## S1　Data coverage without statistical infilling

Data coverage of the non-infilled grids of the GloSATref data set is shown in Fig. S1. For comparison, data coverage for the non-infilled grids of the HadCRUT5 data set (Morice et al., 2021) is shown in Fig. S2. These figures provide a comparison of data coverage for the underpinning observational data that contributes to each data set. Differences in data coverage between
5　these two data sets result from (1) the use of marine air temperature observations in GloSATref and sea-surface temperature in HadCRUT5, (2) land data acquisitions in GloSATref that extend the CRUTEM5 station database and (3) the capability to grid additional land station series in GloSATref through the use of new station climatology estimates. Further discussion of data coverage is provided in the main article.

## S2　Data coverage of the GloSATref analysis

10　Data coverage of the GloSATref statistically infilled analysis fields is shown in Fig. S3. For comparison, data coverage for the HadCRUT5 analysis (Morice et al., 2021) is shown in Fig. S4. The two data sets share common statistical analysis methods but differ in observation data used.

　　Differences in data coverage between the GloSATref and HadCRUT5 analyses result from (1) difference in underpinning observation data coverage described in S1 and (2) differences in fitted statistical model hyperparameters. The marine air
15　temperature analysis parameters of GloSATref result in interpolation over shorter spatial length scales (marine length scale parameter $\rho$ = 1300 km) than the sea-surface temperature analysis of HadCRUT5 (marine length scale parameter $\rho$ = 1550 km). Land surface air temperature (LSAT) analyses for the the two data sets share the same parameter estimate values (LSAT length scale parameter $\rho$ = 1300 km), hence interpolation distances from available observations are broadly equivalent across land and sea ice regions in the two data sets.
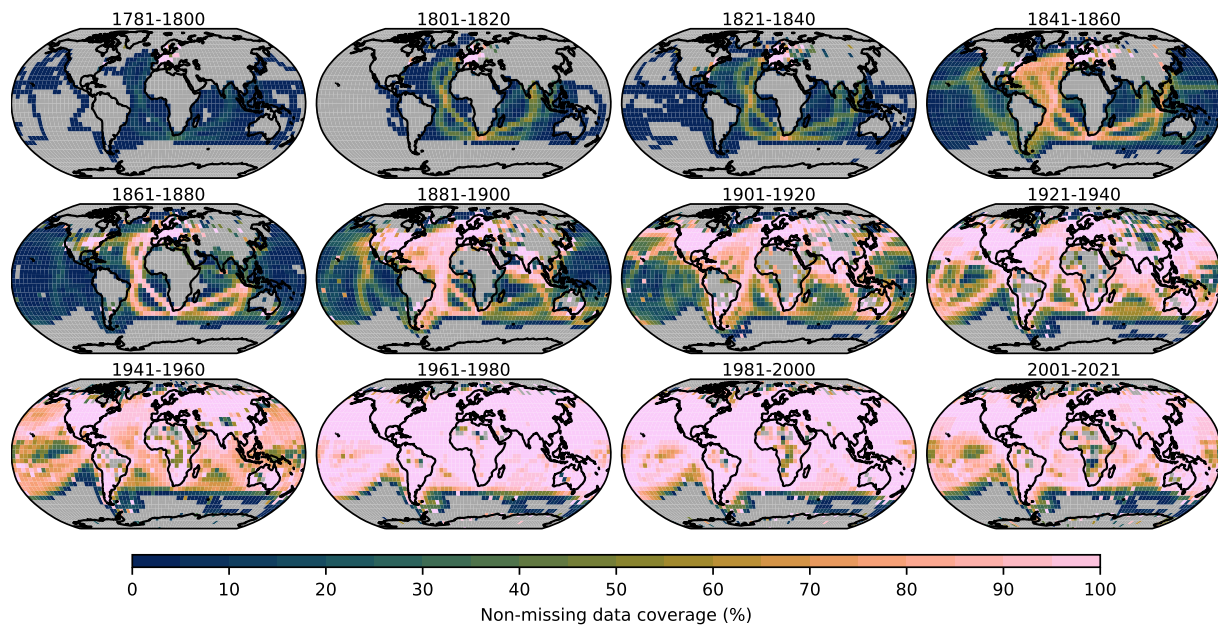
**Figure S1.** Data coverage for the unfilled GloSATref data set, showing the percentage of months with non-missing data for 20-year periods (final panel shows 21-year period). Grey shaded regions have no data in the corresponding period.
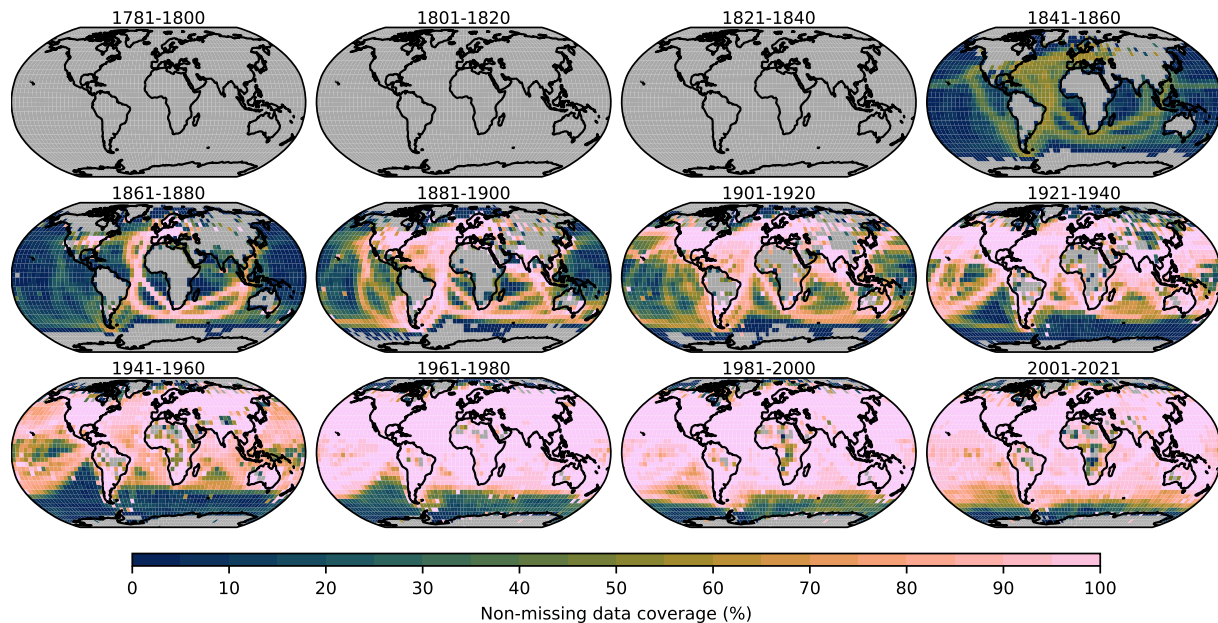


**Figure S2.** Data coverage in the unfilled HadCRUT5 data set for comparison to Fig. S1, showing the percentage of months with non-missing data for 20-year periods (final panel shows 21-year period). Note that the HadCRUT5 grids contain data from 1850 only.
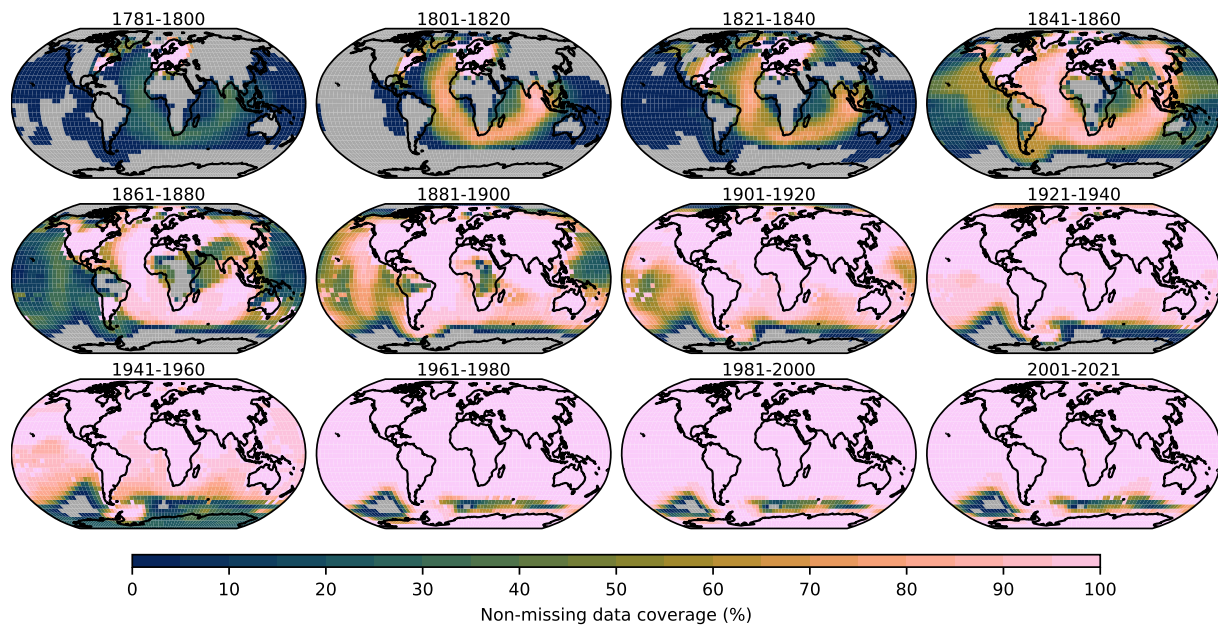
**Figure S3.** Data coverage for the GloSATref infilled analysis, showing the percentage of months with non-missing data for 20-year periods (final panel shows 21-year period). Grey shaded regions have no data in the corresponding period.
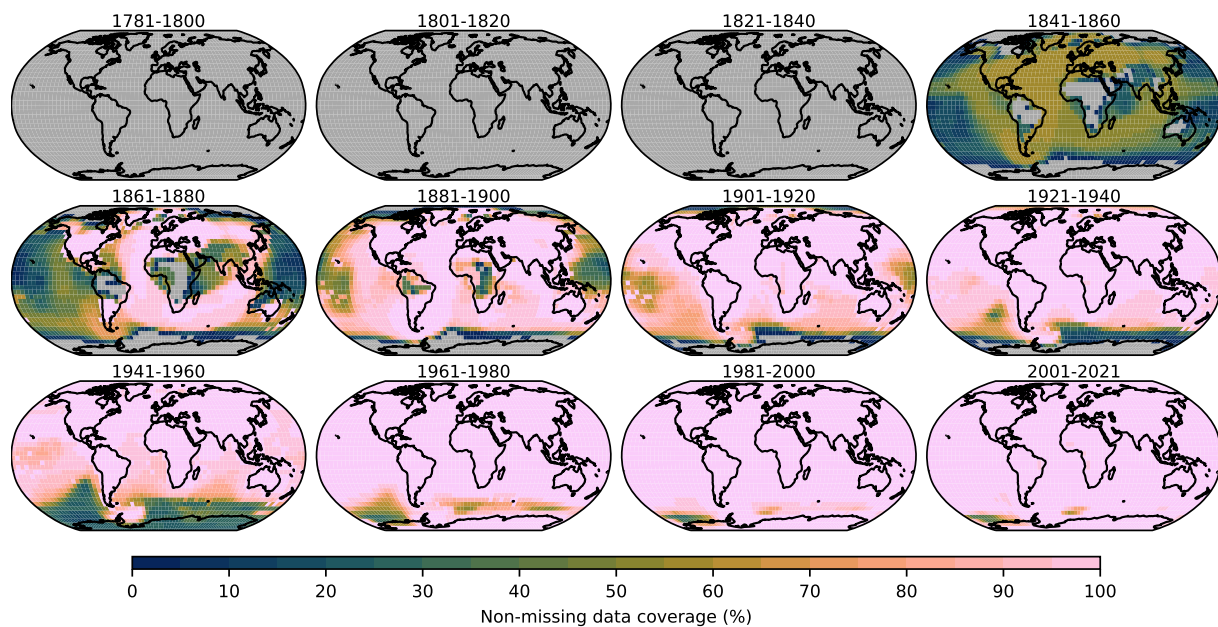


**Figure S4.** Data coverage in the HadCRUT5 infilled analysis for comparison to Fig. S1, showing the percentage of months with non-missing data for 20-year periods (final panel shows 21-year period). Note that the HadCRUT5 grids contain data from 1850 only.

**S3   Comparisons to paleoclimate reconstructions**

Global annual average temperature anomaly series are shown in Fig. S5 in comparison to the PAGES2k multi-method ensemble paleoclimate reconstructions (PAGES2k, 2019). The GloSATref ensemble mean shows greater variability than the PAGES2k median in the early instrumental record. Central estimates exceed the 2.5% and 97.5% confidence limits of each ensemble. Confidence intervals for the GloSATref and PAGES2k ensembles overlap for most years, while departures from confidence
25 intervals occur around the times of the unidentified 1809 volcanic eruption and the 1815 eruption of Mount Tambora. The individual ensemble medians for each paleoclimate reconstruction method have a differing response to these early 19th century volcanoes, with weaker responses than is apparent in the GloSATref analysis global average time series. The GloSATref analysis does, however, have reduced global coverage in this period (see Fig S3), with greater variability resulting from regional sampling.
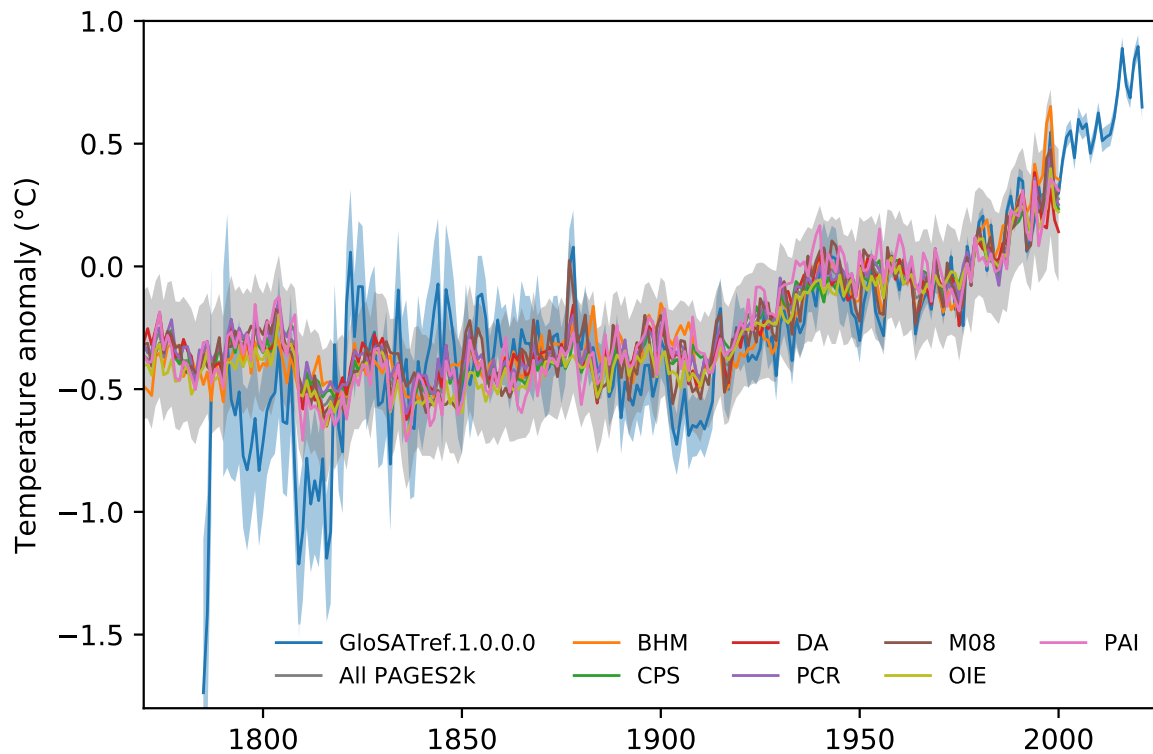


**Figure S5.** Annual global average temperature anomaly time series for the GloSATref analysis and PAGES2k paleoclimate proxy reconstructions (°C relative to 1961-1990). Uncertainties shown as 95% confidence intervals. Ensemble means for each PAGES2k reconstruction method shown: BHM, CPS, DA, PCR, M08, OIE and PAI.

## References

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., and Simpson, I. R.: An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set, Journal of Geophysical Research: Atmospheres, 126, https://doi.org/10.1029/2019jd032361, 2021.

PAGES2k: Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era, Nature Geoscience, 12, 643–649, https://doi.org/10.1038/s41561-019-0400-0, 2019.