https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

Open Access Earth System Science Data Discussions

Berkeley Earth Surface Temperature - High-Resolution (BEST-HR): a 0.25° Global Gridded Temperature Data Set for Climate Monitoring

Robert A. Rohde[1], Bruce Calvert[1], Zeke Hausfather[1,2]

[1]Berkeley Earth, Berkeley, California, USA
[2]Stripe, San Francisco, California, USA

Correspondence to: Robert Rohde (robert@berkeleyearth.org)

Abstract. We present version 2 of the Berkeley Earth Surface Temperature dataset, a global observational temperature record that now features enhanced spatial resolution (0.25° × 0.25°) and a new regression Kriging framework that incorporates physically informed predictive fields. The land analysis uses 58,311 weather stations, updated to reflect the latest source archives, and includes separate reconstructions for TAVG (1750–present), TMAX, and TMIN (both 1850–present). Sea surface temperatures are derived from HadSST4, with spatial detail improved using predictors from the ORAS5 ocean reanalysis. The land and ocean fields are merged into a unified product with a refined coastal blending method. A new uncertainty ensemble is introduced to quantify reconstruction and coverage uncertainty, enabling propagation of errors into derived metrics. The resulting Berkeley Earth Surface Temperature – High-Resolution (BEST-HR) dataset offers a globally complete, monthly resolved temperature product suitable for applications ranging from global trend analysis to regional climate impact assessments. The current BEST-HR dataset described in this paper has been archived at Zenodo DOI: 10.5281/zenodo.20428381 alongside the merged weather station data used to generate it DOI: 10.5281/zenodo.20427092. Monthly updates to these datasets will be provided primarily through the Berkeley Earth website.

1 Introduction

Gridded global temperature data sets that combine land-surface air temperature (LSAT) with sea-surface temperature (SST) underpin almost every quantitative assessment of modern climate change. They form the baseline for tracking the pace of warming, estimating equilibrium climate sensitivity, detecting shifts in extremes, and evaluating the performance of coupled climate models such as those submitted to CMIP6. Recognising this pivotal role, the Global Climate Observing System (GCOS) designates LSAT and SST as essential climate variables (ECVs). Yet the most widely used products were designed for global or continental-scale diagnostics with spatial resolutions (2°–5° latitude × longitude) that are too coarse for many emerging applications–including regional attribution studies, urban-heat assessment, impact modelling, and place-based adaptation or mitigation planning.

Here we introduce Berkeley Earth Surface Temperature – High-Resolution (BEST-HR), a 0.25° × 0.25° global data set that delivers monthly LSAT for mean, max, and min temperature and combined land-and-ocean mean temperature fields from 1750 (land) / 1850 (land and ocean) to the present. BEST-HR is built on 58,311 quality-controlled surface stations and the latest HadSST4 marine observations. These are mapped with a regression-Kriging framework that leverages frequently recurring spatial weather patterns extracted from ERA5 reanalysis and high-resolution topography. Separate land and ocean analyses are first performed–respectively honouring station metadata, lapse-rate effects and marine bias adjustments–and are then merged through a weighted least-squares blend across the coastline buffer to ensure physical continuity.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

1.1 Existing global temperature products

The previous Berkeley Earth data set (Rohde et al., 2013a,b; Rohde & Hausfather, 2020) provides monthly temperature anomalies on a 1° grid. It has been featured in the IPCC Sixth Assessment Report (AR6 WG I) and the World Meteorological Organization's State of the Climate statement alongside other global records: the instrumental blends HadCRUT5 Analysis (Morice et al., 2021), NOAAGlobalTemp v6 (Huang et al., 2022; Yin et al., 2024), NASA GISTEMP v4 (Lenssen et al., 2019), China-MST3-Imax (Sun et al., 2022; Li et al., 2026) and DCENT-I (Chan et al., 2026), plus the reanalysis-based products ERA5 (Hersbach et al., 2020) and the newly released JRA-3Q (Kosaka et al., 2024). A recent community assessment of global warming since 1850 (Thorne et al., 2026) additionally used GloSATref (Morice et al., 2025), COBE-STEMP3 (Ishii et al., 2025), CMA-GMST (Chen et al., 2025), HadCRU_MLE_v1 (Calvert, 2024a), DCENT_MLE_v1 (Calvert, 2024b), and Kadow et al. (2021) to calculate its consensus estimate of global warming. Figure 1 compares global average anomaly time series of BEST-HR with those of other datasets.

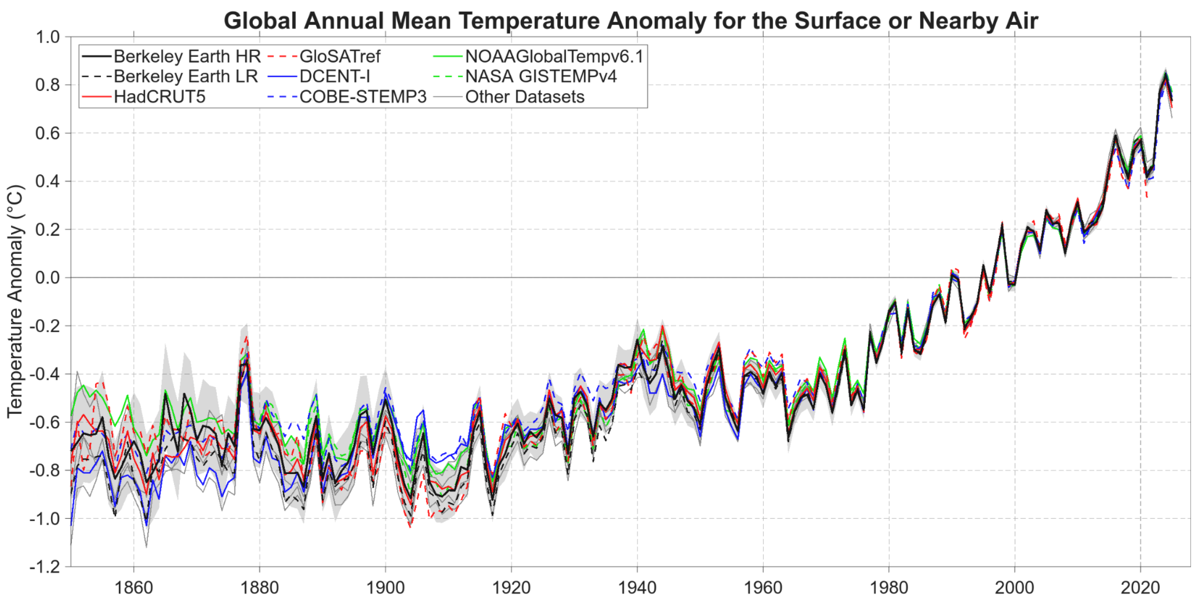


Figure **1:** Time series of global annual mean surface temperature anomalies, except for GloSATref, where surface air temperatures are used for sea regions. The shaded region reflects the 95 % uncertainties of Berkeley Earth HR. Other datasets include China-MST3-Imax, CMA-GMST, HadCRU_MLE, DCENT_MLE, and Kadow *et al*. Anomalies and uncertainties are relative to the 1981–2010 reference period.

As described by Thorne et al. (2026), these datasets differ in their source datasets, bias corrections, and spatial infilling techniques. Sea temperatures are one of the largest sources of differences, with BEST-HR, along with HadCRUT5, HadCRU_MLE, and Kadow et al., relying on sea surface temperatures (SSTs) from HadSST4 (Kennedy et al., 2019). Other datasets use alternative SST datasets (e.g., ERSSTv6, Huang et al., 2025; COBE-SST3, Ishii et al., 2025; DCSST, Chan et al. 2024) or marine air temperature (MAT) datasets (i.e., GloSATMAT, Morice et al., 2025). BEST-HR uses its unique land surface air temperature (LSAT) methodology that automatically detects breakpoints to split and bias correct station time series during the averaging process; LSAT datasets used by other global datasets rely on pre-processed homogenization methods, including automated procedures to adjust station time series (e.g., GHCNv4, Menne et al., 2018; DCLSAT, Chan et al., 2024) and/or adjustments made by national agencies (e.g. CRUTEM5, Osborn et al., 2021). For spatial infilling, BEST-HR relies on Gaussian process regression

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

(i.e. regression Kriging) using distance-based spatial correlation functions and predictive fields derived from historical weather patterns. Kriging-based approaches with distance-based correlation functions have also been adopted by other recent data sets (e.g., HadCRUT5, GloSAT, DCENT-I) though often without the use of predictive fields. Other datasets may use spatial patterns without Kriging (e.g., COBE-STEMP3, NOAA GlobalTemp) or neural networks (Kadow et al., 2018; NOAA GlobalTempv6; Qian et al., 2026) to infill data gaps.

Atmospheric reanalyses such as ERA5 and JRA-3Q provide physically consistent multi-variable fields at high temporal resolution but rely on data-assimilation systems that may under-represent early-period uncertainties and can drift when the observing system changes. ERA5 begins in January 1940 and JRA-3Q in September 1947, both report high spatial resolutions (0.25° and ~0.375°, respectively). Reanalysis in the modern era relies heavily on satellite data, augmented by surface observations–especially pressure, humidity and winds–plus atmospheric profiles from radiosondes / aircraft and marine data from ships and buoys. Before the satellite era (late 1970s) only the surface and upper-air observations were available, necessarily leading to larger analysis errors. The vast and qualitative changes in the inputs available for reanalysis also creates a risk of model drift, which can present as spurious trends.

A further caveat is that ERA5 and JRA-3Q do not directly assimilate most weather-station temperature measurements. Historical surface stations typically reported only daily maximum and minimum temperatures without reporting the time of day that these measurements occurred, yet four-dimensional variational (4D-Var) assimilation schemes require the observation time to apply dynamical constraints. Consequently, land-surface temperature in reanalyses is derived largely from model physics constrained by other variables, rendering the reanalysis fields semi-independent of station temperature observations. The marine component, by contrast, does assimilate SST observations, which is made possible by their lower diurnal variability which makes precise knowledge of the measurement time less important.

Figure 2 provides an example of how BEST-HR compares to the three other high resolution data products (ERA5, JRA-3Q, COBE-STEMP3) and two lower resolution observational data sets, including the previous Berkeley Earth product. We note that COBE's products, which focus on ocean reconstruction, provide 0.25° resolution for oceans but only 1° resolution for land.

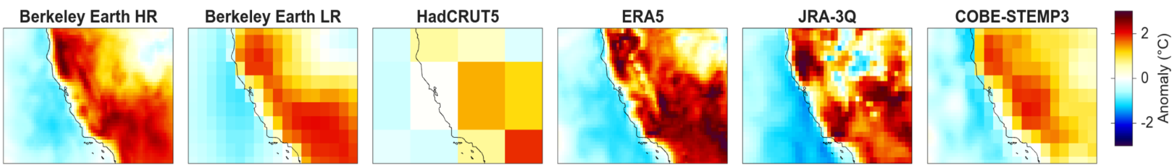


Figure 2: Example comparison of Berkeley Earth HR and Berkeley Earth LR with three other high-resolution datasets and one low resolution dataset (HadCRUT5). April 2021 temperature anomalies in °C for California relative to 1981–2010 are shown. Air temperatures instead of sea surface temperatures were used for ERA5 and JRA-3Q. For COBE-STEMP3, 0.25° resolution COBE-SST3 is shown for ocean cells.

1.2 Motivation for a higher-resolution Berkeley Earth product

User demand for high-resolution, observation-based climate data has grown rapidly in recent years. National meteorological services rely on spatially resolved temperature fields to assess climate trends and extremes at the sub-national level. Cities and regional planners need fine-scale data to design and evaluate heat adaptation strategies, especially in rapidly urbanizing and warming regions. Meanwhile, global and regional climate models require bias correction against observational baselines–

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

preferably with consistent long-term coverage and well-characterized uncertainties–to support multi-sector impact modelling and risk assessments.

However, many of the global temperature products currently available are too coarse in resolution (1°–5°) or too short in duration to support these applications directly. While reanalysis products like ERA5 and JRA-3Q offer higher resolution and physically consistent fields, they rely on complex data assimilation systems and are heavily influenced by satellite-era observations and model dynamics. In particular, reanalyses do not directly assimilate most land-surface temperature observations, which limits their fidelity in replicating true near-surface temperature trends over land. As such, long-term observational temperature records remain the most direct and traceable evidence for quantifying how the surface climate has evolved over the past two centuries.

Yet generating a fine-resolution observational data set from raw station data is non-trivial. Station density is uneven–often sparse in mountainous, polar, and developing regions–and local effects such as elevation, land cover, and orography play a larger role at finer spatial scales. Interpolation methods that ignore these factors will produce biased estimates, especially in regions with strong gradients. A method that is both physically informed and statistically robust is therefore needed.

In the original Berkeley Earth datasets, Kriging with a modelled covariance structure and auxiliary baseline parameters is used as the method of interpolating weather station temperatures while accounting for potential biases and structural variations such as those caused by topographic variability. This has proven very effective, but can produce fields that are unrealistically smooth at small scales, especially when weather station observations are sparse. With a few modest modifications, it is possible to construct fields that more accurately describe small-scale temperature variability.

Regression Kriging offers a natural approach to this challenge. In this method, the existing Kriging approach is supplemented with auxiliary predictors–such as empirically derived recurring weather patterns and topographic information–to better reproduce the spatial structure during interpolation. In this context, the residuals not explained by the predictive fields (i.e., the differences between observed station values and regression estimators) are what we spatially interpolate using our distance-based correlation function, which captures the local variations that cannot be explained by the predictors alone. Some previous studies also combined spatial pattern approaches with distance-based correlation functions (Calvert, 2024a; Ishii et al., 2025).

This two-stage approach combines the predictive power of physically relevant covariates with the flexibility and spatial consistency of geostatistical interpolation, producing a continuous temperature field that:

- Accurately reflects observed station values,
- Captures fine-scale spatial variation in regions with dense station coverage,
- Smoothly extrapolates into data-sparse areas based on physical predictors

When applied consistently over a long time series–such as in the Berkeley Earth Surface Temperature – High-Resolution (BEST-HR) product–this method provides a spatially explicit, observationally anchored record of surface temperature that supports both global climate tracking and localized impact assessments. Moreover, by re-analyzing the full historical archive (back to 1750 over land), BEST-HR enables robust assessments of century-scale warming patterns, regional climate sensitivity, and the evolution of extreme temperatures with more spatial detail than previously possible from traditional global products – with the possible exception of the 0.25° resolution SST estimates in COBE-STEMP3.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

1.3 Contribution of this study

Building on the methodology of Rohde et al. (2013a,b) and Rohde & Hausfather (2020), this study introduces a new high-resolution global temperature data set, Berkeley Earth Surface Temperature – High-Resolution (BEST-HR). Key updates and contributions include:

1. An updated land station archive currently incorporating 58,311 stations, reflecting expanded source data and improved station metadata;
2. A revised characterization of seasonal cycles, now incorporated directly into the iterative reconstruction process rather than handled as a separate preprocessing step;
3. The adoption of regression Kriging at 0.25° spatial resolution, using physically informed predictors including ERA5-derived temperature fields, lapse-rate terms, and high-resolution topography;
4. The introduction of a far-field smoothing approach that increases the accuracy of the modelled field at locations far from observations. This is particularly impactful in the earliest portions of the analysis.
5. A modified sea surface temperature (SST) reconstruction based on HadSST4 and Rohde & Hausfather (2020), which adds large-scale predictive basis functions and local Kriging of residual anomalies to preserve fine-scale structure;
6. A physically consistent merging procedure that blends the land-only TAVG field with the SST reconstruction to produce a seamless global Land + Ocean temperature product;
7. The provision of monthly gridded products:
    - Land-only TAVG from 1750–present,
    - Land-only TMAX and TMIN from 1850–present, and
    - Land + Ocean TAVG from 1850–present;
8. A redesigned uncertainty quantification framework, which generates a set of reduced-resolution ensemble realizations that better represent reconstruction uncertainty while remaining computationally efficient for downstream users;
9. A commitment to open science, with all data and code released under open-access licenses to support transparency, reproducibility, and community use.

While these changes represent substantial enhancements, many core assumptions and algorithms from the original Berkeley Earth methodology are retained, including:

1. The basic approach to building the land station archive and performing de-duplication and cross-referencing, aside from the inclusion of some new and updated sources.
2. The use of the "scalpel" homogenization technique for detecting and correcting non-climatic changes in station records;
3. Most of the quality control procedures for identifying and handling outliers, with a single exception described in Sect. 3;
4. The conceptual framework for decomposing temperature variability into long-term global trends, persistent climatological patterns, and short-term weather fluctuations;
5. The use of HadISST ice cover maps to guide interpolation in sea-ice-covered grid cells, except now updated from HadISST1 (Rayner et al., 2003) to HadISST2 (Titchner & Rayner, 2014).

The remainder of this paper is organized as follows:

- Section 2 describes the input data sets and preprocessing steps;
- Section 3 outlines the regression Kriging methods applied to land;

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

- Section 4 outlines the regression Kriging applied to the ocean and details the land and ocean merging process;
- Section 5 presents the new uncertainty estimation approach and ensemble generation procedure;
- Section 6 documents the characteristics of the resulting data products and compares BEST-HR to existing data sets;
- Section 7 provides summary remarks, usage recommendations, and discussion of limitations.

As discussed below the new BEST-HR analysis is quite similar to the previous BEST products at large scales, and most of the focus is on improving the small scale uses.

By delivering observation-based surface temperature data set at quarter-degree resolution, BEST-HR fills the scale gap between coarse gridded climate products and indirect, time-limited reanalyses – offering new opportunities for robust climate analysis from the global to the local scale.

2 Input data and preprocessing

The Berkeley Earth Surface Temperature – High-Resolution (BEST-HR) product is based on a combination of land-based weather station records, marine sea-surface temperature (SST) observations, and physically relevant predictor fields derived from reanalysis and topographic data sets. The land station component builds on the long-standing Berkeley Earth weather station archive, as described in Rohde et al. (2013a,b), with updates to include the most recent versions of source archives and associated metadata corrections.

2.1 Land station data sources

The Berkeley Earth station archive integrates records from numerous national and international data sets. The most prominent sources used in the current version of BEST-HR include: GHCNv3 (Lawrimore et al., 2011), GHCNv4 (Menne et al., 2018), USHCNv2 (Menne et al., 2009), U.S. COOP (NOAA, 2023a), International Surface Temperature Initiative (ISTI) (Rennie et al., 2014), NERMS (NOAA, 2023b), Global Summary of the Day (GSOD) (NOAA, 2025), World Weather Records (NOAA, 2017), World Monthly Surface Station Climatology (WMSSC) (UCAR, 2026a), U.S. First Order Summary of the Day (UCAR, 2026b), Colonial Era Weather Archive (NOAA, 2008), U.S. FORTS (MRCC, 2026), GSN (GCOS, 2026a), GCOS Monthly CLIMAT Summaries (GCOS, 2026b), MET-READER (British Antarctic Survey, 2026), CRUTEM5 (Osborn et al., 2021), the European Climate Assessment & Dataset (2026), HCLIM (Lundstad et al., 2023), and PROMICE (Fausto et al., 2026).

Daily sources are converted to monthly average sources via averaging. A monthly average is included if no more than 9 days are missing in the month.

All source records undergo automated matching and reconciliation to detect metadata overlaps, duplicated station IDs, and overlapping or redundant time series. In total, over 200,000 source records are ingested and subjected to de-duplication, preliminary outlier screening, and metadata harmonization, following the methodology documented in Rohde et al. (2013a). This results in a consolidated archive of 58,311 distinct station time series, which serve as the basis for the land-temperature reconstruction.

In addition, data sources are subject to preliminary quality control screening following Rohde et al. (2013a). This is not intended to be a comprehensive quality screen, but merely removes some of the most extreme outliers.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

Open Access Earth System Science Data Discussions

Table 1: List of Land Surface Air Temperature Observations used in BEST-HR

Source	Temperature Stations at Source	Temperature Stations Used	TAVG (Months, %)	TMIN (Months, %)	TMAX (Months, %)	Date Range
Global Historical Climatology Network - Daily	41,834	34,711	10,285,726 (43.54%)	10,393,446 (50.97%)	10,256,415 (50.43%)	1763–2026
Global Historical Climatology Network - Monthly v4	27,957	24,124	4,064,724 (17.21%)	0 (0%)	0 (0%)	1701–2026
Global Summary of the Day	27,927	17,900	2,492,901 (10.55%)	3,262,708 (16%)	3,337,970 (16.41%)	1929–2025
US Cooperative Summary of the Day	16,115	13,241	2,033,458 (8.61%)	2,861,882 (14.04%)	2,839,780 (13.96%)	1853–2010
European Climate Assessment & Dataset	10,519	7,635	1,592,391 (6.74%)	1,248,960 (6.13%)	1,265,687 (6.22%)	1756–2026
Hadley Centre Data Release (CRUTEM5)	11,850	5,542	1,293,927 (5.48%)	0 (0%)	0 (0%)	1850–2026
US Cooperative Summary of the Month	12,947	10,805	686,407 (2.91%)	158,351 (0.78%)	164,157 (0.81%)	1831–2011
Global Historical Climatology Network - Monthly v3	7,280	5,612	252,095 (1.07%)	890,234 (4.37%)	904,812 (4.45%)	1701–2019
US Historical Climatology Network - Monthly	1,218	989	194,839 (0.82%)	98,411 (0.48%)	107,359 (0.53%)	1895–2012
International Surface Temperature Initiative (ISTI)	35,909	6,224	175,359 (0.74%)	961,001 (4.71%)	940,104 (4.62%)	1701–2018
Global Historical Climate Database (HCLIM)	3,612	688	151,899 (0.64%)	0 (0%)	0 (0%)	1658–2021
US First Order Summary of the Day	1,600	1,191	144,823 (0.61%)	140,981 (0.69%)	141,040 (0.69%)	1881–2011
GSN Monthly Data	911	289	65,612 (0.28%)	66,104 (0.32%)	66,634 (0.33%)	1861–2005
World Monthly Surface Station Climatology (WMSSC)	5,508	997	59,714 (0.25%)	0 (0%)	0 (0%)	1752–2025
Colonial Archive	1,018	755	45,790 (0.19%)	215,879 (1.06%)	218,979 (1.08%)	1658–2021
Early US Forts	523	357	38,453 (0.16%)	21,655 (0.11%)	22,647 (0.11%)	1791–1932
World Weather Records (WWR)	1,860	998	24,000 (0.1%)	47,304 (0.23%)	48,453 (0.24%)	1961–2000
Reference Antarctic Data for Environmental Research (MET-READER)	119	107	11,351 (0.05%)	0 (0%)	0 (0%)	1903–2026

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

Programme for Monitoring of the Greenland Ice Sheet (PROMICE)	51	48	7,648 (0.03%)	0 (0%)	0 (0%)	1990–2026
GCOS Monthly CLIMAT Summaries	1,213	822	3,344 (0.01%)	24,093 (0.12%)	24,360 (0.12%)	2001–2026
Merged Berkeley Earth Dataset	-	58,311	23,624,462	20,391,007	20,338,395	

In constructing the Berkeley Earth weather station archive, source records are merged according to a defined hierarchy of preference that prioritizes daily-resolution data when available and complete–especially from sources such as GHCN-Daily and GSOD–and favors records based on raw observational data rather than pre-homogenized products. This enables consistent application of the Berkeley Earth quality control and homogenization procedures across all records. For early periods, sources often only provide monthly averages, either because the original data values have not yet been digitized or because the original daily records have been lost. In such cases, monthly archives provide valuable additional data.

A significant challenge in constructing the archive is that many weather stations appear in multiple source datasets, sometimes with overlapping or inconsistent metadata. As such, de-duplication is a critical part of the preprocessing pipeline, responsible for reducing the more than 200,000 initial candidate records down to approximately 58,311 unique station time series used in the final reconstruction.

In some cases, such as ISTI, GCOS CLIMAT, and WMSSC, almost all data in the source archive is redundant with other higher priority sources (e.g. GHCN), and as a result only a few percent of their input is retained (i.e. in the rare cases where they contain something that wasn't also reported by the higher priority source). Another dataset used in Rohde et al. (2013a), Monthly Climate Data of the World (MCDW), was entirely removed as redundant. In some cases, we identified stations that had been dropped from newer archives as an apparent result of incomplete metadata. Consulting multiple sources allowed some of these missing metadata cases to be resolved.

While deduplication is the primary reason for this reduction, some additional filtering occurs due to a minimal continuity requirement: stations must contain at least six months of valid data to be included in the analysis. Although this threshold is intentionally permissive–much lower than that used by most other groups–it still results in the exclusion of many short-duration daily station records, which are common in some modern automated systems and digitized historical archives.

In most cases, only the latest version of a given data archive is used. However, an exception is made for the GHCN-Monthly dataset, where both version 3 and version 4 are considered: version 4 offers more recent and comprehensive coverage of TAVG, while version 3 remains important due to its inclusion of TMAX and TMIN fields, which are currently missing from GHCN-M v4.

The GHCN-Daily and GSOD archives provide the largest share of stations and data volume in the merged product. We note that GSOD has stopped updating in 2025, and is being replaced by Synoptic Summary of the Day (SSOD) (NOAA, 2026a). We anticipate incorporating SSOD in the future, though this is not included at present. Other datasets, such as those from the MET-READER, PROMICE, the Colonial Archive, and HCLIM, contribute a relatively small portion of the overall data but remain important for filling regional and temporal gaps, including early observations in Africa and Asia and continuous records from Antarctica and Greenland. Still other data sources are included not for the uniqueness of their data but the utility of their metadata.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

For example, WMSSC includes historical location reports, while most other datasets only provide the current location. Such historical metadata aids in identifying station moves. In addition, two metadata only sources have also been consulted, the WMO Observing Systems Capability Analysis and Review (OSCAR) Volume A station metadata (WMO, 2021) and NOAA's Enhanced Master Station History Report (Vose et al., 2017), which also provide information on measurement methods and station movements.

A complete summary of station sources, site counts, and monthly data volumes is provided in Table 1.

2.2 Sea-surface temperature data

The SST component of BEST-HR is derived from the HadSST4 data set (Kennedy et al., 2019), which includes bias-corrected marine observations from ships and buoys, with uncertainty estimates and adjustments for changes in measurement methods. These observations are used to reconstruct ocean surface temperatures via large-scale predictive basis fields and local residual Kriging, described further in Sect. 4.

Additionally, sea ice coverage is incorporated from HadISST2 (Titchner & Rayner, 2014), which replaces HadISST1 used in earlier Berkeley Earth products. These ice masks are used to guide the treatment of SSTs in sea-ice-affected grid cells. Due to delays in the updating of HadISST2, the most recent months may substitute sea ice concentrations from the US National Snow and Ice Data Center Climate Data Record v6 (Meier *et al.*, 2026) when necessary.

2.3 Predictor fields for regression Kriging

To guide spatial interpolation via regression Kriging, several physically meaningful predictor variables are used:

- ERA5 reanalysis fields (Hersbach et al., 2020), including monthly 2-meter air temperature and sea surface temperature;
- ORAS5 ocean reanalysis fields (Zuo et al., 2019);
- The SRTM (Shuttle Radar Topography Mission) digital elevation model, providing elevation at high spatial resolution;
- Latitude; and
- A static land/sea mask and near-coast mask derived from the Global Self-consistent, Hierarchical, High-resolution Geography (GSHHS) Database v. 2.0 (Wessel & Smith, 1996).

All predictors are interpolated or averaged to the 0.25° grid and are assumed to be static over time except for the ERA5 and ORAS5 derived fields which are processed separately by month as described below.

As discussed below, additional predictors were considered but the options tried were judged not to provide any significant additional value.

3 Spatial reconstruction over land using regression Kriging

The Berkeley Earth Surface Temperature – High-Resolution (BEST-HR) product reconstructs monthly temperature anomalies over land using a geostatistical interpolation approach based on regression Kriging. This method combines physically informed spatial regression with residual interpolation via Kriging, allowing for a reconstruction that honors both large-scale climatic structure and local observational constraints.

To enable a structured estimation process, the observed temperature field is decomposed into physically interpretable components.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

3.1 Decomposition and modeling framework

At any location $\boldsymbol{x}$ and time t_j, the observed temperature is represented as:

$$T(\boldsymbol{x}, t_j) = \theta(t_j) + B(\boldsymbol{x}, \text{month}(t_j)) + W(\boldsymbol{x}, t_j) \tag{1}$$

where:

- $\theta(t_j)$ is the global mean temperature anomaly for month t_j,
- $B(\boldsymbol{x}, \text{month})$ is the local climatological average for the corresponding calendar month, and
- $W(\boldsymbol{x}, t_j)$ is the weather component, representing short-term spatial and temporal variability.

This decomposition has the desirable property that the weather component approximately satisfies:

$$\frac{1}{A}\int W(\boldsymbol{x}, t_j)\, dA \approx 0 \quad \text{for each } t_j \tag{2}$$

and

$$\frac{1}{T}\sum_j W\,(\boldsymbol{x}, t_j) \approx 0 \quad \text{for each } \boldsymbol{x} \tag{3}$$

implying that the weather term averages to near-zero over space and time, respectively.

To estimate $W(\boldsymbol{x}, t_j)$, we model station observations using a combination of fixed effects and predictor-driven drift, followed by spatial Kriging of the residuals.

Let:

- $d_i(t_j)$: observed temperature at station iii during month t_j,
- $\hat{\theta}_j$: estimated global mean anomaly at time t_j,
- b_i: station-specific baseline (long-term mean),
- $\hat{s}_i(t_j)$: estimated seasonal cycle at station i,
- $q_{k,j}(\boldsymbol{x}_i)$: value of the k-th predictor field at station i,
- $\beta_{k,j}$: regression coefficient for predictor k at time t_j,
- $w_{i,j}(\boldsymbol{x})$: Kriging weight for station i at location $\boldsymbol{x}$ at time t_j.

Then the weather field is reconstructed as:

$$W(\boldsymbol{x}, t_j) = \sum_i w_{i,j}\,(\boldsymbol{x}) \left[d_i(t_j) - \hat{\theta}_j - b_i - \hat{s}_i(t_j) - \sum_k \beta_{k,j}\; q_{k,j}(\boldsymbol{x}_i) \right] + \sum_k \beta_{k,j}\; q_{k,j}(\boldsymbol{x}) \tag{4}$$

This formulation has two key terms:

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

1. Residual interpolation:
   The bracketed quantity represents the de-meaned, de-seasonalized residual at station i, after removing the global signal, station bias, seasonal cycle, and large-scale predictor contributions. These residuals are interpolated via Kriging to estimate the local anomaly contribution.
2. External drift (regression):
   The second term adds back the large-scale spatial trend as predicted by the auxiliary fields evaluated at the target location $\boldsymbol{x}$.

This hybrid structure allows the model to flexibly capture local detail where observations exist, while relying on physically meaningful predictors to guide interpolation in data-sparse regions.

3.2 Predictors used in the regression component

The regression model $\sum_k \beta_{k,j}\ q_{k,j}(\boldsymbol{x})$ incorporates spatially varying predictors that are relevant to surface temperature patterns. For the residual weather component, $W(\boldsymbol{x}, t_j)$, the predictive fields for land TAVG, TMAX, and TMIN are derived solely from the ERA5 2-meter air temperature anomalies. The period January 1940 to December 2023 was used to construct the predictive basis.

3.2.1 Data preparation

ERA5 provides direct monthly TAVG fields. For TMAX and TMIN, we construct synthetic monthly values by averaging the daily maximum and minimum 2-meter air temperatures, respectively. For each grid cell and calendar month, we then compute the monthly anomalies by removing both:

1. The monthly land-average temperature anomaly, and
2. The monthly spatial climatology for that location.

This yields a weather anomaly field:

$$A(\boldsymbol{x}, t_j) = T_{\mathrm{ERA5}}(\boldsymbol{x}, t_j) - \theta_{\mathrm{ERA5}}(t_j) - B_{\mathrm{ERA5}}(\boldsymbol{x}, \mathrm{month}(t_j)) \tag{5}$$

where both the spatial mean for each time and the temporal mean at each location are constructed to be zero. This preprocessing step effectively removes the long-term global warming signal from ERA5 and isolates short-term weather variability.

3.2.2 Principal component decomposition

To extract characteristic patterns of weather variability, we perform a principal component analysis (PCA) using singular value decomposition (SVD). For each calendar month (e.g., all Januaries), we construct a temporal stack containing 252 monthly anomaly fields, spanning 1940–2023. We extend the stack from 84 to 252 members by including anomalies from adjacent months (i.e., ±1 month), enabling smoother transitions between adjacent months and better capturing common patterns.

Each anomaly field in the stack is weighted by grid cell area and is passed into the SVD algorithm. This yields a set of orthogonal spatial basis functions (predictive patterns) and associated explained variances.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

Following common practice, we retain the leading modes sufficient to capture at least 85 % of the total variance in the anomaly stack. Although the number of modes required varies slightly by month, we fix the number of predictive fields at 65, which reflects the mean required across all calendar months.

The leading components typically exhibit interpretable spatial structure. For example, the first mode tends to show spatial patterns of typical variability, i.e. larger inland and at high latitudes and small at the coast, while the later modes may reflect patterns associated with ENSO variability or other signals.

An example of selective maps for TAVG in November is provided in Fig. 3. Leading terms are dominated by typical high-amplitude weather patterns at high northern latitudes, while later modes show more varied features.

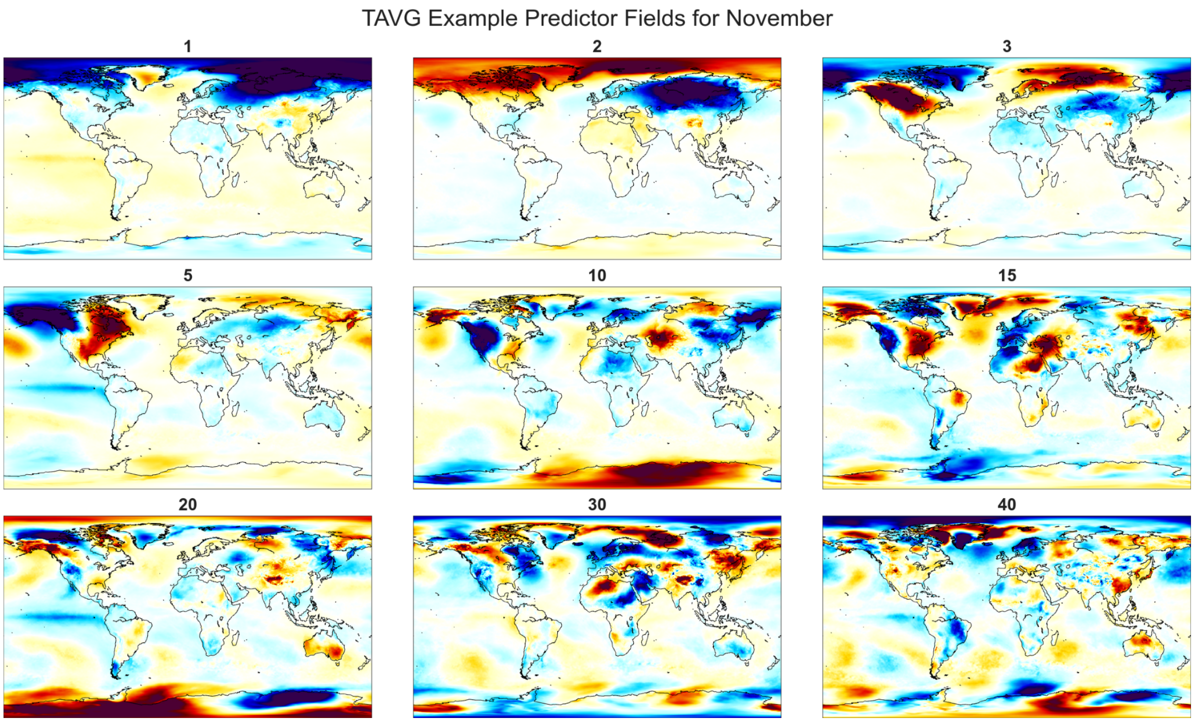


Figure 3: Example of spatial predictive weather patterns extracted from ERA5 for TAVG in the month of November. Numbers above each panel indicate its rank in the sequence of predictive fields, with lower numbers tending to have larger scale features and explaining a larger fraction of monthly average weather variability. The red/blue direction is arbitrary as the patterns will scaled by positive and negative factors during the fitting solution.

3.2.3 Spatial normalization

After decomposition, each principal pattern is:

- Divided by the grid cell area weights previously applied, and
- Renormalized such that the resulting field has unit spatial variance over the land mask used in the temperature reconstruction.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

This ensures that the predictive fields are properly scaled for use in the regression model and that no residual global or regional bias remains embedded in the predictors.

Note that, by construction, each of these spatial patterns has a land-average mean of zero. As a result, the inclusion of these patterns does not affect the land-average mean. This is done intentionally so predictive fields focus on spatial interpolation rather than directly impacting the global average.

These predictive fields are then applied in the time-dependent regression analysis described in Sect. 3.3, serving as physically meaningful spatial covariates that constrain the large-scale structure of temperature anomalies.

3.3 Estimation of regression and global mean parameters

The parameters used in the regression Kriging model–namely, the global mean anomaly $\hat{\theta}_j$ and the regression coefficients $\hat{\beta}_{k,j}$ for each predictor–are estimated jointly using a generalized least squares (GLS) approach that accounts for the spatial covariance of station residuals.

We begin by defining the centered station temperature residual:

$$z_i(t_j) = d_i(t_j) - b_i - \hat{s}_i(t_j) \tag{6}$$

which removes both the station baseline b_i and the estimated seasonal cycle $\hat{s}_i(t_j)$ from the observed temperature $d_i(t_j)$ at station i and time t_j.

To estimate $\hat{\theta}_j$ and $\hat{\beta}_{k,j}$, we construct a design matrix Q_j with elements:

$$Q_{j,i,k} = \begin{cases} 1 & \text{if } k = 1 \\ q_{k-1,j}(\boldsymbol{x}_i) & \text{if } k > 1 \end{cases} \tag{7}$$

where $q_{k-1,j}(\boldsymbol{x}_i)$ denotes the value of the $(k-1)$-th predictor field at the location of station i for time t_j. The first column of Q_j corresponds to a constant term for estimating the global mean, and the remaining columns contain the predictor values used in the spatial regression.

Let $\dot{z}$ be the column vector of $z_i(t_j)$ values across all stations, and let C denote the spatial covariance matrix representing the expected correlation structure of the residuals. Then, the parameter vector:

$$[\hat{\theta}_j \; \hat{\beta}_{1,j} \; \cdots \; \hat{\beta}_{K,j}] \tag{8}$$

is estimated by solving the generalized least squares equation:

$$[\hat{\theta}_j \; \hat{\beta}_{1,j} \; \cdots \; \hat{\beta}_{K,j}] = \left(Q_j^\top C^{-1} Q_j\right)^{-1} Q_j^\top C^{-1} \dot{z} \tag{9}$$

This expression minimizes the weighted sum of squared residuals while accounting for spatial dependence among stations.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

3.3.1 Covariance model

Following Rohde et al. (2013a), the covariance matrix C is approximated using a correlation model drawn from a standard spherical correlation function, commonly used in geostatistical interpolation due to its compact support and smooth behavior. This approximation is sensible as long as the variance structure of the monthly average fluctuations do not vary too rapidly with location.

Let $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ denote the geographic locations of two stations, and define $r = \text{distance}(\boldsymbol{x}_a, \boldsymbol{x}_b)$. The correlation between the residuals at these two stations is given by:

$$C(\boldsymbol{x}_a, \boldsymbol{x}_b) = \begin{cases} c_0 \left(1 - \dfrac{r}{d_{\max}}\right)^2 \left(1 + \dfrac{1}{2}\dfrac{r}{d_{\max}}\right) & \text{if } r \leq d_{\max} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where:

- c_0 is the residual variance (the nugget-plus-sill),
- $d_{\max}$ is the maximum correlation distance beyond which residuals are assumed uncorrelated.

This covariance function ensures that:

- Correlation is strongest at small distances,
- It decays smoothly to zero at $r = d_{\max}$,
- And values beyond $d_{\max}$ are uncorrelated, leading to a sparse matrix structure.

Both c_0 and $d_{\max}$ are empirically estimated from the residuals. In practice, the sparsity of C^{-1} enables computational efficiency when solving the GLS system.
This parameter estimation framework ensures that the large-scale spatial structure of the anomaly field is informed by physically meaningful predictors while respecting the statistical properties of the underlying observational network.

3.3.2 Limiting the predictive fields

Incorporating predictive fields into the regression Kriging framework requires care to ensure that each field is sufficiently constrained by the available observational data. While the full suite of 65 predictive patterns derived from ERA5 (Section 3.2) can be reliably used in the modern era, where thousands of observations provide dense spatial sampling, the situation is more challenging in earlier historical periods, where station networks were sparse and geographically uneven.

To avoid instability in the regression solution, we implement a filtering step that evaluates the statistical constraint on each predictor at each time step and excludes those that cannot be reliably fit.

The generalized least squares solution for the regression parameters above takes the form:

$$\left(Q_j^\top \dot{C}^{-1} Q_j\right)^{-1} Q_j^\top \dot{C}^{-1} \dot{z} \tag{11}$$

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

To assess whether a given predictor $q_{k,j}$ is sufficiently constrained, we perform a diagnostic test: we replace the observation vector $\dot{z}$ with random values drawn from a standard normal distribution (mean zero, unit variance). This simulates the impact of measurement uncertainty on the parameter estimates.

For each regression coefficient $\hat{\beta}_{k,j}$, we compute the implied standard deviation resulting from this synthetic fit. If the standard deviation exceeds 0.5, the predictor is considered ill-constrained and is excluded from the regression model at that time step. This threshold is chosen to approximate the constraint of at least four effectively independent observations influencing each coefficient. Whether and to what extent an observation contributes as an effective constraint will depend on the amplitude of the predictive field at the station's location and how many other stations are nearby. Multiple stations in close proximity will necessarily provide less constraint than stations that are well separated. If no stations are present at locations where the predictive field has large amplitude, then the implied uncertainty in the associated parameter can blow up far greater than 1.0.

This filtering is performed iteratively, starting with the full set of predictors and removing the least-constrained one at each step until all retained predictors fall below the threshold. Because each predictive pattern has been pre-normalized to unit spatial variance, the implied standard deviation directly reflects the confidence in its regression coefficient.

3.3.3 Temporal evolution of constraint

At the earliest time steps, especially in the 18th century for TAVG, only a few predictive patterns are sufficiently constrained by the available data. However, as the station network expands–particularly after 1800–the number of usable patterns grows rapidly. By the mid 19th century, most of the 65 patterns are retained in most months for TAVG. TMIN and TMAX, which have shorter and less complete historical records, reach full constraint somewhat later.

Even after all predictors are retained, their effective constraint varies by pattern. For example, spatial modes associated with Antarctica or the Southern Ocean continue to exhibit relatively high uncertainty until the mid-20th century, due to the scarcity of early observations in those regions. This spatial heterogeneity is incorporated into the overall uncertainty framework described in Sect. 5.

This adaptive filtering process ensures that the regression model remains numerically stable and physically meaningful, avoiding overfitting in poorly observed periods while maximizing explanatory power as data coverage improves.

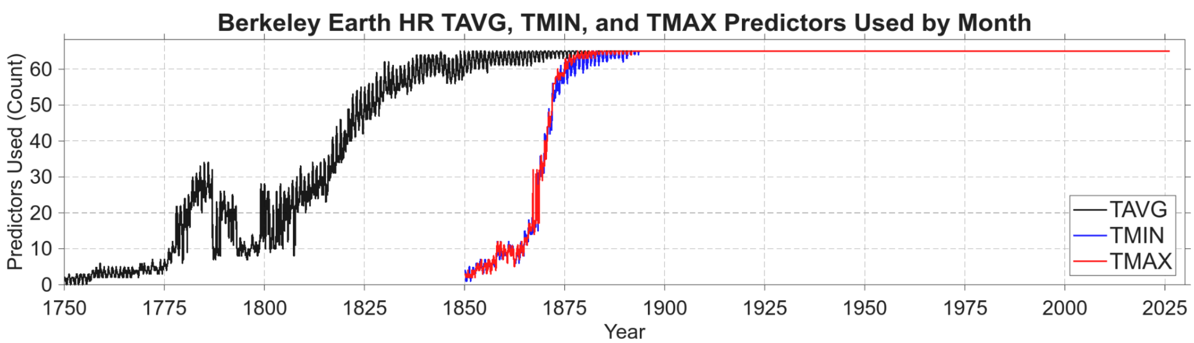


Figure 4: Number of spatial predictive fields used per month for BEST-HR with TAVG, TMAX, and TMIN.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

Open Access Earth System Science Data Discussions

3.4 Estimation of station baselines and seasonal cycles

In the BEST-HR framework, each station's observed temperature record is assumed to include a persistent baseline offset b_i, reflecting the station's long-term mean relative to the global anomaly, and a seasonal cycle $\hat{s}_i(t_j)$, capturing the repeating intra-annual variability. Accurate estimation of these components is critical for correctly isolating the weather term $W(\boldsymbol{x}, t_j)$ during interpolation.

3.4.1 Seasonal model

The seasonal cycle for each station is modeled as a sum of low-order harmonic functions with annual, semiannual, and triannual frequencies. Specifically,

$$\hat{s}_i(t_j) = \sum_{n=1}^{3} \left[\alpha_{i,2n-1} \cos\left(\frac{2\pi n}{12} t_j\right) + \alpha_{i,2n} \sin\left(\frac{2\pi n}{12} t_j\right)\right] \tag{12}$$

where:

- t_j is the month index (1 through 12, or treated continuously),
- $\alpha_{i,1}, \dots, \alpha_{i,6}$ are station-specific harmonic coefficients to be fit,
- and the three harmonic terms correspond to 1-year, 6-month, and 4-month periodicities, respectively.

This form allows each station's seasonal pattern to include both amplitude and phase variability at multiple frequencies, enabling flexible representation of local climatological behavior.

3.4.2 Fitting procedure for stations with ≥5 years of data

For stations with at least five years of valid monthly observations, both b_i and $\hat{s}_i(t_j)$ are estimated using ordinary least squares (OLS). During this step, the global mean anomaly $\hat{\theta}_j$ and the weather field $W(\boldsymbol{x}_i, t_j)$ are held fixed.

We minimize the squared residual:

$$\sum_j \left(d_i(t_j) - b_i - \hat{s}_i(t_j) - \hat{\theta}_j - W(\boldsymbol{x}_i, t_j)\right)^2 \tag{13}$$

where the sum is taken over all available months t_j for the station.

3.4.3 Treatment of stations with <5 years of data

For stations with fewer than five years of data, there is generally insufficient information to robustly estimate a local seasonal cycle. In these cases:

- The baseline offset bib_ibi is still estimated using OLS;

- The seasonal cycle $\hat{s}_i(t_j)$ is not fit independently, but instead inferred by sampling from the seasonal structure of the global regression field at the station location.

This strategy allows even short-duration stations to contribute to the reconstruction while avoiding overfitting of poorly constrained seasonal behavior.

3.5 Iterative solution and outlier detection

The estimation procedures described in previous sections allow for the computation of both large-scale parameters–including the global mean $\hat{\theta}_j$ and regression coefficients $\hat{\beta}_{k,j}$–and station-specific small-scale parameters, including the baselines b_i and seasonal cycles $\hat{s}_i(t_j)$. However, because these parameter sets depend on one another (e.g., estimation of b_i requires knowing $\hat{\theta}_j$, and vice versa), a self-consistent solution must be obtained through iteration.

3.5.1 Iteration and convergence criterion

The BEST-HR land reconstruction is solved via an iterative process, alternating between:

- Updating $\hat{\theta}_j$ and $\hat{\beta}_{k,j}$ using generalized least squares (Section 3.3),
- Updating b_i and $\hat{s}_i(t_j)$ using least squares regression (Section 3.4), and
- Recomputing the weather field $W(\boldsymbol{x}, t_j)$ via regression Kriging (Section 3.1).

This process is repeated until convergence is achieved, defined as:

$$\max_j \left| \overline{T}^{(n)}(t_j) - \overline{T}^{(n-1)}(t_j) \right| < 0.001^{\circ}\mathrm{C} \tag{14}$$

where $\overline{T}^{(n)}(t_j)$ is the global average temperature anomaly at time t_j after iteration n. In practice, convergence is typically achieved after ~60 iterations.

3.5.2 Iterative outlier detection

To improve stability and robustness, an outlier detection and mitigation step is performed between iterations. This step identifies observations that are strongly inconsistent with the current global or local fit and limits their influence in subsequent iterations.

Let the residual at station i and time t_j be:

$$\Delta_i(t_j) = d_i(t_j) - T(\boldsymbol{x}_i, t_j) \tag{15}$$

where $T(\boldsymbol{x}_i, t_j)$ is the reconstructed temperature at the station location from the previous iteration. The distribution of $\Delta_i(t_j)$ is expected to be approximately Gaussian with heavy tails due to outliers or uncorrected errors. A threshold $\Delta_{\max}$ is defined as:

$$\Delta_{\max} = 2.5 \times \sigma_{\Delta} \tag{16}$$

where σ_{Δ} is the empirical standard deviation of the $\Delta_i(t_j)$ distribution across all stations and times. This threshold corresponds roughly to the 99th percentile of residuals.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

Open Access Earth System Science Data Discussions

If an observation exceeds this bound, it is capped in the next iteration:

- If $\Delta_i(t_j) > \Delta_{\max}$, then $d_i(t_j)$ is replaced by $T(\boldsymbol{x}_i, t_j) + \Delta_{\max}$
- If $\Delta_i(t_j) < -\Delta_{\max}$, then $d_i(t_j)$ is replaced by $T(\boldsymbol{x}_i, t_j) - \Delta_{\max}$

This process limits the influence of anomalous data while preserving the overall integrity of the station record. Unlike hard exclusion, this soft clipping approach still retains the data point but bounds its impact.

3.5.3 Secondary outlier test for spatially isolated stations

A second, more conservative quality control test is used to address errors in spatially isolated stations, where a single outlier can disproportionately affect the weather field due to the lack of nearby observations.

Define:

$$G_i(t_j) = d_i(t_j) - \hat{\theta}_j - C(\boldsymbol{x}_i, t_j) \tag{17}$$

where $C(\boldsymbol{x}_i, t_j)$ is the station's estimated climatology. This metric removes both the global and climatological contributions, highlighting potential outliers in the raw residual independent of the local weather fit.

Outliers are flagged if:

$$|G_i(t_j)| > 4 \times \sigma_G \tag{18}$$

where σ_G is the standard deviation of the $G_i(t_j)$ distribution. This corresponds to roughly 0.006 % of values. Observations exceeding this bound are clipped similarly to the primary test.

Although rarely triggered in practice, this additional filter serves to protect the reconstruction from extreme values in low-density regions, where the weather field is otherwise vulnerable to single-point anomalies.

3.6 Estimation of the climatological mean field

To complete the reconstruction of monthly temperatures, the climatological mean field $B(\boldsymbol{x}, \text{month})$ is modeled as the sum of two components:

$$B(\boldsymbol{x}, \text{month}) = \bar{B}(\boldsymbol{x}) + B_s(\boldsymbol{x}, \text{month}) \tag{19}$$

where:

- $\bar{B}(\boldsymbol{x})$ is the non-seasonal mean climatology, representing the annual average temperature at location $\boldsymbol{x}$, and
- $B_s(\boldsymbol{x}, \text{month})$ is the seasonal cycle component, varying across calendar months.

3.6.1 Estimating the non-seasonal climatology $\bar{\boldsymbol{B}}(\boldsymbol{x})$

The baseline field $\bar{B}(\boldsymbol{x})$ is fit using regression Kriging, applied to the station baseline offsets b_i derived in Sect. 3.4. The spatial correlation model used follows the same framework described in Sect. 3.4, with a spherical covariance structure and distance-

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

limited Kriging. Following Rohde et al. (2013a) predictors in the regression component include multiple terms related to latitude, elevation, but in this work we add the long-term climatology of the ERA5 2-meter air temperature, which provides improved constraint on broad spatial patterns.

Additional predictive patterns were considered for constructing $\bar{B}(\boldsymbol{x})$, but were ultimately rejected. These include patterns based on albedo, urbanization, land use type, normalized difference vegetation index (NDVI), and population density. Ultimately, none of these were found to add significantly to the fit and reconstruction beyond what was provided by the use of elevation, latitude, and ERA5 climatology.

The uncertainty associated with each b_i is modeled as inversely proportional to the duration of the station record. Specifically, if a station has N years of valid observations, this is treated as equivalent to N independent estimates of b_i, reducing the corresponding diagonal entry in the covariance matrix and increasing the station's weight in the fit. When multiple stations are located at or near the same geographic location, the estimated value of $\bar{B}(\boldsymbol{x})$ is effectively drawn toward the length-weighted average of the associated b_i values.

As with the broader temperature reconstruction, an outlier detection step is applied iteratively to the b_i values during fitting. Observations that deviate by more than 2.5 standard deviations from the colocated $\bar{B}(\boldsymbol{x})$ are removed. This removes approximately 1 % of the data and helps mitigate the influence of unusual noise or metadata errors arising when station records that are incorrectly documented in location or elevation.

For the reconstruction of TMAX and TMIN climatologies, the regression model for $\bar{B}(\boldsymbol{x})$ also includes an additional coastal proximity predictor–defined as the fractional land area within 60 km and 750 km of the grid cell–to capture known shifts in temperature profiles near oceans. This term was found to be non-significant for TAVG, and is therefore excluded from the TAVG climatology fit.

3.6.2 Estimating the seasonal climatology $B_s(\boldsymbol{x},\text{month})$

The seasonal component $B_s(\boldsymbol{x},\text{month})$ is derived using a similar regression Kriging approach, applied to the estimated seasonal cycles $\hat{s}_i(\text{month})$ from individual stations (Section 3.4). Predictor fields include:

- Latitude and elevation,
- The ERA5 monthly seasonality of 2-meter air temperature anomalies, and
- Two coastal proximity predictors: the fraction of land area within 60 km and within 750 km of the grid cell.

These predictors help capture the sharp attenuation of seasonal amplitude in coastal regions. Empirical analysis shows that the 60 km coastal predictor typically has a value of ~0.5 at the shoreline, ~0.75 at ~24 km inland, and ~0.9 at ~40 km inland, allowing the model to reproduce the observed decay in seasonal variation near ocean boundaries.

To ensure consistency with the station-based seasonality fits, $B_s(\boldsymbol{x},\text{month})$ is constrained to be representable as a third-order sum of harmonic functions (sines and cosines) with annual, semiannual, and triannual periodicities. This mirrors the functional form used to estimate $\hat{s}_i(\text{month})$, ensuring temporal smoothness and preventing unrealistic oscillations.

As with the baseline component, outlier detection is applied to the seasonal cycle amplitudes. Approximately 1 % of the $\hat{s}_i(\text{month})$ series are excluded from the fit based on a 2.5σ threshold against the modeled seasonal trend, improving robustness to misfit or ill-conditioned data. We also note, as before, that $\hat{s}_i(\text{month})$ is only estimated from stations having at least 5 years of observations.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

3.7 Far-field smoothing of the weather field

The final step in the reconstruction of the weather field involves a far-field smoothing correction, applied as a post-processing operation after each iteration of the regression Kriging procedure. This adjustment targets regions far from any weather observations, where the interpolated monthly anomalies can exhibit spurious or poorly constrained variability due to limited data support.

3.7.1 Motivation and formulation

In sparsely observed regions, particularly during early time periods or over remote land areas, the reconstruction of monthly-scale variability becomes increasingly uncertain. However, it is often the case that annual-average anomalies remain more robust and better correlated with large-scale climate signals in these settings. To account for this, we apply a smoothing step that selectively replaces local monthly anomalies with an estimate based on annual averages in poorly constrained areas.

The smoothed temperature field is defined as:

$$\hat{T}^*(\boldsymbol{x}, t_j) = \hat{T}(\boldsymbol{x}, t_j) + \mathrm{FF}(\boldsymbol{x}, t_j)\left[\hat{\theta}^*(t_j) - \hat{\theta}(t_j)\right] \tag{20}$$

where:

- $\hat{T}(\boldsymbol{x}, t_j)$ is the original reconstructed temperature field,
- $\hat{\theta}(t_j)$ is the estimated global mean at time t_j,
- $\hat{\theta}^*(t_j)$ is the 13-point centered moving average of $\theta(t_j)$, approximating the annual global mean,
- $\mathrm{FF}(\boldsymbol{x}, t_j)$ is the far-field weighting function (defined below).

For time steps near the endpoints of the record, where the full 13-month window is not available, $\theta^*(t_j)$ is computed using linear extrapolation from the available values.

This formulation ensures that in regions far from any observations, the local field is gently nudged toward the annual global anomaly, which typically reflects the best-constrained component of the temperature signal. In densely sampled regions, the correction is negligible.

3.7.2 Defining far-field influence via normalized Kriging error

The far-field weighting function $\mathrm{FF}(\boldsymbol{x}, t_j)$ is defined as a function of the effective coverage metric $\mathrm{CC}(\boldsymbol{x}, t_j)$, which quantifies the normalized interpolation uncertainty from the regression Kriging step. This metric replaces the coverage function used in Rohde et al. (2013a), but plays a similar conceptual role in flagging undersampled regions.

Formally:

$$\mathrm{FF}(\boldsymbol{x}, t_j) = f\big(\mathrm{CC}(\boldsymbol{x}, t_j)\big) \tag{21}$$

where f is a tapering function defined such that:

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

- $f(0) = 1$ (full smoothing in the absence of constraint),
- $f(x \geq x_{\mathrm{cutoff}}) = 0$ (no correction when well constrained),
- $f(x)$ smoothly interpolates in between.

The function $f(x)$ is defined using a polynomial taper, empirically fit using reconstructions of simulated model data to minimize reconstruction error. Full details of this fit and the choice of polynomial are provided in the Supplemental Material.

In practice, this correction behaves as follows:

- In regions with good data coverage, typically within a few hundred kilometers of weather stations, $\mathrm{FF}(\boldsymbol{x}, t_j) \approx 0$, and no correction is applied.
- At intermediate distances, the correction becomes moderate.
- At distances near or exceeding the maximum correlation length $d_{\max}$ (≈1650 km), the correction grows large and may fully replace monthly anomalies with their smoothed annual counterparts.

This procedure ensures spatial consistency while preventing overinterpretation of unconstrained monthly anomalies. It is especially beneficial for reducing artificial variability in large-scale averages and trend calculations in early or low-coverage periods.

It is worth noting that since this procedure only affects the temperature field in areas at some distance from the weather stations, it doesn't directly feed back into the computations of any of the weather station fit parameters.

3.8 Implementation notes

The full regression Kriging pipeline is implemented in Matlab, with data ingestion, quality control, predictor computation, regression fitting, and Kriging all executed in parallel by time step to optimize performance. Special care is taken to enforce land/sea masking consistency and to maintain numerical robustness, particularly in the presence of missing data or correlated predictors.

The reconstruction is applied independently for each of the TAVG, TMAX, and TMIN land variables, resulting in three parallel high-resolution monthly gridded datasets with consistent methodology. The procedure for reconstructing ocean surface temperatures and merging the land and ocean fields into a global product is described separately in Sect. 4.

While the regression Kriging framework described in Sect. 3.1–3.5 is mathematically continuous in space, in practice it is necessary to compute solutions on a finite grid due to computational constraints. The final temperature field $T(\boldsymbol{x}, t_j)$ is evaluated on a 0.25° latitude–longitude grid, which also serves as the resolution for the predictor fields. However, to reduce memory and GPU demands, the weather component $W(\boldsymbol{x}, t_j)$ is computed on a sparse, adaptive-resolution grid and interpolated to the full resolution using spherical natural neighbor interpolation.

The adaptive grid used for Kriging has a variable spatial resolution ranging from 0.5° in well-observed regions to 8° in remote areas with few or no station records, particularly over open ocean. Grid refinement is increased in proximity to weather stations and along coastlines where gradients are sharper. This replaces the equal-area gridding scheme used in Rohde et al. (2013a), and yields substantial computational benefits: the number of Kriging locations is reduced from approximately 3.5 billion to 200 million, significantly lowering both memory usage and processing time while maintaining spatial accuracy in regions of interest. Further details on the adaptive grid structure and interpolation method are provided in the Supplemental Material, Sect. S8.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

Open Access Earth System Science Data Discussions

4 Ocean temperature reconstruction and land–ocean merging

The ocean surface temperature reconstruction in BEST-HR extends the methodology developed in Rohde & Hausfather (2020), incorporating new predictive fields and operating at a higher spatial resolution. While the core observational foundation remains the HadSST4 dataset, several enhancements have been introduced to improve spatial detail and consistency with the land analysis described in Sect. 3.

4.1 Sea surface temperature interpolation

The BEST-HR sea surface temperature (SST) reconstruction relies on HadSST4 as its primary observational input. HadSST provides monthly SST anomaly fields on a 5°×5° grid, along with associated bias corrections and measurement uncertainty estimates. Because of this coarse native resolution, much of the fine-scale structure in the BEST-HR ocean product is derived from the application of predictive fields and regression Kriging interpolation, analogous to the approach used over land.

The ocean temperature field is modeled using the decomposition:

$$T(\boldsymbol{x}, t_j) = \theta(t_j) + B(\boldsymbol{x}, \text{month}(t_j)) + W(\boldsymbol{x}, t_j) \tag{22}$$

where:

- $\theta(t_j)$ is the global mean sea surface temperature anomaly at month t_j,
- $B(\boldsymbol{x}, \text{month})$ is the monthly climatology or background spatial field,
- $W(\boldsymbol{x}, t_j)$ is the residual anomaly field, interpolated from observations via regression Kriging.

Unlike the land reconstruction, the ocean model does not include an explicit seasonal fitting step for individual observations. In part this is due to the much smaller seasonal variability present in SST fields. However, both the monthly climatology $B(\boldsymbol{x}, \text{month})$ and the predictive fields used in the regression vary by calendar month, effectively embedding seasonal structure into the baseline field.

Predictive fields for the ocean analysis are constructed in parallel to the land approach but are derived from ORAS5, a fully coupled ocean reanalysis driven by ERA5 atmospheric boundary conditions. ORAS5 provides monthly SST fields that reflect large-scale oceanic processes, such as currents and heat transport, offering more physically realistic predictors than ERA5 SSTs alone.

The generalized least squares regression is used to estimate the global mean $\theta(t_j)$ and predictor coefficients $\hat{\beta}_{k,j}$, and residuals are interpolated using spatial Kriging. With the inclusion of predictive fields, the residual correlation length used in Kriging can be reduced, enabling sharper reconstructions in data-rich areas.

As previously described in Rohde & Hausfather (2020), an additional temporal persistence consideration is applied in regions of the ocean with too little data to directly constrain. Specifically, when possible, spatial gaps in the reconstructed SST field are filled with the average anomaly from one month earlier and one month later. This extra persistence step is possible because SST anomalies often persist for months, though its utility is generally limited to early times and locations of sparse sampling.

The final SST field is reconstructed at 0.25° resolution, matching the land product and enabling natural blending in Sect. 4.3.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

4.2 Uncertainty handling and quality control

Following Rohde & Hausfather (2020), the measurement uncertainties, coverage errors, and bias uncertainties provided by HadSST4 are propagated through the Kriging procedure. This is achieved by including HadSST's per-grid-cell uncertainty estimates and covariance matrices, and by using a measurement-error-aware Kriging model that dampens the influence of grid cells with poor or uncertain observational support. In addition, coverage errors are estimated using simulated data as discussed below.

HadSST microbias uncertainties (i.e., correlated uncertainties due to the measurement practices of a particular ship or buoy) have temporal correlations, which are not reflected in the HadSST covariance matrices. HadCRUT5 and GloSATref attempt to treat microbias errors as perfectly correlated within a calendar year and uncorrelated between calendar years to account for this, although the implementation of this is imperfect and depends on the arbitrary ordering of grid cells prior to Cholesky decomposition (Calvert, 2024). To account for HadSST microbias uncertainties, we draw random microbias realizations such that each area-weighted average microbias uncertainty realization has a temporal month-to-month autocorrelation parameter of 0.9165. This autocorrelation parameter was chosen so that the effective number of independent samples per 12 months of realizations would be 1. This aligns with the assumption in HadSST4 that microbias uncertainties are fully correlated within a year but uncorrelated between years. The validity of this assumption is largely unproven, but we will not attempt to consider alternatives here.

No additional outlier detection is applied in this context, though HadSST has already done such filtering. However, the incorporation of measurement uncertainties into the correlation matrices (Rohde & Hausfather, 2020) also has the effect of smoothing over isolated measurements that are presented with low confidence in favor of larger-scale consistency in the interpolated fields.

4.3 Land–ocean merging

The land-only TAVG reconstruction and the ocean SST field are merged to produce a global surface temperature dataset, consistent with the approach described in Rohde & Hausfather (2020), with some modifications introduced for higher resolution.

Previously, land–ocean blending was performed by computing the land fraction of each 1°×1° grid cell (~120 km), and using this to linearly blend the land and ocean reconstructions. In BEST-HR, the same principle is applied at 0.25° resolution, but with a more nuanced estimate of land influence: the land fraction is computed as the fraction of land area within 60 km of each grid cell center, excluding inland water bodies such as rivers and lakes. This spatial filter produces a smoother and more realistic transition between land and ocean domains, particularly along complex coastlines.

As in previous versions, grid cells covered by sea ice, as defined by the HadISST2 monthly sea ice concentration fields, are treated as "land-like" in the blending step. That is, sea ice-covered ocean regions are assumed to follow the land surface air temperature pattern, rather than open water SSTs. This decision reflects the insulating properties of sea ice and its role in decoupling the ocean from the atmosphere.

Prior to smoothing, our approach estimates a discontinuous temperature field between land, sea ice, and open sea regions, which could cause a double counting of temperature change from sea ice melt as described by Calvert (2024). Furthermore, HadISST2 may substantially overestimate Antarctic sea ice concentrations prior to 1940 due the misinterpretation of a German atlas (Thorne et al., 2026), which could cause BEST-HR to overestimate temperature change in Antarctic sea ice areas. At the same time, our

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

approach excludes regions far from observations from global temperature estimates, which can remove some impacts of Antarctic sea ice melt from our 19th century estimates.

The final merged field provides a globally complete, monthly gridded surface temperature record spanning both land and ocean at 0.25° resolution. This dataset forms the basis for subsequent global average temperature analyses, anomaly calculations, and uncertainty ensemble generation, as discussed in Sect. 5 and 6.

5 Uncertainty analysis and ensemble construction

The uncertainty analysis for BEST-HR builds on the statistical sampling framework developed in Rohde et al. (2013a) and Rohde & Hausfather (2020), but introduces several important modifications. In particular, the new analysis enables the construction of a small ensemble of alternative realizations of the global temperature field, rather than estimating uncertainty solely from pointwise variance. This change improves the utility of the dataset for downstream applications that require uncertainty propagation, such as trend estimation, climate model evaluation, and impact assessments.

Due to the computational demands of the reconstruction process, the primary ensemble is limited to 10 members, a much smaller number than the 100–200 realizations offered by some other datasets. Additionally, to reduce resource requirements, the ensemble is constructed at a coarser 1° × 1° resolution, though it retains the core features of the full-resolution analysis–including predictive fields, Kriging, quality control procedures, and bias adjustments. However, we can combine ensembles from multiple recent months to create a larger superensemble that provides more robust uncertainty estimates. For this paper, 5 recent ensemble sets with end dates ranging from November 2025 to March 2026 are merged into a 50 member superensemble. This superensemble approach has the limitation that the most recent months will have fewer than 50 members, but all other dates will have a full sample.

5.1 Land uncertainty ensemble

For land areas, ensemble members are constructed using a bootstrap resampling approach, in which the station archive is resampled with replacement. This yields alternative realizations of the station network: some stations appear multiple times, while others are omitted. Such bootstrap methods provide a simple method to test the sensitivity of an analysis to data selection and data noise. Each resampled archive is passed through the entire reconstruction pipeline (at reduced resolution). Differences between the resampled fields and the original BEST-HR then reflect the sensitivity to both station selection and measurement noise.

In contrast to Rohde et al. (2013a), the breakpoint detection and homogenization procedures are now re-run for each ensemble member. This allows the ensemble to capture uncertainty in breakpoint timing and alignment, a source of uncertainty previously omitted. Each ensemble member therefore provides an estimate of analysis uncertainty related to noise, metadata, and homogenization.

A second step quantifies coverage uncertainty, i.e., the error associated with the incomplete spatial sampling of the temperature field by the weather station network. Previously (Rohde et al. 2013a) this was estimated using a coverage diagnostic to characterize regions of unsampled variability. BEST-HR replaces this diagnostic with a more robust analysis where potential coverage errors are assessed more directly. This is done by simulating station observations from a known temperature field and comparing the resulting reconstruction to the true field.

In BEST-HR, the known field is drawn from historical runs of the CESM2 climate model, with the long-term forced trend removed. The data are segmented into multi-decade blocks and randomly reordered in time. This preserves realistic patterns of historical variability while decoupling the test from any actual climate change signals. For each ensemble member, simulated data are sampled at the resampled station locations and processed through the same reconstruction pipeline, without adding measurement noise (as that component is already reflected in phase one). The differences between the reconstructed and original model fields provides an estimate of spatial sampling error for each realization.

The two phases–(1) analysis noise and station selection, and (2) coverage error–are added together to create a single uncertainty realization. This process is repeated for all 10 resampled networks, producing a small but comprehensive land uncertainty ensemble.

These 10 members comprise the primary ensemble generated each month. This can then be extended into a superensemble by merging the most recent month of ensembles with ensemble members from prior months.

Uncertainties for land-based metrics are then computed as the ensemble standard deviation relative to the full BEST-HR, scaled to the 95 % confidence range under the assumption of normally distributed errors. We do not apply a Bessel's correction since both our bootstrap method, and the inclusion of an extra degree of freedom in the calculation of the standard deviation, would make it inappropriate.

5.2 Ocean uncertainty ensemble

For the ocean domain, the analysis builds on the HadSST4 ensemble, which provides 200 realizations that sample bias correction uncertainty. However, HadSST's ensemble does not account for correlated and uncorrelated measurement noise, nor does it include coverage uncertainty arising from sparse marine observations.

To construct a fully representative ocean ensemble consistent with the land analysis, a subset of the HadSST bias ensemble is selected at random (matching the size of the land superensemble). For each selected realization, additional perturbations are applied:

- **Uncorrelated measurement errors** are sampled from HadSST's reported variance fields;
- **Correlated measurement errors** are simulated via Cholesky decomposition of the monthly error covariance matrix; in addition, ensemble members for correlated measurement error are generated to preserve a month-to-month temporal autocorrelation parameter of 0.9165 as described in Sect. 4.2;
- **Coverage uncertainty** is assessed via a known-field reconstruction test, using a CESM2-based SST field and comparing the reconstruction to the target, as in the land case.

Because HadSST provides full uncertainty metadata–including bias corrections, measurement error variances, and monthly covariance fields–it is possible to generate full perturbations of the original uninterpolated HadSST fields consistent with the statistical structure of the original data.

The resulting ocean data perturbations are then passed through the BEST-HR ocean reconstruction procedure, including regression Kriging with predictive fields. This produces a set of spatially complete SST ensemble fields, each perturbed by realistic uncertainty components. These perturbed fields are then merged with estimated coverage errors to create a fully realized ensemble of alternative SST fields. Similar to the LSAT uncertainty, SST uncertainties are computed as the ensemble standard deviation

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

relative to BEST-HR, with no Bessel's correction applied, scaled to the 95 % confidence range under the assumption of normally distributed errors.

5.3 Merged land–ocean ensemble

Having constructed separate land and ocean ensembles, corresponding realizations are merged to form a consistent global surface temperature ensemble. The merging procedure follows the blending framework described in Sect. 4.3, with land–ocean weights determined by coastal proximity and sea ice masking.

Each member of the merged ensemble reflects a full-field reconstruction perturbed by:

- Weather station observational noise and metadata uncertainty (from bootstrap resampling and homogenization),
- Coverage limitations (from known-field reconstruction tests), and
- SST bias and measurement error structures from HadSST metadata.

This ensemble enables straightforward estimation of uncertainties in any derived quantity. For example, the uncertainty on global mean temperature anomalies, trend estimates, or regional averages can be obtained by computing the desired metric across the ensemble and taking the standard deviation, scaled to represent the 95 % confidence interval.

6 Results and Evaluation

This section presents the results of the BEST-HR reconstruction, including global and regional temperature trends, internal consistency checks, comparison with existing datasets, and characterization of reconstruction uncertainty. All results are based on the full-resolution 0.25° × 0.25° fields unless otherwise noted.

6.1 Global Time Series

At large scales, Fig. 5 shows that BEST-HR is little changed from the prior low-resolution version of Berkeley Earth's analysis (BEST-LR). Land, ocean, and global trends are each very similar between BEST-HR and BEST-LR, with such differences as do exist generally within the previously reported 95 % uncertainty range.

The most notable global average difference is a slightly warmer 19th century compared to BEST-HR. The largest source of this difference is due to changes in the estimate of temperatures over the Arctic as a consequence of the predictive fields used in the interpolation. Because temperature averages over sea ice are not included in either the land or ocean averages, this difference is not immediately apparent in either the land or ocean averages in Fig. 5.

The other important difference revealed in Fig. 5 is changes in the uncertainty estimation. Prior to ~1920, BEST-HR often has a greater land and global uncertainty estimate than BEST-LR. This is not because the new interpolation method of BEST-HR is less accurate, but rather because the new approach to uncertainty quantification does a better job of assessing the possible errors. In other words, we conclude that BEST-LR was likely somewhat underestimating its uncertainty in the early part of the record. The increased uncertainty estimate arises primarily from two factors, the incorporation of homogenization and breakpoint uncertainties (fully omitted in BEST-LR) and the development of a new coverage uncertainty measurement approach that provides a more robust characterization of the impact of sparse coverage. We also note that the ocean uncertainty around World War II has been reduced as a result of a reduction in the bias uncertainty assessment in HadSST4.2 compared to the earlier HadSST4.1.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

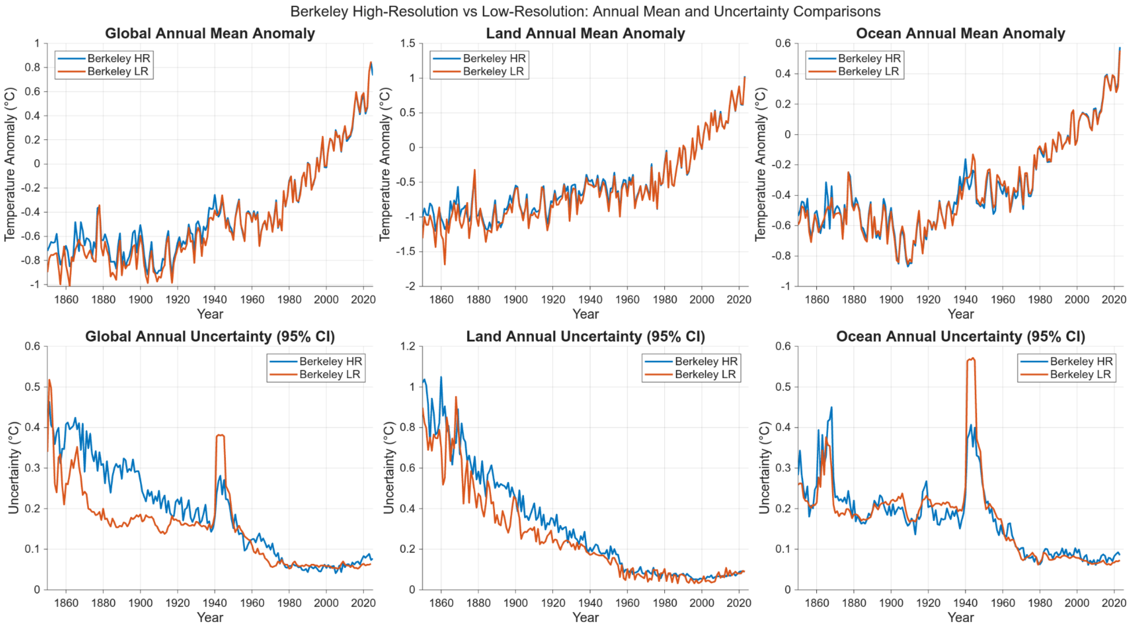


Figure 5: Comparing Global, Land, and Ocean averages from BEST-HR and BEST-LR and their associated uncertainties.

The previously shown Fig. 1 compares global average anomaly time series of BEST-HR with those of other datasets. The data products are in general agreement and generally fall within the 95 % confidence interval of BEST-HR. One exception is that DCENT-I, DCENT_MLE and COBE-STEMP3 are warmer during the 1910s, which could reflect their improved SST bias corrections relative to HadSST4. Similarly, DCENT-I and DCENT_MLE are colder during the 19th century, reflecting improved SST bias corrections during this period.

Figure 6 compares 95 % uncertainty ranges in the annual global average time series. In general, BEST-HR is in alignment with the other datasets on the magnitude of uncertainty. One exception is COBE-STEMP3, which has much less estimated uncertainty than other datasets. This could reflect COBE-STEMP3's use of high 0.25° resolution SST observations, more spatial pattern usage, and use of multiple lines of evidence for SST bias adjustments (i.e., metadata, marine air temperatures, and coastal land air temperatures). Alternatively, this could reflect missing uncertainty components in COBE-STEMP3 (e.g., LSAT homogenization uncertainty, impact of sea ice). During WW2, the two HadSST4.2-based datasets, BEST-HR and HadCRU_MLE, estimate much less uncertainty than the three HadSST4.1-based datasets, BEST-LR, HadCRUT5 Analysis, and Kadow et al. This is because the update from HadSST4.1 to HadSST4.2 reduced the range of engine room intake measurement biases from 0.0-0.5°C to 0.2-0.3°C (Sandford & Rayner, 2026).

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

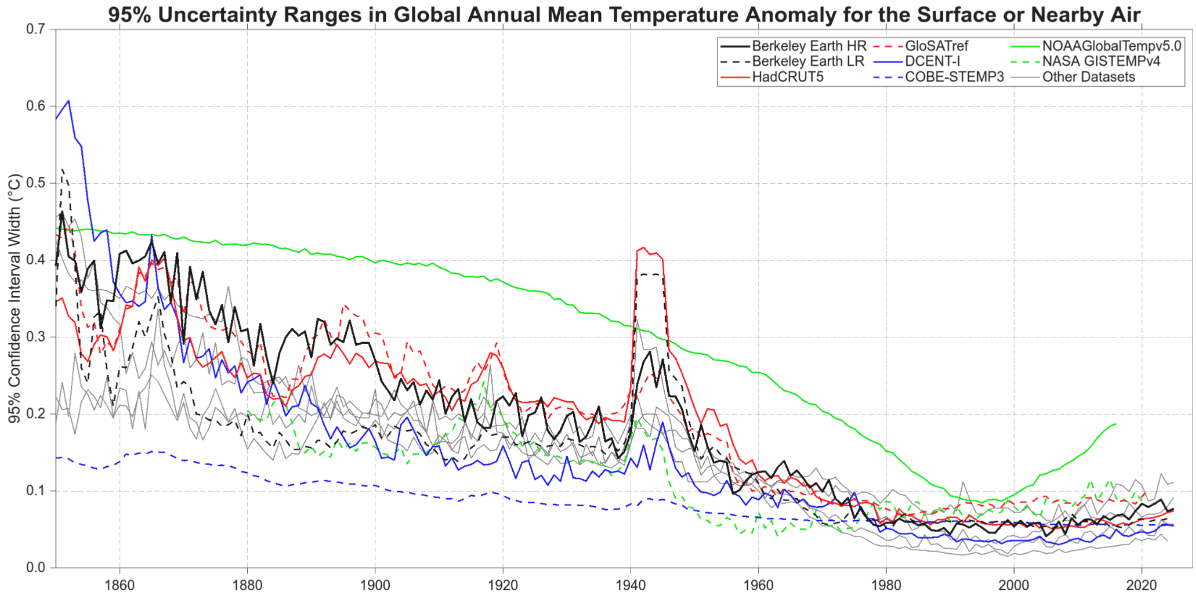


Figure 6: 95 % uncertainty ranges in global annual mean surface temperature anomalies, except for GloSATref, where surface air temperatures are used for sea regions. Other datasets include China-MST3-Imax, HadCRU_MLE, DCENT_MLE, and Kadow *et al*. Most uncertainty ranges (except for COBE-STEMP3 and China-MST3-Imax) were recalculated relative to the 1981–2010 reference period.

Since BEST-HR is an ensemble dataset, it provides temporal uncertainty information, which can be useful for many applications, including when combining with radiative forcing and ocean heat content data to produce improved temperature estimates or estimate climate emulators (e.g., Thorne et al., 2026) or for econometrics (e.g., Bruns et al., 2020; Bilal & Känzig, 2026). We calculated correlation coefficients in uncertainties in the annual global average temperature between consecutive years for each ensemble dataset (not shown). The average of this coefficient from 1850 to 2021 was 0.59 for BEST-HR, consistent with substantial temporal correlations in uncertainty. This value was more than that of GloSATref (0.29), HadCRUT5 Analysis (0.43), and DCENT-I (0.51); but less than that of GISTEMP (0.68), HadCRU_MLE (0.71), DCENT_MLE (0.75), GlobalTemp (0.83), and Kadow et al. (0.83). In general, we expect a significant degree of persistence in uncertainty arising from persistent systematic changes in the data collection networks, such as weather station location moves and instrumentation changes on both land and sea.

Figure 7 shows warming (from 1850–1900 to 2012–2021) by latitude for available temperature datasets. For equatorial- and mid-latitudes, all of the datasets are roughly in agreement on the rate of warming. Interestingly, in the Arctic, BEST-HR estimates much less warming than most other datasets despite its accounting for sea ice loss and climatological differences between sea ice and open sea regions. Perhaps this reduced Arctic warming is due to BEST-HR's use of ERA-based spatial patterns, although other datasets used reanalysis-derived spatial patterns or neural networks (i.e., NOAA GlobalTempv6.1, COBE-STEMP3, China-MST3, CMA-GMST). BEST-HR is consistent with the suggestion of Qian et al. (2026) that many instrumental temperature datasets overestimate Arctic Ocean warming due to high observed anomalies in the Barents Sea. The only dataset with less Antarctic warming is COBE-STEMP3, which is expected since COBE-STEMP3 treats sea ice regions as open sea. In the Antarctic, BEST-HR's estimate of warming is not well constrained due to limited data availability in this region.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

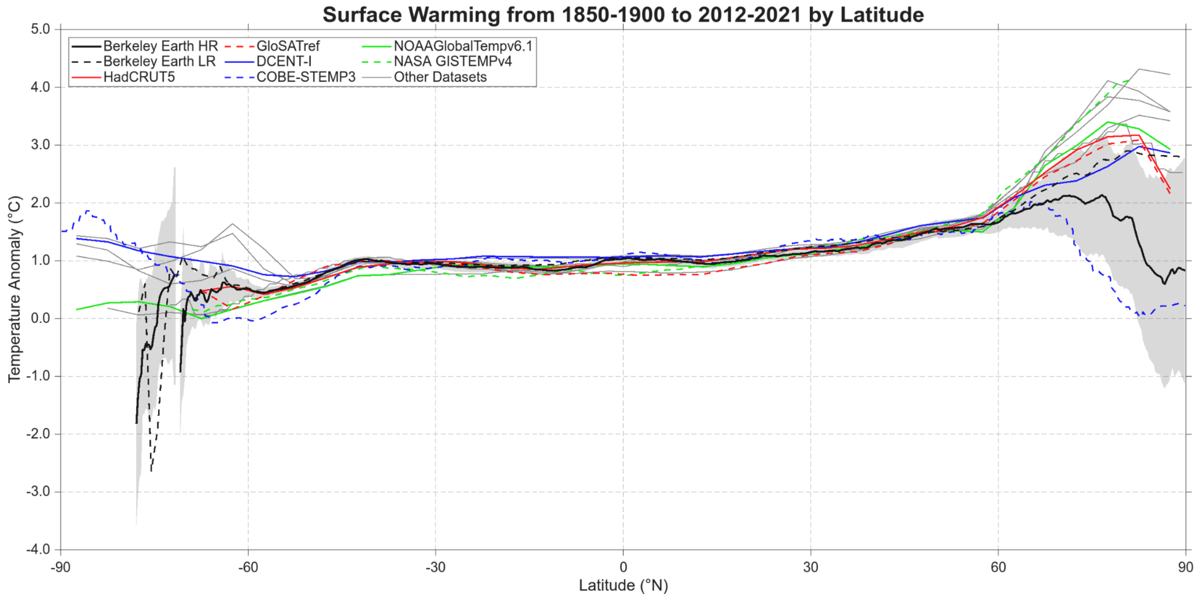
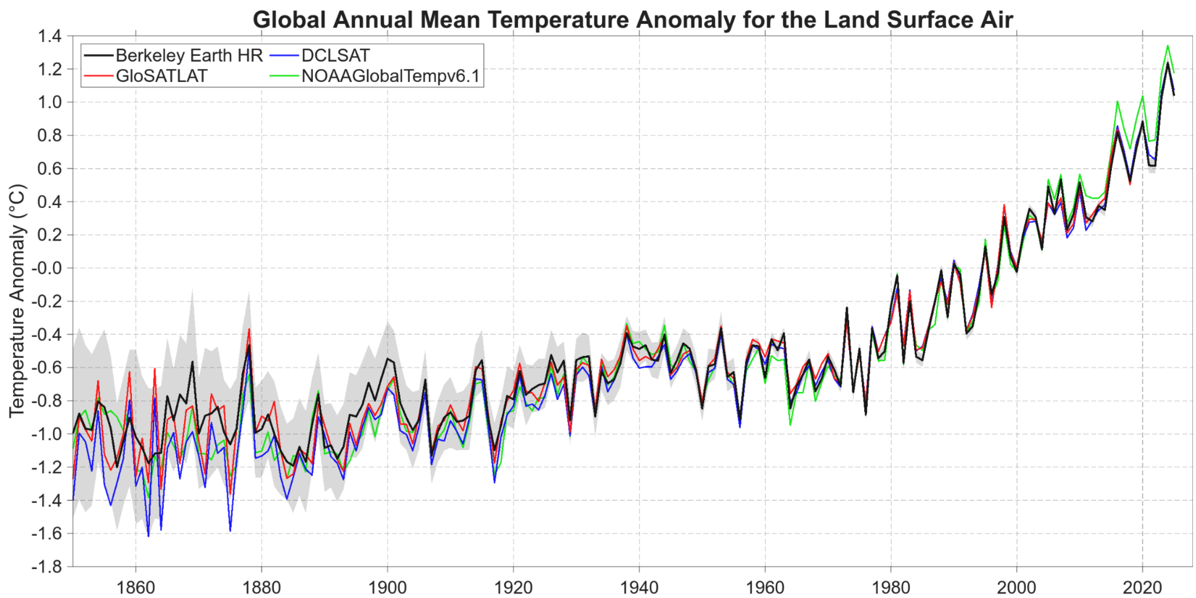


Figure 7: Average warming by latitude from 1850–1900 to 2012–2021. The shaded region reflects the 95 % ensemble uncertainties of BEST-HR and does not include coverage uncertainties. Other datasets include China-MST3-Imax, CMA-GMST, HadCRU_MLE, DCENT_MLE, and Kadow *et al*. Anomalies and uncertainties are relative to the 1981–2010 reference period.

Figure 8 compares global average land surface air temperature anomaly time series of BEST-HR with those of other datasets: GloSATLAT (Taylor et al., 2025), which is an improved version of CRUTEM5; DCLSAT, the land surface component of DCENT (Chan et al., 2024); and NOAAGlobalTempv6.1, which uses an infilled version of GHCNv4. The data products are in general agreement, except for the past decade where NOAAGlobalTempv6.1 is warmer.



https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

Figure 8: Time series of global annual land surface temperature anomalies. The shaded region reflects the 95 % uncertainties of BEST-HR. DCLSAT and GloSATLAT are the land temperature components of DCENT and GLoSAT, respectively. Anomalies and uncertainties are relative to the 1981–2010 reference period.

BEST-HR also provides land-only monthly averages of daily minimum, daily maximum, and daily average temperatures going as far back as the early 18th century for TAVG (Fig. 9), which provides a unique view into the past not available from other global instrumental temperature datasets, although the recent GloSAT project extended temperature estimates back until 1781. Surprisingly, our historic trends suggest that recent daily maximum temperatures have been warming slightly faster than daily minimum temperatures, which is contrary to the expected effects from warming due to greenhouse gas emissions (IPCC, 2021).

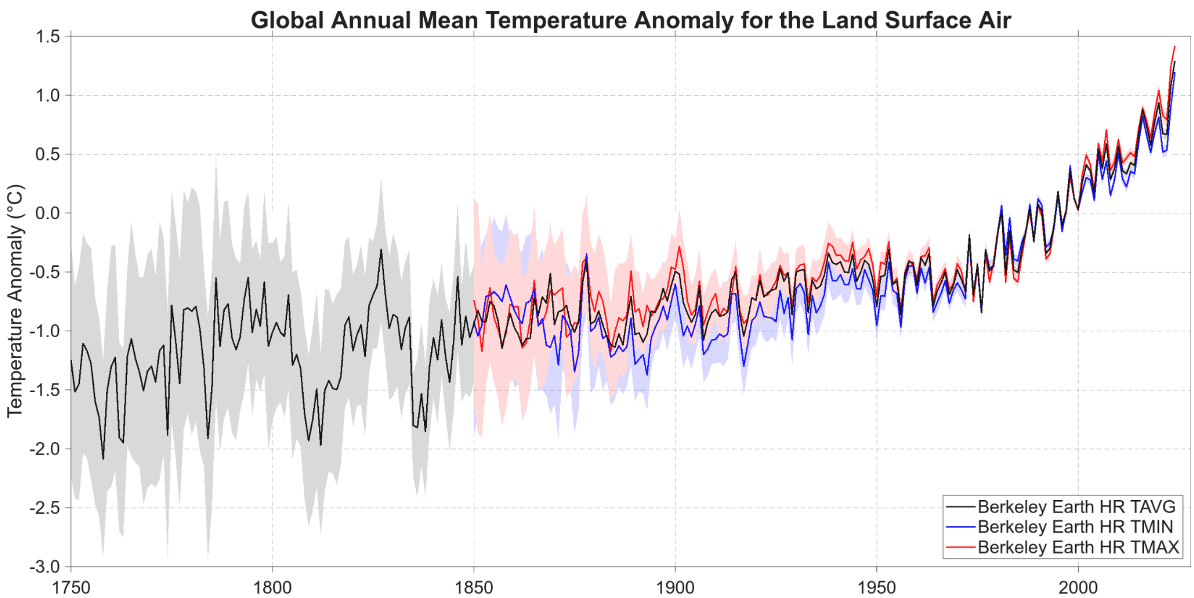


Figure 9: Time series of Berkeley Earth HR global annual average land surface temperature anomalies for maximum daily temperatures (TMAX), minimum daily temperatures (TMIN), and average daily temperatures (TAVG). Shaded regions reflect the 95 % uncertainties. Anomalies and uncertainties are relative to the 1981–2010 reference period.

6.2 Global Gridded Results

Figure 10 shows estimated warming between four different time periods, the early industrial period (1850–1900) frequently used by the Intergovernmental Panel on Climate Change, the late pre-satellite period (1948–1977), the early satellite period (1981–2010) used as a reference period by Thorne et al. (2026), and the most recent decade with data for all datasets (2012–2021). The higher resolution and spatial patterns allow BEST-HR to have more realistic warming patterns. In general, BEST-HR shows similar spatial patterns of warming to other instrumental temperature datasets (see Supplemental Material, Figs. S2–S5), such as land regions generally warming more than sea regions, and warming greatest in Arctic regions of sea ice loss. One exception is that BEST-HR estimates that the Arctic Ocean cooled from the early industrial period to the late pre-satellite period. A second exception is that, outside of the Barents Sea area, and with the exception of COBE-STEMP3 (which treats sea ice regions as open sea), BEST-HR estimates much less Arctic Ocean warming from the early satellite period to the recent decade. The Barents Sea is highly

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

observed and is an area of high warming due to sea ice melt; so it is possible that other instrumental temperature datasets are incorrectly interpolating high sea ice melt anomalies in the Barents Sea to the rest of the Arctic Ocean.

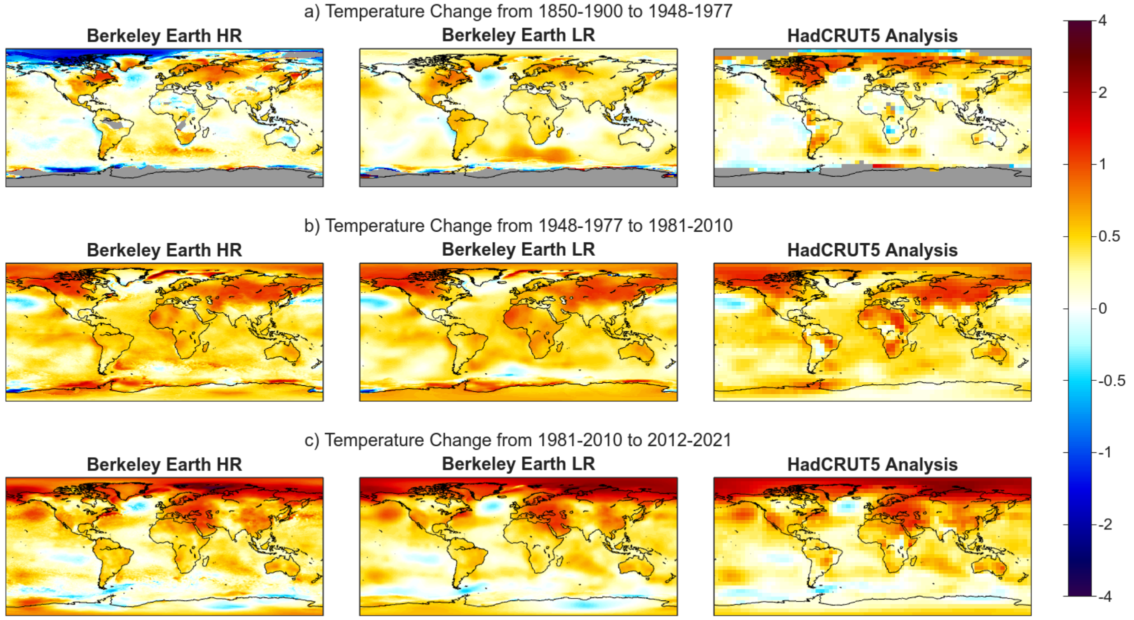
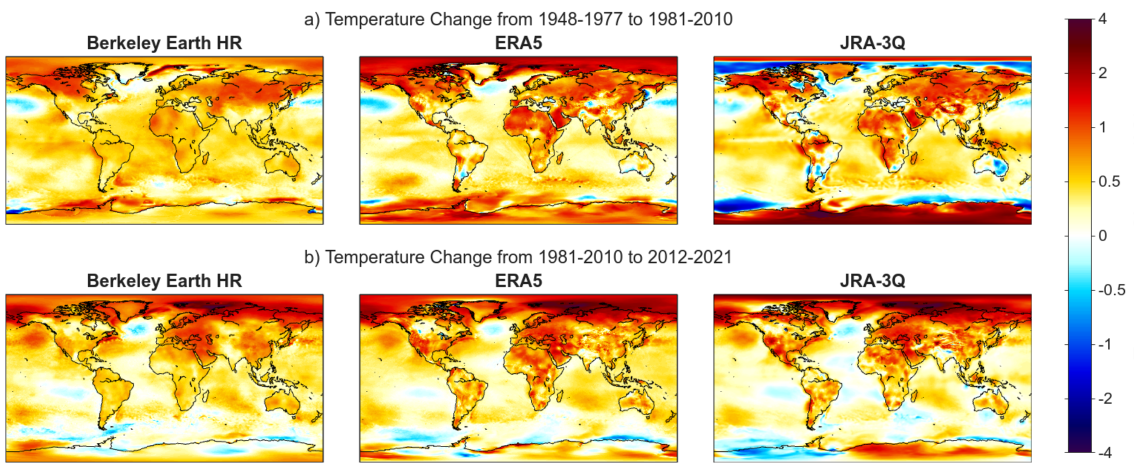


Figure 10: Average warming in °C between four time periods: 1850–1900, 1948–1977, 1981–2010, and 2012–2021.

810 Figure 11 compares estimated warming from BEST-HR with the two reanalysis datasets. The three datasets are in agreement from the early satellite era to the recent decade. However, JRA-3Q and ERA5 strongly disagree with instrumental temperature datasets on the amount and spatial distribution of warming since the pre-satellite era. This could reflect anomalous drifting of the reanalysis datasets in the absence of satellite data.



https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

Open Access Earth System Science Data Discussions

Figure 11: Average warming in °C between three time periods: 1948–1977, 1981–2010, and 2012–2021.

Figure 12 shows the BEST-HR climatology for two months: January and July. This climatology is in agreement with the climatologies of other datasets (see Supplemental Material, Figs. S6–S9). Figure 13 shows BEST-HR's spatial distribution of anomaly uncertainties for two select months: January 2000 and July 2000. BEST-HR shows much more detail, and spatial and seasonal heterogeneity in spatial uncertainty compared to other datasets (see Supplemental Material, Figs. S10–S11). In general, there is not much agreement across datasets in the spatial or seasonal distribution of uncertainty.

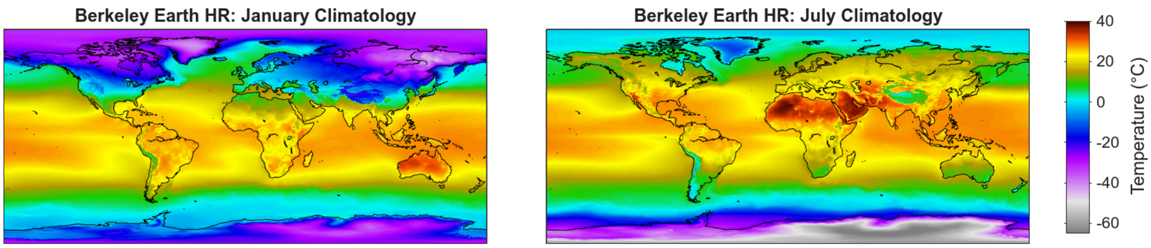


Figure 12: 1981–2010 climatologies of BEST-HR in °C for (a) January and (b) July.

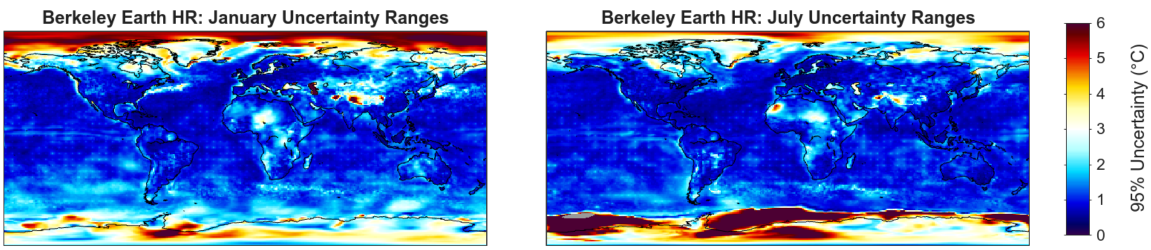


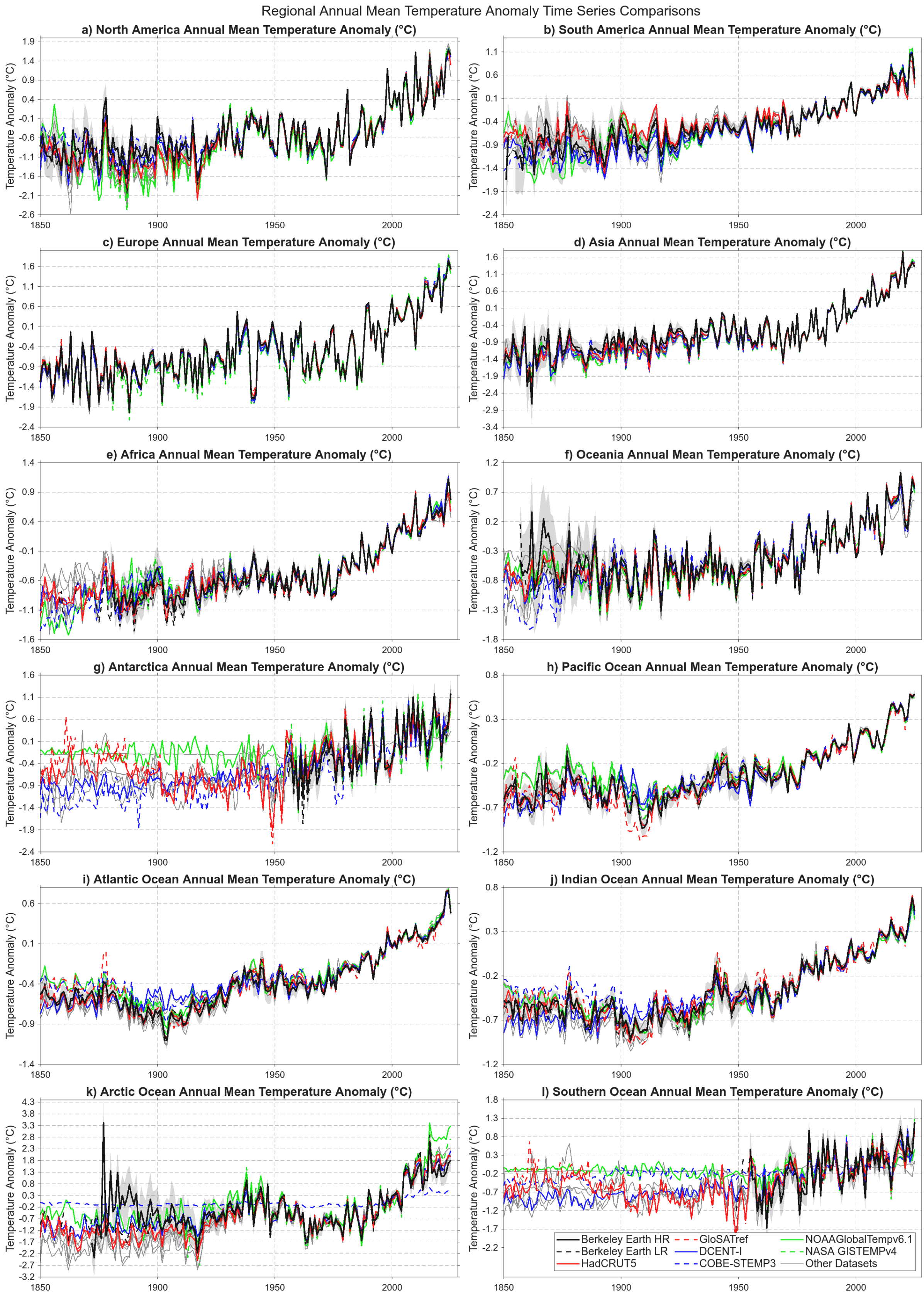
Figure 13: 95 % uncertainty ranges in °C for BEST-HR temperature anomalies relative to 1981–2010 for (a) January 2000 and (b) July 2000.

6.3 Regional Trends

Figure 14 shows temperature anomalies for BEST-HR and other temperature datasets for seven continents and five oceans. BEST-HR is generally in agreement with the trends of other datasets, with the main exceptions being the Arctic Ocean and North America, and Asia to a lesser extent, where BEST-HR estimates less warming than other datasets since the 19th century. DCENT-I, DCENT_MLE, and COBE-STEMP3 show warmer Atlantic conditions during the 1910s, likely due to their improved SST bias corrections; similarly, DCENT-I and DCENT_MLE show colder conditions in the 19th century for the Indian Ocean. COBE-STEMP3 disagrees with other datasets for the Arctic and Southern Oceans as it treats sea ice regions as open sea.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

Open Access Earth System Science Data Discussions



https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

Figure 14: Time series of annual mean surface temperature anomalies for seven continents and five oceans, except for GloSATref, where surface air temperatures are used for sea regions. The shaded region reflects 95 % ensemble uncertainties of BEST-HR. Other datasets include China-MST3-Imax, CMA-GMST, HadCRU_MLE, DCENT_MLE, and Kadow *et al.* Anomalies and uncertainties are relative to the 1981–2010 reference period.

6.4 Examples of Reconstruction Improvements

BEST-HR's use of spatial patterns greatly improves infilling of temperatures in unobserved regions. Figure 15 shows temperature anomalies for two El Niño months when observations were sparse: January 1877 and December 1940. Compared to BEST-LR, BEST-HR reconstructs more realistic El Niño spatial patterns with bands of warm water in the eastern equatorial Pacific. The non-infilled HadCRUT5 anomalies are also shown to indicate the spatial distribution of observations during these months.

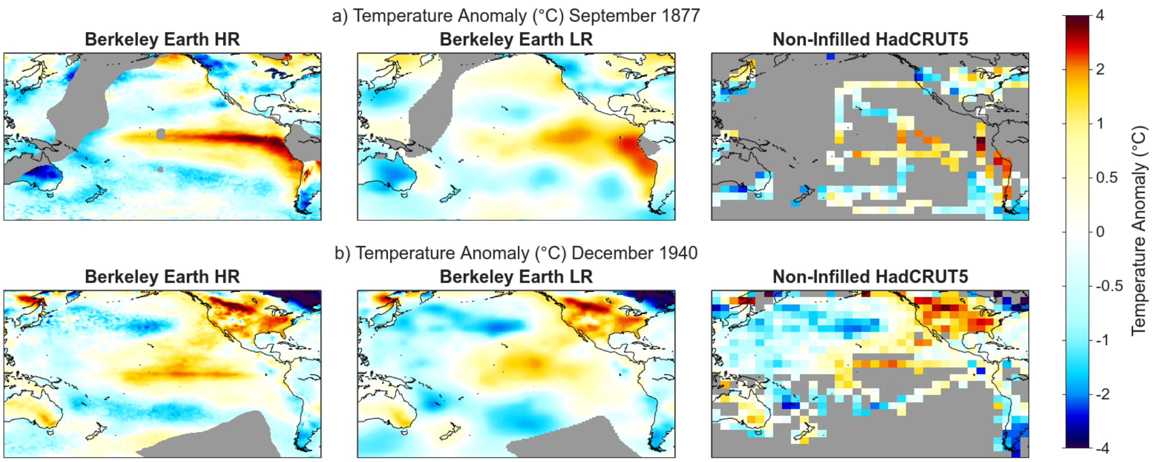


Figure 15: Temperature anomalies relative to 1981–2010 for select months: (a) El Niño of January 1877 and (b) El Niño of December 1940.

Even for years with dense observational coverage, the higher resolution and inclusion of more realistic spatial patterns allows BEST-HR to provide more realistic reconstructions of past temperatures. Figure 16 compares BEST-HR to Berkeley Earth LR, HadCRUT, and the three other high resolution datasets (COBE-STEMP3, ERA5, JRA-3Q). BEST-HR is much more able to capture detailed features such as mountain ranges (panels (a) and (b)) and ocean eddies (panels (c), (d), (e), and (f)). Although, since BEST-HR uses 5° SST data, it's not able to reproduce specific eddies in panels (d) and (f) captured by ERA5 and JRA-3Q. While COBE-STEMP3 uses higher resolution input data, so can sometimes better reconstruct these eddies, BEST-HR has much better resolution and reconstructions for land regions than COBE-STEMP3.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

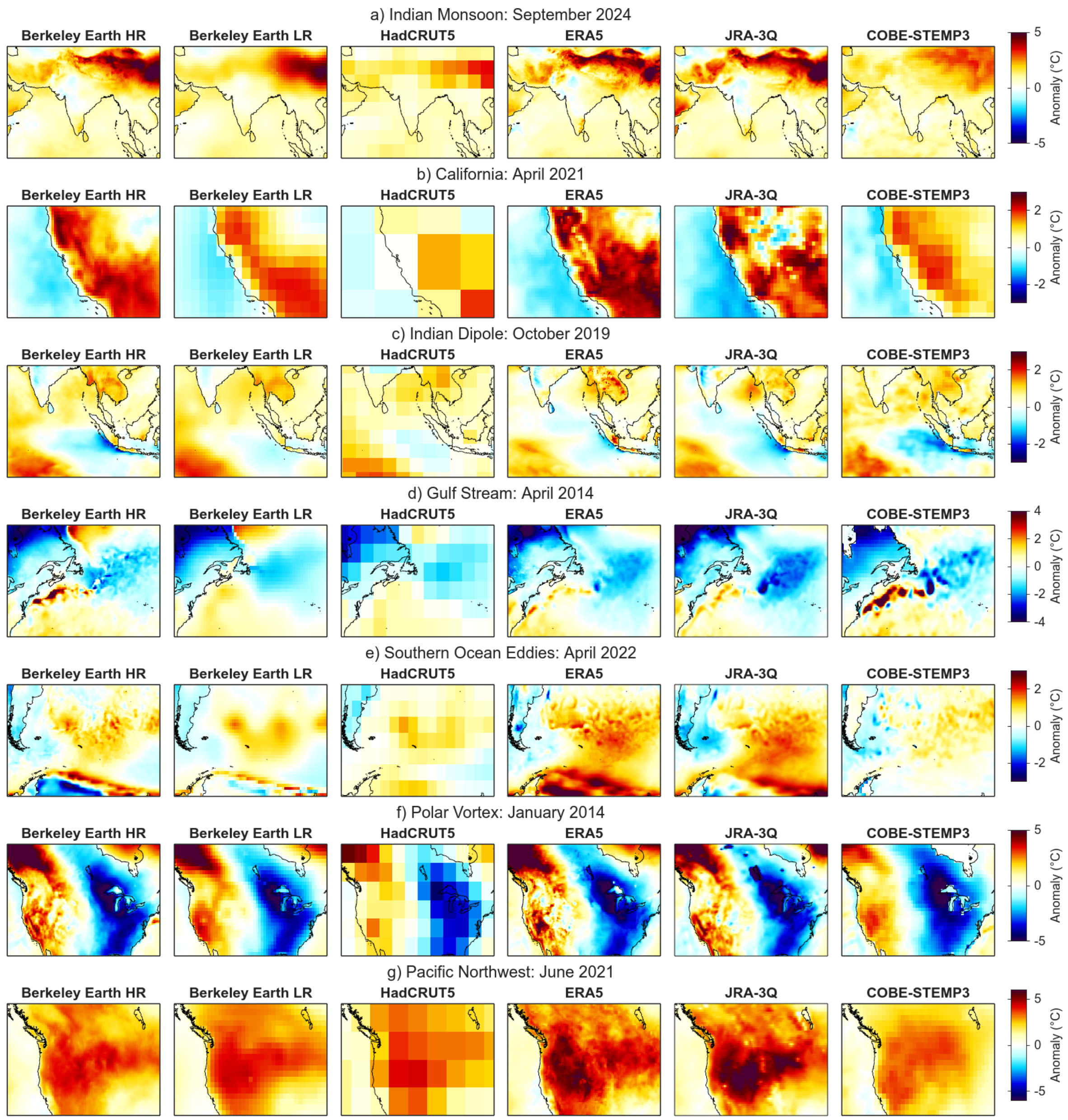


Figure 16: Temperature anomalies relative to 1981–2010 for select months and locations: (a) India September 2024, (b) California
855 December 2020, (c) Indian Dipole November 2023, (d) Gulf Stream April 2014, (e) Kuroshio February 1995, (f) Southern Ocean Eddies April 2022, (g) Polar Vortex January 2014. Air temperatures instead of sea surface temperatures were used for ERA5 and JRA-3Q. For COBE-STEMP3, 0.25° resolution COBE-SST3 is shown for 1° grid cells where COBE-LSAT3 is not available.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

Open Access Earth System Science Data Discussions

7 Conclusions

The Berkeley Earth Surface Temperature – High Resolution (BEST-HR) dataset represents a major step forward in observational climate data products. By applying a revised regression Kriging framework and increasing the native resolution to 0.25° × 0.25°, this new dataset achieves a substantial improvement in spatial detail over previous observational datasets.

One of BEST-HR's core strengths is its broad incorporation of observational data. The flexible station handling procedures and minimal continuity thresholds allow the integration of approximately 58,311 unique station records, far exceeding the coverage of comparable datasets, which typically include between 8,000 and 25,000 stations. This expansive coverage is particularly valuable in regions and periods where other products are sparse.

The methodological updates in BEST-HR primarily affect the interpolation and spatial representation of temperature data. Importantly, the large-scale trends and associated uncertainties are largely consistent with previous Berkeley Earth analyses, affirming the robustness of the long-term conclusions about global climate change. Thus, while BEST-HR enhances the resolution and detail of temperature fields, it does not alter the broader understanding of climate evolution over the instrumental period.

What BEST-HR offers, however, is a critical advance in local and regional-scale climate understanding. The improved spatial resolution enables more detailed analysis of sub-national trends, urban temperature changes, and regional extremes. This granularity is essential for urban planners, regional climate adaptation efforts, and policy development, as well as for climate scientists aiming to validate and calibrate high-resolution climate models.

In many respects, the most appropriate benchmark for comparison is not another observational dataset, but the ERA5 reanalysis, which also operates at a comparable spatial resolution. However, key differences remain: whereas ERA5 provides physically consistent global fields using data assimilation, it does not directly assimilate historical surface weather station temperature measurements–largely due to limitations in temporal resolution and metadata (e.g., daily maximum and minimum temperatures lacking time-of-observation stamps). In contrast, BEST-HR is explicitly constrained by observed surface temperature records and constructed to reflect their values where available.

This distinction is crucial. While ERA5 leverages a diverse array of observational sources–including satellites and radiosondes–it often must rely on model dynamics where direct surface constraints are unavailable. BEST-HR, by contrast, prioritizes fidelity to the observational record, and thus offers more accurate reconstructions of surface climatologies, seasonal cycles, and long-term trends in well-instrumented regions. For users seeking high-resolution observational benchmarks in such areas, BEST-HR is likely to offer superior accuracy.

As with the lower-resolution BEST dataset, we plan to extend BEST-HR with daily gridded fields in the near future, providing higher temporal resolution for applications such as climate extremes, event attribution, and daily-scale climate model validation. In parallel, Berkeley Earth is actively developing a suite of user-facing tools designed to leverage the spatial detail of BEST-HR for climate model bias correction, localized trend analysis, and scenario-based future projections.

Data availability

Ongoing updates to full Berkeley Earth Surface Temperature – High-Resolution (BEST-HR) dataset, including gridded monthly temperature fields, climatologies, and the associated uncertainty ensemble, will generally be available at: https://berkeleyearth.org/data/ In addition, static copies of the key data reported in this paper have been archived at Zenodo. The gridded results files and global summary time series have been archived at https://doi.org/10.5281/zenodo.20428381 (Rohde et al., 2026) and the merged station data used to create these results have been archived at https://doi.org/10.5281/zenodo.20427092 (Rohde, 2026).

The data are released under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Code availability

A snapshot of the source code used in this analysis is available via GitHub at https://github.com/BerkeleyEarthOrg/TemperatureAnalysis-Snapshot

Author contributions

Robert Rohde designed and implemented the updated analysis framework, led the methodological development, and led the manuscript writing. Zeke Hausfather supported scientific validation, interpretation, and manuscript preparation. Bruce Calvert supported manuscript preparation, with a focus on figures, providing context to the broader literature, and Sect. 6. All authors contributed to the writing and revision of the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We thank the many national meteorological services and international data centers that provide open access to historical weather observations. We are also grateful to the developers and maintainers of ERA5, ORAS5, HadSST4, and CESM2, which were critical inputs to this analysis. While nearly all of the development of this analysis predates the existence of AI tools, we note that GitHub Copilot was used to help produce some of the final figure generating code, and ChatGPT was asked to review a draft of this paper to help identify weaknesses. AI played no role in designing the core algorithm or the description of the methods presented here.

Financial support

Funds to support this work were provided by the Patrick J. McGovern Foundation and Quadrature Climate Foundation.

References

Bilal, A. & Känzig, D.R.: The Macroeconomic Impact of Climate Change: Global vs. Local Temperature, NBER Working Paper, 32450, https://doi.org/10.3386/w32450, 2026.

British Antarctic Survey: Met READER, MET-READER [data set], https://legacy.bas.ac.uk/met/READER/surface/stationpt.html, 2020.

Bruns, S.B., Csereklyei, Z. & Stern, D.I.: A multicointegration model of global climate change, J. Econometrics, 214(1), 175-197, https://doi.org/10.1016/j.jeconom.2019.05.010, 2020.

Calvert, B.T.T.: Improving global temperature datasets to better account for non-uniform warming, Q. J. R. Meteorol. Soc., 150(763), 3672-3702. https://doi.org/10.1002/qj.4791, 2024a.

Calvert, B.T.T.: Maximum Likelihood Estimates of Temperatures using Data from the Dynamically Consistent Ensemble of Temperature (Version 1.0), World Data Center for Climate (WDCC) at DKRZ. https://doi.org/10.26050/WDCC/DCENT_MLE_v1_0, 2024b.

Chan, D., Chan, S.C., Siddons, J.T., Cable, A., Faulkner, A., Cornes, R.C., Kent, E.C., Gebbie, G. & Huybers, P.: DCENT-I: A Globally Infilled Extension of the Dynamically Consistent ENsemble of Temperature Dataset, Geosci. Data J., 13(2), e70054, https://doi.org/10.1002/gdj3.70054, 2026.

Chan, D., Gebbie, G., Huybers, P. & Kent, E.: DCENT: Dynamically Consistent ENsemble of Temperature at the Earth surface since 1850 from the DCENT dataset, Sci. Data, 11, 953, https://doi.org/10.1038/s41597-024-03742-x, 2024.

Chen, L., Xu, W., Zhou, Z., Cao, L., Yang, S. & Xu, C.: A new global land–ocean merged surface temperature dataset since the 1850s: the CMA-GMST dataset, Clim. Dyn., 63(4), 187, https://doi.org/10.1007/s00382-025-07614-x, 2025.

European Climate Assessment & Dataset: Daily data, ECA&D [data set], https://www.ecad.eu/dailydata/index.php, 2026.

Fausto, R.S., How, P., Vandecrux, B., Lund, M.C., Box, J.E., Mankoff, K.D., Andersen, S.B., van As, D., Bahbah, R., Citterio, M., Colgan, W., Jakobsgaard, H.T., Karlsson, N.B., Kjeldsen, K.K., Larsen, S.H., Olsen, C., Oraschewski, F., Rutishauser, A., Shields, C.L., Solgaard, A.M., Stevens, I.T., Svendsen, S.H., Langley, K., Messerli, A., Bjørk, A.A., Andersen, J.K., Abermann, J., Steiner, J., Prinz, R., Hynek, B., Lea, J.M., Brough, S., & Ahlstrøm, A.P.: PROMICE | GC-NET automatic weather station data, Earth Syst. Sci. Data, 18(4), 2829-2873, https://doi.org/10.5194/essd-2025-687, 2026.

Global Climate Observing System: GSN Data Sets, The Global Climate Observing System Surface Network (GSN) data set, GSN [data set], https://gcos.dwd.de/DWD-GCOS/EN/nationalcontributions/servicesforgcos/centresforqualityassurance/gsmnc/gsnmc_gsn_datensaetze/gsnmc_gsn_datensaetze_node.html, 2026a.

Global Climate Observing System: Index of /climate_environment/CDC/observations_global/CLIMAT/, GCOS Monthly CLIMAT Summaries [data set], https://opendata.dwd.de/climate_environment/CDC/observations_global/CLIMAT/, 2026b.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Huang, B., Yin, X., Menne, M.J., Vose, R. & Zhang, H.-M.: Improvements to the Land Surface Air Temperature Reconstruction in NOAAGlobalTemp: An Artificial Neural Network Approach, Artif. Intell. Earth Syst., 1(4), e22003, https://doi.org/10.1175/AIES-D-22-0032.1, 2022.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

Open Access Earth System Science Data Discussions

Huang, B., Yin, X., Menne, M.J., Vose, R. & Zhang, H.-M.: Extended Reconstructed Sea Surface Temperature, Version 6 (ERSSTv6). Part I: An Artificial Neural Network Approach, J. Climate, 38(4), 1105-1121, https://doi.org/10.1175/JCLI-D-23-0707.1, 2025.

IPCC (2021): Chapter 11: Weather and Climate Extreme Events in a Changing Climate, 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. Edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R. & Zhou, B., Cambridge University Press, https://doi.org/10.1017/9781009157896.013, 2021.

Ishii, M., Nishimura, A., Yasui, S. & Hirahara, S.: Historical high-resolution daily SST analysis (COBE-SST3) with consistency to monthly land surface air temperature, J. Meteorol. Soc. Jpn., 103(1), 17-44, https://doi.org/10.2151/jmsj.2025-002, 2025.

Kadow, C., Hall, D.M. & Ulbrich, U.: Artificial intelligence reconstructs missing climate information, Nat. Geosci., 13(6), 408-413, https://doi.org/10.1038/s41561-020-0582-5, 2020.

Kennedy, J.J., Rayner, N.A., Smith, R.O., Parker, D.E. & Saunby, M.: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties, J. Geophys. Res. Atmos., 116(D14), D14103, https://doi.org/10.1029/2010JD015218, 2011.

Kennedy, J.J., Rayner, N.A., Atkinson, C.P. & Killick, R.E.: An ensemble data set of sea-surface temperature change from 1850: the Met Office Hadley Centre HadSST.4.0.0.0 data set, J. Geophys. Res. Atmos., 124(14), 7719–7763, https://doi.org/10.1029/2018JD029867, 2019.

Lawrimore, J.H., Menne, M.J., Gleason, B.E., Williams, C.N., Wuertz, D.B., Vose, R.S. & Rennie, J.: An overview of the Global Historical Network monthly mean temperature dataset, version 3, J. Geophys. Res. Atmos., 116(D19), D19121, https://doi.org/10.1029/2011JD016187, 2011.

Lenssen, N.J.L., Schmidt, G.A., Hansen, J.E., Menne, M.J., Persin, A., Ruedy, R. & Zyss, D.: Improvements in the GISTEMP uncertainty model, J. Geophys. Res. Atmos., 124(2), 6307–6326, https://doi.org/10.1029/2018JD029522, 2019.

Li, Z., Li, Q., Jiao, B., Xu, Q., Wei, S., Ru, X., Si, P., Chao, L., Zhang, H., Lin, J., Liao, L., Zhang, H., Huang, B. & Jones, P.: An Integrated Uncertainty Framework for the China-MST 3.0 Global Surface Temperature Data Set, J. Geophys. Res. Atmos., 131(8), e2025JD044732, https://doi.org/10.1029/2025JD044732, 2026.

Lundstad, E., Brugnara, Y., Pappert, D., Kopp, J., Samakinwa, E., Hürzeler, A., Andersson, A., Chimani, B., Cornes, R., Demarée, G., Filipiak, J., Gates, L., Ives, G.L., Jones, J.M., Jourdain, S., Kiss, A., Nicholson, S.E., Przybylak, R., Jones, P., Rousseau, D., Tinz, B., Rodrigo, F.S., Grab, S., Domínguez-Castro, F., Slonosky, V., Cooper, J., Brunet, M., Brönnimann, S.: The global historical climate database HCLIM, Sci. Data, 10, 44, https://doi.org/10.1038/s41597-022-01919-w, 2023.

Meier, W.N., Fetterer, F., Windnagel, A.K., Stewart, J.S., Stafford, T.: NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 6, National Snow and Ice Data Center, https://doi.org/10.7265/b18j-z797, 2026.

Menne, M.J., Williams, C.N.J., & Vose, R.S.: The U.S. Historical Climatology Network Monthly Temperature Data, Version 2, Bull. Am. Meteorol. Soc., 90(7), 993-1008, https://doi.org/10.1175/2008BAMS2613.1, 2009.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E., & Houston, T.G.: An overview of the Global Historical Climatology Network-Daily Database, J. Atmos. Oceanic Technol., 29, 897–910, https://doi.org/10.1175/JTECH-D-11-00103.1, 2012.

Menne, M.J., Williams, C.N., Gleason, B.E., Rennie, J.J., & Lawrimore, J.: The Global Historical Climatology Network Monthly Temperature Dataset, Version 4, J. Clim., 31(24), 9835–9854. https://doi.org/10.1175/JCLI-D-18-0094.1, 2018.

Midwestern Regional Climate Center: Military Forts and Voluntary Observers Database, FORTS [data set], Purdue University & NOAA, https://mrcc.purdue.edu/FORTS, 2026.

Morice, C.P., Berry, D.I., Cornes, R.C., Cowtan, K., Cropper, T., Hawkins, E. Kennedy, J.J., Osborn, T.J., Rayner, N.A., Schurer, A.P., Taylor, M., Rivas, B.R., Teleti, P.R., Wallis, E.J., Winn, J. & Kent, E.C.: An observational record of global gridded near surface air temperature change over land and ocean from 1781, Earth Syst. Sci. Data, 17(12), 7079-7100, https://doi.org/10.5194/essd-17-7079-2025, 2025.

Morice, C.P., Kennedy, J.J., Rayner, N.A., & Jones, P.D.: An updated assessment of near-surface temperature change from 1850: The HadCRUT5 dataset, J. Geophys. Res. Atmos., 126, e2019JD032361, https://doi.org/10.1029/2019JD032361, 2021.

NOAA: The Colonial Era Archive Data Project, Historical Data for Data Sparse Underdeveloped Countries, Colonial Era Weather Archive [data set], https://web.archive.org/web/20170223182307/http://www.ncdc.noaa.gov/oa/climate/research/ghcn/colonialarchive.html, 2008.

NOAA: World Weather Records (WWR) Clearinghouse, WWR [data set], https://www.ncei.noaa.gov/products/world-weather-records, 2017.

NOAA: U.S. Daily Surface Data (COOP Daily/Summary of Day), COOP [data set], https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc%3AC00314, 2023a.

NOAA: Monthly Climatic Data for the World: NERMS - Monthly Climatic Data for the World, NERMS [data set], https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc%3AC01037, 2023b.

NOAA: Global Surface Summary of the Day - GSOD, GSOD [data set], https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc%3AC00516, 2025.

NOAA: README FILE FOR THE SYNOPTIC SUMMARY OF THE DAY (SSOD) DATASET: Version 2.0.0, SSOD [data set], https://www.ncei.noaa.gov/oa/synoptic-summary-of-the-day/v2/doc/ssodv2_DOCUMENTATION.pdf, 2026a.

Osborn, T.J., Jones, P.D., Lister, D.H., Morice, C.P., Simpson, I.R., Winn, J.P., Hogan, E. & Harris, I.C.: Land surface air temperature variations across the globe updated to 2019: the CRUTEM5 data set, J. Geophys. Res. Atmos., 126(2), e2019JD032352, https://doi.org/10.1029/2019JD032352, 2021.

Qian, Z., Liu, T., Bathiany, S., Yang, S., Hess, P., Bochow, N., Burmester, C., Gelbrecht, M., Groenke, B. & Boers, N.: Generative deep learning improves reconstruction of global historical climate records, https://doi.org/10.48550/arXiv.2602.16515, 2026.

Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P., Kent, E.C. & Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, J. Geophys. Res. Atmos., 108(D14), 4407, https://doi.org/10.1029/2002JD002670, 2003.

Rennie, J., Lawrimore, J.H., Gleason, B.E., Thorne, P.W., Morice, C.P., Menne, M.J., Williams, C.N., Gambi de Almeida, W., Christy, J.R., Flannery, M., Ishihara, M., Kamiguchi, K., Klein-Tank, A.M.G., Mhanda, A., Lister, D.H., Razuvaev, V., Renom, M., Rusticucci, M., Tandy, J., Woley, S.J., Venema, V., Angel, W., Brunet, M., Battore, B., Diamond, H., Lazzara, M.A., Le Blancq, F., Luterbacher, J., Mächel, H., Revadekar, J., Vose, R.S. & Yin, X.: The international surface temperature initiative global land surface databank: monthly temperature data release description and methods, Geosci. Data J., 1(2), 75-102, https://doi.org/10.1002/gdj3.8, 2014.

Rohde, R.A., Muller, R.A., Jacobsen, R., Perlmutter, S., Rosenfeld, A.H., Wurtele, J., Groom, D. & Wickham, C.: A new estimate of the average Earth surface land temperature spanning 1753 to 2011, Geoinform. Geostat.: An Overview, 1(1), 1000101, https://berkeley-earth-wp-offload.storage.googleapis.com/wp-content/uploads/2022/12/03232408/Results-Paper-Berkeley-Earth.pdf, 2013a.

Rohde, R.A., Muller, R.A., Jacobsen, R., Perlmutter, S., Rosenfeld, A.H., Wurtele, J., Curry, B., Wickham, C. & Mosher, S.: Berkeley Earth Temperature Averaging Process, Geoinform. Geostat.: An Overview, 1(2), 1000103, https://berkeley-earth-wp-offload.storage.googleapis.com/wp-content/uploads/2022/12/03232409/Methods-GIGS-1-103.pdf, 2013b.

Rohde, R.A., & Hausfather, Z.: The Berkeley Earth Land/Ocean Temperature Record, Earth Syst. Sci. Data, 12, 3469–3502, https://doi.org/10.5194/essd-12-3469-2020, 2020.

Rohde, R.A., Hausfather, Z., & Calvert, B. (2026). Berkeley Earth High Resolution Dataset - April 2026 (Apr 2026) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.20428381

Rohde, R. (2026). Berkeley Earth Merged Weather Station Database - April 2026 (Apr 2026) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.20427092

Sandford, C. & Rayner, N.: Addressing the World War 2 Warm Anomaly in HadSST.4.2.0.0, Int. J. Clim. [preprint], https://arxiv.org/abs/2602.03347, 2026.

Sun, W. Yang, Y., Chao, C., Dong, W., Huang, B., Jones, P. & Li, Q.: Description of the China global Merged Surface Temperature version 2.0, Earth Syst. Sci. Data, 14(4), 1677-1693, https://doi.org/10.5194/essd-14-1677-2022, 2022.

Taylor, M., Osborn, T.J., Cowtan, K., Morice, C.P., Jones, P.D., Wallis, E.J. & Lister, D.H.: GloSAT LATsdb: A Global Compilation of Land Air Temperature Station Records With Updated Climatological Normals From Local Expectation Kriging, Geosci. Data J., 12(4), e70024, https://doi.org/10.1002/gdj3.70024, 2025.

Thorne, P.W., Nicklas, J.M., Kennedy, J.J., Calvert, B., Fox-Kemper, B., Richardson, M.T., Simmons, A., Hawkins, E., Rohde, R., Cowtan, K., Abram, N.J., Andersson, A., Noone, S., Marbaix, P., Lenssen, N., Olonscheck, D., Walsh, T., Outten, S., Bethke, I., Samset, B.H., Smith, C., Pirani, A., Fuglestvedt, J., Rajamani, L., Betts, R.A., Kent, E.C., Trewin, B., Morice, C., Osborn, T., Burgess, S.N., Geden, O., Parnell, A., Forster, P.M., Hewitt, C., Hausfather, Z., Masson-Delmotte, V., Marotzke, J., Gillett, N., Seneviratne, S.I., Schmidt, G.A., Chan, D., Brönnimann, S., Reisinger, A., Menne, M., Rojas Corradi, M., Kadow, C., Huybers, P., Stephenson, D.B., Wallis, E., Rogelj, J., Schurer, A., McKinnon, K., Zhai, P., Driouech, F., Moufouma Okia, W., Vazifehkhah, S., Szopa, S., Merchant, C.J., Hirahara, S., Ishii, M., Engelbrecht, F.A., Li, Q., Lee, J.-Y., Cannon, A.J., Cassou, C., von Schuckmann, K., Delju, A.H. & Murtagh, E.: How well can we quantify when 1.5°C of global warming has been exceeded?, Earth Syst. Sci. Data [preprint], https://doi.org/10.5194/essd-2025-825, 2026.

https://doi.org/10.5194/essd-2026-412
Preprint. Discussion started: 1 September 2026

Titchner, H. A., & Rayner, N. A.: The Met Office HadISST.2.2.0.0 sea ice concentration data set, J. Geophys. Res.: Atmospheres, 119, 2864–2889, https://doi.org/10.1002/2013JD020316, 2014.

UCAR: World Monthly Surface Station Climatology, World Monthly Surface Station Climatology [data set], https://data.ucar.edu/en/dataset/world-monthly-surface-station-climatology, 2026a.

UCAR: NCDC TD3210 U.S. First Order Summary of Day, U.S. First Order Summary of Day [data set], https://doi.org/10.5065/H0J8-V485, 2026b.

Vose, R.S., McNeill, S., Thomas, K. & Shepherd, E.: Enhanced Master Station History Report, EMSHR [data set], https://doi.org/10.7289/V5NV9G8D, 2011.

Wessel, P., & Smith, W.H.F.: A global self-consistent, hierarchical, high-resolution shoreline database, J. Geophys. Res.: Solid Earth, 101(B4), 8741–8743, https://doi.org/10.1029/96JB00104, 1996.

World Meteorological Organization: Pub.9-Vol.A Legacy Page, OSCAR Vol A [data set], https://oscar.wmo.int/surface/#/, 2021.

Yin, X., Huang, B., Menne, M., Vose, R., Zhang, H.-M., Adeyeye, A., Applequist, S., Gleason, K., Liu, Chunying & Sanchez-Lugo, A.: NOAAGlobalTemp Version 6: An AI-Based Global Surface Temperature Dataset, Bull. Amer. Meteorol. Soc., 105(11), E2184-E2193. https://doi.org/10.1175/BAMS-D-24-0012.1, 2024.

Zuo, H., Balmaseda, M.A., Tietsche, S., Mogensen, K., & Mayer, M.: The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: ORAS5, Ocean Sci., 15, 779–808, https://doi.org/10.5194/os-15-779-2019, 2019.