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

Supplemental Material

S.1. Minor methodological notes

For global average temperature anomaly time series presented in this paper that were not readily available from the data source (e.g., regional or zonal averages), these were calculated by first spatially averaging by weighting each observed grid cell by its surface area, and then averaged temporally by weighting each month by its number of days, including leap days.

S.2. Number of observations time series

Figure S1 shows the number of observations used by BEST-HR. In general, BEST-HR uses more sea "observations" than land "observations". Although, a direct comparison is somewhat misleading since multiple LSAT measurements taken by a single weather station in a month is considered a single "observation". In contrast, HadSST4 considers each measurement to be an "observation". The number of SST superobservations is a better analog of land "observations", since multiple measurements by a single observing platform (e.g., ship, buoy) within a single 5° grid cell and month are considered a single superobservation.

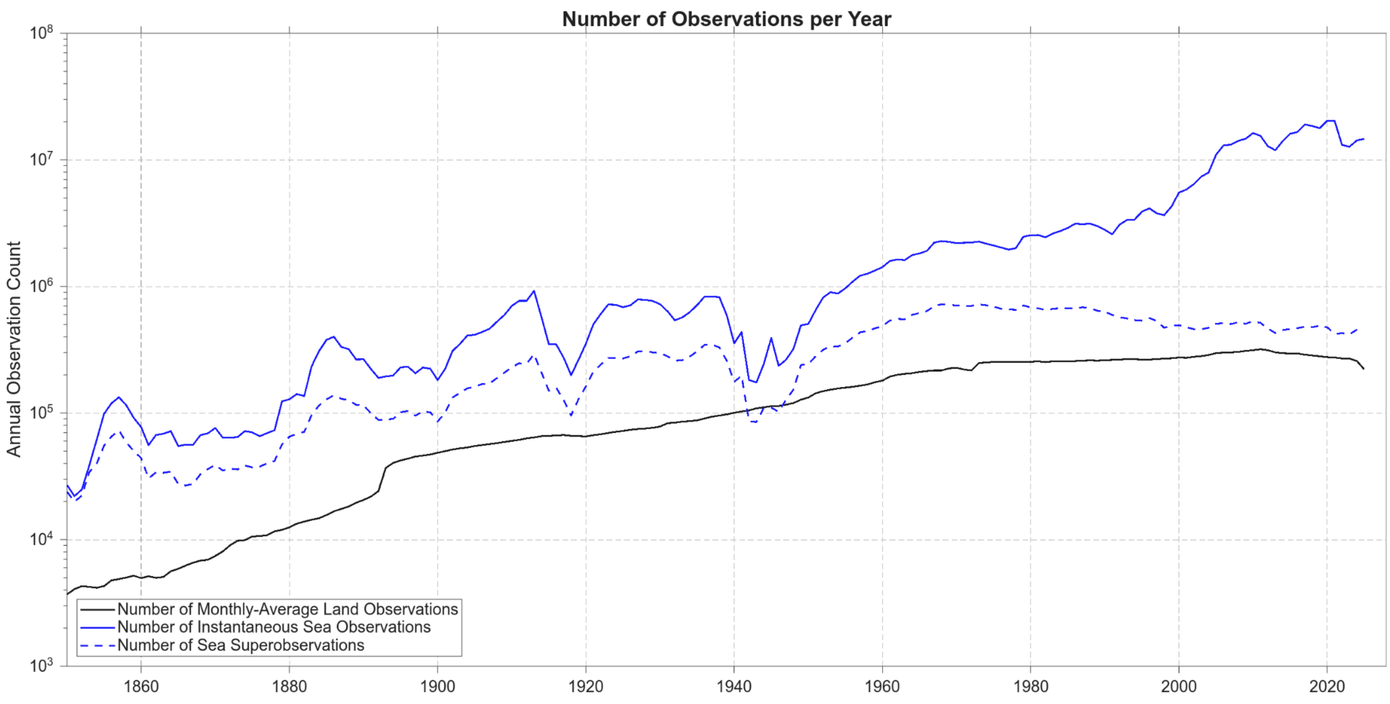

Figure S1: Number of observations used by BEST-HR.

S.3. 1850–1900 to 1948–1977 temperature change

Figure S2 shows average annual warming from 1850–1900 to 1948–1977. BEST-HR provides a higher resolution but similar spatial pattern than most other datasets. The biggest difference is in the Arctic Ocean, where BEST-HR estimates a cooling. COBE-STEMP3 shows very different spatial patterns for Greenland and Siberia compared to other datasets. HadCRU_MLE, DCENT_MLE, and Kadow et al. show bands of warming in the Antarctic, reflecting bands of sea ice loss, based on HadISST2 in HadCRU_MLE and DCENT_MLE. As explained by Thorne et al. (2026), sea ice loss in the Antarctic during this period could have been exaggerated in HadISST2, which could cause exaggerated warming bands in HadCRU_MLE and DCENT_MLE.

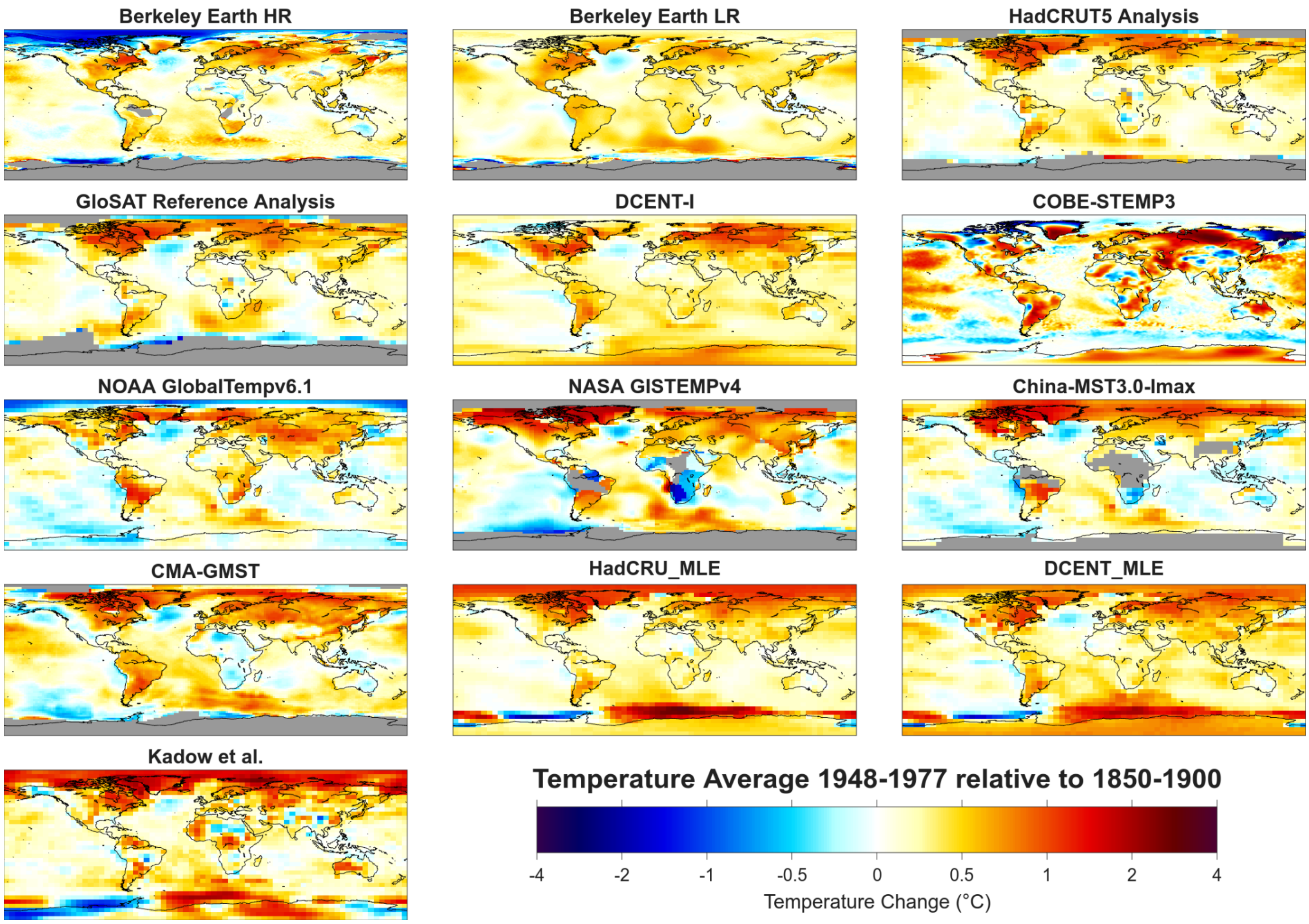

Figure S2: 1850–1900 to 1948–1977 change in annual mean surface temperature (°C) according to 13 datasets. GloSAT uses surface air temperatures rather than sea surface temperatures for open sea regions.

Figure S3 is the same as S2, but emphasizes a region around the United States to better show the differences in resolution.

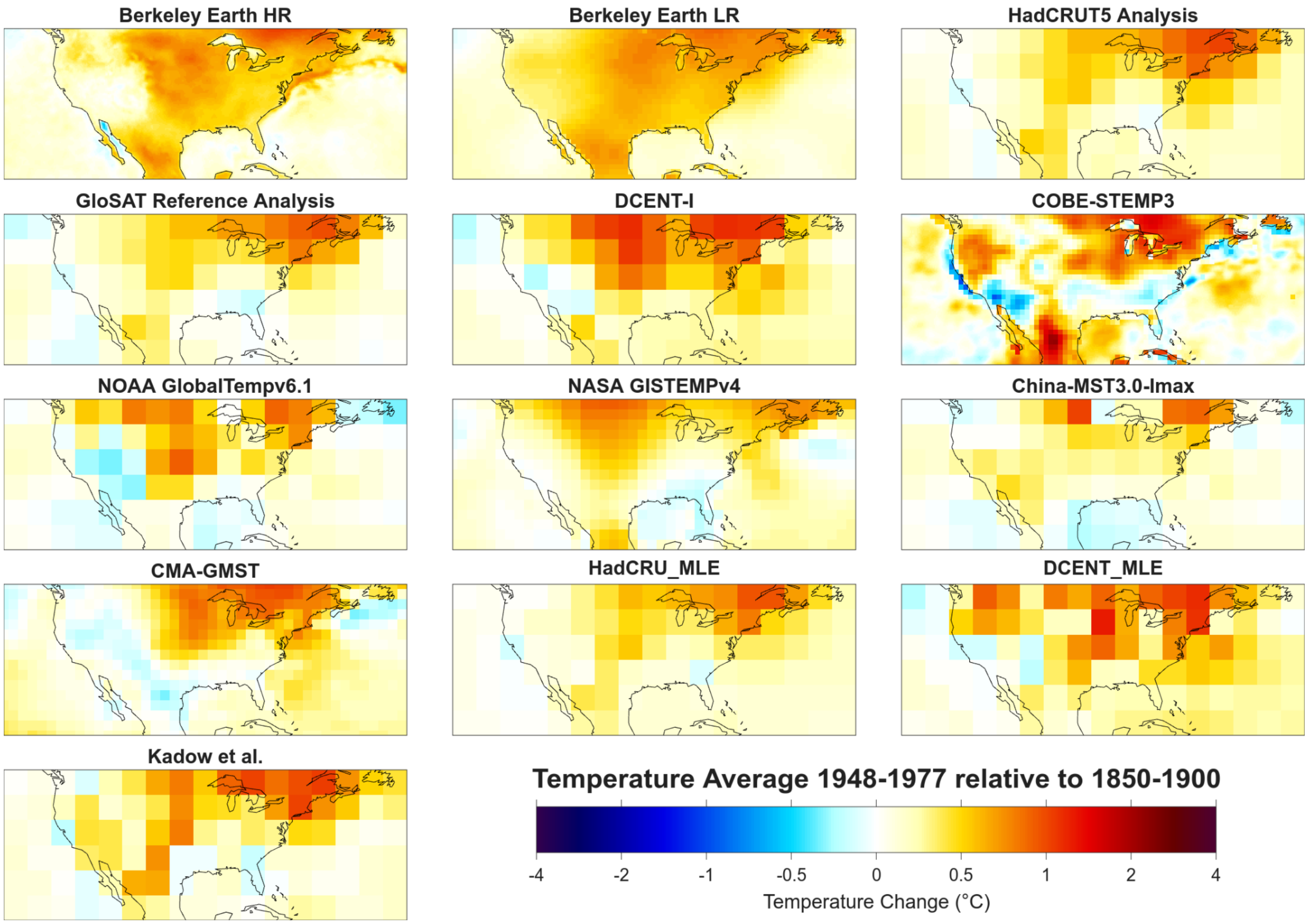

Figure S3: 1850–1900 to 1948–1977 change in annual mean surface temperature (°C) according to 13 datasets, with a focus on the United States to show the impact of differences in resolution. GloSAT uses surface air temperatures rather than sea surface temperatures for open sea regions.

S.4. 1948–1977 to 1981–2010 temperature change

Figure S4 shows average annual warming from 1948–1977 to 1981–2010. BEST-HR provides a higher resolution but similar spatial pattern than most other datasets. The biggest difference is that the two reanalysis datasets, ERA5 and JRA-3Q, show anomalous trends in the Arctic and Antarctic, likely reflecting a lack of data during the pre-satellite era. The four instrumental datasets that account for climatological differences between sea ice and open sea (i.e., BEST-HR, BEST-LR, HadCRU_MLE, and DCENT_MLE), as well as ERA5, estimate warming in the Greenland Sea from sea ice loss, which is missing from other datasets. COBE-STEMP3 shows essentially zero warming in the Arctic Ocean, reflecting its approach of treating sea ice regions as open sea.

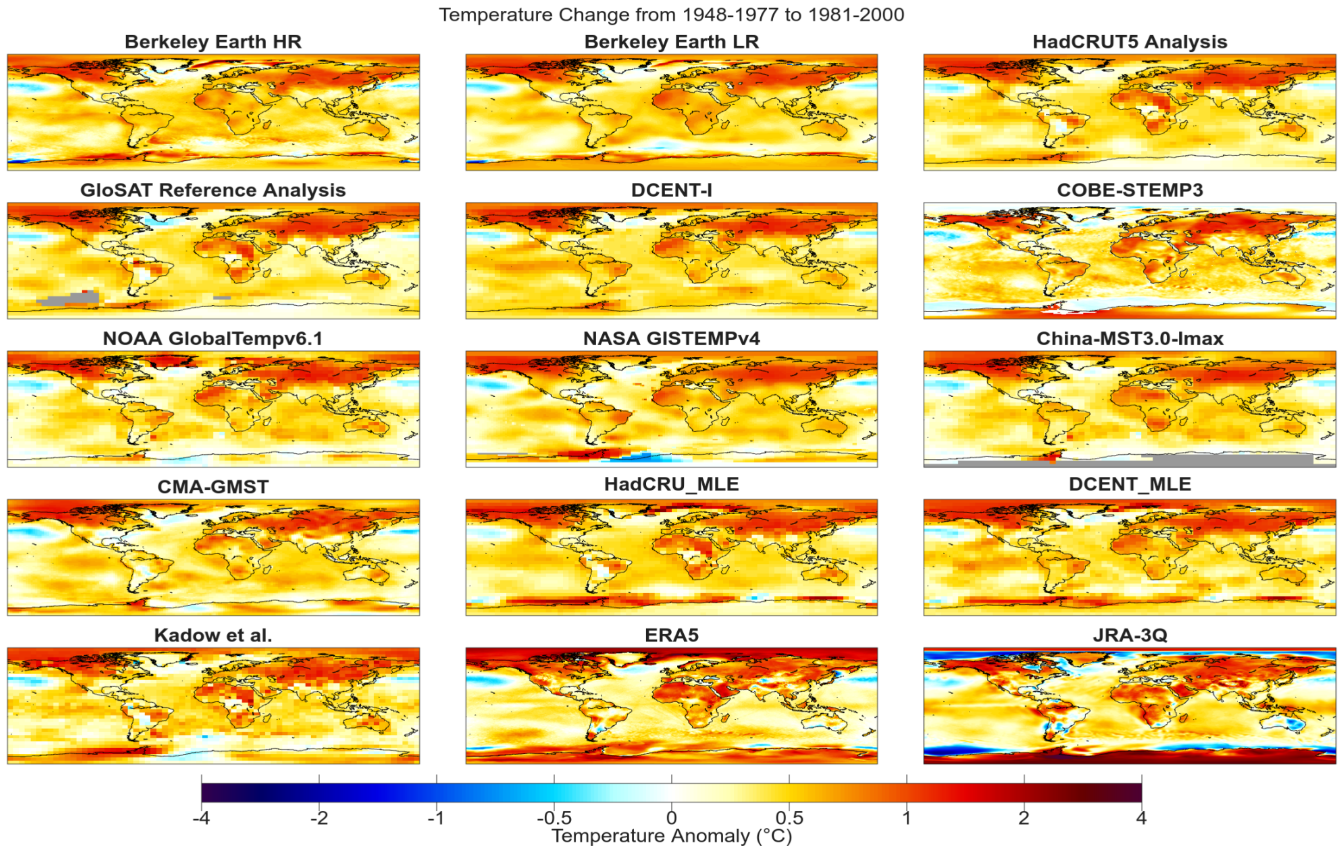

Figure S4: 1948–1977 to 1981–2010 change in annual mean surface temperature (°C) according to 15 datasets. GloSAT, ERA5, and JRA-3Q use surface air temperatures rather than sea surface temperatures for open sea regions.

S.5. 1981–2010 to 2012–2021 temperature change

Figure S5 shows average annual warming from 1981–2010 to 2012–2021. BEST-HR provides a higher resolution but similar spatial pattern than most other datasets. The biggest difference is that, outside of the Barents Sea area, BEST-HR shows less warming in the Arctic Ocean than other datasets, with the exception of COBE-STEMP3, which treats sea ice as open sea.

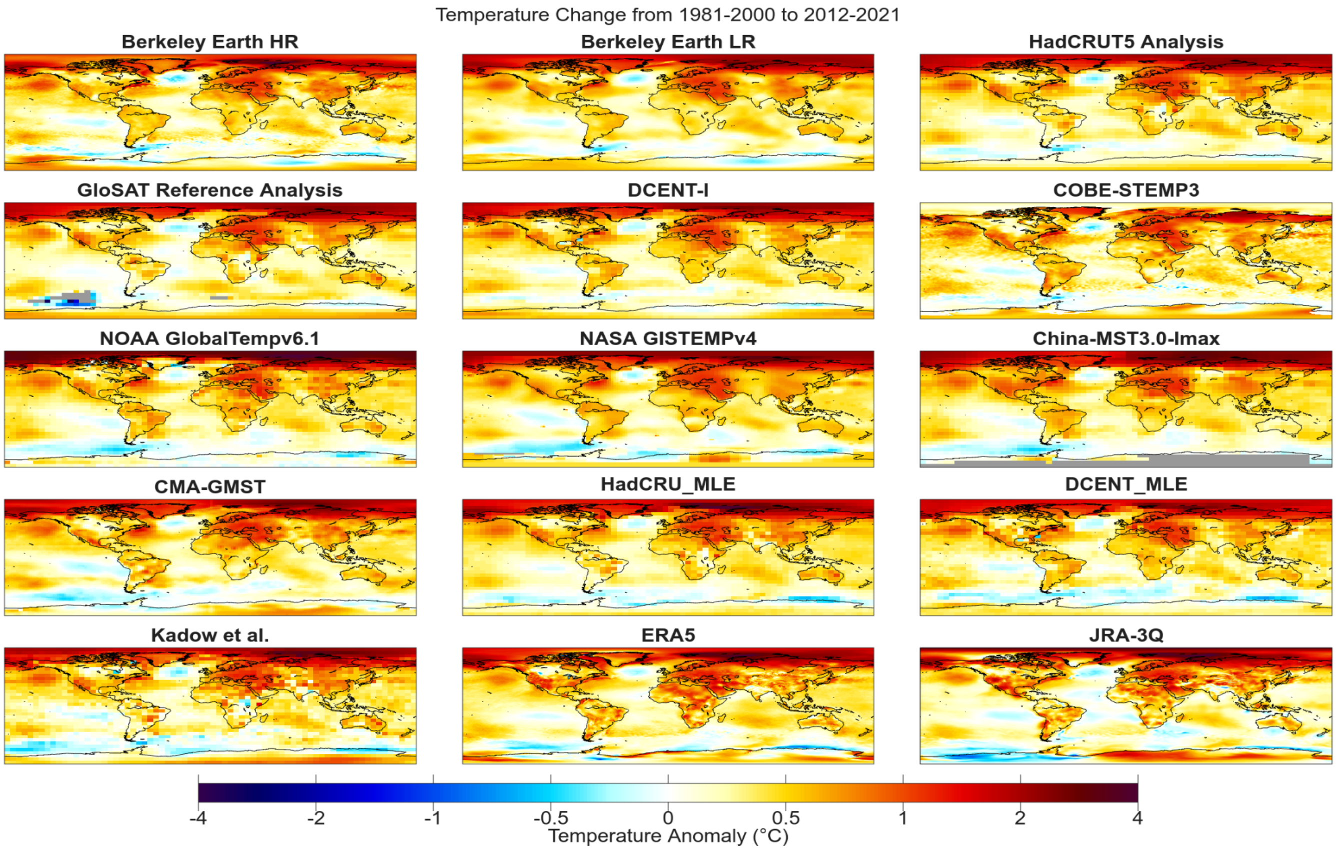

Figure S5: 1981–2010 to 2012–2021 change in annual mean surface temperature (°C) according to 15 datasets. GloSAT, ERA5, and JRA-3Q use surface air temperatures rather than sea surface temperatures for open sea regions.

S.6. 1981–2010 climatologies

Figures S6, S7, S8, and S9 show 1981–2010 climatologies for the 8 available datasets for January and July, respectively. BEST-HR's climatology has a higher resolution but has similar spatial patterns to that of most other datasets. COBE-STEMP3 and DCENT-I use SSTs in sea ice regions, resulting in higher temperature anomalies for the Arctic Ocean.

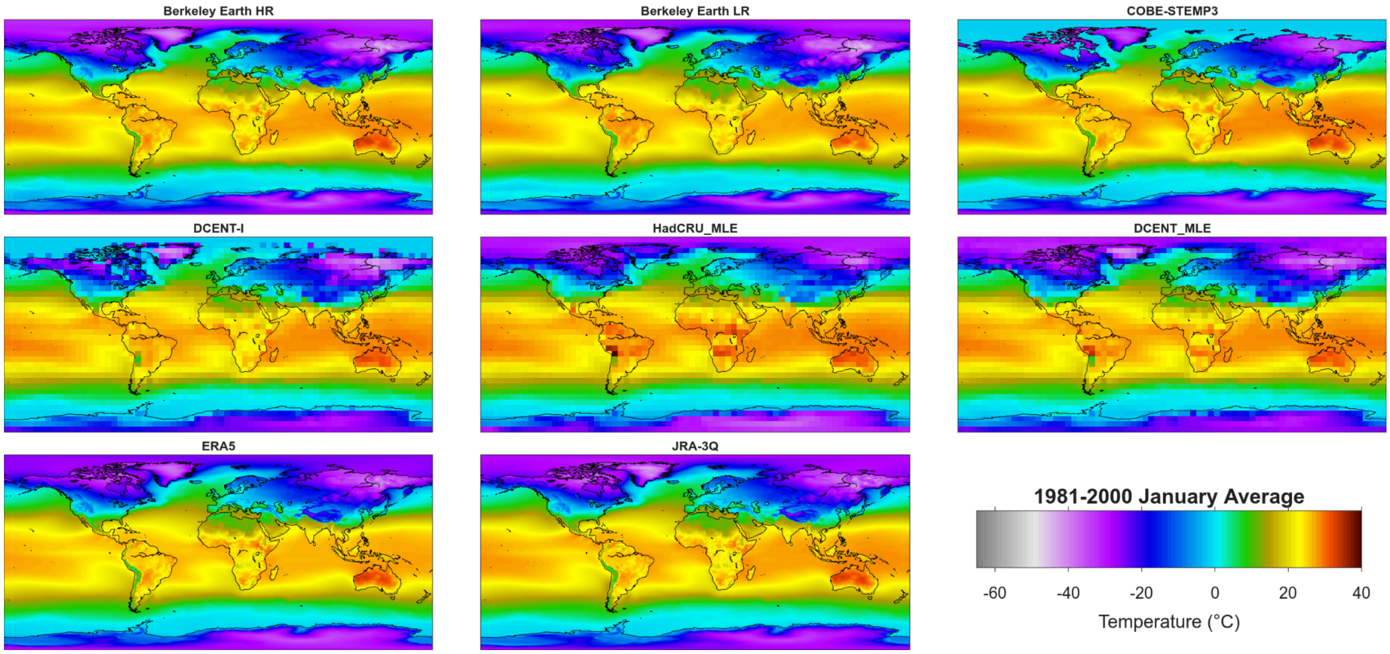

Figure S6: 1981–2010 January climatologies (°C) according to 8 datasets. ERA5 and JRA-3Q use surface air temperatures rather than sea surface temperatures for open sea regions.

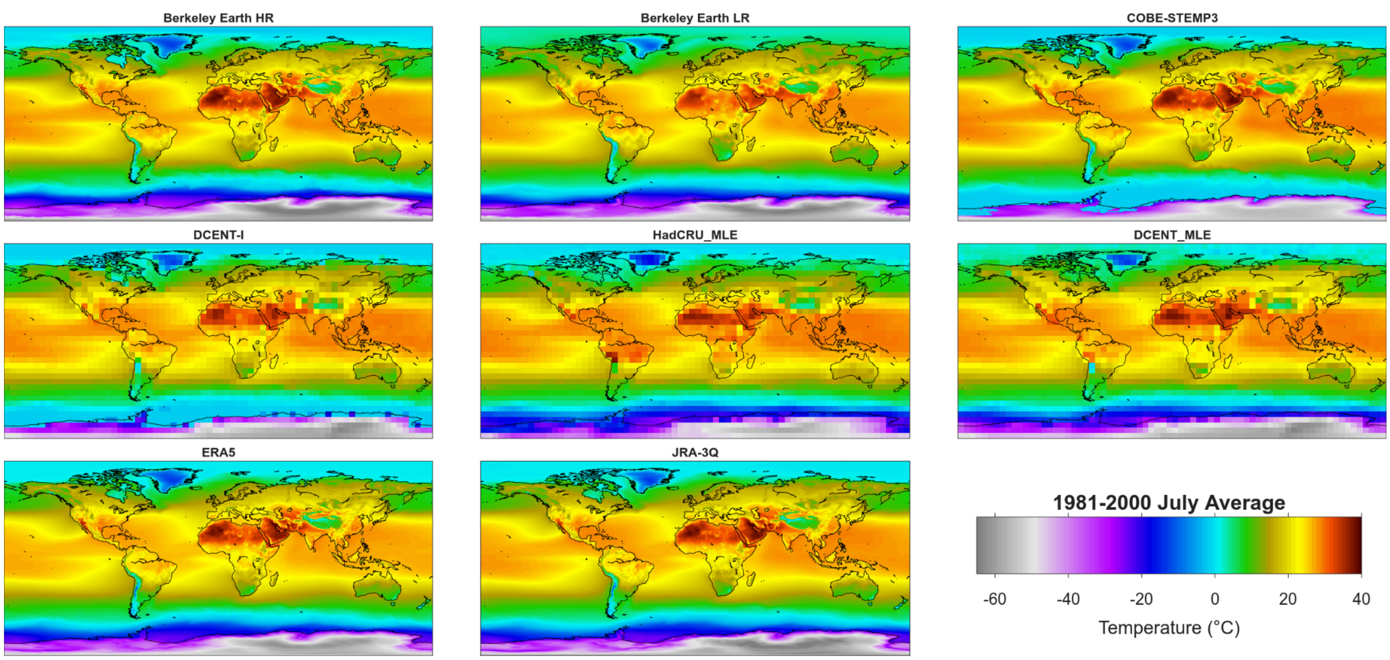

Figure S7: 1981–2010 July climatologies (°C) according to 8 datasets. ERA5 and JRA-3Q use surface air temperatures rather than sea surface temperatures for open sea regions.

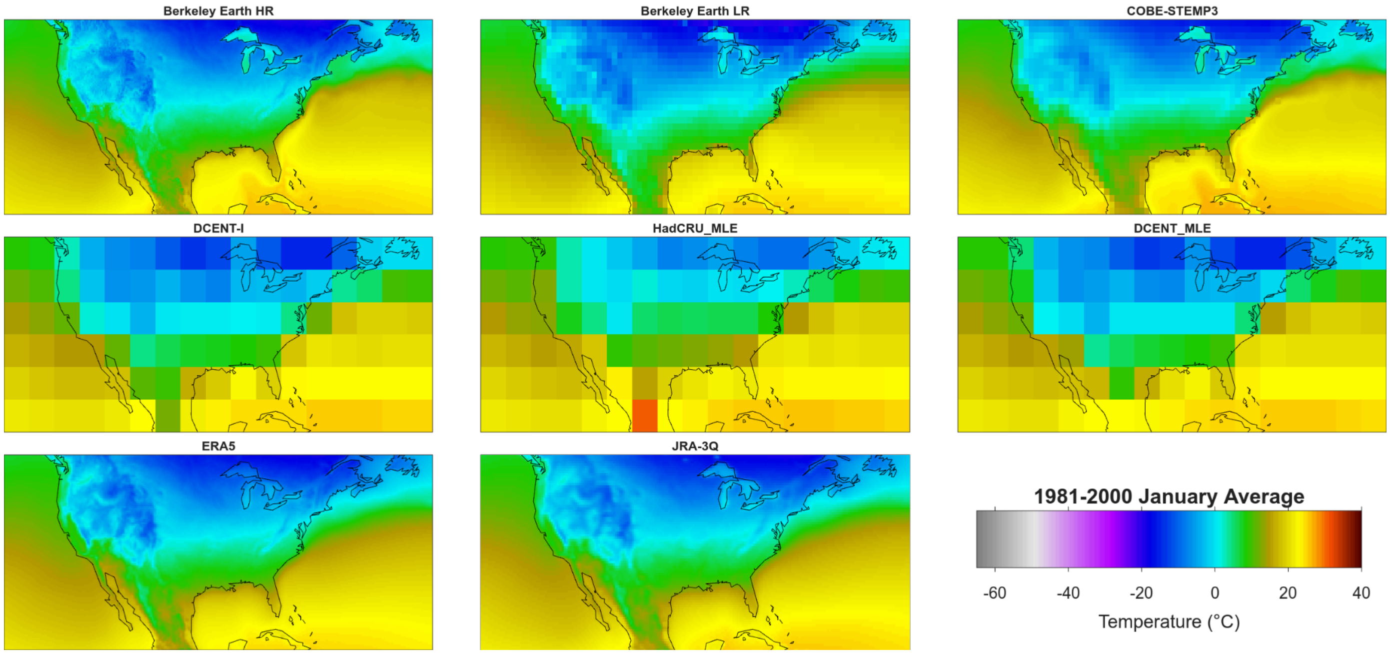

Figure S8: 1981–2010 January climatologies (°C) according to 8 datasets. Same as Fig. S6 but zoomed into the USA to demonstrate the effect of differences in resolution.

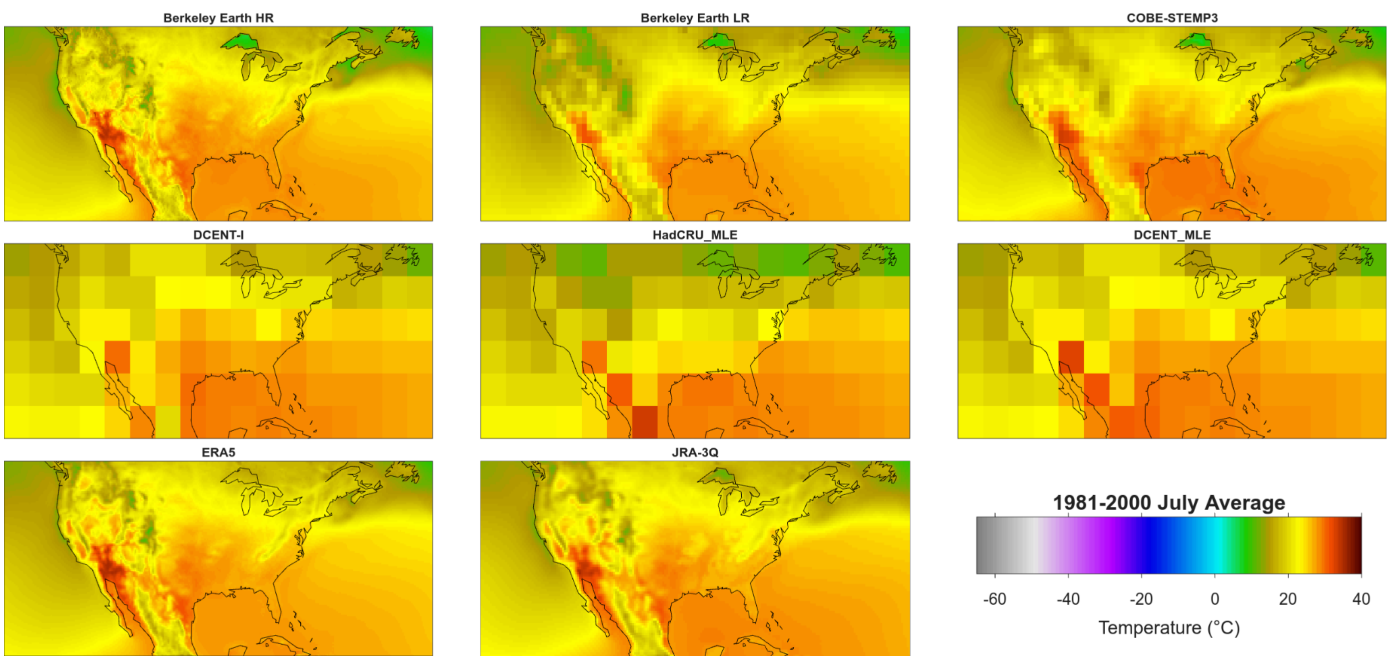

Figure S9: 1981–2010 July climatologies (°C) according to 8 datasets. Same as Fig. S7 but zoomed into the USA to demonstrate the effect of differences in resolution.

S.7. 95 % Uncertainty Ranges in Temperature Anomalies

Figures S10 and S11 show 95 % uncertainty ranges in temperature anomalies relative to 1981–2010 for January 2000 and July 2000, respectively. There is little agreement in the spatial pattern of uncertainty across the 9 datasets, other than generally estimating smaller uncertainty for non-polar sea regions. DCENT_MLE shows unusually high uncertainties in the United States and Europe, which reflects a bug in the DCENT_MLE code, where the calculation of land station climatological uncertainties did not include a division by the number of land stations per grid cell.

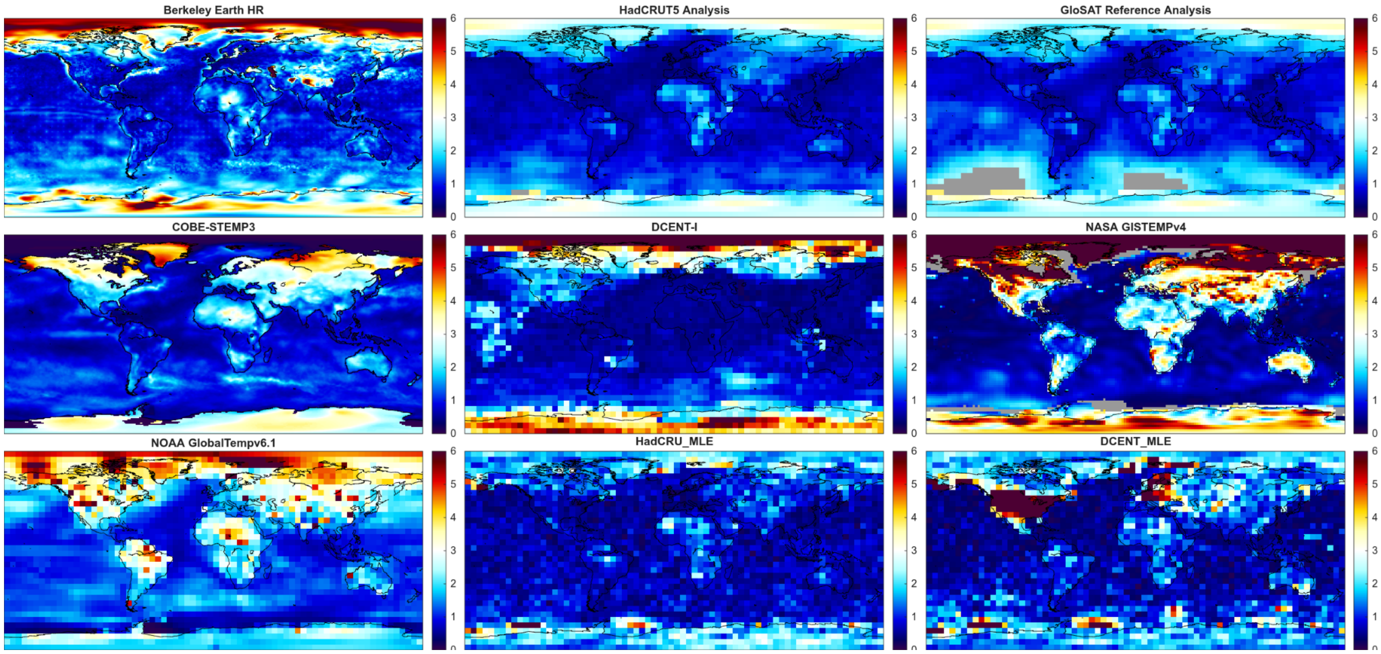

Figure S10: 95 % uncertainty ranges in January 2000 average surface temperature anomalies (°C) according to 9 datasets. Anomalies for all datasets, except for COBE-STEMP3, were calculated relative to 1981–2010 averages. GloSAT, ERA5, and JRA-3Q use surface air temperatures rather than sea surface temperatures for open sea regions.

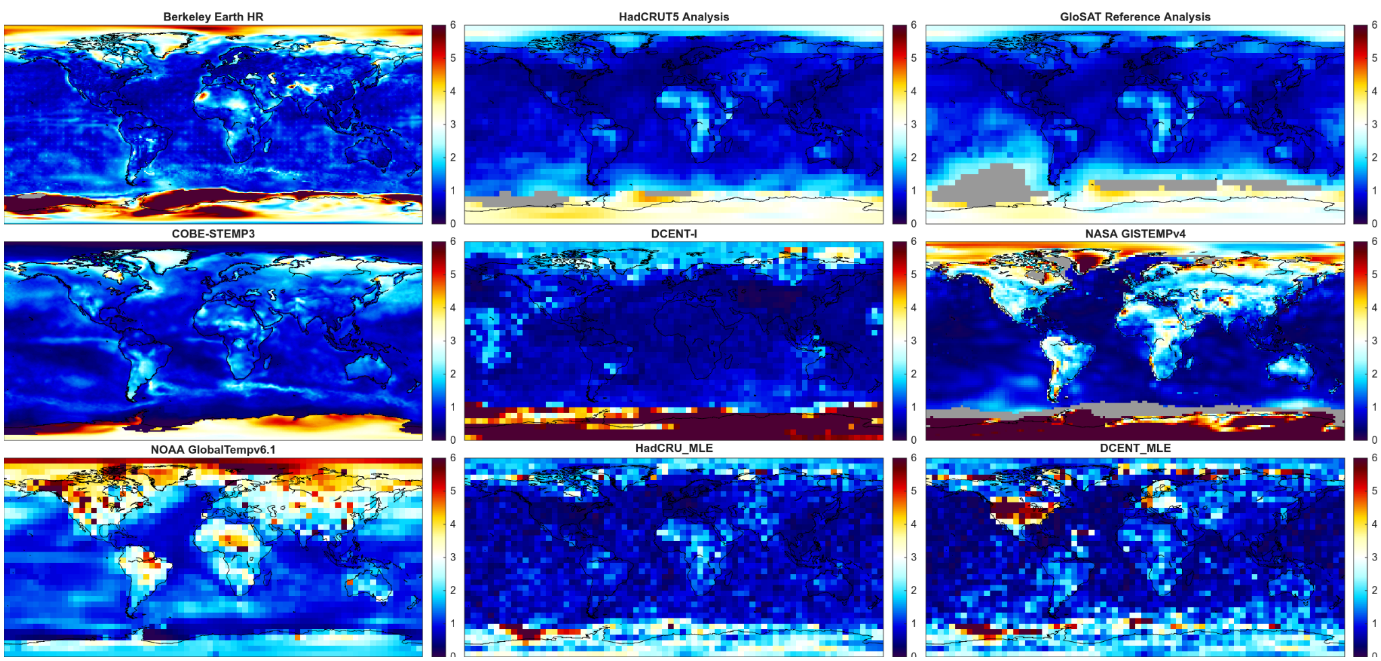

Figure S11: 95 % uncertainty ranges in July 2000 average surface temperature anomalies (°C) according to 9 datasets. Anomalies for all datasets, except for COBE-STEMP3, were calculated relative to 1981–2010 averages. GloSAT, ERA5, and JRA-3Q use surface air temperatures rather than sea surface temperatures for open sea regions.

S.8 Adaptive Gridding

BEST-HR applies an adaptive gridding strategy when computing the Kriging-based weather anomaly field. The sample grid is based on 0.5° spatial resolution with points progressively removed the further one is from any weather station or coastline, so that in remote ocean areas the effective sampling is 8° in latitude or longitude. Between points, spherical natural neighbor interpolation is used to create the final 0.25° weather field. In addition, the predictive fields are also maintained at their full 0.25° resolution throughout. This reduction in map size saves on memory and computation resources. Since ocean areas are replaced with the separately interpolated ocean data, the low density of ocean sampling should have no relevant impact on the computation of global averages. For reasons of practical convenience, the adaptive sampling points on the map are fixed for all time steps and do not vary depending on whether a weather station is active or not. Figure S12 provides a map of the adaptive grid used for TAVG.

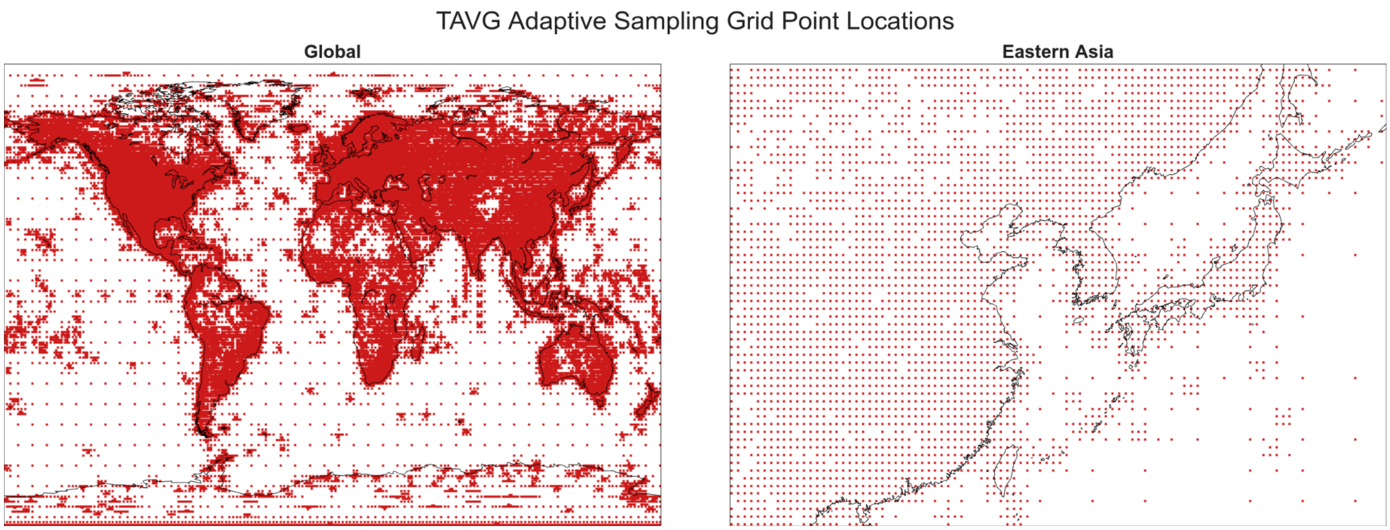

Figure S12: Map of the adaptive grid sampling locations used for the computation of the weather anomaly fields. This is shown for the whole globe (left) and for Eastern Asia (right) as a local example.

S.9. Dataset Usage

The following table shows for TAVG the number of months of data reported by each source as well as the months of data from that source used the final Berkeley Earth Data Set. Except for very short time series, or those with obvious quality control problems, most weather station data is used in its entirety. However, because most weather station time series are reported by multiple sources (on average four copies of each station appears across the input datasets), it is necessary to eliminate duplicates. Which time series are ultimately retained depend on the hierarchy of inputs and the completeness of each source's copy. Further, lower ranked inputs may be used to fill gaps in time series if the primary source chosen is incomplete. As a result, the amount of data originating from each source that is retained in the final repository varies widely.

Source	**Temperature Stations at Source**	**Temperature Stations Used**	**TAVG (Months at Source)**	**TAVG (Months in Merge)**	**Retained Fraction**
Global Historical Climatology Network - Daily	41,834	34,711	20,232,615	12,331,639	61 %
Global Historical Climatology Network - Monthly v4	27,957	24,124	16,523,374	4,064,724	25 %
Global Summary of the Day	27,927	17,900	13,127,598	2,492,901	19 %

US Cooperative Summary of the Day	16,115	13,241	6,341,475	4,031,761	63 %
European Climate Assessment & Dataset	10,519	7,635	7,902,322	1,592,391	20 %
Hadley Centre Data Release (CRUTEM5)	11,850	5,542	8,970,571	1,293,927	14 %
US Cooperative Summary of the Month	12,947	10,805	5,122,590	753,382	15 %
Global Historical Climatology Network - Monthly v3	7,280	5,612	5,413,197	252,095	5 %
US Historical Climatology Network - Monthly	1,218	989	1,545,293	261,814	17 %
International Surface Temperature Initiative (ISTI)	35,909	6,224	16,202,409	175,359	1 %
Global Historical Climate Database (HCLIM)	3,612	688	3,945,492	151,899	4 %
US First Order Summary of the Day	1,600	1,191	577,200	364,113	63 %
GSN Monthly Data	911	289	694,213	66,949	10 %
World Monthly Surface Station Climatology (WMSSC)	5,508	997	2,145,804	61,430	3 %
Colonial Archive	1,018	755	336,016	216,651	64 %
Early US Forts	523	357	109,517	38,453	35 %
World Weather Records (WWR)	1,860	998	206,384	24,666	12 %
Reference Antarctic Data for Environmental Research (MET-READER)	119	107	47,996	11,718	24 %
Programme for Monitoring of the Greenland Ice Sheet (PROMICE)	51	48	8,808	7,648	87 %
GCOS Monthly CLIMAT Summaries	1,213	822	230,054	4,061	2 %

Table S1: List of input datasets used in the Berkeley Earth merged data collection, as well as the original number of stations, the number of stations contributing to the merged dataset, the original number of TAVG months, and the number of TAVG months retained in the merged dataset.

In some cases, similar datasets are considered of equal rank in the sorting hierarchy. In most cases, if they both contain the same weather station then they will report the same data (sometimes differing by rounding precision). However, if two time series of equal rank disagree, the merged dataset averages their values. The number of months in merge in Table S1 counts each source prior to this averaging. In Table 1, the retained totals are fractionally weighted to account for this averaging, and so will be lower than reported here.

S.10. Far Field Weighting

The land component time series apply a far-field weighting adjustment as described in the body of the paper. This is designed to smoothly transition between using estimated monthly average temperature anomalies and using the 13-month moving average of these anomalies when estimating weather anomalies far from all available weather stations. This is only applied to the residual weather anomalies and no smoothing is applied to the predictive field component.

The far field weighting is defined as a smooth function $f(a)$ of the station coverage metric $\mathrm{CC}(\boldsymbol{x}, t_j)$.

$$\mathrm{FF}(\boldsymbol{x}, t_j) = f\big(\mathrm{CC}(\boldsymbol{x}, t_j)\big) \tag{S1}$$

Where

$$\mathrm{CC}\big(\boldsymbol{x}, t_j\big) = \sqrt{1 - \frac{\mathrm{c}_0 - \mathrm{c}_x^T \mathrm{C}^{-1}\, \mathrm{c}_x + (\mathrm{q}_i - \mathrm{Q}_i^T \mathrm{C}^{-1} \mathrm{c}_x)^T (\mathrm{Q}_i^T \mathrm{C}^{-1} \mathrm{Q}_i)^{-1} (\mathrm{q}_i - \mathrm{Q}_i^T \mathrm{C}^{-1} \mathrm{c}_x)}{\mathrm{c}_0 + (\mathrm{q}_i)^T (\mathrm{Q}_i^T \mathrm{C}^{-1} \mathrm{Q}_i)^{-1} (\mathrm{q}_i)}} \tag{S2}$$

And q_i is the vector of Kriging fit parameters, $c_x = C(\boldsymbol{x}, \boldsymbol{x}_i)$ is the vector of expected covariance as location $\boldsymbol{x}$ as a result of the available weather stations $\boldsymbol{x}_i$ at time t_j and C has the same meaning as Equation 10.

The denominator of the fractional expression is the expected variance at $\boldsymbol{x}$ and the numerator is the square of the estimated Kriging error. The resulting ratio is naturally expected to have a range of 0 to 1 with zero indicating a good fit and 1 indicating no constraint at all. The minus 1 reverses this ordering, and the square root converts a variance to an effective error term. In regions near many weather stations $\mathrm{CC}\big(\boldsymbol{x}, t_j\big)$ goes to 1, and it goes to 0 far from any stations. This same coverage metric is also used as a heuristic when deciding how to mask the final fields during export.

The smooth transition function is then given by

$$f(a) = \begin{cases} (a_0 - a)^2 (p_1 a + p_2) & a < a_0 \\ 0 & \text{otherwise} \end{cases} \tag{S3}$$

Which provides smooth derivatives at key transition points. Parameters were estimated empirically based on the difference in correlation length between monthly and annual temperature anomalies with fit values of a_0 = 0.82, $p_1 = 0.0393$, $p_2 = 0.9510$, being adopted. However, the results are largely insensitive to specific parameter choices as long as the transition point $f(a)$ = 0.5 is well within the regime of low CC (i.e. low constraint from measurements).

S.11. Warming by Latitude and Time Interval

The following figure shows reconstructed warming by latitude during the mid to late 20th century as a contrast with the change since the 19th century. The Arctic changes in the Berkeley Earth high-resolution reconstruction are more similar to other groups when only the 20th century is considered.

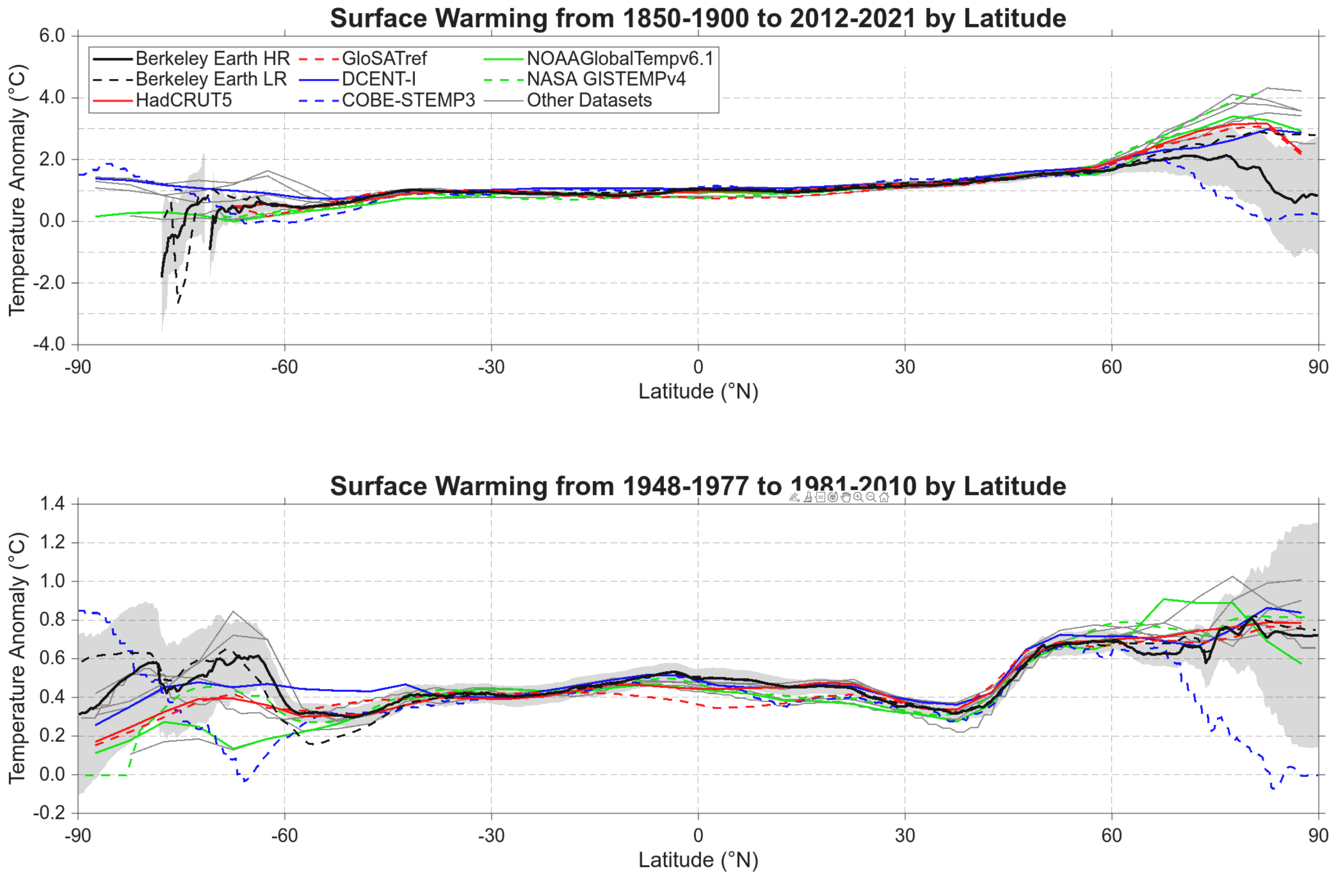

Figure S13: Comparison of warming by latitude for the various time series. The top panel is the same as Fig. 7, and shows the average of 2012–2021 relative to the average 1850–1900. The lower panel shows just the recent change using the time period 1981–2010 relative to 1948–1977.