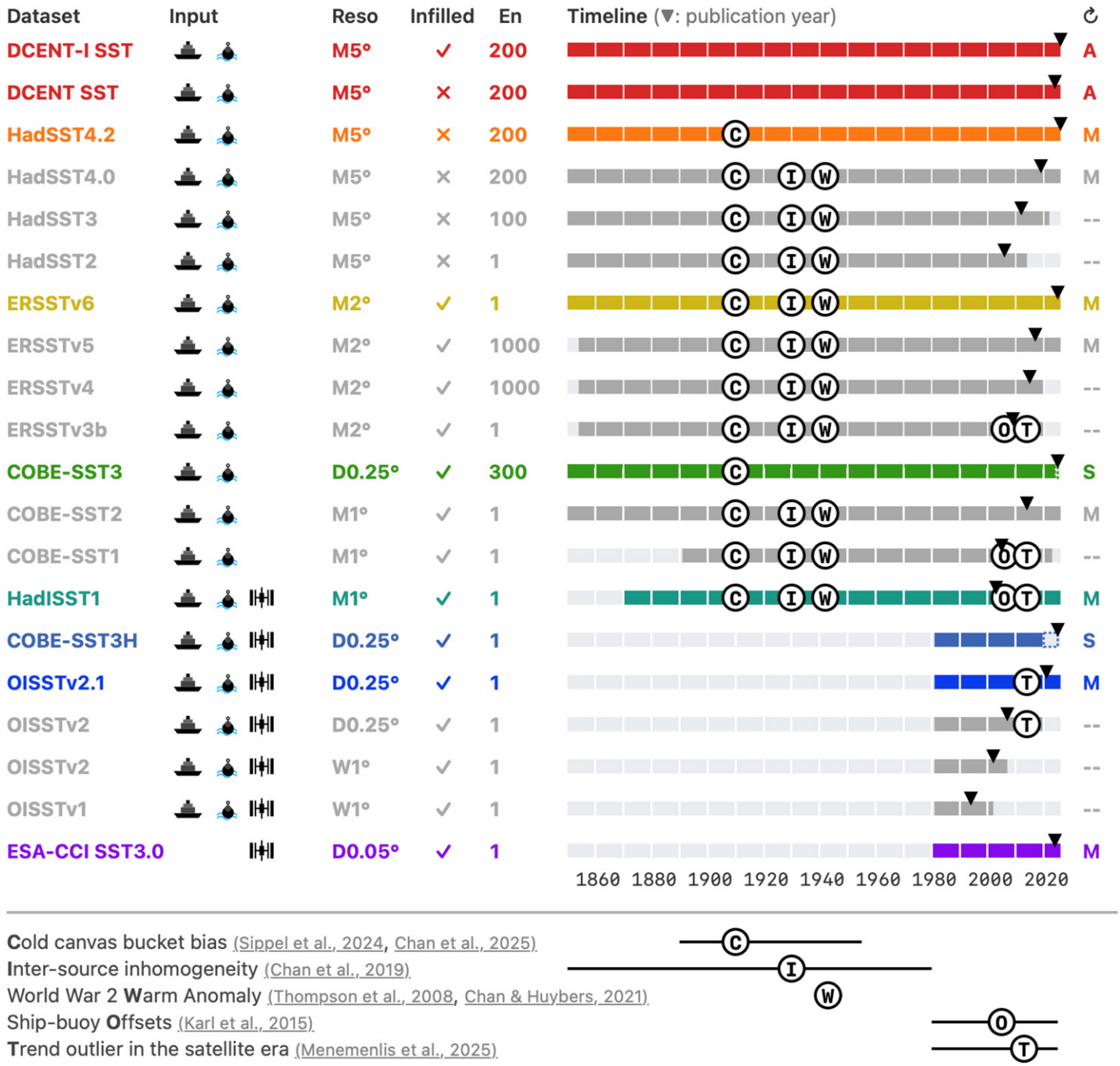


**Fig. 1. Overview of major SST dataset families: DCENT SST (red), HadSST (orange), ERSST (yellow), COBE-SST (green), HadISST (teal), COBE-SST3H (light blue), OISST (blue), and ESA CCI SST (purple). Datasets are grouped by family and ordered by version within each family. Gray indicates noncurrent dataset versions within each family. Columns indicate input data types (input: ship/buoy/satellite), nominal temporal (monthly/weekly/daily) and spatial (°) resolution (Reso), spatial completeness (infilled: ✓/×), ensemble size (En), temporal span (horizontal bar in timeline), publication year (downward black triangle in timeline), and update frequency [↻: annual/monthly/static/discontinued (--)]. Symbols mark major known biases and artifacts that remain in each product: cold canvas bucket bias (C), intersource inhomogeneity (I), WWII warm anomaly (W), ship–buoy offsets (O), and trend outliers in the satellite era (T), based on evaluations presented in sections 2 and 3. Indicative ranges for these issues are shown in the legend. Dataset abbreviations follow those in the text and are expanded in Table 1.**

summaries from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS; Freeman et al. 2017), or older products with outdated bias adjustments, e.g., the Kaplan SST (Kaplan et al. 1998), remain in common use years after release, e.g., for long-term trends in the tropical Pacific zonal SST gradient (Lee et al. 2022). Another example is HadISST1 (released 2003), which is one of few products not to correct for post-1950 ship-based SST biases, leading to lower warming rate estimates since 2000 (Karl et al. 2015) and, thus,


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

systematically different estimates of recent trends (Menemenlis et al. 2025). These are just a few of many examples of the mismatch between the SST products most widely used in research and those that best reflect current understanding of observational biases and uncertainties.

This lag in adoption reflects the reality of research where "switching costs" can be high. Familiarity often shapes dataset choice, while barriers such as nonstandard data formats, large data volumes, difficulty finding the data, and historically fragmented documentation create further friction. The landscape has undoubtedly improved in recent years with comprehensive documentation now available in data journals (see Table 1) and user guides (e.g., HadSST4 and ESA CCI SST). Although fully absorbing the technical details of multiple candidate datasets may not seem an obvious priority, we show in this paper that where scientific analyses depend critically on observational estimates of SST, selecting suitable products is essential for robust and high-quality research. Promoting these improved SST products is also timely as the climate community is determining standards for the upcoming IPCC CMIP7 and AR7 (Beadling et al. 2026).

This paper provides a starting point for SST users in navigating this evolving landscape, enabling them to more easily identify and consult relevant data papers and user guides for informed choices of SST products best suited for their particular application. Specifically, this paper addresses the questions: "Why do datasets differ?" by tracing the evolution of their development in section 2, "What do these differences mean for climate analyses?" by comparing products in section 3, and "How do I pick SST datasets?" by providing guidance on the current state of the art as well as anticipated improvements likely to affect future choices in section 4. Section 5 provides a summary.

Our analysis focuses on long-standing and recently developed SST dataset families that are updated regularly. We exclude older products whose methods have not been updated since before 2000 (e.g., Kaplan SST) or that lack bias adjustments (e.g., gridded ICOADS), due to their limited comparability with modern products. Several high-quality near-real-time analyses are omitted because they are either shorter than 40 years [e.g., the Multiscale Reanalysis by Chin et al. (2017) and the Canadian Meteorological Center analysis by Brasnett (2008)] or built on a significant input of ESA CCI SST data [e.g., the Operational SST and Sea Ice Analysis (OSTIA) reprocessing by Worsfold et al. (2024)]. Operational SST analyses that principally support numerical weather prediction are coordinated by the Group for High Resolution SST (GHRSST; www.ghrsst.org/), and intercomparisons of these datasets have been reported elsewhere (e.g., Fiedler et al. 2019; Yang et al. 2021).

We also exclude hybrid datasets and reanalysis-derived SST fields. This includes blends of different products made for reanalysis (e.g., Hersbach et al. 2020) and surface-forcing datasets for AMIP-style uncoupled simulations that combine HadISST1 and OISSTv2 (Hurrell et al. 2008). Atmospheric reanalyses, such as ERA5, are driven by externally specified SST boundary conditions that often draw on products already discussed here and therefore do not provide independent SST analyses. Ocean reanalyses and state estimates, such as Global Ocean Reanalysis and Simulation, 1/12° resolution, version 1 (GLORYS12V1) (Jean-Michel et al. 2021), and ECCO (Forget et al. 2015), produce essentially model-simulated SST fields that are continually corrected using multiple sources of observations, including satellite SST, subsurface temperature, sea level, sea ice, and salinity. These SST fields are therefore not derived from SST observations alone and are not directly comparable to the products reviewed here.

## 2. History of SST products and recent advances

This section reviews how three core elements of SST product development—bias adjustment, gridding and infilling, and uncertainty quantification—have evolved, with each stage discussed in its own subsection.


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

***a. Bias adjustment.*** Biases in SST records stem from pervasive and systematic errors that differ between measurement methods and platforms, their changing mix over time, and their past data curation and processing (Kent and Kennedy 2021). Ship-based observations made with buckets are typically cold biased because of evaporative cooling, and different bucket types used by various nations and periods left distinct bias signatures. On the other hand, engine-room-intake (ERI) measurements tend to be warm biased owing to heat from the vessel (Kent and Taylor 2006). These biases are often several tenths of a degree Celsius in magnitude and distort long-term trends, making their correction a central task in the development of climate-quality analyses.

Early adjustment efforts concentrated on pre-1940 bucket biases. An initial blanket adjustment (Folland et al. 1984) was followed by land-anchored estimates using coastal station temperatures (Jones et al. 1986) and, soon after, physics-based bucket models that simulated cooling as a function of bucket type and usage (Bottomley et al. 1990; Folland and Parker 1995). Because detailed metadata on bucket types and national practices are sparse, these schemes necessarily assumed simplified and broadly timed transitions, yielding limited regional differentiation, as implemented in, e.g., HadSST2. In parallel, ERSSTv3b pursued an anchoring strategy using nighttime marine air temperatures (Smith and Reynolds 2002; Kent et al. 2013), although adjustments were still only applied prior to 1940.

Thompson et al. (2008) identified a major bias in ERI temperatures that caused a spurious decrease in global-mean surface temperature of approximately 0.3°C following World War II. ERI measurements represent the majority of SST data available between 1930 and 1990 (Kent and Taylor 2006). Subsequent datasets (e.g., HadSST3 and COBE-SST2) extended bias corrections beyond 1940 to account for ERI biases as well as offsets between ship-based and buoy measurements. Time-varying offsets between ship-based and buoy measurements shown to affect post-2000 temperature trends (Karl et al. 2015) were accounted for starting with ERSSTv4, HadSST4.0, and COBE-SST2.

Since 2019, attention has expanded from method-specific biases to finer spatial and platform-dependent structures. HadSST4.0 used near-surface values from marine profile temperatures to estimate regional, ship-related biases after 1940. In parallel, Chan and Huybers (2019) developed an intercomparison framework that quantifies offsets among national groupings and enables pre-1940 comparisons. This framework has revealed a cold truncation bias in part of the Japanese data that contributed to the unusually heterogeneous early-twentieth-century warming pattern (Chan et al. 2019). This truncation bias has recently been adjusted in DCENT(-I) SST, COBE-SST3, and HadSST4.2 through different implementations.

The most recent identification of in situ bias is a global cold bias in decades around the 1910s (Chan et al. 2023; Sippel et al. 2024) that alters estimates of early warming and decadal variability and is attributed to incomplete correction of canvas bucket temperatures (Chan et al. 2025). To date, only DCENT SST and COBE-SST3 implement specific adjustments to account for this global cold bias by reviving the earlier land-anchoring idea (Jones et al. 1986).

Satellite SSTs are obtained from relatively few (~25) missions with differing bias characteristics (e.g., Yang et al. 2021; Fiedler et al. 2019). These platform-dependent effects are also on the order of several tenths of a degree Celsius (Merchant et al. 2008b). Satellite SST records have further required corrections for biases from atypical atmospheric conditions, particularly the stratospheric aerosol from the 1991 Pinatubo eruption (Reynolds 1993; Merchant et al. 1999).

The satellite-only ESA CCI SST is based on physics-based estimation approaches (Merchant et al. 2008a; Embury and Merchant 2012; Merchant et al. 2020) to minimize biases from changing satellite characteristics and from volcanic perturbations to the stratosphere. The local time of satellite overpasses has varied, and the artificial trends arising from changing observation times relative to the daily cycle of SST are also addressed in


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

ESA CCI SST through adjustments to a standard local time of observation. ESA CCI SST also explicitly adjusts the skin temperature observable from space to estimate the SST at 20-cm depth for compatibility with centennial-scale datasets using in situ data from drifting buoys and buckets.

To sum up, both in situ and satellite SST measurements are affected by systematic biases, and recognizing these biases and developing adjustments has taken time. The development of SST analyses has therefore progressed from broad, method-based corrections toward more spatially explicit and platform-specific treatments. This history of improved understanding of the observations over time, alongside advances in numerical techniques and computing capacity, is one key reason why SST products differ.

***b. Construction of gridded fields.*** Typically, individual observations are first assigned to the appropriate grid cell and temporal window (e.g., month/day) and averaged to estimate the mean SST within each cell–window combination. In some products, observations are first averaged at a finer intermediate resolution before aggregation to the final grid, which can reduce biases caused by uneven sampling within a grid cell (Rayner et al. 2006). An obvious difference between products is therefore the spatial and temporal grid resolution (here ranging from 5° monthly to 0.05° daily; Fig. 1). This choice is largely shaped by application needs tempered by data and processing limitations. For example, while monthly 5° or 2° products (such as HadSST and ERSST products) are usually sufficient for studying slowly varying climate backgrounds and large-scale modes of variability, much higher spatial and temporal resolution, such as is available from satellite-based records, is required for resolving extreme events like marine heat waves. Within a family, some products have trended toward finer resolution, as in OISST and COBE-SST (Fig. 1). For example, compared with its predecessor, COBE-SST3 provides a finer daily 0.25° resolution, designed to better capture mesoscale variations and steep meridional gradients that are important for understanding fine-scale air–sea interactions relevant to significant weather events.

A relevant concept is the distinction between nominal grid resolution (the stated grid spacing) and effective resolution (the smallest spatial scale that a dataset can reliably represent; Reynolds et al. 2013). A finer grid hence does not guarantee that smaller-scale physical variations are always resolved. This distinction is particularly important for products that blend in situ and satellite data while aiming to provide a consistent nominal resolution across more than a century. HadISST1, for instance, has an effective resolution of about 4° before 1949, reflecting the reduced-space reconstruction used at that time (Rayner et al. 2003). Some products address this issue by offering separate versions, such as COBE-SST3, which extends back to 1850 without satellite data, and COBE-SST3H, which incorporates satellite measurements but only from 1982 onward (Fig. 1). Fine-scale variability in COBE-SST3 before the satellite era, and especially before the widespread deployment of drifting buoys, can only be reliably estimated in areas neighboring observations; estimates elsewhere should be treated as useful probabilistic reconstructions. Users interested in fronts, coastal upwelling, western boundary currents, equatorial cold tongues and associated upwelling, or other sharp gradients should therefore consider not only the nominal grid spacing but also the period, sampling coverage, and effective resolution of each product.

After gridding data, a further application-oriented difference is how regions and times with no observations are handled. Noninfilled datasets such as the HadSST family leave those grid cells as missing and are often preferred for climate monitoring as they are closer to the original observations. In contrast, infilled products have values for those cells inferred using a variety of statistical methods and are generally more convenient to use, but their completeness weakens traceability to the original observations and requires assumptions about the variability.


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

Infilling methods typically define the expected relationship between conditions at different locations using a covariance matrix. The simplest choice of covariance between locations is isotropic and homogeneous, but more complex empirical relationships can be assumed to better capture regional covariance, as implemented in DCENT-I SST and high-resolution satellite-based products such as ESA CCI SST and COBE-SST3H. Other methods explicitly account for long-range teleconnections, including reduced-space optimal interpolation (Kaplan et al. 1997; e.g., in HadISST1), reconstructions based on empirical orthogonal functions (EOFs; Hirahara et al. 2014; e.g., in COBE-SST2 and 3), and empirical orthogonal teleconnections (EOTs; Smith et al. 1998; e.g., in ERSSTv3–v5).

Improvements in infilling across products in the same family can also be evident. For example, an increasing number of EOT modes have been used in successive ERSST versions to better capture localized variability. In its latest version (ERSSTv6), a three-layer fully connected neural network is used to replace EOT and has yielded better infilling skill (Huang et al. 2025).

***c. Uncertainty estimation.*** Quantifying uncertainty is essential for making appropriate use of SST datasets (Kennedy 2014). Most products provide uncertainty values per grid box and/or time step (e.g., ESA CCI SST, COBE-SST1–2), ensembles of plausible realizations (e.g., ERSSTv4–v5, DCENT(-I) SST, COBE-SST3), or both (e.g., HadSST3–4.2) for purposes of uncertainty quantification. Some older products (such as HadISST1) do not provide uncertainty estimates.

Ensembles are convenient for tracing how uncertainty propagates into climate analyses: A diagnostic is repeated for each member, and the across-member distribution defines confidence intervals (CI) consistent with observational error covariance. Ensembles can quantify complex error structures that cannot be handled analytically. Because individual members may contain more small-scale variability than the ensemble mean or median, variance statistics based on individual members can differ from those on the central measure alone. Moreover, for a given product, the across-member spread reflects only the uncertainty associated with choices internal to that product's particular methodology (known as "parametric uncertainty").

A more complete accounting of uncertainty must also reflect the full range of methodological choices in input data, quality control, bias adjustment, and reconstruction. This "structural uncertainty" is commonly approximated by the spread across independently developed SST datasets (Thorne et al. 2005), assuming they are diverse enough to span the plausible error range. However, many products share observational archives and methodological lineages, leading to common issues. For example, the SST datasets used in the last IPCC assessment all exhibited an early-twentieth-century cold anomaly (Sippel et al. 2024; represented here as the cold canvas bucket bias in Fig. 1), despite their apparent diversity. This cautions data users against treating interproduct agreement as evidence that structural uncertainty has been fully explored and highlights the need for genuine diversity in reconstruction approaches across the entire dataset development cycle.

These advances in bias adjustment and the construction of gridded fields, along with the addition of newly available historical data, provide SST products that better represent the historical evolution of SST than their predecessors, as illustrated in the next section.

## 3. Comparison of products

This section evaluates and compares SST datasets across a range of metrics to help data users determine which products reliably represent the phenomena and scales of variability relevant to their applications. Specifically, we investigate to what extent SST products exhibit bias signatures associated with known data artifacts (section 3a, Fig. 2); the spatial structures of events such as ENSO and marine heat waves (section 3b, Fig. 3); and climate features including long-term warming, major modes of variability, and important regional gradients (section 3c, Fig. 4).


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

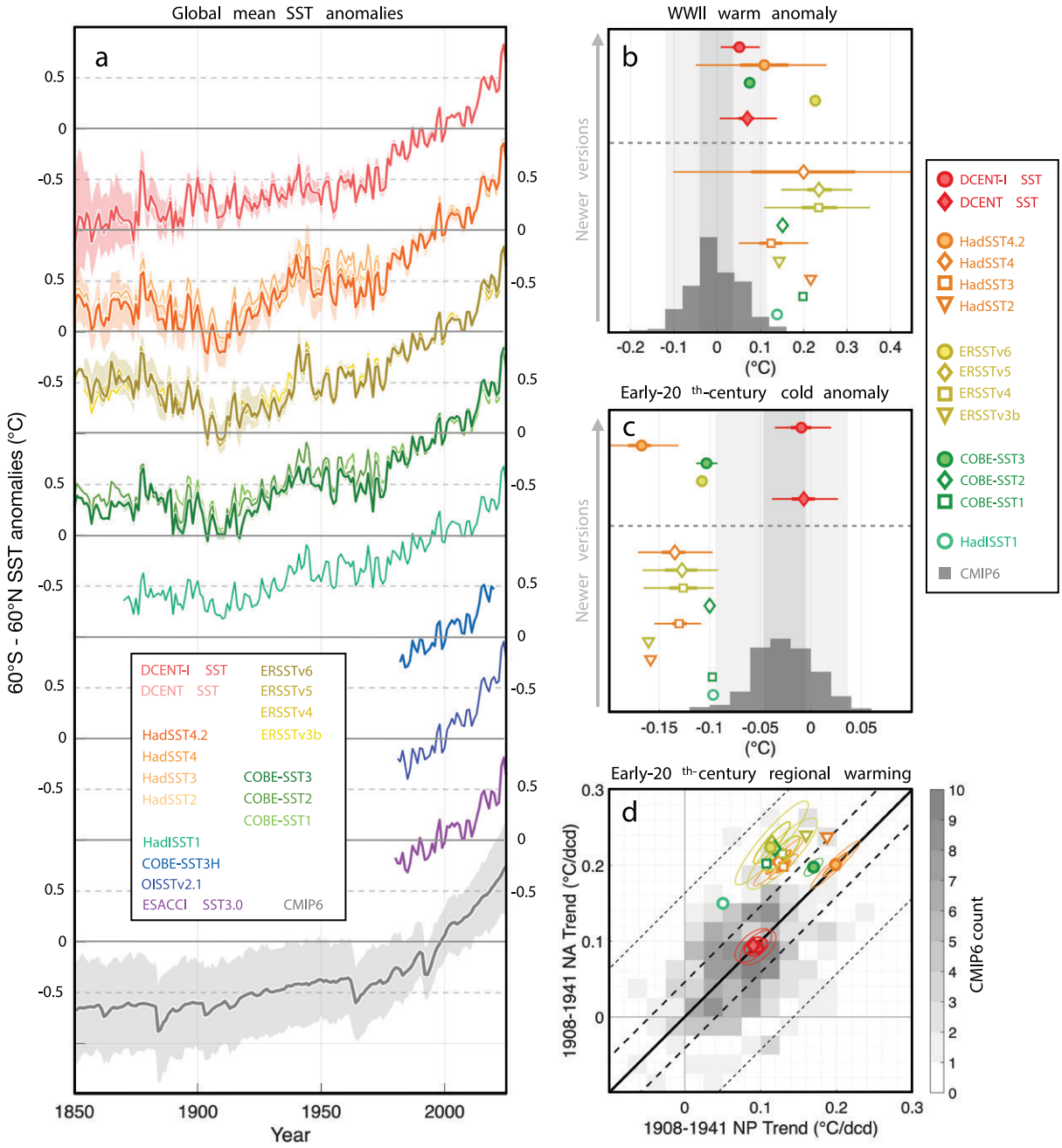


**Fig. 2. Comparison of global-mean SST and data artifacts. (a) Global-mean SST anomalies relative to 1982–2014 climatology. Datasets are grouped and offset by families. Here, "global" is defined as 60°S–60°N to focus on the ice-free ocean, where SST is more consistently defined and comparable across products. Also, only 10% of the global ocean surface is poleward of 60°N, 60°S. Within each family, thick color-coded lines show central estimates for individual versions, and the shading shows the 95% CI for the latest release where an ensemble is available. Simulations from 274 CMIP6 runs (see Table S1 for a detailed list), concatenating historical and shared socioeconomic pathway 2-4.5 (SSP2-4.5) experiments, are shown at the bottom. (b) WWII warm anomaly, calculated as the global-mean SST anomaly over 1941–45, relative to the mean over 1936–40 and 1946–50. Markers, sorted by publication date (descending) on the *y* axis, denote the mean value of a dataset, while thick and thin lines, respectively, denote the interquartile range and 95% CI, where an ensemble is available. The dashed line separates state-of-the-art and older products. The histogram presents the CMIP6 distribution, and the dark and light shading denote, respectively, the interquartile range and 95% CI (c) As in (b), but for early-twentieth-century cold anomaly, defined as the global-mean SST over 1900–30 minus a reference given by a linear trend fitted to the periods 1890–99 and 1931–40. (d) North Atlantic (*y* axis) vs North Pacific (*x* axis) SST trends over 1908–41. Markers are as in (b), and ellipses denote 1 and 2 standard deviation uncertainty using a bivariate Gaussian fit. Heat map squares represent the 2D histogram of CMIP6 historical simulations, the black line depicts the one-to-one relationship, and thick and thin dashed lines denote, respectively, the interquartile range and 95% CI of the simulated interbasin trend difference. CMIP6 values are based on complete coverage within 60°S–60°N and, thus, do not match coverage for noninfilled SST observational estimates.**


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

***a. Bias signatures.*** Figure 2a shows global-mean annual SST anomalies from each of the data products since 1850. The focus in this section is the pre-1960 evolution because this is where differences between SST products are most pronounced due to remaining data biases.

State-of-the-art CMIP6 simulations are shown for context. Compared with observational products, the CMIP6 ensemble mean shows a much smoother evolution of global-mean SST because averaging across many simulations largely cancels internally generated variability, such as ENSO, leaving primarily the response to external forcing. Differences between the CMIP6 ensemble and the observations, however, do not arise from internal variability alone but may also reflect biases in the models, their forcings, or the observational products (Chan 2021). In most contexts, such discrepancies are used to assess model performance. However, if observational biases are not independently identified and accounted for through careful analysis of the constituent data sources and their processing, model–data discrepancies may be misattributed to the models or lack of physical understanding. We illustrate such misleading evaluations below using three examples.

The first evaluation regards the World War II warm anomaly (Fig. 2b; Table 2) that is at least partly associated with wartime changes in measurement practice introducing warm biases (Thompson et al. 2008; Chan and Huybers 2021). Accounting for these issues, COBE-SST3, HadSST4.2, and the new DCENT SST family all show smaller warm anomalies than older products. This reduced warm anomaly is consistent with independent evidence from coral proxies (Pfeiffer et al. 2017) and with a large El Niño event identified in upper-air measurements spanning 1939–42 (Brönnimann et al. 2004). Correcting data, therefore, changes the interpretation of model–data differences during this period, which might otherwise be interpreted, for example, as indications of deficiencies in prescribed external forcings (Gottschalk 2017). ERSST is now the only major family of SST estimates in which a pronounced World War II warm anomaly persists.

Another example is the early-twentieth-century cold anomaly (Fig. 2c), during which both older and some recent products show pronounced negative anomalies. HadSST4.2 appears particularly cold because its 1930s adjustment of the Japanese truncation bias raises the 1931–40 endpoint of the reference fit (Table 2), making the 1900–30 anomaly

**Table 2. Definitions and calculation methods for metrics used in this study.**

| Metric name | How to calculate? |
|---|---|
| Global SST | Area-weighted (cosine latitude) mean over 60°S–60°N oceans, following Menemenlis et al. (2025) |
| Early-twentieth-century cold anomaly | Global-mean SST difference over 1900–30 relative to a linear fit between 1890–99 and 1931–40, following Sippel et al. (2024) |
| WWII warm anomaly | Global-mean SST anomaly averaged over 1941–45 relative to the mean of 1936–40 and 1946–50, following Chan and Huybers (2021) |
| North Pacific SST | Area-weighted mean over 20°–60°N, 100°E–100°W, following Chan et al. (2019) |
| North Atlantic SST | Area-weighted mean over 20°–60°N, 100°W–10°E (excluding Mediterranean), following Chan et al. (2019) |
| Early-twentieth-century warming | Linear trend of global-mean SST over 1908–41, following Chan et al. (2019) |
| Warming level since the preindustrial baseline | Difference in global-mean SST over 2015–24 relative to 1850–1900 |
| Niño-3.4 SST | Area-weighted mean over 5°S–5°N, 170°–120°W, following Trenberth (1997) |
| West equatorial Pacific SST | Area-weighted mean over 5°S–5°N, 110°E–180°, following Watanabe et al. (2021) |
| East equatorial Pacific SST | Area-weighted mean over 5°S–5°N, 180°–80°W, following Watanabe et al. (2021) |
| Southern Ocean | Area-weighted mean over 50°–70°S, following Dong et al. (2023) |
| AMV index | The difference between 20-yr running smoothed monthly North Atlantic SST anomalies (0°–60°N, 80°W–0°) and global SST, following Trenberth and Shea (2006) |


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

appear deeper. As described in section 2a, the cold anomaly likely reflects a residual bias arising from incomplete correction of canvas bucket measurements (Sippel et al. 2024). DCENT SST uses an adjustment anchored to land temperatures (Chan et al. 2024a,b) and shows SST evolution more consistent with collocated coral proxies (Chan et al. 2025). DCENT and DCENT-I SST accordingly show the smallest cold anomaly, reflecting a more moderate but consistent warming from 1850 to 1940. This suggests that the model–data mismatch in earlier attribution studies of this period, in which climate models could explain only about half of the observed temperature trend (Hegerl et al. 2018), may partly reflect residual bias in the SST products used rather than deficiencies in the models or prescribed forcings alone.

On regional scales, correcting the Japanese truncation bias directly alters the contrast in early-twentieth-century warming between the North Pacific and North Atlantic (Chan et al. 2019). For older products, all families show the North Atlantic warming nearly twice as fast as the North Pacific over 1908–41 (Fig. 2d), a contrast that could only be explained by an unusually large expression of internal variability (Delworth and Knutson 2000). In the latest versions, HadSST4.2, COBE-SST3, and DCENT SST correct for this bias, bringing the interbasin warming rates much closer to each other. ERSSTv6, by contrast, still exhibits a pronounced interbasin difference, similar to earlier ERSST releases. Data bias may, therefore, have contributed to the previously inferred interbasin contrast, with implications for assessment of internal variability in models.

Overall, the incorporation of adjustments for newly identified artifacts into SST data production has been gradual. For this reason, strong interproduct agreement may reflect the application of more accurate adjustments over time, but it is also possible to have good agreement when inaccuracies are shared. Interproduct disagreement can occur when some products, but not all, have been corrected for previously shared artifacts (such as the early twentieth-century cold anomaly). Products showing significant differences should, therefore, not automatically be treated as a set of equally plausible alternatives, especially without considering the observational artifacts involved and the methodological choices made to reduce them. Furthermore, evolving interproduct agreement also shows that when models and data disagree, the discrepancies cannot be assumed to arise solely from the models or prescribed forcing. Observational products, especially century-long historical products, should not be treated as fixed benchmarks against which successive generations of climate models are evaluated; using SST products with more complete bias corrections can materially alter the interpretation of model–data differences.

***b. Gridding and infilling.*** The different choices in the reconstruction of gridded products, including resolution and infilling, are important to consider for some applications. When studying historical events with sparse observations, spatial infilling and smoothing can make analyses more convenient, but the resulting fields are highly dependent on the assumptions used to generate complete fields. Taking the 1877 El Niño as an example, only a few ship tracks crossing the equator exist in the Pacific basin, as shown in the noninfilled product HadSST4.2 [Fig. 3a(1)]. Infilled products using isotropic, homogeneous covariance structures, e.g., Berkeley Earth surface temperature[1] (Rohde and Hausfather 2020), produce patterns consistent with their round kernels (Fig. 3b). By contrast, more advanced approaches, including anisotropic kernels [DCENT-I SST, Fig. 3a(2)], artificial intelligence (AI)-based methods [ERSSTv6, Fig. 3a(3)], and EOF-based reconstructions [HadISST1, Fig. 3a(4); COBE-SST3, Fig. 3a(5)], yield more coherent El Niño structures resembling the canonical pattern seen in the satellite

[1] Note that the Berkeley product SST is an infilled version of HadSST4.0 and is shown here to illustrate this effect. As a combined land–sea dataset, it is not used elsewhere in this SST-focused review.


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

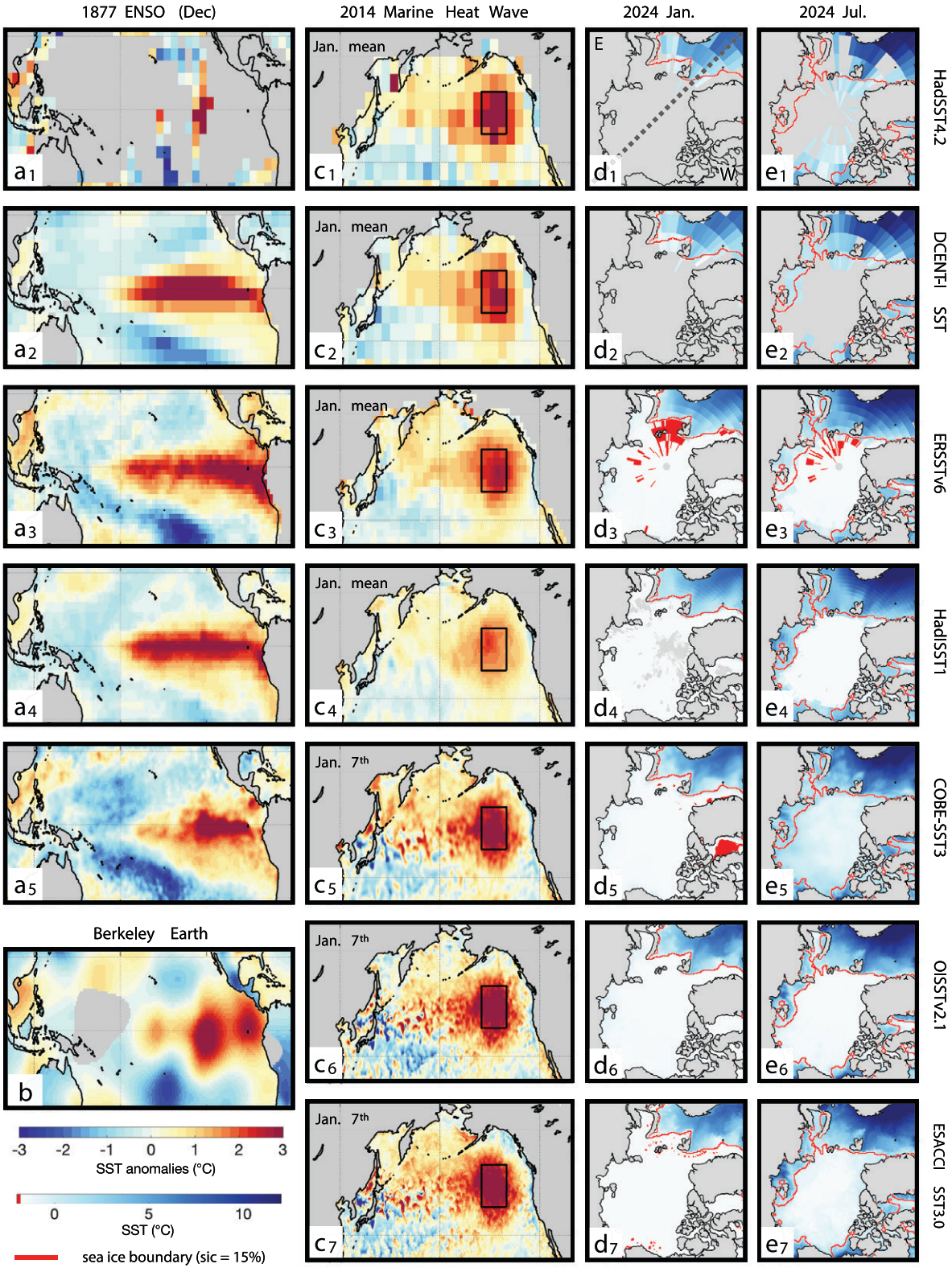


FIG. 3. Comparison of spatial patterns. (a) December 1877 SST anomalies relative to the December 1982–2014 climatology. (from top to bottom) HadSST4.2, DCENT-I SST, ERSSTv6, HadISST1, and COBE-SST3. (b) As in (a), but for the Berkeley Earth surface temperature. (c) January 2014 SST anomalies relative to the January climatology over 1982–2014, highlighting the marine heat wave termed "the Blob" in the North Pacific for datasets as in (a) but also including OISST and ESA CCI SSTv3. For datasets with daily resolution (COBE-SST3, OISSTv2.1, and ESA CCI SSTv3), the maps correspond to 7 Jan 2014, when the event peaked. The black box (140°–155°W, 38°–50°N) marks the region used to calculate the intensity of this event. (d),(e) SSTs (not anomalies) over the Arctic in January and July 2024, respectively. Red curves mark the sea ice edge (≥15% SIC; Cavalieri et al. 2012), gray areas indicate missing values, and red regions denote SSTs below the –1.8°C. In [d(1)], the diagonal dashed line marks the 0° and 180° meridians with the Eastern (E) and Western (W) Hemispheres labeled.


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

era. Fine-scale structure still differs between products as the fields are only tightly constrained near where observations are available. For example, the positive anomaly extends further west in DCENT-I SST and ERSSTv6 than in COBE-SST3, illustrating uncertainty associated with infilling across the datasets.

For contemporary extreme events such as marine heat waves, data availability is not the limiting factor. Rather, the requirement is to resolve fine spatial and temporal scales. Taking the North Pacific "Blob" of January 2014 as an example, all products—including the non-infilled monthly 5° HadSST4.2—show a similar warm anomaly centered near 45°N, 145°W (Fig. 3c). However, monthly fields blur the peak intensity evident in daily analyses. Over a box spanning 140°–155°W, 38°–50°N (black box in Fig. 3c), the mean SST anomaly in January is 2.5°C in DCENT-I SST, 2.4°C in ERSSTv6, 2.9°C in HadSST4.2, and only 1.7°C in HadISST1. Daily products resolve a higher temperature on the peak date, 7 January, corresponding to 3.0°C for COBE-SST3 and 3.1°C for OISSTv2.1 and ESA CCI SST. The daily high-resolution fields in Fig. 3 also show eddy-scale variability and fine filaments, which may be important for understanding the evolution and mechanisms of marine heat waves and their ecosystem impacts (Bian et al. 2023).

Another example of reconstruction differences arises in polar regions, where the open ocean meets sea ice. Due to sparse in situ coverage in polar regions, some products (e.g., DCENT-I SST) omit SST values in grid cells with no open ocean values (Chan et al. 2026). Others, such as the COBE-SST family, use observationally derived sea ice concentration (SIC) with an empirical SIC–SST relationship that anchors SST to a spatially varying freezing point under high SIC (Hirahara et al. 2014). Satellite products such as OISST (Huang et al. 2021) and ESA CCI SST (Embury et al. 2024) adopt similar concepts, using product-specific freezing-point constraints in ice-covered grid cells.

Figure 3d compares absolute Arctic SSTs in January 2024. Infilled products broadly follow the observed ice edge, though ERSSTv6 and COBE-SST3 exhibit below freezing point temperatures within ice-covered regions. Such behavior may not matter for climate analyses where sea ice–covered regions are masked. However, in AMIP simulations, the atmospheric model sees a weighted average of water and sea ice boundary conditions within each atmospheric grid cell. Hence, physically incompatible SST and SIC fields should be used with caution for such applications. Arctic summertime SST estimates in July 2024 show larger differences than in January (Fig. 3e) in both open-ocean regions such as the Laptev–East Siberian Sea (at left of panels) and areas with partial ice cover, indicating that model runs using summertime boundary conditions could be especially sensitive to dataset choice in regions with sea ice.

***c. Climate indicators of variability and change.*** Beyond grid-level maps, widely used climate indicators such as global warming levels, regional trends, and metrics of climate variability could depend strongly on the choice of SST products. Here, we examine several such indicators to demonstrate this sensitivity.

Estimates of long-term warming in global-mean SST (60°S–60°N) remain sensitive to product-specific treatments of nineteenth- and early-twentieth-century biases (Fig. 4a, sections 2a and 3a), with newer products only slightly narrowing the range. Here, we quantify warming using the 2015–24 mean relative to 1850–1900 because this directly measures the present-day state relative to the IPCC preindustrial baseline, while avoiding the assumption that warming evolves linearly over the full record. On this basis, excluding older datasets narrows the estimated SST warming range from 0.7°–1.0°C to 0.8°–1.0°C. This qualitative conclusion is unchanged when warming is instead quantified using the 1870–2024 trend (Fig. S1a in the online supplemental material), 2020–24 mean relative to the 1850–1900 baseline (Fig. S1b), or trends computed from varying start years (Fig. S1c). There is also evidence


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

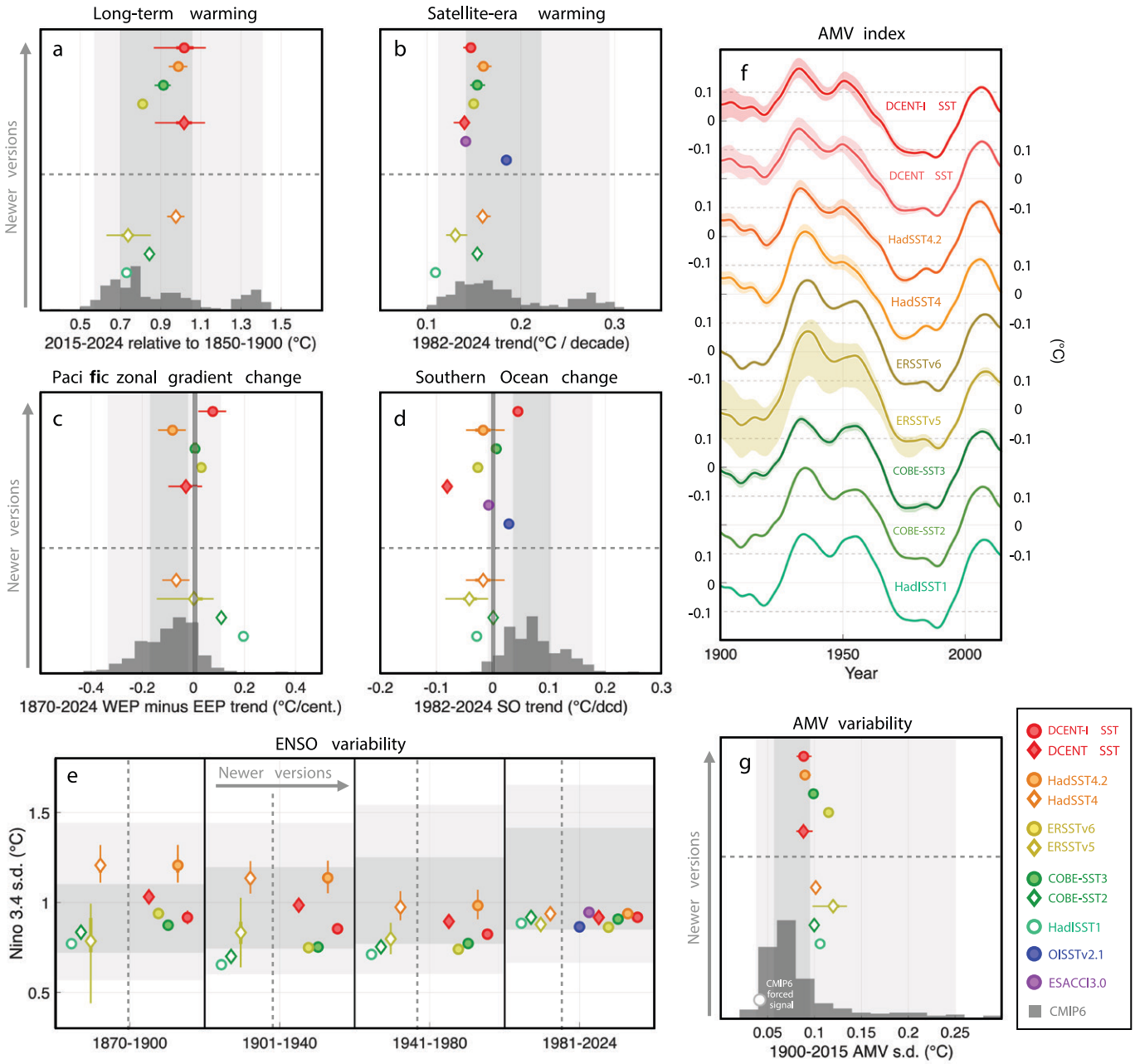


**Fig. 4. Comparison of key climate indices. (a) Mean SST (60°S–60°N) for 2015–24 relative to an 1850–1900 baseline, following the IPCC convention (Allan et al. 2023). Markers are sorted by publication date in descending order, with the dashed line separating state-of-the-art and older products. Thick and thin lines, respectively, give the interquartile range and 95% CI where an ensemble is available, though they are smaller than the symbol size in some cases. Only products extending to at least 2024 are included. The gray histogram shows the CMIP6 distribution, and dark and light shading denote, respectively, the CMIP6 interquartile range and 95% CI (b) As in (a), but for the 1982–2024 linear trend; satellite products are also shown. (c) The 1870–2024 linear trend in the tropical Pacific zonal temperature gradient, defined as the difference between a western box (5°S–5°N, 110°E–180°) and an eastern box (5°S–5°N, 180°–80°W; Table 2). All products covering 1900 to 2024 are shown; the zero line is also highlighted. (d) The 1982–2024 linear trend of SSTs over the Southern Ocean (50°–70°S). (e) Standard deviation of monthly mean SST anomalies in the Niño-3.4 region (5°S–5°N, 170°–120°W), computed within each of four periods (1870–1900, 1901–40, 1941–80, and 1981–2024) after removing the seasonal cycle and linear trend. Products are sorted by publication year from left to right. (f) Twenty-year smoothed monthly AMV index for recently updated century-long products. The series are vertically offset for clarity. (g) Standard deviation of the smoothed AMV index in (f). The white dot marks the standard deviation of the multimodel mean, indicating the amplitude of the forced signal.**

that the ERSST family, which features pronounced cooling over 1850–1910 (Fig. 2a), is likely too warm at the start of that period (Sippel et al. 2024; Chan et al. 2025), leaving scope for further narrowing of the observational range.


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

During the well-sampled satellite era, observational products are expected to agree more closely. Yet, when comparing older and modern datasets, Menemenlis et al. (2025) found a wide spread in the warming trend over the satellite era, reproduced here (Fig. 4b). Restricting this comparison to state-of-the-art products tightens the range among central estimates from 0.11°–0.18°C decade$^{-1}$ to 0.14°–0.18°C decade$^{-1}$. This narrowing is qualitatively robust to the choice of trend window, as shown by sensitivity tests varying the start year (Fig. S1c). Within this group, DCENT SST, ERSSTv6, COBE-SST3, and ESA CCI SST cluster near 0.14°–0.15°C decade$^{-1}$, with HadSST4.2 being slightly higher at 0.16°C decade$^{-1}$. NOAA's daily OISSTv2.1 is the clear remaining outlier (0.18°C decade$^{-1}$), perhaps due to its fixed 0.14°C ship-to-buoy correction from 1981 to 2015. Hence, the observational spread in satellite-era SST warming is narrower than previous comparisons that included older datasets.

On regional scales, an important indicator is the equatorial Pacific zonal SST gradient, which influences circulation, clouds, albedo, and climate sensitivity (Kang et al. 2023). Older products such as HadISST1 and COBE-SST1/2 show a pronounced century-scale strengthening of the zonal SST gradient (Deser et al. 2010), whereas this trend is weaker in newer products, a result that is qualitatively unchanged when the trend start year is varied or different century-scale windows are considered (Fig. 4c; Fig. S2). This weaker century-long enhancement in zonal SST gradient is more consistent with independent data including the sea level pressure gradient and precipitation trend pattern (Deser et al. 2010). Note that these results should not be read as implying improved model–data agreement across all trend windows, as attributing model–data discrepancies exclusively to limitations in older SST products, or as establishing model fidelity. For example, mean-state biases, especially an overly cold equatorial cold tongue, in CMIP6 models can induce excessive eastern Pacific warming and hence an erroneous weakening of the zonal SST gradient (Seager et al. 2019). These findings motivate more detailed future work investigating model–data agreement of tropical Pacific trends using state-of-the-art SST products and, more broadly, highlight how model–data comparisons are sensitive to the choice of SST products.

Southern Ocean SST is another key regional indicator, relevant to Antarctic ice-sheet melt (Dong et al. 2022), heat uptake (Gregory et al. 2024), and the impact of Antarctic stratospheric ozone depletion (Schneider et al. 2015; Seviour et al. 2019). CMIP6 models generally suggest warming (from −0.01° to 0.18°C decade$^{-1}$) over 1982–2024, but observational products show trends closer to zero, with considerable spread (Fig. 4d). Spatial patterns also differ across products, particularly in the magnitude and extent of the cooling band (Fig. S4). Given the sparse in situ sampling in this region, satellite-based products are likely the most reliable for recent Southern Ocean assessments, which further suggests that models may be warming too strongly in recent years.

It is also informative to consider modes of climate variability, particularly ENSO. During the satellite era, observational products consistently show Niño-3.4 variability of 0.85°–0.95°C (1 standard deviation; Fig. 4e). Earlier in the record, however, observational estimates diverge. This divergence could arise from increased sampling and measurement uncertainty, as well as structural differences in interpolation methods. Combined with the intrinsic difficulty of estimating ENSO variance reliably from 30- to 50-yr windows (Wittenberg 2009; Deser et al. 2012), these factors suggest that current SST datasets are unlikely to provide a reliable estimate of long-term changes in ENSO variability. CMIP6 models, which are subject to their own uncertainties and biases, on average suggest an increase over time (Fig. 4e).

Finally, decadal modes of variability such as Atlantic multidecadal variability show broad consistency in phase among products (Fig. 4f) but differ slightly in amplitude (Fig. 4g). These amplitude differences are mainly family specific and vary little across versions within a family. Compared with CMIP6 models, observational amplitudes tend to lie toward the higher end


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

of the model spread, though they remain within the range sampled by individual simulation members.

Overall, state-of-the-art SST datasets now show better agreement with each other than their predecessors across a range of metrics. Differences are larger in long-term trends and in data-sparse regions, but they generally agree on global warming levels and major modes of variability over the satellite era. For climate model evaluation, the result can be sensitive to the choice of observational products, and some model–observation discrepancies reported in the literature (e.g., Chan and Huybers 2021; Sippel et al. 2024; Byrne et al. 2025, as in Figs. 2 and 4) were amplified by biases in older SST products and are less pronounced when assessed using current datasets. Together, these results underscore the importance of being aware of how SST datasets have evolved and adopting up-to-date, well-documented releases matched to the intended analysis.

## 4. How to choose

***a. Principles underpinning dataset choice and usage.*** We have shown that careful dataset choice is crucial for high-quality and robust analyses. Considerations that may influence dataset choice include the length of record, whether fields are spatially complete, spatial and temporal resolution, availability and type of uncertainty estimates, and how recently datasets are updated. All datasets are free for research use, but some have restrictions on commercial use. The web-based selector tool (https://climatedataguide.ucar.edu/climate-data/sst-data-sets-overview-comparison-table; similar to Fig. 1) enables users to quickly view, subset, and access candidate datasets suitable for specific applications. When several viable products exist, results will be more robust if all are used.

Typical dataset choices by application include the following:

- Climate monitoring: compare noninfilled and infilled datasets at monthly or higher resolution.
- AMIP forcing: use infilled datasets at monthly or higher resolution.
- Attribution or model–data comparisons: use ensemble datasets (either infilled or noninfilled) with uncertainty estimates. For noninfilled products, apply the same observational coverage mask to the model output to ensure a fair comparison.
- Western boundary currents, mesoscale eddy signatures, and marine heat waves: use infilled high-resolution datasets (daily, finer than 1° × 1°).
- Paleoproxy calibration: use long instrumental records without known issues during the calibration period and compare noninfilled and infilled products for consistency. The SST datasets reviewed here are instrumental only; estimates that synthesize instruments and proxies can be useful (Tingley and Huybers 2010), but retaining independence during calibration aids intercomparison.

Once candidates are thinned by practical considerations, it is necessary to assess data quality. The analysis in section 3 shows the importance of using the most recent dataset version within each family, with older versions used only alongside their updated counterparts to help interpret past analyses. Even the most recent releases, however, can retain period- or region-specific issues, as described in section 2a and shown in section 3. Datasets with known problems that could affect the analysis should ideally be avoided. If excluding them leaves too few candidates for a robust assessment, they can be used with caution, provided their limitations are clearly acknowledged.

A useful strategy to discriminate among SST products is to evaluate physical consistency with other quantities such as air temperature, sea level pressure, precipitation, and cloudiness (Deser et al. 2010). Yet, this requires understanding how the datasets are constructed


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

and the assumptions involved. For example, bias adjustments in the ERSST family assume a relatively stable difference between SST and nighttime marine air temperature (Smith and Reynolds 2002), such that agreement is prescribed. Similarly, DCENT SST is adjusted to be dynamically consistent with its land counterpart, DCENT land surface air temperature (LSAT) (Chan et al. 2023, 2024a). Another often neglected assumption concerns the spatial covariance embedded in infilling. When records, especially in data-sparse periods, are infilled by projecting onto prescribed EOF patterns, subsequent EOF analyses will only largely recover the imposed covariance structure.

In addition to comparing analysis results across qualifying datasets, uncertainty within each product should also be evaluated to assess robustness. This can be done by analyzing the ensemble or, if no ensemble is available, by perturbing the data using the product's uncertainty estimates. For any metric for which the uncertainty ranges calculated from different data products do not overlap, the results should be interpreted with caution. If practical constraints require using only a subset of an ensemble—such as when running high-resolution AMIP experiments (Chan et al. 2021)—it is important to understand how the ensemble was constructed so the subset still represents the intended uncertainty. In HadSST4, for example, different parts of the 200-member ensemble use different approaches to adjust early SST measurements (Kennedy et al. 2019), so drawing members from across the ensemble gives a more representative sample than simply using the first few members.

Finally, follow data-citation guidelines from journals and dataset producers, citing both the dataset and its associated publication. This ensures producers can track how their datasets are used, which helps maintain the pipeline of high-quality, regularly updated SST products.

***b. Where to find more detailed information and updated advice.*** This paper provides a broad overview of how SST datasets have evolved and their suitability for different climate applications. By design, it cannot cover the full details of individual products and will freeze at the time of publication. To support users beyond this snapshot synthesis, we additionally refer the reader to a set of NSF NCAR Climate Data Guide pages (https://climatedataguide.ucar.edu/; Schneider et al. 2013) that extend guidance in two complementary ways.

First, dataset-specific pages provide summary information on dataset construction, strengths, and known limitations that are more detailed than this synthesis. Written by the developers or expert users and reviewed by leading climate scientists, these resources, accessible through the web-based selector tool (https://climatedataguide.ucar.edu/climate-data/sst-data-sets-overview-comparison-table), help users efficiently evaluate whether any features or issues are critical for their intended analysis. Once candidate datasets have been identified, there is no substitute for a deep dive into the linked dataset papers and product user guides for more detailed usage notes and guidance.

Second, an SST overview page (https://climatedataguide.ucar.edu/climate-data/sst-data-sets-overview-comparison-table) will be updated to provide an evolving summary of the SST dataset and evaluation landscape. By tracking newly released updates, methodological advances, and emerging developments, this page helps ensure that choice and usage guidance remains accurate and relevant as new and improved datasets become available.

***c. Anticipated improvements in SST datasets.***

**1) Better input data and metadata.** Ongoing efforts to rescue historical data (Teleti et al. 2024) and metadata (Carella et al. 2017) provide valuable independent evidence for evaluating historical SST biases including those relevant to the early-twentieth-century cool period and the World War II (WWII) warm anomaly. This may be further accelerated by AI-assisted


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

digitization and metadata extraction (Singh and Middleton 2025). Newly recovered data may also offer opportunities to test the skill of infilling methods in previously data-sparse regions and periods. Equally importantly, modern data infrastructure is needed to ensure that rescued and contemporary observations as well as metadata flow efficiently and transparently into permanent archives and SST dataset production—a gap that currently prevents many recovered measurements from being fully used, including by most of the SST products discussed here. For satellite-era products, fundamental work on the calibration of sensors and SST retrieval methods, particularly in the 1980s and 1990s, will also reduce uncertainty and improve stability.

**2) Better coverage and finer resolution.** Advances in infilling methodology, including AI approaches, together with increased computational capacity will support higher spatial and temporal resolution, with monthly products at 1° or finer as a practical baseline, and 5-day, daily, or subdaily products where supported by the observations and practical data-volume constraints. As these developments progress, the hard boundaries between datasets designed for different purposes are likely to soften, as already seen in COBE-SST3. Feature resolution of satellite-era products pre-2000 should benefit from efforts via international cooperation, such as those spearheaded by ESA (https://ceos.org/news/avhrr-data-recovery), to consolidate full-resolution data from early sensors. These higher-resolution observations have not yet been used in global SST analyses and provide an opportunity for better understanding changes in ecologically important shelf sea regions.

**3) Better bias adjustments.** Beyond the pervasive global- and basin-scale biases discussed in section 2a, progress will require pushing bias estimation further toward ship-specific and hence regional scales, to be enabled by improved metadata such as ship tracking and advanced physical and statistical models. Meanwhile, broader evaluation using independent sources will be essential for assessing and refining bias adjustments. Improved methods are also emerging to improve observational stability in satellite SST records by extending retrieval methods to be "bias aware" (Merchant et al. 2020) and by harmonizing radiance between satellite platforms prior to retrieval of SST.

**4) Better structural uncertainty estimates.** As discussed in section 2c, the current practice of pooling different SST products together does not necessarily provide a comprehensive estimate of structural uncertainty. As understanding of data artifacts improves, clearly inconsistent products are recommended to be excluded from certain analyses. This not only strengthens confidence in the analysis and metrics of interest but also narrows the ensemble and reduces its potential to span the full space of uncertainties that arise across the entire SST production workflow. A more complete characterization will require decomposing that workflow into its major components—input selection, quality control, bias adjustment, gridding, and infilling—and sampling alternative methods and parameter choices within each step. Such a modular approach would help dataset providers explore the widest range of reasonable choices across each component and ensure that all known errors and uncertainties are accounted for and fed through to the infilling schemes. This approach would also lower the barrier for new contributions as novel approaches could be developed for a single component. For satellite datasets, which are constructed from many trillions of radiance measurements, modular exploration of structural uncertainty is challenging in terms of scale and expense because of the large volume, and the community continues to focus on metrological approaches to exposing, quantifying, and correcting effects leading to uncertainty in SST products (Mittaz et al. 2019).


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

**5) User-friendly access and data formats.** In situ SST datasets are currently dispersed across data centers, which complicates comparison and analysis. Moving toward common conventions for both input observations and products, CMIP-style access protocols, regridding and subsetting services such as surftemp.net, and cloud-native formats (e.g., zarr) will further lower these barriers and support more scalable, interoperable use of SST products. For satellite SSTs, products have long been standardized in format, provided with tools, and cataloged through cooperative efforts of the GHRSST international science team (https://www.ghrsst.org/).

**6) Faster-paced innovation.** Delivering the improvements outlined above will require open, standardized, flexible, and streamlined systems that span data intake, processing, and distribution. Such infrastructure would better connect data producers and users, broaden participation in development and evaluation, and ultimately enable users to move from passive recipients of SST products to active participants in improving both the data and the science derived from them.

## 5. Final words

Careful choice of SST datasets is essential for high-quality research. SST datasets have improved in quality over time, and their estimates of important measures of variability have become more consistent. Characterization of dataset uncertainty has also improved, allowing users to assess the sensitivity of their results both within and across datasets. Several key indicators—including recent and centennial global trends—show that the most recent SST versions align more closely with one another than with older products. Observational constraints on projected surface temperature changes are therefore more reliable when using state-of-the-art datasets than older products would suggest.

The above motivates a set of practical steps for SST dataset selection and use:

1) Use the web-based selector tool to identify datasets appropriate for the intended application, taking into account record length, residual biases, spatial and temporal resolution, completeness, uncertainty information, update latency, and specific usage restrictions (e.g., for commercial use).
2) Evaluate these candidates using this paper and the NSF NCAR Climate Data Guide pages to understand their construction, strengths, and known limitations, drawing on the peer-reviewed dataset literature, user guides, and methodological papers to identify issues that may be relevant for specific scientific questions.
3) Wherever possible, analyze the full ensemble and more than one suitable dataset, so that conclusions can be assessed for sensitivity to parametric and structural uncertainty.
4) Cite all datasets in accordance with journal and producer guidelines, including the dataset and its associated publications, to support transparent scientific reuse and sustain the development of high-quality SST products.

At the same time, continued progress in observational coverage, data and metadata rescue, understanding of bias and uncertainty, and infrastructure will enable faster cycles of improvement. As these advances accelerate, the coexistence of multiple approaches with different methodological choices will help quantify structural uncertainty, sharpening our understanding of past climate change and tightening constraints on future projections.

**Acknowledgments.** We acknowledge Nick Rayner of the Met Office for her insights on the HadISST family. D. C. was supported by a UKRI Future Leaders Fellowship (UKRI2360). E. C. K. and R. C. C. were supported by UKRI/NERC via Grant NE/Y005589/1. C. D. and N. L. were supported by the National


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

Center for Atmospheric Research (NCAR), which is sponsored by the National Science Foundation under Cooperative Agreement 1852977. C. J. M. was supported by UKRI/NERC via the National Centre for Earth Observation, Grant NE/RO16518/1. C. S. was supported by the Met Office Hadley Centre Climate Programme, funded by DSIT. P. H. was supported by NSF Grant 2123295. G. G. was supported by U.S. NSF OCE-2122805.

**Data availability statement.** All datasets and model output used in this paper are in the public domain, and citation and access information is summarized in Table 1. A list of individual CMIP6 ensemble members used can be found in Table S1. Code and data for reproducing the main-text figures can be downloaded from https://doi.org/10.7910/DVN/WBBJT2.


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

## References


Abernathey, R. P., and Coauthors, 2021: Cloud-native repositories for big scientific data. *Comput. Sci. Eng.*, **23**, 26–35, https://doi.org/10.1109/MCSE.2021.3059437.

Allan, R. P., and Coauthors, 2023: Summary for policymakers. *Climate Change 2021: The Physical Science Basis*, Cambridge University Press, 3–32.

Beadling, R. L., and Coauthors, 2026: Observational data for next-generation climate model evaluation: Requirements, considerations, and best practices. *Bull. Amer. Meteor. Soc.*, **107**, E813–E835, https://doi.org/10.1175/BAMS-D-25-0079.1.

Bian, C., Z. Jing, H. Wang, L. Wu, Z. Chen, B. Gan, and H. Yang, 2023: Oceanic mesoscale eddies as crucial drivers of global marine heatwaves. *Nat. Commun.*, **14**, 2970, https://doi.org/10.1038/s41467-023-38811-z.

Bottomley, M., C. Folland, J. Hsiung, R. Newell, and D. Parker, 1990: *Global Ocean Surface Temperature Atlas (GOSTA).* MIT and Meteorological Office, 20 pp.

Brasnett, B., 2008: The impact of satellite retrievals in a global sea-surface-temperature analysis. *Quart. J. Roy. Meteor. Soc.*, **134**, 1745–1760, https://doi.org/10.1002/qj.319.

Brönnimann, S., J. Luterbacher, J. Staehelin, T. M. Svendby, G. Hansen, and T. Svenøe, 2004: Extreme climate of the global troposphere and stratosphere in 1940–42 related to El Niño. *Nature*, **431**, 971–974, https://doi.org/10.1038/nature02982.

Byrne, H., R. Seager, and J. E. Smerdon, 2025: CMIP6 models cannot capture long-term forced changes in the tropical Pacific sea surface temperature gradient. *Nat. Commun.*, **17**, 142, https://doi.org/10.1038/s41467-025-66839-w.

Carella, G., E. C. Kent, and D. I. Berry, 2017: A probabilistic approach to ship voyage reconstruction in ICOADS. *Int. J. Climatol.*, **37**, 2233–2247, https://doi.org/10.1002/joc.4492.

Cavalieri, D. J., C. L. Parkinson, N. DiGirolamo, and A. Ivanoff, 2012: Intersensor calibration between F13 SSMI and F17 SSMIS for global sea ice data records. *IEEE Geosci. Remote Sens. Lett.*, **9**, 233–236, https://doi.org/10.1109/LGRS.2011.2166754.

Chan, D., 2021: Combining statistical, physical, and historical evidence to improve historical sea-surface temperature records. *Harv. Data Sci. Rev.*, **3**, 1, https://doi.org/10.1162/99608f92.edcee38f.

——, and P. Huybers, 2019: Systematic differences in bucket sea surface temperature measurements among nations identified using a linear-mixed-effect method. *J. Climate*, **32**, 2569–2589, https://doi.org/10.1175/JCLI-D-18-0562.1.

——, and ——, 2021: Correcting observational biases in sea surface temperature observations removes anomalous warmth during World War II. *J. Climate*, **34**, 4585–4602, https://doi.org/10.1175/JCLI-D-20-0907.1.

——, E. C. Kent, D. I. Berry, and P. Huybers, 2019: Correcting datasets leads to more homogeneous early-twentieth-century sea surface warming. *Nature*, **571**, 393–397, https://doi.org/10.1038/s41586-019-1349-2.

——, G. A. Vecchi, W. Yang, and P. Huybers, 2021: Improved simulation of 19th- and 20th-century North Atlantic hurricane frequency after correcting historical sea surface temperatures. *Sci. Adv.*, **7**, eabg6931, https://doi.org/10.1126/sciadv.abg6931.

——, G. Gebbie, and P. Huybers, 2023: Global and regional discrepancies between early-twentieth-century coastal air and sea surface temperature detected by a coupled energy-balance analysis. *J. Climate*, **36**, 2205–2220, https://doi.org/10.1175/JCLI-D-22-0569.1.

——, ——, ——, and E. C. Kent, 2024a: A Dynamically Consistent ENsemble of Temperature at the Earth surface since 1850 from the DCENT dataset. *Sci. Data*, **11**, 953, https://doi.org/10.1038/s41597-024-03742-x.

——, ——, and ——, 2024b: An improved ensemble of land surface air temperatures since 1880 using revised pair-wise homogenization algorithms accounting for autocorrelation. *J. Climate*, **37**, 2325–2345, https://doi.org/10.1175/JCLI-D-23-0338.1.

——, ——, and ——, 2025: Re-evaluating historical sea surface temperature data sets: Insights from the diurnal cycle, coral proxy data, and radiative forcing. *Geophys. Res. Lett.*, **52**, e2025GL116615, https://doi.org/10.1029/2025GL116615.

——, and Coauthors, 2026: DCENT-I: A globally infilled extension of the Dynamically Consistent ENsemble of Temperature dataset. *Geosci. Data J.*, **13**, e70054, https://doi.org/10.1002/gdj3.70054.

Chin, T. M., J. Vazquez-Cuervo, and E. M. Armstrong, 2017: A multi-scale high-resolution analysis of global sea surface temperature. *Remote Sens. Environ.*, **200**, 154–169, https://doi.org/10.1016/j.rse.2017.07.029.

Delworth, T. L., and T. R. Knutson, 2000: Simulation of early 20th century global warming. *Science*, **287**, 2246–2250, https://doi.org/10.1126/science.287.5461.2246.

Deser, C., A. S. Phillips, and M. A. Alexander, 2010: Twentieth century tropical sea surface temperature trends revisited. *Geophys. Res. Lett.*, **37**, L10701, https://doi.org/10.1029/2010GL043321.

——, and Coauthors, 2012: ENSO and Pacific decadal variability in the Community Climate System Model version 4. *J. Climate*, **25**, 2622–2651, https://doi.org/10.1175/JCLI-D-11-00301.1.

Dong, Y., A. G. Pauling, S. Sadai, and K. C. Armour, 2022: Antarctic ice-sheet meltwater reduces transient warming and climate sensitivity through the sea-surface temperature pattern effect. *Geophys. Res. Lett.*, **49**, e2022GL101249, https://doi.org/10.1029/2022GL101249.

——, L. M. Polvani, and D. B. Bonan, 2023: Recent multi-decadal Southern Ocean surface cooling unlikely caused by Southern Annular Mode trends. *Geophys. Res. Lett.*, **50**, e2023GL106142, https://doi.org/10.1029/2023GL106142.

Embury, O., and C. J. Merchant, 2012: A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: A new retrieval scheme. *Remote Sens. Environ.*, **116**, 47–61, https://doi.org/10.1016/j.rse.2010.11.020.

——, and Coauthors, 2024: Satellite-based time-series of sea-surface temperature since 1980 for climate applications. *Sci. Data*, **11**, 326, https://doi.org/10.1038/s41597-024-03147-w.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

——, and Coauthors, 2023: Human influence on the climate system. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 423–552, https://doi.org/10.1017/9781009157896.005.

Fiedler, E. K., and Coauthors, 2019: Intercomparison of long-term sea surface temperature analyses using the GHRSST Multi-Product Ensemble (GMPE) system. *Remote Sens. Environ.*, **222**, 18–33, https://doi.org/10.1016/j.rse.2018.12.015.

Folland, C. K., and D. E. Parker, 1995: Correction of instrumental biases in historical sea surface temperature data. *Quart. J. Roy. Meteor. Soc.*, **121**, 319–367, https://doi.org/10.1002/qj.49712152206.

——, ——, and F. E. Kates, 1984: Worldwide marine temperature fluctuations 1856–1981. *Nature*, **310**, 670–673, https://doi.org/10.1038/310670a0.

Forget, G., J.-M. Campin, P. Heimbach, C. N. Hill, R. M. Ponte, and C. Wunsch, 2015: ECCO version 4: An integrated framework for non-linear inverse modeling and global ocean state estimation. *Geosci. Model Dev.*, **8**, 3071–3104, https://doi.org/10.5194/gmd-8-3071-2015.

Freeman, E., and Coauthors, 2017: ICOADS release 3.0: A major update to the historical marine climate record. *Int. J. Climatol.*, **37**, 2211–2232, https://doi.org/10.1002/joc.4775.

Gottschalk, B., 2017: Global surface temperature trends and the effect of World War II. arXiv, 1703.09281v1, https://doi.org/10.48550/arXiv.1703.09281.

Gregory, J. M., and Coauthors, 2024: A new conceptual model of global ocean heat uptake. *Climate Dyn.*, **62**, 1669–1713, https://doi.org/10.1007/s00382-023-06989-z.



Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

Hegerl, G. C., S. Brönnimann, A. Schurer, and T. Cowan, 2018: The early 20th century warming: Anomalies, causes, and consequences. *Wiley Interdiscip. Rev.: Climate Change*, **9**, e522, https://doi.org/10.1002/wcc.522.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hirahara, S., M. Ishii, and Y. Fukuda, 2014: Centennial-scale sea surface temperature analysis and its uncertainty. *J. Climate*, **27**, 57–75, https://doi.org/10.1175/JCLI-D-12-00837.1.

Huang, B., and Coauthors, 2015: Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4). Part I: Upgrades and intercomparisons. *J. Climate*, **28**, 911–930, https://doi.org/10.1175/JCLI-D-14-00006.1.

——, and Coauthors, 2017: Extended Reconstructed Sea Surface Temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *J. Climate*, **30**, 8179–8205, https://doi.org/10.1175/JCLI-D-16-0836.1.

——, C. Liu, V. Banzon, E. Freeman, G. Graham, B. Hankins, T. Smith, and H.-M. Zhang, 2021: Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) version 2.1. *J. Climate*, **34**, 2923–2939, https://doi.org/10.1175/JCLI-D-20-0166.1.

——, and Coauthors, 2025: Extended Reconstructed Sea Surface Temperature, version 6 (ERSSTv6). Part I: An artificial neural network approach. *J. Climate*, **38**, 1105–1121, https://doi.org/10.1175/JCLI-D-23-0707.1.

Hurrell, J. W., J. J. Hack, D. Shea, J. M. Caron, and J. Rosinski, 2008: A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model. *J. Climate*, **21**, 5145–5153, https://doi.org/10.1175/2008JCLI2292.1.

Ishii, M., A. Shouji, S. Sugimoto, and T. Matsumoto, 2005: Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the Kobe Collection. *Int. J. Climatol.*, **25**, 865–879, https://doi.org/10.1002/joc.1169.

——, A. Nishimura, S. Yasui, and S. Hirahara, 2025: Historical high-resolution daily SST analysis (COBE-SST3) with consistency to monthly land surface air temperature. *J. Meteor. Soc. Japan*, **103**, 17–44, https://doi.org/10.2151/jmsj.2025-002.

Jean-Michel, L., and Coauthors, 2021: The Copernicus global 1/12° oceanic and sea ice GLORYS12 reanalysis. *Front. Earth Sci.*, **9**, 698876, https://doi.org/10.3389/feart.2021.698876.

Jones, P. D., T. M. L. Wigley, and P. B. Wright, 1986: Global temperature variations between 1861 and 1984. *Nature*, **322**, 430–434, https://doi.org/10.1038/322430a0.

Kang, S. M., Y. Shin, H. Kim, S.-P. Xie, and S. Hu, 2023: Disentangling the mechanisms of equatorial Pacific climate change. *Sci. Adv.*, **9**, eadf5059, https://doi.org/10.1126/sciadv.adf5059.

Kaplan, A., Y. Kushnir, M. A. Cane, and M. B. Blumenthal, 1997: Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. *J. Geophys. Res.*, **102**, 27 835–27 860, https://doi.org/10.1029/97JC01734.

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

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

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

——, N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011a: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res.*, **116**, D14103, https://doi.org/10.1029/2010JD015218.

——, ——, ——, ——, and ——, 2011b: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *J. Geophys. Res.*, **116**, D14104, https://doi.org/10.1029/2010JD015220.

——, ——, C. P. Atkinson, and R. E. Killick, 2019: An ensemble data set of sea surface temperature change from 1850: The Met Office Hadley Centre HadSST.4.0.0.0 data set. *J. Geophys. Res. Atmos.*, **124**, 7719–7763, https://doi.org/10.1029/2018JD029867.

Kent, E. C., and P. K. Taylor, 2006: Toward estimating climatic trends in SST. Part I: Methods of measurement. *J. Atmos. Oceanic Technol.*, **23**, 464–475, https://doi.org/10.1175/JTECH1843.1.

——, and J. J. Kennedy, 2021: Historical estimates of surface marine temperatures. *Ann. Rev. Mar. Sci.*, **13**, 283–311, https://doi.org/10.1146/annurev-marine-042120-111807.

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

Knight, J. R., C. K. Folland, and A. A. Scaife, 2006: Climate impacts of the Atlantic Multidecadal Oscillation. *Geophys. Res. Lett.*, **33**, L17706, https://doi.org/10.1029/2006GL026242.

Kosaka, Y., and Coauthors, 2024: The JRA-3Q reanalysis. *J. Meteor. Soc. Japan*, **102**, 49–109, https://doi.org/10.2151/jmsj.2024-004.

Lee, S., M. L'Heureux, A. T. Wittenberg, R. Seager, P. A. O'Gorman, and N. C. Johnson, 2022: On the future zonal contrasts of equatorial Pacific climate: Perspectives from observations, simulations, and theories. *npj Climate Atmos. Sci.*, **5**, 82, https://doi.org/10.1038/s41612-022-00301-2.

McPhaden, M. J., S. E. Zebiak, and M. H. Glantz, 2006: ENSO as an integrating concept in Earth science. *Science*, **314**, 1740–1745, https://doi.org/10.1126/science.1132588.

Menemenlis, S., G. A. Vecchi, W. Yang, S. Fueglistaler, and S. P. Raghuraman, 2025: Consequential differences in satellite-era sea surface temperature trends across datasets. *Nat. Climate Change*, **15**, 897–903, https://doi.org/10.1038/s41558-025-02362-6.

Merchant, C. J., A. R. Harris, M. J. Murray, and A. M. Závody, 1999: Toward the elimination of bias in satellite retrievals of sea surface temperature: 1. Theory, modeling and interalgorithm comparison. *J. Geophys. Res.*, **104**, 23 565–23 578, https://doi.org/10.1029/1999JC900105.

——, P. Le Borgne, A. Marsouin, and H. Roquet, 2008a: Optimal estimation of sea surface temperature from split-window observations. *Remote Sens. Environ.*, **112**, 2469–2484, https://doi.org/10.1016/j.rse.2007.11.011.

——, and Coauthors, 2008b: Deriving a sea surface temperature record suitable for climate change research from the along-track scanning radiometers. *Adv. Space Res.*, **41** (1), 1–11, https://doi.org/10.1016/j.asr.2007.07.041.

——, S. Saux-Picart, and J. Waller, 2020: Bias correction and covariance parameters for optimal estimation by exploiting matched in-situ references. *Remote Sens. Environ.*, **237**, 111590, https://doi.org/10.1016/j.rse.2019.111590.

Mittaz, J., C. J. Merchant, and E. R. Woolliams, 2019: Applying principles of metrology to historical Earth observations from satellites. *Metrologia*, **56**, 032002, https://doi.org/10.1088/1681-7575/ab1705.

Oliver, E. C. J., J. A. Benthuysen, S. Darmaraki, M. G. Donat, A. J. Hobday, N. J. Holbrook, R. W. Schlegel, and A. Sen Gupta, 2021: Marine heatwaves. *Ann. Rev. Mar. Sci.*, **13**, 313–342, https://doi.org/10.1146/annurev-marine-032720-095144.

Pfeiffer, M., J. Zinke, W.-C. Dullo, D. Garbe-Schönberg, M. Latif, and M. E. Weber, 2017: Indian Ocean corals reveal crucial role of World War II bias for twentieth century warming estimates. *Sci. Rep.*, **7**, 14434, https://doi.org/10.1038/s41598-017-14352-6.

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

——, 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. *J. Climate*, **19**, 446–469, https://doi.org/10.1175/JCLI3637.1.


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC

Reynolds, R. W., 1993: Impact of Mount Pinatubo aerosols on satellite-derived sea surface temperatures. *J. Climate*, **6**, 768–774, https://doi.org/10.1175/1520-0442(1993)006<0768:IOMPAO>2.0.CO;2.

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

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

——, D. B. Chelton, J. Roberts-Jones, M. J. Martin, D. Menemenlis, and C. J. Merchant, 2013: Objective determination of feature resolution in two sea surface temperature analyses. *J. Climate*, **26**, 2514–2533, https://doi.org/10.1175/JCLI-D-12-00787.1.

Rohde, R. A., and Z. Hausfather, 2020: The Berkeley Earth Land/Ocean Temperature Record. *Earth Syst. Sci. Data*, **12**, 3469–3479, https://doi.org/10.5194/essd-12-3469-2020.

Sandford, C., and N. Rayner, 2026: Addressing the World War 2 warm anomaly in HadSST.4.2.0.0. *Int. J. Climatol.*, **46**, e70388, https://doi.org/10.1002/joc.70388.

Schneider, D. P., C. Deser, J. Fasullo, and K. E. Trenberth, 2013: Climate data guide spurs discovery and understanding. *Eos, Trans. Amer. Geophys. Union*, **94**, 121–122, https://doi.org/10.1002/2013EO130001.

——, ——, and T. Fan, 2015: Comparing the impacts of tropical SST variability and polar stratospheric ozone loss on the Southern Ocean westerly winds. *J. Climate*, **28**, 9350–9372, https://doi.org/10.1175/JCLI-D-15-0090.1.

Seager, R., M. Cane, N. Henderson, D.-E. Lee, R. Abernathey, and H. Zhang, 2019: Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. *Nat. Climate Change*, **9**, 517–522, https://doi.org/10.1038/s41558-019-0505-x.

Seviour, W. J. M., and Coauthors, 2019: The Southern Ocean sea surface temperature response to ozone depletion: A multimodel comparison. *J. Climate*, **32**, 5107–5121, https://doi.org/10.1175/JCLI-D-19-0109.1.

Sherwood, S. C., and Coauthors, 2020: An assessment of Earth's climate sensitivity using multiple lines of evidence. *Rev. Geophys.*, **58**, e2019RG000678, https://doi.org/10.1029/2019RG000678.

Singh, L. G., and S. E. Middleton, 2025: Data rescue of historical tables through semi-supervised table structure recognition. *Int. J. Doc. Anal. Recognit.*, **29**, 429–445, https://doi.org/10.1007/s10032-025-00562-6.

Sippel, S., and Coauthors, 2024: Early-twentieth-century cold bias in ocean surface temperature observations. *Nature*, **635**, 618–624, https://doi.org/10.1038/s41586-024-08230-1.

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

——, R. E. Livezey, and S. S. Shen, 1998: An improved method for analyzing sparse and irregularly distributed SST data on a regular grid: The tropical Pacific Ocean. *J. Climate*, **11**, 1717–1729, https://doi.org/10.1175/1520-0442(1998)011<1717:AIMFAS>2.0.CO;2.

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

Teleti, P., E. Hawkins, and K. R. Wood, 2024: Digitizing weather observations from World War II US naval ship logbooks. *Geosci. Data J.*, **11**, 314–329, https://doi.org/10.1002/gdj3.222.

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**, 646–649, https://doi.org/10.1038/nature06982.

Thorne, P. W., D. E. Parker, J. R. Christy, and C. A. Mears, 2005: Uncertainties in climate trends: Lessons from upper-air temperature records. *Bull. Amer. Meteor. Soc.*, **86**, 1437–1442, https://doi.org/10.1175/BAMS-86-10-1437.

Tingley, M. P., and P. Huybers, 2010: A Bayesian algorithm for reconstructing climate anomalies in space and time. Part I: Development and applications to paleoclimate reconstruction problems. *J. Climate*, **23**, 2759–2781, https://doi.org/10.1175/2009JCLI3015.1.

Trenberth, K. E., 1997: The definition of El Niño. *Bull. Amer. Meteor. Soc.*, **78**, 2771–2777, https://doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2.

——, and D. J. Shea, 2006: Atlantic hurricanes and natural variability in 2005. *Geophys. Res. Lett.*, **33**, L12704, https://doi.org/10.1029/2006GL026894.

Watanabe, M., J.-L. Dufresne, Y. Kosaka, T. Mauritsen, and H. Tatebe, 2021: Enhanced warming constrained by past trends in equatorial Pacific sea surface temperature gradient. *Nat. Climate Change*, **11**, 33–37, https://doi.org/10.1038/s41558-020-00933-3.

Wittenberg, A. T., 2009: Are historical records sufficient to constrain ENSO simulations? *Geophys. Res. Lett.*, **36**, L12702, https://doi.org/10.1029/2009GL038710.

Worsfold, M., S. Good, C. Atkinson, and O. Embury, 2024: Presenting a long-term, reprocessed dataset of global sea surface temperature produced using the OSTIA system. *Remote Sens.*, **16**, 3358, https://doi.org/10.3390/rs16183358.

Yang, C., and Coauthors, 2021: Sea surface temperature intercomparison in the framework of the Copernicus Climate Change Service (C3S). *J. Climate*, **34**, 5257–5283, https://doi.org/10.1175/JCLI-D-20-0793.1.


Brought to you by NATIONAL OCEANOGRAPHY CENTRE | Unauthenticated | Downloaded 10/27/26 02:13 PM UTC