A Bayesian ANOVA Scheme for Calculating Climate Anomalies, with Applications to the Instrumental Temperature Record

MARTIN P. TINGLEY*

National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 3 January 2011, in final form 22 July 2011)

ABSTRACT

Climate datasets with both spatial and temporal components are often studied after removing from each time series a temporal mean calculated over a common reference interval, which is generally shorter than the overall length of the dataset. The use of a short reference interval affects the temporal properties of the variability across the records, by reducing the standard deviation within the reference interval and inflating it elsewhere. For an annually averaged version of the Climate Research Unit's (CRU) temperature anomaly product, the mean standard deviation is 0.67°C within the 1961–90 reference interval, and 0.81°C elsewhere.

The calculation of anomalies can be interpreted in terms of a two-factor analysis of variance model. Within a Bayesian inference framework, any missing values are viewed as additional parameters, and the reference interval is specified as the full length of the dataset. This Bayesian scheme is used to re-express the CRU dataset as anomalies with respect to means calculated over the entire 1850–2009 interval spanned by the dataset. The mean standard deviation is increased to 0.69°C within the original 1961–90 reference interval, and reduced to 0.76°C elsewhere. The choice of reference interval thus has a predictable and demonstrable effect on the second spatial moment time series of the CRU dataset. The spatial mean time series is in this case largely unaffected: the amplitude of spatial mean temperature change is reduced by 0.1°C when using the 1850–2009 reference interval, while the 90% uncertainty interval of (−0.03, 0.23) indicates that the reduction is not statistically significant.

1. Introduction

For the purposes of studying climate, space–time datasets are often analyzed after the mean over some specific time interval has been removed from each time series. For example, the gridded temperature anomaly compilation produced by the Climate Research Unit (CRU) is composed of monthly time series of anomalies from a 1961–90 reference interval (Brohan et al. 2006), while the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report plots numerous millennial-scale climate reconstructions as anomalies from that same interval (Fig. 6.10 of Jansen et al. 2007).

There are both technical and scientific reasons for analyzing temperature anomalies rather than the actual values. Jones et al. (1999) argue that use of anomalies avoids a number of problems that can potentially arise when combining daily station data into monthly gridbox averages. The impacts of differences in station elevations, the timings of daily observations, and the methods used to calculate monthly means are minimized by considering anomalies, and the resulting dataset is more homogeneous than the corresponding compilation of actual values (Jones et al. 1999).

In general, climate fields display complex spatial structures, such as a strong dependence on latitude, sharp gradients across land–sea boundaries, and elevation effects—all of which can be seen in the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis (Kalnay et al. 1996; data available at http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html) annual mean temperature field for 1981 (Fig. 1a). Many of these spatial structures are relatively stable as a function of time, and are thus well estimated from the long-term temporal

* Current affiliation: Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts.

Corresponding author address: Martin P. Tingley, National Center for Atmospheric Research, 1850 Table Mesa Drive, Boulder, CO 80305.
E-mail: tingley@fas.harvard.edu

DOI: 10.1175/JCLI-D-11-00008.1

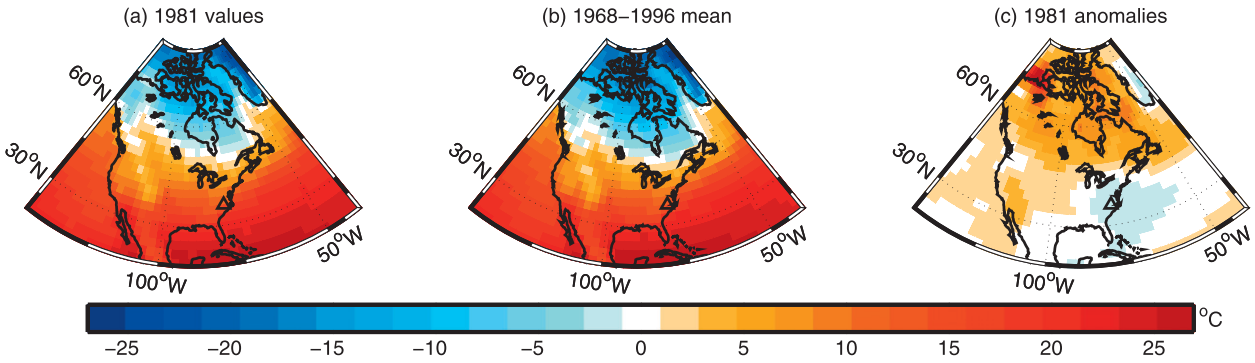

FIG. 1. NCEP reanalysis (Kalnay et al. 1996) for the annual mean temperature field: (a) values for 1981, (b) 1968–96 long-term mean, and (c) 1981 anomalies from the long-term mean.

mean (Fig. 1b). There is generally less structure in the field of anomalies with respect to an estimated long-term mean (Fig. 1c), and the anomalies are generally representative of larger spatial areas (e.g., Hansen and Lebedeff 1987). A sensible analysis of the anomaly field likely requires fewer covariates, as the anomaly field is more likely to be spatially stationary (e.g., Banerjee et al. 2004). As scientific interest often lies in understanding *changes* in climate fields such as surface temperatures, rather than the details of the field itself, the analysis of anomalies allows for a simpler statistical model, which facilitates the identification of trends and patterns of change.

Given the underlying assumption that the climate field is changing in time, it is important to remove from each time series a mean calculated over a common reference interval, and this reference interval is often chosen as a subinterval that minimizes the number of missing values (e.g., Brohan et al. 2006). Using a reference interval that is shorter than the full length of the dataset leads to increased uncertainty in the estimation of the means (smaller sample size), and often does not entirely eliminate the missing data problem. In addition, using a short reference interval results in the variance across the estimated anomaly time series (spatial variance) being reduced within the reference interval, and inflated elsewhere (see section 2). In the extreme case, the spatial variance within a one time-step reference interval is zero.

Any analysis of climate anomalies that depends on second- or higher-moment properties may therefore be affected by the choice of reference interval used to calculate the anomalies. For example, the frequency of climate events that are defined as threshold exceedances, such as heat waves or other extremes, changes if either the mean or the standard deviation changes (see Fig. 4-1 in Watson et al. 2001). As the standard deviation is lower within a short reference interval, any threshold is more likely to be exceeded outside of the reference interval used to calculate the anomalies, and this effect becomes more pronounced the more extreme the threshold.

The scientific interpretation of several figures from the IPCC Fourth Assessment Report (Trenberth et al. 2007) is influenced by the reference interval used to calculate climate anomalies. For example, Fig. 3.5 in the Fourth Assessment Report plots zonally averaged temperature anomalies with respect to a 1961–90 reference interval as a function of latitude and time, while Fig. 3.15 depicts precipitation anomalies in the same manner. The choice of a short reference interval affects the temporal evolution of the spatial standard deviation as a function of time, and this effect is clearly visible in the precipitation plot, which features markedly reduced variability within the 1961–90 reference interval.

As a final example, paleoclimatic field reconstructions are generally calibrated against instrumental anomalies with respect to a reference interval that is a subset of the calibration interval. Mann et al. (2009) presents a reconstruction, calibrated over the 1850–1995 interval, of surface temperature anomalies relative to a 1961–90 reference interval. The spatial pattern of temperature anomalies for the Medieval Climate Anomaly and the Little Ice Age, which are shown in Fig. 2 of Mann et al. (2009), are influenced in part by the choice of reference interval, which affects the spatial variability as a function of time. In addition, the 1850–1995 calibration interval used in Mann et al. (2009) includes both the 1961–90 reference interval used to calculate the instrumental anomalies, as well as times outside of this interval. The reconstruction is thus calibrated against a dataset with nonstationary statistical properties, and this temporal structure may introduce artifacts into the estimated relationship between the two datasets. Note that these issues are not unique to Mann et al. (2009); indeed, all 12 large-scale temperature reconstructions

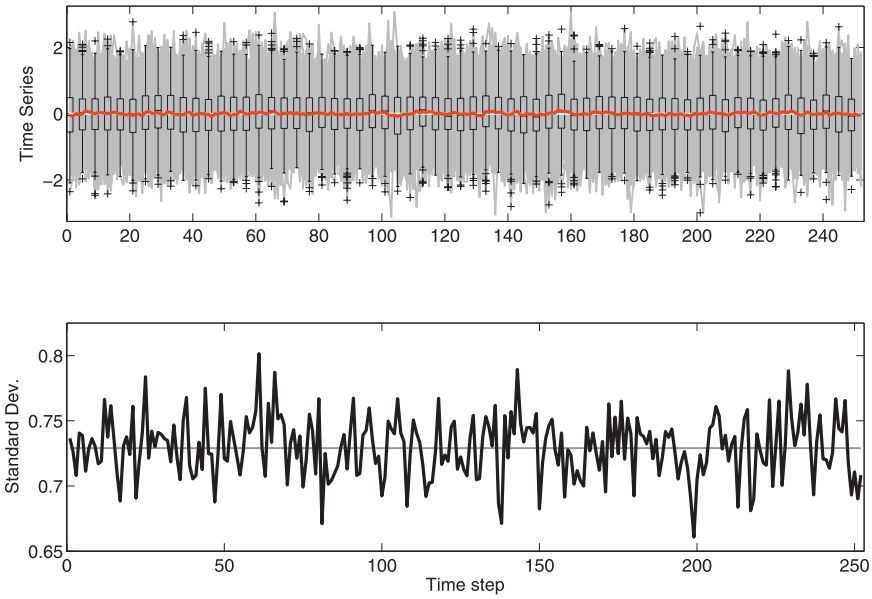

FIG. 2. (top) The 459 independent AR(1) time series ($\alpha = 0.34, \sigma^2 = 0.47$), box plots at every fourth time step, and the sample mean at each time step. (bottom) The sample standard deviation at each time step (black line) and the population standard deviation (gray line) calculated from Eq. (1). The number of time series, the relative length of the reference interval, and the values of α and σ^2 are chosen to correspond to estimates from the CRU dataset analyzed in section 4.

depicted in Figs. 6–10 from Jansen et al. (2007) are calibrated against instrumental temperature anomalies with respect to a 1961–90 reference interval.

An alternative approach to the calculation of anomalies is to maximize the length of the reference interval, thereby mitigating the effects on the second-moment properties of the dataset, while accounting for the increased uncertainty that results from some observations being missing. We propose a two-factor analysis of variance (ANOVA) model for the calculation of climate anomalies, with the factors being location and year. The design is not balanced, as there is either one or zero observations at each combination of factor levels, and this complexity motivates a Bayesian inference scheme. The missing values in the dataset are then treated as additional parameters, the reference interval is specified as the full length of the dataset, and the uncertainty in the estimated anomalies accounts for the fact that the dataset is incomplete.

Section 2 illustrates the effects of a short reference interval on the time series of spatial standard deviations using a simple example dataset. Section 3 presents the ANOVA model for calculating anomalies and details a Bayesian approach to fitting this model in the presence of missing data. Section 4 uses the ANOVA model to re-express the annual mean CRU temperature product as anomalies from the 1850–2009 interval and investigates the effects of the change in reference interval on the time series of means and standard deviations. Section 5 provides the discussion and concluding remarks.

2. Biases produced by a short reference interval

This section will demonstrate that calculating anomalies from a short reference interval introduces spurious structures in the time series of the standard deviations across the series. In the climate context, using a short reference interval introduces *temporal structure* in estimates of the *spatial standard deviation*. Consider a length-N univariate first-order autoregressive [AR(1)] time series $\mathbf{X}$, with AR(1) parameter $|\alpha| < 1$ and independent and identically distributed (IID) normal innovations with mean zero and variance σ^2. The series $\mathbf{X}$ is multivariate normal,

$$\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}), \quad \Omega_{ij} = \frac{\sigma^2}{1 - \alpha^2}\alpha^{|i-j|}, \tag{1}$$

and the diagonal elements of $\mathbf{\Omega}$ are all equal, implying that the variance of $\mathbf{X}$ is constant as a function of time (Fig. 2). Now consider a length N^* reference interval that runs from time points s_1 to s_2, where $1 \le s_1 \le s_2 \le N$. The vector of anomalies from this reference interval, $\mathbf{Y}$, can be written as a linear transformation of $\mathbf{X}$,

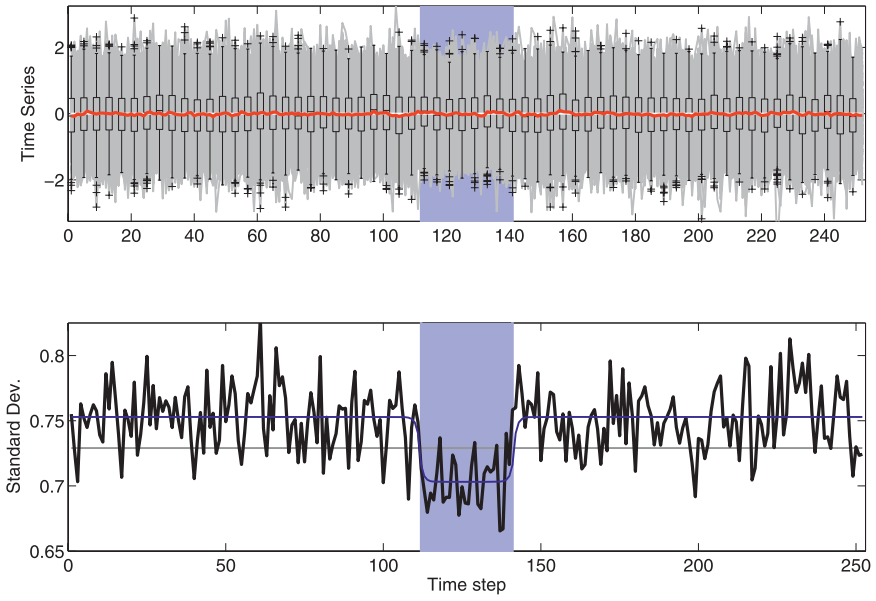

FIG. 3. As in Fig. 2, but after removing from each time series the sample mean calculated over the 30 time step shaded interval. (bottom) The gray line is the population standard deviation of the original series, and the blue line is the population standard deviation of the anomalies calculated from Eq. (4).

$$\mathbf{Y} = (\mathbf{I} + \mathbf{A})\mathbf{X}, \tag{2}$$

$$\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}^*), \quad \mathbf{\Omega}^* = (\mathbf{I} + \mathbf{A})\mathbf{\Omega}(\mathbf{I} + \mathbf{A})^{\mathrm{T}}. \tag{3}$$

where $\mathbf{I}$ is the N by N identity matrix and $\mathbf{A}$ is an N by N matrix composed of columns of zeros outside of the reference interval and columns of $-1/N^*$ inside the reference interval. The distribution of $\mathbf{Y}$ is then

While no structure has been added to the mean vector, the diagonal elements of $\mathbf{\Omega}^*$, which represent the variance at each time point, now vary as a function of time (Fig. 3). The diagonal entries of $\mathbf{\Omega}^*$ can be expressed as

$$\begin{aligned}
\Omega^*_{ii} &= \mathrm{Var}\left(X_i - \frac{1}{N^*}\sum_{k=s_1}^{s_2} X_k\right) \\
\Omega^*_{ii} &= \left(1 + \frac{1}{N^*}\right)\mathrm{Var}(X_i) + \frac{2}{(N^*)^2}\sum_{k=s_1,p>k}^{s_2} \mathrm{Cov}(X_k, X_p) - \frac{2}{N^*}\sum_{k=s_1}^{s_2} \mathrm{Cov}(X_i, X_k) \\
\Omega^*_{ii} &= \left(1 + \frac{1}{N^*}\right)\frac{\sigma^2}{1-\alpha^2} + \frac{2\sigma^2}{1-\alpha^2}\left(\frac{1}{N^{*2}}\sum_{k=s_1,p>k}^{s_2} \alpha^{|k-p|} - \frac{1}{N^*}\sum_{k=s_1}^{s_2} \alpha^{|i-k|}\right).
\end{aligned} \tag{4}$$

In the second line, the first term is the sum of the variances, the second is the sum of the covariances within the reference interval, and the third term is the sum of the covariances between the ith time point and each time point within the reference interval. For points located far from the reference interval (i.e., $i \ll s_1$ or $i \gg s_2$), the second sum essentially drops out, all terms are positive, and $\Omega^*_{ii} > \Omega_{ii}$. For points within the reference interval, the second sum (proportional to $1/N^*$) dominates the first, which is proportional to $1/N^{*2}$ so that $\Omega^*_{ii} < \Omega_{ii}$ (Fig. 3). As the two sums are partial geometric series, a closed-form expression can be derived for the diagonal elements $\mathbf{\Omega}^*$. The resulting expression, however, is no more informative than Eq. (4).

To explore the discrepancy between Ω^*_{ii} and Ω_{ii} as a function of the AR(1) coefficient and the relative length of the reference interval, define the *scaled standard deviation range* as

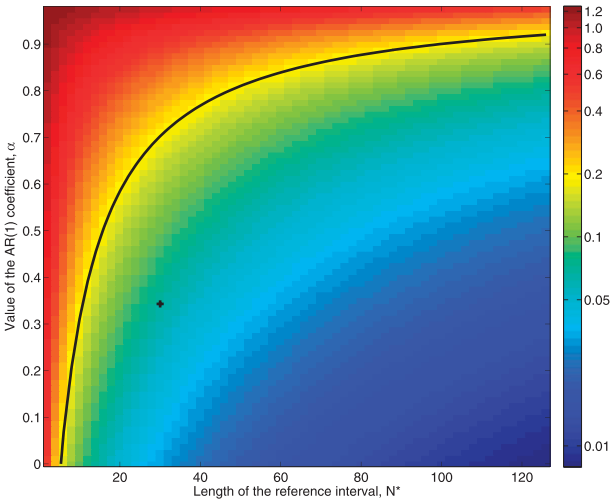

FIG. 4. The scaled standard deviation range, $\Delta(\alpha, N, N^*)$, as a function of both the AR(1) coefficient and the length of the reference interval, for AR(1) time series of length $N = 252$ with reference interval centered on $N/2$. Note that the color scale is logarithmic. The black plus sign designates $N^* = 30$ and $\alpha = 0.34$, which correspond to the example in Figs. 2 and 3 and the parameter values obtained from the analysis of the CRU data in section 4. The black contour corresponds to the actual value of Δ estimated from the CRU dataset in section 4.

$$\Delta(\alpha, N, N^*) = \frac{\sqrt{\max(\Omega^*_{ii})} - \sqrt{\min(\Omega^*_{ii})}}{\sqrt{\Omega_{11}}}. \quad (5)$$

Scaling by $\sqrt{\Omega_{11}}$ eliminates the dependence on σ^2, so that $\Delta(\alpha, N, N^*)$ is the range of standard deviation values induced in the anomalies as a proportion of the standard deviation of the original stationary time series. The value of $\Delta(\alpha, N, N^*)$ for fixed N is a decreasing function of N^* and an increasing function of α (Fig. 4). As a specific example, $\Delta(0.34, 252, 30) = 0.068$ (see Figs. 2 and 3 and section 4c), implying that the anomalies with respect to a 30 time step reference interval feature a structure in the time series of standard deviations with an amplitude that is 6.8% of the original stationary standard deviation.

3. An ANOVA model for calculating climate anomalies

Let $\mathbf{X}$ represent a matrix of observations of a climate variable, with the M rows corresponding to locations and the N columns to equally spaced time points. For example, in section 4, $\mathbf{X}$ will correspond to annual mean temperature anomalies at a number of spatial locations. In general, $\mathbf{X}$ will feature missing values, as different locations have instrumental observations that cover different time periods. We express the elements of $\mathbf{X}$ via a two-way ANOVA decomposition (e.g., Scheffé 1999; Zar 1999), where the factors are location and year:

$$X_{ij} = \gamma + \delta_i + \mu_j + \epsilon_{ij}. \quad (6)$$

Following standard ANOVA terminology, γ represents the grand mean, while the elements of $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$ correspond, respectively, to the M location effects and N year effects. Intuitively, each element of $\boldsymbol{\delta}$ corresponds to the temporal mean (relative to γ) at a particular location, while each element of $\boldsymbol{\mu}$ corresponds to the mean (relative to γ) across available observations at a particular time point.

To ensure identifiability of the parameters, the vectors of location and year effects are subject to sum-to-zero constraints:

$$\sum_{i=1}^{M} \delta_i = 0 \quad \text{and} \quad \sum_{j=1}^{N} \mu_j = 0. \quad (7)$$

As a result, the number of free parameters in $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$ is one less than the length of that vector. The simplest choice for modeling the error terms ϵ_{ij} is to assume that they are IID normal, $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, and this choice is made below. While the assumption of IID errors is likely not correct, it simplifies calculations and is sufficient to demonstrate the effects of a short reference interval; alternatives are discussed in section 5.

If the main goal of the analysis is to arrive at a better estimate of the anomalies, then interest lies primarily in inference on the location effects, $\boldsymbol{\delta}$. Estimates of the anomalies Y_{ij} follow from removing the grand mean and location effect from each observation:

$$Y_{ij} = X_{ij} - \gamma - \delta_i = \mu_j + \epsilon_{ij}. \quad (8)$$

Standard techniques for fitting ANOVA models (e.g., Scheffé 1999; Zar 1999) generally assume a balanced design, meaning that there are the same number of observations at each combination of factor levels. In particular, package ANOVA solutions are not designed to account for factor combinations for which there are no observations, as is the case if there are missing observations in the matrix $\mathbf{X}$.

Bayesian ANOVA with missing data

Fitting the proposed ANOVA model [Eq. (6)] involves parameter estimation in the presence of missing data—a situation amenable to Bayesian analysis. Within the Bayesian framework, the missing values are treated as additional parameters that must be estimated, while the posterior distributions of $\boldsymbol{\delta}$, $\boldsymbol{\mu}$, and γ include the uncertainty introduced by the missing data. Let $\mathbf{X}^m$ and $\mathbf{X}^o$

represent the elements of $\mathbf{X}$ that are missing and observed, respectively, where we require that no row or column be entirely missing. Similarly, let the vectors $\mathbf{X}_j^m$ and $\mathbf{X}_j^o$ represent the observed and missing elements for the jth time step. We seek posterior inference on $\mathbf{X}^m, \boldsymbol{\delta}, \boldsymbol{\mu}, \gamma$, and σ^2, conditional on $\mathbf{X}^o$. Application of Bayes' rule yields

$$P(\mathbf{X}^m, \boldsymbol{\delta}, \boldsymbol{\mu}, \gamma, \sigma^2 | \mathbf{X}^o) \propto P(\mathbf{X}^o | \mathbf{X}^m, \boldsymbol{\delta}, \boldsymbol{\mu}, \gamma, \sigma^2) \times P(\mathbf{X}^m, \boldsymbol{\delta}, \boldsymbol{\mu}, \gamma, \sigma^2) \quad (9)$$

$$= P(\mathbf{X}^o | \mathbf{X}^m, \boldsymbol{\delta}, \boldsymbol{\mu}, \gamma, \sigma^2,) P(\mathbf{X}^m | \boldsymbol{\delta}, \boldsymbol{\mu}, \gamma, \sigma^2) P(\boldsymbol{\delta}, \boldsymbol{\mu}, \gamma, \sigma^2) \quad (10)$$

$$= \prod_{j=1}^{N} P(\mathbf{X}_j^o | \mathbf{X}_j^m, \boldsymbol{\delta}, \mu_j, \sigma^2, \gamma,) \prod_{j=1}^{N} P(\mathbf{X}_j^m | \boldsymbol{\delta}, \mu_j, \sigma^2, \gamma,) \times P(\boldsymbol{\delta}, \boldsymbol{\mu}, \sigma^2, \gamma), \quad (11)$$

where the second line follows from the identity $P(A, B) = P(A|B)P(B)$, and the third from the assumption that the error vectors for each time step are independent.

The first term on the right-hand side of Eq. (11) is the likelihood of the data ($\mathbf{X}^o$) given the unknowns. Under the assumption that the error elements ϵ_{ij} are IID normal, the off-diagonal elements of the covariance matrix of $\mathbf{X}_j$ are zero. The likelihood thus does not depend on the missing values, and can be re-expressed as a double product of univariate normals:

$$\prod_{j=1}^{N} P(\mathbf{X}_j^o | \mathbf{X}_j^m, \boldsymbol{\delta}, \mu_j, \sigma^2, \gamma,) = \prod_{\text{obs}} P(X_{ij}^o | \delta_i, \mu_j, \sigma^2, \gamma,), \quad (12)$$

where the product is over all observed elements of $\mathbf{X}$.

The second product of multivariate normals in Eq. (11) can likewise be expressed as a double product of univariate normals:

$$\prod_{j=1}^{N} P(\mathbf{X}_j^m | \boldsymbol{\delta}, \mu_j, \sigma^2, \gamma,) = \prod_{\text{miss}} P(X_{ij}^m | \delta_i, \mu_j, \sigma^2, \gamma,), \quad (13)$$

where the product is over all missing elements.

The second term on the right-hand side of Eq. (9) gives the joint prior for the unknowns, which is re-expressed in Eq. (11) as the conditional distribution of $\mathbf{X}^m$ given the model parameters multiplied by the joint prior for $\boldsymbol{\delta}$, $\boldsymbol{\mu}$, γ, and σ^2. We specify independent priors for these parameters,

$$P(\boldsymbol{\delta}, \boldsymbol{\mu}, \sigma^2, \gamma) = P(\boldsymbol{\delta}) P(\boldsymbol{\mu}) P(\sigma^2) P(\gamma). \quad (14)$$

Details of the prior specifications, which must enforce the sum-to-zero constraints [Eq. (7)], and the sampling strategy are provided in the appendix. The end result of the analysis is an ensemble of posterior draws of $\mathbf{X}^m$, $\boldsymbol{\delta}$, $\boldsymbol{\mu}$, γ, and σ^2, conditional on the data, priors, and modeling assumptions. The posterior ensemble can be used to specify both point estimates and uncertainties for the anomalies, and for any other function of the unknowns. (A Matlab code package and relevant data files are available online at http://www.ncdc.noaa.gov/paleo/softlib/.)

An alternative inference strategy could make use of a variant of the expectation-maximization algorithm of Dempster et al. (1977). While such a frequentist approach can produce point estimates of the missing values $\mathbf{X}^m$ and the vectors $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$, as well as estimates of the associated uncertainty for these quantities, Bayesian inference is useful in what follows for two reasons. First, draws from the posterior allow for uncertainty estimation in quantities such as the time series of the change in standard deviations after expressing the data as anomalies from the longer interval (Fig. 9). Second, obvious extensions to models with spatially correlated location effects or temporally correlated year effects will be more tractable within a Bayesian framework (see section 5).

4. CRU annual mean temperatures: Anomalies from 1961 to 1990 and from 1850 to 2009

a. Data and basic results

We apply the Bayesian ANOVA model to an annually averaged version of the CRU temperature product version 3 (CRUTEM3) dataset (Brohan et al. 2006; data available at www.cru.uea.ac.uk/cru/data/temperature/) of land surface temperatures; results are qualitatively unchanged when using the variance-adjusted CRUTEM3v. The CRUTEM3 dataset provides monthly mean anomalies with respect to a 1961–90 reference, and we calculate annual anomalies by averaging all available monthly observations for years and locations for which there are at least nine monthly observations. The spatial distribution of data availability indicates that the longer instrumental records are predominantly located in Europe and North America (Fig. 5). Forming the matrix $\mathbf{X}$ from time series at each of the 839 locations for which there is at least one annual mean observation, 45% of the values are missing, with 79% of the missing values occurring in the first half (1850–1929) of the 1850–2009 interval spanned by the dataset. The year 1850 is 111 time steps from the beginning of the 30-yr reference interval, which motivates the symmetric example in Figs. 2 and 3, where each of the 459 (45% of 839) time series

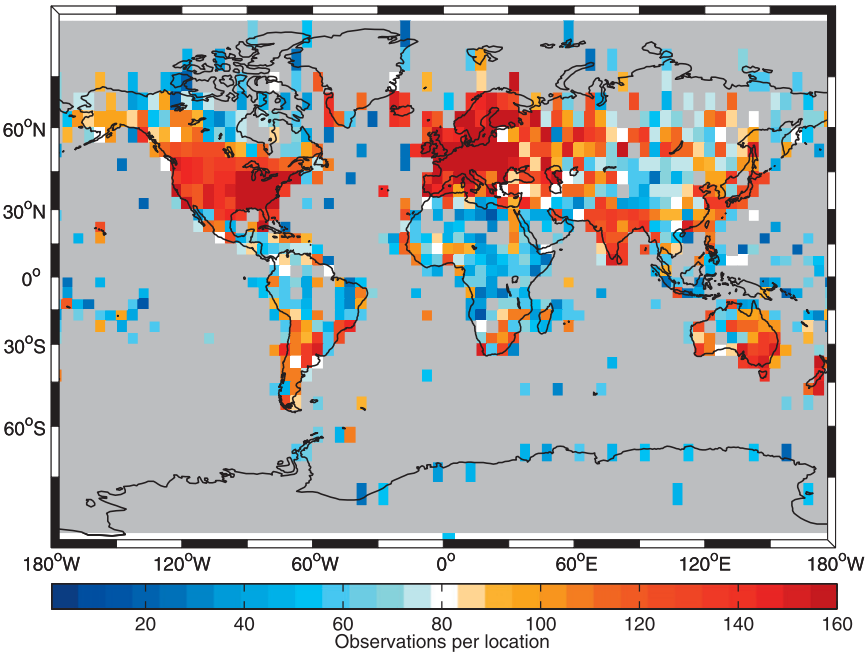

FIG. 5. The number of years in the 1850–2009 interval for which there is an annual mean temperature anomaly observation, as a function of spatial location, calculated from the CRUTEM3 monthly dataset (Brohan et al. 2006).

consists of 111 observations on either side of a 30–time step reference interval.

Results are based on 5000 samples from the posterior distributions of γ, $\boldsymbol{\delta}$, $\boldsymbol{\mu}$, and σ^2, after discarding 600 samples to allow the chain to reach convergence (e.g., Gelman et al. 2004). Details of the hyperparameters used in the prior distributions for the unknown parameters can be found in the appendix.

The elements of the location effects vector $\boldsymbol{\delta}$ are the temporal means of the time series relative to the grand mean γ (Fig. 6a). Point estimates and 90% credible intervals are formed from the median and the 5th and 95th percentiles of the posterior draws, respectively. Nine of the 839 location effects are greater (in magnitude) than 0.5°C, while 108 are greater than 0.25°C. The mean width of the 90% credible intervals is 0.25°C, and 318 of the 90% pointwise credible intervals do not cover zero. The considerable spatial structure in the location effects (Fig. 6) is somewhat surprising, given that the original CRU dataset is already expressed as anomalies from the 1961–90 mean, and that no spatial structure is assumed a priori for the vector $\boldsymbol{\delta}$ (see section 5). The correlation between the medians of the location effects (Fig. 6a) and the number of observations at those locations (Fig. 5) is −0.06, indicating that on a global scale, there is essentially no correlation between data availability and location effect.

The posterior distribution of the year effects (Fig. 6b) is not important in the context of calculating anomalies. Note that the estimated year effects should not be interpreted as an estimate of the temporal evolution of the spatial mean of the temperature field; the year effects $\boldsymbol{\mu}$ are simply the time series that is most common to the dataset, without regard to the spatial distribution of the observations or the pattern of missing data. Estimates of the spatial mean, which make use of simple assumptions about the space–time covariance of the temperature field and account for the temporally changing pattern of data availability, are discussed below (section 4c).

The posterior histograms of the scalar parameters are in all cases sharply peaked relative to the priors, indicating that the posterior is dominated by the information from the data (Fig. 7). Posterior estimates of γ are negative, as the original CRU reference interval of 1961–90 is warm relative to the longer 1850–2009 interval. The parameters σ_δ^2 and σ_μ^2 are related to the variance of the location and mean effects, under priors that enforce the sum-to-zero constraints of Eq. (7); details can be found in appendix A.

b. *Time series of simple means and standard deviations*

For each posterior draw of γ and $\boldsymbol{\delta}$, we calculate the matrix of anomalies $\mathbf{Y}$ via Eq. (8). Although the Bayesian algorithm imputes the missing values $\mathbf{X}^m$ (and thus allows $\mathbf{Y}^m$ to be calculated), for the sake of investigating the effects of changing the reference interval we compare

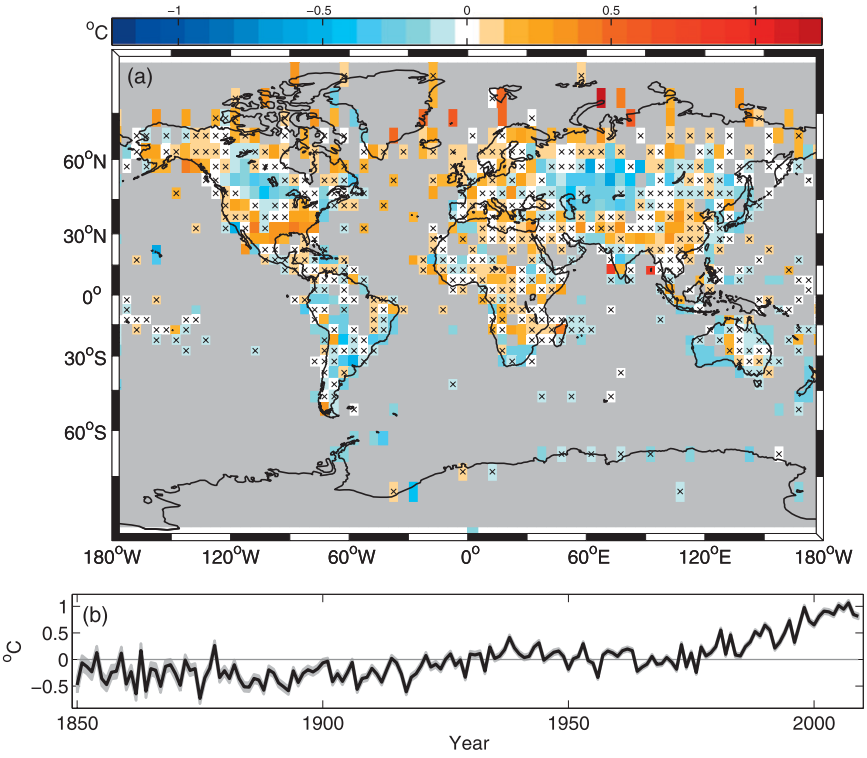

FIG. 6. (a) Posterior medians of the location effects $\boldsymbol{\delta}$. The mean width of the 90% pointwise credible intervals is 0.25°C, and hatching indicates locations where the corresponding 90% credible interval contains zero. (b) Posterior estimates of the year effects, $\boldsymbol{\mu}$. The posterior median is shown in black and 90% pointwise credible intervals in light gray.

means and standard deviations calculated using $\mathbf{X}^o$ (original CRU data; anomalies from 1961 to 1990) and $\mathbf{Y}^o$ (adjusted dataset; anomalies from 1850 to 2009). The time series formed by taking, at each year j, the mean or standard deviation of $\mathbf{X}_j^o$ or $\mathbf{Y}_j^o$ will be referred to simply as the mean or standard deviation time series for that dataset.

Section 2 demonstrated that, for independent AR(1) time series with no missing values, the choice of reference interval does not add temporal structure to the mean time series. Were the CRU data complete, the difference between the mean time series of $\mathbf{X}^o$ and $\mathbf{Y}^o$ would be constant as a function of time and given by the negative of the grand mean, γ (Fig. 8). However, if the spatial distribution of data availability is correlated with the location effects, then the mean time series of $\mathbf{Y}^o$ and $\mathbf{X}^o$ could be very different. For example, if long records have generally positive location effects, then estimates of the mean time series during the early part of the record would be colder, in relation to the later part, after removing the location effects and grand mean from each series. For the CRU dataset, the correlation between the number of observations and the location effects is −0.06, which explains why there is little temporal structure in the difference between the mean time series of $\mathbf{X}^o$ and $\mathbf{Y}^o$ (Fig. 8).

Extending the reference interval from 1961–90 to 1850–2009 increases the standard deviation within the original 1961–90 reference interval, and decreases the standard deviation elsewhere (Fig. 9). Such a result is to be expected given the results from section 2, which indicate that the standard deviation is reduced within a short reference interval and inflated elsewhere. Within (outside of) the original 1961–90 reference interval, the mean of the standard deviation time series of $\mathbf{X}^o$ is 0.67°C (0.81°C), while that of $\mathbf{Y}^o$ is 0.69°C (0.76°C). Re-expressing the dataset as anomalies with respect to the full 1850–1990 interval thus increases the mean standard deviation within the original 1961–90 reference by about 0.02°C and decreases the mean standard deviation elsewhere by about 0.05°C. The total range (difference between highest and lowest value) of the standard deviation time series for $\mathbf{X}^o$ is 0.76°C, while that for $\mathbf{Y}^o$ is 0.66°C. Changing the reference interval thus reduces the total range of standard deviation values by about 13%.

An estimate of the scaled standard deviation range, Δ [see Eq. (5).], for the CRU dataset requires estimates of

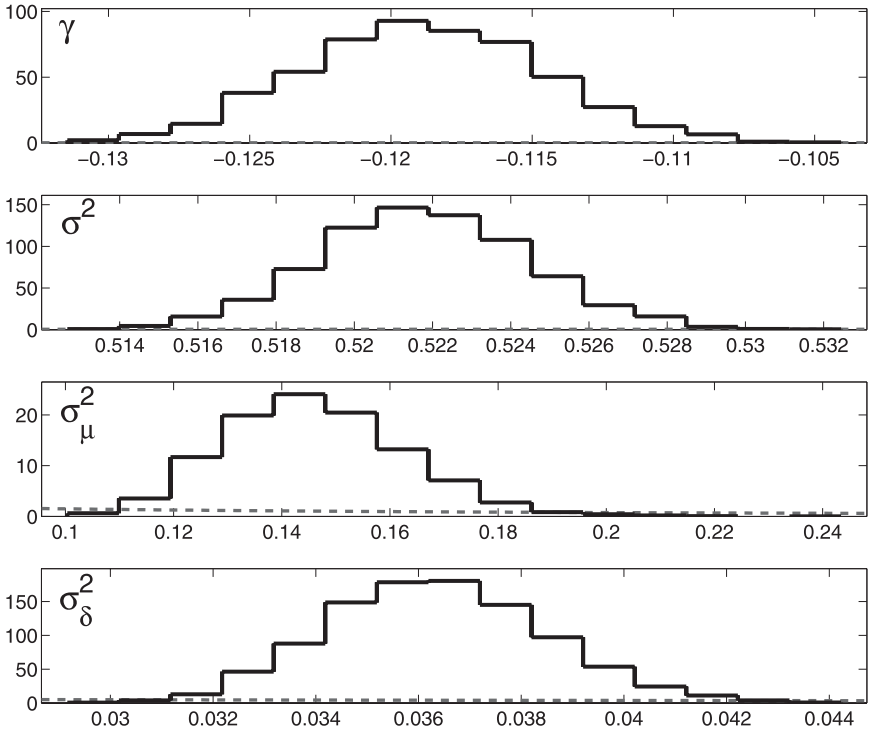

FIG. 7. (top to bottom) Black lines show posterior histograms of the grand mean (γ), error variance (σ^2), and the variances of the year and location effects (σ^2_μ and σ^2_δ). Dashed gray lines show the prior distributions for these parameters.

the maximum and minimum of the standard deviation time series of $\mathbf{X}^o$, and the common standard deviation of $\mathbf{Y}^o$. As each time series of standard deviations is noisy, we estimate $\sqrt{\max(\Omega^*_{ii})}$ and $\sqrt{\min(\Omega^*_{ii})}$ as the mean value of the standard deviation time series of $\mathbf{X}^o$ outside and inside of the 1961–90 reference interval, respectively, and $\sqrt{\Omega_{11}}$ as the mean value of the standard deviation time series of $\mathbf{Y}^o$, giving an estimate of $\Delta = 0.19$ (Fig. 4). Results are unchanged when forming these estimates as the square root of the mean value of the corresponding variance time series. For the CRU data, the anomalies with respect to the 1961–90 reference interval feature a second-moment structure with an amplitude that is 19% of the baseline standard deviation of the anomalies with respect to a reference interval that spans the entire length of the dataset.

c. Time series of spatial means

The posterior distribution of $\boldsymbol{\mu}$ (Fig. 5) gives an estimate of the distribution of the year effects, but should not be interpreted as an estimate of the temporal evolution of the spatial mean of the temperature field. It is, rather, an estimate of the annual effects that are most common to the particular time series under study. Likewise, the mean time series of $\mathbf{X}^o$ and $\mathbf{Y}^o$ (Fig. 8) do not take into account the spatial distribution of the observations, or the spatial and temporal covariance structure of the temperature field.

We estimate the spatial mean time series of the global (ex-Antarctica) land surface temperature anomalies using first the original $\mathbf{X}^o$—the anomalies from 1961 to 1990—and then the pointwise median of $\mathbf{Y}^o$—the anomalies from 1850 to 2009. To account for the changing pattern of data availability and the spatial and temporal covariance of the surface temperature anomaly process, we adopt a hierarchical approach (e.g., Gelman et al. 2004) and infer in each case the spatially and temporally complete field. The *process level* specifies the temperature anomaly field as first-order autoregressive in time, with a spatially constant AR(1) parameter and innovations with covariance that decays exponentially as a function of spatial separation. The resulting space–time covariance form is separable, stationary in both space and time, and isotropic in space (e.g., Banerjee et al. 2004). At the *data level*, the observations are modeled as the true field plus IID normal observational errors. These process- and data-level specifications result in a model that is a special case (no proxy observations) of the Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time (BARCAST) described in Tingley and Huybers (2010),

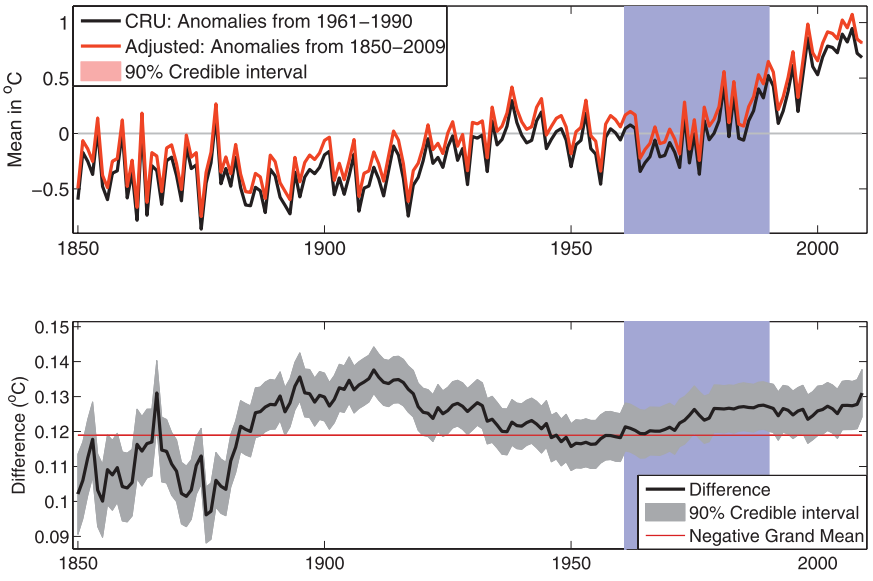

FIG. 8. (top) The mean time series of $\mathbf{X}^o$, the original CRU temperature anomalies using a 1961–90 reference interval (black line) and the mean time series of $\mathbf{Y}^o$, the adjusted anomalies using an 1850–2009 reference interval (red line). Results for the 1850–2009 reference interval are formed by first calculating the anomalies $\mathbf{Y}^o$ via Eq. (8), and then the average across these anomalies, for each posterior draw of γ and $\boldsymbol{\delta}$. Both the medians and the corresponding 90% pointwise credible intervals (light red shading; not readily discernible from the red medians) of the resulting distribution are plotted. (bottom) Median (black line) and 90% credible interval (gray shading) for the difference in the mean time series of each draw of $\mathbf{Y}^o$ and the mean time series of the original $\mathbf{X}^o$.

which is used to infer both model parameters and the spatially complete temperature anomaly field.

To increase the speed of computations, temperatures are inferred at only those 5° by 5°grid boxes that contain a nonzero fraction of land according to a 0.5° by 0.5° land mask (Rodell et al. 2004; data available online at http://ldas.gsfc.nasa.gov/gldas/GLDASvegetation.php). In effect, this decision eliminates a number of the CRU grid

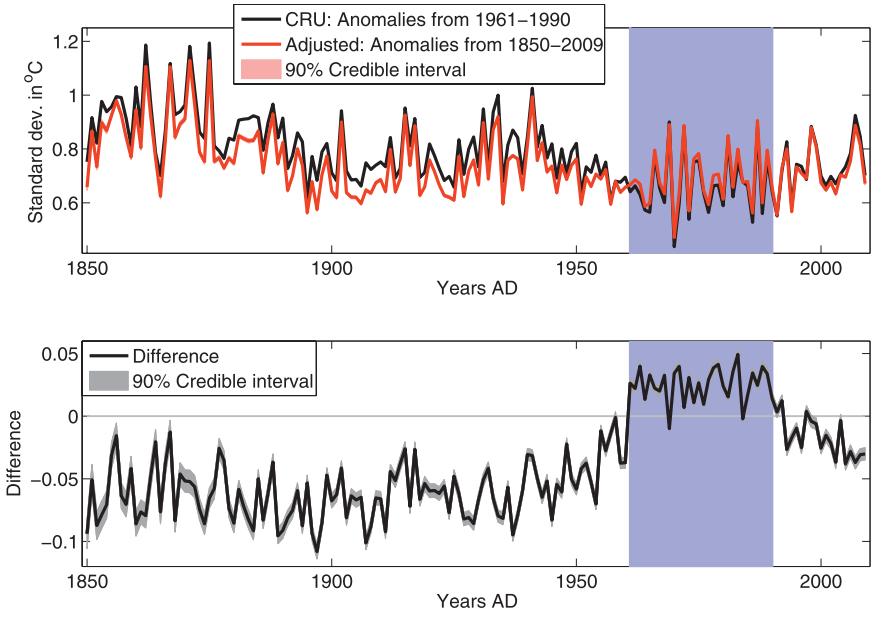

FIG. 9. As in Fig. 8, but for the standard deviation time series of $\mathbf{X}^o$ and $\mathbf{Y}^o$.

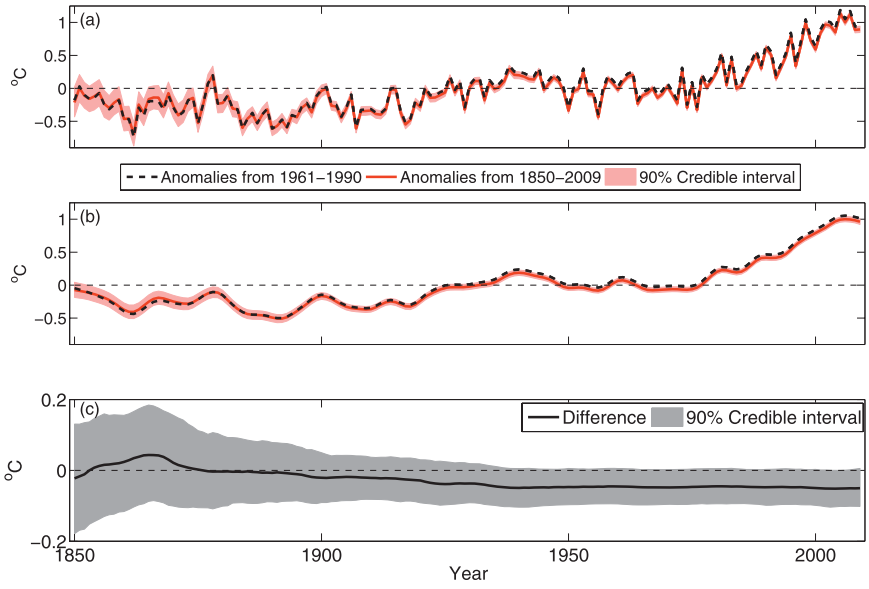

FIG. 10. (a) Time series of spatially averaged global (ex-Antarctica) land surface temperature anomalies based on the annual mean CRU temperature anomalies from the original 1961–90 reference interval (black dashed line) and from the longer 1850–2009 interval (red line). The 90% pointwise credible intervals are shown for the spatial average calculated using the anomalies from the longer 1850–2009 interval (light red shading), and the corresponding uncertainty intervals when using the shorter 1961–90 interval are similar. In both cases, a reduced form of BARCAST (Tingley and Huybers 2010) was used to infer the missing temperature values in space and time, and the temporal mean has been removed from each draw of the spatial mean time series prior to calculating percentiles. (b) As in (a), but each draw of the spatial mean time series from BARCAST is smoothed using a nine-point Hanning window prior to calculating percentiles. (c) The median difference between smoothed draws of the global mean temperature series using the two reference intervals (black line), and the 90% pointwise uncertainty (gray shading).

boxes that are primarily oceanic but contain small remote islands. These grid boxes have a negligible effect on estimates of the spatial mean over land, as they contain very small areas of land and are generally isolated from other landmasses.

To explore the impacts of changing the reference interval on the spatial mean time series, the analysis is conducted in two stages to isolate this effect. First, the 403 annual mean CRU series that are complete from 1950 to 2000 are used to estimate all parameters of the BARCAST model, for both the anomalies from 1961 to 1990 ($\mathbf{X}^o$) and the pointwise median of the anomalies from 1850 to 2009 ($\mathbf{Y}^o$). In each case, we find the posterior sample of the vector of scalar parameters (see Table 1 of Tingley and Huybers 2010) that is closest, according to the Mahalanobis distance, to the median of the ensemble of draws of these parameters. The posterior medians of the AR(1) and variance parameters for the anomalies from the longer 1850–2009 interval, are, respectively, 0.34° and 0.47°C^2, and these values are used in the examples in section 2.

BARCAST is then applied to each anomaly dataset ($\mathbf{X}^o$ and the pointwise median of $\mathbf{Y}^o$), fixing all scalar parameters save the long-term mean according to the first application of BARCAST, to infer temperature anomalies at all nodes of the 5° by 5° grid that contain some fraction of land. The long-term mean parameter is allowed to vary in these second applications, as for both the original and adjusted CRU data, the mean calculated over 1950–2000 is different from that over 1850–2009. The two-stage application of BARCAST allows for much faster computation, as in the first application there is only one pattern of missing data, and in the second only one parameter estimate is updated (see Tingley and Huybers 2010 for details). In effect this is an empirical Bayes' solution, as the uncertainty in the parameter estimates is not taken into account.

Results for each application of BARCAST are based on 2000 draws from the posterior distribution of the temperature process, after discarding 600 samples to allow the chain to reach convergence (e.g., Gelman et al. 2004). The spatial mean time series is calculated for each

posterior draw by weighting each grid box by the area of land it contains. To compare the structure and amplitude of globally (ex-Antarctica) averaged land surface temperature changes inferred using $\mathbf{X}^o$ and the pointwise median of $\mathbf{Y}^o$, we remove from each draw of the spatial mean time series the corresponding draw of the BARCAST mean parameter. A point estimate of each spatial mean time series is then formed by taking the median of the posterior draws for each year, while 90% pointwise credible intervals are formed from the 5th and 95th percentiles (Fig. 10a). We also smooth each of the spatial average time series (after removing the corresponding draw of the BARCAST mean parameter) by a nine-point Hanning window, and calculate the median and 90% pointwise credible intervals (Fig. 10b).

To explore the influence of the reference interval on the rate of long-term temperature change, we take the difference between draws of the temporally smoothed spatial mean time series based on the respective analyses of $\mathbf{X}^o$ and $\mathbf{Y}^o$, and calculate the median and 90% pointwise credible intervals for the time series of differences (Fig. 10c). The smoothed spatial mean time series calculated using $\mathbf{X}^o$ (anomalies from 1961 to 1990) is cooler in the earlier part of the record and warmer in the later part, relative to the spatial mean time series calculated using the pointwise median of $\mathbf{Y}^o$ (anomalies from 1850 to 2009).

The range of the time series of differences in the temporally smoothed spatial means (Fig. 10c) indicates the change in the amplitude of spatial mean temperatures that can be attributed to the choice of reference interval. The range in the pointwise posterior median is 0.1°C, with the maximum at 1866 and the minimum at 2005; the associated 90% credible interval is (−0.03, 0.23). In other words, changing the reference interval from 1961–90 to 1850–2009 reduces the amplitude of the (smoothed) spatial mean time series by about 0.1°C, but as the associated 90% uncertainty interval covers zero, this result is not significant at the 90% level. Indeed, the very weak correlation between the location effects and number of observations at each location suggested that the change in reference interval would not have a large effect on the spatial mean time series.

5. Discussion and conclusions

There are both technical and scientific reasons for analyzing climate datasets after removing from each time series a mean value calculated over a common reference interval. As interest in the temporal evolution of climate variables extends beyond changes in mean values, it is crucial to ensure that the method used to calculate anomalies does not add spurious structures to either the first-moment or higher-moment properties of the dataset. The Bayesian two-factor ANOVA approach to calculating climate anomalies proposed here makes use of all available data, and calculates the location effects over a reference interval that is as long as possible, which avoids the introduction of nonclimatic second-moment structures into the anomalies. Bayesian inference treats the missing values as additional model parameters, and uncertainty estimates for the anomalies include the uncertainty that arises from the missing data.

Several generalizations about the basic analysis model (section 3) are possible. The assumption that the only structure in the location effects is that introduced by the sum-to-zero constraint [Eq. (7)] is plausible for the analysis presented here, as the original CRU data are already expressed as anomalies from a common interval. However, there is clear spatial coherence to the estimated location effects (Fig. 6a), which could be accounted for in future work by modeling the location effects as a spatial process with a standard spatial covariance form (e.g., Banerjee et al. 2004), modified to account for the sum-to-zero constraint. When calculating anomalies from actual values (rather than adjusting the reference interval, as is done here), the model for the location effects should also take into account expected spatial structures by including latitude, elevation, and perhaps other variables as covariates in the expression for the mean of the location effects. The treatment of the year effects could likewise be generalized to include temporal trends in the mean structure and a covariance matrix that accounts for temporal autocorrelation. Finally, the assumption that the error terms are IID normal could be modified to account for any observed spatial or temporal patterns in the residuals, though care must be taken to ensure identifiability when adding structure to the errors as well as the location and year effects.

Using the basic Bayesian ANOVA scheme introduced here to re-express an annually averaged version of the CRU's gridded temperature product as anomalies with respect to means calculated over the entire 1850–2009 interval demonstrates the influence that the choice of reference interval can have on the statistical properties of the anomaly dataset. Relative to the original anomalies with respect to 1961–90, the anomalies with respect to the longer interval display larger spatial variance within the original 1961–90 reference interval, and smaller spatial variance elsewhere. Any analysis of the original CRU data that depends on second-moment properties, such as estimates of spatial patterns of variability or the spatial distribution of extreme values, will thus be affected by second-moment features that are directly attributable to the choice of reference interval. Measured by an

estimate of the scaled standard deviation range [Δ from Eq. (5)], the anomalies with respect to the original 1961–90 interval feature a second-moment structure with an amplitude of about 19% of the magnitude of the mean standard deviation of the anomalies with respect to the longer 1850–2009 interval. Calculations for AR(1) time series with the AR(1) parameter estimated from the CRU anomalies with respect to the longer 1850–2009 interval predicted a qualitatively similar result, but a smaller-scaled standard deviation range of about 6.8%. The larger value found in practice could result from the spatial covariance of the CRU data, which was not accounted for in the ANOVA model or the experiments with AR(1) data in section 2.

For the CRU data, the location effects from the ANOVA analysis are essentially uncorrelated with the number of observations at each location, and as a result, the choice of reference interval has little effect on either the time series of simple means or estimates of the time series of spatial means. In terms of the spatial means, re-expressing the CRU data as anomalies with respect to the longer 1850–2009 reference interval reduces the amplitude of temperature change over the past 160 yr by about 0.1°C, but as the 90% uncertainty interval, $(-0.03, 0.23)$, covers zero, this result is not statistically significant. It is important to emphasize that for other climate datasets, where the location effects may be more strongly correlated with data availability, the choice of reference interval used to calculate the anomalies will influence both the first- and second-moment structures of the resulting anomalies.

Acknowledgments. The content and presentation of the article benefited from discussions with T. Greasby, P. Huybers, D. Nychka, J. Rougier, S. Sain, and B. Shaby, and from the comments of two anonymous reviewers.

APPENDIX

Prior Specification and Posterior Sampling

A Gibbs Sampler (e.g., Gelman et al. 2004) is used to sample from the posterior distribution of the missing values, year and location effects, and scalar parameters of the ANOVA model. We specify conjugate priors for all scalar parameters; given the structure of Eq. (10), it is not necessary to specify priors for the values of the missing observations, $\mathbf{X}^m$. We first specify the functional forms of the priors and full conditional posteriors, and then discuss the hyperparameters used in the analysis of the CRU data. The notation $A|\cdot$ will denote the distribution of the variable A conditional on all other variables.

a. The missing values, X_{ij}^m

No prior is necessary for the missing values; the full conditional posterior for each is normal:

$$X_{ij}^m|\cdot \sim \mathcal{N}(\gamma + \delta_i + \mu_j, \sigma^2). \tag{A1}$$

b. Grand mean, γ

The normal distribution is the conjugate prior:

$$\gamma \sim \mathcal{N}(\mu_\gamma, \sigma_\gamma^2). \tag{A2}$$

The conditional posterior is likewise normal:

$$\gamma|\cdot \sim \mathcal{N}(\Psi_\gamma V_\gamma, \Psi_\gamma), \tag{A3}$$

where

$$V_\gamma = \frac{1}{\sigma^2}\sum_{\text{all}} X_{ij} + \frac{\mu_\gamma}{\sigma_\gamma^2} \tag{A4}$$

and

$$\Psi_\gamma = \left(\frac{NM}{\sigma^2} + \frac{1}{\sigma_\gamma^2}\right)^{-1}. \tag{A5}$$

c. Error variance, σ^2

The inverse-gamma distribution is the conjugate prior:

$$\sigma^2 \sim \text{Inverse-Gamma}(\lambda, \nu)$$
$$P(\sigma^2) \propto (\sigma^2)^{-(\lambda+1)} \exp(-\nu/\sigma^2). \tag{A6}$$

Note that the inverse-gamma prior can be interpreted as 2λ prior observations with an average squared deviation of ν/λ (e.g., Gelman et al. 2004). The conditional posterior is likewise inverse-gamma:

$$\sigma^2|\cdot \sim \text{Inverse-Gamma}\left(\frac{NM}{2} + \lambda, \frac{1}{2}\sum_{\text{all}}(X_{ij} - \gamma - \delta_i - \mu_j)^2 + \nu\right). \tag{A7}$$

d. The location and year effects, $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$

The priors for $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$ must take into account the sum-to-zero constraints of Eq. (7). We follow Kaufman and Sain (2010), setting $\delta \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}^{\delta})$ and $\mu \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}^{\mu})$. Using $\mathbf{I}_M$ to represent the M by M identity and $\mathbf{J}_M$ to represent the M by M matrix of ones, we specify the prior covariances as

$$\boldsymbol{\Sigma}^{\delta} = \sigma_{\delta}^2\left(\mathbf{I}_M - \frac{1}{M}\mathbf{J}_M\right) \quad \text{and} \quad \boldsymbol{\Sigma}^{\mu} = \sigma_{\mu}^2\left(\mathbf{I}_N - \frac{1}{N}\mathbf{J}_N\right). \tag{A8}$$

In other words,

$$\boldsymbol{\Sigma}_{ij}^{\delta} = \begin{cases} \left(1 - \dfrac{1}{M}\right)\sigma_{\delta}^2, & \text{if} \quad i = j \\ -\dfrac{1}{M}\sigma_{\delta}^2, & \text{if} \quad i \neq j \end{cases} \tag{A9}$$

and similarly for $\boldsymbol{\Sigma}^{\mu}$. The sum of the elements of $\boldsymbol{\delta}$ is then distributed as $\mathcal{N}(\mathbf{1}'\mathbf{0}, \mathbf{1}'\boldsymbol{\Sigma}^{\delta}\mathbf{1}) \equiv \mathcal{N}(0, 0)$, which ensures that the sum-to-zero condition is enforced. Posterior sampling is complicated by the singularity of the prior covariance matrices $\boldsymbol{\Sigma}^{\delta}$ and $\boldsymbol{\Sigma}^{\mu}$. Our strategy for sampling the vectors of location and year effects, $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$, which we detail for $\boldsymbol{\delta}$, is based on the presentation in Kaufman and Sain (2010).

The sum-to-zero constraint implies that there are only $M - 1$ free parameters in the length M vector $\boldsymbol{\delta}$. We therefore seek to express $\boldsymbol{\delta}$ as a linear transform of an $M - 1$ dimensional vector of IID normal variables. Letting $\boldsymbol{\delta}^*|\sigma_{\delta}^2 \sim \mathcal{N}(\mathbf{0}_{M-1}, \sigma_{\delta}^2\mathbf{I}_{M-1})$, we must find an M by $M - 1$ matrix $\mathbf{Q}_M$ such that $\boldsymbol{\delta} = \mathbf{Q}_M\boldsymbol{\delta}^*$ has the required covariance form [Eq. (A9)].

Following Kaufman and Sain (2010), we define the M by $M - 1$ matrix $\mathbf{Q}_M^*$ as columns of Helmert contrasts, which compare the effect of one level of a factor to the mean of the preceding factors (e.g., Ruberg 1989). As an example, for $M = 4$,

$$\mathbf{Q}_4^* = \begin{pmatrix} 0 & 0 & -3 \\ 0 & -2 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \tag{A10}$$

Now solve for an $M - 1$ by $M - 1$ diagonal matrix $\mathbf{R}_M$, where $\mathbf{Q}_M = \mathbf{Q}_M^*\mathbf{R}_M$, such that $\boldsymbol{\delta} = \mathbf{Q}_M\boldsymbol{\delta}^*$ has the correct covariance form. That is, solve for $\mathbf{R}_M$ that satisfies

$$\mathbf{Q}_M^*\mathbf{R}_M\mathbf{R}_M\mathbf{Q}_M^{*\mathrm{T}} = \boldsymbol{\Sigma}^{\delta} = \mathbf{I}_M - \frac{1}{M}\mathbf{J}_M. \tag{A11}$$

Solving for $\mathbf{R}_M$ yields

$$\begin{aligned} R_M &= \left[(\mathbf{Q}_M^{*\mathrm{T}}\mathbf{Q}_M^*)^{-1}\mathbf{Q}_M^{*\mathrm{T}}\left(\mathbf{I}_M - \frac{1}{M}\mathbf{J}_M\right)\mathbf{Q}_M^*(\mathbf{Q}_M^{*\mathrm{T}}\mathbf{Q}_M^*)^{-1}\right]^{1/2} \\ &= \left[(\mathbf{Q}_M^{*\mathrm{T}}\mathbf{Q}_M^*)^{-1}\mathbf{Q}_M^{*\mathrm{T}}\mathbf{Q}_M^*(\mathbf{Q}_M^{*\mathrm{T}}\mathbf{Q}_M^*)^{-1}\right]^{1/2} \\ &= (\mathbf{Q}_M^{*\mathrm{T}}\mathbf{Q}_M^*)^{-1/2}. \end{aligned}$$

The first line follows from left-multiplying Eq. (A11) by $(\mathbf{Q}_M^{*\mathrm{T}}\mathbf{Q}_M^*)^{-1}\mathbf{Q}_M^{*\mathrm{T}}$, right-multiplying by $\mathbf{Q}_M^*(\mathbf{Q}_M^{*\mathrm{T}}\mathbf{Q}_M^*)^{-1}$, and then taking the matrix square root. The second line follows from the fact that each column of $\mathbf{Q}_M^*$ sums to zero, so that $\mathbf{Q}_M^*\mathbf{J}_M = \mathbf{0}_M$. Inserting this form for $\mathbf{R}_M$ into the definition of $\mathbf{Q}_M$ gives

$$\mathbf{Q}_M = \mathbf{Q}_M^*(\mathbf{Q}_M^{*\mathrm{T}}\mathbf{Q}_M^*)^{-1/2}. \tag{A13}$$

Kaufman and Sain (2010) indicate that each column of a matrix of Helmert contrasts needs to be scaled by some factor, and demonstrate this for the $M = 3$ case, but do not provide the general formulas derived here.

To produce samples from the conditional posterior of $\boldsymbol{\delta}$, we first draw from the conditional posterior of $\boldsymbol{\delta}^*$, and then transform it using the expression for $\mathbf{Q}_M$. The prior for $\boldsymbol{\delta}^*$ given σ_{δ}^2 is normal:

$$\delta^*|\sigma_{\delta}^2 \sim \mathcal{N}(\mathbf{0}_{M-1}, \sigma_{\delta}^2\mathbf{I}_{M-1}). \tag{A14}$$

The full conditional posterior for $\boldsymbol{\delta}^*$ is likewise normal:

$$\delta^*|\cdot \sim \mathcal{N}(\Psi_{\delta^*}\mathbf{V}_{\delta^*}, \Psi_{\delta^*}\mathbf{I}_{M-1}), \tag{A15}$$

where

$$\mathbf{V}_{\delta^*} = \frac{1}{\sigma^2} \cdot \mathbf{Q}_M^{\mathrm{T}} \cdot \sum_{j=1}^{N}(\mathbf{X}_{\cdot j} - \gamma\mathbf{1}_M) \tag{A16}$$

and

$$\Psi_{\delta^*} = \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_{\delta}^2}\right)^{-1}. \tag{A17}$$

The calculation proceeds by substituting $\mathbf{Q}_M$ for $\boldsymbol{\delta}$ in Eq. (11), and factoring the joint distribution of the elements of $\mathbf{X}$ as a product of N multivariate normals, one for each time interval.

The treatment of $\boldsymbol{\mu}$ is equivalent to that for $\boldsymbol{\delta}$. The prior for $\boldsymbol{\mu}^*$, given σ_{μ}^2, is normal:

$$\mu^*|\sigma_{\mu}^2 \sim \mathcal{N}(\mathbf{0}_{N-1}, \sigma_{\mu}^2\mathbf{I}_{N-1}). \tag{A18}$$

The full conditional posterior is likewise normal:

$$\mu^*|\cdot \sim \mathcal{N}(\Psi_{\mu^*}\mathbf{V}_{\mu^*}, \Psi_{\mu^*}\mathbf{I}_{N-1}), \tag{A19}$$

where

$$\mathbf{V}_{\mu^*} = \frac{1}{\sigma^2}\cdot \mathbf{Q}_N^{\mathrm{T}} \sum_{i=1}^{M} (\mathbf{X}_{i\cdot} - \gamma \mathbf{1}_N) \tag{A20}$$

and

$$\Psi_{\mu^*} = \left(\frac{M}{\sigma^2} + \frac{1}{\sigma_\mu^2}\right)^{-1}. \tag{A21}$$

e. Variances of the location and year effects, σ_δ and σ_μ

The conjugate priors are inverse-gamma:

$$\begin{aligned}\sigma_\delta^2 &\sim \text{Inverse-Gamma}(\lambda_\delta, \nu_\delta)\\ \sigma_\mu^2 &\sim \text{Inverse-Gamma}(\lambda_\mu, \nu_\mu).\end{aligned} \tag{A22}$$

The full conditional posteriors are likewise inverse-gamma:

$$\begin{aligned}\sigma_\delta^2|\cdot &= \text{Inverse-Gamma}\left(\frac{M-1}{2} + \lambda_\delta, \frac{\delta^{*\mathrm{T}}\delta^*}{2} + \nu_\delta\right)\\ \sigma_\mu^2|\cdot &= \text{Inverse-Gamma}\left(\frac{N-1}{2} + \lambda_\mu, \frac{\mu^{*\mathrm{T}}\mu^*}{2} + \nu_\mu\right).\end{aligned} \tag{A23}$$

f. Hyperparameters for the analysis of the CRU data

Preliminary analysis of the observed values is used to set the parameters of the prior distributions for γ, σ^2, σ_δ^2, and σ_μ^2. An initial estimate of the grand mean (γ) is formed as the mean of $\mathbf{X}^o$, while the location effects are estimated as the temporal means of $\mathbf{X}^o$ minus the estimated grand mean, and likewise for the year effects. An estimate of the error variance can be formed by first estimating each element of $\mathbf{X}^o$ as the sum of the grand mean and corresponding location and year effects, taking the difference between these estimates and the observed values, and then taking the variance of the resulting residuals. The hyperparameters are then set as follows:

- Grand mean, γ: set the prior mean, μ_γ, to the mean of all available observations, and the prior variance σ_γ^2 to 16 times the estimated variance.
- Prior for the error variances, σ^2: set λ to 1/2, and ν to half the estimated residual variance. These parameters correspond to one prior observation with an average squared deviation given by the estimated residual variance.
- Factor variances, σ_δ^2 and σ_μ^2: set $\lambda_{\delta,\mu}$ to 1/4 and set $\mu_{\delta,\mu}$ to one-fourth the variance of the estimated effect vectors. In each case, these parameters correspond to half a prior observation with an average squared deviation given by the sample variance.

REFERENCES

Banerjee, S., B. P. Carlin, and A. E. Gelfand, 2004: *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC, 452 pp.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.,* **111,** D12106, doi:10.1029/2005JD006548.

Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977: Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc.,* **39,** 1–38.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 2004: *Bayesian Data Analysis*. 2nd ed. Chapman & Hall/CRC, 668 pp.

Hansen, J., and S. Lebedeff, 1987: Global trends of measured surface air temperature. *J. Geophys. Res.,* **92,** 13 345–13 372.

Jansen, E., and Coauthors, 2007: Palaeoclimate. *Climate Change 2007: The Physical Science Basis,* S. Solomon et al., Eds., Cambridge University Press, 435–497.

Jones, P., M. New, D. Parker, S. Martin, and I. Rigor, 1999: Surface air temperature and its changes over the past 150 years. *Rev. Geophys.,* **37,** 173–199.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Kaufman, C., and S. Sain, 2010: Bayesian functional ANOVA modeling using Gaussian process prior distributions. *Bayesian Anal.,* **5** (1), 123–150.

Mann, M. E., and Coauthors, 2009: Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly. *Science,* **326,** 1256–1260.

Rodell, M., and Coauthors, 2004: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.,* **85,** 381–394.

Ruberg, S., 1989: Contrasts for identifying the minimum effective dose. *J. Amer. Stat. Assoc.,* **84,** 816–822.

Scheffé, H., 1999: *The Analysis of Variance*. Wiley-Interscience, 477 pp.

Tingley, M. P., and P. Huybers, 2010: A Bayesian algorithm for reconstructing climate anomalies in space and time. Part I: Development and applications to paleoclimate reconstruction problems. *J. Climate,* **23,** 2759–2781.

Trenberth, K. E., and Coauthors, 2007: Observations: Surface and atmospheric climate change. *Climate Change 2007: The Physical Science Basis,* S. Solomon et al., Eds., Cambridge University Press, 235–336.

Watson, R. T., and Coauthors, Eds., 2001: *Climate Change 2001: Synthesis Report*. Cambridge University Press, 397 pp.

Zar, J. H., 1999: *Biostatistical Analysis*. 4th ed. Prentice Hall, 929 pp.