Journal of Geophysical Research: Atmospheres

Supporting Information for

An updated assessment of near-surface temperature change from 1850: the HadCRUT5 dataset

C. P. Morice[1], J. J. Kennedy[1], N. A. Rayner[1], J. P. Winn[1], E. Hogan[1], R. E. Killick[1], R. J. H. Dunn[1], T. J. Osborn[2], P. D. Jones[2] and I. R. Simpson[1]

1. Met Office, Hadley Centre, Exeter, EX1 3PB, UK

2 Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

Corresponding author: Colin Morice (colin.morice@metoffice.gov.uk)

Contents of this file

Figure S13 Example monthly temperature anomaly fields for 2000.

Table S1. HadCRUT5 uncertainty model components for the non-filled dataset and the HadCRUT5 analysis.

Introduction

In this document, we provide additional methodological details and results in support of information provided in the main article and as a guide to users of the HadCRUT5 dataset. We include diagnostics supporting the values of analysis hyperparameters used, results of a sensitivity analysis for choices in spatial analysis masking, additional comparisons to the results of other analyses and example comparisons of monthly fields for the HadCRUT5 non-infilled dataset and HadCRUT5 analysis.

Text S1 presents an overview of the error model structure for the HadCRUT5 non-infilled data set and for the HadCRUT5 analysis, with a summary of error model components provided in Table S1. Text S2 presents methodological descriptions for merging of land and ocean data while methods for time series calculation are described in Text S3 and S4, based on methods published in Kennedy et al. (2011), Morice et al. (2012) and Kennedy et al. (2019). Where there are modifications to the previously published methods those modifications are stated in these sections. Text S5 provides additional information on the estimation of covariance function hyperparameters for the HadCRUT5 analysis. Text S6 describes diagnostics in support of the analysis masking criteria.

Figures S1 and S2 show the results of monthly hyperparameter optimization for land air temperature and sea-surface temperature anomaly fields, along with the fixed hyperparameter values used in the HadCRUT5 analysis, derived as the average of monthly hyperparameters in the 1961-1990 climatology period.

Figures S3 to S6 demonstrate the sensitivity analysis of our key results to choices in our analysis masking criteria. Figures S7 and S8 demonstrate the ensemble spread for the HadCRUT5 analysis without and with masking, showing the limitations of our prior model for anomaly variability and the sensitivity of the analysis to our simple prior model in regions with weak observational constraint. These figures are provided in support of the choice to mask regions of weak observational constraint from the HadCRUT5 analysis.

We include a comparison of global average temperature anomaly series and uncertainties for the HadCRUT5 analysis with those of other studies in Figure S9. Examples of monthly temperature anomaly fields for the non-infilled HadCRUT5 dataset, the HadCRUT5 analysis and fields of differences between the two are show in Figures S10, S11, S12 and S13. These are shown to demonstrate the differences in spatial coverage for monthly fields and the effects of the analysis methods for grid cells that are populated in the non-infilled dataset.

Text S1 Error model structure for HadCRUT5

This section outlines the terms of the uncertainty model for the HadCRUT5 non-infilled data set and the HadCRUT5 analysis grids and time series.

The error model is split into components according to the way that uncertainty information is presented in the HadCRUT5 dataset. The sources of uncertainty that are modelled in HadCRUT5 are grouped according to their error correlation structure to allow uncertainties to be propagated appropriately into derived diagnostics such as regional average time series. A summary of the sources of uncertainty within each component of the error models of both the HadCRUT5 non-infilled dataset and the HadCRUT5 analysis are presented in Table S1.

Error model for the HadCRUT5 non-infilled dataset

The error model for the non-infilled dataset describes the estimate of temperature anomaly $\hat{T}(s,t)$, at spatial location s and time t, as a sum of the true temperature anomaly $T(s,t)$ and three error terms: a bias term $\varepsilon_b(s,t)$ representing biases with large-scale spatial and temporal structure; a partially correlated error term $\varepsilon_p(s,t)$ for errors with typically local structure; and an uncorrelated error term $\varepsilon_u(s,t)$ describing errors that are independent between spatial and temporal locations. The full error model for non-infilled fields is given by:

$$\hat{T}(s,t) = T(s,t) + \varepsilon_b(s,t) + \varepsilon_p(s,t) + \varepsilon_u(s,t) \tag{1}$$

The bias term $\varepsilon_b(s,t)$ models systematic biases from land station homogenization error, urbanization and non- standard measurement enclosures (Morice et al., 2012) and adjustments applied to correct for changes in marine measurement methods (Kennedy et al., 2019). The partially correlated error term $\varepsilon_p(s,t)$ models the effects of biases in observations from individual marine observing platforms. The uncorrelated error term $\varepsilon_u(s,t)$ models the effects of random measurement errors and sampling errors from estimating grid cell average temperature anomalies from a finite number of observations.

The HadCRUT5 non-infilled ensemble samples the uncertainties for the combination $T(s,t) + \varepsilon_b(s,t)$. The uncertainties for partially correlated and uncorrelated errors are not encoded into the ensemble. Uncertainty information for partially correlated errors $\varepsilon_p(s,t)$ are provided as spatial error covariance matrices while uncertainties for uncorrelated errors $\varepsilon_u(s,t)$ are provided for each observed grid cell in the non-infilled dataset.

The error model for estimates of spatial average time series $\hat{T}(t)$, derived from the gridded data, is given as a sum of the true temperature anomaly time series $T(t)$ and four error terms:

$$\hat{T}(t) = T(t) + \varepsilon_b(t) + \varepsilon_p(t) + \varepsilon_u(t) + \varepsilon_c(t) \tag{2}$$

Here $\varepsilon_b(t)$ is the effect of the bias term propagated into the spatial average, $\varepsilon_p(t)$ is the effect of the partially correlated term, $\varepsilon_u(t)$ the effect of the uncorrelated error term, and the fourth error term $\varepsilon_c(t)$ is the error in estimating the spatial average from incomplete spatial coverage, with missing grid cells resulting from limited spatial sampling of the observation network. Methods for propagation of uncertainty associated with each of these terms are given in Text S3.

Error model for the HadCRUT5 analysis

The error model for the HadCRUT5 analysis describes the analysis estimate $\hat{T}(s,t)$ of the temperature anomaly, at spatial location s and time t, as the sum of the true temperature $T(s,t)$ and the analysis error $\varepsilon_a(s,t)$:

$$\hat{T}(s,t) = T(s,t) + \varepsilon_a(s,t) \tag{3}$$

The analysis error term combines all errors that are modelled in the Gaussian process analysis, both spatial reconstruction errors and observational errors. This includes the propagation of the partially correlated and the uncorrelated observational error terms through the analysis equations and the effects of the observational bias term, through application of the analysis method to each HadCRUT5 non-infilled ensemble member, as described in Appendix A. The analysis ensemble samples the analysis uncertainty such that each ensemble member is a sample of $T(s,t) + \varepsilon_a(s,t)$.

Time series of global and regional average temperatures include an additional coverage error term, $\varepsilon_c(t)$. Like the coverage error term for the non-infilled dataset, this coverage error term arises from estimation of averages from spatially incomplete grids of temperature anomaly data.

$$\hat{T}(t) = T(t) + \varepsilon_a(t) + \varepsilon_c(t) \tag{4}$$

The ensemble time series for the HadCRUT5 analysis sample the uncertainty associated with $T(t) + \varepsilon_a(t)$. Uncertainty associated with coverage error $\varepsilon_c(t)$ is derived as an additional uncertainty term that accompanies the ensemble time series. The calculation of analysis time series, uncertainties and the derivation of summary statistics is described in Text S4.

Text S2 Merging land and ocean data products

The following subsections describe the calculation of merged temperature anomaly fields from the land air temperature and sea surface temperature ensemble data sets and, for the non-infilled dataset, the corresponding propagation of uncertainty for the uncorrelated and partially correlated components. The computation of terms of the uncertainty model for the land air temperature data are described in Morice et al. (2012) and those of the sea surface temperature data are described in Kennedy et al (2019). The merging method follows that of Morice et al. (2012). Methods for deriving weights for land and sea data are described in Section 3 of the main article, including modifications to the weighting method for the HadCRUT5 analysis to weight towards use of the land air temperature analysis in sea-ice regions.

Ensemble temperature anomaly fields

The merged global fields are computed as a weighted average of land air temperature and sea-surface temperature anomalies. For data at spatial locations s and time t, we define the land air temperature anomaly as $T_L(s,t)$ and the marine temperature anomaly as $T_M(s,t)$. We define the weighting given to land data as $f(s,t)$ and the weighting given to the marine data as $1-f(s,t)$. The temperature anomaly for the merged field is then calculated as:

$$T(s,t) = f(s,t)T_L(s,t) + \big(1 - f(s,t)\big)T_M(s,t) \qquad (5)$$

The values of the weights are dependent on the fraction of land in a grid cell and data availability. Where there is no land data available $f(s,t) = 0.0$ and where there is no marine data $f(s,t) = 1.0$. For the HadCRUT5 analysis the weighting is also dependent on sea ice coverage, with sea ice regions treated as land. This results in the air temperature analysis being prioritized in sea ice regions, rather than the sea-surface temperature analysis. As in Morice et al. (2012), grid cells that contain reporting land stations receive a minimum weighting of $f(s,t) = 0.25$ to ensure that island and coastal station data receive a non-negligible weighting. Further details are provided in Section 3 of the main article.

When the weighting is applied to ensemble members, either those of the HadCRUT5 non-infilled dataset or the HadCRUT5 analysis, we define $T_{Ld}(s,t)$ as the dth ensemble member in the land air temperature ensemble and $T_{Md}(s,t)$ as the dth member of the sea-surface temperature ensemble. The merged temperature anomaly fields for ensemble member d, denoted as $T_d(s,t)$, are then computed through the following weighted average:

$$T_d(s,t) = f(s,t)T_{Ld}(s,t) + \big(1 - f(s,t)\big)T_{Md}(s,t) \qquad (6)$$

For the non-infilled dataset, $T_d(s,t)$ is a sample of the uncertainty associated with $T(s,t) + \varepsilon_b(s,t)$, where $T(s,t)$ is the true temperature anomaly and $\varepsilon_b(s,t)$ is the error associated with systematic biases. It samples the uncertainty in temperature anomaly fields associated with systematic measurement biases.

The same merging method is used to merge land and ocean ensemble analysis fields. In the case of the analysis fields, $T_d(s,t)$ samples the uncertainty in $T(s,t) + \varepsilon_a(s,t)$, where $\varepsilon_a(s,t)$

is the error associated with the spatial analysis. Hence, for the HadCRUT5 analysis, the merged ensemble encapsulates the full uncertainty model for the analysis fields.

Uncorrelated component

Both the land air temperature uncertainty model defined in Morice et al. (2012) and the sea-surface temperature uncertainty model in Kennedy et al. (2019) include uncertainty fields that describe measurement errors and grid cell sampling errors that are fully uncorrelated between grid cells. The standard uncertainty, $\sigma_u(s,t)$, in the merged analysis that arises from these uncorrelated errors, $\varepsilon_u(s,t)$, is computed by propagating the standard uncertainty for land air temperature, $\sigma_L(s,t)$, and that for sea-surface temperature, $\sigma_M(s,t)$, as follows:

$$\sigma_u(s,t) = \sqrt{f(s,t)^2\sigma_L(s,t)^2 + \left(1 - f(s,t)\right)^2\sigma_M(s,t)^2} \tag{7}$$

The merged uncorrelated uncertainty fields are only relevant to the non-infilled dataset (whereas the effects of these error sources are represented within the analysis ensemble for the HadCRUT5 analysis).

Partially correlated component

The partially correlated component describes the errors $\varepsilon_p(s,t)$ that exhibit correlations between spatial locations. This information is provided through provision of monthly error covariance matrices for the non-infilled dataset. There is no equivalent uncertainty component for the HadCRUT5 analysis because the effects of these partially correlated errors are represented in the analysis ensemble (see Text S1).

The partially correlated error term arises from marine observing platform 'micro biases' that impart correlated error structures between spatial locations as observing platforms move. As there is no error covariance term for the land dataset, the error covariance $\boldsymbol{C}$ for the non-infilled merged dataset is defined only as a function of the weights and the HadSST4 marine error covariance matrices $\boldsymbol{C}_M$.

$$\boldsymbol{C} = (\boldsymbol{I} - \boldsymbol{F})\boldsymbol{C}_M(\boldsymbol{I} - \boldsymbol{F}) \tag{8}$$

where $\boldsymbol{I}$ is an identity matrix and $\boldsymbol{F}$ is a matrix with the land weights on the leading diagonal and zeros elsewhere:

$$\boldsymbol{F} = \begin{bmatrix} f(s_1,t) & & 0 \\ & \ddots & \\ 0 & & f(s_N,t) \end{bmatrix} \tag{9}$$

Text S3 Time series calculation – HadCRUT5 non-infilled dataset

Time series calculation for the non-infilled HadCRUT5 dataset follows the method described in Morice et al. (2012) with only minor modifications that are outlined in this section. The following subsections describe the application of spatial and temporal averaging to each component of the HadCRUT5 uncertainty model. All reported time average diagnostics are computed by first computing any spatial averaging required to produce monthly time series and then computing temporal averages. This order of operation is adopted to equally weight the contribution of each month in time averaged diagnostics.

Global and regional time series

Here we present the methodology for computing regional and temporal average temperature anomaly series from the non-infilled HadCRUT5 dataset. For each component of the HadCRUT5 uncertainty model, the following text describes uncertainty propagation under the described spatial and temporal averaging operations.

Spatial average time series

A spatial average temperature anomaly at time t is computed as the grid cell area weighted average of $i = 1, \dots, N$ non-missing grid cells at spatial locations s_i. Denoting the grid cell temperature anomalies as $T(s_i, t)$ and the normalized grid cell weights as $w(s_i, t)$, the regional average anomaly is defined as:

$$T(t) = \sum_{i=1}^{N} w(s_i, t) T(s_i, t) \quad (10)$$

Each grid cell weight $w(s_i, t)$ is computed as the area of the respective grid cell normalized by the sum of the grid cell areas for non-missing grid cells in the region.

For global averages we adopt the equal hemispheric weighting method of Morice et al. (2012), with hemispheric weights $\mathrm{b}_N = b_S = 0.5$ applied to Northern Hemisphere and Southern Hemisphere averages, $T_N(t)$ and $T_S(t)$.

$$T(t) = b_N T_N(t) + b_S T_S(t) \quad (11)$$

Annual average time series

Annual average time series are computed by first applying spatial averaging to obtain monthly time series and then averaging the monthly series to obtain annual series. After the application of spatial averaging, we denote the values of the monthly temperature anomaly time series as $T(t_{jm})$, where subscripts index the year j and month m. The annual average $T(t_j)$ for year j is computed from the $m = 1, \dots, M_j$ monthly time series values for year j, noting that $M_j = 12$ for a complete year of data:

$$T(t_j) = \frac{1}{M_j} \sum_{m=1}^{M_j} T(t_{jm}) \tag{12}$$

Uncorrelated component

Uncertainty in spatial average time series

The uncorrelated component of the non-infilled dataset describes measurement error and grid cell sampling error that are fully uncorrelated between grid cells and between months. When propagated into an area average the resulting uncertainty in that spatial average is given by:

$$\sigma_u(t) = \sqrt{\sum_{i=1}^{N} w(s_i, t)^2 \sigma_u(s_i, t)^2} \tag{13}$$

where $\sigma_u(s_i, t)$ is the 1-sigma measurement and sampling uncertainty for grid cell i and $\sigma_u(t)$ is the uncertainty propagated into the spatial average.

When computing global average time series as an average of hemispheric series, denoting the value of the northern and southern hemispheric series for the uncorrelated component as $\sigma_{Nu}(t)$ and $\sigma_{Su}(t)$, the resulting uncertainty in the global average series, $\sigma_u(t)$, is calculated as follows:

$$\sigma_u(t) = \sqrt{b_{\mathrm{N}}^2 \sigma_{Nu}(t)^2 + b_S^2 \sigma_{Su}(t)^2} \tag{14}$$

for hemispheric weights $b_N = b_S = 0.5$.

Uncertainty in annual average time series

The contribution to total uncertainty in annual average time series from the uncorrelated uncertainty term is derived from monthly series of the uncorrelated component. The resulting propagated uncertainty for year j, denoted as $\sigma_u(t_j)$, is calculated as follows from $m = 1, \dots, M_j$ monthly values, each denoted as $\sigma_u^2(t_{jm})$:

$$\sigma_u(t_j) = \sqrt{\left(\frac{1}{M_j}\right)^2 \sum_{m=1}^{M_j} \sigma_u^2(t_{jm})} \tag{15}$$

Partially correlated errors represented by spatial error covariance matrices

The error covariance matrices for the non-infilled HadCRUT5 data set represent the uncertainty in the non-infilled grids that arises from persistent biases that are associated with individual marine observing platforms (e.g. an individual ship). These error covariance matrices are derived from the HadSST4 error covariance matrices, which are reweighted to account for the land-sea weighting in HadCRUT5 (as described in Text S2).

Uncertainty in spatial average time series

If $\boldsymbol{w}$ is a vector of normalized grid cell weights $\boldsymbol{w} = [w(s_1, t), \quad \dots, \quad w(s_N, t)]^T$ and $\boldsymbol{C}$ is the spatial error covariance matrix for the HadCRUT5 non-infilled data set, the uncertainty in a spatial average resulting from this uncertainty term, denoted $\sigma_p(t)$, is given by:

$$\sigma_p(t) = \sqrt{\boldsymbol{w}^T \boldsymbol{C} \boldsymbol{w}} \qquad (16)$$

For this partially correlated error term, the uncertainty in the global average requires computation of hemispheric variances and cross covariances before applying hemispheric weighting. The error covariance matrices for populated northern and southern hemisphere grid cells are notated as $\boldsymbol{C}_{NN}$ and $\boldsymbol{C}_{SS}$ and the cross covariances between grid cells of each hemisphere as $\boldsymbol{C}_{NS}$ and $\boldsymbol{C}_{SN}$. Hemispheric variances and cross covariances between hemispheres are calculated through multiplication by the normalized grid cell weight vectors for grid cells in each hemisphere, denoted $\boldsymbol{w}_N$ and $\boldsymbol{w}_S$. The hemispheres are then weighted equally by applying the hemispheric weight vector $\boldsymbol{b} = [b_N \quad b_S]^T = [0.5 \quad 0.5]^T$. The contribution to total uncertainty from the partially correlated term is then:

$$\sigma_p(t) = \left(\boldsymbol{b}^T \begin{bmatrix} \boldsymbol{w}_N^T \boldsymbol{C}_{NN} \boldsymbol{w}_N & \boldsymbol{w}_N^T \boldsymbol{C}_{NS} \boldsymbol{w}_S \\ \boldsymbol{w}_S^T \boldsymbol{C}_{SN} \boldsymbol{w}_N & \boldsymbol{w}_S^T \boldsymbol{C}_{SS} \boldsymbol{w}_S \end{bmatrix} \boldsymbol{b} \right)^{0.5} \qquad (17)$$

Uncertainty in annual average time series

For HadCRUT5, the propagation of uncertainty reported in error covariance matrices, resulting from marine platform micro biases, is simplified from the approach reported in Morice et al. (2012), following changes in time series calculation in Kennedy et al. (2019). In this simplified error model, a conservative estimate of uncertainty is made by treating this source of error as fully correlated throughout a year and independent between years.

For the partially correlated component, the uncertainty $\sigma_p(t_j)$ in an annual average for year *j* is calculated from monthly uncertainty series values $\sigma_p(t_{jm})$, for the $m = 1, \dots, M_j$ monthly series values in year *j*, as follows:

$$\sigma_p(t_j) = \frac{1}{M_j} \sum_{m=1}^{M_j} \sigma_p(t_{jm}) \qquad (18)$$

Coverage uncertainty

The coverage uncertainty calculations presented here follow the method described in Morice et al. (2012). This uncertainty term represents uncertainty arising from regions that do not contain data in the HadCRUT5 temperature anomaly fields. The coverage uncertainty estimates are computed with use of a globally complete reanalysis dataset. For HadCRUT5 coverage uncertainty estimates the reanalysis dataset used is ERA5.

We create an ensemble of $p = 1, \dots, P$ reference datasets from P complete years of temperature anomaly fields in the reanalysis dataset. Each reference ensemble member is constructed by repeating the temperature anomalies for year p in the globally complete reference dataset to cover the time period of HadCRUT5. We denote the temperature anomalies for the reference constructed from year p of the globally complete reference dataset as $R_p(s,t)$. We then mask the globally complete fields $R_p(s,t)$ to the coverage of HadCRUT5 at time t and denote the values of the masked fields as $\tilde{R}_p(s,t)$.

The ensemble of P spatial fields at time t provides P samples of temperature anomaly variability, with each sample derived from a different year of reanalysis data with appropriate variability for each calendar month, as represented in the reanalysis dataset, that can be used to assess the error in monthly or annual time series for a given grid coverage. Errors in temperature anomaly time series are computed by calculating time series for the globally complete and the masked reference fields. Denoting the time series value at time t for the pth globally complete reference data set as $R_p(t)$ and that derived from the coverage reduced field as $\tilde{R}_p(t)$, the error associated with the omitted grid cells is calculated as:

$$\epsilon_p(t) = \tilde{R}_p(t) - R_p(t) \tag{19}$$

The coverage uncertainty is then computed as the root mean square of the P error samples. This approach differs from Morice et al. (2012), which used the standard deviation rather than root mean square of the error samples, and results in larger uncertainty estimates. Using the root mean square metric, the estimate of coverage uncertainty, $\sigma_c(t)$, is calculated as:

$$\sigma_c(t) = \sqrt{\frac{1}{P}\sum_{p=1}^{P}\left(\tilde{R}_p(t) - R_p(t)\right)^2} \tag{20}$$

Ensemble statistics (mean and spread)

For the non-infilled HadCRUT5 dataset, as in Morice et al. (2012), the ensemble spread represents the uncertainty arising from systematic measurement biases. Here we describe the calculation of summary statistics from the ensemble. For $D = 200$ ensemble members, diagnostics (e.g. a regional average monthly or annual series) are computed for the $d = 1, \dots, D$ ensemble members. Summary statistics are then computed from the D ensemble member diagnostics.

Ensemble statistics for spatial fields

When deriving ensemble statistics for spatial fields, summary statistics are computed directly from the ensemble grids. Values of the ensemble mean, $\mu_e(s,t)$, at spatial location s and time t, are computed from the corresponding ensemble member temperature anomalies, $T_d(s,t)$, as follows:

$$\mu_e(s,t) = \frac{1}{D}\sum_{d=1}^{D} T_d(s,t) \tag{21}$$

Similarly, uncertainty ranges for individual grid cells, $\sigma_e(s,t)$, are derived from the ensemble fields as:

$$\sigma_e(s,t) = \sqrt{\frac{1}{D-1}\sum_{d=1}^{D}\left(T_d(s,t) - \mu_e(s,t)\right)^2} \tag{22}$$

Ensemble statistics for time series

For global and regional average time series diagnostics, we first compute time series for each ensemble member and then compute summary statistics. Denoting the value of a diagnostic time series for an individual ensemble member at time t as $T_d(t)$, the ensemble mean, $\mu_e(t)$, is defined as:

$$\mu_e(t) = \frac{1}{D}\sum_{d=1}^{D} T_d(t) \tag{23}$$

The uncertainty derived from the ensemble, $\sigma_e(t)$, is computed as the ensemble standard deviation as:

$$\sigma_e(t) = \sqrt{\frac{1}{D-1}\sum_{d=1}^{D}\left(T_d(t) - \mu_e(t)\right)^2} \tag{24}$$

Best estimate time series and total uncertainty: HadCRUT5 non-infilled dataset

We define the 'best estimate' time series for the non-infilled HadCRUT5 dataset as the ensemble mean of the $D = 200$ ensemble member time series, with equal weighting given to each ensemble member. The values of this time series $\mu(t)$ are therefore equal to the ensemble mean values $\mu_e(t)$ and are computed as:

$$\mu(t) = \mu_e(t) = \frac{1}{D}\sum_{d=1}^{D} T_d(t) \tag{25}$$

The full uncertainty model for the non-infilled HadCRUT5 dataset comprises four distinct components: observational bias uncertainty represented in the ensemble, $\sigma_e(t)$, uncertainty from uncorrelated measurement and sampling errors, $\sigma_u(t)$, uncertainties arising from individual marine observing platform biases, $\sigma_p(t)$, and uncertainty arising from incomplete observational sampling of the globe, $\sigma_c(t)$.

These terms are combined in quadrature to obtain $\sigma(t)$, the total uncertainty in time series for the non-infilled dataset:

$$\sigma(t) = \sqrt{\sigma_e(t)^2 + \sigma_u(t)^2 + \sigma_p(t)^2 + \sigma_c(t)^2} \tag{26}$$

Text S4 Time series calculation – HadCRUT5 analysis

The HadCRUT5 analysis methods encode uncertainties represented in the non-infilled ensemble members, the uncorrelated measurement and sampling error component, and the partially correlated component into the analysis ensemble fields. This results in fewer terms in the error model for the HadCRUT5 analysis than for the non-infilled dataset (see Text S1 for details) because the effects of these terms are represented the analysis ensemble. The resulting uncertainty model for the HadCRUT5 analysis has two terms: the uncertainty that is encoded into the analysis ensemble and the remaining uncertainty from incomplete coverage of the analysis fields.

Best estimate time series and total uncertainty: HadCRUT5 analysis

The best estimate of a regional or global average statistic is computed as the ensemble mean or time series derived from each ensemble member. These ensemble member time series are computed using the methods for spatial and temporal averaging described in Text S3. The 'best estimate' time series is then computed as the ensemble mean time series as:

$$\mu(t) = \mu_e(t) = \frac{1}{D}\sum_{d=1}^{D} T_d(t) \qquad (27)$$

This is the same as for the non-infilled dataset, but now the terms refer to the diagnostic (regional or global average) computed from the HadCRUT5 analysis ensemble.

The total uncertainty is defined by the combination of the uncertainty represented in the analysis ensemble, $\sigma_e(t)$, and the coverage uncertainty associated with regions omitted from the masked HadCRUT5 analysis, $\sigma_c(t)$. The ensemble uncertainty is calculated following the methods described in Text S3, noting that these values are now computed for the analysis temperature anomaly ensemble fields. The coverage uncertainty calculations follow those described in Text S3, with reference reanalysis fields masked to the HadCRUT5 analysis coverage. Hence, the coverage uncertainty represents the uncertainty arising from missing data regions in the analysis, where the analysis is masked because of a weak local observational constraint.

The uncertainty in the 'best estimate' temperature anomaly time series, $\sigma(t)$, is then computed by combining the two uncertainty components: the ensemble uncertainty, $\sigma_e(t)$, and coverage uncertainty, $\sigma_c(t)$, in quadrature:

$$\sigma(t) = \sqrt{\sigma_e(t)^2 + \sigma_c(t)^2} \qquad (28)$$

Text S5 HadCRUT5 analysis hyperparameter estimation

Here we provide additional information of the land air temperature and sea-surface temperature analysis hyperparameter estimates. Analysis hyperparameters are fit using the maximum marginal likelihood method that is described in Appendix A of the main manuscript.

We use a Matérn covariance function, for which the covariance $k(s_m, s_n)$ is computed as a function of distance $d(s_m, s_n)$ between locations s_m and s_n. Here we measure this distance as the Euclidian or chordal distance between the two locations on the surface of a spherical Earth. The Matérn covariance function is parameterized as a function of a range hyperparameter r, scale hyperparameter σ and smoothing hyperparameter ν, and is defined as:

$$k(s_m, s_n) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}}{r} d(s_m, s_n) \right)^{\nu} K_{\nu} \left(\frac{\sqrt{2\nu}}{r} d(s_m, s_n) \right) \tag{29}$$

where Γ is the gamma function and K_ν is the modified Bessel function of the second kind of order ν. We use a fixed Matérn smoothing parameter of $\nu = 1.5$ and optimize the scale and range parameters (σ, r).

We computed the marginal maximum likelihood optimization for each monthly field of land air temperature anomalies and sea-surface temperature anomalies as described in Appendix A. Our land and marine hyperparameter estimates are computed as an average of the monthly optimized values for the 360 monthly fields in the 1961 to 1990 climatology period. Finally, we rounded the scale parameter estimates to the nearest 0.05 ˚C and range parameters estimates to the nearest 50 km. We have chosen to use parameter estimates based on 1961-1990 data for the following reasons: (i) this is a period of good global observational coverage for land and ocean; (ii) observing methods in the early record are less well understood (e.g. see Osborn et al 2020 and Kennedy et al 2019); and (iii) parameter estimates in the climatology period are less likely to be effected by differences in regional trends.

Monthly values of optimized hyperparameters are shown in Figures S1 and S2, along with the average parameters in the 1961-1990 period. We note that there is non-stationarity in the hyperparameter fits over time for marine parameter estimates. It is not clear whether this is a real feature of the temperature field or a result of e.g. differences in variability as fully unsampled regions become sampled or due to changes in observation methods that are not described by our uncertainty model. We also note that marine correlation structure shows significant spatial anisotropy, as demonstrated in Kennedy et al. (2019), with regional variation and much longer correlation ranges in zonal directions than meridional, which may also have an impact when combined with changes in spatial sampling over time.

Text S6 Masking of the HadCRUT5 analysis fields by observational constraint

The HadCRUT5 analysis fields are generally not globally complete because regions with weak observational constraint are masked from the analysis fields. The masking is controlled by a parameter α which sets a threshold for the ratio of the spatial Gaussian process' posterior distribution variance to its prior distribution variance (see Appendix A.4 of the main article). The HadCRUT5 analysis uses a value of this threshold of $\alpha = 0.25$, which corresponds to masking the analysis in regions where there is a reduction in variance from prior to posterior of less than 25%.

Figures S3 to S5 show global and hemispheric annual time series and uncertainties for varying values of analysis masking parameter ranging from $\alpha = 0.0$ (no masking) to $\alpha = 0.5$. Figure S6 shows the corresponding global and hemispheric coverage for the masked grids contributing to the annual averages. At values of $\alpha = 0.5$ and above, the analysis is masked at grid cells that are actually observed in the non-infilled dataset. At these large values of α, these observed grid cells are masked because observational uncertainty is sufficiently great that the observations do not provide enough information to achieve the required analysis constraint. Hence, we do not consider values of α larger than $\alpha = 0.5$.

There is little sensitivity of global and hemispheric diagnostics to variation in the range $0.1 \leq \alpha \leq 0.5$. When $\alpha = 0.0$ the coverage is global, however, this unmasked analysis may not faithfully represent regional trends for the unconstrained regions. Uncertainties in global and regional time series are also insensitive to variations in α in the range $0.1 \leq \alpha \leq 0.5$. However, for $\alpha = 0.0$, the uncertainty is notably larger due to the sampling strategy for the analysis ensemble (described in Appendix A) where the analysis errors are modelled treating persistence in temperature anomalies as being fully correlated in time during a year. This results in conservative estimates of uncertainty for annual averages in regions where the analysis uncertainty is large (i.e. regions with a weak observational constraint).

Figure S7 shows the ensemble spread for the unmasked analysis over various time periods while the ensemble spread for the masked analysis is shown in Figure S8. The metric of ensemble spread used here is computed by combining the following two quantities in quadrature: (1) the standard deviation across the time period of the grid cell ensemble means and (2) the mean across the time period of the monthly ensemble standard deviations for each grid cell.

The unmasked land and ocean analyses each tend toward a uniform ensemble spread in regions of weak observational constraint (Figure S7). This happens because the Gaussian process priors, fit separately for land and ocean, model typical variability in each domain. This effect is mitigated by masking regions of weak observational constraint (Figure S8), and the uncertainty in spatial average time series that results from the masking is represented by the coverage uncertainty estimates. It should be noted, however, that the analysis is able to represent variations in regional variability for regions where the analysis is constrained by local observations.

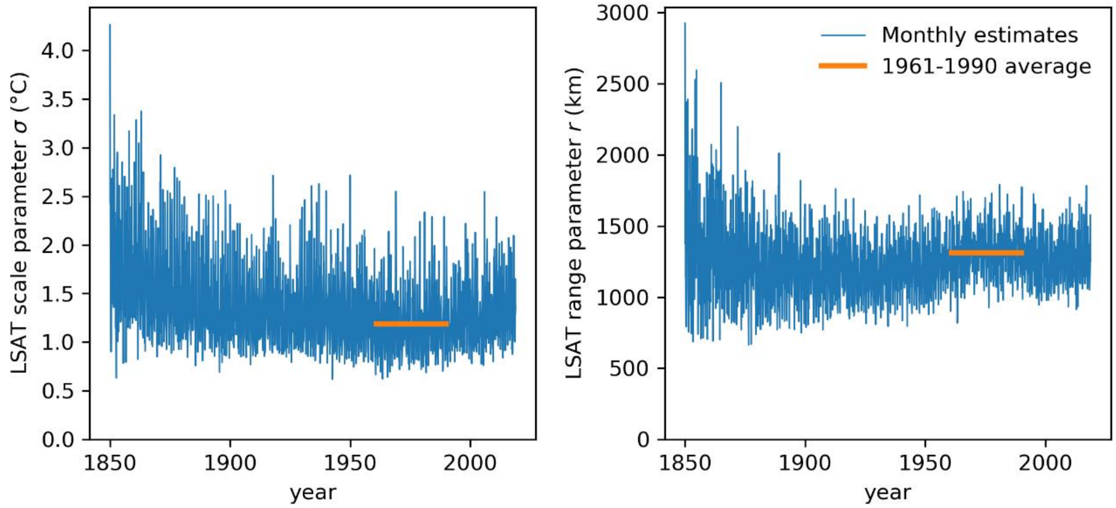

Figure S1. Estimates of land surface air temperature (LSAT) analysis scale (left) and range (right) hyperparameters. Monthly estimates by maximum marginal log likelihood optimization are shown in blue. The average of the monthly estimates over the 1961-1990 climatology period are shown in orange.

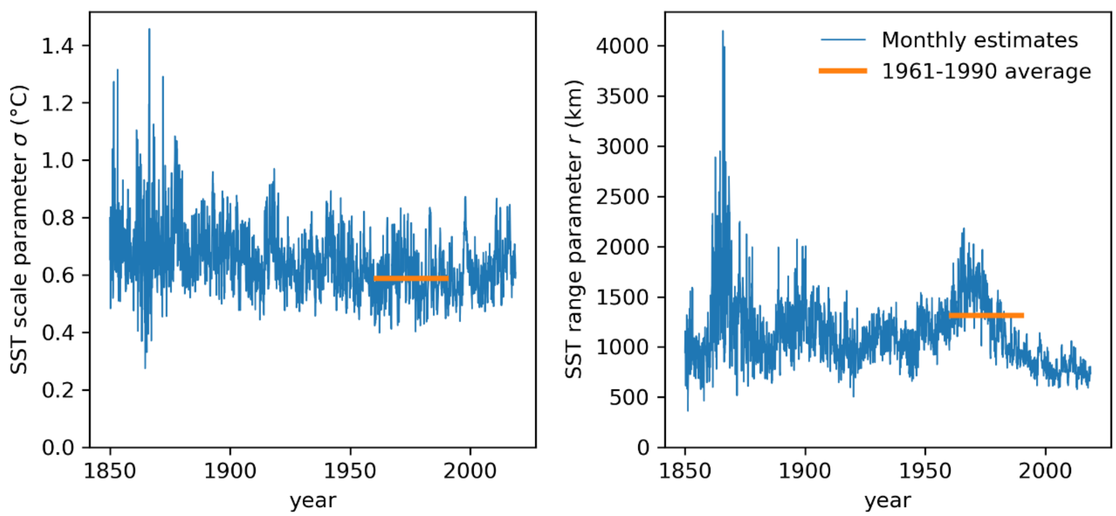

Figure S2. Estimates of sea-surface temperature (SST) analysis scale (left) and range (right) hyperparameters. Monthly estimates by maximum marginal log likelihood optimization are shown in blue. The average of the monthly estimates over the 1961-1990 climatology period are shown in orange.

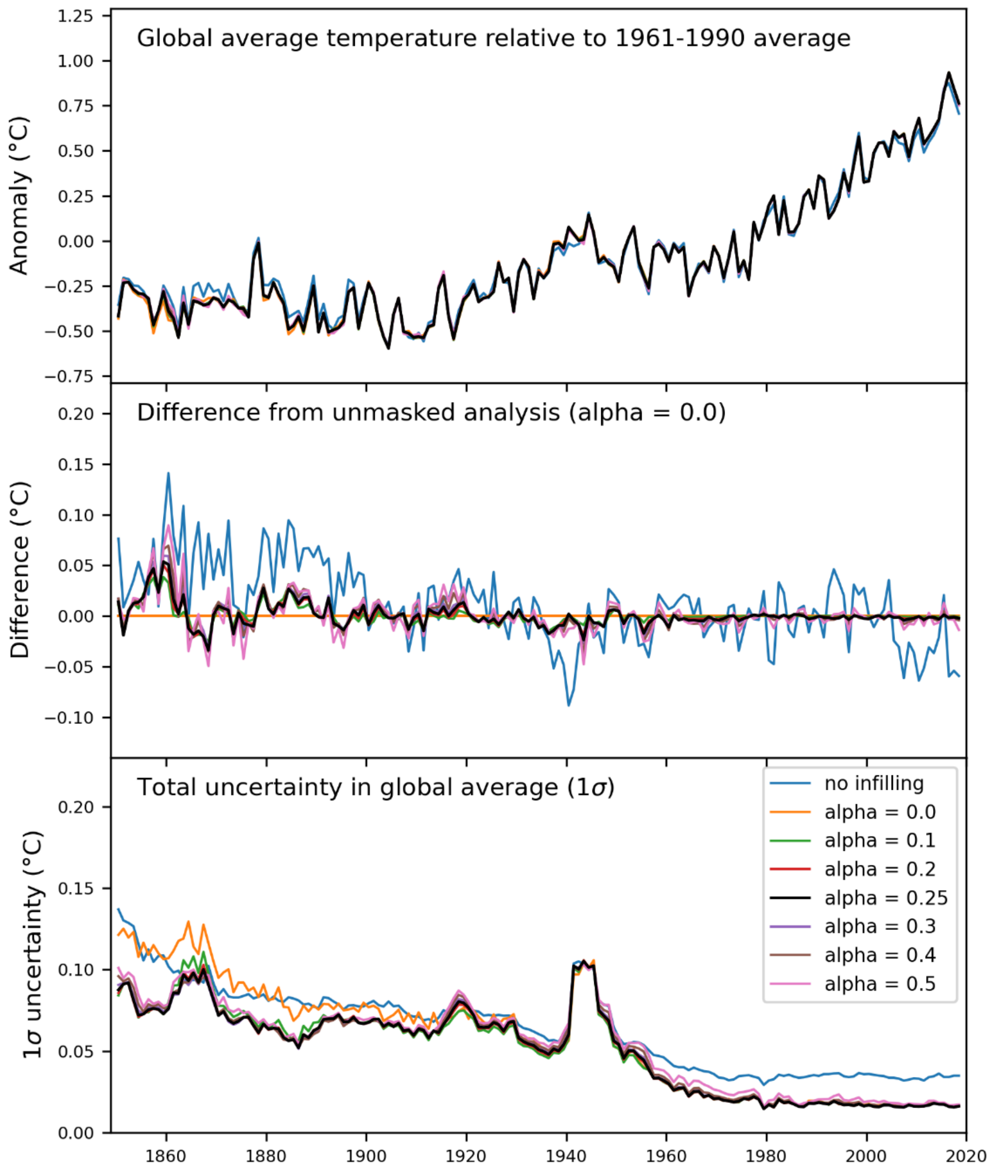

Figure S3. Comparison of annual global mean surface temperature diagnostics for varying levels of masking of the analysis, controlled by the observation constraint threshold $\boldsymbol{\alpha}$. Higher values of alpha indicate more masking. Also shown is the time series for the non-infilled HadCRUT5 dataset (blue). (Top) annual global average temperature anomaly series (^{0}C). (Middle) difference in global average series from the fully unmasked analysis ($\boldsymbol{\alpha = 0.0}$, orange). (Bottom) estimated uncertainty ranges ($\boldsymbol{1\ \sigma}$) in annual average global mean surface temperature for varying values of observation constraint threshold.

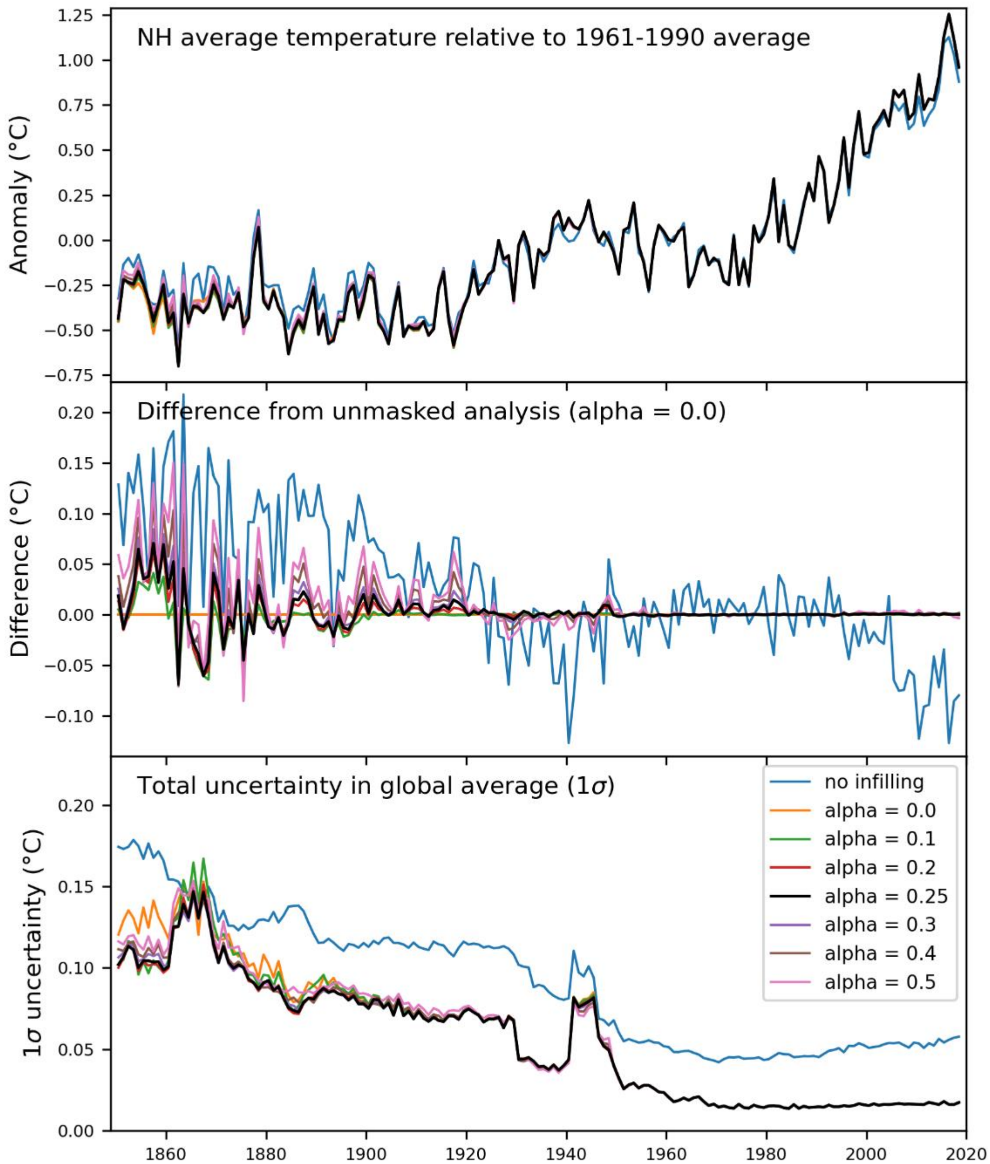

Figure S4. As Figure S3 for the Northern Hemisphere.

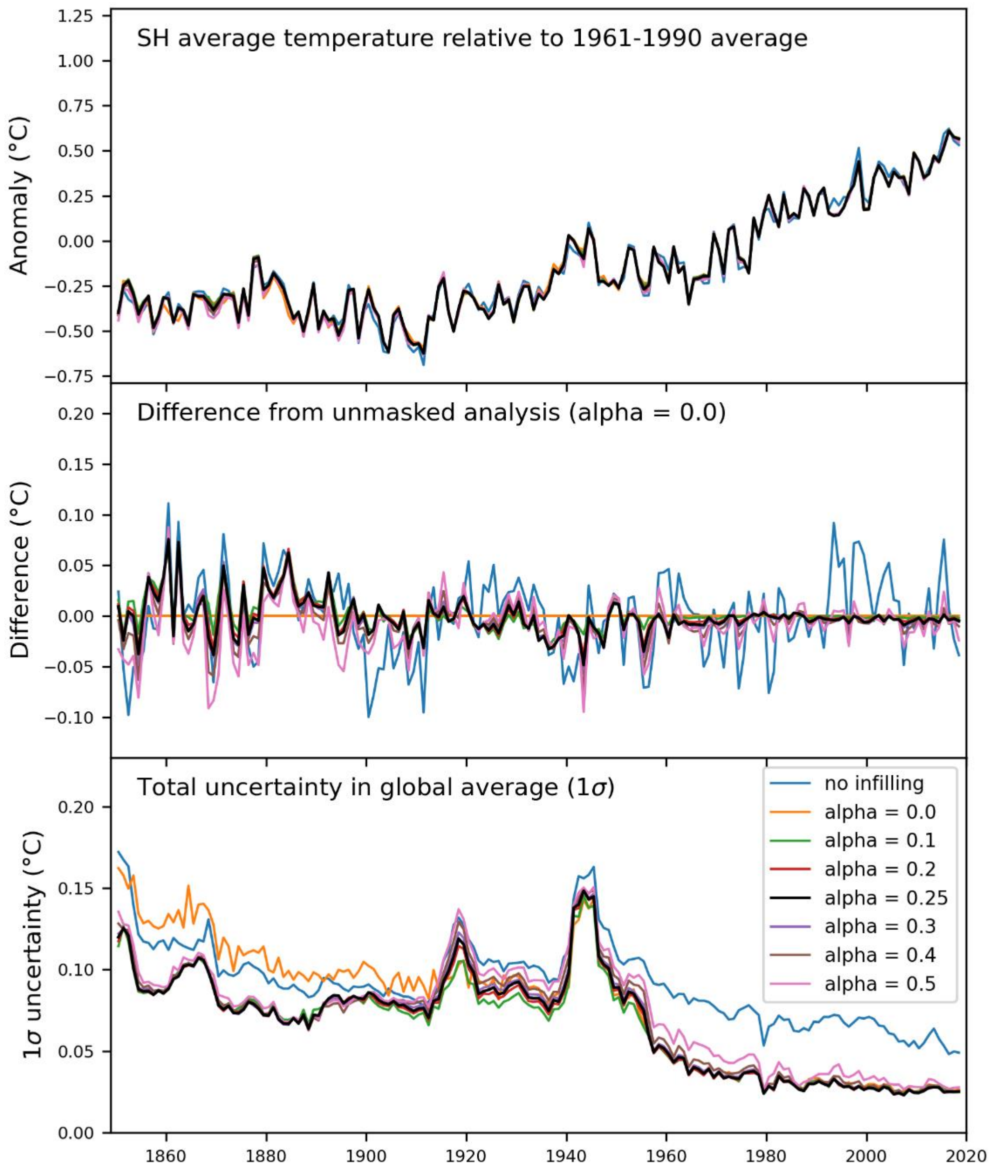

Figure S5. As Figure S3 for the Southern Hemisphere.

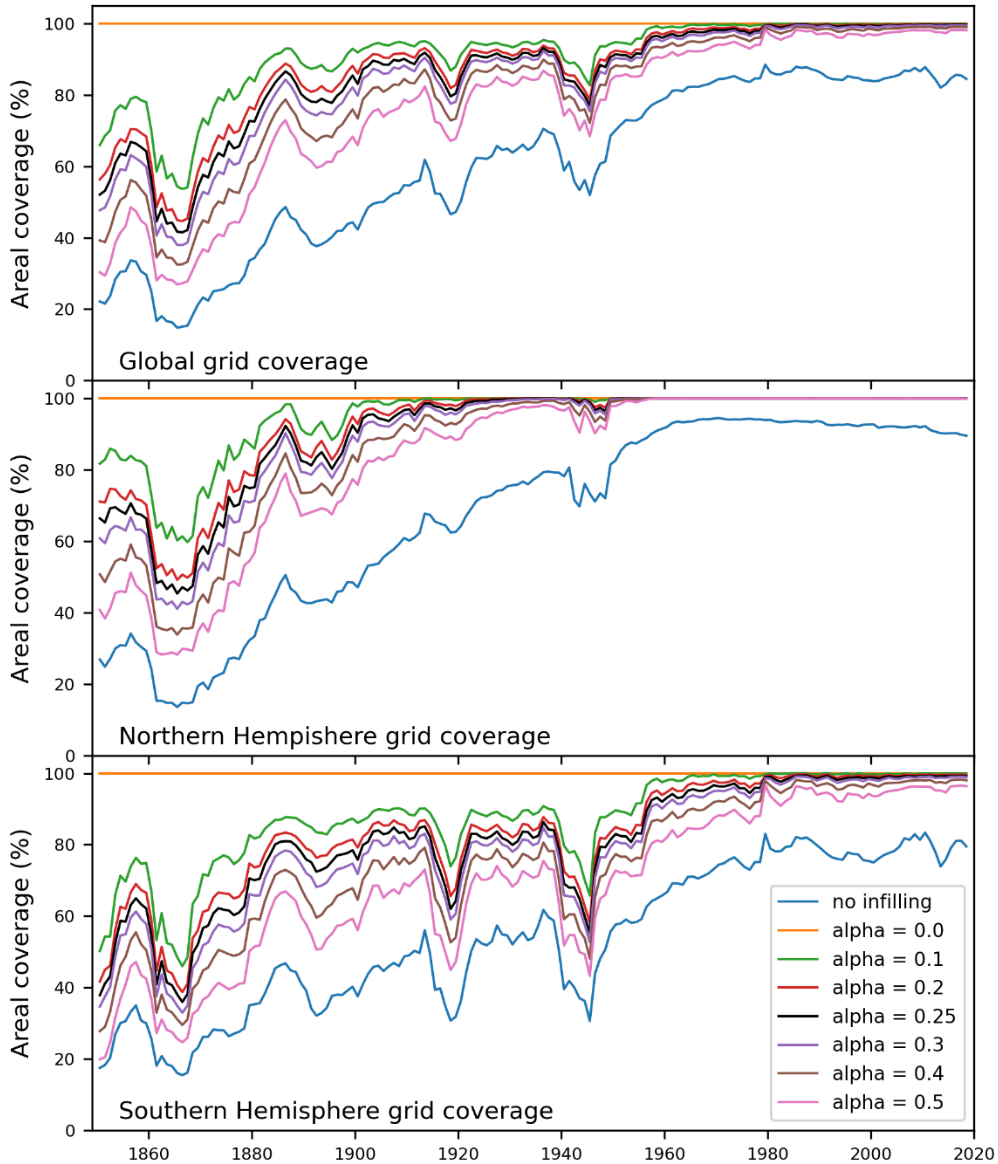

Figure S6. Percentage area of the globe represented by the analysis for varying values of the masking parameter $\boldsymbol{\alpha}$.

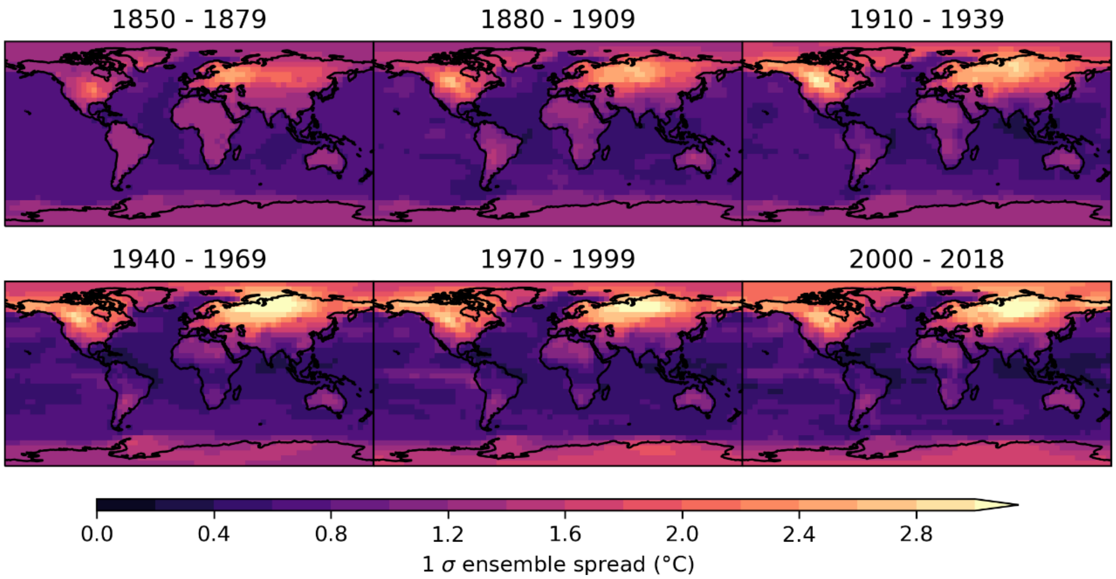

Figure S7. Ensemble spread ($\mathbf{1}\ \boldsymbol{\sigma}$) for the global analysis without masking regions with weak observational constraint ($\boldsymbol{\alpha} = \mathbf{0.0}$). The ensemble spread for each grid cell is computed here as the time average of the monthly ensemble standard deviations summed in quadrature with the standard deviation of the monthly ensemble means.

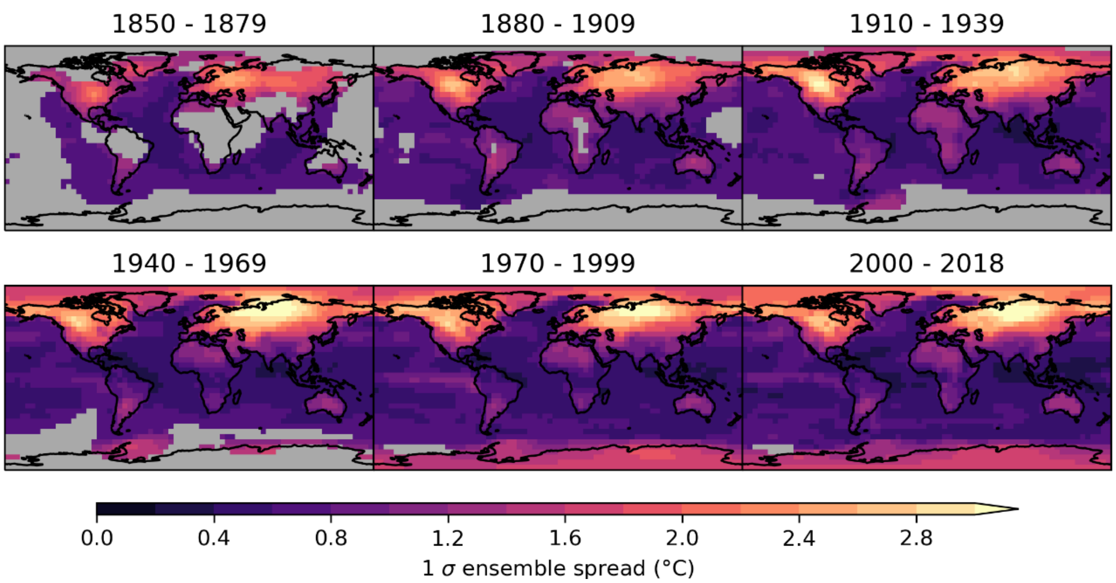

Figure S8. Ensemble spread ($\mathbf{1}\ \boldsymbol{\sigma}$) for the masked HadCRUT5 analysis, with observation constraint ($\boldsymbol{\alpha} = \mathbf{0.25}$). Grid cells are plotted where at least 50% of grid cells are populated during each time period. The ensemble spread is computed as in Figure S7.

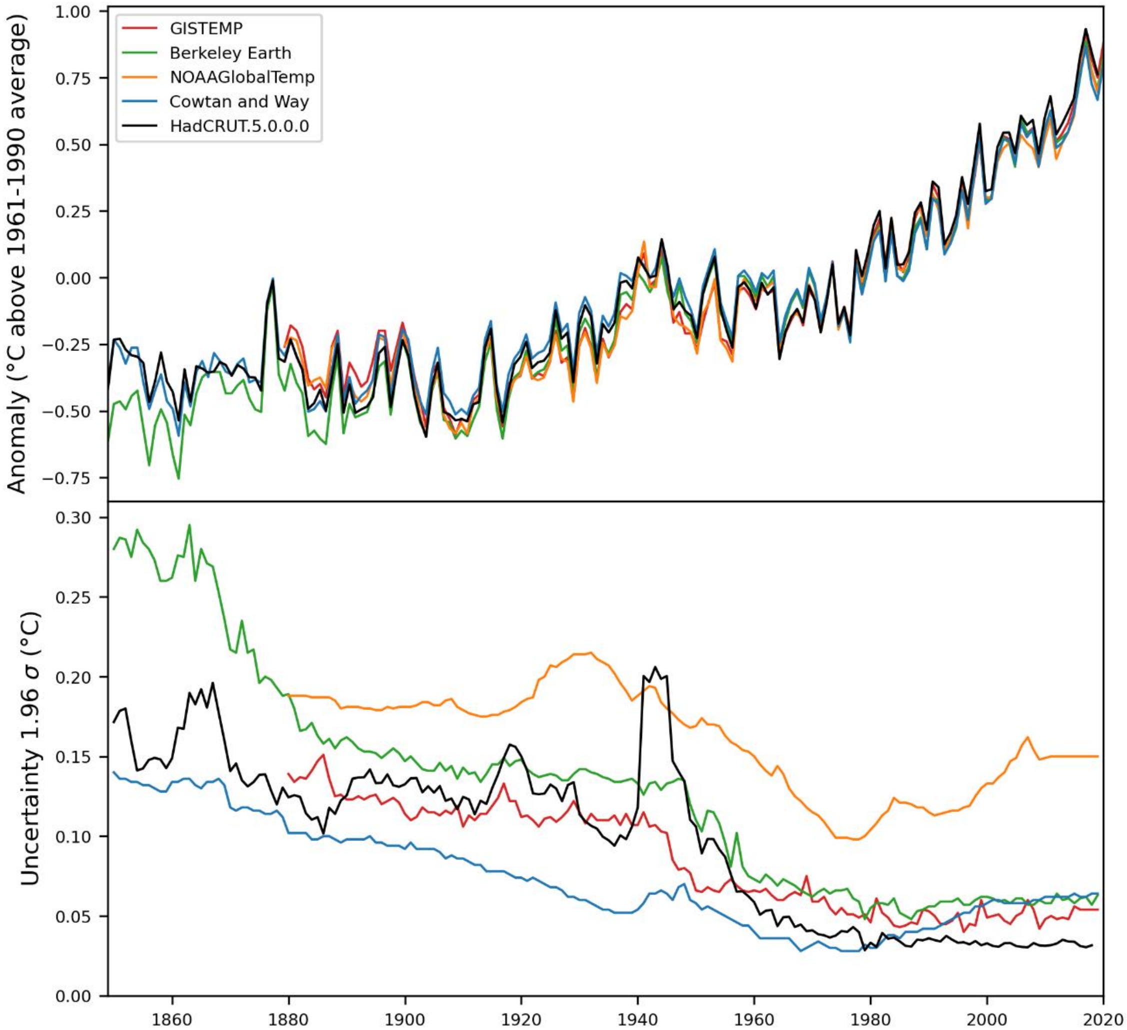

Figure S9. Global average temperature anomaly series (upper panel) and uncertainties (lower panel, ^{0}C) reported for a range of datasets. Series are as provided by producers of each dataset, with anomalies adjusted to a common reference period of 1961-1990 and uncertainties expressed as 1.96 sigma or 95% confidence range. For NOAAGlobalTemp, the uncertainty range is taken from version 4 of the data set as the v5 uncertainties were not available at the time of writing.

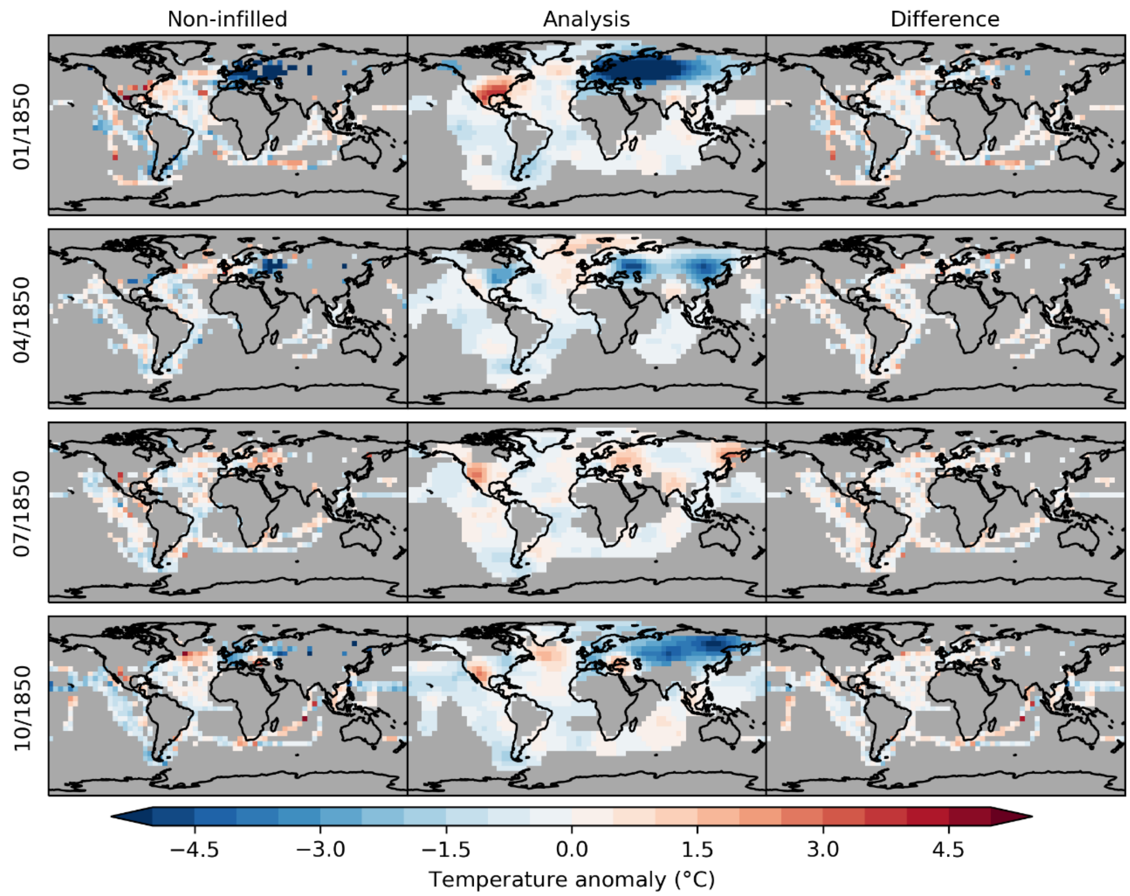

Figure S10. Monthly temperature anomaly fields (°C relative to 1961 to 1990 average) for January, April, July and October 1850, showing the non-infilled dataset (left), the HadCRUT5 analysis (middle) and differences between the non-infilled dataset and the analysis (right).

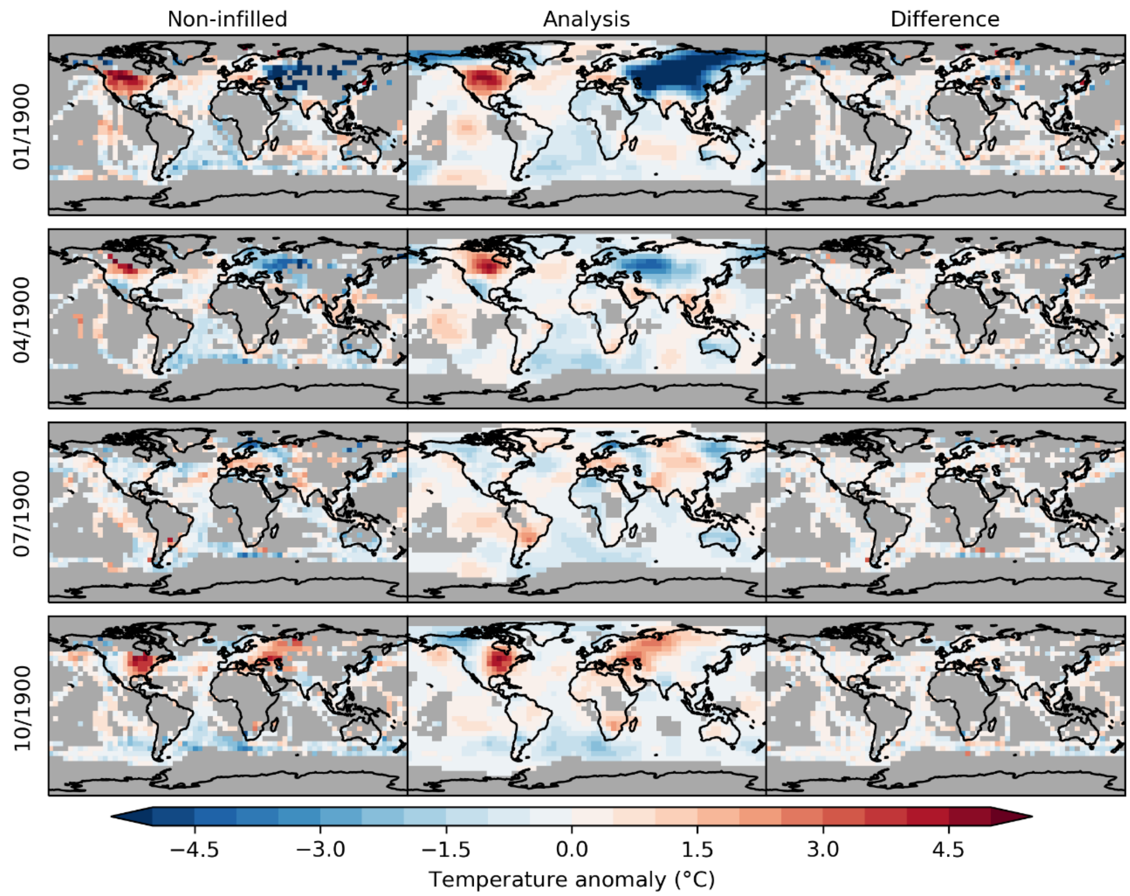

Figure S11. Monthly temperature anomaly fields (°C relative to 1961 to 1990 average) for January, April, July and October 1900, showing the non-infilled dataset (left), the HadCRUT5 analysis (middle) and differences between the non-infilled dataset and the analysis (right).

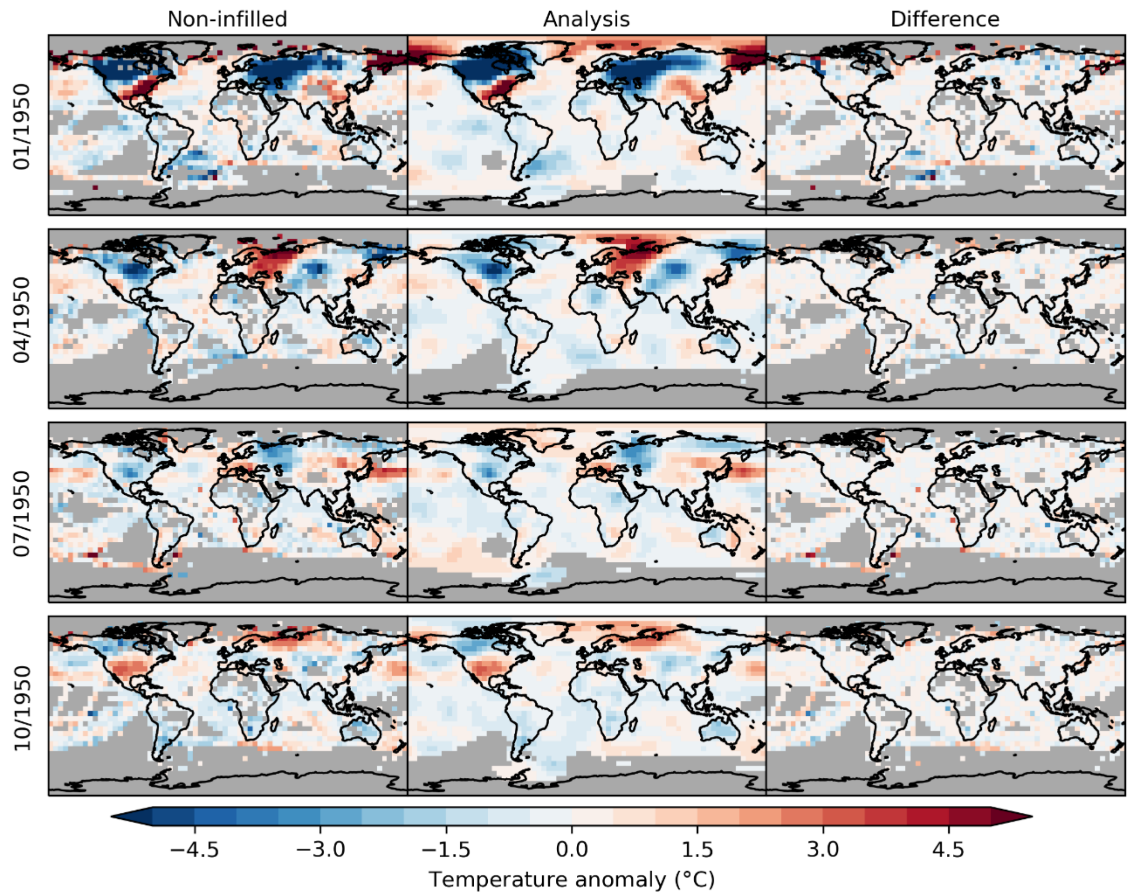

Figure S12. Monthly temperature anomaly fields (°C relative to 1961 to 1990 average) for January, April, July and October 1950, showing the non-infilled dataset (left), the HadCRUT5 analysis (middle) and differences between the non-infilled dataset and the analysis (right).

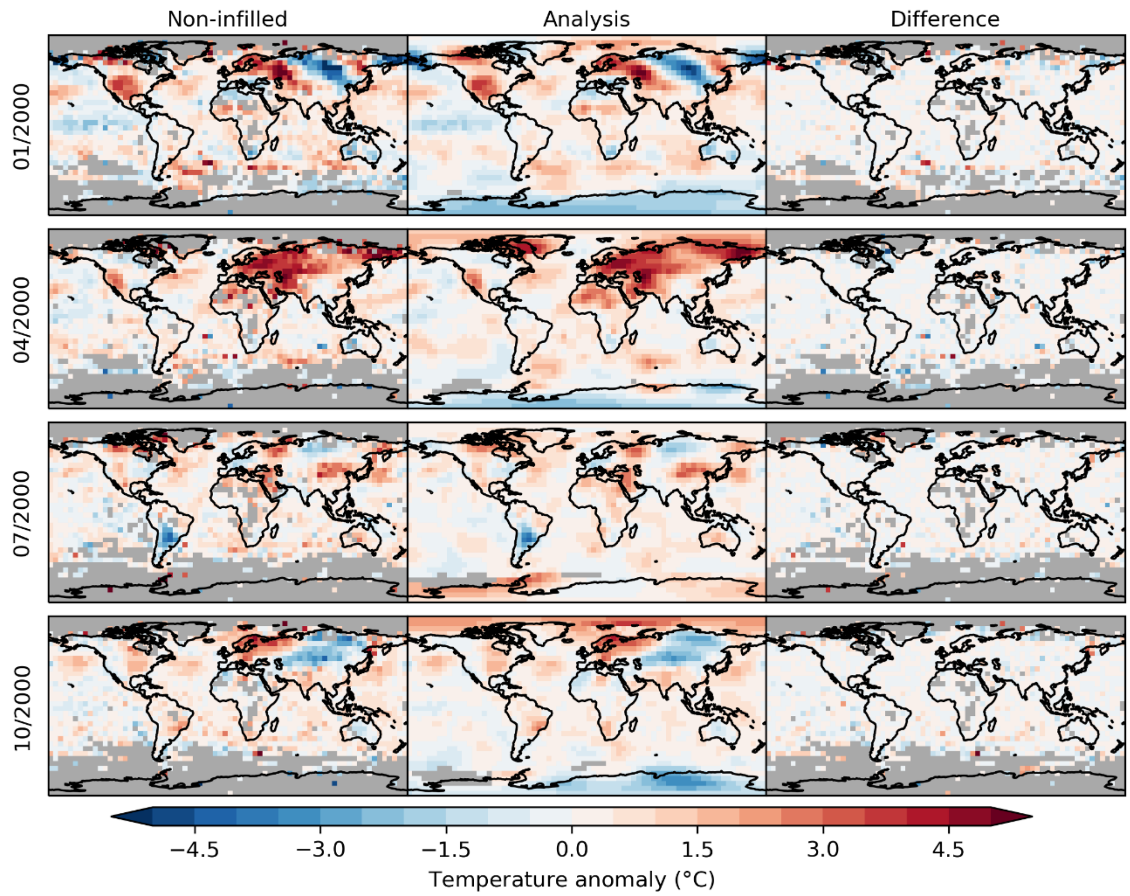

Figure S13. Monthly temperature anomaly fields (°C relative to 1961 to 1990 average) for January, April, July and October 2000, showing the non-infilled dataset (left), the HadCRUT5 analysis (middle) and differences between the non-infilled dataset and the analysis (right).

	Ensemble	**Uncorrelated uncertainty** **(per grid-cell)**	**Error covariance matrix** **(inter-grid-cell)**	**Coverage uncertainty** **(regional average time series only)**
HadCRUT5 non-infilled dataset	Samples uncertainty from systematic biases from land station homogenization error, urbanization and non-standard measurement enclosures (Morice et al., 2012) and adjustments applied to correct for changes in marine measurement methods (Kennedy et al., 2019).	Per-grid-cell 1-sigma uncertainty associated with within-grid-cell observational sampling and random measurement error.	Between-grid-cell and within-grid-cell partially-correlated uncertainties. Describes the effect of individual marine measurement platform-specific 'micro biases' (Kennedy et al., 2019) as platforms move between grid cells.	Uncertainty in regional average time series arising from incomplete global coverage of the non-infilled grids.
HadCRUT5 analysis	Samples all uncertainties modelled for the HadCRUT5 analysis temperature anomaly fields. Includes the effects of systematic measurement biases, per-grid-cell uncorrelated uncertainties, inter-grid-cell error covariances associated with the gridded observations and the uncertainty associated with the statistical analysis method.	Included in the analysis ensemble.	Included in the analysis ensemble.	Uncertainty in regional average time series arising from incomplete global coverage from masking the analysis in regions of weak observational constraint.

Table S1. HadCRUT5 uncertainty model components provided for the non-infilled dataset and the HadCRUT5 analysis.