Geophysical Research Letters

RESEARCH LETTER
10.1029/2025GL116615

Re-Evaluating Historical Sea Surface Temperature Data Sets: Insights From the Diurnal Cycle, Coral Proxy Data, and Radiative Forcing

Duo Chan[1,2], **Geoffrey Gebbie**[2], **and Peter Huybers**[3]

[1]School of Ocean and Earth Science, University of Southampton, Southampton, UK, [2]Department of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, MA, USA, [3]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA

Key Points:

- Changes in the diurnal cycle of ship-based sea surface temperature (SST) measurements indicate that a wooden-to-canvas bucket transition occurred by 1910
- Leading SST products apply corrections for a later transition, leading to the appearance of excess 19c cooling and early 20c warming
- An SST product capturing the early bucket transition is more consistent with coral proxies and expected responses from anthropogenic forcing

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

D. Chan,
Duo.Chan@soton.ac.uk

Citation:

Chan, D., Gebbie, G., & Huybers, P. (2025). Re-evaluating historical sea surface temperature data sets: Insights from the diurnal cycle, coral proxy data, and radiative forcing. *Geophysical Research Letters*, *52*, e2025GL116615. https://doi.org/10.1029/2025GL116615

Received 3 MAY 2025
Accepted 10 JUN 2025

Author Contributions:

Conceptualization: Duo Chan, Geoffrey Gebbie, Peter Huybers
Data curation: Duo Chan
Formal analysis: Duo Chan
Funding acquisition: Geoffrey Gebbie, Peter Huybers
Methodology: Duo Chan
Resources: Peter Huybers
Software: Duo Chan
Supervision: Geoffrey Gebbie, Peter Huybers
Validation: Duo Chan
Visualization: Duo Chan
Writing – original draft: Duo Chan

Abstract Discrepancies in historical global mean surface temperature (GMST) estimates largely stem from differences in bias corrections applied to sea surface temperature (SST) records. Here, using the amplitude of the diurnal cycle in SST, we provide evidence that wooden-to-canvas bucket transitions were mostly complete by the early 1900s, earlier than commonly assumed by two decades, resulting in strong early 20th century cold biases. We then use this diurnal evidence, together with coral δ^{18}O and Sr/Ca proxies and expected GMST responses to external radiative forcing, to evaluate four different SST estimates: HadSST4, ERSST5, COBESST2, and DCSST. Of these, DCSST, developed by our team through adjustment of SSTs to match coastal land surface air temperatures, shows the closest overall agreement with all three lines of evidence. DCSST features a larger and steadier warming since the 1900s relative to the other estimates and indicates somewhat higher transient climate sensitivity and smaller decadal variability.

Plain Language Summary Sea surface temperature (SST) records are essential for tracking climate change. Differences in correcting historical SST biases, however, can bias global temperature trends by up to 0.1°C per century—a small but important amount when assessing warming limits and climate sensitivity. We evaluate four major SST data sets using three independent methods: day–night temperature differences, coral-based temperature proxies, and expected warming from greenhouse gases. These approaches indicate that most SST data sets underestimate early 20th-century biases from less-insulated canvas buckets, distorting historical trends. One data set, named DCSST and developed by our team, appears to correct this bias more fully than the others and shows steadier historical warming.

1. Introduction

Sea surface temperature (SST) is crucial for estimating historical global mean surface temperature (GMST), a metric widely used to quantify climate variability (Gulev et al., 2021; Kennedy et al., 2024), attribute anthropogenic climate change (Eyring et al., 2021), constrain climate sensitivity (Sherwood et al., 2020), and inform future projections (Cannon, 2025; Tokarska et al., 2020). Before 1940, most SSTs were measured using buckets, with temperatures recorded on deck (Kent & Taylor, 2006). During this measuring process, water temperatures are often altered by wind-induced evaporation or solar heating (Kent et al., 2010).

Correcting biases in SST observations is essential for purposes of documenting and interpreting historical SST variability. Two broad strategies are commonly used. The first relies on physics-based bucket models, where a transition from more-insulated wooden to less-insulated canvas buckets is typically assumed to occur beginning late 19th century and concluding in the 1920s (Folland & Parker, 1995). This approach underpins data sets such as the Centennial in-situ Observation-Based Estimate SST version 2 (COBESST2; Hirahara et al., 2014). The second strategy anchors SSTs to temperature records thought to be less biased. For example, the Extended Reconstructed SST version 5 (ERSST5) ensemble uses nighttime marine air temperatures (NMATs; Huang et al., 2017), while the Dynamically Consistent SST (DCSST; developed by our team) infers corrections from coastal weather stations temperatures (Chan et al., 2023, Chan, Gebbie, Huybers, et al., 2024). The Hadley Center SST version 4 (HadSST4; Kennedy et al., 2019) ensemble adopts a mixed approach before 1940, designed to span both strategies. Half of its 200 members rely solely on bucket models, while the other half blend model-based and NMAT-based corrections. HadSST4 also explores two bucket-transition patterns, with 100 members applying a

Writing – review & editing: Duo Chan, Geoffrey Gebbie, Peter Huybers

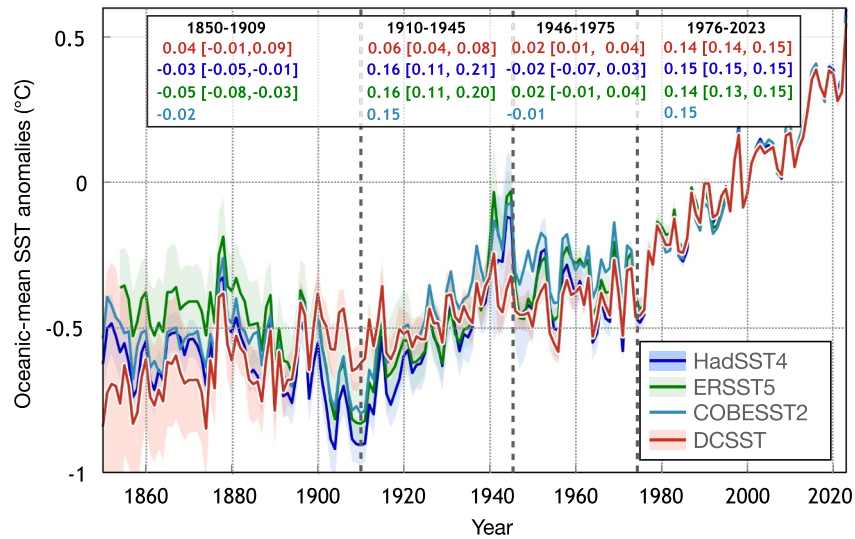

Figure 1. Ocean-mean sea surface temperature from different products. Anomalies are relative to the 1982–2014 mean, for HadSST4 (blue), ERSST5 (green), COBESST2 (teal), and DCSST (red), each masked to their least common spatial coverage before averaging. Listed values show linear trends (°C per decade) for 1850–1909, 1910–1945, 1946–1975, and 1976–2023. Shading and brackets denote 95% confidence intervals when ensembles are available (HadSST4: 200, ERSST5: 500, DCSST: 200 members). ERSST5 originally had 1,000 members, but only 500 extend beyond 2016.

gradual, linear change from 1850 to 1920, and the other 100 assuming a more abrupt transition centered around 1910.

Differences in bias corrections are the major source of divergence among these leading SST products, especially given that each is based upon very similar data deriving from the International Comprehensive Ocean-Atmosphere Data Set (ICOADS) (Freeman et al., 2017). When referenced to an 1980–2014 mean, SST anomalies before 1945 can differ by up to 0.3°C (Figure 1). HadSST4, ERSST5, and COBESST2 show early 20th century warming of ∼0.16°C per decade—30% faster than contemporaneous land warming (Sippel et al., 2024), which is physically inconsistent due to land's smaller heat capacity and lower water availability (Manabe et al., 1991). In contrast, DCSST indicates warming being only 0.06°C per decade over 1910–1945. Before 1910, DCSST shows a consistent warming of 0.04°C per decade while other products have cooling.

Given the discrepancies among SST estimates and their differing corrections, an obvious question arises: How well do existing SST products remove historical SST biases? We pursue this question by evaluating the four data sets using three independent lines of evidence. First, we introduce a diagnostic, the century-long evolution of the amplitude of observed SST diurnal-cycle, that offers a direct, physics-based fingerprint of bucket-type changes (Section 2). The other two approaches (Section 3) are complementary to those used in another recent study (Sippel et al., 2024) and involve comparing instrumental SSTs to independent paleo-proxy records from corals and against expected temperature changes from external radiative forcing. Taken together, these lines of evaluation provide practical guidance on the relative reliability and physical consistency among currently available SST data sets.

2. Amplitude of the SST Diurnal Cycle

Diurnal amplitudes are estimated using ICOADS SSTs by calculating sub-daily SST anomalies relative to daily means for individual ships based on the Carella et al. (2017) ship-track reconstructions. Note that ICOADS SSTs are ungridded in-situ measurements, distinct from gridded monthly products such as HadSST4. These anomalies are binned by local hour, month, and latitude, and a once-per-day sinusoidal function is fit via least squares to estimate amplitude (Chan & Huybers, 2020). These amplitudes are indicative of measurement methods and instrumental characteristics (Carella et al., 2018; Chan & Huybers, 2021). For example, engine-room-intake (ERI) measurements, which draw water from greater depth, record smaller diurnal amplitude but are, on average, warm biased due to engine heat (Carella et al., 2018). In contrast, less-insulated canvas buckets show larger diurnal amplitudes from daytime solar heating that peaks around noon, but cooler daily averages due to stronger wind-driven cooling throughout the day and night (Carella et al., 2018; Chan & Huybers, 2020; Folland

& Parker, 1995; Kent et al., 2010). Thus, the diurnal cycle permits us to infer which instruments were used even when meta-data is missing.

Diurnal amplitude and ship coverage vary by region, season, and time (Freeman et al., 2017; Kennedy et al., 2007). We focus on regions where diurnal cycles are strongest and most robust (Chan & Huybers, 2020; Morak-Bozzo et al., 2016): the tropics (20°S–20°N), subtropics (20°–40°) in summer/winter, and extratropical summer (40°–60°). The analysis covers 1880–2009, excluding earlier years due to sparse data and later years because the ship-track reconstruction from Carella et al. (2017) ends in 2009.

Anomalous diurnal amplitudes from ship observations are defined relative to a 1990–2014 climatology based on ICOADS SSTs from drifting buoys at each monthly 5° grid box (Chan & Huybers, 2019). Buoy-based amplitudes show minimal long-term changes (less than 0.002°C per decade since 1980), especially compared to the much larger variability observed in ship-based records (Figure S1 in Supporting Information S1). The stability of the buoy records supports using the relative anomalies from ship-based diurnal amplitudes as a diagnostic for measurement changes in historical SSTs.

Ship-based diurnal amplitude anomalies show clear temporal structure (Figures 2a–2d), rising rapidly from approximately 0.04°C in the 1880s to more than 0.1°C in the early 1900s, and then plateauing through the early 20th century. A sharp decrease to −0.03°C during the 1940s coincides with World War II when many engine-room-intake temperatures were measured (Chan & Huybers, 2021). After the 1950s, diurnal amplitudes remain nearly flat and likely reflect an inconsistent mix of measurement methods including buckets, ERI, and hull-based sensors across fleets and regions (Kennedy et al., 2011).

The change in diurnal amplitude before 1940 may offer insights into the poorly understood transition from more-insulated wooden to less-insulated canvas buckets in the late 19th century (Bottomley et al., 1990; Folland & Parker, 1995). The early 20th-century plateau suggests this transition completed earlier than the widely assumed date of 1920 (Folland & Parker, 1995; Kennedy et al., 2011). Accurately specifying this transition is important because canvas buckets can introduce cooling biases that are as much as 0.5°C larger than wooden buckets (Chan & Huybers, 2020; Folland & Parker, 1995). To quantitatively estimate the transition timing, we fit a piecewise linear model to individual diurnal amplitude series. The piecewise model consists of an upward rise beginning in 1880 that transitions to level values that persist to 1940. Residual errors from varying the transition year between 1890 and 1930 show a minimum between 1896 and 1910 that averages to 1901 (Figure 2e). This result is robust across different data groupings and supports the presence of an early transition from wooden to canvas buckets.

Diurnal amplitude variations also permit the evaluation of some assumptions in the corrections of gridded SST products, calculated as the difference between each product and a corresponding gridded field of unadjusted ICOADS SSTs. HadSST4 provides such an unadjusted SST estimates (Kennedy et al., 2019), but for other gridded estimates we compute a corresponding unadjusted gridded data set using ICOADS SSTs. ICOADS gridding is performed using the same methodology as used for gridding DCSST (Chan, Gebbie, Huybers, et al., 2024).

In the tropics, COBESST2 and the HadSST4 ensemble mean have increasing corrections until 1920, lagging behind the stabilization suggested by diurnal amplitudes (Figure 2a). The ERSST5 ensemble average, meanwhile, applies a nearly constant correction of 0.3°C from 1880 to 1940 and shows limited correspondence. DCSST corrections increase rapidly before 1900 and plateau around 0.5°C, better aligning with the rise and stabilization in diurnal amplitudes implied by variations in diurnal amplitude.

Pearson's correlation is further used to quantify agreement between variations in diurnal amplitude and bias correction. We note that correlations can be degraded by a number of factors, such as effective wind exposure that can alter daily-mean biases without much influencing diurnal amplitude (Chan & Huybers, 2020). Nevertheless, the correlations with diurnal amplitudes (1880–1940) are highest for DCSST (0.70), followed by HadSST4 (0.64), COBESST2 (0.49), and ERSST5 (0.08) in the tropics (Figure 2a). DCSST also shows relatively higher correlations across subtropical seasons (Figures 2b and 2c) and hemispheres (Figure S2 in Supporting Information S1). The overall better skill across multiple regions and seasons supports DCSST, and to a large extent also HadSST, as better capturing changes in instrumentation.

It is also useful to examine the distribution of correlations that one obtains comparing diurnal amplitudes against the ensemble of SST corrections available for DCSST, HadSST4, and ERSST5. Taking the tropics as an example,

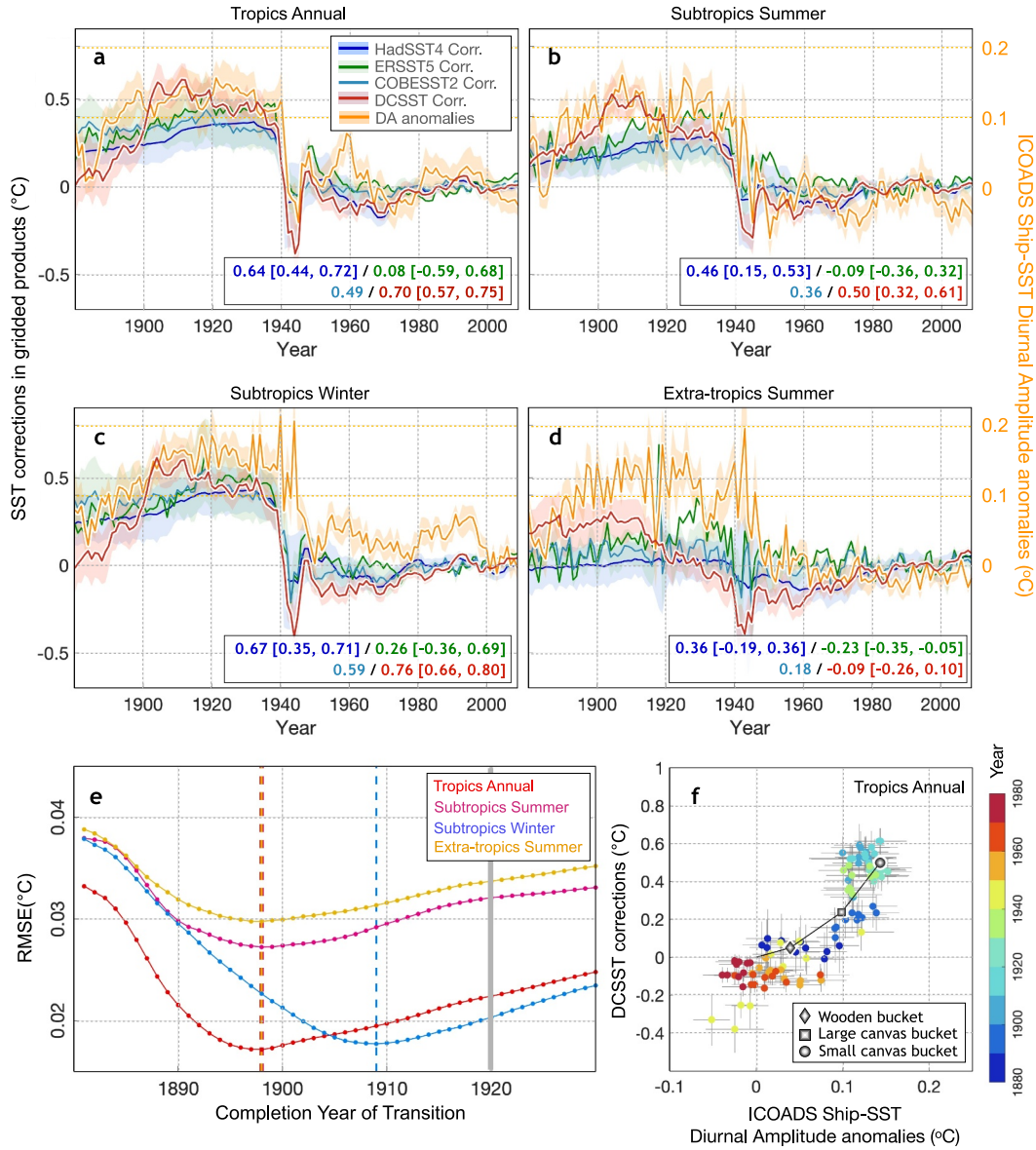

Figure 2. Sea surface temperature (SST) corrections and the amplitude of the diurnal cycle. (a) Tropical (20°S-20°N) product SST corrections relative to the 1982–2009 mean correction (left *y*-axis) in HadSST4 (blue), ERSST5 (green), COBESST2 (teal), and DCSST (red), alongside anomalous International Comprehensive Ocean-Atmosphere Data Set Ship-SST diurnal amplitude relative to a 1900–2014 buoy climatology (orange, right *y*-axis). Shading represents 95% confidence intervals. Pearson's correlation with diurnal amplitudes (HadSST4 | ERSST5 | COBESST2 | DCSST; 1880–1940) are listed at the bottom. (b–d) as (a) but for (b) subtropical summer (20°-40°; JJA for NH and DJF for SH) (c) subtropical winter (DJF for NH and JJA for SH), and (d) extra-tropical summer (40°-60°). Values are calculated separately by hemisphere and averaged. (e) Residual errors from segmented linear fits to diurnal amplitude time series as a function of transition year. Dashed lines indicate regional best-fit transition years; the solid gray line marks the commonly assumed transition year of 1920. (f) Tropical DCSST corrections (*y*-axis) versus diurnal amplitude anomalies (*x*-axis; colored dots), compared with simulated daily-mean biases against changes in diurnal amplitude for wooden (diamond), large canvas (square), and small canvas (circle) buckets. Simulated bias signs are reversed to indicate required corrections. Yellow, orange, and red markers represent a mixture of bucket, ERI, and hull sensor measurements after 1940, giving some negative corrections and diurnal amplitude anomalies outside the range bucket models can capture.

DCSST shows the highest consistency (95% c.i.: [0.57, 0.75]; Figure S3 in Supporting Information S1). HadSST4 correlations range from 0.44 to 0.72, with members assuming a linear bucket transition to 1920 averaging 0.69, compared to 0.58 for those with a more abrupt 1910 shift (Figure S3 in Supporting Information S1). This result suggests that an early but gradual transition is more realistic. ERSST5 spans a wider range (95% c.i.: [–0.59, 0.68]; Figure 2a) with a distinct tri-modal structure centered, respectively, at −0.5, 0.1, and 0.6 (Figure S3 in Supporting Information S1). This wide spread likely arises from different NMAT sources and reflects uncertainties that could be further constrained.

In the extra-tropics, overall correlations between diurnal amplitude and temperature corrections are low, particularly during Northern Hemisphere summer. Because bucket heating by sunlight is almost exactly canceled by wind-driven cooling for all bucket types (Figure S4C in Supporting Information S1), most products do not apply much correction during extra-tropical summer. DCSST, however, applies a 0.2–0.3°C correction between 1880 and 1910 (Figure 2d) and has the lowest correlation with diurnal amplitudes. We hypothesize that DCSST likely includes a spurious warm offset for methodological reasons involving the fact that a bucket-bias pattern is rescaled without estimating an intercept before the 1930s (Chan, Gebbie, Huybers, et al., 2024). Under this approach, any offset present in the satellite-based climatology is treated as a bucket bias, with a scaling factor determined primarily by the data-rich tropical and subtropical regions. Because extra-tropical summertime bucket SSTs are expected to be less biased due to the net effect of increased solar heating and reduced evaporative cooling (Folland & Parker, 1995), this climatological offset persists and appears as a positive correction relative to later periods (Figure 2d). Further analysis of extra-tropical trends is warranted including with respect to using a more complete bias model.

To further examine whether the relationship between diurnal amplitudes and SST corrections is physically consistent with known biases, we use a process-based bucket model simulating sub-daily and daily-mean SST biases for different bucket types (Chan & Huybers, 2020). Driven by modern climatological inputs, the model accounts for solar heating, evaporative cooling, and heat exchange with air (see Table S1 in Supporting Information S1 for model parameters). When a wooden-to-canvas bucket transition is imposed, the model reproduces the observed relationship between diurnal amplitude and DCSST correction in the tropics (Figure 2f) and subtropics (Figure S4 in Supporting Information S1). These results further support the plausibility of an early gradual transition and highlight the importance of the transition pattern and timing.

3. Additional Lines of Evidence From Proxies and Radiative Forcing

Having established evidence for early SST biases using diurnal amplitudes, we next assess SST corrections using coral-based $\delta^{18}O$ and Sr/Ca records. Complementing previous work (Sippel et al., 2024) that used basin-scale reconstructions from older proxy compilations (Tierney et al., 2015), we directly compare co-located coral $\delta^{18}O$ and Sr/Ca records to gridded instrumental SSTs at 5° resolution, allowing for a more localized comparison without spatial interpolation. We also use more recent and extensive proxy archives, that is, PAGES2k2017 (Consortium et al., 2017) and Iso2kV1.0.0 (Konecky et al., 2020).

We begin with 196 annually and sub-annually resolved coral records, retaining 96 that overlap with instrumental SSTs for at least 30 years since 1910. Proxy signals in coral skeleton covary negatively with temperatures (Figure S5 in Supporting Information S1) but are also influenced by other factors such as source water properties, including salinity (Lee & Fung, 2008). To ensure temperature sensitivity, we retain only proxies with a Spearman's rank correlation more negative than −0.4 with local instrumental SSTs after 1950 (Spearman, 1910). When multiple instrumental SST products are available, correlations are averaged. This screening yields 26 proxy records: 21 $\delta^{18}O$ and 5 Sr/Ca, all located in the tropics (Figure 3a). As a result, this evaluation does not inform DCSST's extratropical summertime discrepancies noted earlier. Each selected proxy is scaled to match the amplitude of local instrumental temperatures using the ratio of their 1950–2000 standard deviations. This approach is equivalent to total least squares (Golub & Van Loan, 1980), accounting for uncertainty in both proxies and observations. Signs are reversed to express proxy anomalies in temperature units (Figure S5 in Supporting Information S1).

Upon averaging over grid boxes containing both proxy and instrumental records, these proxy-based temperatures indicate a steady warming from 1850 to 1940, a pattern that is slightly more consistent with DCSST than other estimates (Figures 3b–3e). Quantitatively, the Pearson's correlation between 20-year low-pass filtered instrumental and proxy SSTs is 0.91 [0.78, 0.94] for DCSST, 0.87 [0.77, 0.92] for HadSST4, 0.87 for COBESST2, and only 0.67

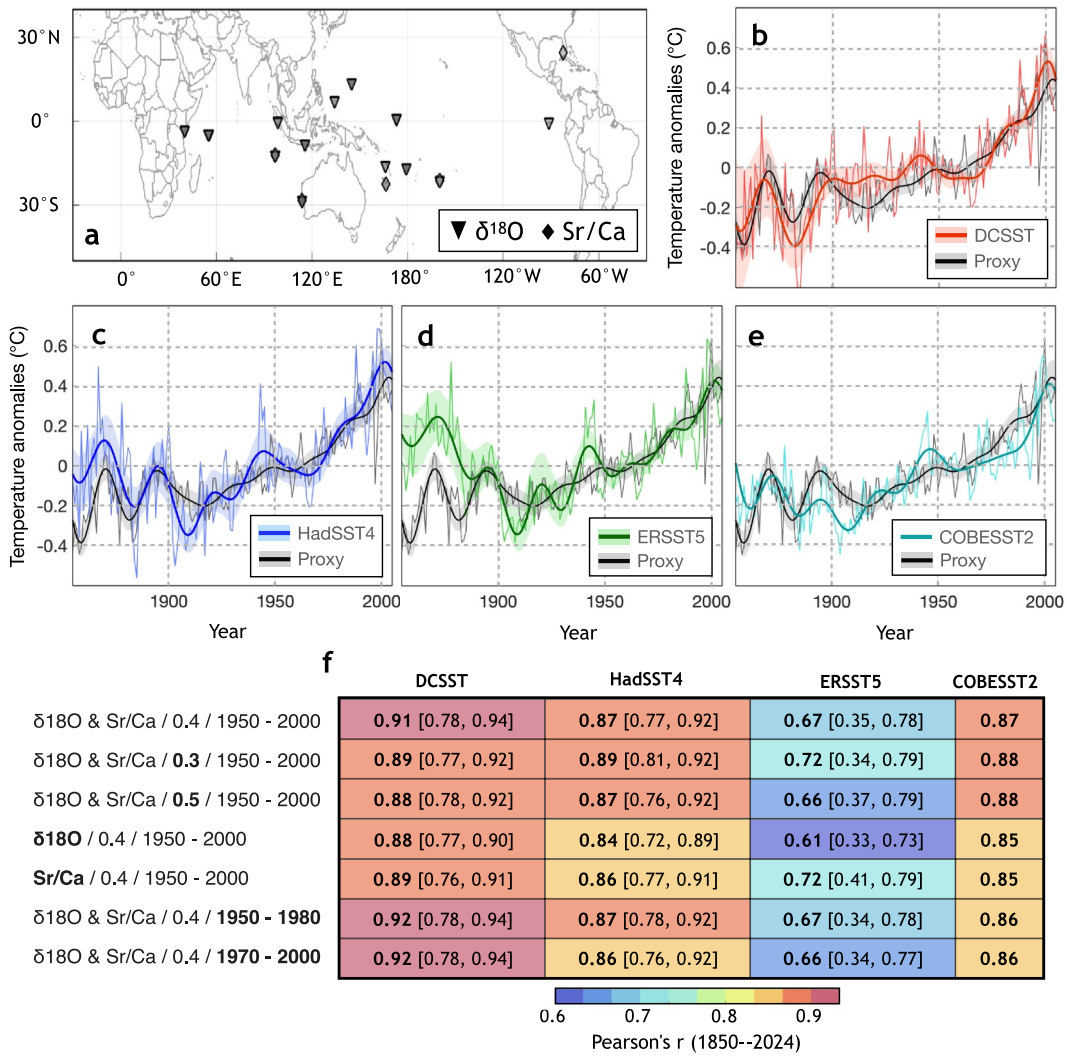

Figure 3. Comparison with paleo-proxies. (a) Temperature-indicating coral records used, including 21 δ^{18}O (triangles) and 5 Sr/Ca (diamonds). (b) DCSST anomalies (red) and coral-based reconstructions (black), averaged over all sites with both data types. To highlight low-frequency variability, annual signals (light lines) are 20-year low-pass filtered (dark lines) using a Fourier filter (Gonzalez, 2009). Anomalies are relative to 1930–1980 mean. Shading denotes 95% confidence intervals from DCSST's ensemble spreads and proxy calibration residuals. (c–e) as in panel (b), but for (c) HadSST4 (d) ERSST5, and (e) COBESST2. (f) Pearson's correlation between instrumental and proxy SSTs across sensitivity tests. Numbers in brackets indicate the 95% c.i. across ensemble members. Experiment labels indicate proxy type, correlation threshold, and calibration period.

[0.35, 0.78] for ERSST5 (Figure 3f). An evaluation of root-mean-square error gives qualitatively similar results (Figure S6 in Supporting Information S1). These proxy results are robust to a variety of plausible methodological differences, including using different correlation cut-offs for proxy selection, calibration intervals, and comparing against only Sr/Ca or δ^{18}O records (Figures 3f, S6, and S7 in Supporting Information S1). Despite methodological differences, our localized proxy-instrument comparison yields conclusions similar to Sippel et al. (2024), that is, the 1910s cooling in several instrumental SST data sets is not supported by proxy records.

As our final evaluation, we assess how well GMSTs incorporating different SST estimates align with temperatures expected from external radiative forcing. We compare GMST estimates from DCENT (using DCSST), HadCRUT5 analysis (using HadSST4; Morice et al., 2021), and NOAA Global Temp V5 (using ERSST5; Zhang et al., 2020). Because early observations are sparse over land, high latitudes, and the Pacific, spatial infilling is critical. NOAA Global Temp is spatially infilled to provide global coverage. For DCENT and HadCRUT5, which

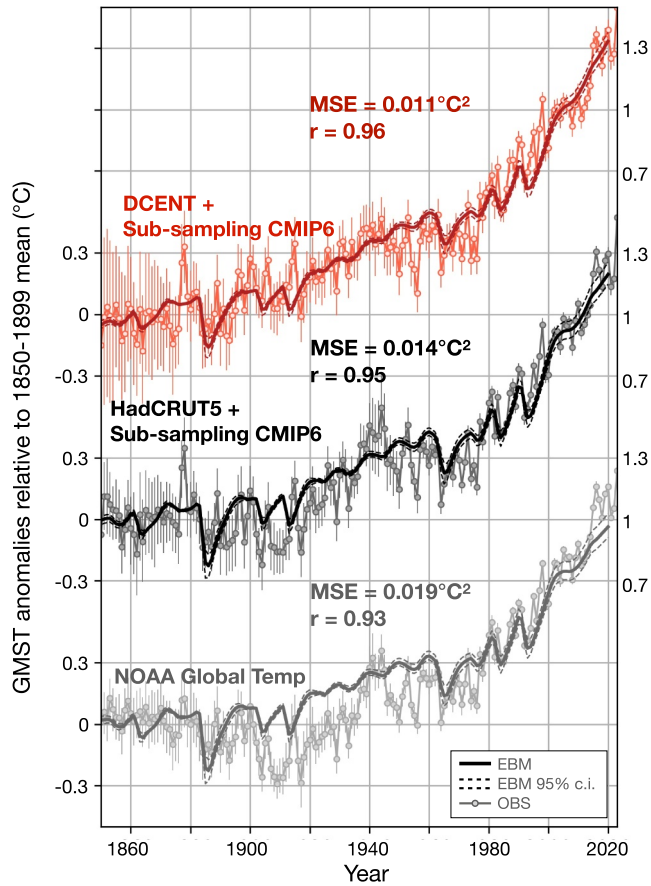

Figure 4. Energy-balance model (EBM) simulations. Observed global mean surface temperature (GMST) (markers) and 2-box EBM simulations (solid curve). From top to bottom, DCENT (with sampling uncertainty from CMIP6 masking), HadCRUT5 analysis, and NOAA Global Temperature 5. Dashed lines show 95% Bayesian confidence intervals. Error bars on markers indicate 95% c.i. for sampling uncertainty and ensemble spread. Listed numbers are the mean square error and Pearson's correlation (r) between observations and modeled GMST.

are not globally complete, we correct for sampling bias by comparing full CMIP6 simulations with those masked to observational coverage (see Text S1, Figures S8 and S9 in Supporting Information S1).

We use a two-box Energy-balance model (EBM) (Geoffroy et al., 2013) to estimate expected GMST responses to changes in external radiative forcing. This widely used approach (Haustein et al., 2019) provides a simple yet effective representation of ocean heat uptake, which governs the transient climate response (Geoffroy et al., 2013). The model equations are,

$$\begin{aligned} c_p \rho d_s \frac{dT_s}{dt} &= -\lambda T_s + F - \kappa(T_s - T_d), \\ c_p \rho d_d \frac{dT_d}{dt} &= \kappa(T_s - T_d), \end{aligned} \tag{1}$$

where T_s and T_d are surface and deep ocean temperatures. The heat capacity of seawater (c_p) is 4,180 J/kg/°C, and seawater density (ρ) is 1,030 kg/m^3. The effective layer depths (d_s, d_d), climate feedback parameter (λ), and ocean heat uptake coefficient (κ) are to be conditioned using observations. External radiative forcing (F) is prescribed from the year 1500, following Haustein et al. (2019).

Compared to CMIP6 models used in Sippel et al. (2024), the EBM offers a complementary perspective by explicitly conditioning key parameters to observational data sets, enabling a more flexible evaluation of SST corrections. Specifically, we use Bayesian inference to condition model parameters (d_s, d_d, λ, κ) on each GMST

estimate and observed ocean heat content estimate since 1960 (Cheng et al., 2017). Normal priors, $N(1, 0.5)$, that are truncated to be greater than zero are prescribed for λ and κ, as well as $N(100, 50)$ for d_s and $N(1000, 500)$ for d_d. Exponential priors are prescribed for σ_s (mean of 1°C) and σ_d (mean of 1×10^{23}J). We implement the model in PyMC (Abril-Pla et al., 2023) using Hamiltonian Markov Chain Monte Carlo with a No-U-Turn sampler (Hoffman & Gelman, 2014).

All data sets capture the overall warming trend as reproduced by the EBM (Figure 4), but only DCENT aligns with the model's continuous warming through 1850 to 1940. In contrast, HadCRUT5 and NOAA Global Temperature V5 show a 1910s cooling followed by a rapid warming into World War II. Quantitatively, DCENT shows the lowest mean-square error (0.011°C^2), compared to 0.014°C^2 for HadCRUT5 ($p < 0.15$) and 0.019°C^2 for NOAA ($p < 0.01$), and the highest correlation ($r = 0.96$), despite the model being re-tuned to each individual GMST estimate. When restricted to pre-1940, DCENT's correlation (0.81) remains strongest, compared with 0.72 for HadCRUT5 and 0.52 NOAA Global TempV5.

Notable is that both the EBM and DCENT indicate that the GMST warming between 1976 and 2023 is twice as fast as that between 1910 and 1945, whereas other estimates indicate that recent warming is only 25% faster. Although aerosol radiative forcing remains uncertain for interpreting early 20th-century temperature variations (Albright et al., 2021), the closer correspondence between the EBM results and the DCENT estimate suggests that GMST variations may be more directly attributable to external forcing with a smaller role of internal climate variability (Haustein et al., 2019).

4. Discussion and Conclusions

In this study, we used three independent lines of evidence, diurnal-cycle changes, coral proxy records, and an EBM, to reassess early 20th-century SST bias corrections. Among these, the evolution of diurnal amplitude covaries most closely with DCSST's corrections and suggests a gradual wooden-to-canvas bucket transition completed earlier than the previously assumed 1920 date. The coral and EBM approaches also suggest somewhat better agreement with DCSST as compared to other products. Relative to other SST estimates, DCSST indicates that historical warming was steadier and that the pre-industrial baseline was cooler, both of which have implications for understanding decadal variability (Haustein et al., 2019), climate sensitivity (Proistosescu & Huybers, 2017), and policy-relevant warming thresholds (Cannon, 2025).

Several caveats relating to DCSST deserve note. First, most of our evaluations are determined by variability on decadal and longer timescales. Given noise and limited record length, differences between SST products are not necessarily statistically significant, especially between DCSST and HadSST4. Second, EBM parameters are tuned to each GMST series, making the model–data check a consistency test rather than a fully independent evaluation. More generally, the DCSST corrections comes at the expense of losing independence between land and ocean data archives. The quality of DCSST, therefore, relies on the accuracy and representativeness of coastal station measurements that require their own bias adjustments and have limited coverage in early records (Chan, Gebbie, & Huybers, 2024; Cowtan et al., 2018; Menne et al., 2018). Finally, the lack of accounting for offsets in climatological estimates has led to non-physical adjustments in the summertime extra-tropics that can be addressed in future updates.

It is useful to further consider some specific issues in other SST estimates as well. The evolution of the diurnal amplitude helps explain the residual cold bias in the HadSST4 ensemble and COBESST2, which assume either a gradual bucket-type transition ending in 1920 or an earlier but more abrupt transition. For the ERSST5 ensemble, biases remain because NMAT itself shows a minimum in the 1910s. A cause may be the late-19th-century transition from coal-to oil-powered ships, which produce less deck heat and may lead to cooler NMAT readings on deck (Fletcher, 1975; Gray, 2017). This potential NMAT cooling bias could covary with SST errors associated with bucket-type transitions, leading to SST biases being incorrectly retained. On a separate note, neither ERSST nor HadSST ensembles directly account for the cold bias due to truncation of decimal points in the Japanese Kobe data collection (Chan et al., 2019), which affects regional SST patterns and has been adjusted in DCSST (Chan, Gebbie, Huybers, et al., 2024).

Improving historical SST estimates requires integrating multiple independent lines of evidence, as no single correction method fully resolves uncertainties. To further advance toward this goal, future work could also incorporate additional constraints such as sea-level pressure records, expand temporal coverage by integrating

more coral proxy data (Pfeiffer et al., 2017), and leverage isotope-enabled climate models (e.g., Nusbaumer et al., 2017) for a more comprehensive assessment of historical climate variations.

Data Availability Statement

All data used in this study and code used to generate them are available at Harvard Dataverse via https://doi.org/10.7910/DVN/OHCGPN under a CC0 1.0 License (Chan et al., 2025).

Acknowledgments
We thank Elizabeth C. Kent for discussing the transition from coal-to oil-powered ships. G.G. is supported by NSF OCE-2122805 and OCE-2103049. P.H. was supported by NSF Grant 2123295.

References

Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesbeck, C. J., Kochurov, M., et al. (2023). PyMC: A modern, and comprehensive probabilistic programming framework in Python. *PeerJ Computer Science*, *9*, e1516. https://doi.org/10.7717/peerj-cs.1516

Albright, A. L., Proistosescu, C., & Huybers, P. (2021). Origins of a relatively tight lower bound on anthropogenic aerosol radiative forcing from Bayesian analysis of historical observations. *Journal of Climate*, *34*(21), 8777–8792. https://doi.org/10.1175/jcli-d-21-0167.1

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

Cannon, A. J. (2025). Twelve months at 1.5°C signals earlier than expected breach of Paris agreement threshold. *Nature Climate Change*, *15*(3), 1–4. https://doi.org/10.1038/s41558-025-02247-8

Carella, G., Kennedy, J., Berry, D., Hirahara, S., Merchant, C. J., Morak-Bozzo, S., & Kent, E. (2018). Estimating sea surface temperature measurement methods using characteristic differences in the diurnal cycle. *Geophysical Research Letters*, *45*(1), 363–371. https://doi.org/10.1002/2017gl076475

Carella, G., Kent, E. C., & Berry, D. I. (2017). A probabilistic approach to ship voyage reconstruction in ICOADS. *International Journal of Climatology*, *37*(5), 2233–2247. https://doi.org/10.1002/joc.4492

Chan, D., Gebbie, G., & Huybers, P. (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. https://doi.org/10.1175/jcli-d-22-0569.1

Chan, D., Gebbie, G., & Huybers, P. (2024). An improved ensemble of land-surface air temperatures since 1880 using revised pair-wise homogenization algorithms accounting for autocorrelation. *Journal of Climate*, *37*(7), 2325–2345. https://doi.org/10.1175/jcli-d-23-0338.1

Chan, D., Gebbie, G., Huybers, P., & Kent, E. C. (2024). A dynamically consistent ENsemble of temperature at the earth surface since 1850 from the DCENT dataset. *Scientific Data*, *11*(1), 953. https://doi.org/10.1038/s41597-024-03742-x

Chan, D., Geoffrey, G., & Huybers, P. (2025). Replication data and code for "re-evaluating historical SST Datasets: Insights from diurnal cycle, bucket models, paleo-records, and radiative forcing" [Dataset]. *Harvard Dataverse*. https://doi.org/10.7910/DVN/OHCGPN

Chan, D., & Huybers, P. (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. https://doi.org/10.1175/jcli-d-18-0562.1

Chan, D., & Huybers, P. (2020). Systematic differences in bucket sea surface temperatures caused by misclassification of engine room intake measurements. *Journal of Climate*, *33*(18), 7735–7753. https://doi.org/10.1175/jcli-d-19-0972.1

Chan, D., & Huybers, P. (2021). Correcting observational biases in sea surface temperature observations removes anomalous warmth during World War II. *Journal of Climate*, *34*(11), 4585–4602. https://doi.org/10.1175/jcli-d-20-0907.1

Chan, D., Kent, E. C., Berry, D. I., & Huybers, P. (2019). Correcting datasets leads to more homogeneous early-twentieth-century sea surface warming. *Nature*, *571*(7765), 393–397. https://doi.org/10.1038/s41586-019-1349-2

Cheng, L., Trenberth, K. E., Fasullo, J., Boyer, T., Abraham, J., & Zhu, J. (2017). Improved estimates of ocean heat content from 1960 to 2015. *Science Advances*, *3*(3), e1601545. https://doi.org/10.1126/sciadv.1601545

Consortium, P., McKay, N. P., Kaufman, D. S., von Gunten, L., Wang, J., Anchukaitis, K. J., et al. (2017). A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data*, *4*(1), 170088. https://doi.org/10.1038/sdata.2017.88

Cowtan, K., Rohde, R., & Hausfather, Z. (2018). Evaluating biases in sea surface temperature records using coastal weather stations. *Quarterly Journal of the Royal Meteorological Society*, *144*(712), 670–681. https://doi.org/10.1002/qj.3235

Eyring, V., Gillett, N. P., Achuta Rao, K. M., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., et al. (2021). Human influence on the climate system. (Chapter 3).

Fletcher, M. E. (1975). From coal to oil in British shipping. *The Journal of Transport History*, *ss-3*(1), 1–19. https://doi.org/10.1177/002252667500300101

Folland, C., & Parker, D. (1995). Correction of instrumental biases in historical sea surface temperature data. *Quarterly Journal of the Royal Meteorological Society*, *121*(522), 319–367. https://doi.org/10.1002/qj.49712152206

Freeman, E., Woodruff, S. D., Worley, S. J., Lubker, S. J., Kent, E. C., Angel, W. E., et al. (2017). ICOADS release 3.0: A major update to the historical marine climate record. *International Journal of Climatology*, *37*(5), 2211–2232. https://doi.org/10.1002/joc.4775

Geoffroy, O., Saint-Martin, D., Olivié, D. J., Voldoire, A., Bellon, G., & Tytéca, S. (2013). Transient climate response in a two-layer energy-balance model. Part I: Analytical solution and parameter calibration using CMIP5 AOGCM experiments. *Journal of Climate*, *26*(6), 1841–1857. https://doi.org/10.1175/jcli-d-12-00195.1

Golub, G. H., & Van Loan, C. F. (1980). An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis*, *17*(6), 883–893. https://doi.org/10.1137/0717073

Gonzalez, R. C. (2009). *Digital image processing*. Pearson Education India.

Gray, S. (2017). Fuelling mobility: Coal and Britain's naval power, c. 1870–1914. *Journal of Historical Geography*, *58*, 92–103. https://doi.org/10.1016/j.jhg.2017.06.013

Gulev, S. K., Thorne, P. W., Ahn, J., Dentener, F. J., Domingues, C. M., Gerland, S., et al. (2021). Changing state of the climate system.

Haustein, K., Otto, F. E., Venema, V., Jacobs, P., Cowtan, K., Hausfather, Z., et al. (2019). A limited role for unforced internal variability in twentieth-century warming. *Journal of Climate*, *32*(16), 4893–4917. https://doi.org/10.1175/jcli-d-18-0555.1

Hirahara, S., Ishii, M., & Fukuda, Y. (2014). Centennial-scale sea surface temperature analysis and its uncertainty. *Journal of Climate*, *27*(1), 57–75. https://doi.org/10.1175/jcli-d-12-00837.1

Hoffman, M. D., & Gelman, A. (2014). The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*(1), 1593–1623.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017). Extended reconstructed sea surface temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *Journal of Climate*, *30*(20), 8179–8205. https://doi.org/10.1175/jcli-d-16-0836.1

Kennedy, J., Brohan, P., & Tett, S. (2007). A global climatology of the diurnal variations in sea-surface temperature and implications for MSU temperature trends. *Geophysical Research Letters*, *34*(5). https://doi.org/10.1029/2006gl028920

Kennedy, J., Rayner, N., Atkinson, C., & Killick, R. (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. https://doi.org/10.1029/2018jd029867

Kennedy, J., Rayner, N., Smith, R., Parker, D., & Saunby, M. (2011). 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), D14104. https://doi.org/10.1029/2010jd015220

Kennedy, J., Trewin, B., Betts, R., Thorne, P., Foster, P., Siegmund, P., et al. (2024). State of the climate 2024. update for cop29.

Kent, E. C., Kennedy, J. J., Berry, D. I., & Smith, R. O. (2010). Effects of instrumentation changes on sea surface temperature measured in situ. *Wiley Interdisciplinary Reviews: Climate Change*, *1*(5), 718–728. https://doi.org/10.1002/wcc.55

Kent, E. C., & Taylor, P. K. (2006). Toward estimating climatic trends in SST. Part I: Methods of measurement. *Journal of Atmospheric and Oceanic Technology*, *23*(3), 464–475. https://doi.org/10.1175/jtech1843.1

Konecky, B. L., McKay, N. P., Churakova, O. V., Comas-Bru, L., Dassié, E. P., Delong, K. L., et al. (2020). The Iso2k database: A global compilation of paleo-δ 18 O and δ 2 H records to aid understanding of common era climate. *Earth System Science Data Discussions*, *2020*, 1–49.

Lee, J.-E., & Fung, I. (2008). "Amount effect" of water isotopes and quantitative analysis of post-condensation processes. *Hydrological Processes: International Journal*, *22*(1), 1–8. https://doi.org/10.1002/hyp.6637

Manabe, S., Stouffer, R. J., Spelman, M. J., & Bryan, K. (1991). Transient responses of a coupled ocean–atmosphere model to gradual changes of atmospheric CO2. Part I. Annual mean response. *Journal of Climate*, *4*(8), 785–818. https://doi.org/10.1175/1520-0442(1991)004<0785:troaco>2.0.co;2

Menne, M. J., Williams, C. N., Gleason, B. E., Rennie, J. J., & Lawrimore, J. H. (2018). The global historical climatology network monthly temperature dataset, version 4. *Journal of Climate*, *31*(24), 9835–9854. https://doi.org/10.1175/jcli-d-18-0094.1

Morak-Bozzo, S., Merchant, C., Kent, E., Berry, D., & Carella, G. (2016). Climatological diurnal variability in sea surface temperature characterized from drifting buoy data. *Geoscience Data Journal*, *3*(1), 20–28. https://doi.org/10.1002/gdj3.35

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J., Hogan, E., Killick, R., et al. (2021). An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set. *Journal of Geophysical Research: Atmospheres*, *126*(3), e2019JD032361. https://doi.org/10.1029/2019jd032361

Nusbaumer, J., Wong, T. E., Bardeen, C., & Noone, D. (2017). Evaluating hydrological processes in the Community Atmosphere Model Version 5 (CAM5) using stable isotope ratios of water. *Journal of Advances in Modeling Earth Systems*, *9*(2), 949–977. https://doi.org/10.1002/2016ms000839

Pfeiffer, M., Zinke, J., Dullo, W.-C., Garbe-Schönberg, D., Latif, M., & Weber, M. (2017). Indian Ocean corals reveal crucial role of World War II bias for twentieth century warming estimates. *Scientific Reports*, *7*(1), 14434. https://doi.org/10.1038/s41598-017-14352-6

Proistosescu, C., & Huybers, P. (2017). Slow climate mode reconciles historical and model-based estimates of climate sensitivity. *Science Advances*, *3*(7), e1602821. https://doi.org/10.1126/sciadv.1602821

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, *58*(4), e2019RG000678. https://doi.org/10.1029/2019rg000678

Sippel, S., Kent, E. C., Meinshausen, N., Chan, D., Kadow, C., Neukom, R., et al. (2024). Early-twentieth-century cold bias in ocean surface temperature observations. *Nature*, *635*(8039), 618–624. https://doi.org/10.1038/s41586-024-08230-1

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, *3*(3), 271–295. https://doi.org/10.1111/j.2044-8295.1910.tb00206.x

Tierney, J. E., Abram, N. J., Anchukaitis, K. J., Evans, M. N., Giry, C., Kilbourne, K. H., et al. (2015). Tropical sea surface temperatures for the past four centuries reconstructed from coral archives. *Paleoceanography*, *30*(3), 226–252. https://doi.org/10.1002/2014pa002717

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., & Knutti, R. (2020). Past warming trend constrains future warming in CMIP6 models. *Science Advances*, *6*(12), eaaz9549. https://doi.org/10.1126/sciadv.aaz9549

Zhang, H., Huang, B., Lawrimore, J., Menne, M., & Smith, T. (2020). *NOAA global surface temperature dataset (NOAAGlobalTemp)*. Version.

References From the Supporting Information

Cowtan, K., & Way, R. G. (2014). Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society*, *140*(683), 1935–1944. https://doi.org/10.1002/qj.2297

Rohde, R. A., & Hausfather, Z. (2020). The Berkeley Earth land/ocean temperature record. *Earth System Science Data*, *12*(4), 3469–3479. https://doi.org/10.5194/essd-12-3469-2020