



## Research Article

A SCITECHNOL JOURNAL

# Correcting for Amplification Bias and Sea Ice Bias in Global Temperature Datasets

Bruce T. T. Calvert\*

### Abstract

To estimate changes in global mean surface temperature (GMST), one must infer past temperatures for regions of the planet that lacked observations. However, current global instrumental temperature datasets (GITDs) do not adequately account for the tendency of different regions of the planet to warm at different rates, creating a bias in their estimates, which this paper calls amplification bias. In addition, most GITDs do not adequately account for changes in sea ice, creating a bias in their estimates, which this paper calls sea ice bias. To estimate the impact of these two biases, a new GITD was created that used maximum likelihood estimation (MLE) to combine the land surface air temperature (LSAT) anomalies of HadCRUT4 with the sea surface temperature (SST) anomalies of HadSST4. The new GITD has improvements compared to the Cowtan and Way version 2 dataset, including an improved statistical foundation for estimating model parameters, taking advantage of temporal correlations of observations, taking advantage of correlations between land and sea observations, accounting for more sources of uncertainty, and better treatment of the El Niño Southern Oscillation (ENSO). Corrections for amplification bias and sea ice bias in the new dataset increase the estimate of GMST change from the late 1800s (1850-1899) to 2018 by 0.01°C and 0.08°C respectively, although tests suggest that there may be an overcorrection by a factor of two for sea ice bias. Overall, the median estimate of GMST change from the late 1800s to 2018 is 1.20°C, with a 95% confidence interval of (1.11°C, 1.30°C).

### Keywords

Climate Change; Global Warming; Instrumental; Temperature; Amplification; Sea Ice; Bias; Maximum Likelihood; Internal Variability; El Niño

## Introduction

Estimating the evolution of global surface temperatures since the industrial revolution is important as it can help improve understanding of the Earth's climate system. Currently, there are various GITDs, including HadCRUT4 [1] by the UK Met Office Hadley Centre (Met Office) in conjunction with the Climatic Research Unit (CRU) of the University of East Anglia, GISTEMPv4 [2] by NASA's Goddard Institute of Space Studies, and NOAA GlobalTempv5 [3-5] by NOAA's National Climate Data Centre. One major issue with estimating past temperatures is that not all regions of the Earth have complete observational records. As a result, one must statistically infer

temperatures of unobserved regions of the planet using available data to properly understand the evolution of global surface temperatures since the industrial revolution.

In the past decade, two new GITDs have been created that use a geostatistical technique known as kriging [6,7] to infill temperatures for unobserved regions of the Earth: Berkeley Earth Surface Temperature [8-10], and Cowtan and Way [11,12]. For convenience, this paper refers to Berkeley Earth Surface Temperature using LSATs for sea ice regions as BEST, and refers to the long reconstructions of Cowtan and Way version 2 using HadSST3 [13,14] and HadSST4 [15] as C&W-HadSST3 and C&W respectively. Kriging has desirable statistical properties: if covariances between temperature observations are known with certainty, then kriging provides the most-efficient linear unbiased estimates of temperatures of unobserved regions. In addition, the kriging estimates are identical to the maximum likelihood estimates if the distribution of residuals is multivariate normal.

While kriging has desirable statistical properties when covariances between temperature observations are known with certainty, in reality, these covariances are unknown and must be empirically estimated from available data. Ordinary kriging, by itself, does not provide a method to empirically estimate these covariances, nor does it account for the uncertainty due to estimating these covariances. To address these issues, this study uses MLE to estimate parameters in the covariance function and quantify their uncertainty. MLE has very desirable statistical properties when the number of observations is large: consistency, asymptotic efficiency, and asymptotic normality.

The underlying statistical models of BEST and C&W do not adequately account for the statistical tendency of different regions of the planet to warm at different rates and, therefore, are subject to amplification bias. In addition, C&W does not account for changes in sea ice and so is subject to sea ice bias. To quantify these biases, the Akaike information criterion (AIC) [16] was used to select the best of 16 different temperature anomaly models and the best of 24 different temperature climatology models. Using AIC to pick from a large set of models helps reduce omitted variable bias and avoid the statistical issue of overfitting. The best models were combined to construct a new GITD. This GITD is given the name HadCRU\_MLE, to reflect its use of MLE and data primarily from the Met Office and CRU. Similar to C&W, LSAT anomalies of HadCRUT4 and SST anomalies of HadSST4 were used, so the results of this study are most comparable to the results of C&W.

## Data

5° by 5° gridded monthly temperature anomalies of HadCRUT4 and HadSST3 were obtained from the Met Office website. This includes 100 ensemble members as well as ensemble medians. These datasets were used to infer LSAT anomalies of HadCRUT4, which are essentially CRUTEM4 [17] but with adjustments to account for biases and uncertainties such as due to temperature homogenization and the urban heat island effect. Measurement and sampling uncertainties of CRUTEM4 were also obtained.

5° by 5° gridded monthly temperature anomalies of HadSST4 were obtained from the Met Office website. This includes 200

\*Corresponding author: Bruce T. T. Calvert, Independent Researcher, Ottawa, Canada, E-mail: brucetcalvert@gmail.com

Received: December 09, 2020 Accepted: January 21, 2021 Published: May 31, 2021

ensemble members, which account for bias uncertainties, as well as the ensemble median. The 100 ensemble members of HadCRUT4 LSAT anomalies were combined with the 200 ensemble members of HadSST4 to produce 200 ensemble members of temperature anomaly observations; each LSAT ensemble member was used twice. Error covariance matrices and total uncertainty of HadSST4, which account for measurement and sampling uncertainties, were also obtained.

0.25° by 0.25° gridded land mask data of OSTIA [18] was obtained from the Copernicus Marine Environment Monitoring Service website. OSTIA data was integrated to a 5° by 5° grid to replicate the land fraction data used in HadCRUT4. OSTIA data was also integrated to a 1° by 1° grid. OSTIA was used in combination with the HadCRUT4, HadSST3, and CRUTEM4 data to infer LSAT anomalies of HadCRUT4.

1° by 1° gridded monthly sea ice concentrations (SICs) of HadISST1 [19] and HadISST2 [20] were obtained from the Met Office website. These SICs were component-wise multiplied by the non-land fraction to obtain sea ice fractions (SIFs). This paper defines the SIF of a grid cell as the fraction of the total surface area of a grid cell covered by sea ice. SIFs were integrated to a 5° by 5° grid. HadISST2 was used as the main SIC dataset of this study; HadISST1 was used only to determine how much sea ice extent varies between datasets.

The 5° by 5° gridded 1961-1990 temperature climatology of Jones et al. [21] and 100 ensemble members of the 5° by 5° gridded 1961-1990 temperature climatology of HadSST3 were obtained from the Met Office website. These temperature climatologies were used to help estimate and correct for sea ice bias. The 1° by 1° gridded 1979-1998 temperature climatology of IABP-POLES [22] was obtained from the University of Washington website. IABP-POLES was used to test for the impact of spatial resolution on the estimate of sea ice bias.

Time series data of the ensemble medians of C&W-HadSST3 and C&W was obtained from the University of York website. In addition, time series data of BEST was obtained from the Berkeley Earth website. These datasets were used for comparison with the results of this study.

30 arc-second resolution GMTED2010 surface elevation data [23] was obtained from the US Geological Survey website. This was integrated to a 0.25° by 0.25° grid. This was combined with the land fraction data and integrated to a 5° by 5° grid to produce average surface elevation data for land regions and non-land regions. This average surface elevation data was used to help explain variation in the temperature climatology.

CCSM4 output for the pre-industrial control scenario and for the RCP6.0 scenario was obtained from the University Corporation of Atmospheric Research website. This includes monthly gridded surface air temperatures (SATs) 2 m above the surface and monthly gridded SICs. The climate model output for the pre-industrial control scenario was used to estimate temperature patterns of internal variability. As the RCP6.0 scenario roughly corresponds to historic forcing, the output for the RCP6.0 scenario was used for comparison with the corrections for sea ice bias of this study. CCSM4 was chosen because it “has been analysed the most extensively of any current climate model with regards to [interdecadal Pacific variability] processes and mechanisms, and compares favourably in those aspects to observations” [24]. Only the last 501 years (800-1300) of the pre-industrial control scenario were used because the model was not necessarily in equilibrium at the beginning of the model run.

SICs of CCSM4 are given in a Greenland pole grid, where the

North Pole is moved to Greenland to avoid singularity problems in the sea ice model. Nearest neighbour interpolation was performed to convert SICs to a 0.125° by 0.125° grid; this small grid size was used to reduce interpolation error. These SICs were integrated to a 0.25° by 0.25° grid and component-wise multiplied by the non-land fraction to obtain SIFs. Similarly, SATs, which are given in a 1.25° longitude by 0.9375° latitude grid, were converted to a 0.125° by 0.125° grid using nearest neighbour interpolation. SIFs and SATs were integrated to a 5° by 5° grid. 5° by 5° climatologies of SIFs and SATs for the pre-industrial control scenario were constructed by taking averages by calendar month.

Monthly Southern Oscillation Index (SOI) [25] data was obtained from the Australian Government website. The SOI is based upon the pressure difference between Darwin, Australia and Tahiti, and is strongly correlated with ENSO. This index was compared with the estimated ENSO behaviour of this study.

The versions of the datasets used in this study were their most up-to-date versions at the time of download. Data after December 2018 was not used as HadSST4 data after December 2018 was not available at the time of download. Results of this study are often given as the change from the late 1800s to 2018; the entire late 1800s is used as uncertainties of annual GMST anomalies in the late 1800s are large and 2018 is used as it is the most recent year with data for all calendar months.

## Methods

### Amplification bias

Adopting the notation of Berkeley Earth, the temperature  $T_{\vec{x},t}$  at a position on the Earth's surface  $\vec{x}$  and time in months  $t$  is

$$T_{\vec{x},t} = C_{\vec{x},m_t} + \theta_t + W_{\vec{x},t} \quad (1)$$

where  $m_t$  is the calendar month of  $t$ ,  $C_{\vec{x},m_t}$  is the temperature climatology at  $\vec{x}$  for calendar month  $m_t$ ,  $\theta_t$  is a constant for month  $t$ , and  $W$  is a weather residual term.

For this paper,  $\theta$  can differ from the GMST and  $\vec{x}$  is replaced with  $(\vec{x},s)$ , where  $\vec{x}$  corresponds to the mid latitude and mid longitude of a 5° by 5° grid cell, and  $s$  is the surface type;  $s=0$  for open sea (referred to as sea) and  $s=1$  for land or sea ice (referred to as land-ice). For convenience,  $\vec{x}$  is called a grid cell and  $(\vec{x},s)$  is called a grid subcell. This study adopts the convention of GISTEMP, BEST, and C&W of treating SATs above sea ice similarly to LSATs since SATs above sea ice tend to more closely resemble LSATs than SATs above sea (also called marine air temperatures (MATs)). Thus, (1) becomes

$$T_{\vec{x},s,t} = C_{\vec{x},s,m_t} + \theta_t + W_{\vec{x},s,t} \quad (2)$$

Similar to C&W,  $C$  is taken to be the 1961-1990 temperature climatology. Thus, (2) simplifies to become

$$\Delta T_{\vec{x},s,t} = T_{\vec{x},s,t} - C_{\vec{x},s,m_t} = \theta_t + W_{\vec{x},s,t} \quad (3)$$

where  $\Delta T_{\vec{x},s,t}$  is the local temperature anomaly at  $(\vec{x},s)$  for month  $t$ .

In reality, observations of temperature anomalies are subject to measurement, sampling, and bias uncertainties. Thus, (3) is modified to become

$$Y_{\vec{x},s,t} = \Delta T_{\vec{x},s,t} + E_{\vec{x},s,t} = \theta_t + W_{\vec{x},s,t} + E_{\vec{x},s,t} \quad (4)$$

where  $Y_{\vec{x},s,t}$  is the temperature anomaly observation at  $(\vec{x},s)$  for month  $t$  and  $E_{\vec{x},s,t}$  is the error due to measurement, sampling, and

bias uncertainties at  $(\bar{x}, s)$  for month  $t$ .

Equation (4) implicitly assumes that all regions of the Earth have a statistical tendency to warm at the same rate. However, polar regions tend to warm faster due to the snow-albedo feedback and because global warming increases heat transfer from equatorial regions to polar regions. This increased heat transfer occurs partly because warmer air can hold more water vapour and thus can have a higher heat capacity. In addition, sea regions tend to warm more slowly than land regions, as water has a higher heat capacity than land. As a result, the assumption that all regions have a statistical tendency to warm at the same rate can lead to amplification bias. When amplification bias is ignored, unobserved regions are infilled with temperature anomalies that are biased towards the GMST anomaly. Since polar regions such as Antarctica, the Southern Ocean, and the Arctic Ocean were poorly observed during the instrumental period, not accounting for amplification bias should cause an underestimation of the rate of GMST change as these regions tend to have high local amplification factors (LAFs).

To try to correct for amplification bias, (4) is modified to become

$$Y_{\bar{x},s,t} = A_{\bar{x},s,m} \theta_t + W_{\bar{x},s,t}^- + E_{\bar{x},s,t}^- \quad (5)$$

where  $A_{\bar{x},s,m}$  is called the LAF at  $(\bar{x}, s)$  for calendar month  $m$  and is defined to have a surface area weighted mean of one over the Earth's surface for each calendar month (assuming that SICs are the 1961-1990 HadISST2 mean by calendar month). In reality, the amplification function  $A$  could change from one year to the next. However, modelling studies find that the assumption of unchanging LAFs over the instrumental period is reasonable [26,27] and find that LAFs are mostly independent of the type of climate driver [28]. This suggests that (5) should correct for most amplification bias.

This study tested four different functional forms of the amplification function (Table 1) to determine if LAFs are affected by calendar month, latitude, or surface type. For simplicity, the impact of latitude on  $A$  was treated as a polynomial of order at most four. In addition, the derivative of  $A$  with respect to latitude was restricted to zero at the poles to ensure smoothness at the poles. All four functional forms can be expressed as

$$A_{\bar{x},s,m} = 1 + F_{\bar{x},s,m} f \quad (6)$$

where  $f \in \mathbb{R}^{n_f}$ ,  $n_f \in \mathbb{N}$ ,  $F_{\bar{x},s,m} \in \mathbb{R}^{1 \times n_f}$ , and  $F$  has a surface area weighted mean of zero for each calendar month.

Internal variability, such as from ENSO, may cause issues in estimating  $A$ . Polar regions tend to warm faster than equatorial regions. However, an El Niño event causes significant short term

warming that is concentrated in equatorial regions, particularly in the Eastern Equatorial Pacific Ocean. As a result, if one estimates  $A$  using (5) and (6), then ENSO could bias the estimate of  $A$  towards the identity function.

To try to avoid this ENSO bias and allow for internal variability to be incorporated into the model, (5) is modified to become

$$Y_{\bar{x},s,t} = (1 + F_{\bar{x},s,m}^- f) \theta_t + Q_{\bar{x}} q_t + W_{\bar{x},s,t}^- + E_{\bar{x},s,t}^- \quad (7)$$

where  $q_t = \begin{bmatrix} q_{t,1} \\ \vdots \\ q_{t,n_q} \end{bmatrix} \in \mathbb{R}^{n_q}$ ,  $n_q \in \mathbb{N}$  is the number of internal variability patterns (IVPs) in the model,  $q_{t,i}$  is an index for month  $t$  that corresponds to the  $i^{\text{th}}$  IVP,  $Q_{\bar{x}} = [Q_{\bar{x},1} \cdots Q_{\bar{x},n_q}] \in \mathbb{R}^{1 \times n_q}$ , and  $Q_{\bar{x},i}$  is the value of the  $i^{\text{th}}$  IVP at  $\bar{x}$ . One IVP was obtained (see later subsection on IVPs), corresponding to ENSO, and this study tested models that contain zero or one IVPs. While only one IVP was used, the statistical framework of this study can be easily extended to incorporate more IVPs.

To estimate model parameters in (7) using MLE, a model of  $W$  is needed.  $W$  is assumed to be multivariate normal with a mean of zero and with covariance function

$$\text{cov}(W_{\bar{x}_i,s_i,t_i}, W_{\bar{x}_j,s_j,t_j}) = \lambda(r)^{|t_i-t_j|} \lambda(k_1)^{\|\bar{x}_i-\bar{x}_j\|} \lambda(k_2)^{|s_i-s_j|} \exp\left(-\frac{v_{\bar{x}_i,s_i,m_{t_i}} + v_{\bar{x}_j,s_j,m_{t_j}}}{2} z\right) \quad (8)$$

where  $r, k_1, k_2 \in \mathbb{R}$ ,  $z = \begin{bmatrix} z_1 \\ \vdots \\ z_{n_z} \end{bmatrix} \in \mathbb{R}^{n_z}$ ,  $n_z \in \mathbb{Z}_+$ ,  $v_{\bar{x},s,m} \in \mathbb{R}^{1 \times n_z}$ ,  $\lambda$

is the standard logistic function, and  $\|\bar{x}_i - \bar{x}_j\|$  is the great circle distance between  $\bar{x}_i$  and  $\bar{x}_j$ . In this model,  $r$  determines the temporal autocorrelation of weather residuals,  $k_1$  determines the spatial correlation of weather residuals,  $k_2$  affects the correlation between land-ice and sea weather residuals, and  $z$  determines the variances. The standard logistic function constrains temporal and spatial correlations between zero and one, which ensures physically sensible results and improves numeric stability of the estimation procedure.

Two different models of the covariance function were tested (Table 2) to determine if the variance of weather residuals is affected by surface type. The models of the covariance function are simple compared to the models of the amplification function since additional parameters in the covariance function are more computationally intensive than additional parameters in the amplification function.

GISTEMP, BEST, and C&W infill land-ice and sea regions separately as it was found to give better results. However, the

**Table 1:** The different models of the amplification function. These models were used to estimate temperature anomalies.

Model	Amplification function depends on latitude?	Amplification function depends on surface type?	Amplification function depends on calendar month?
1	No	No	No
2	No	Yes	Yes
3	Polynomial of order four	No	Yes
4	Polynomial of order four	Yes	Yes

**Table 2:** The different models of the covariance function of weather residuals. These models were used to estimate temperature anomalies.

Model	Covariance function depends on latitude?	Covariance function depends on surface type?	Covariance function depends on calendar month?
1	No	No	No
2	No	Yes	No

introduction of the amplification function puts this practice into doubt. Perhaps this approach gave better results due to differences in LAFs between land-ice and sea regions. The statistical models tested in this study account for correlations between land-ice and sea weather residuals, which can improve temperature anomaly estimates.

Alternatively, perhaps infilling land-ice and sea regions separately gave better results due to differences in their correlation decay lengths. This paper defines the correlation decay length as the distance between two locations required to reduce the spatial correlation of their residuals to  $e^{-1}$ . However, C&W estimate similar correlation decay lengths for weather residuals: 767 km and 860 km for land-ice and sea regions respectively. Similarly, BEST has similar maximum correlation lengths: 3310 km and 2680 km for land-ice and sea regions respectively, corresponding to correlation decay lengths of 1497 km and 1212 km respectively. Having different correlation decay lengths for land-ice and sea regions was considered, but this approach was not taken due to its high computational cost. As C&W and BEST estimate similar correlation decay lengths for land-ice and sea regions, the impact of using identical correlation decay lengths for land-ice and sea regions should be minimal.

The four models of the amplification function were combined with the two models of the covariance function and using zero or one IVPs to produce 16 different temperature anomaly models. To estimate a temperature anomaly model, an iterative approach to MLE was taken, where a reasonable first guess was calculated and used in combination with a saddle-free Newton's method [29,30]. To make the procedure computationally feasible, numeric approximations to the log-likelihood function were used and their derivations are given in the supplementary information. As the bias uncertainties of  $E$  have a complicated covariance structure, do not have a multivariate normal distribution, and have significant temporal correlation, numeric approximations that account for bias uncertainties could not be derived. As a result, bias uncertainties were neglected in the estimation of the temperature anomaly model. However, bias uncertainties were still taken into account in the estimation of total uncertainty.

The best of the 16 models was chosen by minimizing the AIC. To reduce computation time, only six of the 16 models were estimated and the best temperature anomaly model was chosen in a stepwise manner. Firstly, the best covariance function was chosen by estimating the two covariance functions with the first amplification function and zero IVPs and picking the model with the lowest AIC. Secondly, the best amplification function was chosen by estimating the best covariance function with the other three amplification functions and zero IVPs, and picking the model with the lowest AIC. Finally, the best covariance function was estimated with the best amplification function and the IVP for ENSO. For comparison, Bayesian information criterion (BIC) [31] values were also calculated. This paper presents normalized AIC and BIC values (nAIC and nBIC), which are normalized by the number of observations.

For each of the six models, maximum likelihood estimates of temperature anomalies were calculated for all combinations of grid subcells and months. Land-ice and sea temperature anomalies were then blended, by weighting observations using SICs of the 1961-1990 HadISST2 mean by calendar month. For the model with the lowest AIC, its uncertainty was evaluated using a Monte Carlo approach. The Monte Carlo approach produced 200 ensemble members of temperature anomalies, which account for measurement, sampling, model parameter, infilling, and bias uncertainties.

## Sea Ice Bias

When sea ice is replaced with open sea, an increase in SATs is observed empirically [22] and in climate models [32,33]. This is because sea water has a lower albedo than sea ice and because sea ice acts as an insulator that prevents heat transfer between colder air and warmer water. As a result, not accounting for changes in sea ice can lead to sea ice bias. In particular, as warming has caused sea ice to melt, neglecting sea ice bias leads to an underestimation of GMST change over the instrumental period. C&W assumes constant SICs for each calendar month and uses the 1981-2010 HadISST1 median by calendar month. As a result, C&W may underestimate GMST change over the instrumental period. In comparison, BEST varies SICs according to HadISST2 and, partially as a result, shows more warming over the instrumental period than C&W.

To correct for sea ice bias, a temperature climatology model was combined with the best temperature anomaly model to estimate temperature differences between sea ice and open sea. The estimated temperature differences were used to estimate and correct for sea ice bias. Since HadSST4 and CRUTEM4 are temperature anomaly datasets, suitable temperature climatology datasets that correspond to these temperature anomaly datasets should be used to estimate a temperature climatology model. Since a temperature climatology for HadSST4 was not available, the HadSST3 climatology was used as the temperature climatology corresponding to HadSST4.

CRUTEM4 does not have a corresponding LSAT climatology. However, the Met Office website lists the 1961-1990 temperature climatology of Jones et al. along with its main temperature anomaly datasets of HadCRUT4, HadSST3, and CRUTEM4. The Jones et al. climatology was constructed using CRUCL1 [34] for land regions north of 60°S, IABP-POLES for sea ice regions north of 60°N, a combination of MAT and SST observations for sea regions north of 60°S, and LSAT observations of Antarctica for land regions south of 60°S.

Climatologies corresponding to CRUTEM4 were inferred from the Jones et al. climatology. The following model was used to explain variation in the temperature climatology

$$C_{\bar{x},m} = X_{\bar{x},m}B + U_{\bar{x},m} \quad (9)$$

where  $C_{\bar{x},m}$  is the temperature climatology at  $\bar{x}$  for calendar month  $m$ ,  $B \in \mathbb{R}^{n_B}$ ,  $n_B \in \mathbb{N}$ ,  $X_{\bar{x},m} \in \mathbb{R}^{1 \times n_B}$  contains explanatory variables at  $\bar{x}$  for calendar month  $m$ , and  $U$  is a term for the temperature climatology residual. In particular,  $X_{\bar{x},m}B$  is given the functional form

$$X_{\bar{x},m}B = e_{\bar{x}}B_1 + \overline{sea}_{\bar{x},m}X_{\bar{x},m}^{sea}B + \overline{land}_{\bar{x}}X_{\bar{x},m}^{land}B + \overline{ice}_{\bar{x},m}X_{\bar{x},m}^{ice}B \quad (10)$$

where  $e_{\bar{x}}$  is the average surface elevation at  $\bar{x}$ ,  $\overline{land}_{\bar{x}}$  is the fraction of grid cell  $\bar{x}$  covered by land,  $\overline{ice}_{\bar{x},m}$  is the SIF of grid cell  $\bar{x}$  for calendar month  $m$  using SICs of the 1961-1990 HadISST2 mean by calendar month,  $\overline{sea}_{\bar{x},m} = 1 - \overline{land}_{\bar{x}} - \overline{ice}_{\bar{x},m}$ , and  $X^{sea}$ ,  $X^{land}$  and  $X^{ice}$  are defined implicitly.

$U$  is assumed to be multivariate normal with a mean of zero and with covariance function

$$\text{cov}(U_{\bar{x}_i,m_i}, U_{\bar{x}_j,m_j}) = \lambda(r)^{1-\delta_{m_i,m_j}} \lambda(k)^{\|\bar{x}_i - \bar{x}_j\|} \exp\left(-\frac{v_{\bar{x}_i,m_i} + v_{\bar{x}_j,m_j}}{2} z\right) \quad (11)$$



where  $r, k \in \mathbb{R}$ ,  $z \in \mathbb{R}^{n_z}$ ,  $n_z \in \mathbb{Z}_+$ ,  $v_{\bar{x},m} \in \mathbb{R}^{1 \times n_z}$ , and  $\delta$  is the Kronecker delta. In this model,  $r$  determines the temporal correlation of residuals,  $k$  determines the spatial correlation of residuals, and  $z$  determines the variances. Variables in (8) are repeated in (11) for ease of communication; however, temperature climatology models and temperature anomaly models were estimated independently.

Six different models of the climatology function (Table 3) and four different models of the covariance function (Table 4) were combined to produce a total of 24 models. The different models of the climatology function determine how the temperature climatology varies by surface elevation, latitude, calendar month, and surface type. The different models of the covariance function determine how variance varies by latitude and surface type. The derivatives of the climatology function and the covariance function with respect to latitude were restricted to zero at the poles. Model parameters were estimated using MLE.

For some of the climatology functions, the temperature climatology can differ between sea ice and land regions for the northern hemisphere. This allows for the possibility that temperatures of land and sea ice regions may behave differently. However, this distinction is not made for the southern hemisphere since, unlike in the northern hemisphere, Jones et al. did not use direct observations of temperatures of sea ice regions in the southern hemisphere. Instead, Jones et al. extended nearby LSAT estimates to sea ice regions in the southern hemisphere.

Jones et al. linearly interpolated between sea temperatures at 60°S and estimated temperatures at the sea ice edge to estimate sea temperatures south of 60°S. However, Jones et al. noted that their temperature estimates for sea ice regions might be unphysical as “the field is likely to exhibit a sharp increase in air temperature at the margin of the continental ice sheet.” To prevent sea temperatures south of 60°S from biasing estimates of sea ice bias, the Jones et al. climatology was considered missing for grid cells south of 60°S that have less 90% of their area covered by land or sea ice for any calendar month.

To reduce computation time, only 9 of the 24 models were estimated and the best temperature climatology model was chosen in a stepwise manner. Firstly, the best covariance function was chosen by estimating the four covariance functions with the first climatology

function and picking the model with the lowest AIC. Next, the best climatology function was chosen by estimating the best covariance function with the five other climatology functions and picking the model with the lowest AIC.

The temperature climatology of sea regions  $C^{sea}$ , the temperature climatology of land regions  $C^{land}$ , and the temperature climatology of sea ice regions  $C^{ice}$  were estimated as

$$C_{\bar{x},m}^{sea} = \mu_{\bar{x},m} \left( e_{\bar{x}}^{not\ land} B_1 + X_{\bar{x},m}^{sea} B + U_{\bar{x},m} \right) \quad (12)$$

$$C_{\bar{x},m}^{land} = e_{\bar{x}}^{land} B_1 + X_{\bar{x},m}^{land} B + U_{\bar{x},m} \quad (13)$$

$$C_{\bar{x},m}^{ice} = e_{\bar{x}}^{not\ land} B_1 + X_{\bar{x},m}^{ice} B + U_{\bar{x},m} \quad (14)$$

where  $\mu_{\bar{x},m}$  is a function that returns the HadSST3 climatology at  $\bar{x}$  for calendar month  $m$  if it is available, but is otherwise the identity function,  $e_{\bar{x}}^{land}$  is the average land elevation at  $\bar{x}$  and  $e_{\bar{x}}^{not\ land}$  is the average surface elevation of non-land regions at  $\bar{x}$ .  $C^{sea}$  mostly reflects the HadSST3 climatology, except for a few grid cells where the HadSST3 climatology is not available.

Sea ice bias can be estimated as the change in SIF multiplied by the difference between sea temperatures and sea ice temperatures. For this study, sea ice bias was estimated as

$$SIB_{\bar{x},t} = \left( ice_{\bar{x},t} - \overline{ice_{\bar{x},m_t}} \right) \left( \frac{\left( C_{\bar{x},m_t}^{sea} + \Delta T_{\bar{x},0,t} \right)}{\left( C_{\bar{x},m_t}^{ice} + \Delta T_{\bar{x},1,t} \right)} \right) \quad (15)$$

where  $SIB_{\bar{x},t}$  is the sea ice bias at  $\bar{x}$  for month  $t$ ,  $ice_{\bar{x},t}$  is the SIF at  $\bar{x}$  for month  $t$ . Sea ice bias-corrected temperatures of grid cells were estimated as

$$T_{\bar{x},t} = + sea_{\bar{x},t} \left( C_{\bar{x},m_t}^{sea} + \Delta T_{\bar{x},0,t} \right) + land_{\bar{x}} \left( C_{\bar{x},m_t}^{land} + \Delta T_{\bar{x},1,t} \right) + ice_{\bar{x},t} \left( C_{\bar{x},m_t}^{ice} + \Delta T_{\bar{x},1,t} \right) \quad (16)$$

Temperatures inferred using (15) may be physically unrealistic, such as inferred sea temperatures below the freezing temperature of sea water or inferred sea ice temperatures above the freezing temperature of fresh water. Therefore, one may want to use the following to estimate sea ice bias while preventing unphysical temperatures.

$$SIB_{\bar{x},t} = \left( ice_{\bar{x},t} - \overline{ice_{\bar{x},m_t}} \right) \left( \frac{\max \left( C_{\bar{x},m_t}^{sea} + \Delta T_{\bar{x},0,t}, -1.8^\circ C \right)}{-\min \left( C_{\bar{x},m_t}^{ice} + \Delta T_{\bar{x},1,t}, 0^\circ C \right)} \right) \quad (17)$$

**Table 3:** The different models of the temperature climatology function. These models were used to quantify the sea ice bias of temperature anomaly estimates.

Model	Climatology function depends on latitude?	Climatology function depends on surface type?	Special treatment for northern hemisphere sea ice?	Climatology function depends on calendar month?
1	Polynomial of order four	No	No	Yes
2	Polynomial of order four	Yes	No	Yes
3	Polynomial of order four	Yes	Yes	Yes
4	Polynomial of order six	No	No	Yes
5	Polynomial of order six	Yes	No	Yes
6	Polynomial of order six	Yes	Yes	Yes

**Table 4:** The different models of the covariance function of temperature climatology residuals. These models were used to quantify the sea ice bias of temperature anomaly estimates.

Model	Covariance function depends on latitude?	Covariance function depends on surface type?	Covariance function depends on calendar month?
1	No	No	No
2	No	Yes	No
3	Polynomial of order four	No	No
4	Polynomial of order four	Yes	No

Bias-corrected temperatures that are physically realistic could then be estimated as

$$T_{\bar{x},t} = \begin{aligned} & sea_{\bar{x},t} \max \left( C_{\bar{x},m_t}^{sea} + \Delta T_{\bar{x},0,t}, -1.8^\circ\text{C} \right) \\ & + land_{\bar{x}} \left( C_{\bar{x},m_t}^{land} + \Delta T_{\bar{x},1,t} \right) \\ & + ice_{\bar{x},t} \min \left( C_{\bar{x},m_t}^{ice} + \Delta T_{\bar{x},1,t}, 0^\circ\text{C} \right) \end{aligned} \quad (18)$$

The maximum function restricts temperatures of sea water to at least  $-1.8^\circ\text{C}$ , which is the freezing temperature of sea water with a salinity of 35 parts per thousand, and the minimum function restricts temperatures of sea ice to at most  $0^\circ\text{C}$ , which is the freezing temperature of fresh water. While SATs above sea and sea ice can exceed these limits, they cannot significantly exceed these limits for monthly temperature averages. Restricting sea temperatures to at least  $-1.8^\circ\text{C}$  is standard practice and is performed in HadISST1, KOBE-SST1 [35], ERSSTv5 [4], and BEST. Also, sea ice regions tend to have isothermal summers with temperatures close to  $0^\circ\text{C}$  [22].

While sea ice bias and temperature estimates using (17) and (18) would be physically realistic, they might be statistically biased. Estimating (9) and (11) using MLE would correspond to  $B$  being estimated using generalized least squares (GLS) given the estimate of the covariance function. GLS is unbiased if the covariance function is known; therefore, estimates of (15) and (16) may be relatively unbiased. Applying minimum and maximum functions to GLS estimates could result in biased estimates; in particular, the magnitude of sea ice bias could be overestimated. In addition, it is unclear if the use of freezing temperatures in (17) and (18) is appropriate as the temperature climatology datasets used have measurement error and do not correspond perfectly with the temperature anomaly datasets. Therefore, the results of (15) and (16) are presented as the main estimates of this study and the results of (17) and (18) are presented as alternate estimates.

Changes in SICs can cause changes in local temperatures. Changes in local temperatures can cause changes in SICs. In addition, changes in SICs can cause temperature changes in nearby regions due to heat transfer. The sea ice bias estimated using the above methodology may capture all three of these mechanisms and does not necessarily reflect the impact of SICs on local temperatures alone. This lack of distinction of these mechanisms in the statistical model can make physically interpretability of results difficult. However, for the purpose of statistically inferring temperatures using available data, it is not necessary to identify how much of the correlation between SICs and local temperatures is caused by each of these mechanisms since what matters is the existence of the correlation.

If SICs are correlated with the temperatures of nearby regions and if the statistical model used assumes a temperature field that is discontinuous between sea and sea ice regions, then this could cause an overcorrection for sea ice bias. This is because the temperature anomaly observations of nearby land or sea regions may already reflect the higher regional warming due to melting sea ice. If the temperature climatology function is estimated from blended data and a coarse resolution is used, then this may cause the temperature climatology model to overestimate the impact of SICs on local temperatures, resulting in a further overcorrection for sea ice bias. To test for this potential issue, IABP-POLES data north of  $60^\circ\text{N}$  was used. Using ordinary least squares, the temperature climatology function was re-estimated twice: once using a  $5^\circ$  resolution of IABP-POLES and once using a  $1^\circ$  resolution of IABP-POLES. For simplicity, these tests neglect differences in  $\Delta T$  between land-ice and sea regions in

(15). The impact of sea ice bias on GMST change due to changes in SICs north of  $60^\circ\text{N}$  was estimated for each of these resolutions. To be consistent with Jones et al., IABP-POLES was treated as a 1961-1990 climatology.

As an additional test, empirical LAFs of regions of sea ice loss were calculated for HadCRU\_MLE and for CCSM4 under the RCP6.0 scenario using a  $5^\circ$  resolution. For each calendar month, weighted averages of temperature change from the late 1800s to 2018, weighted by the surface area of net sea ice loss for each grid cell, were calculated. These weighted temperature changes were averaged over the calendar year and divided by the annual GMST change to produce empirical LAFs of regions of sea ice loss. For simplicity, leap days were neglected in this calculation. While these empirical LAFs may underestimate the true LAF of regions of sea ice loss due to using a coarse resolution of  $5^\circ$ , they provide useful metrics for comparison.

For each of the temperature climatology models, its uncertainty was evaluated using a Monte Carlo approach. The Monte Carlo approach produced 200 ensemble members of temperature climatologies. To generate these ensemble members, each of the 100 ensemble members of the HadSST3 climatology was used twice. The HadSST3 ensemble accounts for bias uncertainties, but not other uncertainties such as measurement or sampling uncertainties. 200 ensemble members of sea ice bias and 200 ensemble members of blended temperatures were then calculated. While this Monte Carlo approach accounts for parameter estimation uncertainty and infilling uncertainty, bias and uncertainty due to a possible methodological overcorrection for sea ice bias were neglected. In addition, measurement uncertainties of temperature climatologies and uncertainties of SICs were neglected since they were not quantified. Uncertainties of SICs may be large as SICs vary greatly between datasets; the change in average global sea ice extent between the late 1800s and 2018 is  $-0.7\%$  of the Earth's surface according to HadISST1, which is approximately half of the  $-1.3\%$  change according to HadISST2.

## Patterns of Internal Variability

Temperature patterns that correspond to modes of internal variability are needed to use (7) to obtain amplification bias-corrected estimates of past temperatures. One approach to obtain IVPs is to use the method of empirical orthogonal teleconnections (EOTs) [36]. This is used extensively by NOAA in ERSSTv5, where 140 EOTs are used to reconstruct SSTs. NOAA calculates its EOTs using OISSTv2 [37] data for the period 1982-2011. OISSTv2 combines satellite data with SST data and is spatially complete. A second approach is to use empirical orthogonal functions (EOFs) [38]. Methodologies where one first obtains internal variability indices from observed temperatures and then looks for IVPs that are orthogonal to these indices have also been used [24,39,40].

The approach of NOAA of using OISSTv2 data to calculate IVPs has three problems. Firstly, as global warming has occurred over the instrumental period, estimates of these IVPs might contain warming signals. Secondly, estimates of these IVPs are not independent of instrumental temperature data, thus proper error analysis using these IVPs becomes more difficult. Thirdly, estimates of these IVPs include weather residuals for the period 1982-2011. Thus, the use of such estimated IVPs in (7) could result in zero-biased estimates of internal variability indices due to attenuation bias [41]. Even worse, because of how such IVPs are calculated, estimates of internal variability indices could contain more attenuation bias prior to 1982 than after 1982, which might affect estimates of changes in ENSO variability over

time. The use of climate model output for scenarios with constant forcing over time can avoid these three problems. This approach has been performed using CMIP5 output for pre-industrial control scenarios [24,39,40].

Warmer temperatures due to internal variability can cause sea ice to melt, which could cause high localized warming in areas where sea ice melts. However, the correlation between SICs and temperature is included in the correction for sea ice bias. If IVPs include temperature changes explained by SIC changes, then this could result in double counting. Therefore, it is desirable to remove temperature changes explained by SIC changes from temperature anomalies before obtaining IVPs.

To remove temperature changes explained by SIC changes, the results of the best climatology model estimated using Jones et al. data were used. In particular, sea ice detrended temperature anomalies for CCSM4 under the pre-industrial control scenario were estimated as

$$\widehat{\Delta T}_{\bar{x},t} = \Delta T_{\bar{x},t} + \left( ice_{\bar{x},t} - \overline{ice}_{\bar{x},m_t} \right) \left( C_{\bar{x},m_t}^{sea} - C_{\bar{x},m_t}^{ice} \right) \tag{19}$$

where  $\Delta T_{\bar{x},t}$  is the CCSM4 temperature anomaly at  $\bar{x}$  for month  $t$ , and  $\widehat{\Delta T}_{\bar{x},t}$  is the detrended temperature anomaly at  $\bar{x}$  for month  $t$ . SICs of CCSM4 were used for  $ice$  and  $\overline{ice}$ , whereas the ensemble of the best temperature climatology model estimated using Jones et al. data was used for  $C^{sea}$  and  $C^{ice}$ .

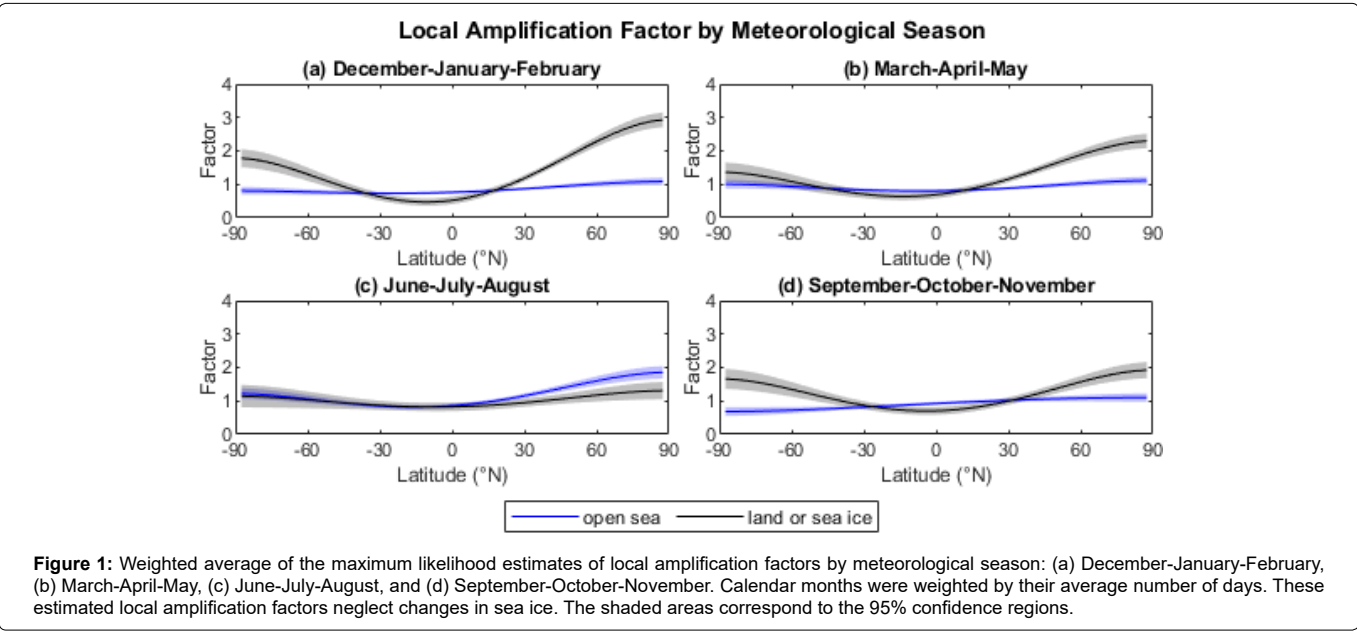
EOFs were calculated from annual averages of the ensemble median of detrended CCSM4 temperature anomalies using the approach given in the supplementary information. Annual averages were used to remove temperature variation on a timescale of less than a year. Throughout this study, annual averages were calculated by weighted each month by its number of days while accounting for leap years. An EOF approach was used instead of an EOT approach since the EOT approach would involve the selection of a set of reference grid cells, which can be somewhat arbitrary.

Results and Discussion

Table 5 summarizes the results of the estimated temperature anomaly models. For the best (lowest AIC) model with a constant amplification function (amplification function 1), the maximum likelihood estimate of GMST change from the late 1800s to 2018 is 1.11°C, which is similar to the 1.11°C estimate of GMST change of C&W. When the best amplification function is introduced into the model, the estimate of GMST change increases by 0.01°C to 1.12°C. This is expected as polar regions, which tend to be poorly observed, have high LAFs. Introducing the IVP for ENSO leaves the estimate of GMST change relatively unchanged at 1.12°C. For the best temperature anomaly model, the maximum likelihood estimate of GMST change is slightly greater than the ensemble median estimate of GMST change: 1.11°C, with a 95% confidence interval of (1.04°C, 1.20°C).

Table 5: Summary of results of the estimated temperature anomaly models. Estimates listed correspond to maximum likelihood estimates. The global mean surface temperature (GMST) change values listed in this table do not correct for sea ice bias and include the impact of the internal variability pattern (IVP) if applicable.

Amplification function	Covariance function	Number of IVPs	nAIC	nBIC	GMST change from the late 1800s to 2018 (°C)	Correlation of residuals between consecutive months	Correlation decay length (km)	Correlation of residuals between land-ice and sea regions
1	1	0	2.279	2.288	1.103	0.08	970	0.36
1	2	0	2.194	2.203	1.107	0.08	975	0.53
2	2	0	2.194	2.203	1.108	0.08	974	0.53
3	2	0	2.194	2.203	1.106	0.08	974	0.53
4	2	0	2.194	2.203	1.116	0.08	969	0.53
4	2	1	2.190	2.207	1.117	0.07	945	0.52



**Figure 1** shows averages of the estimated amplification functions for each meteorological season; for simplicity, amplification functions are shown by meteorological season instead of by calendar month. As expected, polar regions tend to have higher estimated LAFs than equatorial regions and land-ice regions tend to have higher estimated LAFs than sea regions. The northern hemisphere tends to have higher estimated LAFs than the southern hemisphere.

The estimated LAFs are often far from unity, with Arctic land regions having LAFs of two or greater. This suggests that accounting for LAFs is important for temperature estimates of unobserved regions. Of the GITDs discussed in the introduction, none of their statistical models adequately account for LAFs, so all should be subject to amplification bias. GISTEMPv4 and NOAAGlobalTempv5 use ERSSTv5 for SSTs; ERSSTv5 might indirectly account for some amplification bias through its use of EOTs. In addition, GISTEMPv4 averages temperature anomalies zonally before averaging globally, which may cause GISTEMPv4 to have reduced amplification bias compared to other GITDs.

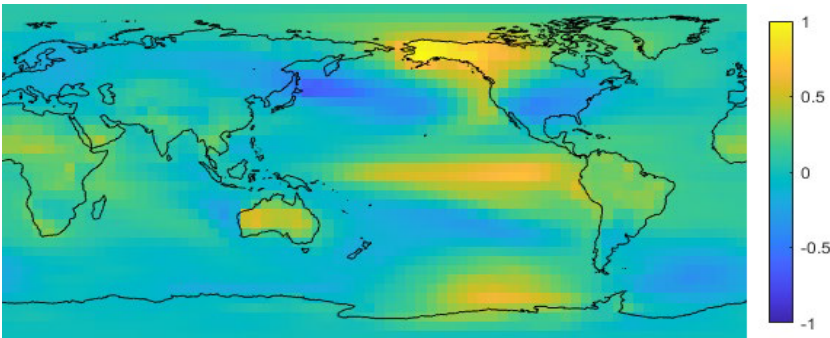
**Figure 2** shows the first EOF of the CCSM4 detrended temperature anomalies under the pre-industrial control scenario. This EOF appears to correspond to the temperature pattern of ENSO due to its strong warming band in the Eastern Equatorial Pacific. If temperature variation explained by SICs is excluded, this EOF explains 18% of the variation in local annual mean SATs and 51% of the variation

in annual GMSTs for the CCSM4 output under the pre-industrial control scenario.

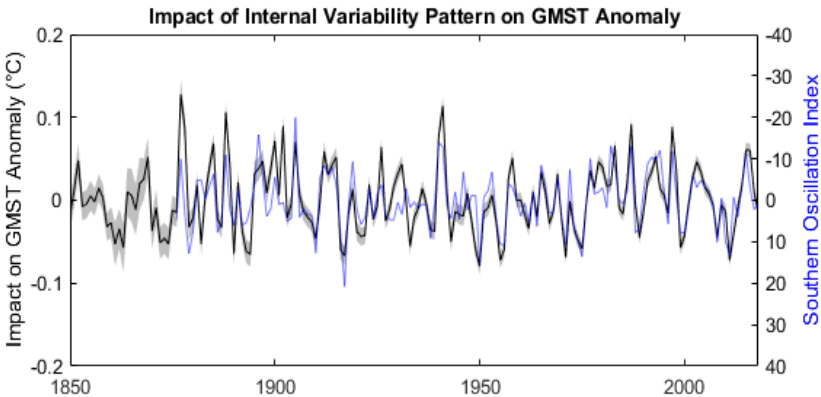
**Figure 3** shows the estimated impact of the IVP on annual GMST anomalies. It appears that this IVP causes fluctuations in annual GMSTs with an amplitude of approximately 0.1°C. The GMST variability explained by this IVP corresponds well with the SOI, suggesting that this mode of internal variability corresponds well to ENSO.

**Table 6** summarizes the results of the estimated temperature climatology models. For the model with the lowest AIC, the median estimate of the impact of sea ice bias on GMST change from the late 1800s to 2018 is -0.08°C, with a 95% confidence interval of (-0.04°C, -0.13°C). The maximum likelihood estimate of the lapse rate of the best model is 6.4°C km<sup>-1</sup>, corresponding to a typical wet lapse rate for Earth; the retrieval of a physically sensible lapse rate adds confidence to the results of the temperature climatology model. When the freezing temperatures of sea water and fresh water are applied, the median estimate remains relatively unchanged at -0.09°C. Thus, while the use of the freezing temperatures has an impact on the estimated bias, it does not change the order of magnitude of estimated bias.

To test for a possible overestimation of sea ice bias due to using a temperature field that is discontinuous between sea and sea ice regions, the climatology function was re-estimated twice using IABP-POLES. When the resolution is 5° and when only accounting for changes in



**Figure 2:** The dominant temperature pattern of internal variability obtained from CCSM4 output under the pre-industrial control scenario. The strong warming band in the Eastern Equatorial Pacific suggests that this pattern corresponds to the El Niño Southern Oscillation.

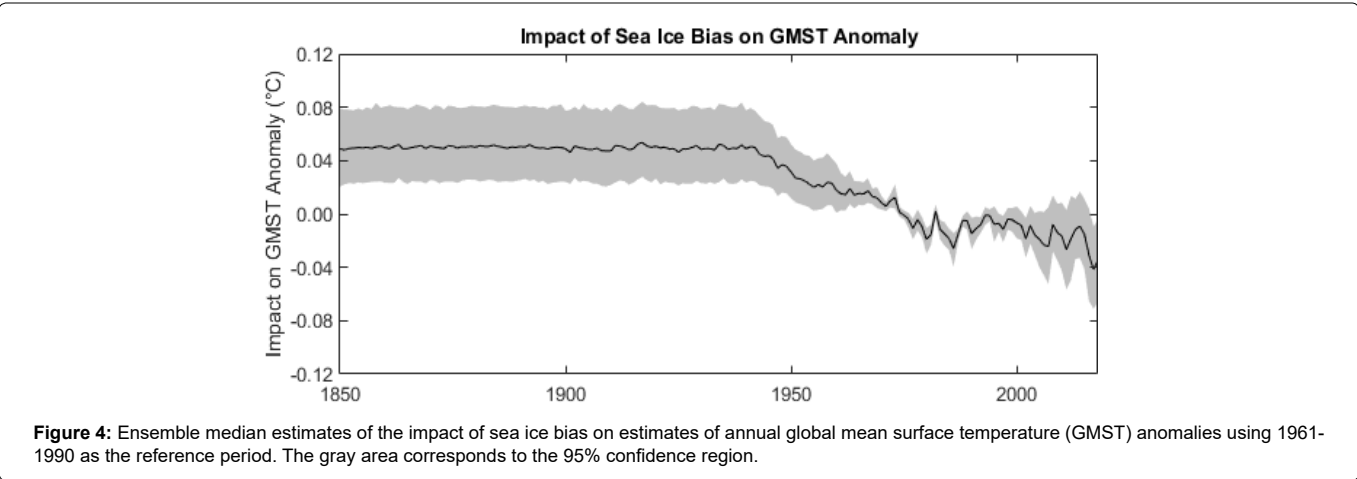


**Figure 3:** Maximum likelihood estimate of the impact of the internal variability pattern on annual global mean surface temperature (GMST) anomalies using 1961-1990 as the reference period. This estimate neglects the indirect impact on the GMST due to how internal variability may affect sea ice concentrations. The internal variability pattern corresponds well with the negative of the Southern Oscillation Index, suggesting that this mode of internal variability corresponds to the El Niño Southern Oscillation. The gray area corresponds to the 95% confidence region.



**Table 6:** Summary of results of the estimated temperature climatology models. The best (lowest nAIC) temperature anomaly model of Table 5 was used in the estimation of the impact of sea ice bias on global mean surface temperature (GMST) change. Sea ice bias estimates listed correspond to ensemble median estimates; other estimates listed correspond to maximum likelihood estimates.

Climatology function	Covariance function	nAIC	nBIC	Impact of sea ice bias on the GMST change from the late 1800s to 2018 (°C)		Correlation of residuals between calendar months	Correlation decay length (km)	Lapse Rate (°C km <sup>-1</sup> )
				Neglecting freezing temperatures	Using freezing temperatures			
1	1	2.21	2.22	-0.061	-0.070	0.67	2481	5.86
1	2	1.74	1.76	-0.057	-0.063	0.70	2879	6.10
1	3	2.03	2.04	-0.057	-0.070	0.78	2506	6.36
1	4	1.56	1.58	-0.056	-0.063	0.75	2906	6.47
2	4	1.54	1.56	-0.087	-0.097	0.74	2334	6.46
3	4	1.50	1.54	-0.080	-0.089	0.74	2315	6.41
4	4	1.55	1.57	-0.057	-0.064	0.74	2924	6.44
5	4	1.51	1.54	-0.084	-0.094	0.75	2210	6.45
6	4	1.47	1.53	-0.085	-0.094	0.75	2228	6.40



sea ice north of 60°N, the estimate of the impact of sea ice bias on GMST change between the late 1800s and 2018 is -0.045°C. When the resolution is changed to 1°, the estimate is reduced in magnitude to -0.030°C. Since the use of a coarser resolution should result in a greater overestimation of sea ice bias, this suggests that there could be a 50% overestimation of sea ice bias. In addition, some overestimation may remain as a 1° resolution is still somewhat coarse.

As an additional test, empirical LAFs of regions of sea ice loss from the late 1800s to 2018 were calculated. Combining the best temperature anomaly model with the best temperature climatology model constructs HadCRU\_MLE. According to HadCRU\_MLE, the median estimate of the empirical LAF is 1.5 if no sea ice bias correction is included and 4.6 if the sea ice bias correction is included. In comparison, the empirical LAF is 2.7 for CCSM4 under the RCP6.0 scenario using a 5° resolution. This suggests that while a sea ice bias correction is needed, HadCRU\_MLE may overcorrect for sea ice bias by a factor of two.

Both local observations of temperatures in sea ice regions as well as results of climate models can help quantify the magnitude of sea ice bias. Rayner et al. [19] use local temperature observations to estimate the relationship between temperature and SIC. Using their figure 7, I estimate by visual inspection that the difference in temperature in going from 0% SIC to 100% SIC is likely between 2°C and 6°C. Multiplying this by the change in average global sea ice extent between the late 1800s and 2018 suggests a sea ice bias of between -0.03°C

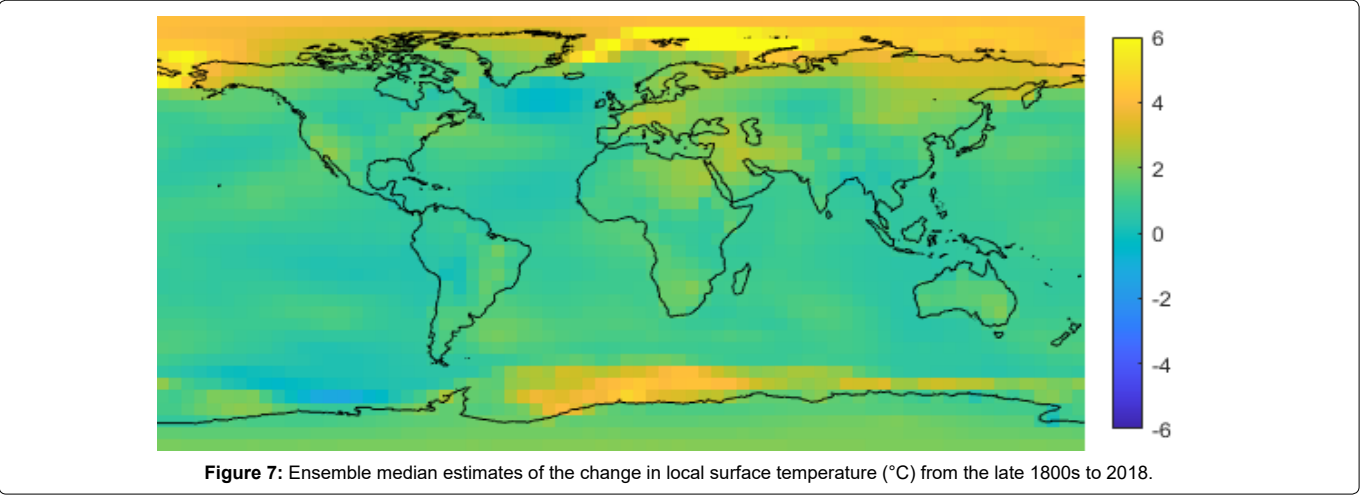
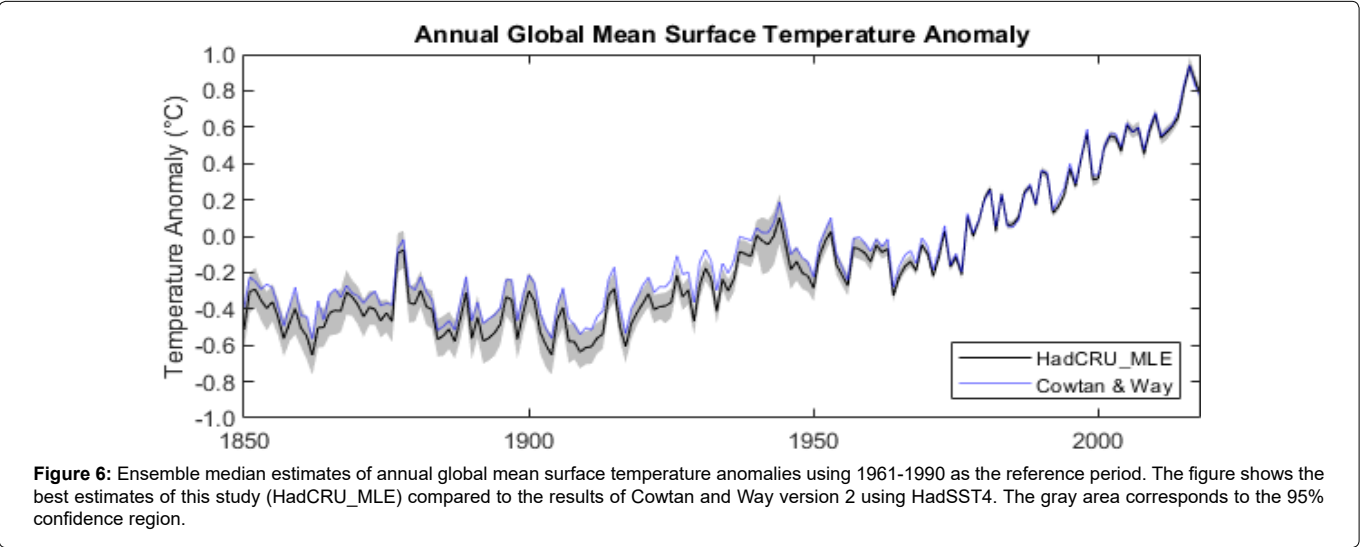
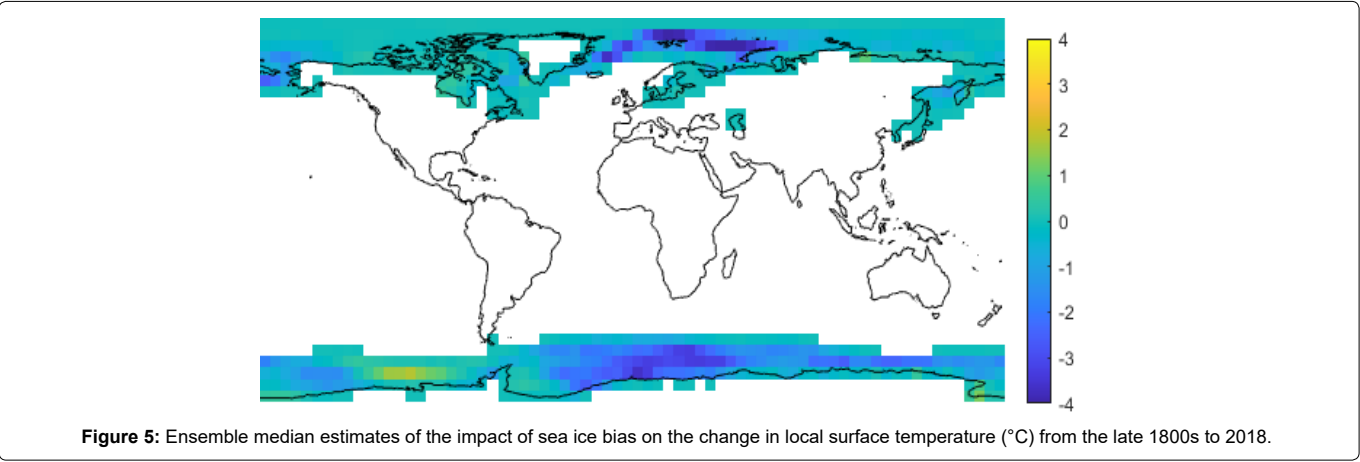
and -0.08°C. Alternatively, Cowtan et al. [32] find, using CMIP5 output, that blending land and sea absolute temperatures instead of blending land and sea temperature anomalies increases the estimated GMST change by 3%. Since the GMST change over the instrumental period is approximately 1°C, this suggests that sea ice bias may cause an underestimation of GMST change of 0.03°C. Thus, both local observations of temperatures in sea ice regions and results of climate models confirm the order of magnitude of sea ice bias estimated in this study.

Figure 4 shows the estimated impact of sea ice bias on annual GMST anomalies. Figure 5 shows the distribution of the estimated impact of sea ice bias on surface temperature change from the late 1800s to 2018. Regions that have experienced significant ice melt over the instrumental period, such as the Norwegian Sea and parts of the Southern Ocean, show the most bias.

It should be noted that HadISST2 is imperfect. In the northern hemisphere, it has no interannual change in SICs prior to 1901 and little interannual change in SICs prior to 1953. In the southern hemisphere, it has no interannual change in SICs prior to 1939 and little interannual change in SICs prior to 1973. This is caused by a lack of data and infilling missing data with SIC climatologies estimated using data from other years. As a result, HadISST2 underestimates SIC change prior to 1973. Thus, the results of this study likely underestimate variation in sea ice bias prior to 1973.

The results of HadCRU\_MLE give this study's best estimate of GMST change from the late 1800s to 2018: 1.20°C with a 95% confidence interval of (1.11°C, 1.30°C). As an alternate estimate, if the freezing temperatures of sea water and fresh water are applied, then the median estimate of GMST change remains relatively unchanged at 1.20°C.

Figure 6 shows this study's best estimates of annual GMST anomalies. Figure 7 shows the distribution of local surface temperature change from the late 1800s to 2018. HadCRU\_MLE estimates more warming than the 1.16°C of GMST change from the late 1800s to 2018 of BEST. This may be because BEST uses HadSST3, which shows less



warming over the instrumental period than HadSST4. The estimate of GMST change increases from 1.03°C in C&W-HadSST3 to 1.11°C in C&W, suggesting that the estimate of GMST change of BEST would increase by a comparable amount if it is updated to use HadSST4. As BEST accounts for changes in SICs, the larger estimate of GMST change of BEST compared to C&W is partially explained by the sea ice bias of C&W; the remaining difference could be due to different homogenization procedures of LSATs or differences in statistical infilling methods.

**Figure 7** shows a large contrast in temperature change in the Southern Ocean. This contrast is likely exaggerated since HadISST2 neglects interannual variation in SICs for the late 1800s. The large change in SICs for the Southern Ocean is largely the result of a single German 1929-1939 Antarctic sea ice climatology. Thus, the reliability of this study's estimate of temperature change since the late 1800s for the Southern Ocean is largely dependent on this single 1929-1939 temperature climatology.

HadCRU\_MLE may have a remaining bias due to using SSTs instead of MATs. According to the results of climate models, this bias may cause an underestimation of GMST change of 5% to 9% [32,33]. However, the top ocean layer of climate models is typically 10 m deep, whereas the SSTs of HadSST4 use a buoy reference depth 20 cm below the sea surface; so, the actual bias of HadCRU\_MLE should be smaller than the bias suggested by climate models. In addition, HadSST4 assumes that MATs and SSTs warm at similar rates and uses MATs to correct for SST biases. In particular, MATs are used to determine the fractions of canvas and wooden buckets in the early part of the instrumental period. So, it is unclear if there is any significant bias due to differences in warming rates of SSTs and MATs when using HadSST4.

While this paper outlines some improvements to the statistical methodology of estimating past temperatures, there are still areas where further improvements could be made. Firstly, uncertainties in the estimates of the temperature climatology and SICs could be quantified and taken into account. Secondly, the magnitude of sea ice bias could be better quantified and verified using different lines of evidence. Thirdly, attempts to better incorporate bias errors of instrumental observations into the statistical model could be made. Fourthly, amplification and temperature climatology functions of this study are discontinuous between land-ice and sea regions, which is physically unrealistic; further improvements to the functional forms could be made. Lastly, other sources of observations, such as satellite and paleoclimate data, could be used to further improve temperature estimates.

## Conclusions

This paper identifies two biases in GITDs: amplification bias and sea ice bias. Amplification bias occurs when the underlying statistical model of the GITD neglects the tendency of different regions of the planet to warm at different rates. Since polar regions, which tend to warm the fastest, were poorly observed during the instrumental period, not accounting for amplification bias causes an underestimation of GMST change over the instrumental period. Sea ice bias occurs when the GITD neglects the impact of changes in sea ice on temperatures. Since SICs have decreased over the instrumental period, neglecting changes in SICs causes an underestimation of GMST change over the instrumental period. To quantify the impact of these biases, a new bias-corrected GITD was constructed, called HadCRU\_MLE, which used MLE to combine the LSAT anomalies of HadCRUT4 with the SST anomalies of HadSST4.

HadCRU\_MLE has improvements compared to C&W. Model parameters were estimated using MLE, which provides a better statistical foundation for parameter estimates and allows for quantification of parameter uncertainty. HadCRU\_MLE takes advantage of temporal correlations and correlations between land-ice and sea regions to improve temperature anomaly estimates. More sources of uncertainty are accounted for, including parameter estimation uncertainty and infilling uncertainty as well as all uncertainties of HadCRUT4 and HadSST4.

To correct for amplification bias, an amplification function was incorporated into the temperature anomaly model, which was fit to observations. As El Niño events correspond to warming concentrated in the Eastern Equatorial Pacific Ocean, neglecting the behaviour of ENSO could cause an attenuation bias towards the identity function of the estimate of the amplification function. To avoid this attenuation bias, an IVP for ENSO was obtained from CCSM4 output for a pre-industrial control scenario and incorporated into the temperature anomaly model. The inclusion of the amplification function and the IVP for ENSO increases the estimate of GMST change from the late 1800s to 2018 by 0.01°C.

To correct for sea ice bias, a temperature climatology model was fit to observations. Estimates of temperature climatologies and temperature anomalies were combined to quantify sea ice bias. Accounting for sea ice bias increases the estimate of GMST change from the late 1800s to 2018 by 0.08°C. The use of a temperature field that is discontinuous between sea and sea ice regions may cause an overcorrection for sea ice bias. The results of sensitivity tests using IABP-POLES data suggest that there may be a 50% overcorrection for sea ice bias. A comparison of HadCRU\_MLE with CCSM4 output under the RCP6.0 scenario, which roughly corresponds to historic forcing, suggests that there may be an overcorrection by a factor of two for sea ice bias. Overall, the median estimate of GMST change from the late 1800s to 2018 is 1.20°C, with a 95% confidence interval of (1.11°C, 1.30°C).

## Acknowledgments

I thank all who contributed to the source datasets used in this study, the editor and reviewers of this paper, the World Data Center for Climate at DKRZ for archiving the code and data of HadCRU\_MLE [42], and Steven Mosher and Kevin Cowtan for providing inspiration. This research received no funding.

## References

1. Morice CP, Kennedy JJ, Rayner NA, Jones PD (2012) Quantifying Uncertainties in Global and Regional Temperature Change using an Ensemble of Observational Estimates: The HadCRUT4 Dataset. *J Geophys Res Atmos* 117: D08101.
2. Lenssen NJL, Schmidt GA, Hansen JE, Menne MJ, Persin A, et al. (2019) Improvements in the GISTEMP uncertainty model. *J Geophys Res Atmos* 124: 6307-6326.
3. Vose RS, Arndt D, Banzon VF, Easterling DR, Gleason BE, et al. (2012) NOAA's Merged Land-Ocean Surface Temperature Analysis. *Bull Am Meteorol Soc* 93: 1677-1685.
4. Huang B, Thorne PW, Banzon VF, Boyer T, Chepurin G, et al. (2017) Extended Reconstructed Sea Surface Temperature version 5 (ERSSTv5): Upgrades, validations, and intercomparisons *J Clim* 30: 8179-8205.
5. Menne MJ, Williams CN, Gleason BE, Rennie JJ, Lawrimore JH (2018) The Global Historical Climatology Network Monthly Temperature Dataset, Version 4. *J Clim* 31: 9835-9854.
6. Krige DG (1951) A Statistical Approach to some Basic Mine Valuation Problems on the Witwatersrand. *J Chem Metall Min Soc S Afr* 52: 119-139.
7. Matheron G (1962) *Traité de Géostatistique Appliquée* (Vol 1). Éditions Technip, Paris, France.

8. Rohde RA, Muller RA, Jacobsen R, Muller E, Perlmutter S, et al. (2013) A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011. *Geoinf Geostat Overview 1*.
9. Rohde RA, Muller RA, Jacobsen R, Perlmutter S, Rosenfeld A, et al. (2013) Berkeley Earth Temperature Averaging Process. *Geoinf Geostat Overview 1*.
10. Rohde RA, Hausfather Z (2020) The Berkeley Earth Land/Ocean Temperature Record. *Earth Syst Sci Data 12*: 3469-3479.
11. Cowtan K, Way RG (2014) Coverage Bias in the HadCRUT4 Temperature Series and its Impact on Recent Temperature Trends. *Q J R Meteorol Soc 140*: 1935-1944.
12. Cowtan K, Way RG (2014) Coverage Bias in the HadCRUT4 Temperature Series and its Impact on Recent Temperature Trends. UPDATE. Temperature Reconstruction by Domain: Version 2.0 Temperature Series. University of York Press, York, UK.
13. Kennedy JJ, Rayner NA, Smith RO, Parker DE, Saunby M (2011) Reassessing Biases and Other Uncertainties in Sea-Surface Temperature Observations since 1850 Part 1: Measurement and Sampling Errors. *J Geophys Res Atmos 116*: D14103.
14. Kennedy JJ, Rayner NA, Smith RO, Parker DE, Saunby M (2011) Reassessing Biases and Other Uncertainties in Sea-Surface Temperature Observations since 1850 Part 2: Biases and Homogenisation. *J Geophys Res Atmos 116*: D14104.
15. Kennedy JJ, Rayner NA, Atkinson CP, Killick RE (2019) 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*: 7719-7763.
16. Akaike H (1974) A New Look at the Statistical Model Identification. *IEEE Trans Automat Contr 19*: 716-723.
17. Jones PD, Lister DH, Osborn TJ, Harpham C, Salmon M, et al. (2012) Hemispheric and Large-scale Land Surface Air Temperature Variations: An Extensive Revision and an Update to 2010. *J Geophys Res Atmos 117*: D05127.
18. Donlon CJ, Marin M, Stark J, Roberts-Jones J, Fiedler E, et al. (2011) The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA). *Remote Sens Environ 116*: 140-158.
19. Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, et al. (2003) Global Analyses of Sea Surface Temperature, Sea Ice, and Night Marine Air Temperature since the Late Nineteenth Century. *J Geophys Res Atmos 108*: 4407.
20. Titchner HA, Rayner NA (2014) The Met Office Hadley Center Sea Ice and Sea Surface Temperature Data Set, Version 2: 1. Sea Ice Concentrations. *J Geophys Res Atmos 119*: 2864-2889.
21. Jones PD, New M, Parker DE, Martin S, Rigor IG (1999) Surface Air Temperature and its Changes over the Past 150 Years. *Rev Geophys 37*: 173-199.
22. Rigor IG, Colony RL, Martin S (2000) Variations in surface air temperature in the Arctic, 1979-1997. *J Clim 13*: 896-914.
23. Danielson JJ, Gesch DB (2011) Global multi-resolution terrain elevation data 2010 (GMTED2010). US Geol Surv Open-File Rep: 2011-1073.
24. Meehl GA, Hu A, Santer BD, Xie S-P (2016) Contribution of the Interdecadal Pacific Oscillation to Twentieth-Century Global Surface Temperature Trends. *Nat Clim Change 6*: 1005-1008.
25. Ropelewski CF, Jones PD (1987) An Extension of the Tahiti-Darwin Southern Oscillation Index. *Mon Weather Rev 115*: 2161-2165.
26. Serreze MC, Barry RG (2011) Processes and impacts of Arctic amplification: A research synthesis. *Glob Planet Change 77*: 85-96.
27. Cohen J, Screen JA, Furtado JC, Barlow M, Whittleston D, et al. (2014) Recent Arctic amplification and extreme mid-latitude weather. *Nat Geosci 7*: 627-637.
28. Stjern CW, Lund MT, Samset BH, Myhre G, Forster PM, et al. (2019) Arctic Amplification Response to Individual Climate Drivers. *J Geophys Res Atmos 124*: 6698-6717.
29. Pascanu R, Dauphin YN, Ganguli S, Bengio Y (2014) On the Saddle Point Problem for Non-Convex Optimization.
30. Dauphin YN, Pascanu R, Gulcehre C, Cho K, Ganguli S, et al. (2014) Identifying and Attacking the Saddle Point Problem in High-Dimensional Non-Convex Optimization. *Proceedings of Advances in Neural Information Processing Systems 27 (NIPS 2014)*.
31. Schwarz G (1978) Estimating the Dimension of a Model. *Ann Stat 6*: 461-464.
32. Cowtan K, Hausfather Z, Hawkins E, Jacobs P, Mann ME, et al. (2015) Robust Comparison of Climate Models with Observations using Blended Land Air and Ocean Sea Surface Temperatures. *Geophys Res Lett 42*: 6526-6534.
33. Richardson M, Cowtan K, Hawkins E, Stolpe MB (2016) Reconciled Climate Response Estimates from Climate Models and the Energy Budget of Earth. *Nat Clim Change 6*: 931-935.
34. New M, Hulme M, Jones PD (1999) Representing Twentieth-Century Space-Time Climate Variability. Part I: Development of a 1961-90 Mean Monthly Terrestrial Climatology. *J Clim 13*: 2217-2238.
35. Ishii M, Shouji A, Sugimoto S, Matsumoto T (2005) Objective Analyses of Sea-Surface Temperature and Marine Meteorological Variables for the 20<sup>th</sup> Century using ICOADS and the Kobe Collection. *Int J Climatol 25*: 865-879.
36. van den Dool HM, Sasha S, Johansson Å (2000) Empirical Orthogonal Teleconnections. *J Clim 13*: 1421-1435.
37. Reynolds RW, Rayner NA, Smith TM, Stokes DC, Wang W (2002) An Improved In Situ and Satellite SST Analysis for Climate. *J Clim 15*: 1609-1625.
38. Lorenz EN (2015) Empirical Orthogonal Functions and Statistical Weather Prediction. Massachusetts Institute of Technology, Department of Meteorology, Cambridge, USA.
39. Huber M, Knutti R (2014) Natural variability, radiative forcing and climate response in the recent hiatus reconciled. *Nat Geosci 7*: 651-656.
40. Stolpe MB, Medhaug I, Knutti R (2017) Contribution of Atlantic and Pacific Multidecadal Variability to Twentieth-Century Temperature Changes. *J Clim 30*: 6279-6295.
41. Riggs DS, Guarnieri JA, Addelman S (1978) Fitting straight lines when both variables are subject to error. *Life Sci 22*: 1305-1360.
42. Calvert BTT (2021) Maximum Likelihood Estimates of Temperatures using Data from the Hadley Centre and the Climate Research Unit (Version 1.0). World Data Center for Climate (WDCC) at DKRZ, Hamburg, Germany.

## Author Affiliation

Top

Independent Researcher, Ottawa, Canada



# Supplementary Information for “Correcting for Amplification Bias and Sea Ice Bias in Global Temperature Datasets”

by Bruce T. T. Calvert

## Source Data Availability

All source datasets used to produce HadCRU\_MLE\_v1.0 are publicly available; instructions for obtaining these datasets are given in the code zip file of HadCRU\_MLE\_v1.0, which is available at the World Data Center for Climate at DKRZ. Table S1 provides additional details about the source datasets used.

Dataset	Version	Types of Data Used	First Month of Data	Last Month of Data
HadCRUT4	4.6.0.0	Surface temperature anomalies	January 1850	September 2019
CRUTEM4	4.6.0.0	LSAT anomalies	January 1850	September 2019
HadSST3	3.1.1.0	SST anomalies	January 1850	October 2019
HadSST3	3.1.1.0	SST climatology	January 1961	December 1990
HadSST4	4.0.0.0	SST anomalies	January 1850	December 2018
HadISST1	1.0	Sea ice concentrations	January 1870	July 2019
HadISST2	2.2.0.0	Sea ice concentrations	January 1850	January 2019
OSTIA	1.0	Land mask	N/A	N/A
Jones et al. (1999)	N/A	Surface temperature climatology	January 1961	December 1990
IABP-POLES	N/A	Surface temperature climatology	January 1979	December 1998
Cowtan and Way using HadSST3 (long reconstruction)	2.0	Surface temperature anomalies	January 1850	August 2019
Cowtan and Way using HadSST4 (long reconstruction)	2.0	Surface temperature anomalies	January 1850	December 2018
Berkeley Earth Surface Temperature using LSATs for sea ice regions	N/A	Surface temperature anomalies	January 1850	October 2019
GMTED2010	N/A	Surface elevation	N/A	N/A
CCSM4 output for the pre-industrial control scenario (ensemble member r1i1p1)	N/A	SATs and sea ice concentrations	January 800	December 1300
CCSM4 output for the RCP6.0 scenario (ensemble member r1i1p1)	N/A	SATs and sea ice concentrations	January 1850	December 2100
Southern Oscillation Index	N/A	Southern Oscillation Index	January 1876	June 2019

**Table S1:** Additional details about the source datasets used to produce HadCRU\_MLE\_v1.0.

## Code Availability

The code used to produce HadCRU\_MLE\_v1.0 was written in MATLAB and is publicly available; the code is available in the code zip file of HadCRU\_MLE\_v1.0, which is available at the World Data Center for Climate at DKRZ. Due to the high computational requirements of the code, the code must be run using a graphics processing unit. Using a GeForce GTX 1050Ti graphics card, it takes about 44 days to execute the code. In addition, after the code is run once, it takes about 12 days to update the HadCRU\_MLE\_v1 dataset using updated source data. All random numbers generated by the code are controlled using a single random seed, which was generated using the system clock. Thus, results can be easily reproduced using the same random seed or verified using a different random seed. In addition, the random numbers are generated in such a way that the results should be relatively comparable if the dataset is updated with newer source data provided that the same random seed is used.

## Inferring Adjusted CRUTEM4

HadCRUT4 data blends HadSST3 data with CRUTEM4 data (with adjustments to account for biases and uncertainties such as due to temperature homogenization and the urban heat island effect). If observations exist for both HadSST3 and CRUTEM4 at a given grid cell and month, then the HadCRUT4 temperature anomaly is a weighted average of the HadSST3 temperature anomaly and the adjusted CRUTEM4 temperature anomaly. In this case, the adjusted CRUTEM4 weight is equal to the land fraction of the grid cell or 0.25, whichever is greater.

Suppose that observations for both HadSST3 and CRUTEM4 exist at  $\vec{x}$  for month  $t$ .

Let  $landweight_{\vec{x}} = \max(0.25, land_{\vec{x}})$ . Then the adjusted CRUTEM4 temperature anomaly  $Y_{\vec{x},1,t}$  can be calculated from the HadCRUT4 temperature anomaly  $HadCRUT4_{\vec{x},t}$  and the HadSST3 temperature anomaly  $HadSST3_{\vec{x},t}$  using  $Y_{\vec{x},1,t} = (HadCRUT4_{\vec{x},t} - (1 - landweight_{\vec{x}})HadSST3_{\vec{x},t})/landweight_{\vec{x}}$ .

## Temperature Climatology Model

### Functional Forms of the Temperature Climatology Model

Let  $G = 36 \cdot 72$  be the number of grid cells. For convenience, the convention  $m = 1$  corresponds to January and the normalization operator  $\langle \rangle$  are used.  $\langle \rangle$  is the operator that returns the function minus its weighted mean, where grid cells are weighted by their surface area. Definitions of the models of the temperature climatology function are given by Table S2. Definitions of the models of the covariance function of the temperature climatology residuals are given by Table S3.

Model	$X_{\bar{x},m}B$
1	$(e_{\bar{x}}^{land} land_{\bar{x}} + e_{\bar{x}}^{not\ land}(1 - land_{\bar{x}})) B_1 + (B_{1+m} + \alpha_{1,\bar{x}} B_{13+m} + \alpha_{2,\bar{x}} B_{25+m})$
2	$(e_{\bar{x}}^{land} land_{\bar{x}} + e_{\bar{x}}^{not\ land}(1 - land_{\bar{x}})) B_1 + \overline{se} \bar{a}_{\bar{x},m} (B_{1+m} + \alpha_{1,\bar{x}} B_{13+m} + \alpha_{2,\bar{x}} B_{25+m})$ $+ (land_{\bar{x}} + \overline{ice} \bar{e}_{\bar{x},m}) (B_{37+m} + \alpha_{1,\bar{x}} B_{49+m} + \alpha_{2,\bar{x}} B_{61+m})$
3	$(e_{\bar{x}}^{land} land_{\bar{x}} + e_{\bar{x}}^{not\ land}(1 - land_{\bar{x}})) B_1 + \overline{se} \bar{a}_{\bar{x},m} (B_{1+m} + \alpha_{1,\bar{x}} B_{13+m} + \alpha_{2,\bar{x}} B_{25+m})$ $+ (land_{\bar{x}} + SH_{\bar{x}} \overline{ice} \bar{e}_{\bar{x},m}) (B_{37+m} + \alpha_{1,\bar{x}} B_{49+m} + \alpha_{2,\bar{x}} B_{61+m})$ $+ NH_{\bar{x}} \overline{ice} \bar{e}_{\bar{x},m} (B_{73+m} + \alpha_{1,\bar{x}} B_{85+m} + \alpha_{2,\bar{x}} B_{97+m})$
4	$(e_{\bar{x}}^{land} land_{\bar{x}} + e_{\bar{x}}^{not\ land}(1 - land_{\bar{x}})) B_1$ $+ (B_{1+m} + \alpha_{1,\bar{x}} B_{13+m} + \alpha_{2,\bar{x}} B_{25+m} + \alpha_{3,\bar{x}} B_{37+m} + \alpha_{4,\bar{x}} B_{49+m})$
5	$(e_{\bar{x}}^{land} land_{\bar{x}} + e_{\bar{x}}^{not\ land}(1 - land_{\bar{x}})) B_1$ $+ \overline{se} \bar{a}_{\bar{x},m} (B_{1+m} + \alpha_{1,\bar{x}} B_{13+m} + \alpha_{2,\bar{x}} B_{25+m} + \alpha_{3,\bar{x}} B_{37+m} + \alpha_{4,\bar{x}} B_{49+m})$ $+ (land_{\bar{x}} + \overline{ice} \bar{e}_{\bar{x},m}) (B_{61+m} + \alpha_{1,\bar{x}} B_{73+m} + \alpha_{2,\bar{x}} B_{85+m} + \alpha_{3,\bar{x}} B_{97+m} + \alpha_{4,\bar{x}} B_{109+m})$
6	$(e_{\bar{x}}^{land} land_{\bar{x}} + e_{\bar{x}}^{not\ land}(1 - land_{\bar{x}})) B_1$ $+ \overline{se} \bar{a}_{\bar{x},m} (B_{1+m} + \alpha_{1,\bar{x}} B_{13+m} + \alpha_{2,\bar{x}} B_{25+m} + \alpha_{3,\bar{x}} B_{37+m} + \alpha_{4,\bar{x}} B_{49+m})$ $+ (land_{\bar{x}} + SH_{\bar{x}} \overline{ice} \bar{e}_{\bar{x},m}) (B_{61+m} + \alpha_{1,\bar{x}} B_{73+m} + \alpha_{2,\bar{x}} B_{85+m} + \alpha_{3,\bar{x}} B_{97+m} + \alpha_{4,\bar{x}} B_{109+m})$ $+ NH_{\bar{x}} \overline{ice} \bar{e}_{\bar{x},m} (B_{121+m} + \alpha_{1,\bar{x}} B_{133+m} + \alpha_{2,\bar{x}} B_{145+m} + \alpha_{3,\bar{x}} B_{157+m} + \alpha_{4,\bar{x}} B_{169+m})$

**Table S2:** The different models of the temperature climatology function. These models were used to quantify the sea ice bias of temperature anomaly estimates.  $\alpha_{1,\bar{x}} = \varphi_{\bar{x}}^3 - 3 \frac{\pi^2}{4} \varphi_{\bar{x}}$ ,  $\alpha_{2,\bar{x}} = \langle \varphi_{\bar{x}}^4 - \frac{\pi^2}{2} \varphi_{\bar{x}}^2 \rangle$ ,  $\alpha_{3,\bar{x}} = \varphi_{\bar{x}}^5 - 5 \frac{\pi^4}{16} \varphi_{\bar{x}}$ ,  $\alpha_{4,\bar{x}} = \langle \varphi_{\bar{x}}^6 - 3 \frac{\pi^4}{16} \varphi_{\bar{x}}^2 \rangle$ ,  $\varphi_{\bar{x}}$  is the latitude of  $\bar{x}$  in radians,  $SH_{\bar{x}}$  is 1 if  $\bar{x}$  is in the southern hemisphere and is otherwise 0, and  $NH_{\bar{x}} = 1 - SH_{\bar{x}}$ .

Model	$v_{\bar{x},m}Z$
1	$z_1$
2	$z_1 + (land_{\bar{x}} + \overline{ice} \bar{e}_{\bar{x},m}) z_2$
3	$z_1 + \alpha_{1,\bar{x}} z_2 + \alpha_{2,\bar{x}} z_3$
4	$z_1 + \alpha_{1,\bar{x}} z_2 + \alpha_{2,\bar{x}} z_3 + (land_{\bar{x}} + \overline{ice} \bar{e}_{\bar{x},m}) (z_4 + \alpha_{1,\bar{x}} z_5 + \alpha_{2,\bar{x}} z_6)$

**Table S3:** The different models of the covariance function of temperature climatology residuals. These models were used to quantify the sea ice bias of temperature anomaly estimates.

### Notation for the Temperature Climatology Model

For convenience, the following notation is used:

- ' is the transpose operator
- $\otimes$  is the Kronecker product
- $\mathbb{O}_{a,b}$  is the zeros matrix of size  $a \times b$
- $\circ$  is the Hadamard product
- $I_a$  is the identity matrix of size  $a \times a$
- $\mathbb{1}_{a,b}$  is the ones matrix of size  $a \times b$

- $tr$  is the trace operator
- $vec$  is the vectorization operator
- $sum$  is the operator that returns the sum of the elements of a matrix
- $c$  is the operator that returns the lower triangular Cholesky matrix
- $c'$  is the operator that returns the upper triangular Cholesky matrix
- $\lambda$  is the standard logistic function
- $G = 36 \cdot 72$  is the number of grid cells

Let  $\{\vec{x}_1, \dots, \vec{x}_G\}$  be an ordering of the set of all grid cells,  $\rho = \lambda(r)$ ,  $\kappa = \lambda(\kappa)$ ,

$$R = \begin{bmatrix} \rho^{1-\delta_{1,1}} & \dots & \rho^{1-\delta_{1,12}} \\ \vdots & \ddots & \vdots \\ \rho^{1-\delta_{12,1}} & \dots & \rho^{1-\delta_{12,12}} \end{bmatrix}, \text{ and } \tilde{K} = \begin{bmatrix} \kappa^{\|\vec{x}_1 - \vec{x}_1\|} & \dots & \kappa^{\|\vec{x}_1 - \vec{x}_G\|} \\ \vdots & \ddots & \vdots \\ \kappa^{\|\vec{x}_G - \vec{x}_1\|} & \dots & \kappa^{\|\vec{x}_G - \vec{x}_G\|} \end{bmatrix}. \text{ To account for missing}$$

observations, let  $n$  be the number of observed grid subcells and  $N = 12n$  be the number of temperature climatology observations. Define  $\tilde{\eta}$  as the  $G \times n$  restriction matrix constructed by starting with  $I_G$  and then  $\forall i \in \{1, \dots, G\}$  removing the  $i^{\text{th}}$  column from  $I_G$  if  $\vec{x}_i$  is unobserved, and define  $\tilde{\zeta}$  as the  $G \times (G - n)$  restriction matrix constructed by starting with  $I_G$  and then  $\forall i \in \{1, \dots, G\}$  removing the  $i^{\text{th}}$  column from  $I_G$  if  $\vec{x}_i$  is observed.

$$\forall i, j \in \{1, \dots, 12\} \text{ let } C_i = \tilde{\eta}' \begin{bmatrix} C_{\vec{x}_1, i} \\ \vdots \\ C_{\vec{x}_G, i} \end{bmatrix}, \tilde{U}_i = \begin{bmatrix} U_{\vec{x}_1, i} \\ \vdots \\ U_{\vec{x}_G, i} \end{bmatrix}, U_i = \tilde{\eta}' \tilde{U}_i, \tilde{U}_i = \tilde{\zeta}' \tilde{U}_i, X_i = \tilde{\eta}' \begin{bmatrix} X_{\vec{x}_1, i} \\ \vdots \\ X_{\vec{x}_G, i} \end{bmatrix}, X_{i,a} \text{ be the } a^{\text{th}}$$

column of  $X_i \forall a \in \{1, \dots, n_B\}$ ,  $\Lambda_{i,j} = cov(\tilde{U}_i, \tilde{U}_j)$ ,  $\Omega_{i,j} = cov(U_i, U_j) = \tilde{\eta}' \Lambda_{i,j} \tilde{\eta}$ ,

$$v_{i,a} = \tilde{\eta}' \begin{bmatrix} \left(\frac{1}{2} v_{\vec{x}_1, i}\right)_a & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \left(\frac{1}{2} v_{\vec{x}_G, i}\right)_a \end{bmatrix} \tilde{\eta} \forall a \in \{1, \dots, n_z\},$$

$$\tilde{V}_i = \begin{bmatrix} \exp\left(\frac{1}{2} v_{\vec{x}_1, iZ}\right) & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \exp\left(\frac{1}{2} v_{\vec{x}_G, iZ}\right) \end{bmatrix}, V_i = \tilde{\eta}' \tilde{V}_i \tilde{\eta}, \text{ and } K = \tilde{\eta}' \tilde{K} \tilde{\eta}.$$

$$\text{In addition, let } \eta = I_{12} \otimes \tilde{\eta}, \zeta = I_{12} \otimes \tilde{\zeta}, C = \begin{bmatrix} C_1 \\ \vdots \\ C_{12} \end{bmatrix}, U = \begin{bmatrix} U_1 \\ \vdots \\ U_{12} \end{bmatrix}, \tilde{U} = \begin{bmatrix} \tilde{U}_1 \\ \vdots \\ \tilde{U}_{12} \end{bmatrix}, X = \begin{bmatrix} X_1 \\ \vdots \\ X_{12} \end{bmatrix},$$

$$\Lambda = \begin{bmatrix} \Lambda_{1,1} & \dots & \Lambda_{1,12} \\ \vdots & \ddots & \vdots \\ \Lambda_{12,1} & \dots & \Lambda_{12,12} \end{bmatrix}, \Omega = \eta' \Lambda \eta, \tilde{V} = \begin{bmatrix} \tilde{V}_1 & \mathbb{O}_{G,G} & \dots & \mathbb{O}_{G,G} \\ \mathbb{O}_{G,G} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbb{O}_{G,G} \\ \mathbb{O}_{G,G} & \dots & \mathbb{O}_{G,G} & \tilde{V}_{12} \end{bmatrix}, \text{ and } V = \eta' \tilde{V} \eta. \text{ For convenience,}$$

$\forall i \in \{1, \dots, 12\}$  let  $\Lambda_i = \Lambda_{i,i}$  and  $\Omega_i = \Omega_{i,i}$ .



### Calculating the Log-Likelihood of the Temperature Climatology Model

The log-likelihood of the temperature climatology model is  $l = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln|\Omega| - \frac{1}{2} U' \Omega^{-1} U$ .  $\Omega$  is a matrix with nearly 1 billion numbers and, as a result, takes up nearly 8 GB of RAM. This is too large to easily work with on today's standard desktop computer, so  $\Omega$  is broken down into smaller matrices.

$$\Omega = \begin{bmatrix} R_{1,1} V_1^{-1} K V_1^{-1} & \cdots & R_{1,12} V_1^{-1} K V_{12}^{-1} \\ \vdots & \ddots & \vdots \\ R_{12,1} V_{12}^{-1} K V_1^{-1} & \cdots & R_{12,12} V_{12}^{-1} K V_{12}^{-1} \end{bmatrix} = V^{-1} (R \otimes K) V^{-1}.$$

$$\begin{aligned} \Rightarrow \ln|\Omega| &= \ln(|V^{-1}(R \otimes K)V^{-1}|) = 2 \ln(|V^{-1}|) + \ln(|R \otimes K|) \\ &= -2 \ln(|V|) + n \ln(|R|) + 12 \ln(|K|) = -2 \sum_{i=1}^{12} \ln(|V_i|) + 12 \ln(|K|) + n \ln(|R|) \\ \text{But } R \text{ has determinant } |R| &= (1 - \rho)^{11} (1 + 11\rho). \\ \Rightarrow \ln|\Omega| &= -2 \sum_{i=1}^{12} \ln(|V_i|) + 12 \ln(|K|) + n \ln((1 - \rho)^{11} (1 + 11\rho)) \end{aligned}$$

$$\begin{aligned} \Rightarrow \Omega^{-1} &= (V^{-1})^{-1} (R \otimes K)^{-1} (V^{-1})^{-1} = V (R^{-1} \otimes K^{-1}) V \\ \Rightarrow U' \Omega^{-1} U &= \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} U_i' V_i K^{-1} V_j U_j \end{aligned}$$

$$\begin{aligned} \Rightarrow l &= -\frac{N}{2} \ln(2\pi) + \sum_{i=1}^{12} \ln(|V_i|) - 6 \ln(|K|) - \frac{n}{2} \ln((1 - \rho)^{11} (1 + 11\rho)) \\ &- \frac{1}{2} \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} U_i' V_i K^{-1} V_j U_j, \text{ where } (R^{-1})_{i,j} = \begin{cases} \frac{1+10\rho}{(1-\rho)(1+11\rho)} & \text{if } i = j \\ \frac{-\rho}{(1-\rho)(1+11\rho)} & \text{else} \end{cases}. \end{aligned}$$

For large matrices, calculating the natural logarithm of the determinant can result in significant computational error. However,  $\ln(|K|) = \ln(|c(K)c'(K)|) = 2 \text{sum} \left( \text{eln} \left( \text{diag}(c(K)) \right) \right)$ , where  $\text{eln}$  is the element-wise natural logarithm function and  $\text{diag}$  is the operator that converts a matrix into a vector of its diagonal elements. Calculating 2 times the sum of the natural logarithm of the diagonals of the lower triangular Cholesky matrix is more numerically stable than calculating the natural logarithm of the determinant, so is performed instead. This method for computing the natural logarithm of the determinant is used for other large matrices as well. Various other methods to reduce the

computational cost are employed in practice; some of these methods are not listed in this supplementary information as they are relatively trivial and would reduce clarity.

### Initial Guess of the Temperature Climatology Model

To create an initial guess of  $(B, r, k, z)$ , the following method is used:

- $B$  is initially estimated using ordinary least squares (OLS) regression:  $C_{\vec{x},m} = X_{\vec{x},m}B + U_{\vec{x},m}$ .
- The residuals  $U$  of the OLS regression are obtained.
- $z_2, \dots, z_{n_z}$  are initially estimated as zero.
- If  $z_2, \dots, z_{n_z}$  are zero, then the expected value of  $U_{\vec{x},m}^2$  is  $e^{-z_1} \forall (\vec{x}, m) \in \mathcal{E}$ , where  $\mathcal{E}$  is the set of all combinations of grid cells and calendar months that have observations. As a result,  $z_1$  is initially estimated as  $-\ln(\sum_{(\vec{x},m) \in \mathcal{E}} (U_{\vec{x},m}^2) / \sum_{(\vec{x},m) \in \mathcal{E}} 1)$ .
- $r$  is initially estimated by comparing residuals  $U$  within the same grid cell and in different calendar months.  $\forall (\vec{x}, m_i), (\vec{x}, m_j) \in \mathcal{E}$ , if  $m_i \neq m_j$ , then  $\text{cov}(U_{\vec{x},m_i}, U_{\vec{x},m_j}) = \lambda(r) \exp(-0.5(v_{\vec{x},m_i} + v_{\vec{x},m_j})z)$ . As a result,  $r$  is initially estimated as

$$\lambda^{-1} \left( \sum_{(\vec{x},m_i) \in \mathcal{E}} \sum_{\substack{(\vec{x},m_j) \in \mathcal{E} \\ m_i \neq m_j}} \frac{U_{\vec{x},m_i} U_{\vec{x},m_j}}{\exp(-0.5(v_{\vec{x},m_i} + v_{\vec{x},m_j})z)} \bigg/ \sum_{(\vec{x},m_i) \in \mathcal{E}} \sum_{\substack{(\vec{x},m_j) \in \mathcal{E} \\ m_i \neq m_j}} 1 \right).$$

- $k$  is initially estimated by comparing residuals  $U$  within the same calendar month, within the same longitudinal band, and in different and adjacent latitudinal bands.  $\forall (\vec{x}_i, m), (\vec{x}_j, m) \in \mathcal{E}$ , if  $\psi_{\vec{x}_i} = \psi_{\vec{x}_j}$  and  $|\varphi_{\vec{x}_i} - \varphi_{\vec{x}_j}| = \frac{5\pi}{180}$ , where  $\varphi_{\vec{x}}$  and  $\psi_{\vec{x}}$  are the latitude and longitude of  $\vec{x}$  in radians respectively, then  $\text{cov}(U_{\vec{x}_i,m}, U_{\vec{x}_j,m}) = \lambda(k)^{\frac{5\pi}{180}} \exp(-0.5(v_{\vec{x}_i,m} + v_{\vec{x}_j,m})z)$ . As a result,  $k$  is estimated as

$$\lambda^{-1} \left( \left( \sum_{(\vec{x}_i,m) \in \mathcal{E}} \sum_{\substack{(\vec{x}_j,m) \in \mathcal{E} \\ \psi_{\vec{x}_i} = \psi_{\vec{x}_j} \\ |\varphi_{\vec{x}_i} - \varphi_{\vec{x}_j}| = \frac{5\pi}{180}}} \frac{U_{\vec{x}_i,m} U_{\vec{x}_j,m}}{\exp(-0.5(v_{\vec{x}_i,m} + v_{\vec{x}_j,m})z)} \bigg/ \sum_{(\vec{x}_i,m) \in \mathcal{E}} \sum_{\substack{(\vec{x}_j,m) \in \mathcal{E} \\ \psi_{\vec{x}_i} = \psi_{\vec{x}_j} \\ |\varphi_{\vec{x}_i} - \varphi_{\vec{x}_j}| = \frac{5\pi}{180}}} 1 \right)^{\frac{180}{5\pi}} \right).$$

- $B$  is re-estimated using GLS regression:  $C_{\vec{x},m} = X_{\vec{x},m}B + U_{\vec{x},m}$ . In particular, the GLS estimate is  $(X' \Omega^{-1} X)^{-1} X' \Omega^{-1} C = (\sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} X_i' V_i K^{-1} V_j X_j)^{-1} \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} X_i' V_i K^{-1} V_j C_j$ .

### Estimating the Temperature Climatology Model

The first and second derivatives of the log-likelihood function are

- $\frac{\partial l}{\partial B_a} = -U' \Omega^{-1} \frac{\partial U}{\partial B_a} = \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} U'_i V_i K^{-1} V_j X_{j,a} \quad \forall a \in \{1, \dots, n_B\}$
- $\frac{\partial l}{\partial r} = -\frac{1}{2} \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial r} \right) - \frac{1}{2} U' \frac{\partial \Omega^{-1}}{\partial r} U = \frac{66n\rho^2}{(1+11\rho)} - \frac{1}{2} \sum_{i=1}^{12} \sum_{j=1}^{12} \left( \frac{\partial R^{-1}}{\partial r} \right)_{i,j} U'_i V_i K^{-1} V_j U_j,$   
 where  $\left( \frac{\partial R^{-1}}{\partial r} \right)_{i,j} = \left( \frac{\partial R^{-1}}{\partial \rho} \rho(1-\rho) \right)_{i,j} = \begin{cases} \frac{22\rho^2(1+5\rho)}{(1-\rho)(1+11\rho)^2} & \text{if } i = j \\ \frac{-\rho(1+11\rho^2)}{(1-\rho)(1+11\rho)^2} & \text{else} \end{cases}.$
- $\frac{\partial l}{\partial k} = -\frac{1}{2} \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial k} \right) - \frac{1}{2} U' \frac{\partial \Omega^{-1}}{\partial k} U = -6 \text{tr} \left( K^{-1} \frac{\partial K}{\partial k} \right) - \frac{1}{2} \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} U'_i V_i \frac{\partial K^{-1}}{\partial k} V_j U_j,$   
 where  $\frac{\partial K}{\partial k} = \frac{\partial K}{\partial \kappa} \kappa(1-\kappa)$  and  $\frac{\partial K^{-1}}{\partial k} = -K^{-1} \frac{\partial K}{\partial k} K^{-1}.$
- $\frac{\partial l}{\partial z_a} = -\frac{1}{2} \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial z_a} \right) - \frac{1}{2} U' \frac{\partial \Omega^{-1}}{\partial z_a} U$   
 $= \sum_{i=1}^{12} \text{tr}(v_{i,a}) - \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} U'_i V_i K^{-1} v_{j,a} V_j U_j \quad \forall a \in \{1, \dots, n_z\}$
- $\frac{\partial^2 l}{\partial B_a \partial B_b} = -\frac{\partial U'}{\partial B_a} \Omega^{-1} \frac{\partial U}{\partial B_b} = -\sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} X'_{i,a} V_i K^{-1} V_j X_{j,b} \quad \forall a, b \in \{1, \dots, n_B\}$
- $\frac{\partial^2 l}{\partial B_a \partial r} = -\frac{\partial U'}{\partial B_a} \frac{\partial \Omega^{-1}}{\partial r} U = \sum_{i=1}^{12} \sum_{j=1}^{12} \left( \frac{\partial R^{-1}}{\partial r} \right)_{i,j} X'_{i,a} V_i K^{-1} V_j U_j \quad \forall a \in \{1, \dots, n_B\}$
- $\frac{\partial^2 l}{\partial B_a \partial k} = -\frac{\partial U'}{\partial B_a} \frac{\partial \Omega^{-1}}{\partial k} U = \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} X'_{i,a} V_i \frac{\partial K^{-1}}{\partial k} V_j U_j \quad \forall a \in \{1, \dots, n_B\}$
- $\frac{\partial^2 l}{\partial B_a \partial z_b} = -\frac{\partial U'}{\partial B_a} \frac{\partial \Omega^{-1}}{\partial z_b} U$   
 $= \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} X'_{i,a} V_i (v_{i,b} K^{-1} + K^{-1} v_{j,b}) V_j U_j \quad \forall a \in \{1, \dots, n_B\} \quad \forall b \in \{1, \dots, n_z\}$
- $\frac{\partial^2 l}{\partial r^2} = -\frac{1}{2} \text{tr} \left( \Omega^{-1} \left( \frac{\partial^2 \Omega}{\partial r^2} - \frac{\partial \Omega}{\partial r} \Omega^{-1} \frac{\partial \Omega}{\partial r} \right) \right) - \frac{1}{2} U' \frac{\partial^2 \Omega^{-1}}{\partial r^2} U$   
 $= \frac{66n\rho^2(1-\rho)(2+11\rho)}{(1+11\rho)^2} - \frac{1}{2} \sum_{i=1}^{12} \sum_{j=1}^{12} \left( \frac{\partial^2 R^{-1}}{\partial r^2} \right)_{i,j} U'_i V_i K^{-1} V_j U_j,$   
 where  $\left( \frac{\partial^2 R^{-1}}{\partial r^2} \right)_{i,j} = \left( \frac{\partial^2 R^{-1}}{\partial \rho \partial r} \rho(1-\rho) \right)_{i,j} = \begin{cases} \frac{44\rho^2(1+7\rho+28\rho^2)}{(1-\rho)(1+11\rho)^3} & \text{if } i = j \\ \frac{-\rho(1-11\rho+55\rho^2-99\rho^3)}{(1-\rho)(1+11\rho)^3} & \text{else} \end{cases}.$
- $\frac{\partial^2 l}{\partial r \partial k} = -\frac{1}{2} \text{tr} \left( \Omega^{-1} \left( \frac{\partial^2 \Omega}{\partial r \partial k} - \frac{\partial \Omega}{\partial r} \Omega^{-1} \frac{\partial \Omega}{\partial k} \right) \right) - \frac{1}{2} U' \frac{\partial^2 \Omega^{-1}}{\partial r \partial k} U$   
 $= -\frac{1}{2} \sum_{i=1}^{12} \sum_{j=1}^{12} \left( \frac{\partial R^{-1}}{\partial r} \right)_{i,j} U'_i V_i \frac{\partial K^{-1}}{\partial k} V_j U_j$

- $\frac{\partial^2 l}{\partial r \partial z_a} = -\frac{1}{2} \text{tr} \left( \Omega^{-1} \left( \frac{\partial^2 \Omega}{\partial r \partial z_a} - \frac{\partial \Omega}{\partial r} \Omega^{-1} \frac{\partial \Omega}{\partial z_a} \right) \right) - \frac{1}{2} U' \frac{\partial^2 \Omega^{-1}}{\partial r \partial z_a} U$   
 $= -\sum_{i=1}^{12} \sum_{j=1}^{12} \left( \frac{\partial R^{-1}}{\partial r} \right)_{i,j} U_i' V_i K^{-1} v_{j,a} V_j U_j \quad \forall a \in \{1, \dots, n_z\}$
- $\frac{\partial^2 l}{\partial k^2} = -\frac{1}{2} \text{tr} \left( \Omega^{-1} \left( \frac{\partial^2 \Omega}{\partial k^2} - \frac{\partial \Omega}{\partial k} \Omega^{-1} \frac{\partial \Omega}{\partial k} \right) \right) - \frac{1}{2} U' \frac{\partial^2 \Omega^{-1}}{\partial k^2} U$   
 $= -6 \text{tr} \left( K^{-1} \frac{\partial^2 K}{\partial k^2} - K^{-1} \frac{\partial K}{\partial k} K^{-1} \frac{\partial K}{\partial k} \right) - \frac{1}{2} \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} U_i' V_i \frac{\partial^2 K^{-1}}{\partial k^2} V_j U_j,$   
where  $\frac{\partial^2 K}{\partial k^2} = \frac{\partial^2 K}{\partial \kappa^2} \kappa^2 (1 - \kappa)^2 + \frac{\partial K}{\partial \kappa} (1 - 2\kappa)$  and  $\frac{\partial^2 K^{-1}}{\partial k^2} = K^{-1} \left( 2 \frac{\partial K}{\partial \kappa} K^{-1} \frac{\partial K}{\partial \kappa} - \frac{\partial^2 K}{\partial \kappa^2} \right) K^{-1}.$
- $\frac{\partial^2 l}{\partial k \partial z_a} = -\frac{1}{2} \text{tr} \left( \Omega^{-1} \left( \frac{\partial^2 \Omega}{\partial k \partial z_a} - \frac{\partial \Omega}{\partial k} \Omega^{-1} \frac{\partial \Omega}{\partial z_a} \right) \right) - \frac{1}{2} U' \frac{\partial^2 \Omega^{-1}}{\partial k \partial z_a} U$   
 $= -\sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} U_i' V_i \frac{\partial K^{-1}}{\partial k} v_{j,a} V_j U_j \quad \forall a \in \{1, \dots, n_z\}$
- $\frac{\partial^2 l}{\partial z_a \partial z_b} = -\frac{1}{2} \text{tr} \left( \Omega^{-1} \left( \frac{\partial^2 \Omega}{\partial z_a \partial z_b} - \frac{\partial \Omega}{\partial z_a} \Omega^{-1} \frac{\partial \Omega}{\partial z_b} \right) \right) - \frac{1}{2} U' \frac{\partial^2 \Omega^{-1}}{\partial z_a \partial z_b} U$   
 $= -\sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} U_i' V_i (v_{i,a} K^{-1} + K^{-1} v_{j,a}) v_{j,b} V_j U_j \quad \forall a, b \in \{1, \dots, n_z\}$

The derivatives of the log-likelihood function could, in principle, be used to obtain the maximum likelihood estimates. However, it can be easier to obtain maximum likelihood estimates by instead using the derivatives of a concentrated log-likelihood function (also called a profile log-likelihood function), which is obtained by reducing the number of parameters in the log-likelihood function. One of the conditions of the maximum likelihood estimates is  $\frac{\partial l}{\partial B} = 0$ , which implies that  $B$  is its GLS estimate  $(X' \Omega^{-1} X)^{-1} X' \Omega^{-1} C$ . Substituting this GLS estimate of  $B$  into the log-likelihood function yields a concentrated log-likelihood function  $\tilde{l}$ .

The first and second derivatives of the concentrated log-likelihood function are

- $\frac{\partial \tilde{l}}{\partial r} = \frac{\partial l}{\partial r} + U' \Omega^{-1} X \frac{\partial B}{\partial r} = \frac{\partial l}{\partial r} + \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} U_i' V_i K^{-1} V_j X_j \frac{\partial B}{\partial r},$   
where  $\frac{\partial B}{\partial r} = (X' \Omega^{-1} X)^{-1} X' \frac{\partial \Omega^{-1}}{\partial r} U.$
- $\frac{\partial \tilde{l}}{\partial k} = \frac{\partial l}{\partial k} + U' \Omega^{-1} X \frac{\partial B}{\partial k} = \frac{\partial l}{\partial k} + \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} U_i' V_i K^{-1} V_j X_j \frac{\partial B}{\partial k},$   
where  $\frac{\partial B}{\partial k} = (X' \Omega^{-1} X)^{-1} X' \frac{\partial \Omega^{-1}}{\partial k} U.$
- $\frac{\partial \tilde{l}}{\partial z_a} = \frac{\partial l}{\partial z_a} + U' \Omega^{-1} X \frac{\partial B}{\partial z_a} = \frac{\partial l}{\partial z_a} + \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} U_i' V_i K^{-1} V_j X_j \frac{\partial B}{\partial z_a} \quad \forall a \in \{1, \dots, n_z\},$   
where  $\frac{\partial B}{\partial z_a} = (X' \Omega^{-1} X)^{-1} X' \frac{\partial \Omega^{-1}}{\partial z_a} U \quad \forall a \in \{1, \dots, n_z\}.$



- $$\begin{aligned} \frac{\partial^2 \bar{l}}{\partial r^2} &= \frac{\partial^2 l}{\partial r^2} + \frac{\partial B'}{\partial r} X' \frac{\partial \Omega^{-1}}{\partial r} U - \frac{\partial B'}{\partial r} X' \Omega^{-1} X \frac{\partial B}{\partial r} + \frac{\partial^2 B'}{\partial r^2} X' \Omega^{-1} U \\ &= \frac{\partial^2 l}{\partial r^2} + \sum_{i=1}^{12} \sum_{j=1}^{12} \left( \frac{\partial R^{-1}}{\partial r} \right)_{i,j} \frac{\partial B'}{\partial r} X_i' V_i K^{-1} V_j U_j - \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} \frac{\partial B'}{\partial r} X_i' V_i K^{-1} V_j X_j \frac{\partial B}{\partial r} \\ &\quad + \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} \frac{\partial^2 B'}{\partial r^2} X_i' V_i K^{-1} V_j U_j, \text{ where } \frac{\partial^2 B}{\partial r^2} = (X' \Omega^{-1} X)^{-1} X' \left( \frac{\partial^2 \Omega^{-1}}{\partial r^2} U - 2 \frac{\partial \Omega^{-1}}{\partial r} X \frac{\partial B}{\partial r} \right). \end{aligned}$$
- $$\begin{aligned} \frac{\partial^2 \bar{l}}{\partial r \partial k} &= \frac{\partial^2 l}{\partial r \partial k} + \frac{\partial B'}{\partial k} X' \frac{\partial \Omega^{-1}}{\partial r} U - \frac{\partial B'}{\partial r} X' \Omega^{-1} X \frac{\partial B}{\partial k} + \frac{\partial^2 B'}{\partial r \partial k} X' \Omega^{-1} U \\ &= \frac{\partial^2 l}{\partial r \partial k} + \sum_{i=1}^{12} \sum_{j=1}^{12} \left( \frac{\partial R^{-1}}{\partial r} \right)_{i,j} \frac{\partial B'}{\partial k} X_i' V_i K^{-1} V_j U_j - \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} \frac{\partial B'}{\partial r} X_i' V_i K^{-1} V_j X_j \frac{\partial B}{\partial k} \\ &\quad + \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} \frac{\partial^2 B'}{\partial r \partial k} X_i' V_i K^{-1} V_j U_j, \\ \text{where } \frac{\partial^2 B}{\partial r \partial k} &= (X' \Omega^{-1} X)^{-1} X' \left( \frac{\partial^2 \Omega^{-1}}{\partial r \partial k} U - \frac{\partial \Omega^{-1}}{\partial r} X \frac{\partial B}{\partial k} - \frac{\partial \Omega^{-1}}{\partial k} X \frac{\partial B}{\partial r} \right). \end{aligned}$$
- $$\begin{aligned} \frac{\partial^2 \bar{l}}{\partial r \partial z_a} &= \frac{\partial^2 l}{\partial r \partial z_a} + \frac{\partial B'}{\partial z_a} X' \frac{\partial \Omega^{-1}}{\partial r} U - \frac{\partial B'}{\partial r} X' \Omega^{-1} X \frac{\partial B}{\partial z_a} + \frac{\partial^2 B'}{\partial r \partial z_a} X' \Omega^{-1} U \\ &= \frac{\partial^2 l}{\partial r \partial z_a} + \sum_{i=1}^{12} \sum_{j=1}^{12} \left( \frac{\partial R^{-1}}{\partial r} \right)_{i,j} \frac{\partial B'}{\partial z_a} X_i' V_i K^{-1} V_j U_j - \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} \frac{\partial B'}{\partial r} X_i' V_i K^{-1} V_j X_j \frac{\partial B}{\partial z_a} \\ &\quad + \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} \frac{\partial^2 B'}{\partial r \partial z_a} X_i' V_i K^{-1} V_j U_j \quad \forall a \in \{1, \dots, n_z\}, \\ \text{where } \frac{\partial^2 B}{\partial r \partial z_a} &= (X' \Omega^{-1} X)^{-1} X' \left( \frac{\partial^2 \Omega^{-1}}{\partial r \partial z_a} U - \frac{\partial \Omega^{-1}}{\partial r} X \frac{\partial B}{\partial z_a} - \frac{\partial \Omega^{-1}}{\partial z_a} X \frac{\partial B}{\partial r} \right) \quad \forall a \in \{1, \dots, n_z\}. \end{aligned}$$
- $$\begin{aligned} \frac{\partial^2 \bar{l}}{\partial k^2} &= \frac{\partial^2 l}{\partial k^2} + \frac{\partial B'}{\partial k} X' \frac{\partial \Omega^{-1}}{\partial k} U - \frac{\partial B'}{\partial k} X' \Omega^{-1} X \frac{\partial B}{\partial k} + \frac{\partial^2 B'}{\partial k^2} X' \Omega^{-1} U \\ &= \frac{\partial^2 l}{\partial k^2} + \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} \frac{\partial B'}{\partial k} X_i' V_i \frac{\partial K^{-1}}{\partial k} V_j U_j - \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} \frac{\partial B'}{\partial k} X_i' V_i K^{-1} V_j X_j \frac{\partial B}{\partial k} \\ &\quad + \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} \frac{\partial^2 B'}{\partial k^2} X_i' V_i K^{-1} V_j U_j, \text{ where } \frac{\partial^2 B}{\partial k^2} = (X' \Omega^{-1} X)^{-1} X' \left( \frac{\partial^2 \Omega^{-1}}{\partial k^2} U - 2 \frac{\partial \Omega^{-1}}{\partial k} X \frac{\partial B}{\partial k} \right). \end{aligned}$$
- $$\begin{aligned} \frac{\partial^2 \bar{l}}{\partial k \partial z_a} &= \frac{\partial^2 l}{\partial k \partial z_a} + \frac{\partial B'}{\partial z_a} X' \frac{\partial \Omega^{-1}}{\partial k} U - \frac{\partial B'}{\partial k} X' \Omega^{-1} X \frac{\partial B}{\partial z_a} + \frac{\partial^2 B'}{\partial k \partial z_a} X' \Omega^{-1} U \\ &= \frac{\partial^2 l}{\partial k \partial z_a} + \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} \frac{\partial B'}{\partial z_a} X_i' V_i \frac{\partial K^{-1}}{\partial k} V_j U_j - \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} \frac{\partial B'}{\partial k} X_i' V_i K^{-1} V_j X_j \frac{\partial B}{\partial z_a} \\ &\quad + \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} \frac{\partial^2 B'}{\partial k \partial z_a} X_i' V_i K^{-1} V_j U_j \quad \forall a \in \{1, \dots, n_z\}, \\ \text{where } \frac{\partial^2 B}{\partial k \partial z_a} &= (X' \Omega^{-1} X)^{-1} X' \left( \frac{\partial^2 \Omega^{-1}}{\partial k \partial z_a} U - \frac{\partial \Omega^{-1}}{\partial k} X \frac{\partial B}{\partial z_a} - \frac{\partial \Omega^{-1}}{\partial z_a} X \frac{\partial B}{\partial k} \right) \quad \forall a \in \{1, \dots, n_z\}. \end{aligned}$$
- $$\begin{aligned} \frac{\partial^2 \bar{l}}{\partial z_a \partial z_b} &= \frac{\partial^2 l}{\partial z_a \partial z_b} + \frac{\partial B'}{\partial z_b} X' \frac{\partial \Omega^{-1}}{\partial z_a} U - \frac{\partial B'}{\partial z_a} X' \Omega^{-1} X \frac{\partial B}{\partial z_b} + \frac{\partial^2 B'}{\partial z_a \partial z_b} X' \Omega^{-1} U \\ &= \frac{\partial^2 l}{\partial z_a \partial z_b} + \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} \frac{\partial B'}{\partial z_b} X_i' V_i (v_{i,a} K^{-1} + K^{-1} v_{j,a}) V_j U_j \\ &\quad - \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} \frac{\partial B'}{\partial z_a} X_i' V_i K^{-1} V_j X_j \frac{\partial B}{\partial z_b} \\ &\quad + \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} \frac{\partial^2 B'}{\partial z_a \partial z_b} X_i' V_i K^{-1} V_j U_j \quad \forall a, b \in \{1, \dots, n_z\}, \\ \text{where } \frac{\partial^2 B}{\partial z_a \partial z_b} &= (X' \Omega^{-1} X)^{-1} X' \left( \frac{\partial^2 \Omega^{-1}}{\partial z_a \partial z_b} U - \frac{\partial \Omega^{-1}}{\partial z_a} X \frac{\partial B}{\partial z_b} - \frac{\partial \Omega^{-1}}{\partial z_b} X \frac{\partial B}{\partial z_a} \right) \quad \forall a, b \in \{1, \dots, n_z\}. \end{aligned}$$

These derivatives, combined with the initial guess of  $(r, k, z)$ , are used to iteratively converge to the maximum likelihood estimates using the saddle-free Newton's method. The new value of  $(r, k, z)$  is

$(r, k, z) + \min \left( 1, \frac{-\tilde{l}}{J[\tilde{H}]^{-1} \tilde{J}'} \right) [\tilde{H}]^{-1} \tilde{J}'$ , where  $\tilde{H} = \tilde{D}' \tilde{J}$  is the Hessian of the concentrated log-likelihood function,  $\tilde{J} = \tilde{D} \tilde{l}$  is the Jacobian of the concentrated log-likelihood function,  $\tilde{D} = \begin{bmatrix} \frac{\partial}{\partial r} & \frac{\partial}{\partial k} & \frac{\partial}{\partial z} \end{bmatrix}$ , and  $[\cdot]$  is the operator that returns the matrix with the same eigen-decomposition, but after taking the absolute value of each eigenvalue. The minimum function is used to reduce the step size if the step size is large compared to the magnitude of the log-likelihood function; this increases the stability of the iteration procedure when the determinant of the Hessian is close to zero. The new value of  $B$  is then calculated using GLS:  $B = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} C$ . 100 iterations are performed to achieve adequate convergence to the maximum likelihood estimates.

### Selecting the Temperature Climatology Model

After the estimation of each model is performed, the model with the lowest AIC is selected. More complicated models may be better able to explain observations, but they come at the cost of higher complexity and thus an increased risk of overfitting. The AIC is roughly a measure of the information lost by the model and balances the increase in the ability of the model to explain observations with the number of parameters in the model. The formula for the AIC is  $AIC = 2n_{parameters} - 2l$ , where  $n_{parameters} = n_{\beta} + n_z + 2$  is the number of model parameters. The BIC is calculated for comparison; it is similar to the AIC except  $2n_{parameters}$  is replaced with  $\ln(N) n_{parameters}$ .

$$AIC = 2n_{parameters} - 2l = 2(n_B + n_z + 2) + N \ln(2\pi) - 2 \sum_{i=1}^{12} \ln(|V_i|) + 12 \ln(|K|) + n \ln((1 - \rho)^{11} (1 + 11\rho)) + \sum_{i=1}^{12} \sum_{j=1}^{12} (R^{-1})_{i,j} U_i' V_i K^{-1} V_j U_j$$

### Parameter Uncertainty of the Temperature Climatology Model

The maximum likelihood estimate of  $(B, r, k, z)$  is asymptotically multivariate normal with covariance matrix  $(-H)^{-1}$ , where  $H = D'J$  is the Hessian of the log-likelihood function,  $J = D\tilde{l}$  is the Jacobian of the log-likelihood function, and  $D = \begin{bmatrix} \frac{\partial}{\partial B} & \frac{\partial}{\partial r} & \frac{\partial}{\partial k} & \frac{\partial}{\partial z} \end{bmatrix}$ . Since  $H$  is symmetric, the covariance matrix has eigen-decomposition  $\Phi \Delta \Phi^{-1} = \Phi \Delta \Phi'$ , where  $\Phi$  is a matrix of eigenvectors and  $\Delta$  is a diagonal matrix of eigenvalues. Thus  $(-H)^{-1} = \Phi (-\Delta)^{-1} \Phi' = \Phi (-\Delta)^{-0.5} (\Phi (-\Delta)^{-0.5})'$ . This multivariate normal approximation is used to quantify the uncertainty of the maximum likelihood estimates. For the selected temperature climatology model, 200 vectors of random numbers of length  $n_{parameters}$  are generated from independent standard normal probability distributions. For each of these 200 vectors,  $\Phi (-\Delta)^{-0.5}$  is multiplied by the random vector to obtain an ensemble member of the parameters  $(B, r, k, z)$  of the temperature climatology model.

### Estimating Residuals of the Temperature Climatology Model

Since some combinations of grid cells and calendar months are unobserved, their residuals must be estimated. The simple kriging estimate of the unobserved climatology residuals  $\tilde{U}$  is  $\zeta' \Lambda \eta \Omega^{-1} U$ . Conditional on the estimate of the model parameters, the kriging estimate is the most-efficient linear unbiased estimate as well as the maximum likelihood estimate of the unobserved climatology residuals.

$$\begin{aligned} \zeta' \Lambda \eta \Omega^{-1} U &= \begin{bmatrix} \sum_{i=1}^{12} \sum_{j=1}^{12} R_{1,i} \tilde{\zeta}' \tilde{V}_1^{-1} \tilde{K} \tilde{V}_i^{-1} \tilde{\eta} (R^{-1})_{i,j} V_i K^{-1} V_j U_j \\ \vdots \\ \sum_{i=1}^{12} \sum_{j=1}^{12} R_{12,i} \tilde{\zeta}' \tilde{V}_{12}^{-1} \tilde{K} \tilde{V}_i^{-1} \tilde{\eta} (R^{-1})_{i,j} V_i K^{-1} V_j U_j \end{bmatrix} \\ &= \begin{bmatrix} \sum_{j=1}^{12} \tilde{\zeta}' \tilde{V}_1^{-1} \tilde{K} \tilde{\eta} K^{-1} V_j U_j \sum_{i=1}^{12} R_{1,i} (R^{-1})_{i,j} \\ \vdots \\ \sum_{j=1}^{12} \tilde{\zeta}' \tilde{V}_{12}^{-1} \tilde{K} \tilde{\eta} K^{-1} V_j U_j \sum_{i=1}^{12} R_{12,i} (R^{-1})_{i,j} \end{bmatrix}. \text{ But } \sum_{i=1}^{12} R_{a,i} (R^{-1})_{i,j} = \delta_{a,j} \forall a \in \{1, \dots, 12\} \text{ since } R \text{ and } \\ R^{-1} &\text{ are inverses of each other. Thus, the simple kriging estimate is } \zeta' \Lambda \eta \Omega^{-1} U = \begin{bmatrix} \tilde{\zeta}' \tilde{V}_1^{-1} \tilde{K} \tilde{\eta} K^{-1} V_1 U_1 \\ \vdots \\ \tilde{\zeta}' \tilde{V}_{12}^{-1} \tilde{K} \tilde{\eta} K^{-1} V_{12} U_{12} \end{bmatrix}. \end{aligned}$$

### Infilling Uncertainty of the Temperature Climatology Model

The covariance matrix of  $\zeta' \Lambda \eta \Omega^{-1} U$  is  $\zeta' \Lambda \zeta - \zeta' \Lambda \eta \Omega^{-1} \eta' \Lambda \zeta$ . To account for uncertainties in the simple kriging estimate of  $\tilde{U}$ , the 200 ensemble members of parameters are used to obtain 200 ensemble

members of  $\tilde{U}$ . For each ensemble member of  $U$ , a vector  $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{12} \end{bmatrix}$ , is generated from an independent standard normal probability distribution, where  $\varepsilon_i \in \mathbb{R}^{G-n} \forall i \in \{1, \dots, 12\}$ .

For each ensemble member of  $U$ , an ensemble member of  $\tilde{U}$  can be generated as  $\zeta' \Lambda \eta \Omega^{-1} U +$

$$\begin{aligned} &c(\zeta' \Lambda \zeta - \zeta' \Lambda \eta \Omega^{-1} \eta' \Lambda \zeta) \varepsilon. \text{ Note that } \zeta' \Lambda \zeta - \zeta' \Lambda \eta \Omega^{-1} \eta' \Lambda \zeta = \tau = \begin{bmatrix} \tau_{1,1} & \cdots & \tau_{1,12} \\ \vdots & \ddots & \vdots \\ \tau_{12,1} & \cdots & \tau_{12,12} \end{bmatrix}, \text{ where} \\ \tau_{i,j} &= R_{i,j} \tilde{\zeta}' \tilde{V}_i^{-1} \tilde{K} \tilde{V}_j^{-1} \tilde{\zeta} - \sum_{a=1}^{12} \sum_{b=1}^{12} (R_{i,a} \tilde{\zeta}' \tilde{V}_i^{-1} \tilde{K} \tilde{V}_a^{-1} \tilde{\eta}) ((R^{-1})_{a,b} V_a K^{-1} V_b) (R_{b,j} \tilde{\eta}' \tilde{V}_b^{-1} \tilde{K} \tilde{V}_j^{-1} \tilde{\zeta}) \\ &= R_{i,j} \tilde{\zeta}' \tilde{V}_i^{-1} \tilde{K} \tilde{V}_j^{-1} \tilde{\zeta} - \tilde{\zeta}' \tilde{V}_i^{-1} \tilde{K} \tilde{\eta} K^{-1} \tilde{\eta}' \tilde{K} \tilde{V}_j^{-1} \tilde{\zeta} \sum_{a=1}^{12} \sum_{b=1}^{12} R_{i,a} (R^{-1})_{a,b} R_{b,j} \\ &= R_{i,j} \tilde{\zeta}' \tilde{V}_i^{-1} (\tilde{K} - \tilde{K} \tilde{\eta} K^{-1} \tilde{\eta}' \tilde{K}) \tilde{V}_j^{-1} \tilde{\zeta} \forall i, j \in \{1, \dots, 12\}. \end{aligned}$$

To reduce computational rounding error,  $c(\tau)$  is calculated as

$$\begin{bmatrix} c(\tau_{1,1}) & \mathbb{O}_{G-n,G-n} & \cdots & \mathbb{O}_{G-n,G-n} \\ \mathbb{O}_{G-n,G-n} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbb{O}_{G-n,G-n} \\ \mathbb{O}_{G-n,G-n} & \cdots & \mathbb{O}_{G-n,G-n} & c(\tau_{12,12}) \end{bmatrix} \begin{bmatrix} \Psi_{1,1} & \mathbb{O}_{G-n,G-n} & \cdots & \mathbb{O}_{G-n,G-n} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbb{O}_{G-n,G-n} \\ \Psi_{12,1} & \cdots & \cdots & \Psi_{12,12} \end{bmatrix}.$$

$\Rightarrow \Psi$  can be calculated iteratively using  $\Psi_{i,i} = c(I_{G-n} - \sum_{j=1}^{i-1} \Psi_{i,j} \Psi'_{i,j}) \forall i \in \{1, \dots, 12\}$  and

$$\Psi_{i,j} = \left( \left( c(\tau_{i,i}) \right)^{-1} \tau_{i,j} \left( c'(\tau_{j,j}) \right)^{-1} - \sum_{a=1}^{j-1} \Psi_{i,a} \Psi'_{j,a} \right) \Psi'_{j,j}^{-1} \forall i \in \{1, \dots, 12\} \forall j \in \{1, \dots, i-1\}.$$

After  $c(\tau)$  is calculated,  $\gamma = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_{12} \end{bmatrix} = c(\tau)\varepsilon$  is calculated using  $\gamma_i = c(\tau_{i,i})(\sum_{j=1}^i \Psi_{i,j}\varepsilon_j) \forall i \in \{1, \dots, 12\}$ .

The ensemble member of  $\tilde{U}$  is then calculated as  $\zeta' \Lambda \eta \Omega^{-1} U + \gamma$ .

These 200 ensemble members of climatology residuals are combined with the 200 ensemble members of the temperature climatology function to produce 200 ensemble members of temperature climatologies.

### Obtaining Patterns of Internal Variability

Let  $G = 36 \cdot 72$  be the number of grid cells,  $t = 1$  correspond to January 800, and  $M = 6012$  be the number months used from the pre-industrial control scenario of the CCSM4 model.  $\forall j \in \{1, \dots, \frac{M}{12}\}$  let  $\overline{\Delta T}_{\vec{x},j}$  be the weighted average of  $(\overline{\Delta T}_{\vec{x},12j-11}, \dots, \overline{\Delta T}_{\vec{x},12j})$ , where each month is weighted by its number of days. February is given a weight of 28 days for all years since CCSM4 does not contain leap years. To calculate EOFs of these sea ice detrended annual average temperature anomalies, the

weighted observation matrix  $O = \begin{bmatrix} Area_{\vec{x}_1}^{0.5} \overline{\Delta T}_{\vec{x}_1,1} & \cdots & Area_{\vec{x}_G}^{0.5} \overline{\Delta T}_{\vec{x}_G,1} \\ \vdots & \ddots & \vdots \\ Area_{\vec{x}_1}^{0.5} \overline{\Delta T}_{\vec{x}_1,\frac{M}{12}} & \cdots & Area_{\vec{x}_G}^{0.5} \overline{\Delta T}_{\vec{x}_G,\frac{M}{12}} \end{bmatrix}$  is calculated, where  $Area_{\vec{x}}$  is

the proportion of the Earth's surface covered by grid cell  $\vec{x}$ . Observations are weighted by the square roots of the surface areas of their grid cells because this is standard practice and prevents the smaller grid cells near the poles from having disproportionate weight. Since  $\sum_{j=1}^{M/12} \overline{\Delta T}_{\vec{x},j} = 0$  for all grid cells  $\vec{x}$ , the EOFs can be calculated as the eigenvectors of the uncorrected sample covariance matrix,  $\frac{12}{M} O' O$ . The EOFs are ranked in descending order by eigenvalue to find the EOFs that contribute the most to explaining the area-weighted variance of  $\overline{\Delta T}$ . In addition, the EOFs are normalized to have maximum values of 1. The EOF that explains the most variance is used as the IVP to represent ENSO.

## Temperature Anomaly Model

### Functional Forms of the Temperature Anomaly Model

Definitions of the models of the amplification function are given by Table S4. Definitions of the models of the covariance function of the weather residuals are given by Table S5. In Table S4, the convention  $m = 1$  corresponds to January, and the normalization operators  $\langle \rangle_m$  and  $\langle \rangle$  are used.  $\langle \rangle_m$  is the operator that returns the function minus its weighted mean, where grid subcells are weighted by their surface area for calendar month  $m$ .  $\langle \rangle$  is the operator that returns the function minus its weighted mean, where grid cells are weighted by their surface area. These normalization operators ensure that the amplification function averages to one when integrated over the surface of the Earth for each calendar month. If the amplification function were normalized to a different non-zero real number, then this would not affect the final temperature estimates since the change in the definition of  $A$  would be offset by a change in the estimates of  $\theta$ . However, having the amplification function be normalized to one allows for a more intuitive interpretation of results.

Model	$F_{\vec{x},s,m} f$
1	0
2	$\langle s \rangle_m f_m$
3	$\alpha_{1,\vec{x}} f_m + \alpha_{2,\vec{x}} f_{12+m}$
4	$\alpha_{1,\vec{x}} f_m + \alpha_{2,\vec{x}} f_{12+m} + \langle s \rangle_m f_{24+m} + \langle s \cdot \alpha_{1,\vec{x}} \rangle_m f_{36+m} + \langle s \cdot \alpha_{2,\vec{x}} \rangle_m f_{48+m}$

**Table S4:** The different models of the amplification function.  $\alpha_{1,\vec{x}} = \varphi_{\vec{x}}^3 - 3 \frac{\pi^2}{4} \varphi_{\vec{x}}$  and  $\alpha_{2,\vec{x}} = \langle \varphi_{\vec{x}}^4 - \frac{\pi^2}{2} \varphi_{\vec{x}}^2 \rangle$ , where  $\varphi_{\vec{x}}$  is the latitude of  $\vec{x}$  in radians. These models were used to estimate temperature anomalies.

Model	$v_{\vec{x},s,m} \mathbf{z}$
1	$\mathbf{z}_1$
2	$\mathbf{z}_1 + \mathbf{s} \cdot \mathbf{z}_2$

**Table S5:** The different models of the covariance function of weather residuals. These models were used to estimate temperature anomalies.

### Notation for the Temperature Anomaly Model

For convenience, the following notation is used:

- ' is the transpose operator

-  $\circ$  is the Hadamard product

- $\otimes$  is the Kronecker product
- $\mathbb{O}_{a,b}$  is the zeros matrix of size  $a \times b$
- $tr$  is the trace operator
- $sum$  is the operator that returns the sum of the elements of a matrix
- $c$  is the operator that returns the lower triangular Cholesky matrix
- $c'$  is the operator that returns the upper triangular Cholesky matrix
- $\lambda$  is the standard logistic function
- $M$  is the number of months in the temperature anomaly dataset
- The convention  $t = 1$  corresponds to January 1850 is used
- $I_a$  is the identity matrix of size  $a \times a$
- $\mathbb{1}_{a,b}$  is the ones matrix of size  $a \times b$
- $vec$  is the vectorization operator
- $G = 36 \cdot 72$  is the number of grid cells

Let  $\{(\tilde{x}_1, s_1), \dots, (\tilde{x}_{2G}, s_{2G})\}$  be an ordering of the set of all grid subcells. Let  $g = \begin{bmatrix} k \\ z \end{bmatrix}$ ,  $\rho = \lambda(r)$ ,

$$\kappa_1 = \lambda(k_1), \kappa_2 = \lambda(k_2), R = \begin{bmatrix} \rho^{|1-1|} & \dots & \rho^{|1-M|} \\ \vdots & \ddots & \vdots \\ \rho^{|M-1|} & \dots & \rho^{|M-M|} \end{bmatrix}, \text{ and } \tilde{K} \text{ be the } 2G \times 2G \text{ matrix with entries}$$

$\tilde{K}_{i,j} = \kappa_1 \|\tilde{x}_i - \tilde{x}_j\| \kappa_2^{|s_i - s_j|} \forall i, j \in \{1, \dots, 2G\}$ . To account for missing observations,  $\forall t \in \{1, \dots, M\}$  let  $n_t$  be the number of observed grid subcells for month  $t$ ,  $N = \sum_{t=1}^M n_t$  be the number of temperature anomaly observations, and define  $\eta_t$  as the  $2G \times n_t$  restriction matrix constructed by starting with  $I_{2G}$  and then  $\forall i \in \{1, \dots, 2G\}$  removing the  $i^{\text{th}}$  column from  $I_{2G}$  if  $(\tilde{x}_i, s_i)$  is unobserved for month  $t$ .

$$\forall i, j \in \{1, \dots, M\} \text{ let } K_{i,j} = \eta_i' \tilde{K} \eta_j, W_i = \begin{bmatrix} W_{\tilde{x}_1, s_1, i} \\ \vdots \\ W_{\tilde{x}_{2G}, s_{2G}, i} \end{bmatrix}, E_i = \eta_i' \begin{bmatrix} E_{\tilde{x}_1, s_1, i} \\ \vdots \\ E_{\tilde{x}_{2G}, s_{2G}, i} \end{bmatrix}, U_i = \eta_i' W_i + E_i,$$

$$\Lambda_{i,j} = cov(W_i, W_j), \Gamma_{i,j} = cov(E_i, E_j), \Omega_{i,j} = cov(U_i, U_j) = \eta_i' \Lambda_{i,j} \eta_j + \Gamma_{i,j}, S_{i,j} = \frac{\Omega_{i,j}}{R_{i,j}},$$

$$Y_i = \eta_i' \begin{bmatrix} Y_{\tilde{x}_1, s_1, i} \\ \vdots \\ Y_{\tilde{x}_{2G}, s_{2G}, i} \end{bmatrix}, Q_i = \eta_i' \begin{bmatrix} Q_{\tilde{x}_1} \\ \vdots \\ Q_{\tilde{x}_{2G}} \end{bmatrix}, F_i = \eta_i' \begin{bmatrix} F_{\tilde{x}_1, s_1, m_i} \\ \vdots \\ F_{\tilde{x}_{2G}, s_{2G}, m_i} \end{bmatrix}, F_{i,a} \text{ be the } a^{\text{th}} \text{ column of } F_i \forall a \in \{1, \dots, n_f\},$$

$$A_i = \mathbb{1}_{n_i, 1} + F_i f, P_i = [A_i \quad Q_i], v_{i,a} = \eta_i' \begin{bmatrix} \left(\frac{1}{2} v_{\tilde{x}_1, s_1, m_i}\right)_a & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \left(\frac{1}{2} v_{\tilde{x}_{2G}, s_{2G}, m_i}\right)_a \end{bmatrix} \eta_i \forall a \in \{1, \dots, n_z\},$$

$$\tilde{V}_i = \begin{bmatrix} \exp\left(\frac{1}{2} v_{\tilde{x}_1, s_1, m_i} z\right) & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \exp\left(\frac{1}{2} v_{\tilde{x}_{2G}, s_{2G}, m_i} z\right) \end{bmatrix}, p_i = \begin{bmatrix} \theta_i \\ q_i \end{bmatrix}, V_i = \eta_i' \tilde{V}_i \eta_i, \text{ and } Z_i = V_i^{-1} K_i V_i^{-1}.$$

In addition, let  $\eta = \begin{bmatrix} \eta_1 & \mathbb{O}_{2G,n_1} & \cdots & \mathbb{O}_{2G,n_1} \\ \mathbb{O}_{2G,n_2} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbb{O}_{2G,n_{M-1}} \\ \mathbb{O}_{2G,n_M} & \cdots & \mathbb{O}_{2G,n_M} & \eta_M \end{bmatrix}$ ,  $W = \begin{bmatrix} W_1 \\ \vdots \\ W_M \end{bmatrix}$ ,  $E = \begin{bmatrix} E_1 \\ \vdots \\ E_M \end{bmatrix}$ ,  $U = \begin{bmatrix} U_1 \\ \vdots \\ U_M \end{bmatrix}$ ,

$\Lambda = \begin{bmatrix} \Lambda_{1,1} & \cdots & \Lambda_{1,M} \\ \vdots & \ddots & \vdots \\ \Lambda_{M,1} & \cdots & \Lambda_{M,M} \end{bmatrix}$ ,  $\Gamma = \begin{bmatrix} \Gamma_{1,1} & \cdots & \Gamma_{1,M} \\ \vdots & \ddots & \vdots \\ \Gamma_{M,1} & \cdots & \Gamma_{M,M} \end{bmatrix}$ ,  $\Omega = \eta' \Lambda \eta + \Gamma$ ,  $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_M \end{bmatrix}$ ,

$P = \begin{bmatrix} P_1 & \mathbb{O}_{n_1,n_q+1} & \cdots & \mathbb{O}_{n_1,n_q+1} \\ \mathbb{O}_{n_2,n_q+1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbb{O}_{n_{M-1},n_q+1} \\ \mathbb{O}_{n_M,n_q+1} & \cdots & \mathbb{O}_{n_M,n_q+1} & P_M \end{bmatrix}$ ,  $\tilde{V} = \begin{bmatrix} \tilde{V}_1 & \mathbb{O}_{2G,2G} & \cdots & \mathbb{O}_{2G,2G} \\ \mathbb{O}_{2G,2G} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbb{O}_{2G,2G} \\ \mathbb{O}_{2G,2G} & \cdots & \mathbb{O}_{2G,2G} & \tilde{V}_M \end{bmatrix}$ ,

$q = \begin{bmatrix} q_1 \\ \vdots \\ q_M \end{bmatrix}$ ,  $p = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}$ , and  $V = \eta' \tilde{V} \eta$ . For convenience,  $\forall i \in \{1, \dots, M\}$  let  $K_i = K_{i,i}$ ,  $\Lambda_i = \Lambda_{i,i}$ ,  $\Gamma_i = \Gamma_{i,i}$ ,  $\Omega_i = \Omega_{i,i}$ , and  $S_i = S_{i,i}$ .

As the bias uncertainties have a complicated covariance structure, do not have a multivariate normal distribution, and have significant temporal correlation, only measurement and sampling uncertainties are taken into account in  $E$ . Instead, bias uncertainties are taken into account later on in the estimation of total uncertainty. Since the measurement and sampling uncertainties of HadCRUT4 and HadSST4 neglect temporal correlations between months,  $\Gamma$  is a block diagonal matrix.

#### Approximating the Log-Likelihood of the Temperature Anomaly Model

The log-likelihood of the temperature anomaly model is  $l = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln|\Omega| - \frac{1}{2} U' \Omega^{-1} U$ .  $\Omega$  is a matrix with more than 10 trillion numbers and, as a result, would take up more than 80 TB of memory. Given current personal computer technology, it is not feasible to compute, store, or perform computations with such a large matrix, so  $\Omega$  needs to be approximated and broken down into smaller matrices.

The second order Taylor approximation of  $|\Omega|$  around  $\rho = 0$  is  $|\Omega|_{\rho=0} + \frac{\partial |\Omega|}{\partial \rho} \Big|_{\rho=0} \rho + \frac{\partial^2 |\Omega|}{\partial \rho^2} \Big|_{\rho=0} \frac{\rho^2}{2}$

$= |\Omega|_{\rho=0} + |\Omega| \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \rho} \right) \Big|_{\rho=0} \rho + \left( |\Omega| \text{tr}^2 \left( \Omega^{-1} \frac{\partial \Omega}{\partial \rho} \right) + |\Omega| \text{tr} \left( \Omega^{-1} \frac{\partial^2 \Omega}{\partial \rho^2} - \Omega^{-1} \frac{\partial \Omega}{\partial \rho} \Omega^{-1} \frac{\partial \Omega}{\partial \rho} \right) \right) \Big|_{\rho=0} \frac{\rho^2}{2}$



$$\begin{aligned}
&= |\Omega| \Big|_{\rho=0} \left( 1 - \frac{\rho^2}{2} \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \rho} \Omega^{-1} \frac{\partial \Omega}{\partial \rho} \right) \right) \Big|_{\rho=0} = \left( \prod_{i=1}^M |S_i| \right) \left( 1 - \frac{\rho^2}{2} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M \text{tr}(S_i^{-1} S_{i,j} S_j^{-1} S_{j,i}) \right) \\
&= \left( \prod_{i=1}^M |S_i| \right) \left( 1 - \frac{\rho^2}{2} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M \text{sum}(S_{i,j}^* \circ S_{i,j}) \right), \text{ where } S_{i,j}^* = S_i^{-1} S_{i,j} S_j^{-1} \forall i, j \in \{1, \dots, M\}.
\end{aligned}$$

In the special case where  $\eta_i = \eta_{i+1} \forall i \in \{1, \dots, M-1\}$  and  $E_i = \mathbb{O}_{n_i,1} \forall i \in \{1, \dots, M\}$ , one can calculate

$$|\Omega| \text{ exactly since, in this case, } \Omega = \begin{bmatrix} R_{1,1} V_1^{-1} K_{1,1} V_1^{-1} & \dots & R_{1,M} V_1^{-1} K_{1,M} V_M^{-1} \\ \vdots & \ddots & \vdots \\ R_{M,1} V_M^{-1} K_{M,1} V_1^{-1} & \dots & R_{M,M} V_M^{-1} K_{M,M} V_M^{-1} \end{bmatrix} = V^{-1} (R \otimes K_1) V^{-1}.$$

Thus  $|\Omega| = |V^{-1} (R \otimes K_1) V^{-1}| = |V^{-1}| |R|^{n_1} |K_1|^M |V^{-1}|$ .

But  $R$  is the Kac-Murdock-Szegő matrix, which has determinant  $|R| = (1 - \rho^2)^{M-1}$ .

$$\Rightarrow |\Omega| = |V^{-1}| (1 - \rho^2)^{n_1(M-1)} |K_1|^M |V^{-1}|$$

$$\begin{aligned}
\text{But } n_1(M-1) &= \frac{1}{2} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M n_i = \frac{1}{2} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M \text{tr}(I_{n_i}) \\
&= \frac{1}{2} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M \text{tr}(S_i^{-1} S_{i,j} S_j^{-1} S_{j,i}) = \frac{1}{2} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M \text{sum}(S_{i,j}^* \circ S_{i,j}) \text{ and}
\end{aligned}$$

$$|V^{-1}| |K_1|^M |V^{-1}| = \left( \prod_{i=1}^M |V_i^{-1}| \right) \left( \prod_{i=1}^M |K_i| \right) \left( \prod_{i=1}^M |V_i^{-1}| \right) = \left( \prod_{i=1}^M |S_i| \right).$$

$$\text{Thus } |\Omega| = \left( \prod_{i=1}^M |S_i| \right) (1 - \rho^2)^{\frac{1}{2} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M \text{sum}(S_{i,j}^* \circ S_{i,j})}.$$

$$\text{But } \frac{\partial}{\partial \rho} (1 - \rho^2)^{n_1(M-1)} \Big|_{\rho=0} = n_1(M-1)(1 - \rho^2)^{n_1(M-1)-1} (-2\rho) \Big|_{\rho=0} = 0 \text{ and}$$

$$\frac{\partial^2}{\partial \rho^2} (1 - \rho^2)^{n_1(M-1)} \Big|_{\rho=0} = \frac{4\rho^2 n_1(M-1)(n_1(M-1)-1)(1 - \rho^2)^{n_1(M-1)-2}}{-2n_1(M-1)(1 - \rho^2)^{n_1(M-1)-1}} \Big|_{\rho=0} = -2n_1(M-1).$$

Thus, the second order Taylor approximation of the above formula for  $|\Omega|$  around  $\rho = 0$  in the special

$$\text{case where } E_i = \mathbb{O}_{n_i,1} \forall i \in \{1, \dots, 12\} \text{ is } |\Omega| \Big|_{\rho=0} + \frac{\partial |\Omega|}{\partial \rho} \Big|_{\rho=0} \rho + \frac{\partial^2 |\Omega|}{\partial \rho^2} \Big|_{\rho=0} \frac{\rho^2}{2} =$$

$$\prod_{i=1}^{12} |S_i| (1 - \rho^2 n_1(M-1)) = \prod_{i=1}^{12} |S_i| \left( 1 - \frac{\rho^2}{2} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M \text{sum}(S_{i,j}^* \circ S_{i,j}) \right), \text{ which is the same as}$$

the second order Taylor approximation of  $|\Omega|$  around  $\rho = 0$  in the general case.

Since the above formula for  $|\Omega|$  in the special case has the same second order Taylor approximation around  $\rho = 0$  as  $|\Omega|$  does in the general case, it is used as the approximation of  $|\Omega|$ .

$$\Rightarrow \ln(|\Omega|) \approx \sum_{i=1}^M \ln(|S_i|) + \frac{1}{2} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M \text{sum}(S_{i,j}^* \circ S_{i,j}) \ln(1 - \rho^2)$$

The first order Taylor approximation of  $\Omega^{-1}$  around  $\rho = 0$  is

$$\Omega^{-1}|_{\rho=0} + \frac{\partial \Omega^{-1}}{\partial \rho} \Big|_{\rho=0} \rho = \Omega^{-1}|_{\rho=0} - \Omega^{-1} \frac{\partial \Omega}{\partial \rho} \Omega^{-1} \Big|_{\rho=0} \rho =$$

$$\begin{bmatrix} S_{1,1}^{-1} & -\rho S_{1,1}^{-1} S_{1,2} S_{2,2}^{-1} & \mathbb{O}_{n_1, n_3} & \cdots & \mathbb{O}_{n_1, n_M} \\ -\rho S_{2,2}^{-1} S_{2,1} S_{1,1}^{-1} & \ddots & \ddots & \ddots & \vdots \\ \mathbb{O}_{n_3, n_1} & \ddots & \ddots & \ddots & \mathbb{O}_{n_{M-2}, n_M} \\ \vdots & \ddots & \ddots & \ddots & -\rho S_{M-1, M-1}^{-1} S_{M-1, M} S_{M, M}^{-1} \\ \mathbb{O}_{n_M, n_1} & \cdots & \mathbb{O}_{n_M, n_{M-2}} & -\rho S_{M, M}^{-1} S_{M, M-1} S_{M-1, M-1}^{-1} & S_{M, M}^{-1} \end{bmatrix}.$$

In the special case where  $\eta_i = \eta_{i+1} \forall i \in \{1, \dots, M-1\}$  and  $E_i = \mathbb{O}_{n_i, 1} \forall i \in \{1, \dots, M\}$ , one can calculate  $\Omega^{-1}$  exactly since, in this case,  $\Omega^{-1} = (V^{-1}(R \otimes K_1)V)^{-1} = V(R^{-1} \otimes K_1^{-1})V$ . But  $R$  is the Kac-

Murdock-Szegő matrix, which has inverse  $R^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1+\rho^2 & -\rho \\ 0 & \cdots & 0 & -\rho & 1 \end{bmatrix}.$

$$\Rightarrow \Omega^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} V_1 K_1^{-1} V_1 & -\rho V_1 K_1^{-1} V_2 & \mathbb{O}_{n_1, n_1} & \cdots & \mathbb{O}_{n_1, n_M} \\ -\rho V_2 K_1^{-1} V_1 & (1+\rho^2) V_2 K_1^{-1} V_2 & \ddots & \ddots & \vdots \\ \mathbb{O}_{n_1, n_1} & \ddots & \ddots & \ddots & \mathbb{O}_{n_1, n_1} \\ \vdots & \ddots & \ddots & (1+\rho^2) V_{M-1} K_1^{-1} V_{M-1} & -\rho V_{M-1} K_1^{-1} V_M \\ \mathbb{O}_{n_1, n_1} & \cdots & \mathbb{O}_{n_1, n_1} & -\rho V_M K_1^{-1} V_{M-1} & V_M K_1^{-1} V_M \end{bmatrix}$$

$$= \frac{1}{1-\rho^2} \begin{bmatrix} S_{1,1}^{-1} & -\rho S_{1,2}^* & \mathbb{O}_{n_1, n_3} & \cdots & \mathbb{O}_{n_1, n_M} \\ -\rho S_{2,1}^* & (1+\rho^2) S_{2,2}^{-1} & \ddots & \ddots & \vdots \\ \mathbb{O}_{n_3, n_1} & \ddots & \ddots & \ddots & \mathbb{O}_{n_{M-2}, n_M} \\ \vdots & \ddots & \ddots & (1+\rho^2) S_{M-1, M-1}^{-1} & -\rho S_{M-1, M}^* \\ \mathbb{O}_{n_M, n_1} & \cdots & \mathbb{O}_{n_M, n_{M-2}} & -\rho S_{M, M-1}^* & S_{M, M}^{-1} \end{bmatrix}.$$

But  $\frac{\partial}{\partial \rho} \frac{1}{1-\rho^2} \Big|_{\rho=0} = \frac{2\rho}{(1-\rho^2)^2} \Big|_{\rho=0} = 0$ ,  $\frac{\partial}{\partial \rho} \frac{1+\rho^2}{1-\rho^2} \Big|_{\rho=0} = \frac{4\rho}{(1-\rho^2)^2} \Big|_{\rho=0} = 0$ , and

$\frac{\partial^2}{\partial \rho^2} \frac{-\rho}{1-\rho^2} \Big|_{\rho=0} = -\frac{1+\rho^2}{(1-\rho^2)^2} \Big|_{\rho=0} = -1$ . Thus, the first order Taylor approximation of the above formula for

$\Omega^{-1}$  around  $\rho = 0$  in the special case where  $E_i = \mathbb{O}_{n_i, 1} \forall i \in \{1, \dots, 12\}$  is  $\Omega^{-1}|_{\rho=0} + \frac{\partial \Omega^{-1}}{\partial \rho} \Big|_{\rho=0} \rho$

$$= \begin{bmatrix} S_{1,1}^{-1} & -\rho S_{1,2}^* & \mathbb{O}_{n_1, n_3} & \cdots & \mathbb{O}_{n_1, n_M} \\ -\rho S_{2,1}^* & \ddots & \ddots & \ddots & \vdots \\ \mathbb{O}_{n_3, n_1} & \ddots & \ddots & \ddots & \mathbb{O}_{n_{M-2}, n_M} \\ \vdots & \ddots & \ddots & \ddots & -\rho S_{M-1, M}^* \\ \mathbb{O}_{n_M, n_1} & \cdots & \mathbb{O}_{n_M, n_{M-2}} & -\rho S_{M, M-1}^* & S_{M, M}^{-1} \end{bmatrix}, \text{ which is the same as the first order Taylor}$$

approximation of  $\Omega^{-1}$  around  $\rho = 0$  in the general case.

Since the above formula for  $\Omega^{-1}$  in the special case has the same first order Taylor approximation around  $\rho = 0$  as  $\Omega^{-1}$  does in the general case, it is used as the approximation of  $\Omega^{-1}$ . For convenience, let  $\Omega^*$  correspond to this approximation of  $\Omega^{-1}$ .

$$\Rightarrow U' \Omega^{-1} U \approx \sum_{i=1}^M (R^{-1})_{i,i} U'_i S_i^{-1} U_i + \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} U'_i S_{i,j}^* U_j,$$

$$\text{where } (R^{-1})_{i,j} = \begin{cases} \frac{1}{1-\rho^2} & \text{if } i = j \in \{1, M\} \\ \frac{1+\rho^2}{1-\rho^2} & i = j \in \{2, \dots, M-1\} \\ \frac{-\rho}{1-\rho^2} & |i-j| = 1 \\ 0 & \text{else} \end{cases}.$$

$$\begin{aligned} \Rightarrow l &= -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln|\Omega| - \frac{1}{2} U' \Omega^{-1} U \approx -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^M \ln|S_i| \\ &- \frac{\ln(1-\rho^2)}{4} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M \text{sum}(S_{i,j}^* \circ S_{i,j}) - \frac{1}{2} \sum_{i=1}^M (R^{-1})_{i,i} U'_i S_i^{-1} U_i - \frac{1}{2} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} U'_i S_{i,j}^* U_j \end{aligned}$$

### Initial Guess of the Temperature Anomaly Model

To create an initial guess of  $(f, r, g, p)$ , the following method is used:

- $f$  is initially estimated as a zero vector.
- $p$  is initially estimated using OLS regression:  $Y_{\vec{x},s,t} = \theta_t + Q_{\vec{x}} q_t + U_{\vec{x},s,t}$ .
- The residuals  $U$  of the OLS regression are obtained.
- $z_2, \dots, z_{n_z}$  are initially estimated as zero.
- If  $z_2, \dots, z_{n_z}$  are zero, then the expected value of  $U_{\vec{x},s,t}^2$  is  $e^{-z_1} + \sigma_{E_{\vec{x},s,t}}^2 \forall (\vec{x}, s, t) \in \mathcal{E}$ , where  $\mathcal{E}$  is the set of all combinations of grid subcells and months that have observations and  $\sigma_{E_{\vec{x},s,t}}^2$  is the variance of the measurement and sampling errors at  $(\vec{x}, s)$  for month  $t$ . As a result,  $z_1$  is initially estimated as  $-\ln\left(\left(\sum_{(\vec{x},s,t) \in \mathcal{E}} \left(U_{\vec{x},s,t}^2 - \sigma_{E_{\vec{x},s,t}}^2\right)\right) / \sum_{(\vec{x},s,t) \in \mathcal{E}} 1\right)$ .
- $r$  is initially estimated by comparing residuals  $U$  within the same grid subcell and in different and consecutive months.  $r$  is estimated as

$$\lambda^{-1} \left( \sum_{(\vec{x},s,t_i) \in \mathcal{E}} \sum_{\substack{(\vec{x},s,t_j) \in \mathcal{E} \\ |t_i-t_j|=1}} \frac{U_{\vec{x},s,t_i} U_{\vec{x},s,t_j}}{\exp(-0.5(v_{\vec{x},s,t_i} + v_{\vec{x},s,t_j})z)} \right) / \sum_{(\vec{x},s,t_i) \in \mathcal{E}} \sum_{\substack{(\vec{x},s,t_j) \in \mathcal{E} \\ |t_i-t_j|=1}} 1.$$

- $k_1$  is initially estimated by comparing residuals  $U$  within the same surface type, within the same month, within the same longitudinal band, and in different latitudinal bands.  $k_1$  is estimated as

$$\lambda^{-1} \left( \left( \sum_{(\tilde{x}_i, s, t) \in \mathcal{E}} \sum_{\substack{(\tilde{x}_j, s, t) \in \mathcal{E} \\ \psi_{\tilde{x}_i} = \psi_{\tilde{x}_j} \\ |\varphi_{\tilde{x}_i} - \varphi_{\tilde{x}_j}| = \frac{5\pi}{180}}} \frac{U_{\tilde{x}_i, s, t} U_{\tilde{x}_j, s, t}}{\exp(-0.5(v_{\tilde{x}_i, s, t} + v_{\tilde{x}_j, s, t})z)} \right) / \left( \sum_{(\tilde{x}_i, s, t) \in \mathcal{E}} \sum_{\substack{(\tilde{x}_j, s, t) \in \mathcal{E} \\ \psi_{\tilde{x}_i} = \psi_{\tilde{x}_j} \\ |\varphi_{\tilde{x}_i} - \varphi_{\tilde{x}_j}| = \frac{5\pi}{180}}} 1 \right)^{\frac{180}{5\pi}} \right).$$

- $k_2$  is initially estimated by comparing residuals  $U$  within the same grid cell, within the same month, and in different grid subcells.  $k_2$  is estimated as

$$\lambda^{-1} \left( \sum_{(\tilde{x}, s, t) \in \mathcal{E}} \sum_{\substack{(\tilde{x}, s, j, t) \in \mathcal{E} \\ s_i \neq s_j}} \frac{U_{\tilde{x}, s, i, t} U_{\tilde{x}, s, j, t}}{\exp(-0.5(v_{\tilde{x}, s, i, t} + v_{\tilde{x}, s, j, t})z)} / \sum_{(\tilde{x}, s, i, t) \in \mathcal{E}} \sum_{\substack{(\tilde{x}, s, j, t) \in \mathcal{E} \\ s_i \neq s_j}} 1 \right).$$

- $p$  is re-estimated using GLS regression:  $Y_{\tilde{x}, s, t} = (1 + F_{\tilde{x}, s, m_t} f) \theta_t + Q_{\tilde{x}} q_t + W_{\tilde{x}, s, t} + E_{\tilde{x}, s, t}$ . This is done by taking the estimate of  $f$  as given. In particular, the GLS estimate is

$$(P' \Omega^{-1} P)^{-1} P' \Omega^{-1} Y \approx (P' \Omega^* P)^{-1} P' \Omega^* Y =$$

$$\begin{bmatrix} P'_1 S_{1,1}^{-1} P_1 & -\rho P'_1 S_{1,2}^* P_2 & \mathbb{O}_{n_q+1, n_q+1} & \cdots & \mathbb{O}_{n_q+1, n_q+1} \\ -\rho P'_2 S_{2,1}^* P_1 & (1 + \rho^2) P'_2 S_{2,2}^* P_2 & \ddots & \ddots & \vdots \\ \mathbb{O}_{n_q+1, n_q+1} & \ddots & \ddots & \ddots & \mathbb{O}_{n_q+1, n_q+1} \\ \vdots & \ddots & \ddots & (1 + \rho^2) P'_{M-1} S_{M-1, M-1}^* P_{M-1} & -\rho P'_{M-1} S_{M-1, M}^* P_M \\ \mathbb{O}_{n_q+1, n_q+1} & \cdots & \mathbb{O}_{n_q+1, n_q+1} & -\rho P'_M S_{M, M-1}^* P_{M-1} & P'_M S_{M, M}^{-1} P_M \end{bmatrix}^{-1}$$

$$(1 - \rho^2) \begin{bmatrix} (R^{-1})_{1,1} P'_1 S_{1,1}^{-1} Y_1 + \sum_{\substack{j=1 \\ |1-j|=1}}^M (R^{-1})_{1,j} P'_1 S_{1,j}^* Y_j \\ \vdots \\ (R^{-1})_{M,M} P'_M S_{M,M}^{-1} Y_M + \sum_{\substack{j=1 \\ |M-j|=1}}^M (R^{-1})_{M,j} P'_M S_{M,j}^* Y_j \end{bmatrix}.$$

### Estimating the Temperature Anomaly Model

The first and second derivatives of the approximation to the log-likelihood function are

- $\frac{\partial l}{\partial f_a} = U' \Omega^{-1} \frac{\partial P}{\partial f_a} p$   
 $\approx \sum_{i=1}^M (R^{-1})_{i,i} U'_i S_i^{-1} F_{i,a} \theta_i + \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} U'_i S_{i,j}^* F_{j,a} \theta_j \quad \forall a \in \{1, \dots, n_f\}$

- $$\frac{\partial l}{\partial r} = -\frac{1}{2} \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial r} \right) - \frac{1}{2} U' \frac{\partial \Omega^{-1}}{\partial r} U \approx \frac{\rho^2}{2(1+\rho)} \sum_{i=1}^M \sum_{j=1}^M \sum_{|i-j|=1} \text{sum}(S_{i,j}^* \circ S_{i,j})$$

$$- \frac{1}{2} \sum_{i=1}^M \left( \frac{\partial R^{-1}}{\partial r} \right)_{i,i} U_i' S_i^{-1} U_i - \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M \sum_{|i-j|=1} \left( \frac{\partial R^{-1}}{\partial r} \right)_{i,j} U_i' S_{i,j}^* U_j,$$

where  $\left( \frac{\partial R^{-1}}{\partial r} \right)_{i,j} = \left( \frac{\partial R^{-1}}{\partial \rho} \rho(1-\rho) \right)_{i,j} = \begin{cases} \frac{2\rho^2}{(1+\rho)(1-\rho^2)} & \text{if } i = j \in \{1, M\} \\ \frac{4\rho^2}{(1+\rho)(1-\rho^2)} & i = j \in \{2, \dots, M-1\} \\ \frac{-\rho(1+\rho^2)}{(1+\rho)(1-\rho^2)} & |i-j| = 1 \\ 0 & \text{else} \end{cases}.$
- $$\frac{\partial l}{\partial g_a} = -\frac{1}{2} \ln \left| \Omega^{-1} \frac{\partial \Omega}{\partial g_a} \right| - \frac{1}{2} U' \frac{\partial \Omega^{-1}}{\partial g_a} U$$

$$\approx -\frac{1}{2} \sum_{i=1}^M \text{tr} \left( S_i^{-1} \frac{\partial S_i}{\partial g_a} \right) - \frac{\ln(1-\rho^2)}{4} \sum_{i=1}^M \sum_{j=1}^M \sum_{|i-j|=1} \text{sum} \left( \frac{\partial S_{i,j}^*}{\partial g_a} \circ S_{i,j} + S_{i,j}^* \circ \frac{\partial S_{i,j}}{\partial g_a} \right)$$

$$- \frac{1}{2} \sum_{i=1}^M (R^{-1})_{i,i} U_i' \frac{\partial S_i^{-1}}{\partial g_a} U_i - \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M \sum_{|i-j|=1} (R^{-1})_{i,j} U_i' \frac{\partial S_{i,j}^*}{\partial g_a} U_j \quad \forall a \in \{1, \dots, n_z + 2\}, \text{ where}$$

$$\frac{\partial S_{i,j}}{\partial g_a} = V_i^{-1} \frac{\partial K_{i,j}}{\partial \kappa_a} V_j^{-1} \kappa_a (1 - \kappa_a) \quad \forall a \in \{1, 2\}, \quad \frac{\partial S_i}{\partial g_a} = -v_{i,a-2} Z_i - Z_i v_{i,a-2} \quad \forall a \in \{3, \dots, n_z + 2\},$$

$$\frac{\partial S_{i,j}}{\partial g_a} = -v_{i,a-2} S_{i,j} - S_{i,j} v_{j,a-2} \quad \forall a \in \{3, \dots, n_z + 2\} \text{ if } i \neq j, \quad \frac{\partial S_i^{-1}}{\partial g_a} = -S_i^{-1} \frac{\partial S_i}{\partial g_a} S_i^{-1}, \text{ and}$$

$$\frac{\partial S_{i,j}^*}{\partial g_a} = S_i^{-1} \frac{\partial S_{i,j}}{\partial g_a} S_j^{-1} - S_i^{-1} \frac{\partial S_i}{\partial g_a} S_{i,j}^* - S_{i,j}^* \frac{\partial S_j}{\partial g_a} S_j^{-1} \text{ if } i \neq j.$$
- $$\frac{\partial l}{\partial p_i} = U' \Omega^{-1} P \approx (R^{-1})_{i,i} U_i' S_i^{-1} P_i + \sum_{j=1}^M \sum_{|i-j|=1} (R^{-1})_{i,j} U_j' S_{j,i}^* P_i \quad \forall i \in \{1, \dots, M\}$$
- $$\frac{\partial^2 l}{\partial f_a \partial f_b} = -p' \frac{\partial P'}{\partial f_a} \Omega^{-1} \frac{\partial P}{\partial f_b} p$$

$$\approx -\sum_{i=1}^M (R^{-1})_{i,i} \theta_i F_{i,a}' S_i^{-1} F_{i,b} \theta_i - \sum_{i=1}^M \sum_{j=1}^M \sum_{|i-j|=1} (R^{-1})_{i,j} \theta_i F_{i,a}' S_{i,j}^* F_{j,b} \theta_j \quad \forall a, b \in \{1, \dots, n_f\}$$
- $$\frac{\partial^2 l}{\partial f_a \partial r} = p' \frac{\partial P'}{\partial f_a} \frac{\partial \Omega^{-1}}{\partial r} U$$

$$\approx \sum_{i=1}^M \left( \frac{\partial R^{-1}}{\partial r} \right)_{i,i} \theta_i F_i' S_i^{-1} U_i + \sum_{i=1}^M \sum_{j=1}^M \sum_{|i-j|=1} \left( \frac{\partial R^{-1}}{\partial r} \right)_{i,j} \theta_i F_i' S_{i,j}^* U_j \quad \forall a \in \{1, \dots, n_f\}$$
- $$\frac{\partial^2 l}{\partial f_a \partial g_b} = p' \frac{\partial P'}{\partial f_a} \frac{\partial \Omega^{-1}}{\partial g_b} U \approx \sum_{i=1}^M (R^{-1})_{i,i} \theta_i F_{i,a}' \frac{\partial S_i^{-1}}{\partial g_b} U_i + \sum_{i=1}^M \sum_{j=1}^M \sum_{|i-j|=1} (R^{-1})_{i,j} \theta_i F_{i,a}' \frac{\partial S_{i,j}^*}{\partial g_b} U_j$$

$$\quad \forall a \in \{1, \dots, n_f\} \quad \forall b \in \{1, \dots, n_z + 2\}$$
- $$\frac{\partial^2 l}{\partial f_a \partial p_i} = U' \Omega^{-1} \frac{\partial P}{\partial f_a} - p' \frac{\partial P'}{\partial f_a} \Omega^{-1} P$$

$$\approx (R^{-1})_{i,i} U_i' S_i^{-1} [F_{i,a} \quad \mathbb{0}_{n_i, n_q}] + \sum_{j=1}^M \sum_{|i-j|=1} (R^{-1})_{i,j} U_j' S_{j,i}^* [F_{j,a} \quad \mathbb{0}_{n_j, n_q}]$$

$$- (R^{-1})_{i,i} \theta_i F_{i,a}' S_i^{-1} P_i - \sum_{j=1}^M \sum_{|i-j|=1} (R^{-1})_{i,j} \theta_j F_{j,a}' S_{j,i}^* P_i \quad \forall a \in \{1, \dots, n_f\} \quad \forall i \in \{1, \dots, M\}$$

- $$\begin{aligned} \frac{\partial^2 l}{\partial r^2} &= -\frac{1}{2} \text{tr} \left( \Omega^{-1} \left( \frac{\partial^2 \Omega}{\partial r^2} - \frac{\partial \Omega}{\partial r} \Omega^{-1} \frac{\partial \Omega}{\partial r} \right) \right) - \frac{1}{2} U' \frac{\partial^2 \Omega^{-1}}{\partial r^2} U \\ &\approx \frac{\rho^2(2+\rho)(1-\rho)}{2(1+\rho)^2} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M \text{sum}(S_{i,j}^* \circ S_{i,j}) - \frac{1}{2} \sum_{i=1}^M \left( \frac{\partial^2 R^{-1}}{\partial r^2} \right)_{i,i} U_i' S_i^{-1} U_i \\ &\quad - \frac{1}{2} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M \left( \frac{\partial^2 R^{-1}}{\partial r^2} \right)_{i,j} U_i' S_{i,j}^* U_j, \end{aligned}$$

where  $\left( \frac{\partial^2 R^{-1}}{\partial r^2} \right)_{i,j} = \left( \frac{\partial^2 R^{-1}}{\partial \rho \partial r} \rho(1-\rho) \right)_{i,j} = \begin{cases} \frac{2\rho^2(2-\rho+\rho^2)}{(1+\rho)^2(1-\rho^2)} & \text{if } i = j \in \{1, M\} \\ \frac{4\rho^2(2-\rho+\rho^2)}{(1+\rho)^2(1-\rho^2)} & i = j \in \{2, \dots, M-1\} \\ \frac{-\rho(1+\rho^2)(1-\rho+5\rho^2-\rho^3)}{(1+\rho)^2(1-\rho^2)} & |i-j| = 1 \\ 0 & \text{else} \end{cases}$
- $$\begin{aligned} \frac{\partial^2 l}{\partial r \partial g_a} &= -\frac{1}{2} \text{tr} \left( \Omega^{-1} \left( \frac{\partial^2 \Omega}{\partial r \partial g_a} - \frac{\partial \Omega}{\partial r} \Omega^{-1} \frac{\partial \Omega}{\partial g_a} \right) \right) - \frac{1}{2} U' \frac{\partial^2 \Omega^{-1}}{\partial r \partial g_a} U \\ &\approx \frac{\rho^2}{2(1+\rho)} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M \text{sum} \left( \frac{\partial S_{i,j}^*}{\partial g_a} \circ S_{i,j} + S_{i,j}^* \circ \frac{\partial S_{i,j}}{\partial g_a} \right) - \frac{1}{2} \sum_{i=1}^M \left( \frac{\partial R^{-1}}{\partial r} \right)_{i,i} U_i' \frac{\partial S_i^{-1}}{\partial g_a} U_i \\ &\quad - \frac{1}{2} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M \left( \frac{\partial R^{-1}}{\partial r} \right)_{i,j} U_i' \frac{\partial S_{i,j}^*}{\partial g_a} U_j \quad \forall a \in \{1, \dots, n_z + 2\} \end{aligned}$$
- $$\frac{\partial^2 l}{\partial r \partial p_i} = U' \frac{\partial \Omega^{-1}}{\partial r} P \approx \left( \frac{\partial R^{-1}}{\partial r} \right)_{i,i} U_i' S_i^{-1} P_i + \sum_{\substack{j=1 \\ |i-j|=1}}^M \left( \frac{\partial R^{-1}}{\partial r} \right)_{i,j} U_j' S_{j,i}^* P_i \quad \forall i \in \{1, \dots, M\}$$
- $$\begin{aligned} \frac{\partial^2 l}{\partial g_a \partial g_b} &= -\frac{1}{2} \text{tr} \left( \Omega^{-1} \left( \frac{\partial^2 \Omega}{\partial g_a \partial g_b} - \frac{\partial \Omega}{\partial g_a} \Omega^{-1} \frac{\partial \Omega}{\partial g_b} \right) \right) - \frac{1}{2} U' \frac{\partial^2 \Omega^{-1}}{\partial g_a \partial g_b} U \\ &\approx -\frac{1}{2} \sum_{i=1}^M \text{sum} \left( S_i^{-1} \circ \frac{\partial^2 S_i}{\partial g_a \partial g_b} - \left( S_i^{-1} \frac{\partial S_i}{\partial g_b} \right) \circ \left( \frac{\partial S_i}{\partial g_a} S_i^{-1} \right) \right) \\ &\quad - \frac{\ln(1-\rho^2)}{4} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M \text{sum} \left( \frac{\partial^2 S_{i,j}^*}{\partial g_a \partial g_b} \circ S_{i,j} + \frac{\partial S_{i,j}^*}{\partial g_a} \circ \frac{\partial S_{i,j}}{\partial g_b} + \frac{\partial S_{i,j}^*}{\partial g_b} \circ \frac{\partial S_{i,j}}{\partial g_a} + S_{i,j}^* \circ \frac{\partial^2 S_{i,j}}{\partial g_a \partial g_b} \right) \\ &\quad - \frac{1}{2} \sum_{i=1}^M (R^{-1})_{i,i} U_i' \frac{\partial^2 S_i^{-1}}{\partial g_a \partial g_b} U_i - \frac{1}{2} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} U_i' \frac{\partial^2 S_{i,j}^*}{\partial g_a \partial g_b} U_j \quad \forall a, b \in \{1, \dots, n_z + 2\}, \end{aligned}$$

where  $\frac{\partial^2 S_{i,j}}{\partial g_a^2} = V_i^{-1} \frac{\partial^2 K_{i,j}}{\partial \kappa_a^2} V_j^{-1} \kappa_a^2 (1 - \kappa_a)^2 + \frac{\partial S_{i,j}}{\partial g_a} (1 - 2\kappa_a) \quad \forall a \in \{1, 2\}$ ,  
 $\frac{\partial^2 S_{i,j}}{\partial g_a \partial g_b} = V_i^{-1} \frac{\partial^2 K_{i,j}}{\partial \kappa_a \partial \kappa_b} V_j^{-1} \kappa_a (1 - \kappa_a) \kappa_b (1 - \kappa_b) \quad \forall a, b \in \{1, 2\} \text{ if } a \neq b$ ,  
 $\frac{\partial^2 S_{i,j}}{\partial g_a \partial g_b} = -v_{i,b-2} \frac{\partial S_{i,j}}{\partial g_a} - \frac{\partial S_{i,j}}{\partial g_a} v_{j,b-2} \quad \forall a \in \{1, \dots, n_z + 2\} \quad \forall b \in \{3, \dots, n_z + 2\}$ ,  
 $\frac{\partial^2 S_i^{-1}}{\partial g_a \partial g_b} = S_i^{-1} \frac{\partial S_i}{\partial g_a} S_i^{-1} \frac{\partial S_i}{\partial g_b} S_i^{-1} + S_i^{-1} \frac{\partial S_i}{\partial g_b} S_i^{-1} \frac{\partial S_i}{\partial g_a} S_i^{-1} - S_i^{-1} \frac{\partial^2 S_i}{\partial g_a \partial g_b} S_i^{-1}$ , and  
 $\frac{\partial^2 S_{i,j}^*}{\partial g_a \partial g_b} = \frac{\partial S_i^{-1}}{\partial g_b} \frac{\partial S_{i,j}}{\partial g_a} S_j^{-1} + S_i^{-1} \frac{\partial^2 S_{i,j}}{\partial g_a \partial g_b} S_j^{-1} + S_i^{-1} \frac{\partial S_{i,j}}{\partial g_a} \frac{\partial S_j^{-1}}{\partial g_b} - \frac{\partial S_i^{-1}}{\partial g_b} \frac{\partial S_i}{\partial g_a} S_{i,j}^* - S_i^{-1} \frac{\partial^2 S_i}{\partial g_a \partial g_b} S_{i,j}^*$   
 $- S_i^{-1} \frac{\partial S_i}{\partial g_a} \frac{\partial S_{i,j}^*}{\partial g_b} - \frac{\partial S_{i,j}^*}{\partial g_b} \frac{\partial S_i}{\partial g_a} S_j^{-1} - S_{i,j}^* \frac{\partial^2 S_j}{\partial g_a \partial g_b} S_j^{-1} - S_{i,j}^* \frac{\partial S_j}{\partial g_a} \frac{\partial S_j^{-1}}{\partial g_b}$ . To reduce the computational

burden of calculating this second derivative, the approximation  $0 = \frac{\partial^2}{\partial g_a \partial g_b} \text{tr}(I_{n_i}) \approx \frac{\partial^2}{\partial g_a \partial g_b} \text{tr}(S_i^{-1} S_{i,j} S_j^{-1} S_{j,i}) = \text{sum} \left( \frac{\partial^2 S_{i,j}^*}{\partial g_a \partial g_b} \circ S_{i,j} + \frac{\partial S_{i,j}^*}{\partial g_a} \circ \frac{\partial S_{i,j}}{\partial g_b} + \frac{\partial S_{i,j}^*}{\partial g_b} \circ \frac{\partial S_{i,j}}{\partial g_a} + S_{i,j}^* \circ \frac{\partial^2 S_{i,j}}{\partial g_a \partial g_b} \right)$  is used.

- $\frac{\partial^2 l}{\partial g_a \partial p_i} = U' \frac{\partial \Omega^{-1}}{\partial g_a} P$   
 $\approx (R^{-1})_{i,i} U_i' \frac{\partial S_i^{-1}}{\partial g_a} P_i + \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} U_j' \frac{\partial S_{j,i}^*}{\partial g_a} P_i \quad \forall a \in \{1, \dots, n_z + 2\} \quad \forall i \in \{1, \dots, M\}$
- $\frac{\partial^2 l}{\partial p_i \partial p_j} = -P' \Omega^{-1} P \approx -\delta_{i,j} (R^{-1})_{i,i} P_i' S_i^{-1} P_i - \delta_{|i-j|=1} (R^{-1})_{i,j} P_i' S_{i,j}^* P_j \quad \forall i, j \in \{1, \dots, M\}$

The derivatives of the log-likelihood function could, in principle, be used to obtain the maximum likelihood estimates. However, it can be easier to obtain maximum likelihood estimates by instead using the derivatives of a concentrated log-likelihood function. One of the conditions of the maximum likelihood estimates is  $\frac{\partial l}{\partial p} = 0$ , which implies that  $p$  is its GLS estimate  $(P' \Omega^{-1} P)^{-1} P' \Omega^{-1} Y \approx (P' \Omega^* P)^{-1} P' \Omega^* Y$ . Substituting this GLS estimate of  $p$  into the log-likelihood function yields a concentrated log-likelihood function  $\tilde{l}$ .

The first and second derivatives of the approximation to the concentrated log-likelihood function are

- $\frac{\partial \tilde{l}}{\partial f_a} = \frac{\partial l}{\partial f_a} + U' \Omega^{-1} P \frac{\partial p}{\partial f_a} \approx \frac{\partial l}{\partial f_a} + \sum_{i=1}^M (R^{-1})_{i,i} U_i' S_i^{-1} P_i \frac{\partial p_i}{\partial f_a} + \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} U_i' S_{i,j}^* P_j \frac{\partial p_j}{\partial f_a}$   
 $\forall a \in \{1, \dots, n_f\}$ , where  $\frac{\partial p}{\partial f_a} \approx (P' \Omega^* P)^{-1} \left( \frac{\partial P'}{\partial f_a} \Omega^* U - P' \Omega^* \frac{\partial P}{\partial f_a} p \right) \quad \forall a \in \{1, \dots, n_f\}$ .
- $\frac{\partial \tilde{l}}{\partial r} = \frac{\partial l}{\partial r} + U' \Omega^{-1} P \frac{\partial p}{\partial r} \approx \frac{\partial l}{\partial r} + \sum_{i=1}^M (R^{-1})_{i,i} U_i' S_i^{-1} P_i \frac{\partial p_i}{\partial r} + \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} U_i' S_{i,j}^* P_j \frac{\partial p_j}{\partial r}$ ,  
where  $\frac{\partial p}{\partial r} \approx (P' \Omega^* P)^{-1} P' \frac{\partial \Omega^*}{\partial r} U$ .
- $\frac{\partial \tilde{l}}{\partial g_a} = \frac{\partial l}{\partial g_a} + U' \Omega^{-1} P \frac{\partial p}{\partial g_a} \approx \frac{\partial l}{\partial g_a} + \sum_{i=1}^M (R^{-1})_{i,i} U_i' S_i^{-1} P_i \frac{\partial p_i}{\partial g_a} + \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} U_i' S_{i,j}^* P_j \frac{\partial p_j}{\partial g_a}$   
 $\forall a \in \{1, \dots, n_z + 2\}$ , where  $\frac{\partial p}{\partial g_a} \approx (P' \Omega^* P)^{-1} P' \frac{\partial \Omega^*}{\partial g_a} U \quad \forall a \in \{1, \dots, n_z + 2\}$ .
- $\frac{\partial^2 \tilde{l}}{\partial f_a \partial f_b} = \frac{\partial^2 l}{\partial f_a \partial f_b} - \frac{\partial p'}{\partial f_a} P' \Omega^{-1} P \frac{\partial p}{\partial f_b} - p' \frac{\partial P'}{\partial f_a} \Omega^{-1} P \frac{\partial p}{\partial f_b} + U' \Omega^{-1} \frac{\partial P}{\partial f_a} \frac{\partial p}{\partial f_b} + \frac{\partial^2 p'}{\partial f_a \partial f_b} P' \Omega^{-1} U$   
 $\approx \frac{\partial^2 l}{\partial f_a \partial f_b} - \sum_{i=1}^M (R^{-1})_{i,i} \frac{\partial p_i'}{\partial f_a} P_i' S_i^{-1} P_i \frac{\partial p_i}{\partial f_b} - \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} \frac{\partial p_i'}{\partial f_a} P_i' S_{i,j}^* P_j \frac{\partial p_j}{\partial f_b}$   
 $- \sum_{i=1}^M (R^{-1})_{i,i} p_i' \frac{\partial P_i'}{\partial f_a} S_i^{-1} P_i \frac{\partial p_i}{\partial f_b} - \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} p_i' \frac{\partial P_i'}{\partial f_a} S_{i,j}^* P_j \frac{\partial p_j}{\partial f_b}$   
 $+ \sum_{i=1}^M (R^{-1})_{i,i} U_i' S_i^{-1} \frac{\partial P_i}{\partial f_a} \frac{\partial p_i}{\partial f_b} + \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} U_i' S_{i,j}^* \frac{\partial P_j}{\partial f_a} \frac{\partial p_j}{\partial f_b}$





$$\begin{aligned}
& - \sum_{i=1}^M (R^{-1})_{i,i} \frac{\partial p'_i}{\partial r} P'_i S_i^{-1} P_i \frac{\partial p_i}{\partial g_a} - \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} \frac{\partial p'_i}{\partial r} P'_i S_{i,j}^* P_j \frac{\partial p_j}{\partial g_a} \\
& + \sum_{i=1}^M (R^{-1})_{i,i} \frac{\partial^2 p'_i}{\partial r \partial g_a} P'_i S_i^{-1} U_i + \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} \frac{\partial^2 p'_i}{\partial r \partial g_a} P'_i S_{i,j}^* U_j \quad \forall a \in \{1, \dots, n_z + 2\}, \\
& \text{where } \frac{\partial^2 p}{\partial r \partial g_a} \approx (P' \Omega^* P)^{-1} P' \left( \frac{\partial^2 \Omega^*}{\partial r \partial g_a} U - \frac{\partial \Omega^*}{\partial r} P \frac{\partial p}{\partial g_a} - \frac{\partial \Omega^*}{\partial g_a} P \frac{\partial p}{\partial r} \right) \quad \forall a \in \{1, \dots, n_z + 2\}.
\end{aligned}$$

$$\begin{aligned}
\bullet \quad & \frac{\partial^2 \tilde{l}}{\partial g_a \partial g_b} = \frac{\partial^2 l}{\partial g_a \partial g_b} + \frac{\partial p'}{\partial g_a} P' \frac{\partial \Omega^{-1}}{\partial g_b} U - \frac{\partial p'}{\partial g_a} P' \Omega^{-1} P \frac{\partial p}{\partial g_b} + \frac{\partial^2 p'}{\partial g_a \partial g_b} P' \Omega^{-1} U \\
& \approx \frac{\partial^2 l}{\partial g_a \partial g_b} + \sum_{i=1}^M (R^{-1})_{i,i} \frac{\partial p'_i}{\partial g_a} P'_i \frac{\partial S_i^{-1}}{\partial g_b} U_i + \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} \frac{\partial p'_i}{\partial g_a} P'_i \frac{\partial S_{i,j}^*}{\partial g_b} U_j \\
& - \sum_{i=1}^M (R^{-1})_{i,i} \frac{\partial p'_i}{\partial g_a} P'_i S_i^{-1} P_i \frac{\partial p_i}{\partial g_b} - \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} \frac{\partial p'_i}{\partial g_a} P'_i S_{i,j}^* P_j \frac{\partial p_j}{\partial g_b} \\
& + \sum_{i=1}^M (R^{-1})_{i,i} \frac{\partial^2 p'_i}{\partial g_a \partial g_b} P'_i S_i^{-1} U_i + \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} \frac{\partial^2 p'_i}{\partial g_a \partial g_b} P'_i S_{i,j}^* U_j \quad \forall a, b \in \{1, \dots, n_z + 2\}, \\
& \text{where } \frac{\partial^2 p}{\partial g_a \partial g_b} \approx (P' \Omega^* P)^{-1} P' \left( \frac{\partial^2 \Omega^*}{\partial g_a \partial g_b} U - \frac{\partial \Omega^*}{\partial g_a} P \frac{\partial p}{\partial g_b} - \frac{\partial \Omega^*}{\partial g_b} P \frac{\partial p}{\partial g_a} \right) \quad \forall a, b \in \{1, \dots, n_z + 2\}.
\end{aligned}$$

These derivatives, combined with the initial guess of  $(f, r, g)$ , are used to iteratively converge to the maximum likelihood estimates using the saddle-free Newton's method. The new value of  $(f, r, g)$  is  $(f, r, g) + \min \left( 1, \frac{-\tilde{l}}{\tilde{J}[\tilde{H}]^{-1} \tilde{J}'} \right) [\tilde{H}]^{-1} \tilde{J}'$ , where  $\tilde{H} = \tilde{D}' \tilde{J}$  is the Hessian of the concentrated log-likelihood function,  $\tilde{J} = \tilde{D} \tilde{l}$  is the Jacobian of the concentrated log-likelihood function, and  $\tilde{D} = \begin{bmatrix} \frac{\partial}{\partial f} & \frac{\partial}{\partial r} & \frac{\partial}{\partial g} \end{bmatrix}$ . The new value of  $p$  is then calculated using GLS:  $p = (P' \Omega^{-1} P)^{-1} P' \Omega^{-1} Y \approx (P' \Omega^* P)^{-1} P' \Omega^* Y$ . Only 10 iterations are performed for each model due to the high computational cost of each iteration. Fewer iterations are needed compared to the temperature climatology model due to the use of more data to estimate the temperature anomaly model. In all cases, after 10 iterations the hessian is negative definite, suggesting that 10 iterations provide satisfactory convergence to the maximum likelihood estimates.

### Selecting the Temperature Anomaly Model

After the estimation of each model is performed, the model with the lowest AIC is selected. The BIC is also calculated for comparison.

$$\begin{aligned}
AIC &= 2n_{\text{parameters}} - 2l \approx 2(n_f + M \cdot (n_q + 1) + n_z + 3) + N \ln(2\pi) + \sum_{i=1}^M \ln(|S_i|) \\
&+ \frac{\ln(1-\rho^2)}{2} \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M \text{sum}(S_{i,j}^* \circ S_{i,j}) + \sum_{i=1}^M (R^{-1})_{i,i} U'_i S_i^{-1} U_i + \sum_{i=1}^M \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} U'_i S_{i,j}^* U_j
\end{aligned}$$

### Parameter Uncertainty of the Temperature Anomaly Model

The maximum likelihood estimate of  $(f, r, g, p)$  is asymptotically multivariate normal with covariance matrix  $(-H)^{-1}$ , where  $H = D'J$  is the Hessian of the log-likelihood function,  $J = Dl$  is the Jacobian of the log-likelihood function, and  $D = \begin{bmatrix} \frac{\partial}{\partial f} & \frac{\partial}{\partial r} & \frac{\partial}{\partial g} & \frac{\partial}{\partial p} \end{bmatrix}$ . This multivariate normal approximation is used to quantify the uncertainty of the maximum likelihood estimates. From the covariance matrix  $(-H)^{-1}$ , the sub-covariance matrices  $\text{cov}\left(\begin{bmatrix} r \\ g \end{bmatrix}, \begin{bmatrix} r \\ g \end{bmatrix}\right)$ ,  $\text{cov}\left(\begin{bmatrix} f \\ p \end{bmatrix}, \begin{bmatrix} f \\ p \end{bmatrix}\right)$ , and  $\text{cov}\left(\begin{bmatrix} r \\ g \end{bmatrix}, \begin{bmatrix} f \\ p \end{bmatrix}\right)$  can be obtained.

To greatly reduce the computational time of estimating infilling uncertainty while still accounting for the uncertainty of the covariance function, only five ensemble members of the covariance function parameters  $(r, g)$  are generated. Since  $\text{cov}\left(\begin{bmatrix} r \\ g \end{bmatrix}, \begin{bmatrix} r \\ g \end{bmatrix}\right)$  is symmetric, it has eigen-decomposition  $\Phi\Delta\Phi'$ , where  $\Phi$  is a matrix of eigenvectors and  $\Delta$  is a diagonal matrix of eigenvalues. For the selected temperature model, five vectors of random numbers of length  $(n_z + 3)$  are generated from independent standard normal probability distributions. For each of these five vectors,  $\Phi\Delta^{0.5}$  is multiplied by the random vector to obtain an ensemble member of the parameters  $(r, g)$  of the temperature anomaly model. The five ensemble members of  $(r, g)$  are then normalized such that their sample means are equal to the means of the maximum likelihood estimates of  $(r, g)$ , and their uncorrected sample variances are equal to the variances of the maximum likelihood estimates of  $(r, g)$ ; this ensures that the small sample size does not significantly skew the results.

Given an ensemble member  $(r, g)$  and maximum likelihood estimates  $(\hat{f}, \hat{r}, \hat{g}, \hat{p})$ , the expected value of  $(f, p)$  under the conditional probability distribution is  $\begin{bmatrix} \hat{f} \\ \hat{p} \end{bmatrix} + \text{cov}\left(\begin{bmatrix} f \\ p \end{bmatrix}, \begin{bmatrix} r \\ g \end{bmatrix}\right) \left( \text{cov}\left(\begin{bmatrix} r \\ g \end{bmatrix}, \begin{bmatrix} r \\ g \end{bmatrix}\right) \right)^{-1} \begin{bmatrix} r - \hat{r} \\ g - \hat{g} \end{bmatrix}$ . In addition, under this conditional probability distribution, the covariance matrix of  $(r, g)$  is  $\text{cov}\left(\begin{bmatrix} f \\ p \end{bmatrix}, \begin{bmatrix} f \\ p \end{bmatrix}\right) - \text{cov}\left(\begin{bmatrix} f \\ p \end{bmatrix}, \begin{bmatrix} r \\ g \end{bmatrix}\right) \left( \text{cov}\left(\begin{bmatrix} r \\ g \end{bmatrix}, \begin{bmatrix} r \\ g \end{bmatrix}\right) \right)^{-1} \text{cov}\left(\begin{bmatrix} r \\ g \end{bmatrix}, \begin{bmatrix} f \\ p \end{bmatrix}\right)$ . Since the covariance matrix of  $(r, g)$  under the conditional probability distribution is symmetric, it has eigen-decomposition  $\tilde{\Phi}\tilde{\Delta}\tilde{\Phi}'$ , where  $\tilde{\Phi}$  is a matrix of eigenvectors and  $\tilde{\Delta}$  is a diagonal matrix of eigenvalues. For each of the ensemble members of  $(r, g)$ , 40 vectors of length  $(n_{parameters} - n_z - 3)$  are generated from independent standard normal probability distributions. For each of these 40 vectors,  $\text{cov}\left(\begin{bmatrix} f \\ p \end{bmatrix}, \begin{bmatrix} r \\ g \end{bmatrix}\right) \left( \text{cov}\left(\begin{bmatrix} r \\ g \end{bmatrix}, \begin{bmatrix} r \\ g \end{bmatrix}\right) \right)^{-1} \begin{bmatrix} r \\ g \end{bmatrix}$  is added to the product of  $\tilde{\Phi}\tilde{\Delta}^{0.5}$  with the random vector to obtain an ensemble member of the parameters  $(f, p)$  of the temperature anomaly model. Overall, this produces 200 ensemble members of the parameters  $(f, r, g, p)$  of the temperature anomaly model.

### Estimating Residuals of the Temperature Anomaly Model

The simple kriging estimate of the weather residuals  $W$  is  $\Lambda\eta\Omega^{-1}U$ . Conditional on the estimate of the model parameters, the kriging estimate is the most-efficient linear unbiased estimate as well as the maximum likelihood estimate of the weather residuals. To calculate an approximation to the simple

kriging estimate, first an approximation to  $\eta\Omega^{-1}U$ ,  $\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_M \end{bmatrix}$ , is calculated, where  $\beta_i \in \mathbb{R}^{n_i}$

$$\forall i \in \{1, \dots, M\} \text{ and } \beta = \eta \frac{1}{1-\rho^2} \begin{bmatrix} S_{1,1}^{-1} & -\rho S_{1,2}^* & \mathbb{O}_{n_1, n_3} & \dots & \mathbb{O}_{n_1, n_M} \\ -\rho S_{2,1}^* & (1+\rho^2)S_{2,2}^{-1} & \ddots & \ddots & \vdots \\ \mathbb{O}_{n_3, n_1} & \ddots & \ddots & \ddots & \mathbb{O}_{n_{M-2}, n_M} \\ \vdots & \ddots & \ddots & (1+\rho^2)S_{M-1, M-1}^{-1} & -\rho S_{M-1, M}^* \\ \mathbb{O}_{n_M, n_1} & \dots & \mathbb{O}_{n_M, n_{M-2}} & -\rho S_{M, M-1}^* & S_{M, M}^{-1} \end{bmatrix} U.$$

$$\Rightarrow \beta_i = \eta_i \left( (R^{-1})_{i,i} S_i^{-1} U_i + \sum_{\substack{j=1 \\ |i-j|=1}}^M (R^{-1})_{i,j} S_{i,j}^* U_j \right) \forall i \in \{1, \dots, M\}$$

Note that  $\Lambda\eta\Omega^{-1}U \approx \Lambda\beta = \tilde{V}^{-1}(c(R) \otimes I_{2G})(c'(R) \otimes I_{2G})(I_M \otimes \tilde{K})\tilde{V}^{-1}\beta$ . Next,  $\omega = \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_M \end{bmatrix} = (c'(R) \otimes I_{2G})(I_M \otimes \tilde{K})\tilde{V}^{-1}\beta$  is calculated, where  $\omega_i \in \mathbb{R}^{2G} \forall i \in \{1, \dots, M\}$ . The lower triangular Cholesky

matrix of the Kac-Murdock-Szegö matrix  $R$  is  $c(R) = \begin{bmatrix} \rho^{1-1} & 0 & \dots & 0 \\ \rho^{2-1} & \rho^{2-2}\sqrt{1-\rho^2} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \rho^{M-1} & \rho^{M-2}\sqrt{1-\rho^2} & \dots & \rho^{M-M}\sqrt{1-\rho^2} \end{bmatrix}$ .

$$\Rightarrow \omega = \begin{bmatrix} \rho^{1-1} I_{2G} & \rho^{2-1} I_{2G} & \dots & \rho^{M-1} I_{2G} \\ 0 & \rho^{2-2}\sqrt{1-\rho^2} I_{2G} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \rho^{M-2}\sqrt{1-\rho^2} I_{2G} \\ 0 & 0 & \dots & \rho^{M-M}\sqrt{1-\rho^2} I_{2G} \end{bmatrix} (I_M \otimes \tilde{K})\tilde{V}^{-1}\beta$$

$$\Rightarrow \omega_1 = \sum_{j=1}^M \rho^{M-1} \tilde{K} \tilde{V}_j^{-1} \beta_j, \omega_i = \sqrt{1-\rho^2} \sum_{j=i}^M \rho^{M-i} \tilde{K} \tilde{V}_j^{-1} \beta_j \forall i \in \{2, \dots, M\}$$

$$\Rightarrow \omega \text{ can be calculated iteratively using } \omega_M = \sqrt{1-\rho^2} \tilde{K} \tilde{V}_M^{-1} \beta_M, \omega_i = \sqrt{1-\rho^2} \tilde{K} \tilde{V}_i^{-1} \beta_i + \rho \omega_{i+1} \forall i \in \{2, \dots, M-1\}, \text{ and } \omega_1 = \tilde{K} \tilde{V}_1^{-1} \beta_1 + \frac{\rho}{\sqrt{1-\rho^2}} \omega_2.$$

Next,  $\xi = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_M \end{bmatrix} = (c(R) \otimes I_{2G})\omega$  is calculated, where  $\xi_i \in \mathbb{R}^{2G} \forall i \in \{1, \dots, M\}$ .

$$\xi = \begin{bmatrix} \rho^{1-1} I_{2G} & 0 & \dots & 0 \\ \rho^{2-1} I_{2G} & \rho^{2-2} \sqrt{1-\rho^2} I_{2G} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \rho^{M-1} I_{2G} & \rho^{M-2} \sqrt{1-\rho^2} I_{2G} & \dots & \rho^{M-M} \sqrt{1-\rho^2} I_{2G} \end{bmatrix} \omega$$

$$\Rightarrow \xi_i = \rho^{i-1} \omega_1 + \sqrt{1-\rho^2} \sum_{j=2}^i \rho^{i-j} \omega_j \quad \forall i \in \{1, \dots, M\}$$

$$\Rightarrow \xi \text{ can be calculated iteratively using } \xi_1 = \omega_1 \text{ and } \xi_i = \sqrt{1-\rho^2} \omega_i + \rho \xi_{i-1} \quad \forall i \in \{2, \dots, M\}.$$

Finally, the approximation to the simple kriging estimate is calculated as  $\tilde{V}^{-1} \xi = \begin{bmatrix} \tilde{V}_1^{-1} \xi_1 \\ \vdots \\ \tilde{V}_M^{-1} \xi_M \end{bmatrix}$ .

For Table 5 of the paper, the maximum likelihood estimates of model parameters and the kriging estimates of the unobserved weather residuals conditional on the maximum likelihood estimates of model parameters were combined to approximate the maximum likelihood estimates of temperature changes. The true maximum likelihood estimates of temperature changes may differ slightly due to a combination of uncertainties in the estimates of model parameters and non-linearities in how the estimates of model parameters interact with the estimates of the weather residuals. The generation of ensemble members discussed in the next section does not depend on this approximation.

### Infilling Uncertainty of the Temperature Anomaly Model

The covariance matrix of  $\Lambda \eta \Omega^{-1} U$  is  $\Lambda - \Lambda \eta \Omega^{-1} \eta' \Lambda$ . To account for uncertainties in the simple kriging estimate of  $W$ , the 200 ensemble members of parameters are combined with the 200 ensemble members of temperature anomaly data to obtain 200 ensemble members of  $W$ . For each ensemble

member of  $U$ , a vector  $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_M \end{bmatrix}$ , is generated from an independent standard normal probability distribution, where  $\varepsilon_i \in \mathbb{R}^{2G} \forall i \in \{1, \dots, M\}$ .

$\Lambda - \Lambda \eta \Omega^{-1} \eta' \Lambda = \Lambda c(\Lambda^{-1} - \eta \Omega^{-1} \eta') c' (\Lambda^{-1} - \eta \Omega^{-1} \eta')' \Lambda$ . Thus, for each ensemble member of  $U$ , an ensemble member of  $W$  can be generated as  $\Lambda \eta \Omega^{-1} U + \Lambda c(\Lambda^{-1} - \eta \Omega^{-1} \eta') \varepsilon$ . To calculate an approximation to  $c(\Lambda^{-1} - \eta \Omega^{-1} \eta')$ , note that  $\Lambda^{-1} = (\tilde{V}^{-1} (R \otimes \tilde{K}) \tilde{V}^{-1})^{-1} = \tilde{V} (R^{-1} \otimes \tilde{K}^{-1}) \tilde{V}$  and

$$\Omega^{-1} \approx \frac{1}{1-\rho^2} \begin{bmatrix} S_{1,1}^{-1} & -\rho S_{1,2}^* & \mathbb{O}_{n_1,n_3} & \cdots & \mathbb{O}_{n_1,n_M} \\ -\rho S_{2,1}^* & (1+\rho^2)S_{2,2}^{-1} & \ddots & \ddots & \vdots \\ \mathbb{O}_{n_3,n_1} & \ddots & \ddots & \ddots & \mathbb{O}_{n_{M-2},n_M} \\ \vdots & \ddots & \ddots & (1+\rho^2)S_{M-1,M-1}^{-1} & -\rho S_{M-1,M}^* \\ \mathbb{O}_{n_M,n_1} & \cdots & \mathbb{O}_{n_M,n_{M-2}} & -\rho S_{M,M-1}^* & S_{M,M}^{-1} \end{bmatrix}.$$

$$\Rightarrow \Lambda^{-1} - \eta \Omega^{-1} \eta' \approx \tau = \begin{bmatrix} \tau_{1,1} & \tau_{1,2} & \mathbb{O}_{n_1,n_3} & \cdots & \mathbb{O}_{n_1,n_M} \\ \tau_{2,1} & \ddots & \ddots & \ddots & \vdots \\ \mathbb{O}_{n_3,n_1} & \ddots & \ddots & \ddots & \mathbb{O}_{n_{M-2},n_M} \\ \vdots & \ddots & \ddots & \ddots & \tau_{M-1,M} \\ \mathbb{O}_{n_M,n_1} & \cdots & \mathbb{O}_{n_M,n_{M-2}} & \tau_{M,M-1} & \tau_{M,M} \end{bmatrix}, \text{ where } \forall i, j \in \{1, \dots, M\}$$

$$\tau_{i,i} = (R^{-1})_{i,i} (\tilde{V}_i \tilde{K}^{-1} \tilde{V}_i - \eta_i S_{i,i}^{-1} \eta_i') \text{ and } \tau_{i,j} = (R^{-1})_{i,j} (\tilde{V}_i \tilde{K}^{-1} \tilde{V}_j - \eta_i S_{i,j}^* \eta_j') \text{ if } |i-j| = 1.$$

To reduce computational rounding error,  $c(\tau)$  is calculated as

$$\begin{bmatrix} c(\tau_{1,1}) & \mathbb{O}_{n_1,n_2} & \cdots & \mathbb{O}_{n_1,n_M} \\ \mathbb{O}_{n_2,n_1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbb{O}_{n_{M-1},n_M} \\ \mathbb{O}_{n_M,n_1} & \cdots & \mathbb{O}_{n_M,n_{M-1}} & c(\tau_{M,M}) \end{bmatrix} \begin{bmatrix} \Psi_{1,1} & \mathbb{O}_{n_1,n_2} & \mathbb{O}_{n_1,n_3} & \cdots & \mathbb{O}_{n_1,n_M} \\ \Psi_{2,1} & \ddots & \ddots & \ddots & \vdots \\ \mathbb{O}_{n_3,n_1} & \ddots & \ddots & \ddots & \mathbb{O}_{n_{M-2},n_M} \\ \vdots & \ddots & \ddots & \ddots & \mathbb{O}_{n_{M-1},n_M} \\ \mathbb{O}_{n_M,n_1} & \cdots & \mathbb{O}_{n_M,n_{M-2}} & \Psi_{M,M-1} & \Psi_{M,M} \end{bmatrix}.$$

$$\Rightarrow \Psi \text{ can be calculated iteratively using } \Psi_{1,1} = I_{n_1}, \Psi_{i,i} = c(I_{n_i} - \Psi_{i,i-1} \Psi_{i,i-1}') \forall i \in \{2, \dots, M\}, \text{ and}$$

$$\Psi_{i,i-1} = \left( c(\tau_{i,i}) \right)^{-1} \tau_{i,i-1} \left( c(\tau_{i-1,i-1}) \right)^{-1} (\Psi_{i-1,i-1}')^{-1} \forall i \in \{2, \dots, M\}.$$

After  $c(\tau)$  is calculated,  $\gamma = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_M \end{bmatrix} = c(\tau) \varepsilon$  is calculated using  $\gamma_1 = c(\tau_{1,1}) \Psi_{1,1} \varepsilon_1$  and

$\gamma_i = c(\tau_{i,i}) (\Psi_{i,i-1} \varepsilon_{i-1} + \Psi_{i,i} \varepsilon_i) \forall i \in \{2, \dots, M\}$ . Using the approximation  $\beta \approx \eta \Omega^{-1} U$  from the previous section, one obtains  $\Lambda \eta \Omega^{-1} U + \Lambda c(\Lambda^{-1} - \eta \Omega^{-1} \eta') \varepsilon \approx \Lambda(\beta + \gamma)$ .  $\Lambda(\beta + \gamma)$  is calculated similarly to the previous section. First  $\omega$  is calculated iteratively using  $\omega_M = \sqrt{1 - \rho^2} \tilde{K} \tilde{V}_M^{-1} (\beta_M + \gamma_M)$ ,

$\omega_i = \sqrt{1 - \rho^2} \tilde{K} \tilde{V}_i^{-1} (\beta_i + \gamma_i) + \rho \omega_{i+1} \forall i \in \{2, \dots, M-1\}$ , and  $\omega_1 = \tilde{K} \tilde{V}_1^{-1} \beta_1 + \frac{\rho}{\sqrt{1-\rho^2}} \omega_2$ .

Next,  $\xi$  is calculated iteratively using  $\xi_1 = \omega_1$  and  $\xi_i = \sqrt{1 - \rho^2} \omega_i + \rho \xi_{i-1} \forall i \in \{2, \dots, M\}$ .

Finally,  $\Lambda(\beta + \gamma)$  is calculated as  $\tilde{V}^{-1} \xi = \begin{bmatrix} \tilde{V}_1^{-1} \xi_1 \\ \vdots \\ \tilde{V}_M^{-1} \xi_M \end{bmatrix}$ .

These 200 ensemble members of weather residuals are combined with the 200 ensemble members of the temperature anomaly model parameters to produce 200 ensemble members of temperature anomalies. The 200 ensemble members of temperature anomalies are then rebaselined to the 1961-1990 reference period. Land-ice and sea temperature anomalies are then blended by using the average



sea ice concentrations of 1961-1990 by calendar month. These 200 ensemble members are combined with the 200 ensemble members of temperature climatologies to produce 200 ensemble members of bias-corrected temperature anomalies.

After running the code, three minor issues were found related to the generation of ensemble members of temperature anomalies. A future version of HadCRU\_MLE is planned to correct for these issues as well as update the dataset to use newer source data.

- In the estimation of the temperature anomaly model and in the generation of infilling uncertainty, the homogenization and climatological uncertainties of LSAT anomalies of HadCRUT4 were unintentionally included in the error covariance matrix  $I'$  for the measurement and sampling uncertainties. However, the ensemble members of LSAT anomalies of HadCRUT4 also include these homogenization and climatological uncertainties. As a result, the uncertainties of LSAT anomalies in HadCRU\_MLE may be slightly overestimated. As the homogenization and climatological uncertainties are small relative to the measurement and sampling uncertainties, this issue negligibly affects temperature estimates and does not affect the conclusions of the study.
- An error in the code caused the equation  $\Psi_{i,i} = c(I_{n_i} - \Psi_{i,i-1}\Psi'_{i,i-1})$  on the previous page to be calculated as  $\Psi_{i,i} = c(I_{n_i} - \Psi_{i,i-2}\Psi'_{i,i-2})$  in most cases. As a result, the estimated infilling uncertainty may not appropriately reflect the temporal correlation of weather residuals. The maximum likelihood estimates of HadCRU\_MLE still appropriately account for the temporal correlation of weather residuals. As the estimated correlation of weather residuals between consecutive months is very small (0.07), this issue negligibly affects temperature estimates and does not affect the conclusions of the study.
- For one in every 300 months, an error in the code caused  $(R^{-1})_{i,i+1}S_{i,i+1}^*U_{i+1}$  to be incorrectly added to the wrong ensemble members in the calculation of  $\beta_i$  on page 26. The maximum likelihood estimates of HadCRU\_MLE are not affected by this error. As the estimated correlation of weather residuals between consecutive months is very small (0.07), this issue negligibly affects temperature estimates and does not affect the conclusions of the study.