Choosing the Right Sea Surface Temperature Dataset: 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] , Geoffrey Gebbie[m] , Nick Rayner[i]

[a] *School of Ocean and Earth Science, University of Southampton, UK*

[b] *National Oceanography Centre, Southampton, UK*

[c] *Colorado School of Mines, Applied Mathematics and Statistics, USA*

[d] *National Center for Atmospheric Research, USA*

[e] *National Centre for Earth Observation, Reading, UK*

[f] *Department of Meteorology, University of Reading, Reading, UK*

[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, UK*

[j] *NOAA / National Centers for Environmental Information, North Carolina, USA*

[k] *Independent researcher, Verdun, France*

[l] *Department of Earth and Planetary Sciences, Harvard University, USA*

[m] *Department of Physical Oceanography, Woods Hole Oceanographic Institution, USA*

Corresponding author: Duo Chan, Duo.Chan@soton.ac.uk

ABSTRACT: Sea surface temperature (SST) datasets underpin many climate applications, including monitoring, attribution, model evaluation, ecosystem assessment, and boundary conditions for atmospheric 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-20th-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.

CAPSULE: A user-oriented synthesis of the evolution of sea surface temperature (SST) datasets, how their differences influence climate analyses, and practical guidance and tools to help users choose appropriate products.

Significance Statement

Sea surface temperature is an "Essential Climate Variable" used for tracking climate change, evaluating models, and understanding events such as marine heatwaves 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.

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 heatwaves (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), atmospheric model simulations, e.g., the Atmospheric Model Intercomparison Project (AMIP, Eyring et al. 2016), and represents key modes of climate variability such as the El Niño-Southern Oscillation (ENSO, McPhaden et al. 2006) and the 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 target one of three main user requirements (Fig. 1):

Table 1: SST datasets used in this paper

Dataset	Dataset Name	Citation	Available from
DCSST	Dynamically Consistent SST	Chan et al. (2024a)	https://doi.org/10.7910/DVN/NU4UGW
DCSST-I	Dynamically Consistent SST - Infilled	Chan et al. (2026)	https://doi.org/10.7910/DVN/ROG38Q
HadSST4.2	Met Office Hadley Centre SST	Sandford and Rayner (in review)	https://www.metoffice.gov.uk/hadobs/hadsst4
HadSST4.1		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/products/extended-reconstructed-sst
ERSSTv5		Huang et al. (2017)	https://www.ncei.noaa.gov/pub/data/cmb/ersst/v5/netcdf
ERSSTv4		Huang et al. (2015)	https://www.ncei.noaa.gov/pub/data/cmb/ersst/v4/netcdf
ERSSTv3b		Smith et al. (2008)	https://www.ncei.noaa.gov/pub/data/cmb/ersst/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-SST		Ishii et al. (2005)	https://ds.data.jma.go.jp/tcc/tcc/products/elnino/cobesst_doc.html
HadISST1	Hadley Centre Sea Ice and Sea Surface Temperature data set	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 Sea Surface Temperature	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

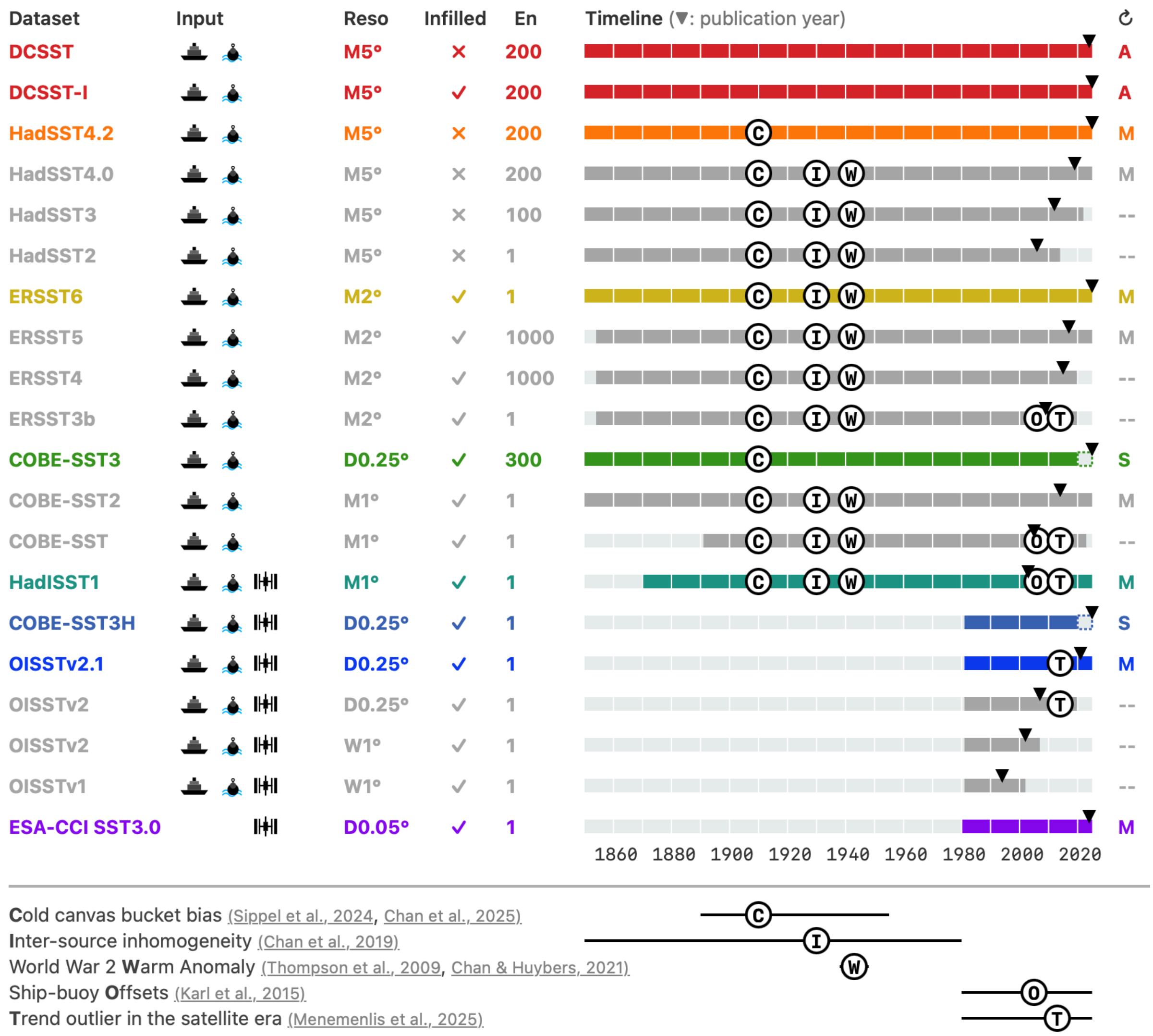

Fig. 1: **Overview of major SST dataset families:** DCSST (red); HadSST (orange); ERSST (yellow); COBE-SST (green); HadISST (teal); COBE-SSTH (light blue); OISST (blue) and ESA CCI SST (purple). Datasets are grouped by family and ordered by version within each family. Columns indicate input data types (ship/ buoy/ satellite), nominal temporal (Monthly/ Weekly/ Daily) and spatial resolution (°), spatial completeness (✓/×), ensemble size, temporal span (horizontal bar), publication year (downward triangle), and update frequency (Annual/ Monthly/ Static/ Discontinued (–)). Symbols mark major known biases and artifacts that remain in each product: cold canvas bucket bias (C), inter-source inhomogeneity (I), World War II warm anomaly (W), ship–buoy offsets (O), and trend outliers in the satellite era (T), based on evaluations presented in section 2 and 3. Reference ranges for these issues are shown in the legend below. Dataset abbreviations follow those in the text and are expanded in Table 1

1. Long historical *in situ* records beginning before 1900, e.g. HadSST, ERSST and 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 six years aligned with the Intergovernmental Panel on Climate Change (IPCC) assessment cycle.

2. High-resolution SST analyses, e.g. ESA 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 multi-decadal records at intermediate resolution designed for input to atmospheric reanalyses or as boundary conditions for other atmosphere-only dynamical models. An example is HadISST which blends *in situ* observations with satellite measurements to reconstruct global fields back to the 1870s.

The creation of an SST product generally requires four elements: (1) data selection (and, in the case of satellite-based products, inference of SST from top-of-atmosphere measurements); (2) bias corrections to 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. DCSST.

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 summaries from the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, Freeman et al. 2017), or legacy products with outdated bias adjustments, e.g., the Kaplan SST (Kaplan et al. 1998), remain highly cited and widely used 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 remains among the most cited SST datasets in 2025, but is one of few products to not correct post-1950 ship-based SST biases. This bias leads to lower warming rate estimates since 2000 (Karl et al. 2015) and thus 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 infrastructure where “switching costs” can be high. Familiarity often shapes dataset choice, while barriers such as non-standard 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, shaping the next years of climate science (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. Legacy products whose methods have not been updated since before 2000 (e.g., Kaplan SST) or those lacking any bias adjustments (e.g., gridded ICOADS) are excluded due to limited comparability. Several high-quality near-real-time analyses are omitted because they are either shorter than forty years (e.g., the Multiscale Reanalysis by Chin et al., 2017 and the Canadian Meteorological Center analysis by Brasnett et al., 2018) or built on a significant input of ESA CCI SST data (e.g., the OSTIA reprocessing by Worsfold et al., 2024). Operational SST analyses that principally support numerical weather prediction are coordinated by the Group for High Resolution Sea Surface Temperature (GHRSST, `www.ghrsst.org/`), and inter-comparisons of these datasets have been reported elsewhere (e.g., Fiedler et al. 2019; Yang et al. 2021). We also do not consider hybrid datasets that combine other products, for example, blends of different products made for reanalysis (e.g. Hersbach et al. 2020), or surface-forcing data sets for AMIP-style uncoupled simulations that combine HadISST1 and OISSTv2 (Hurrell et al. 2008).

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.

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 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, ERSST3b 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.

A major indication of errors present in engine-room-intake (ERI) temperatures, which caused a spurious decrease in global mean surface temperature by approximately 0.3°C following World War II, was discovered by Thompson et al. (2008). 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 ERSST4, HadSST4.0 and COBE-SST2.

Since 2019, attention has expanded from method-specific biases to finer spatial and platform-dependent structures. HadSST4.0 used 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-20th-century warming pattern (Chan et al. 2019). This truncation bias has recently been adjusted in DCSST(-I), 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 DCSST 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. 2020a) 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 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 the SST at 20 cm depth for compatibility with centennial-scale datasets using *in situ* data from drifting buoys and buckets.

b. Construction of gridded fields

A variety of approaches are used to construct gridded fields from individual measurements. Obvious differences between products are 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 products are usually sufficient for studying slowly varying climate backgrounds, much higher spatial and temporal resolution is required for studying extreme events like marine heat waves. Within a family, some products have trended toward finer resolution, as in OISST and COBE-SST (Fig. 1).

A relevant concept is the distinction between nominal grid resolution and effective resolution (Reynolds et al. 2013). In other words, a finer grid 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. 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).

Another application-oriented difference is whether unsampled grid cells remain missing or are infilled to be globally complete. Non-infilled datasets such as the HadSST family are often preferred for climate monitoring as they are closer to the original observations; infilled fields are generally more convenient to use, but weaken the traceability to original observations by making assumptions about the variability to gain the spatial completeness.

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 variations in covariance, as implemented in DCSST-I 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 ERSST v3–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 (v6), 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 the data (Kennedy 2014). Most products provide uncertainty values per grid box and/or time step (e.g., ESA CCI SST, COBE-SST1–2) or ensembles of plausible realizations (e.g., ERSSTv4–5, DCSST(-I), COBE-SST3) or both (e.g., HadSST3–4.2) for uncertainty quantification. Some older products such as HadISST1 do not provide uncertainty estimates.

Uncertainty 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 consistent with observational error covariance. Ensembles can quantify complex error structures which cannot be handled analytically. Because individual members often 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 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 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-20$^{\text{th}}$-century cold bias (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 inter-product 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. Evaluation and 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, Figure 2), the spatial structures of events such as ENSO and marine heatwaves (Figure 3), and climate features including long-term warming, major modes of variability, and important regional gradients (Figure 4).

We additionally compare the observed SST metrics with state-of-the-art CMIP6 simulations (Figures 2 and 4). Ideally, observational datasets should be evaluated independent of model-based expectations insomuch as they are to be used as checks of these models or assumptions that go into construction of such models. That said, climate models are useful for highlighting unexpected features in the datasets. Model-data discrepancies have been important for identifying systematic errors in observations, particularly prior to the satellite era. However, better agreement with CMIP6 alone does not imply that a product is more accurate and model–data consistency is not used as a formal criterion in SST product development. Adjustments in SST data are only made when multiple lines of evidence — physical, statistical, or documentary — indicate data issues with a known cause.

a. Bias signatures

Global-mean SST anomalies (see Table 2 for definition) are visually similar after 1980, indicating broad consistency in the satellite era (Fig. 2a). Earlier periods, however, show clear differences, largely due to biases in observations. For example, the World War II warm anomaly (Fig. 2b; Table 2) is due to wartime changes in measurement practice that introduced warm biases (Thompson et al. 2008; Chan and Huybers 2021). In legacy HadISST1, ERSST and COBE-SST versions, this anomaly amplitude lies outside the -0.12 to 0.11°C (95% c.i.) range from CMIP6 historical simulations. In COBE-SST3 and the new DCSST family, bias corrections reduce the warm

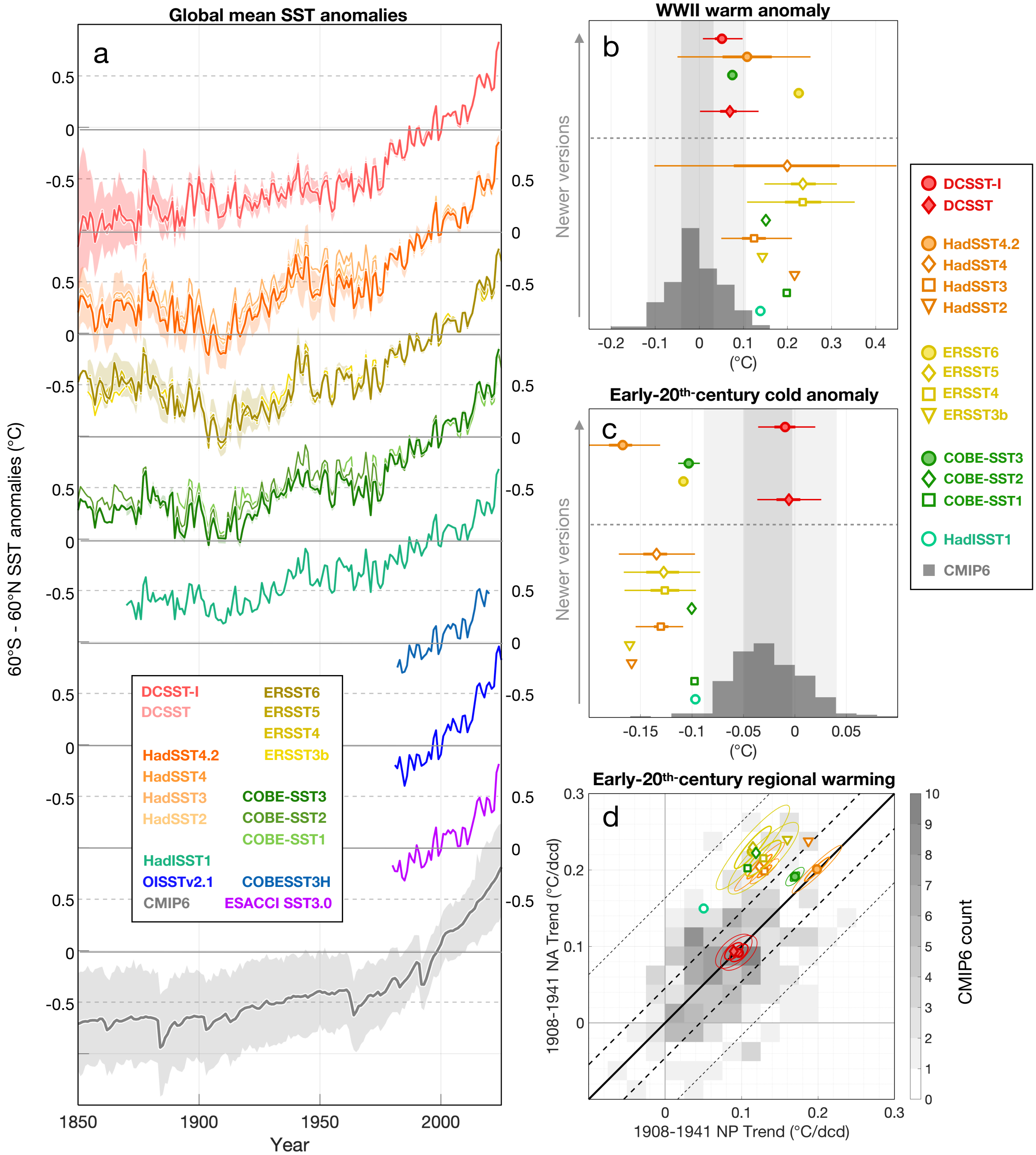

Fig. 2: **Comparison of global mean SST and data artifacts.** (a) global mean SST (60°S-60°N) anomalies relative to 1982–2014 climatology. Datasets are grouped and offset by families. Within each family, thick lines show the central estimate of individual versions (color-coded), and the shading shows the 95% c.i. for the most recent release where an ensemble is available. Simulations from 229 CMIP6 runs, concatenating historical and SSP2-4.5 experiments, are shown at the bottom. (b) World War II warm anomaly, calculated as the global mean SST anomaly over 1941-1945, relative to the mean over 1936-1940 and 1946-1950. Markers, sorted by publication dates (descending) in the y-axis, denote the mean value of a dataset, while thick and thin lines, respectively, denote the interquartile range and 95% confidence interval (c.i.), where an ensemble is available. Dashed line separates state-of-the-art and legacy products. The histogram presents the CMIP6 distribution, and the dark and light shading denotes, respectively, the interquartile and 95% c.i. (2.5%-97.5%). (c) as (b) but for early-20th-century cold SST anomaly, defined as the global-mean SST over 1900–1930 minus a reference SST given by a linear trend fitted to the periods 1890–1899 and 1931–1940. (d) North Atlantic (y-axis) versus North Pacific (x-axis) SST trends over 1908-1941. Markers are as (b), and ellipses denote 1 s.d. and 2 s.d. uncertainty using a bi-variate Gaussian fit. The heat map squares represent the 2D histogram of CMIP6 historical simulations and the black line depicts the one-to-one relationship, and thick and thin dashed lines denote, respectively, the interquartile range and 95% c.i. of the simulated inter-basin trend difference.

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.
Early-20$^{\text{th}}$-Century Cold Bias	Global mean SST difference over 1900–1930 relative to a linear fit between 1890–1899 and 1931–1940, following Sippel et al. (2024).
WWII Warm Anomaly	Global mean SST anomaly averaged over 1941–1945 relative to the mean of 1936–1940 and 1946–1950, following Chan and Huybers (2021)
North Pacific SST	Area-weighted mean over 20°N–60°N, 100°E–100°W, following Chan et al. (2019).
North Atlantic SST	Area-weighted mean over 20°N–60°N, 100°W–10°E (excluding Mediterranean), following Chan et al. (2019).
Early-20$^{\text{th}}$-Century Warming	Linear trend of global mean SST over 1908–1941, following Chan et al. (2019).
Niño3.4 SST	Area-weighted mean over 5°S–5°N, 170°W–120°W.
West Equatorial Pacific SST	Area-weighted mean over 5°S–5°N, 120°E–170°E.
East Equatorial Pacific SST	Area-weighted mean over 5°S–5°N, 150°W–80°W.
Southern Ocean	Area-weighted mean over 50°S–70°S.
AMV Index	The difference between 20-year running smoothed monthly North Atlantic SST anomalies (0°–60°N, 80°W–0°E) and global SST, following Trenberth and Shea (2006).

anomaly to within the CMIP6 envelope, suggesting better physical consistency. HadSST4.2 similarly improved estimates of engine-room-intake bias, reducing the anomaly from 0.18 (-0.10–0.45)°C in HadSST4.0 to 0.11 (-0.05–0.25)°C (95% c.i.), closer to the CMIP6 range. ERSSTv6 is now the only major product family in which a pronounced WWII warm anomaly persists.

Farther back in time, the evolution from 1850 to 1940 differs substantially across product families, but is relatively stable within each family. DCSST shows nearly continuous warming whereas ERSST exhibits the strongest cooling from 1850 to about 1910 before warming quickly. HadSST and COBE-SST lie between these endpoints (Fig. 2a). These four products differ due to the treatment of early bucket biases, modulating the magnitude of the early 20$^{\text{th}}$-century cold anomaly (Sippel et al. 2024). In COBE-SST3, this cold anomaly is similar to earlier COBE-SST releases (~0.1°C, Fig. 2c). HadSST4.2 appears particularly cold by this measure because an adjustment applied to data after 1930 increases the SST in the period used as a reference (Table 2). ERSSTv6, in contrast, produces cooler 1930s SSTs and thus a smaller anomaly relative to HadSST4.2. Nevertheless, most products remain outside the CMIP6 range with only DCSST and its infilled derivative exhibiting early 20th century SSTs consistent with model simulations.

On regional scales, correcting the Japanese truncation bias directly alters the contrast in early-20$^{\text{th}}$-century warming between the North Pacific and North Atlantic (Chan et al. 2019). In legacy products, all families show the North Atlantic warming nearly twice as fast as the North Pacific over 1908–1941 (Fig. 2d), a phenomenon which would require an unusually large expression of internal variability to explain (Delworth and Knutson 2000). In the latest versions, DCSST, HadSST4.2 and COBE-SST3 correct for this bias, bringing the inter-basin warming rates into much closer agreement with each other and with the expected warming pattern under greenhouse-gas forcing. ERSSTv6 still exhibits a pronounced contrast between basins, similar to earlier ERSST releases. Fig. 2d also shows differences in the overall magnitude of early-20$^{\text{th}}$-century warming: DCSST estimates (~0.1°C per decade) fall within the CMIP6 range whereas HadSST4.2 and COBE-SST3 values (~0.2°C per decade) remain on the warm end of the model distribution and exceed observed contemporary land warming (Sippel et al. 2024).

In general, incorporating adjustments for newly identified artifacts in data production has been gradual. Yet, recent versions generally apply more complete corrections, are more internally consistent, and better agree with CMIP simulations.

b. Gridding and Infilling

The different choices in the reconstruction of gridded products, including resolution and infilling, are important to consider for specific 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 non-infilled 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, state-of-the-art approaches, including anisotropic kernels (DCSST-I, Fig. 3a_2), AI-based methods (ERSST6, 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 era. Fine-scaled structure still differs between products as the fields are only tightly constrained by observations

[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.

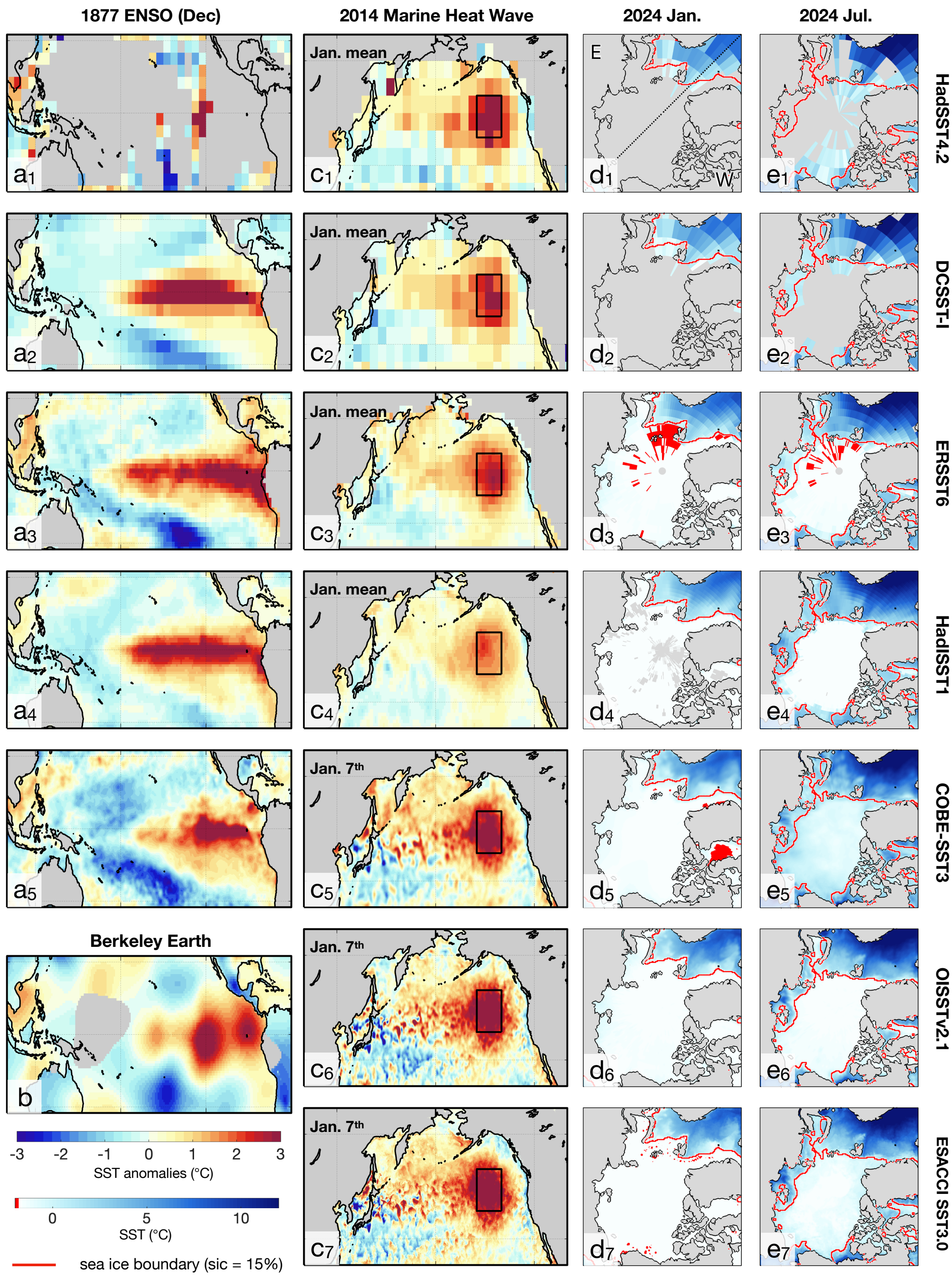

FIG. 3: **Comparison of spatial patterns.** Column (a) shows December 1877 SST anomalies relative to the 1982–2014 December climatology. Rows (top to bottom) show HadSST4.2, DCSST-I, ERSST6, HadISST1, and COBE-SST3. (b) as (a) but for the Berkeley Earth surface temperature. Column (c) shows January 2014 SST anomalies relative to the January climatology over the same period, highlighting the marine heatwave 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 January 7, 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. Columns (d) and (e) show actual SSTs over the Arctic in January and July 2024, respectively. Red curves mark the sea-ice edge (≥ 15% sea-ice concentration, Cavalieri et al. 2011), gray areas indicate missing values, and red regions denote SSTs below the -1.8°C freezing point. In panel d_1, the diagonal dashed line marks the 0° and 180° meridians. The Eastern and Western Hemispheres are labeled "E" and "W," respectively.

near to where they exist. For example, the positive anomaly extends further west in DCSST-I and ERSST6 than in COBE-SST3, contributing to the structural uncertainty across the datasets.

For contemporary extreme events such as marine heatwaves, 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 145°W, 45°N (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 DCSST-I, 2.4°C in ERSST6, 2.9°C in HadSST4.2, and only 1.7°C in HadISST1, whereas on the peak date (January 7) in daily products the corresponding values are usually higher (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 reveal eddy-scale variability and fine filaments, which may be important for understanding the evolution and mechanisms of such events 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., DCSST-I) 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.

Fig. 3d compares absolute Arctic SSTs in January 2024. Infilled products broadly follow the observed ice edge, though ERSST6 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 diverge even more (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 sea-ice–affected regions.

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 trends in the global mean SST (60°S–60°N) remain sensitive to product-specific treatments of nineteenth- and early-20$^{\text{th}}$-century biases (Fig. 4a, sections 2a and 3a); yet excluding legacy datasets only narrows the estimated 2019–2023 warming level, relative to the 1850–1900 baseline, from 0.7–1.0 to 0.8–1.0°C. There is also evidence that the ERSST family, which features pronounced cooling over 1850–1910 (Fig. 2a), is likely too warm in the late nineteenth century (Sippel et al. 2024; Chan et al. 2025), providing scope for further narrowing of the observational range.

During the well-sampled satellite era, observational products are expected to agree more closely. Yet, when comparing legacy and modern datasets, Menemenlis et al. (2025) found a wide spread in the 1982–2024 warming trend, reproduced here (Fig. 4b). Restricting this comparison to state-of-the-art products tightens the range among central estimates from 0.39–0.63 to 0.49–0.63°C per decade. Within this group, DCSST, ERSST6, COBE-SST3, and ESA CCI SST cluster near 0.51°C per decade, with HadSST4.2 being slightly higher at 0.56°C per decade. NOAA's daily OISSTv2.1 is the clear remaining outlier (0.63°C per decade), perhaps due to its fixed 0.14°C ship-to-buoy correction from 1981–2015. Hence, the observational spread in satellite-era SST warming is narrower than previous comparisons that included legacy datasets.

On regional scales, an important indicator is the Equatorial Pacific zonal SST gradient which influences circulation, cloud, albedo, and climate sensitivity (Kang et al. 2023). Model–data discrepancies in the sign of the satellite-era trend in this index have been widely noted (e.g., Lee et al. 2022) and the large influence of internal variability on trends over short time periods motivates examination of century-long trends. Over 1900–2010, CMIP6 models simulate an west-minus-east trend difference of –0.54 to 0.20°C per century (95%c.i.), which, on average, weakens the gradient. However, legacy products such as HadISST1, HadSST2, and COBE-SST2

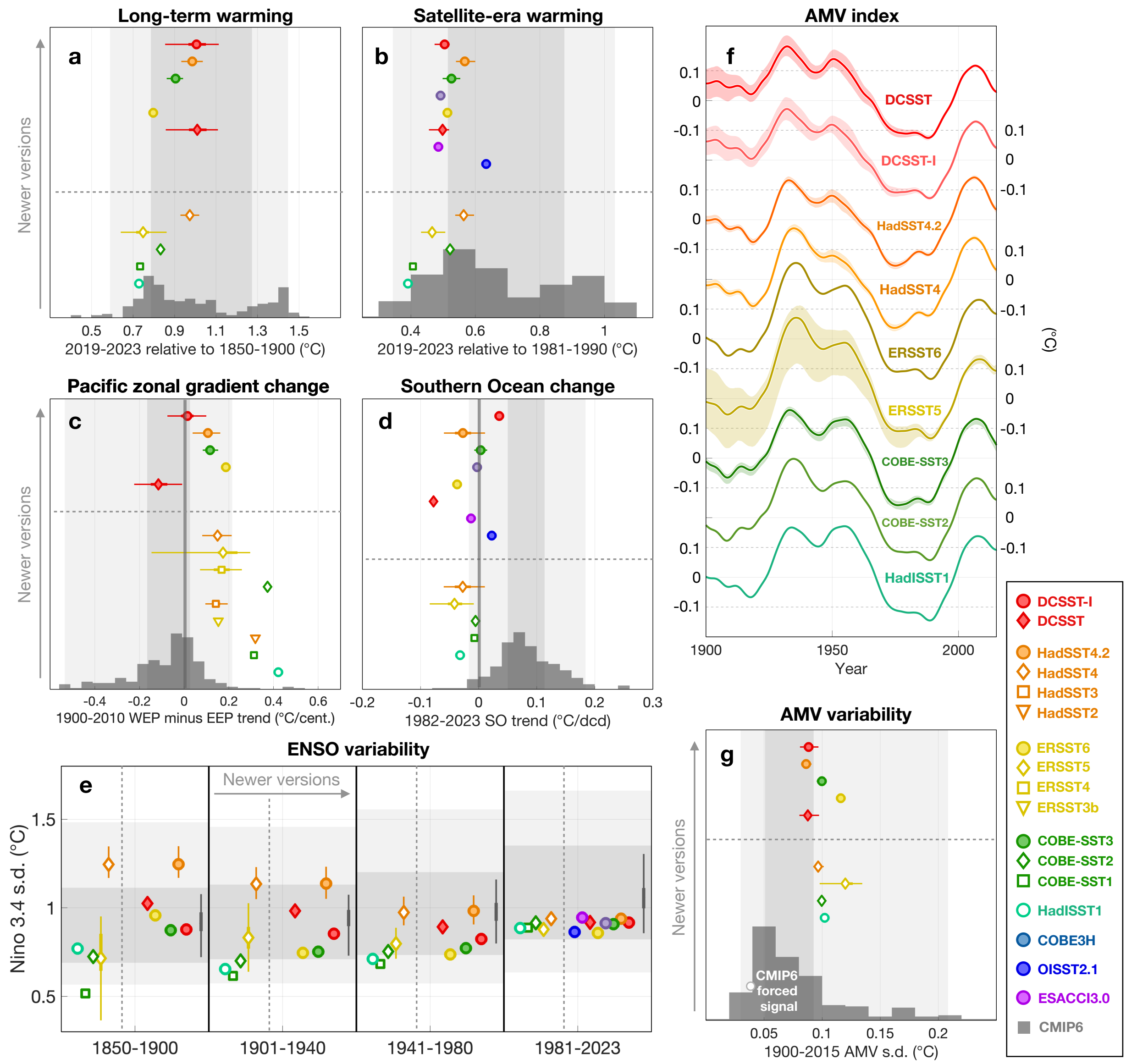

Fig. 4: **Comparison of key climate indices.** (a) Mean SST (60°S–60°N) for 2019–2023 relative to the 1850–1900 baseline. Layout follows Fig. 2b and the dashed lines separate the legacy products from the most recent versions. Only products extending to 2023 are included. Error bars are shown where visible; in some cases they are smaller than the symbol size. (b) As in (a), but relative to the 1981–1990 mean; satellite products are also shown. (c) As in (a), but showing the 1900–2010 linear trend in Tropical Pacific zonal temperature gradient, defined as the difference between the west (5°S–5°N, 120°E–170°E) and the east (5°S–5°N, 150°W–80°W). All products covering 1900 to 2010 are shown; the zero line is also highlighted. (d) As in (c), but for the 1982–2023 linear trend of SSTs over the Southern Ocean (50°S–70°S). (e) Standard deviation of monthly mean SST anomalies in the Niño 3.4 region (5°S–5°N, 170°W–120°W), computed after removing the seasonal cycle and linear trend, for four periods: 1850–1900, 1901–1940, 1941–1980, and 1981–2023. Products are sorted by publication year in the direction indicated by gray arrows. Dark and light shading show the interquartile and 95% confidence intervals across CMIP6 simulations. Gray bars in (e) indicate the spread after removing each model's period mean, emphasizing internal variability (offset from zero for visualization purposes only). (f) Twenty-year smoothed monthly AMV index for recently-updated century-long products. Series are vertically offset for clarity. (g) Standard deviation of the smoothed AMV index in (f). The white dot marks the s.d. of the multi-model mean, indicating the amplitude of the forced signal.

suggests enhanced gradient (positive trends) falling outside the simulated range (Fig. 4c). Newer SST versions indicate relatively weakened gradient than earlier releases; although observational estimates still suggest at most a near-neutral trend, they now fall within the CMIP6 spread. Despite better agreement in the west-minus-east trend difference in newer products, spatial patterns of trends still differ (Fig. S1), underscoring the need to further understand how bias adjustments and infilling choices affect observational estimates, as well as how model structure and configuration shape simulated trends.

Average Southern Ocean SST is another key regional indicator, relevant to Antarctic sea-ice melt (Dong et al. 2022) and heat uptake (Gregory et al. 2024). CMIP6 models generally suggest warming (–0.02 to 0.18°C per decade) over 1982–2023, but observational products show trends closer to zero (Fig. 4d). Spatial patterns also differ across products, particularly in the magnitude and extent of the cooling band (Fig. S2). 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 s.d.), well within the CMIP6 spread (Fig. 4e). Earlier in the record, however, observational estimates diverge to a greater spread than model internal variability after removing model-specific biases. 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 30–50-year 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.

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 on the higher end 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 variability modes over the satellite era. State-of-the-art SST datasets also suggest better agreement with known physical process as presented in CMIP6 simulations. This improving agreement suggests that some model-observation discrepancies in the literature reflect now-resolved data limitations. 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. With that in mind, there are practical considerations that may restrict dataset choice. These include the length of record, whether fields are spatially complete, spatial and temporal resolution, the availability and type of uncertainty estimates, and the immediacy of updates to include the most recent data. All datasets are free to use for research, but some have restrictions for other purposes such as commercial use that need to be checked and adhered to. The **web-based selector tool** (similar to Figure 1) enables users to quickly view, subset, and access candidate datasets suitable for specific applications. When several products exist, results will be more robust if all are used. Typical dataset choices by application include:

- **Climate monitoring**: compare non-infilled 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 non-infilled) with uncertainty estimates. For non-infilled products, apply the same observational coverage mask to the model output to ensure a fair comparison.
- **Western boundary currents, mesoscale eddy signatures, marine heatwaves**: use infilled high-resolution datasets (daily, finer than $1^\circ \times 1^\circ$).
- **Paleo proxy calibration**: use long records without known issues during the calibration period, and compare non-infilled and infilled products for consistency.

Once candidates are thinned by practical considerations, it is necessary to assess data quality. The analysis presented in section 3 shows the importance of choosing the most recent dataset versions in any family. Typically, older, deprecated products should be used only alongside their updated counterparts to aid interpretation of past analyses.

Moreover, even the most recent releases can retain period- or region-specific issues, as described in section 2a and their impacts shown in section 3. If a dataset has known problems that could affect the analysis, it should be excluded. However, if removing such datasets results in too few candidates for a robust assessment, they should be used with caution, provided their limitations are clearly acknowledged in the interpretation.

Another 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 and the assumptions involved. For example, ERSST family's bias adjustment assumes a relatively stable difference between SST and nighttime marine air temperatures (Smith and Reynolds 2002); so agreement with those temperatures is not independent support. Similarly, DCSST is adjusted to be dynamically consistent with its land counterpart DCLSAT (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 largely recover the imposed covariance structure, rather than reveal additional information about the underlying variability.

In addition to checking robustness across qualified datasets, results should also be tested against the estimated uncertainty within each product. This can be done by perturbing the data using the product's uncertainty estimates or by analyzing the ensemble. 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, the first and second sets of 100 members use different approaches to adjust early SST measurements Kennedy et al. (2019), so drawing members from both sub-sets provides a more representative sample.

Finally, follow the data-citation instructions provided by journals, which typically require citing both the dataset and its associated publication, and any additional information requested by dataset producers. Accurate citation does more than acknowledge the source: it helps dataset providers secure support for ongoing maintenance and understand how their products are being used, allowing the datasets to evolve in ways aligned with scientific needs. In turn, this benefits users by increasing the likelihood that high-quality, regularly updated SST datasets remain available.

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 provide 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, 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, 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 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

Better input data and metadata: ongoing efforts to rescue historical data (Teleti et al. 2024) and metadata (Carella et al. 2017) are essential for extending coverage and clarifying bias structures in the early record. Meanwhile, modern data infrastructure is needed to ensure that both rescued and contemporary observations and metadata flow efficiently and transparently into permanent archives and SST dataset production — a gap that currently prevents many recovered measurements from being fully used. For satellite-era products, fundamental work on the calibration of early sensors and SST retrieval methods will also reduce uncertainty and improve stability in the 1980s and 1990s particularly.

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 at least 1° monthly as a baseline and finer daily or sub-daily products where observations permit. 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 co-operation spearheaded by ESA, https://ceos.org/news/avhrr-data-recovery/) to consolidate full-resolution data from early sensors. These higher resolution observations have not been exploited in global SST analyses before, and provide an opportunity for better understanding changes in ecologically-important shelf sea regions.

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. 2020b) and by harmonizing irradiance between satellite platforms prior to retrieval of SST.

Better structural uncertainty estimates: As discussed in section 2c, the current practice of estimating structural uncertainty from an ad hoc ensemble of SST datasets is limited. As understanding of data artifacts improves, clearly inconsistent products are recommended to be excluded from certain analyses. This 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).

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 (for example, 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 catalogued through cooperative efforts of the GHRSST international science team `https://www.ghrsst.org/`.

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

This paper has shown how careful choice of SST datasets is essential for robust research. Over time, SST datasets have improved in quality, and their estimates of important measures of variability have become more consistent. Characterization of dataset uncertainty has also improved, enabling users to understand the sensitivity of their results to uncertainty within each dataset as well as between a selection of different datasets. A number of important indicators, including recent and centennial global trends and the Tropical Pacific trend contrast, show that the most recent SST dataset versions align more closely with one another and with the latest generation of climate models, compared with legacy products. Observational constraints on future projected surface temperature changes are therefore more robust when using state-of-the-art datasets than might be inferred from the use of legacy products.

These considerations can be summarized in a set of practical steps to support effective SST dataset selection and use:

1. **Use the data-selection 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. **Use this paper and NSF NCAR Climate Data Guide pages to evaluate these candidates,** gaining an understanding of their construction, strengths, and known limitations.
3. **Draw on the peer-reviewed dataset literature for the shortlisted products,** including user guides and methodological papers, to identify issues that may be relevant for the specific scientific question.
4. **Wherever possible, analyze the entire uncertainty ensemble and more than one suitable dataset,** so that conclusions can be assessed for robustness to parametric and structural uncertainty.
5. **Finally, cite all datasets in accordance with journal and producer guidelines,** including both the dataset and its associated publications, which supports not only transparent scientific reuse but also continued maintenance and improvements.

Taken together, these practices help ensure that present-day analyses make the best possible use of available SST datasets. At the same time, continued progress in observational coverage, data and metadata rescue, understanding of bias and uncertainty, and infrastructure capability will enable increasingly rapid cycles of improvement. As these advances accelerate, the coexistence of multiple approaches, each making different methodological choices, will help to better quantify structural uncertainty, supporting a more robust understanding of past climate change as well as improved constraints on future projections.

Acknowledgments. D.C. was supported by a UKRI Future Leaders Fellowship (UKRI2360). E.K. and R.C. were supported by UKRI/NERC via grant NE/Y005589/1. C.D. and N.L. were supported by the National 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 summarized in Table 1.

577 ## Supplementary Figures

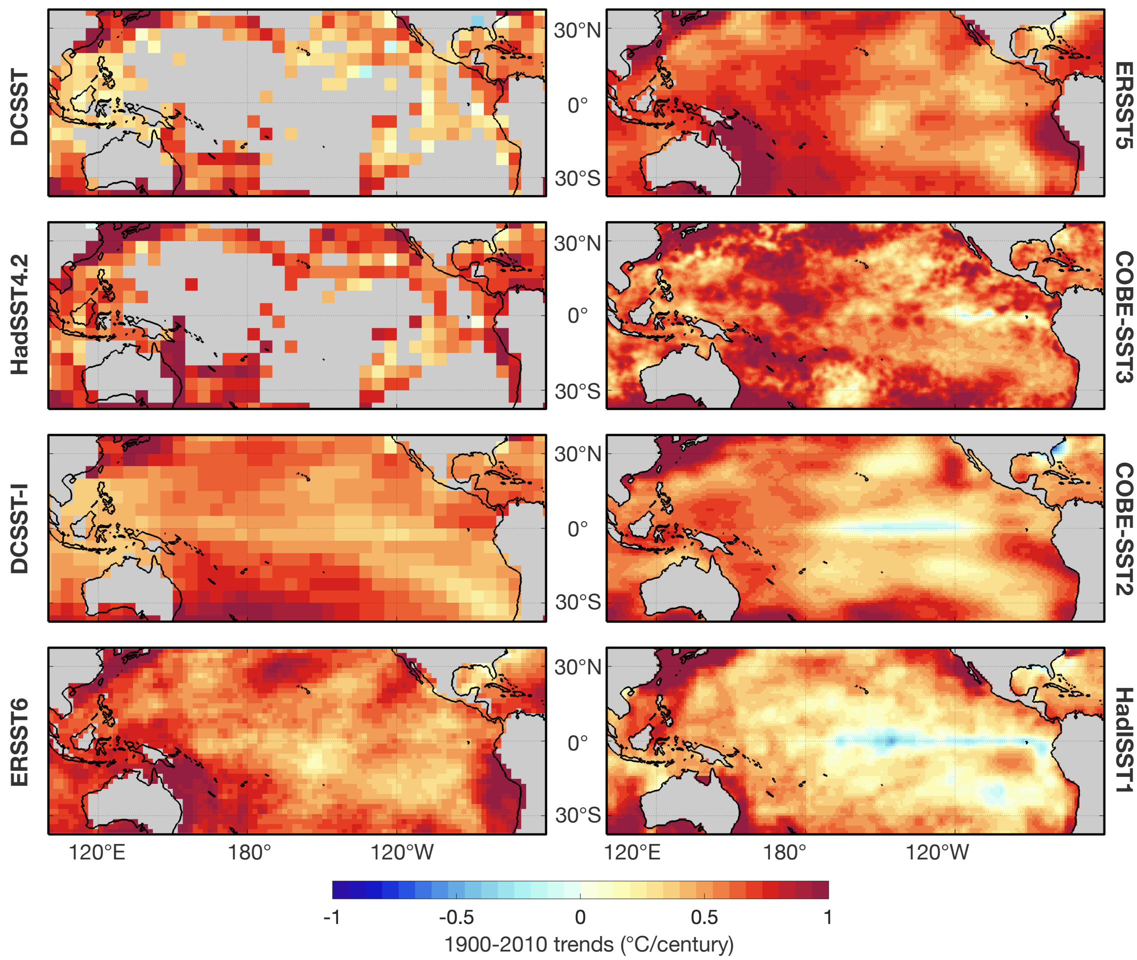

Fig. S1: **Patterns of 1900–2010 SST trends over the Tropical Pacific.** From top to bottom and left to right, the datasets shown are DCSST, HadSST4.2, DCSST-I, ERSST6, ERSST5, COBE-SST3, COBE-SST2, and HadISST1. For the non-infilled datasets (DCSST and HadSST4.2), a grid cell is considered to have a valid trend if it contains at least three valid years in each decade from the 1900s to the 2000s, where a valid year is defined as having at least three months of data.

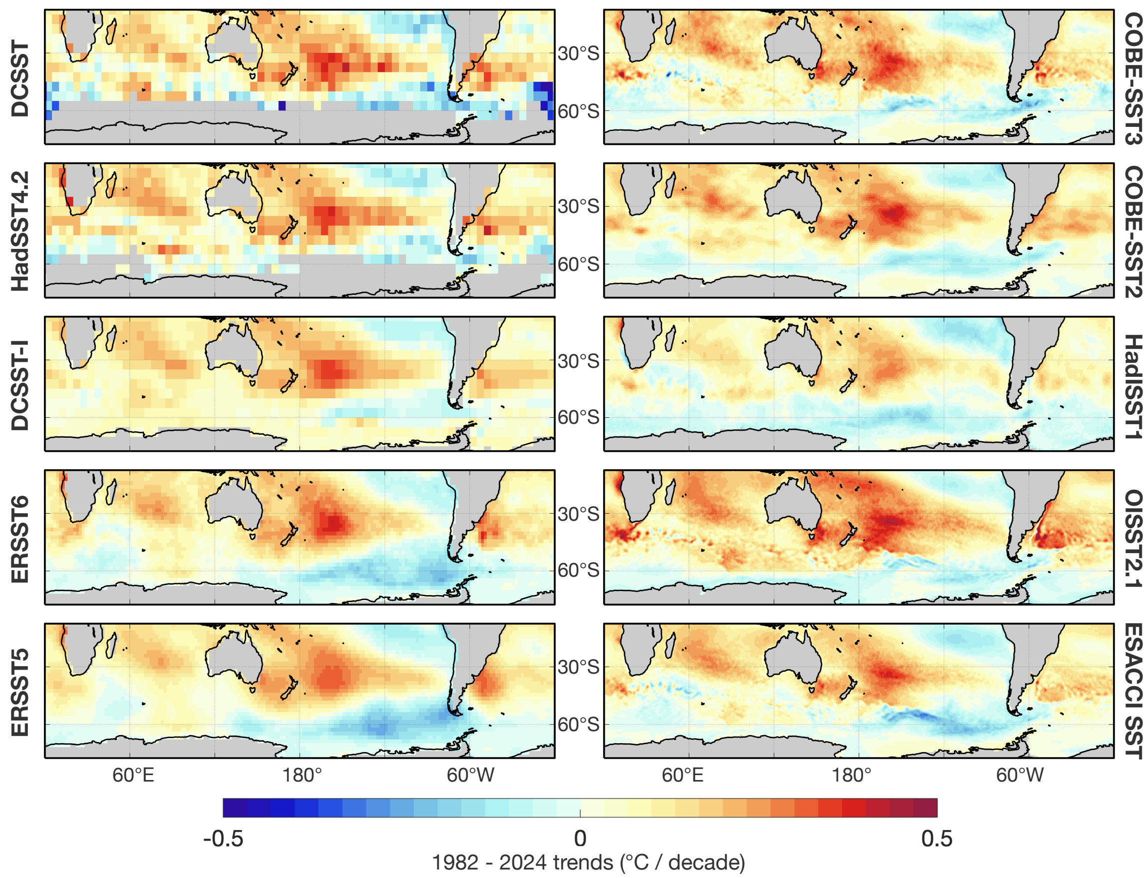

Fig. S2: **Patterns of 1982-2024 SST trends over the Southern Ocean.** From top to bottom and left to right, the datasets shown are DCSST, HadSST4.2, DCSST-I, ERSST6, ERSST5, COBE-SST3, COBE-SST2, HadISST1, OISSTv2.1, and ESA CCI SST3. Similar to Fig. S1, for the non-infilled datasets (DCSST and HadSST4.2), a grid cell is considered to have a valid trend if it contains at least three valid years in each decade from the 1980s to the 2010s, where a valid year is defined as having at least three months of data.

References

Abernathey, R. P., and Coauthors, 2021: Cloud-native repositories for big scientific data. *Computing in Science & Engineering*, **23 (2)**, 26–35.

Beadling, R. L., and Coauthors, 2026: Observational data for next generation climate model evaluation: Requirements, considerations, and best practices. *BAMS (under review)*.

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. *Nature Communications*, **14 (1)**, 2970.

Bottomley, M., C. Folland, J. Hsiung, R. Newell, and D. Parker, 1990: Global ocean surface temperature atlas (GOSTA). *Meteorological Office, Bracknell, UK*.

Brasnett, B., 2008: The impact of satellite retrievals in a global sea-surface-temperature analysis. *Quarterly Journal of the Royal Meteorological Society*, **134 (636)**, 1745–1760.

Carella, G., E. C. Kent, and D. I. Berry, 2017: A probabilistic approach to ship voyage reconstruction in ICOADS. *International Journal of Climatology*, **37 (5)**, 2233–2247.

Cavalieri, D. J., C. L. Parkinson, N. DiGirolamo, and A. Ivanoff, 2011: Intersensor calibration between F13 SSMI and F17 SSMIS for global sea ice data records. *IEEE Geoscience and Remote Sensing Letters*, **9 (2)**, 233–236.

Chan, D., S. C. Chan, J. T. Siddons, A. Cable, R. C. Cornes, E. C. Kent, G. Gebbie, and P. Huybers, 2026: DCENT-I: A globally infilled extension of the Dynamically Consistent ENsemble of Temperature dataset. *Geoscience Data Journal*, **In press**.

Chan, D., G. Gebbie, and P. Huybers, 2023: Global and regional discrepancies between early 20th century coastal air and sea-surface temperature detected by a coupled energy-balance analysis. *Journal of Climate*, **36 (7)**, 2205–2220.

Chan, D., G. Gebbie, and P. Huybers, 2025: Re-evaluating historical sea surface temperature data sets: Insights from the diurnal cycle, coral proxy data, and radiative forcing. *Geophysical Research Letters*, **52 (13)**, e2025GL116 615.

Chan, D., G. Gebbie, P. Huybers, and E. C. Kent, 2024a: A Dynamically Consistent ENsemble of Temperature at the Earth surface since 1850 from the DCENT dataset. *Scientific Data*, **11 (1)**, 953.

Chan, D., and P. Huybers, 2019: Systematic differences in bucket sea surface temperature measurements among nations identified using a linear-mixed-effect method. *Journal of Climate*, **32 (9)**, 2569–2589.

Chan, D., and P. Huybers, 2021: Correcting observational biases in sea surface temperature observations removes anomalous warmth during World War II. *Journal of Climate*, **34 (11)**, 4585–4602.

Chan, D., E. C. Kent, D. I. Berry, and P. Huybers, 2019: Correcting datasets leads to more homogeneous early-twentieth-century sea surface warming. *Nature*, **571 (7765)**, 393.

Chan, D., 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. *Science Advances*, **7 (26)**, eabg6931.

Chin, T. M., J. Vazquez-Cuervo, and E. M. Armstrong, 2017: A multi-scale high-resolution analysis of global sea surface temperature. *Remote sensing of environment*, **200**, 154–169.

Delworth, T. L., and T. R. Knutson, 2000: Simulation of early 20th century global warming. *Science*, **287 (5461)**, 2246–2250.

Deser, C., A. S. Phillips, and M. A. Alexander, 2010: Twentieth century tropical sea surface temperature trends revisited. *Geophysical Research Letters*, **37 (10)**.

Deser, C., and Coauthors, 2012: ENSO and Pacific decadal variability in the Community Climate System Model version 4. *Journal of Climate*, **25 (8)**, 2622–2651.

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. *Geophysical Research Letters*, **49 (24)**, e2022GL101 249.

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 Sensing of Environment*, **116**, 47–61, doi:10.1016/j.rse.2010.11.020, URL http://dx.doi.org/10.1016/j.rse.2010.11.020.

Embury, O., and Coauthors, 2024: Satellite-based time-series of sea-surface temperature since 1980 for climate applications. *Scientific Data*, **11 (1)**, 326.

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. *Geoscientific Model Development*, **9 (5)**, 1937–1958.

Eyring, V., and Coauthors, 2023: *Human influence on the climate system, Chapter 3 in Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]*, 423–552. Cambridge University Press, doi: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 Sensing of Environment*, **222**, 18–33, doi:10.1016/j.rse.2018.12.015, URL http://dx.doi.org/10.1016/j.rse.2018.12.015.

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

Folland, C. K., D. Parker, and F. Kates, 1984: Worldwide marine temperature fluctuations 1856–1981. *Nature*, **310 (5979)**, 670–673.

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

Gregory, J. M., and Coauthors, 2024: A new conceptual model of global ocean heat uptake. *Climate Dynamics*, **62 (3)**, 1669–1713.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146 (730)**, 1999–2049, doi:10.1002/qj.3803, URL http://dx.doi.org/10.1002/qj.3803.

Hirahara, S., M. Ishii, and Y. Fukuda, 2014: Centennial-scale sea surface temperature analysis and its uncertainty. *Journal of Climate*, **27 (1)**, 57–75.

Huang, B., 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. *Journal of Climate*, **34 (8)**, 2923–2939.

Huang, B., and Coauthors, 2015: Extended reconstructed sea surface temperature version 4 (ERSST. v4). Part I: Upgrades and intercomparisons. *Journal of Climate*, **28 (3)**, 911–930.

Huang, B., and Coauthors, 2017: Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons. *Journal of Climate*, **30 (20)**, 8179–8205.

Huang, B., and Coauthors, 2025: Extended reconstructed sea surface temperature, version 6 (ERSSTv6). part I: an artificial neural network approach. *Journal of Climate*, **38 (4)**, 1105–1121.

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. *Journal of Climate*, **21 (19)**, 5145–5153, doi:10.1175/2008jcli2292.1, URL http://dx.doi.org/10.1175/2008JCLI2292.1.

Ishii, M., A. Nishimura, S. Yasui, and S. Hirahara, 2025: Historical high-resolution daily SST analysis (COBE-SST3) with consistency to monthly land surface air temperature. *Journal of the Meteorological Society of Japan. Ser. II*, **103 (1)**, 17–44.

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. *International Journal of Climatology*, **25 (7)**, 865–879.

Jones, P. D., T. M. Wigley, and P. B. Wright, 1986: Global temperature variations between 1861 and 1984. *Nature*, **322 (6078)**, 430–434.

Kang, S. M., Y. Shin, H. Kim, S.-P. Xie, and S. Hu, 2023: Disentangling the mechanisms of equatorial pacific climate change. *Science Advances*, **9 (19)**, eadf5059.

Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856–1991. *Journal of Geophysical Research: Oceans*, **103 (C9)**, 18 567–18 589.

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. *Journal of Geophysical Research: Oceans*, **102 (C13)**, 27 835–27 860.

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

Kennedy, J., N. Rayner, C. Atkinson, and R. 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. *Journal of Geophysical Research: Atmospheres*, **124 (14)**, 7719–7763.

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

Kennedy, J., N. Rayner, R. Smith, D. Parker, and M. Saunby, 2011b: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. biases and homogenization. *Journal of Geophysical Research: Atmospheres*, **116 (D14)**.

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

Kent, E. C., and J. J. Kennedy, 2021: Historical estimates of surface marine temperatures. *Annual Review of Marine Science*, **13**, 283–311.

Kent, E. C., N. A. Rayner, D. I. Berry, M. Saunby, B. I. Moat, J. J. Kennedy, and D. E. Parker, 2013: Global analysis of night marine air temperature and its uncertainty since 1880: The HadNMAT2 data set. *Journal of Geophysical Research: Atmospheres*, **118 (3)**, 1281–1298.

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

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

Kosaka, Y., and Coauthors, 2024: The JRA-3Q Reanalysis. *Journal of the Meteorological Society of Japan. Ser. II*, **102 (1)**, 49–109, doi:10.2151/jmsj.2024-004, URL http://dx.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 and Atmospheric Science*, **5 (1)**, doi:10.1038/s41612-022-00301-2, URL http://dx.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 (5806)**, 1740–1745.

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

Merchant, C., P. Le Borgne, A. Marsouin, and H. Roquet, 2008a: Optimal estimation of sea surface temperature from split-window observations. *Remote Sensing of Environment*, **112 (5)**, 2469–2484, doi:10.1016/j.rse.2007.11.011, URL http://dx.doi.org/10.1016/j.rse.2007.11.011.

Merchant, C., and Coauthors, 2008b: Deriving a sea surface temperature record suitable for climate change research from the along-track scanning radiometers. *Advances in Space Research*, **41 (1)**, 1–11.

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. *Journal of Geophysical Research: Oceans*, **104 (C10)**, 23 565–23 578, doi:10.1029/1999jc900105, URL http://dx.doi.org/10.1029/1999JC900105.

Merchant, C. J., S. Saux-Picart, and J. Waller, 2020a: Bias correction and covariance parameters for optimal estimation by exploiting matched in-situ references. *Remote Sensing of Environment*, **237**, 111 590, doi:10.1016/j.rse.2019.111590, URL http://dx.doi.org/10.1016/j.rse.2019.111590.

Merchant, C. J., S. Saux-Picart, and J. Waller, 2020b: Bias correction and covariance parameters for optimal estimation by exploiting matched in-situ references. *Remote Sensing of Environment*, **237**, 111 590.

Mittaz, J., C. J. Merchant, and E. R. Woolliams, 2019: Applying principles of metrology to historical earth observations from satellites. *Metrologia*, **56 (3)**, 032 002.

Oliver, E. C., J. A. Benthuysen, S. Darmaraki, M. G. Donat, A. J. Hobday, N. J. Holbrook, R. W. Schlegel, and A. Sen Gupta, 2021: Marine heatwaves. *Annual Review of Marine Science*, **13 (1)**, 313–342.

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

Rayner, N., D. E. Parker, E. Horton, C. Folland, L. Alexander, D. Rowell, E. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, **108 (D14)**.

Reynolds, R. W., 1993: Impact of Mount Pinatubo aerosols on satellite-derived sea surface temperatures. *Journal of Climate*, **6 (4)**, 768–774.

Reynolds, R. W., 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. *Journal of Climate*, **26 (8)**, 2514–2533, doi:10.1175/jcli-d-12-00787.1, URL http://dx.doi.org/10.1175/JCLI-D-12-00787.1.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *Journal of Climate*, **15 (13)**, 1609–1625.

Reynolds, R. W., 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. *Journal of Climate*, **20 (22)**, 5473–5496.

Rohde, R. A., and Z. Hausfather, 2020: The Berkeley Earth land/ocean temperature record. *Earth System Science Data*, **12 (4)**, 3469–3479.

Sandford, C., and N. Rayner, in review: Addressing the World War 2 Warm Anomaly in HadSST.4.2.0.0 . *In review for International Journal of Climatology*.

Schneider, D. P., C. Deser, J. Fasullo, and K. E. Trenberth, 2013: Climate data guide spurs discovery and understanding. *Eos, Transactions American Geophysical Union*, **94 (13)**, 121–122.

Sherwood, S. C., and Coauthors, 2020: An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, **58 (4)**, e2019RG000 678.

Sippel, S., and Coauthors, 2024: Early-twentieth-century cold bias in ocean surface temperature observations. *Nature*, **635 (8039)**, 618–624.

Smith, T. M., 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. *Journal of Climate*, **11 (7)**, 1717–1729.

Smith, T. M., and R. W. Reynolds, 2002: Bias corrections for historical sea surface temperatures based on marine air temperatures. *Journal of Climate*, **15 (1)**, 73–87.

Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006). *Journal of Climate*, **21 (10)**, 2283–2296.

Teleti, P., E. Hawkins, and K. R. Wood, 2024: Digitizing weather observations from World War II US naval ship logbooks. *Geoscience Data Journal*, **11 (3)**, 314–329.

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

Thorne, P. W., D. E. Parker, J. R. Christy, and C. A. Mears, 2005: Uncertainties in climate trends: Lessons from upper-air temperature records. *Bulletin of the American Meteorological Society*, **86 (10)**, 1437–1442.

Trenberth, K. E., and D. J. Shea, 2006: Atlantic hurricanes and natural variability in 2005. *Geophysical Research Letters*, **33 (12)**.

Wittenberg, A. T., 2009: Are historical records sufficient to constrain ENSO simulations? *Geophysical Research Letters*, **36 (12)**.

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 Sensing*, **16 (18)**, 3358.

Yang, C., and Coauthors, 2021: Sea surface temperature intercomparison in the framework of the Copernicus Climate Change Service (C3S). *Journal of Climate*, **34 (13)**, 5257–5283, doi: 10.1175/jcli-d-20-0793.1, URL http://dx.doi.org/10.1175/JCLI-D-20-0793.1.