A Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time. Part II: Comparison with the Regularized Expectation–Maximization Algorithm

Martin P. Tingley

Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, North Carolina, and Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts

Peter Huybers

Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts

ABSTRACT

Part I presented a Bayesian algorithm for reconstructing climate anomalies in space and time (BARCAST). This method involves specifying simple parametric forms for the spatial covariance and temporal evolution of the climate field as well as "observation equations" describing the relationships between the data types and the corresponding true values of the climate field. As this Bayesian approach to reconstructing climate fields is new and different, it is worthwhile to compare it in detail to the more established regularized expectation–maximization (RegEM) algorithm, which is based on an empirical estimate of the joint data covariance matrix and a multivariate regression of the instrumental time series onto the proxy time series. The differing assumptions made by BARCAST and RegEM are detailed, and the impacts of these differences on the analysis are discussed. Key distinctions between BARCAST and RegEM include their treatment of spatial and temporal covariance, the prior information that enters into each analysis, the quantities they seek to impute, the end product of each analysis, the temporal variance of the reconstructed field, and the treatment of uncertainty in both the imputed values and functions of these imputations. Differences between BARCAST and RegEM are illustrated by applying the two approaches to various surrogate datasets. If the assumptions inherent to BARCAST are not strongly violated, then in scenarios comparable to practical applications BARCAST results in reconstructions of both the field and the spatial mean that are more skillful than those produced by RegEM, as measured by the coefficient of efficiency. In addition, the uncertainty intervals produced by BARCAST are narrower than those estimated using RegEM and contain the true values with higher probability.

1. Introduction

From a statistical perspective, reconstructing the temporal evolution of a climate field from overlapping time series of instrumental and proxy observations is a challenging problem. Instrumental and proxy records of climate fields are invariably incomplete with respect to their coverage in both time and space, necessitating some statistical method for spatial and temporal in filling. In addition, the instrumental records are used to both estimate the climate field under analysis and to determine the relationship between the available proxy records and the field. The goal in this context is to assimilate the available instrumental and proxy information to estimate, with uncertainties, the temporal evolution of a climate field in some optimal manner. While various methodologies have been explored, it is safe to say that there remains significant scope for further testing and development of methodologies for reconstructing and interpreting past climate variability (NRC 2006; Jansen et al. 2007; Jones et al. 2009).

Tingley and Huybers (2010, hereafter Part I) developed a hierarchical Bayesian approach to reconstructing climate fields, referred to as BARCAST for "A Bayesian algorithm for reconstructing climate anomalies in space and time." This approach is based on specifying parametric forms for the spatial covariance and temporal

Corresponding author address: Martin P. Tingley, Department of Earth and Planetary Sciences, Harvard University, 20 Oxford St., Cambridge, MA 02138.
E-mail: tingley@fas.harvard.edu

evolution of the field as well as the relationships between the data types and the field. (See Part I for a detailed description of BARCAST. A package of Matlab code that implements the algorithm is available at ftp://ftp.ncdc. noaa.gov/pub/data/paleo/softlib/barcast/.) As BARCAST is new and different from other approaches to reconstructing climate fields, it makes sense to compare it against a method that is well established.

Most approaches to the climate field reconstruction problem are based on a multivariate regression of the instrumental time series onto the proxy time series during a calibration period (e.g., Mann et al. 1998; Schneider 2001; Cook et al. 1999; Luterbacher et al. 2004; Jones et al. 2009). The coefficients are then used to predict the values of the missing instrumental observations back through time using the available proxy time series. At the heart of these methods is the estimation of the mean of each time series and the joint covariance matrix of the instrumental and proxy datasets—a submatrix of which must be inverted to calculate the regression coefficients. If the length of the overlap between the instrumental and proxy datasets is short relative to the number of time series—as is often the case—the estimate of the covariance matrix of the instrumental and proxy datasets is far from certain, and the requisite matrix inversion generally not possible without some form of conditioning or regularization. In addition, the proxy time series are generally of different lengths, which complicates the estimation of the mean and covariance.

The regularized expectation–maximization (RegEM) algorithm (Schneider 2001), developed to overcome these difficulties, has been applied extensively to climate field reconstruction problems (e.g., Rutherford et al. 2003, 2005; Mann et al. 2007b, 2008; Steig et al. 2009; Zhang et al. 2004). This algorithm combines several well known statistical techniques: the expectation–maximization algorithm (Dempster et al. 1977) and regularized regression, either ridge regression (Hoerl and Kennard 1970) or truncated total least squares regression (van Huffel and Vandewalle 1991; Fierro et al. 1997).

In this study, we compare the assumptions and behavior of RegEM and BARCAST to provide insight into the novel features, strengths, and weaknesses of this new approach to the climate reconstruction problem and to position these developments within the context of previous work. While RegEM is by no means the only technique being used to reconstruct climate fields—other methods include those of Mann et al. (1998); Cook et al. (1999), and Luterbacher et al. (2004)—it is well established in the literature and seems the most statistically sophisticated method that has been widely applied.

The more theoretical comparisons in this study complement the practical comparisons of the two methods presented in Part I.

Section 2 briefly describes the technical aspects of RegEM in a manner that facilitates comparisons with BARCAST, section 3 compares the assumptions and methods of BARCAST to those of RegEM, section 4 compares the results of applying variants of the two analysis strategies to simple surrogate datasets, and section 5 provides discussion and concluding remarks.

2. The RegEM algorithm

While the technical details of RegEM are described in detail elsewhere (see, e.g., Schneider 2001; Mann et al. 2007b), it is convenient for the purposes of comparison to summarize the main ideas behind this approach. We first describe the expectation–maximization (EM) algorithm and explain its shortcomings in the context of climate reconstructions, then briefly describe the two regularized regression techniques and how each of them influences the results of the EM algorithm.

a. Expectation–maximization algorithm

The EM algorithm (Dempster et al. 1977; Gelman et al. 2003) is an iterative technique for estimating distribution parameters and imputing missing values for incomplete datasets. To illustrate the main concepts, consider a number M of variables, assumed to follow a multivariate normal distribution, and a number N of independent samples of these variables. In the climate context, the variables could be, for example, annual mean temperature observations, both instrumental and proxy, at a large number of spatial locations, and the samples correspond to observations for different years. Some percentage of the dataset is missing, and the missing data mechanism is assumed to be *ignorable*. Ignorability requires, in a Bayesian sense, that the probability that data points are missing be a function only of fully observed covariates, the observed data, and the parameters governing the missing data process (Rubin 1976; Gelman et al. 2003).

Given a complete data matrix in which all time series span the same years and there are no missing values, the mean vector and covariance matrix can be estimated in a straightforward manner. Similarly, given a year of incomplete data, the missing values can be imputed using the available values for that year, the mean vector, and the covariance matrix as the conditional expectation of the missing values given the observed values. The EM algorithm, initialized with some estimate of the full mean vector and covariance matrix of the incomplete dataset (e.g., using all available data), iterates two steps:

1) In the expectation step, missing values for each incomplete sample are imputed as the conditional expectation of the missing variables given the observed variables and the current estimates of the mean and covariance matrix.

2) In the maximization step, the maximum likelihood estimates (MLEs) of the mean and covariance matrix are formed from the data matrix completed with the most recently imputed values, noting that, as the imputed values are conditional expectations, the conditional variances of the missing values must be added to the estimate of the covariance matrix.

Details of the formulas involved can be found in standard references (e.g., Gelman et al. 2003). For the purposes of this development, the key idea is that the expectation step is a multiple linear regression. We make use of the notation

$$[\mathbf{X}_o, \mathbf{X}_m] \sim \mathrm{N}\left[(\mu_o, \mu_m), \begin{pmatrix} \mathbf{\Sigma}_{oo} & \mathbf{\Sigma}_{om} \\ \mathbf{\Sigma}_{mo} & \mathbf{\Sigma}_{mm} \end{pmatrix}\right], \qquad (1)$$

where $\mathbf{X}_o$ and $\mathbf{X}_m$ are row vectors of length M_o and M_m (where $M_o + M_m = M$) and represent the observed and missing values, respectively, for a particular year, and μ and $\mathbf{\Sigma}$ are the population mean and covariance (which have been partitioned). The distribution of $\mathbf{X}_m|\mathbf{X}_o$ is normal, with the mean and variance following standard forms (e.g., Anderson 2003):

$$\mathbf{X}_m|\mathbf{X}_o, \mu, \mathbf{\Sigma}$$
$$\sim \mathrm{N}(\mu_m + (\mathbf{X}_o - \mu_o)\mathbf{\Sigma}_{oo}^{-1}\mathbf{\Sigma}_{om}, \mathbf{\Sigma}_{oo} - \mathbf{\Sigma}_{mo}\mathbf{\Sigma}_{oo}^{-1}\mathbf{\Sigma}_{om}). \quad (2)$$

The expectation step of the EM algorithm uses the current estimates of the joint mean vector ($\hat{\mu}$; quantities estimated from data will be indicated with hats) and covariance matrix ($\hat{\mathbf{\Sigma}}$) to impute the missing values at each year as the conditional expectation of the missing values, given the observed values. The imputation has the form of an MLE prediction from a linear regression, with the estimate of the coefficient matrix given by $\hat{\mathbf{\beta}} \equiv \hat{\mathbf{\Sigma}}_{oo}^{-1}\hat{\mathbf{\Sigma}}_{om}$.

If the number of predictor variables exceeds the number of samples of the variables available to estimate the covariance matrix of the predictors, then the sample estimate of the predictor covariance matrix $\hat{\mathbf{\Sigma}}_{oo}$ is rank deficient, its inverse does not exist, and the estimate $\hat{\mathbf{\beta}}$ is undefined. This is the case in the climate reconstruction problem if the total number of proxy variables exceeds the number of years in the overlap between the instrumental and proxy datasets. A similar problem can arise if the data time series are correlated, in which case $\hat{\mathbf{\Sigma}}_{oo}$ can be nearly singular (at least one eigenvalue very close to zero).

A number of techniques exist to regularize underdetermined or ill-posed regression problems; we describe both ridge regression and truncated total least squares (T-TLS) regression. In the original description of the RegEM algorithm, Schneider (2001) makes use of ridge regression to provide the regularization, arguing that the continuous eigenvalue filtering offered by ridge regression has advantages over the discrete set of truncation values offered by T-TLS (see below). Several studies have found that reconstructions performed with the ridge-regularized RegEM are sensitive to the standardization applied to the data prior to analysis, and that the estimation of the optimal ridge parameter can be poorly constrained (Mann et al. 2007a,b; Smerdon and Kaplan 2007). These issues lead Mann et al. (2007b) to use T-TLS to provide the regularization in RegEM. There is a suggestion, however, that the shortcomings identified in ridge-regularized RegEM result from a nonignorable missing data structure, rather than the method, so will be present regardless of the regularization strategy (Smerdon et al. 2008).

b. Ridge regression

The basic idea behind ridge regression (Hoerl and Kennard 1970), also called Tikhonov regression (Tikhonov and Arsenin 1977), is the substitution of

$$(\hat{\mathbf{\Sigma}}_{oo} + h^2\mathbf{D})^{-1} \quad \text{for} \quad \hat{\mathbf{\Sigma}}_{oo}^{-1}, \qquad (3)$$

where $\mathbf{D}$ is a diagonal matrix. Schneider (2001) sets $\mathbf{D}$ to the diagonal of $\hat{\mathbf{\Sigma}}_{oo}$, and we follow this choice in the development below. By inflating the diagonal of $\hat{\mathbf{\Sigma}}_{oo}$, the ridge procedure regularizes the regression by ensuring that the necessary matrix inverse exists.

The ridge regularized estimate of the regression coefficient matrix can be written as

$$\mathbf{\beta}_h^* = \mathbf{D}^{-1/2}(\mathbf{D}^{-1/2}\hat{\mathbf{\Sigma}}_{oo}\mathbf{D}^{-1/2} + h^2\mathbf{I})^{-1}\mathbf{D}^{-1/2}\hat{\mathbf{\Sigma}}_{om}, \qquad (4)$$

where $\mathbf{D}^{-1/2}\hat{\mathbf{\Sigma}}_{oo}\mathbf{D}^{-1/2} \equiv \tilde{\mathbf{\Sigma}}_{oo}$ is the sample correlation matrix of the predictors. In other words, ridge regression, as applied by Schneider (2001), involves adding a matrix proportional to the identity to the sample correlation matrix of the predictors. In the paleoclimate context, the predictors are the observed proxy and instrumental variables for a given year, and for most years in a reconstruction, only proxy observations will be available. The ridge estimate $\mathbf{\beta}_h^*$ is biased toward underestimating the magnitude of the elements of $\mathbf{\beta}$. In the case of a univariate response, so that the coefficient matrix reduces to a vector, setting $h > 0$ results in a smaller solution, in the sense that $\beta_h^{*\mathrm{T}}\beta_h^* \leq \hat{\beta}^{\mathrm{T}}\hat{\beta}$ (Hoerl and Kennard 1970).

The ridge estimate, however, can reduce the expected mean squared error (MSE) of predictions relative to the standard MLE solution, which results if h is set to zero (Hoerl and Kennard 1970). The appendix presents a geometric interpretation of ridge regression and an example illustrating the effects of the regularization in a simple case.

Poor regression models are known to result if the predictor variables are strongly correlated or if the number of predictor variables is not much smaller than the number of replicates (e.g., Zar 1999; Devore 1995); in both of these cases, at least one of the eigenvalues of the sample covariance matrix is small. The resulting MLE of the regression vector can be large and is sensitive to small changes in the values of the predictor variables. A positive value of h stabilizes the matrix inversion by putting a lower bound on the eigenvalues of the sample covariance matrix, which reduces the magnitude of the estimated regression vector. As the value of h increases, the *bias* in the estimate of the regression vector monotonically increases, while the *variance* decreases (Hoerl and Kennard 1970). In the limit $h \to \infty$, the estimates of the coefficients converge to zero, as does the variance of these estimates, while the bias saturates. As the MSE is given by the sum of the squared bias and the variance (e.g., Casella and Berger 2002), the possibility exists that a positive value of h will result in an estimate with lower MSE than the MLE (see Fig. 1 in Hoerl and Kennard 1970). In practice, accepting a small amount of bias often permits a substantial reduction in the variance of the estimated regression vector, and thus reduces the MSE of the estimate. Intuitively, the idea is to limit the sensitivity of the estimates of the regression coefficients to noise and spurious correlations, thereby reducing the expected MSE of predictions while limiting the bias.

Ridge regression, as applied by Schneider (2001), can be interpreted as smoothly scaling the weights associated with the eigenvectors of the sample correlation matrix of the observed values, $\tilde{\boldsymbol{\Sigma}}_{oo}$. Weights corresponding to eigenvalues of $\tilde{\boldsymbol{\Sigma}}_{oo}$ that are large relative to h^2 are only mildly affected, while weights corresponding to eigenvalues small relative to h^2 are smoothly scaled toward zero (Schneider 2001). In the climate reconstruction context, RegEM generally uses proxy observations to impute the missing instrumental observations so that $\tilde{\boldsymbol{\Sigma}}_{oo}$ is the sample correlation matrix of the proxy time series. RegEM with ridge regularization thus involves smoothly filtering the weights associated with the eigenvectors of the sample correlation matrix of the proxy time series.

To estimate the ridge-regularization parameter, Schneider (2001) makes use of a generalized cross validation procedure (Golub et al. 1979; Krakauer et al.

2004a,b) based on minimizing the expected MSE of predictions. In practice, all the missing values for a given year can be imputed using one ridge parameter and a multiple regression, or the missing values for a given year can each be imputed separately using a number of simple regressions, each with a distinct ridge parameter.

c. Truncated total least squares

To illustrate T-TLS, consider the standard regression problem:

$$\mathbf{X}_m = \mathbf{X}_o \boldsymbol{\beta} + \epsilon,$$

where $\mathbf{X}_m$ is an N by M_m response matrix, $\mathbf{X}_o$ is an N by M_o predictor matrix, $\boldsymbol{\beta}$ is the M_o by M_m coefficient matrix, and ϵ is an N by M_m noise term. Total least squares regression seeks an estimate of the coefficient matrix $\boldsymbol{\beta}$ that solves

$$\min \|(\mathbf{X}_o, \mathbf{X}_m) - (\hat{\mathbf{X}}_o, \hat{\mathbf{X}}_m)\|_F \quad \text{subject to} \quad \hat{\mathbf{X}}_m = \hat{\mathbf{X}}_o \hat{\boldsymbol{\beta}}, \quad (5)$$

where $\|\cdot\|_F$ indicates the Frobenius norm (Fierro et al. 1997); see Golub and Van Loan (1980) for a more general description of total least squares regression. This is in contrast to ordinary least squares regression, which seeks only to minimize the variance of the residual $\hat{\mathbf{X}}_m - \mathbf{X}_m$, while assuming the predictor matrix $\mathbf{X}_o$ is constant or fixed. The total least squares approach is designed for so-called errors in variables models, in which the predictor variables, as well as the response variables, are assumed to contain errors (van Huffel and Vandewalle 1991).

There are many ways of describing the T-TLS approach (see Fierro et al. 1997, for an alternative description). For the sake of simplicity, we assume that each variable has a mean of zero. In this case, the scaled inner products between the columns of the joint data matrix are estimates of the elements of the joint covariance matrix:

$$\frac{1}{N-1}(\mathbf{X}_o, \mathbf{X}_m)^{\mathrm{T}}(\mathbf{X}_o, \mathbf{X}_m) = \begin{pmatrix} \hat{\boldsymbol{\Sigma}}_{oo} & \hat{\boldsymbol{\Sigma}}_{om} \\ \hat{\boldsymbol{\Sigma}}_{mo} & \hat{\boldsymbol{\Sigma}}_{mm} \end{pmatrix} \equiv \mathbf{V}\boldsymbol{\Lambda}^2\mathbf{V}^{\mathrm{T}},$$

$$(6)$$

where $\boldsymbol{\Lambda}^2$ is diagonal and composed of the eigenvalues, arranged from largest to smallest, of the joint covariance matrix of the response and predictor variables, and $\mathbf{V}$ is the corresponding matrix of eigenvectors. The same eigenvector matrix $\mathbf{V}$ can be obtained from a singular value decomposition of $[\mathbf{X}_o, \mathbf{X}_m]$. If the problem is underdetermined, some of the eigenvalues will be zero; if the problem is poorly conditioned (nearly colinear predictors or response variables), some of the eigenvalues will be very small. The idea behind T-TLS is to retain only

the eigenvectors corresponding to eigenvalues above some cutoff. In other words, some number $\nu < M_o + M_m$ of the eigenvectors of the joint [predictor, response] sample covariance matrix are used to predict the regression coefficient matrix $\boldsymbol{\beta}$. Denoting the upper-left $\nu \times \nu$ submatrix of $\boldsymbol{\Lambda}^2$ by $\boldsymbol{\Lambda}_\nu^2$, and the first ν columns of $\mathbf{V}$ by $\mathbf{V}_\nu$, define $\mathbf{H}_\nu$ as

$$\mathbf{V}_\nu \boldsymbol{\Lambda}_\nu^2 \mathbf{V}_\nu^T \equiv \mathbf{H}_\nu = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix}, \qquad (7)$$

where $\mathbf{H}_{11}$ is M_o by M_o and $\mathbf{H}_{22}$ is M_m by M_m. $\mathbf{H}_\nu$ is the truncated representation of $\hat{\boldsymbol{\Sigma}}$ using only the ν largest eigenvalues, and thus the pseudoinverse $\mathbf{H}_{11}^\dagger$ of $\mathbf{H}_{11}$ is an approximation of $\hat{\boldsymbol{\Sigma}}_{oo}^{-1}$, and $\mathbf{H}_{12}$ is an approximation of $\hat{\boldsymbol{\Sigma}}_{om}$. The T-TLS estimate of the regression coefficient matrix $\boldsymbol{\beta}$ is then [cf. Eq. (2)]

$$\hat{\boldsymbol{\beta}}_\nu = \mathbf{H}_{11}^\dagger \mathbf{H}_{12}. \qquad (8)$$

If the regression problem is overdetermined (i.e., there are more records than variables, $N > M_m + M_o = M$) and well conditioned (predictors or responses not close to colinear, so the eigenvalues are not too close to zero), then there are several special cases of the T-TLS approach:

- If $\nu = M$, the resulting $\hat{\boldsymbol{\beta}}_M$ is simply the MLE $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Sigma}}_{oo}^{-1}\hat{\boldsymbol{\Sigma}}_{om}$ [cf. Eq. (2)].
- If $\nu = M_o$, the resulting $\hat{\boldsymbol{\beta}}_{M_o}$ is the standard total least squares estimate, which, if the uncertainties in both predictor and response variables are the same, minimizes the mean square orthogonal distance from the data points to the line of best fit (van Huffel and Vandewalle 1991; Golub and Van Loan 1980).
- If $\nu < M_o$, the resulting $\hat{\boldsymbol{\beta}}_\nu$ is labeled by Fierro et al. (1997) as a truncated total least squares estimate.

If the system is underdetermined—that is, the rank of $[\mathbf{X}_o, \mathbf{X}_m] < M$, which results if $N < M$, or the predictors are colinear—then the truncated total least squares solution results if $\nu < \min(M_o, \text{rank}\ [\mathbf{X}_o, \mathbf{X}_m])$.

The formulation $\hat{\boldsymbol{\beta}}_\nu = \mathbf{H}_{11}^\dagger \mathbf{H}_{12}$ shows that T-TLS results in a filtered solution, in the sense that the eigenvectors of the joint covariance matrix corresponding to small eigenvalues are not used in the estimation of $\hat{\boldsymbol{\beta}}_\nu$. In the context of climate reconstruction problems, RegEM regularized with T-TLS involves predicting the missing instrumental values using only the leading patterns of the joint instrumental and proxy covariance matrix. The T-TLS regularization parameter, which gives the number of eigenvectors retained in the estimate of the joint covariance matrix, can take on only a finite number of values and in the context of RegEM is set a priori. This is in contrast to the regularization parameter in ridge regression, which can take on any value and is chosen adaptively by RegEM.

d. Uncertainty estimation in RegEM

Estimation of the uncertainty in the values imputed by RegEM, using either ridge or T-TLS regularization, is nontrivial. If the covariance matrix and mean of the joint data matrix are known, and no regularization is used, then the estimated uncertainty in the imputed values follows directly from Eq. (2). However, both the mean and covariance structure are estimated from the data, and at least one regularization parameter that modifies the estimate of the covariance matrix is either specified (T-TLS) or estimated from the data (ridge). The RegEM uncertainty estimate takes the form of the regularized sample estimate of the conditional variance [i.e., the variance form in Eq. (2) estimated using RegEM], scaled to account for the loss of degrees of freedom due to the estimation of the regularization parameters as well as the uncertainty in these parameters. The resulting estimates of the uncertainties in the imputed values are lower bounds, and are generally too small (Schneider 2001). To correct for this bias, Schneider (2001) suggests inflating the regularized estimate of the conditional covariance matrix by some additional factor that is determined via numerical simulations.

3. Comparing BARCAST and RegEM: Assumptions and methodology

The RegEM approaches, which are generalizations of the EM algorithm, assume that the dataset is composed of a series of independent, identically distributed draws, some of which are incomplete, from a multivariate normal distribution. BARCAST likewise assumes that the data vector for each year is a (possibly incomplete) draw from a multivariate normal distribution, but makes a number of additional assumptions about the temporal and spatial covariance structure of the underlying field. We now turn to a point-by-point comparison of the assumptions, methodologies, and end products of these two approaches. Unless otherwise specified, RegEM will refer in this section to the family of reconstruction techniques that includes the EM algorithm and RegEM regularized using either ridge regression or T-TLS.

a. Treatment of missing data

A key assumption made by both BARCAST and RegEM is that the distribution of the missing observations is ignorable (Rubin 1976; Gelman et al. 2003). An example of a dataset with a nonignorable missing data structure would be an ice core that features missing values

related to surface melting events during particularly warm years. More generally, the amount of available climate data has increased over the last 150 years, as have both greenhouse gas concentrations and temperatures. In addition, proxy records such as tree rings are only available in geographical regions with particular climates amenable to the development of the proxy—there are no tree ring records from Greenland, for example. These facts suggest that the assumption of ignorability is likely incorrect. While not explored here, the influences of the missing data structure on climate reconstructions is a topic that warrants, and is beginning to receive, further investigation (e.g., Smerdon et al. 2008).

b. Covariance matrices

BARCAST and RegEM make use of different covariance matrices, the implications of which will be discussed in several contexts below. BARCAST estimates the parameters of a specified spatial covariance form, which can then be used to specify the covariance matrix of the underlying true field values at the locations of the data time series and any other target locations of interest. RegEM, in contrast, is based on an estimate of the joint covariance matrix of the proxy and instrumental time series.

c. Local versus global relationships

BARCAST assumes that observations reflect information about the local field values and then makes use of a parametric spatial covariance form to allow the observations at one location to influence predictions of the field value at other locations. We currently specify the spatial covariance to follow an exponential decay of correlation with separation, so that the weight of each observation in estimating the field at a particular location decreases with distance from that location. RegEM, in contrast, makes use of all linear relationships between the proxy and instrumental time series, as estimated by the sample cross-covariance matrix, and so can exploit strong covariances between distantly separated proxy and instrumental time series. While BARCAST implicitly assumes that the spatial correlation length scale of the field is constant through time, RegEM makes the same stationarity assumption with regards to the more complex patterns of covariance between the proxy and instrumental time series.

We stress here and below that, in any particular scenario, one analysis might be more appropriate than the other. In particular, if the field values at pairs of distantly located points are often more correlated than at pairs of more closely located points, then the field estimates produced by the current implementation of BARCAST

will suffer, while those from RegEM will not. In such a scenario, the simple spatial structure currently assumed by BARCAST prohibits the algorithm from exploiting covariance structures between distant points. That said, the examples presented in Part I show that, using reasonable pseudoproxy data, BARCAST produces reconstructions of North American surface temperatures that are demonstrably superior to those produced by RegEM.

d. Regularization and prior covariance information

BARCAST parameterizes the structure of the spatial covariance matrix of the field with two unknowns: the covariance at zero separation and an inverse length scale that describes the exponential decay of covariance as a function of separation. Specification of a parametric form for the spatial covariance matrix of the field regularizes the analysis by reducing the total number of parameters that must be estimated from the data and by ensuring that the estimated covariance matrix is not singular. These assumptions can be thought of as placing a prior on the structure of the covariance of the field, and section 4 will explore the performance of BARCAST when this prior is clearly incorrect. The physically based assumption that climate fields display covariance that decays as a function of separation, while not likely perfect in any given situation, is likely adequate in many (see Fig. 2 of Part I). As discussed in Part I, more complicated spatial relationships could be incorporated into BARCAST by modifying the parametric from of the covariance matrix.

BARCAST also involves prior information in the form of the prior distributions for the scalar parameters and the field values for the first year of the reconstruction. In realistic applications we find that these priors are sufficiently diffuse as to have no noticeable influence on the posterior distributions (see Fig. 8 of Part I and Tingley 2009).

RegEM, in contrast, is based on empirical estimates of the joint proxy–instrumental covariance matrix and so involves estimating the covariance between each pair of data time series. While RegEM exploits all linear relationships between the proxy and instrumental time series, there is often insufficient data to adequately constrain the covariance matrix. The techniques used to regularize the regression, both T-TLS and ridge regression, can be interpreted in terms of prior constraints. T-TLS limits the number of distinct patterns in the joint data covariance matrix used in the analysis and so can be interpreted as constraining a priori the complexity of the data structure. Similarly, ridge regression can be interpreted as down weighting the contributions of the eigenvectors of the proxy covariance matrix associated

with small eigenvalues, so a priori emphasizes a smaller number of patterns in the data structure.

In the case of a univariate response, the ridge estimate of the coefficient vector, β_h^* [cf. Eq. (4)], has a simple Bayesian interpretation. Given the regression model $\mathbf{X}_m = \mathbf{X}_o\beta + \epsilon$, with $\epsilon \sim N(0, \sigma^2\mathbf{I})$, then β_h^* is equivalent to the posterior mean which results from placing a $N(0, \sigma^2(h^2\mathbf{D})^{-1})$ prior on β (Hoerl and Kennard 1970). That is, ridge regression implicitly assumes a prior that generally reduces the magnitudes of the estimated regression coefficients. While BARCAST and the RegEM techniques differ with regards to the covariance matrix that underpins the analysis—BARCAST makes use of a spatial covariance matrix whereas RegEM considers the sample covariance matrix of the data time series—each makes use of what can be interpreted as prior information to ensure that a required matrix inverse exists.

Prior to analysis with RegEM, it is sometimes useful to reduce the number of proxy time series using principal component analysis (PCA) or other techniques. This has been done in practice, for example, when dealing with large numbers of nearby tree ring records (e.g., Rutherford et al. 2005). The resulting reduced dataset requires less regularization, as a large number of highly correlated time series are replaced by a much smaller number of weighting time series, each associated with a dominant mode of variability of the network. BARCAST, in contrast, is designed to impute spatially and temporally complete fields from spatially incomplete instrumental and proxy observations. There is no need for data reduction with BARCAST, which makes explicit use of the location of each time series and a parametric spatial covariance form that anticipates that nearby observation time series will be highly correlated. A cluster of observations will result in the field estimates in that region having low uncertainty, but these observations will only affect estimates of the field at other locations according to the assumed exponential decay of spatial covariance.

To insure adequate regularization, RegEM has also been applied after using PCA to reduce the instrumental dataset to a smaller number of loading time series, each corresponding to a dominant mode of spatial variability (e.g., Mann et al. 2007b; Steig et al. 2009). As the instrumental time series are generally of different lengths, it can be necessary to impute the missing instrumental values prior to performing the PCA, and this has been done using an additional application of RegEM (e.g., Rutherford et al. 2005). As BARCAST makes explicit use of the spatial location of each instrumental time series, it is not an appropriate tool for reconstructing loading time series resulting from a PCA, nor is there a need to use PCA when data are closely spaced.

e. Temporal autocorrelation

Most climate time series feature nonzero temporal autocorrelation, and BARCAST includes this information in the analysis by specifying that the field evolves according to a first-order multivariate autoregressive process. While this assumption is not likely to be exact in many applications, climate time series do tend to have red spectra (e.g., Hegerl et al. 2007), suggesting it is a better assumption than zero autocorrelation.

RegEM, in contrast, assumes that the observations at subsequent years are independent, which has at least two important ramifications if the system does in fact have nonzero temporal autocorrelation. First, RegEM does not exploit the information available form observations at neighboring years in the prediction of the field for each year. Second, the estimated uncertainties in the sample mean vector and covariance matrix will be biased toward low values, as temporal autocorrelation reduces the degrees of freedom available for estimating these quantities. In practice, temporal dependencies have been incorporated into RegEM by considering lags in the relationship between the proxy and instrumental observations. For example, Rutherford et al. (2005) use proxy observations at times $t-1$, t, and $t+1$ to infer the instrumental observations at time t, but find that the additional predictors do not increase the skill of the reconstructions. In contrast, Schneider (2001) suggests incorporating temporal autocorrelation into RegEM by augmenting the vector $[\mathbf{X}_o, \mathbf{X}_m]$ for each year t to include observations from years $t-1$ and $t+1$ as well, which has not yet been done in practical applications.

f. End products

The end products of a RegEM analysis are the completed-by-imputation data matrix, estimates of the uncertainty in the imputed values, and estimates of the mean vector and covariance matrix. The end product from BARCAST is an ensemble of draws of the space–time field and scalar parameters, each of which is consistent with the data and model assumptions. This ensemble can be used to estimate the full posterior distributions of any number of quantities, from simple measures like the temporal evolution of the field at each location to more exotic quantities like, for example, the probability that the mean (spatially and temporally) of the surface temperature field was more extreme over the most recent decade than over any other decade covered by the reconstruction (see Tingley 2009).

g. Target quantities and locations

RegEM, as currently implemented, imputes missing instrumental observations and so can estimate the field

only at spatial locations where instrumental observations are available during the calibration period. BARCAST seeks to impute the underlying true field values and can do so at any set of spatial locations. This results from the differing assumptions regarding the spatial covariance structure. As BARCAST assumes a parametric from, and then uses the data to estimate the parameters, the conditional distribution of the field at any set of locations, given the observations, is readily specified (see Part I). RegEM is based on the sample covariances between each pair of data time series and so does not predict the field at locations without observations (see Fig. 5 of Part I).

The missing instrumental values imputed by RegEM at each year could be interpolated spatially via kriging (e.g., Banerjee et al. 2004) to give a complete field. The ability of RegEM to exploit strong correlations between time series at distantly separated locations is often cited as an advantage of the method (e.g., Jones et al. 2009), whereas simple kriging techniques tend to be based on stationary, and usually isotropic, parametric spatial covariance forms that describe a decay of correlation with increased separation (e.g., Banerjee et al. 2004). Using a simple kriging procedure on the output from RegEM to produce a spatially complete field estimate thus involves two different views of the spatial covariance: the first an empirical data estimate and the second based on a simpler parametric form. Apart from these issues, such a multistep analysis complicates the propagation of uncertainty estimates.

h. Error estimation for the imputed values

Uncertainties for the field estimates produced by BARCAST are estimated by calculating the 5th and 95th (or any other) percentiles of the ensemble of posterior draws and thus account for the uncertainty in all other parameters of the model. Those from RegEM, in contrast, do not account for the uncertainty in the estimation of the covariance matrix or the regularization parameters. As a result, the basic uncertainty estimates from RegEM tend to be too small, so a variance inflation factor must be estimated to ensure that confidence intervals have the correct coverage rates (Table 3 of Part I; Schneider 2001). For the trials reported in Part I, the uncertainty estimates from BARCAST have the correct coverage rates, while those from RegEM, when the variance inflation factor is set to the default of one, do not (Table 3, Part I).

Neither BARCAST nor RegEM account for errors in the *structure* of the estimation model—results using BARCAST are conditional on the assumptions made about the covariance matrix, autoregressive temporal evolution, and observation equations. If, for example, the assumption that the proxies have a linear relationship with the true values (BARCAST) or instrumental observations (RegEM) is incorrect, then the uncertainty estimates will tend to be biased low. In addition, BARCAST, but not RegEM, makes a simplifying assumption about the spatial covariance of the field which is unlikely to hold exactly in practice, and as a result, uncertainty estimates from BARCAST could be biased low in practical applications. Section 4 explores the robustness of BARCAST to deviations from the model assumptions using simple surrogate datasets, and future work will investigate these issues in more realistic scenarios, as has been done for RegEM (Rutherford et al. 2005; Mann et al. 2007b).

Differences in error estimation also arise as a consequence of the reconstruction techniques having different target quantities. As RegEM seeks to impute the missing instrumental observations, there is no uncertainty associated with the reconstruction over the calibration interval (see Part I, Figs. 5 and 7). BARCAST, in contrast, treats the instrumental time series as noisy observations and seeks to estimate the underlying true field. Note, however, that if the estimate of the instrumental error variance is small relative to that for the proxies, this distinction between RegEM and BARCAST will likewise be small.

The ensemble of posterior draws produced by BARCAST can be used to estimate the uncertainty in the imputed instrumental values by adding to each ensemble member white noise draws with variance given by the corresponding draw of the instrumental observational error, and then taking the percentiles of the resulting distributions. RegEM, in contrast, does not infer observational errors and so can only estimate the uncertainty associated with imputations of the missing instrumental values, which, in the presence of uncertainty in the instrumental observations, will be larger than those for the missing field values.

i. Estimating functions of the field and the associated uncertainty

It is often of interest to estimate the time evolution of the spatial mean of a climate quantity over a particular region and the associated uncertainty. BARCAST produces both these quantities by specifying as the target locations a number of evenly distributed points in the region and then taking, for each posterior draw of the field, the mean across these target locations. The percentiles of the resulting distribution can be used to produce both an estimate of the time evolution of the spatial mean and an estimate of the associated uncertainty. RegEM can likewise estimate the spatial mean and associated uncertainty at each year as a linear function of the imputed instrumental values, where the weights given to each imputed

observations can be set to take into account the heterogeneous distribution of instrumental observations.

The ensemble of draws of the space–time field produced by BARCAST can be used to calculate the probability distribution of any function of this field by applying the function to each ensemble member and then estimating percentiles. One function that is frequently plotted (e.g., NRC 2006) is the estimate of the spatial mean, smoothed through time (see Part I, Fig. 7). RegEM allows for a point estimate of any function of the space–time field, with the caveat that the target locations are limited by the spatial distribution of the instrumental time series. There is, however, no general way of estimating the uncertainty in functions of the imputed values. As an example, it is straightforward to smooth (through time) the estimate of the spatial mean, which can be plotted together with the smoothed uncertainty envelope for the spatial mean, but this smoothed uncertainty envelope is biased to be wider than the uncertainty in the smoothed quantity (see Part I, Fig. 7).

j. Regression dilution and the temporal variance of the reconstruction

Climate reconstruction approaches based on ordinary least squares regression, which minimize the error sum of squares, result in reconstructions of a field or spatial mean biased toward having lower temporal variance over the interval when only proxy observations are available than over the calibration interval (NRC 2006). Ordinary least squares regression assumes that the predictor variables, which are generally the proxy observations in the paleoclimate reconstruction problem, are error free. If this assumption holds, then the paleoclimate problem reduces to inferring the field at the target locations given error-free information at the proxy locations, which are generally sparse and heterogeneously distributed. Even in this idealized circumstance, not all of the variability of the climate field at the target locations will be captured by the information provided at the proxy locations. As a result, the predictions from an ordinary least squares regression of the field values at the target location will in general be less variable than the true values.

Furthermore, the assumption that the proxy observations are error free is clearly wrong, and errors in the predictor values result in ordinary least squares estimates of the regression coefficients that are biased toward zero—this is the so-called *regression dilution* problem (e.g., Frost and Thompson 2000). There are two issues that warrant discussion with regards to regression dilution: estimation of the regression coefficients and inference on the response variables using predictors that contain errors—the latter being the main goal of climate reconstructions.

To explore the first issue, consider the simple linear regression problem $\mathbf{X}_m = \mathbf{X}_o\beta + \epsilon$, where β is a scalar and for the sake of simplicity we assume that the vectors $\mathbf{X}_m$ and $\mathbf{X}_o$ both have means of zero. The ordinary least squares estimate of the regression coefficient can be written as $\hat{\beta}_{X_o} = \mathrm{Cov}[\mathbf{X}_m,\mathbf{X}_o]/\mathrm{Var}[\mathbf{X}_o]$. Assuming instead that we observe $\mathbf{Z} \equiv \mathbf{X}_o + \eta$, where the elements of η are independent and identically distributed (iid) draws from a normal distribution with a mean of zero and variance δ^2, the estimate of the regression coefficient linking the elements $\mathbf{X}_m$ to the elements of $\mathbf{Z}$ is then $\hat{\beta}_Z = \mathrm{Cov}[\mathbf{X}_m,\mathbf{Z}]/\mathrm{Var}[\mathbf{Z}] = \mathrm{Cov}[\mathbf{X}_m,\mathbf{X}_o]/(\delta^2 + \mathrm{Var}[\mathbf{X}_o])$. If the goal of the analysis is to estimate the relationship between $\mathbf{X}_o$ and $\mathbf{X}_m$ based on the observations of $\mathbf{Z}$ and $\mathbf{X}_m$, then the ordinary least squares approach results in an estimate of the coefficient $\hat{\beta}$ that is biased toward zero.

If there are multiple measurements of the predictor variables for each measurement of the response variable, a number of techniques exist to correct for regression dilution based on inferring the variance of the errors in the predictor values (Frost and Thompson 2000). If this is not the case, the regression dilution effect can be mitigated by using more robust alternatives to ordinary least squares regression. Total least squares regression, which results from setting $\nu = M_m$ in Eq. (8), provides unbiased estimates of the regression coefficients when the predictor and response vectors each contain iid errors with the same variances (Fierro et al. 1997; Golub and Van Loan 1980). A simple modification can produce unbiased estimates if the errors in the two variables have different variances, but with a known ratio. These regression techniques can thus correct for the bias in the estimate of the slope, but they require an assumption about the relative errors in the response and predictor variables.

BARCAST, in contrast, accounts for errors in the predictor variables by explicitly modeling both the proxy and instrumental observations as containing errors. The assumptions made by BARCAST about the spatial structure of the field allow the algorithm to infer both the relationship between the field at the observation and target locations and the errors associated with the proxy and instrumental observations.

To compare the reconstruction methods with regards to this issue, we recast aspects of the BARCAST formalism developed in Part I to mimic the RegEM approach, assuming for the sake of simplicity that the autoregressive coefficient α is zero. The goal is to predict a set of instrumental observations $\mathbf{W}_I$ given a number of proxy observations $\mathbf{W}_P$ (i.e., $\mathbf{W}_{P,I}$ correspond to $\mathbf{X}_{o,m}$ from section 2). The joint distribution of the two types of observations is multivariate normal, and the mean and covariance follow from taking the expectation and

variance of Eq. (6) of Part I with respect to the true field values:

$$\begin{pmatrix} \mathbf{W}_I \\ \mathbf{W}_P \end{pmatrix} \sim N\left[\begin{pmatrix} (\beta_1 \mu + \beta_0)\mathbf{1} \\ \mu\mathbf{1} \end{pmatrix}, \begin{pmatrix} \mathbf{\Sigma}_{I,I}^s + \tau_I^2\mathbf{I} & \mathbf{\Sigma}_{I,P}^s \\ \mathbf{\Sigma}_{P,I}^s & \mathbf{\Sigma}_{P,P}^s + \tau_P^2\mathbf{I} \end{pmatrix} \right].$$

(9)

Here τ_I^2 and τ_P^2 are the error variances for the proxy and instrumental observations, $\mathbf{\Sigma}_{I,P}^s$ is the cross spatial covariance of the true field at the locations corresponding to the observations $\mathbf{W}_I$ and $\mathbf{W}_P$, and similarly for the other $\mathbf{\Sigma}^s$. A superscript s is used here to distinguish the spatial covariance matrix used by BARCAST from the covariance of the joint proxy and instrumental dataset used by RegEM. The conditional distribution of the instrumental observations, given the proxy observations, is likewise normal:

$$\mathbf{W}_I | \mathbf{W}_P \sim N[\mu\mathbf{1} + \mathbf{\Sigma}_{I,P}^s (\mathbf{\Sigma}_{P,P}^s + \tau_P^2\mathbf{I})^{-1} (W_P - (\beta_1\mu - \beta_0)\mathbf{1}),$$
$$(\mathbf{\Sigma}_{I,I}^s + \tau_I^2\mathbf{I}) - \mathbf{\Sigma}_{I,P}^s (\mathbf{\Sigma}_{P,P}^s + \tau_P^2\mathbf{I})^{-1} \mathbf{\Sigma}_{P,I}^s],$$

(10)

where the term $\mathbf{\Sigma}_{I,P}^s (\mathbf{\Sigma}_{P,P}^s + \tau_P^2\mathbf{I})^{-1}$ is a matrix of regression coefficients. As BARCAST estimates the spatial covariance matrix $\mathbf{\Sigma}^s$ separately from the observational error variances τ_I^2 and τ_P^2, it is possible to disentangle the uncertainty introduced by the observational errors from the uncertainty introduced by the predictors not being collocated with the response variables. RegEM, in contrast, estimates the joint covariance of the proxy and instrumental time series, which includes these observational variances.

Setting τ_I^2 to zero in Eq. (10) gives the conditional distribution of instrumental observations with zero observational error, which are equivalent to the underlying field values, given the proxy observations. Note that τ_I^2 appears in the expression for the conditional variance but not the conditional mean—the presence of instrumental observational error does not change the estimate of the mean value of the unknown quantity, but it does change the associated uncertainty estimate. If τ_P^2 is also set to zero, then Eq. (10) gives the conditional distribution of a number of true field values given (a linear transformation of error free) observations of other field values. If the error-free elements of $\mathbf{W}_I$ and $\mathbf{W}_P$ are collocated, the regression is then ill posed as the joint covariance matrix is singular. In this context, the presence of observational error plays a similar role to the ridge parameter in regularizing the regression and reducing the estimates of the regression coefficients.

The second issue with respect to regression dilution concerns the estimation of the response variables given noisy estimates of the predictors. In this instance, the optimal solution from the perspective of minimizing expected MSE of prediction is to use estimates of the regression coefficients that do not correct for regression dilution (Frost and Thompson 2000). While BARCAST disentangles the spatial effect (predictors and response variables not collocated) from the uncertainty introduced by observational errors, the predictions from the noisy proxy observations are made using regression coefficients reduced by the presence of errors in the predictors [Eq. (10)]. The variance of a reconstruction from either BARCAST or RegEM will be biased low, relative to the variance of the true values, by the presence of errors in the proxy observations. This is not necessarily a flaw, and the ideal balance between producing a reconstruction with unbiased temporal variance and one that minimizes the expected MSE of predictions will likely depend upon the goals of the analysis.

4. Comparing BARCAST and RegEM: Numerical experiments

To facilitate the comparison between RegEM and BARCAST we make use of trials conducted on variations of a simple surrogate dataset specifically constructed in accordance or discordance with the assumptions made by BARCAST. Initially, nine true field time series, specified as being located at unit-spaced nodes on a line, are generated according to a multivariate first-order autoregressive process driven by innovations with covariance that decays exponentially as a function of separation. See Table 1 for parameter values and Table 1 from Part I for a description of the notation.

Surrogate "instrumental" time series with a signal-to-noise ratio (SNR) of three are produced from the last half of each true value time series by adding iid draws from a mean-zero normal distribution to the true values. Surrogate "proxy" time series are produced from the full length of every second surrogate time series by adding iid draws from a mean-zero normal distribution to the true values. These proxy time series are then standardized by removing the common mean and dividing by the common standard deviation of all proxy observations (cf. the algorithm presented in Osborn and Briffa 2006). The second half of each dataset acts as a calibration period, while the first half of each dataset is used to test the reconstructions. We will vary the length of the dataset and the proxy SNR, which is calculated in terms of standard deviations, while keeping the number of locations fixed at nine. Each set of experiments discussed below is based on applying the reconstruction

TABLE 1. Posterior percentiles of the eight scalar parameters estimated by BARCAST, for a typical analysis of a surrogate dataset of length 100 and a proxy SNR of 1. See Table 1 of Part I for definitions of the parameters. The first set of percentiles refers to the analysis performed with the time series tagged with the correct locations, and the second set, denoted with (L), refers to the analysis performed after switching the location tags. The parameter values used to construct the dataset are listed in the second column. The "true" values of the last three parameters, τ_P^2, β_0 and β_1, are in bold, as these are not specified in the data construction—the proxy records are simply standardized after the addition of noise [see Part I, Eq. (9)].

Parameter	Truth	Percentiles			Percentiles (L)		
		0.05	0.50	0.95	0.05	0.50	0.95
α	0.50	0.34	0.49	0.57	0.49	0.57	0.65
μ	0	-0.22	0.06	0.33	-0.15	0.08	0.30
σ^2	1	0.81	1.02	1.29	0.77	0.94	1.15
ϕ	0.25	0.20	0.28	0.36	0.65	0.93	1.33
τ_I^2	0.1481	0.09	0.14	0.21	0.13	0.22	0.34
τ_P^2	**0.5**	0.57	0.65	0.74	0.54	0.63	0.72
β_1	**0.6124**	0.43	0.50	0.58	0.45	0.53	0.62
β_0	**0**	-0.12	-0.03	0.05	-0.13	-0.04	0.05

techniques to 100 surrogate datasets for each length or proxy SNR.

BARCAST makes explicit use of the spatial separation between time series in calculating the covariance matrix, so the results are dependent on the spatial locations of the data time series. To explore the sensitivity of BARCAST to the correctness of the assumptions made about the spatial covariance, we apply BARCAST to both the surrogate datasets with the time series tagged with the "correct" locations (those used to construct the data), and with "incorrect" locations, formed by switching the locations assigned to the second and fourth time series with those assigned to the eighth and sixth time series. The spatial covariance of the resulting dataset strongly violates the structure assumed by BARCAST.

RegEM requires the specification of a variance inflation factor, and either ridge parameter(s) or a truncation parameter, for regularization via ridge regression or T-TLS, respectively (Schneider 2001). In the experiments below, the variance inflation parameters are set to give reasonable results in terms of the coverage rates of the resulting confidence intervals, the generalized cross validation presented in Schneider (2001) is used to choose values of ridge parameters, and we explore several choices for the truncation parameter required by T-TLS. The objective selection of these parameters remains something of an open question and will not be addressed in this study; see Mann et al. (2007b) for further discussion.

We will distinguish below between a number of different reconstruction methods, summarized here for

convenience. Ridge-I refers to RegEM regularized with a separate regression and ridge parameter estimation for each missing value, while Ridge-M refers to RegEM regularized with a multiple regression and single ridge parameter estimation for each year with missing values. T-TLS(k) refers to RegEM regularized with truncated total least squares regression, retaining k eigenvectors of the joint covariance matrix, while EM refers to application of the standard EM algorithm- that is, RegEM applied without regularization. BARCAST refers to the application of the Bayesian algorithm to the data tagged with the correct locations while BARCAST(L) indicates that the data time series are tagged with the incorrect locations. The results from BARCAST and BARCAST(L) are tested against the same set of true values: the only difference between the two cases is that in the first the location information corresponding to each time series is the same as was used to generate the data, while in the second the location information is switched between the generation and analysis of the data.

a. Coefficient of efficiency as a function of dataset length

It is of interest to understand how the reconstructive skill of the various methods depends on the length of the dataset, which determines the amount of information available for inferring the relationships between the data types (RegEM) or the relationships between the data types and the true field (BARCAST). Here and below, "field" refers to the true values of the nine time series generated in each experiment, and we will limit the reconstructions to inferences on the missing field or instrumental values at the locations of these surrogate time series. We assess the skill of the reconstructions of both the field and the spatial mean using the coefficient of efficiency (CE) statistic (Cook et al. 1994; Rutherford et al. 2005):

$$\text{and} \quad \text{CE} = 1 - \frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2}. \quad (11)$$

If each estimate, $\hat{y}_i$, is set to the true mean, $\bar{y}_i$, of that variable over the testing interval, then the CE is zero. A positive value indicates that the reconstruction contains information about the variation of the true values about the mean.

We consider surrogate datasets ranging in length from 4 to 100, with a proxy SNR of 1, and report the 25% trimmed mean of the CE values that result from producing and analyzing 100 realizations of the surrogate

dataset for each length (Fig. 1). As the CE values are occasionally very large and negative for each method, we use a trimmed mean (formed in this case by discarding the largest and smallest 25% of the values) as a robust measure of the center (e.g., Devore 1995). The trimming percentage does not affect the main results reported below, but a lower trimming percentage requires a greater number of realizations for these results to become apparent.

For most dataset lengths, BARCAST and BARCAST(L) result in the highest trimmed mean CE values for both the reconstruction of the mean and the field (Fig. 1). Results for BARCAST(L) are comparable to those for BARCAST for the mean reconstructions but are uniformly lower for the field reconstructions. The inferences made by BARCAST and BARCAST(L) about the scalar parameters lend insight into the influences of the covariance misspecification on the results (Table 1). BARCAST(L) infers a much larger value for the inverse spatial range parameter ϕ, which is to be expected as switching the location tags results in the spatial covariance having a shorter length scale. BARCAST(L) also results in larger estimates of the instrumental observational error variance, τ_I^2. As the dataset does not follow the assumptions about the spatial covariance, the analysis infers larger errors in the data time series to bring the estimates of the field in line with the assumptions.

The CE values for BARCAST and BARCAST(L) plateau once the series length reaches about 50. With this much data, the algorithm can accurately identify the scalar parameters of the analysis model so that the estimated uncertainty in each reconstruction is primarily the result of the proxy data being both noisy and incomplete. The CE values for T-TLS(1) likewise plateau once the series length reaches about 65, with the values being uniformly lower than those for BARCAST(L); results are similar for T-TLS(2). The plateau indicates that the retained eigenmode(s) are well estimated by the data, but do not capture the full covariance structure of the surrogate instrumental and proxy observations. On the other hand, retaining more eigenmodes than can be well constrained by the data leads to less skillful reconstructions—for this reason, the field CE values for T-TLS(2) are smaller than those for T-TLS(1) for short series lengths (not shown).

For both the mean and the field, the Ridge-M reconstructions are initially more skillful—then less skillful—than the T-TLS reconstructions, and for series lengths greater than about 65 (95) for the mean (field), they are more skillful. Results are similar for Ridge-I. For series lengths longer than about 90, the Ridge-M field reconstructions are more skillful than those from BARCAST(L). Both versions of RegEM result in higher CE values than EM for datasets shorter than about 15 and 25, for the mean and field, respectively, indicating the utility of the regularization.

With increasing series length, the regularization aspect of RegEM becomes less necessary, as there is sufficient data available to accurately estimate the covariance structure of the datasets. The CE values for the field become higher for EM than for BARCAST(L) at about series length 60, and by series length 100 the CE values for the EM mean reconstruction are comparable to those for BARCAST and BARCAST(L). As the series length increases past about 50, the CE values for Ridge-M begin to converge with those for EM. The ridge parameters(s) are picked adaptively by RegEM (see above) and as the amount of data increases, RegEM selects smaller and smaller regularization parameter(s), resulting in reconstructions that are progressively closer to those from EM.

Whereas the CE values for the BARCAST and T-TLS reconstructions plateau before series length 100, the CE values for the Ridge-M and EM reconstructions are clearly still rising at series length 100. With sufficient data, the EM approach and BARCAST should have similar reconstructive skill, provided the datasets meet the assumptions made by BARCAST about the spatial covariance of the field. If the assumptions made by BARCAST are incorrect, then with sufficient data the EM algorithm should be more skillful, as can be seen at series length 100 for the BARCAST(L) field reconstruction (Fig. 1). Applying the EM algorithm to surrogate datasets of length 500 results in 25% trimmed mean CE values of 0.70 for the mean and 0.52 for the field, which are comparable to the values from BARCAST at series length 100.

When applying these different methods to actual data, a key constraint is the length of the overlap between the proxy and instrumental datasets compared to the number of time series (in the case of RegEM) or spatial locations (in the case of BARCAST) involved in the reconstruction. As a reference, the North American example presented in Part I involved 163 spatial locations, 102 instrumental time series, 20 proxy time series, and a 67 year overlap between the two types of data. In a real reconstruction scenario for this area, the entire length of the instrumental record would be used for calibration, so the overlap between the two data types would increase to about 157 years—about 1.2 times the number of time series. While the number of time series, both instrumental and proxy, increases as the spatial domain of the reconstruction increases, the length of the overlap tends to be capped at around 150 years by the length of most instrumental records, so the ratio will decrease as the domain expands.
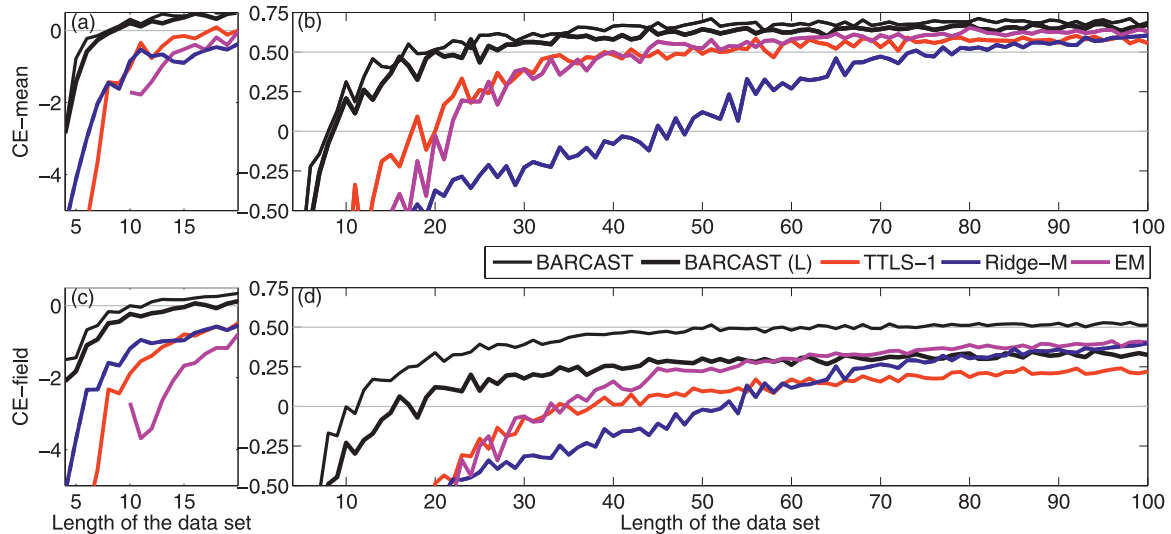
FIG. 1. CE as a function of the length of the surrogate dataset, for a number of different reconstruction approaches. BARCAST and BARCAST(L) refer to analysis with BARCAST with the datasets tagged with the correct and incorrect locations, respectively; T-TLS(1) refers to truncated total least squares–regularized RegEM making use of one eigenmode of the joint covariance matrix; Ridge-M refers to ridge-regularized RegEM with one regularization parameter calculated for each year for which there are missing observations; EM refers to the unregularized EM algorithm. Results using T-TLS(2) are similar to those shown for T-TLS(1), and results for Ridge-I are similar to those shown for Ridge-M. (a),(b) The CE for the reconstructions of the spatial mean; (c),(d) the CE for the reconstructions of the field. Panels (a) and (c) show the first parts of panels (b) and (d), respectively, but with the y axes expanded to show how the algorithms differ when applied to short datasets.

In many real applications, then, the number of years for which both proxy and instrumental observations are available will be at best slightly larger, and often smaller, than the number of data time series. The relevant test of the various algorithms is thus their performance when the overlap between the proxy and instrumental observations is similar to or smaller than the number of locations involved in the reconstruction. The surrogate datasets are composed of nine instrumental time series and five proxy time series, while the overlap between the two is half the length of the dataset. As such, dataset lengths less than about 30 are comparable with actual climate reconstruction problems. The simple examples analyzed here suggest that in this range of dataset lengths, both BARCAST and BARCAST(L) produce higher CE values than RegEM (Fig. 1).

It is of course possible to set up a test where any particular method results in the highest CE values. Violations of the assumptions made by BARCAST about the spatial covariance structure are detrimental to the reconstructive skill, and given a dataset with a covariance structure sufficiently different from that assumed by BARCAST, the RegEM approaches will result in the most skillful reconstructions. In the context of climate field reconstructions, the validity of the assumption that covariance decays exponentially with distance will depend on the particulars of the field and spatial domain

under consideration. Nonetheless, the BARCAST(L) experiments indicate that BARCAST is robust to considerable deviation from the assumptions made about the covariance, at least in terms of the reconstruction of the mean.

b. Temporal standard deviation of the reconstructed mean

To investigate the influence of the proxy observational errors on the estimated temporal standard deviation of a reconstruction, we apply the various approaches to surrogate datasets of length 80, with values of the proxy SNR ranging from 0.25 to 3 (Fig. 2), where the SNR is measured in terms of standard deviations. Datasets of length 80 involve 40 years of overlap between the two data types, so according to the discussion in section 4a the ratio of years of overlap to numbers of locations is somewhat larger than for many practical applications. The main conclusions of this section are robust to the length of the dataset, which is set to 80 to provide a sufficiently long testing interval from which to estimate temporal standard deviations and a calibration interval that is sufficiently long to ensure the parameters of each analysis scheme are well estimated. This allows us to isolate the effects of varying the proxy SNR and to explore via simulations the issues discussed in section 3j with regards to regression dilution in the presence of
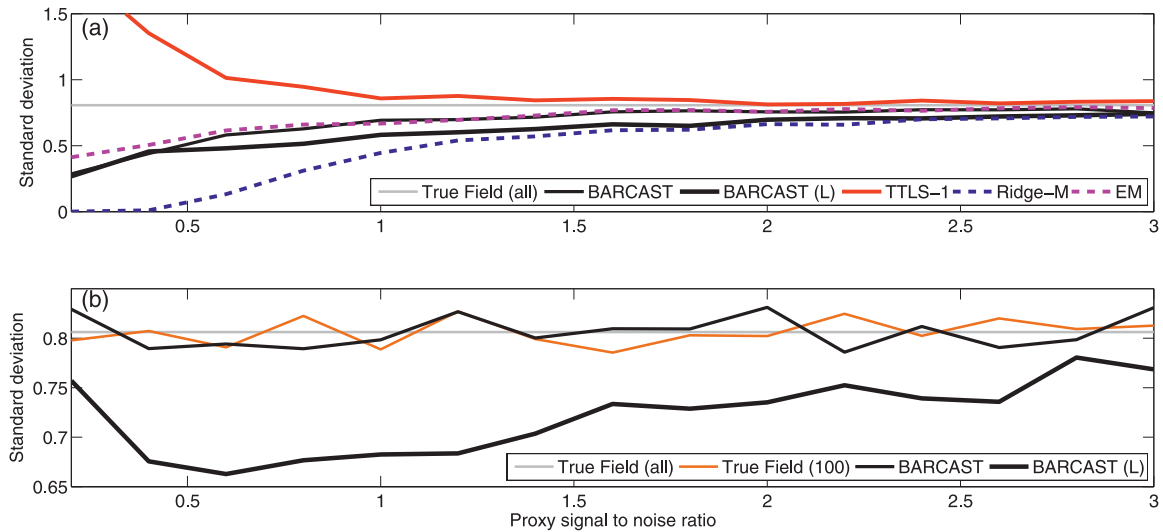
FIG. 2. Temporal standard deviation of the reconstructed mean as a function of the proxy SNR, for a number of different reconstruction approaches. (a) Temporal standard deviation of the reconstructed mean, calculated from 30 time steps during which only proxy observations are available. True field (all) refers to the expected value of the standard deviation of a length 30 realization of the underlying true mean time series. Results using T-TLS(2) are similar to those shown for T-TLS1, and results for Ridge-I are similar to those shown for Ridge-M. (b) The mean of the standard deviations of each of the ensemble members produced from applying BARCAST to each of 100 surrogate datasets at each SNR. True field (100) refers to the mean from estimating the standard deviation of different sets of 100 realizations, each of length 30, of the underlying true time series.

errors in the predictor variables. Note that, in most practical applications, the SNR will be below one (e.g., von Storch et al. 2009; Lee et al. 2008; Mann et al. 2007b); we include higher values to demonstrate how differences in the results from the various methods become less pronounced as the SNR increases.

For each proxy SNR we generate and analyze 100 surrogate datasets and calculate the temporal standard deviation of the reconstructed mean between time points 6 and 35 (recall that no instrumental observations are available before the 41st time point). As BARCAST models the temporal autocorrelation, values near the beginning of the proxy or instrumental observations are not used, as the estimates at these time points are affected by nearby changes in data availability. We then take the mean of the 100 standard deviation estimates at each SNR (Fig. 2).

BARCAST and EM result in similar standard deviations, which, for small values of the SNR, are biased low relative to the expected standard deviation of the mean of the true values. The standard deviations approach the true values as the SNR increases, indicating that, in the absence of proxy observational error, the uncertainty in the mean across the nine target locations is small. The standard deviation using BARCAST(L) is smaller than that for BARCAST, indicating that the model misspecification weakens the relationship between the proxy observations and the mean across the

target locations. T-TLS(1) results in larger standard deviations estimates, which is to be expected as T-TLS involves a correction for regression bias and therefore produces larger estimates of the coefficients and predictions with higher variance. T-TLS drastically overestimates the standard deviation for small SNR values, indicating that, in this range of the SNR, even the first eigenvector is not well estimated by the data. Finally, Ridge-M results in standard deviations lower or on par with those from BARCAST(L). The low standard deviations produced by Ridge-M are expected, as the ridge parameter has a similar impact on the estimate of the regression coefficients as do errors in the proxy observations.

BARCAST, unlike RegEM, results in an ensemble of draws of the space–time field (and the scalar parameters) consistent with the model assumptions and data. Provided the model is well specified and there is sufficient data to constrain the scalar parameters, each of these draws should have, on average, the same temporal variance as the true values, regardless of the SNR. To demonstrate this feature, we estimate the standard deviation of the mean time series for each of a large number of ensemble members for each realization of the surrogate dataset, and then take the mean of these estimated standard deviations at each SNR (Fig. 2). As a comparison, we produce different sets of 100 realizations of the surrogate dataset, each of length 30, calculate the spatial

mean of the true field values for each, and then calculate the standard deviation of these time series. The mean across these standard deviations for each set of 100 realizations (Fig. 2) gives an indication of the expected variability in estimating the population temporal standard deviation from 100 realizations of the true field, each of length 30.

The temporal standard deviations of the ensemble members produced by BARCAST are essentially constant as a function of the proxy SNR, are centered on the expected value, and display a variability comparable to the estimates from realizations of the true mean time series (Fig. 2). BARCAST(L), however, results in ensemble members with standard deviations biased low relative to the true values, with the discrepancy generally larger for lower values of the SNR. In practical applications, we expect that the spatial covariance assumptions inherent to BARCAST are at best an approximation to the truth and so conclude that the ensemble members produced in real applications will likely have temporal standard deviations that are biased low, with the extent of the bias a function of the proxy SNR and the extent to which the assumed spatial covariance structure is incorrect.

c. Confidence interval widths and coverage rate

To investigate the widths and coverage rates of the 90% confidence or credible intervals estimated by the different methods, we use surrogate datasets of length 80 and vary the proxy SNR between 0.25 and 3. Ideally, 90% of the withheld values should fall within the 90% uncertainty intervals, while the extents to which the coverage rates differ from 90% are indications of biases in the estimated intervals. For each of 100 surrogate datasets at each signal SNR, we calculate the average width of the confidence interval for imputing the missing instrumental observations between time points 6 and 35 and then take the mean across these widths (Fig. 3a). The confidence intervals for RegEM and EM are the standard error estimates scaled by 2.71 (the distance between the 5th and 95th percentiles of the standard normal). The credible interval widths for BARCAST are calculated by adding to each draw of the field white noise with variance given by the corresponding draw of τ_I^2, and then taking the distance between the 5th and 95th percentiles of the resulting distributions.

As both variants of RegEM require the specification of a variance inflation factor, the widths and coverage rates of the resulting confidence intervals are somewhat arbitrary. We have chosen values of 1.1 for Ridge-M and 1.2 for T-TLS(1), as these give reasonable confidence intervals for these trials and permit us to focus on how the widths and coverage rates of the intervals vary as a function of the proxy SNR.

For all methods save T-TLS(1), the width of the estimated intervals for the missing instrumental observations decreases steadily as the SNR increases. T-TLS(1) uses a single eigenvector of the joint covariance matrix to predict the missing values, and the width of the estimated uncertainty interval for the imputed instrumental values does not vary with the SNR. BARCAST results in narrower credible intervals for the imputed instrumental values than does BARCAST(L), while EM produces confidence intervals with nearly the same widths as the credible intervals from BARCAST. The confidence intervals from Ridge-M are wider than those from BARCAST(L) for SNR values less than one, and are narrower for larger values of the SNR.

The rates at which the estimated intervals cover the missing instrumental values are relatively constant as a function of the SNR, save for those produced by T-TLS(1), which increase dramatically before reaching a plateau at an SNR of about two (Fig. 3b). Compared to the other methods, T-TLS(1) does not seem to capture important behavior of the field (cf. Fig. 1d). The coverage rates for the intervals produced by EM are much lower than 90%, which is a result of the covariance matrix not being sufficiently well estimated from a dataset of length 80. Indeed, if the experiment is repeated using a dataset of length 1000, the coverage rate of the EM intervals for the missing instrumental values is the stated 90%. The BARCAST and BARCAST(L) intervals have coverage rates very close to 90%, while the coverage rates of the wider Ridge-M intervals are generally just under 90%.

For EM and the various versions of RegEM, the rates at which the estimated intervals cover the underlying true field values are higher than those for the missing instrumental values (Fig. 3c). This is to be expected, as the instrumental values are more variable than the missing true values, and neither EM nor RegEM can adjust the intervals to account for this distinction. The credible intervals from BARCAST cover the true values nearly 90% of the time, regardless of the SNR, while the coverage rates from BARCAST(L) are biased low. That the intervals for BARCAST(L) are overly narrow is not surprising, as they do not take into account the uncertainty produced by model misspecification, in the form of incorrect assumptions about the spatial covariance. Recall that BARCAST(L) generally overestimates the instrumental observational error (Table 1), and in this particular case produces credible intervals for the missing instrumental values with the correct coverage rate, despite the model misspecification which biases the coverage rates for the true field values.
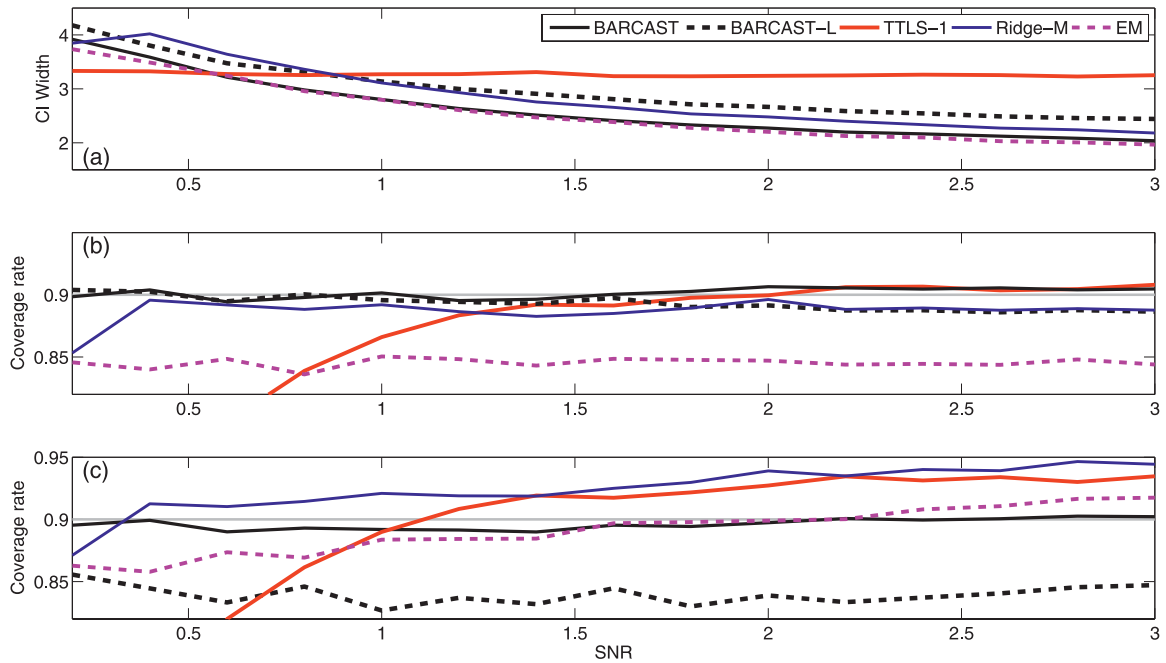
FIG. 3. Widths and coverage rates of estimated 90% uncertainty intervals as a function of the proxy SNR for a number of different reconstruction approaches. (a) Average width of the estimated 90% confidence or credible intervals associated with the estimates of the missing instrumental observations. The variance inflation was set to 1.1 for Ridge-M, and results are similar for Ridge-I. The variance inflation was set to 1.2 for T-TLS(1), and results for T-TLS(2) have a similar shape but require a different variance inflation for the coverage rates to be reasonable. (b) Percentage of the missing instrumental observations that fall within these 90% confidence or credible intervals. If the estimated intervals are accurate, they should cover the missing values 90% of the time. (c) Percentage of the unobserved true values which fall within the 90% confidence or credible intervals. Using the EM or RegEM approaches, the confidence intervals themselves are the same as for (b). The credible intervals for BARCAST in this case are simply the 5th and 95th percentiles of the ensemble of posterior draws of the field.

5. Discussion and conclusions

BARCAST, as analyzed here and presented in Part I, is a simple implementation of a general approach to the analysis of climate data—a number of possible extensions are discussed in Part I. Likewise, the various RegEM implementations analyzed here should be considered as simple implementations of a more general approach. Alternative choices for the form of the matrix **D** [Eq. (3)] when regularizing via ridge regression, for example, can be used to enforce spatial smoothing or other characteristics on the imputations of the missing values. In addition, Schneider (2001) lists a number of extensions to RegEM, including suggestions on how to incorporate temporal correlations into the model. In this work, we have focused on what we consider to be the simplest implementation of BARCAST that is applicable to paleoclimate reconstruction problems and the basic variants of the RegEM algorithm that have been used in the analysis of proxy data (e.g., Rutherford et al. 2003, 2005; Mann et al. 2007b; Steig et al. 2009; Zhang et al. 2004).

One of the more readily apparent differences between the two methods concerns their treatment of covariance matrices. Whereas BARCAST assumes a parametric form for the spatial covariance of the field, RegEM is based on an empirical estimate of the joint covariance matrix of the proxy and instrumental observations. As a result, RegEM can exploit any linear relationships between the proxy and instrumental time series to impute the missing instrumental values. In many practical applications, however, the empirical covariance estimates at the heart of RegEM will require some form of regularization, as there is generally insufficient information to adequately constrain these estimates. Both ridge regression and T-TLS have been used to provide the regularization in RegEM, and both of these techniques can be interpreted in terms of prior constraints on sample covariances matrices. In short, both BARCAST and RegEM make use of prior information in estimating the covariance matrix used in the analysis, but the prior is more readily apparent in the BARCAST formalism.

There are many other differences between RegEM and BARCAST, and we consider the inclusion of a temporal model within BARCAST to be one of the most significant, as this allows the estimates of the field for a given year to be influenced by observations from neighboring

years. Another key distinction is that BARCAST produces an ensemble of draws of the spatially complete field through time, each one of which is consistent with the data and the modeling assumptions. This ensemble allows for the investigation of novel questions and can be used, for example, to estimate the probability that the spatial mean for a given year was the warmest in the interval covered by a reconstruction.

We have explored the impacts of the differing assumptions made by RegEM and BARCAST using simple surrogate datasets and a number of measures of the performance of the various methods. If the datasets are generated according to the assumptions made by BARCAST, then for the experiments we have conducted BARCAST results in more skillful reconstructions than does RegEM, as measured by the coefficient of efficiency statistic. In addition, BARCAST produces narrower uncertainty intervals with higher coverage rates, while the ensemble members produced by BARCAST have, on average, the same temporal standard deviation as the true field values.

As it is unlikely that assumptions made by BARCAST about the spatial covariance of the field are correct in any given application, we also apply BARCAST to the surrogate datasets after corrupting the spatial information to violate those assumptions. The resulting reconstructions are inferior to those that result from applying BARCAST to the uncorrupted datasets. As measured by the coefficient of efficiency, however, these reconstructions remain superior to those from RegEM, at least for dataset lengths comparable to realistic climate reconstruction scenarios. BARCAST thus shows a certain robustness to violations of the assumptions made about the covariance of the field. Taken together with the demonstrations of RegEM and BARCAST in Part I, these results suggest that the assumptions made by BARCAST about the spatial structure of the field under analysis do not prevent the algorithm from arriving at reasonable results in realistic situations.

The testing and intercomparison of different climate field reconstruction methods is an ongoing area of research, and more work must be done to fully characterize the strengths and weaknesses of BARCAST and other approaches. A number of studies have used pseudoproxies constructed from climate model output to assess and compare the performance of climate reconstruction methods (for a review of recent work, see Jones et al. 2009), and similar testing is needed to further assess the performance of BARCAST in a wider range of situations. The description and demonstration of BARCAST in Part I and the code package posted online should facilitate the inclusion of BARCAST in future such studies, while the theoretical discussions and analyses of simple surrogate datasets presented here should facilitate the interpretation of results, and allow them to be tied back to the fundamental assumptions of the methods. Although more work is required to fully establish this new approach, BARCAST appears to be a useful tool for exploring the climate of the past.

APPENDIX

Regression as Projection and a Ridge Regression Example

The interpretation of the ridge procedure can be clarified by thinking in terms of projections. Consider the simplest regression formulation, where there is one response variable $\mathbf{Y}$ and one predictor $\mathbf{X}$, so that $\mathbf{Y} = \mathbf{X}\beta + \epsilon$, where β is a scalar. The MLE of β is $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$, and the regression estimate of $\mathbf{Y}$ is then $\hat{\mathbf{Y}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$. Note that $(\mathbf{X}^T\mathbf{X})^{1/2} \equiv |\mathbf{X}|$ is the length of $\mathbf{X}$, and the unit vector in the direction of $\mathbf{X}$ is $\mathbf{X}_u \equiv \mathbf{X}/|\mathbf{X}|$. The regression estimate $\hat{\mathbf{Y}}$ has two parts: the inner product between $\mathbf{Y}$ and $\mathbf{X}_u$ gives the length of the projection of $\mathbf{Y}$ in the direction $\mathbf{X}_u$, and this scalar is then multiplied by the unit vector in the $\mathbf{X}$ direction. In the more general context, where $\mathbf{Y}$ and/or $\mathbf{X}$ are matrices, the term $(\mathbf{X}^T\mathbf{X})^{-1/2}\mathbf{X}^T\mathbf{Y}$ is the inner product between the columns of $\mathbf{Y}$ and the unit vectors spanning the space of the columns of $\mathbf{X}$, and $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1/2}$ then multiplies these inner products by the unit vectors spanning the column space of $\mathbf{X}$.

The effects of the ridge parameter are readily understood in this geometric context. Consider the simplest case when the columns of $\mathbf{X}$ are all orthogonal, so that the matrix $\mathbf{X}^T\mathbf{X}$ is diagonal. The ridge parameter scales up each diagonal element by some small factor. This is akin to reducing the length of the unit vectors in the directions of the columns of $\mathbf{X}$, as the elements of $(\mathbf{X}^T\mathbf{X} + h^2\text{Diag}(\mathbf{X}^T\mathbf{X}))^{-1}$ are reduced. Ergo, as discussed in Hoerl and Kennard (1970), the ridge regression estimate of β is shorter than the MLE. If several of the eigenvalues of $\mathbf{X}^T\mathbf{X}$ are zero or nearly so, then the inner products between columns are nearly as large as the squared norms of the columns. Adding a small amount
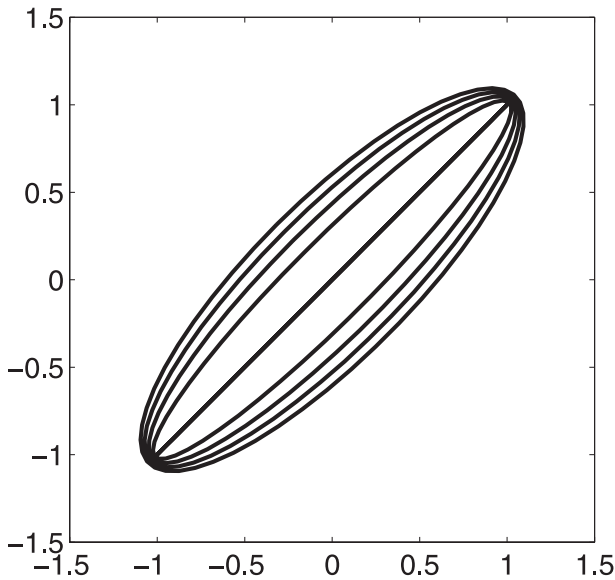
FIG. A1. Ellipses of constant one standard deviation uncertainty for values of h ranging from 0 to 0.2, for the covariance matrix $(\mathbf{X}^T\mathbf{X})$ given in Eq. (A2). The units are arbitrary.

to the diagonal of the covariance matrix inflates the squared norms of the columns, ensuring that they are larger than the inner products between columns. As a concrete example, consider the case

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{so that} \quad \mathbf{X}^T\mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \text{(A1)}$$

which is singular. The ridge procedure adds a small amount to the diagonal, proportional to the values on the diagonal:

$$\mathbf{X}^T\mathbf{X} + h^2\text{Diag}(\mathbf{X}^T\mathbf{X}) = \begin{pmatrix} 1+h^2 & 1 \\ 1 & 1+h^2 \end{pmatrix} \equiv \mathbf{X}_r^T\mathbf{X}_r,$$

where (A2)

$$\mathbf{X}_r = \frac{1}{\sqrt{2}}\begin{pmatrix} \sqrt{2+h^2} & \sqrt{2+h^2} \\ -h & h \end{pmatrix}. \quad \text{(A3)}$$

In the limit $h \rightarrow 0$, $\mathbf{X}_r$ reverts back to $\mathbf{X}$. The effects of the ridge procedure are now apparent: it adds to the columns of $\mathbf{X}$ small perturbations, of opposite sign, in the direction perpendicular to the colinear columns. The vectors are lengthened (meaning that the estimate of β will be shorter), and the colinearity is destroyed. In general, the effect of the ridge parameter is to "spread out" the columns of $\mathbf{X}$, lengthening them and making them closer to orthogonal.

A surface of constant uncertainty for a given two dimensional covariance matrix is an ellipse. In the case of

a singular covariance matrix such as that in Eq. (A1), the semiminor axis is zero and the ellipse collapses into a line. As h increases from zero, the ellipse begins to fill out and become two-dimensional (Fig. A1).

REFERENCES

Anderson, T., 2003: *An Introduction to Multivariate Statistical Analysis*. 3rd ed. Wiley, 752 pp.

Banerjee, S., B. P. Carlin, and A. E. Gelfand, 2004: *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall, 472 pp.

Casella, G., and R. Berger, 2002: *Statistical Inference*. Thomson Learning Pacific Grove, 700 pp.

Cook, E., K. Briffa, and P. Jones, 1994: Spatial regression methods in dendroclimatology: A review and comparison of two techniques. *Int. J. Climatol.,* **14,** 379–402.

——, D. Meko, D. Stahle, and M. Cleaveland, 1999: Drought reconstructions for the continental United States. *J. Climate,* **12,** 1145–1162.

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.

Devore, J., 1995: *Probability and Statistics for Engineering and the Sciences*. 4th ed. Brooks/Cole, 750 pp.

Fierro, R., G. Golub, P. Hansen, and D. O'Leary, 1997: Regularization by truncated total least squares. *SIAM J. Sci. Comput.,* **18,** 1223–1241.

Frost, C., and S. Thompson, 2000: Correcting for regression dilution bias: Comparison of methods for a single predictor variable. *J. Roy. Stat. Soc.,* **163,** 173–189.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 2003: *Bayesian Data Analysis*. 2nd ed. Chapman & Hall, 696 pp.

Golub, G., and C. Van Loan, 1980: An analysis of the total least squares problem. *SIAM J. Numer. Anal.,* **17,** 883–893.

——, M. Heath, and G. Wahba, 1979: Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics,* **21,** 215–223.

Hegerl, G., and Coauthors, 2007: Understanding and attributing climate change. *Climate Change 2007: The Physical Science Basis.* S. Solomon et al., Eds., Cambridge University Press, 663–745.

Hoerl, A., and R. Kennard, 1970: Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics,* **12,** 55–67.

Jansen, E., and Coauthors, 2007: Palaeoclimate. *Climate Change 2007: The Physical Science Basis,* S. Solomon et al., Eds., Cambridge University Press, 433–498.

Jones, P., and Coauthors, 2009: High-resolution palaeoclimatology of the last millennium: A review of current status and future prospects. *Holocene,* **19,** 3–49, doi:10.1177/0959683608098952.

Krakauer, N., T. Schneider, J. Randerson, and S. Olsen, 2004a: Linear inversion methods and generalized cross-validation. Online supplement to "Using generalized cross-validation to select parameters in inversions for regional carbon fluxes," 7 pp. [Available online at http://ce193.engr.ccny.cuny.edu/papers/GCV_appendix.pdf.]

——, ——, ——, and ——, 2004b: Using generalized cross-validation to select parameters in inversions for regional carbon fluxes. *Geophys. Res. Lett.,* **31,** L19108, doi:10.1029/2004GL020323.

Lee, T., F. Zwiers, and M. Tsao, 2008: Evaluation of proxy-based millennial reconstruction methods. *Climate Dyn.,* **31,** 263–281.

Luterbacher, J., D. Dietrich, E. Xoplaki, M. Grosjean, and H. Wanner, 2004: European seasonal and annual temperature variability, trends, and extremes since 1500. *Science,* **303,** 1499–1503.

Mann, M., R. Bradley, and M. Hughes, 1998: Global-scale temperature patterns and climate forcing over the past six centuries. *Nature,* **392,** 779–787.

——, S. Rutherford, E. Wahl, and C. Ammann, 2007a: Reply. *J. Climate,* **20,** 5671–5674.

——, ——, ——, and ——, 2007b: Robustness of proxy-based climate field reconstruction methods. *J. Geophys. Res.,* **112,** D12109, doi:10.1029/2006JD008272.

——, Z. Zhang, M. Hughes, R. Bradley, S. Miller, S. Rutherford, and F. Ni, 2008: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc. Natl. Acad. Sci. USA,* **105,** 13 252–13 257.

NRC, 2006: *Surface Temperature Reconstructions for the Last 2000 Years.* The National Academies Press, 160 pp.

Osborn, T. J., and K. R. Briffa, 2006: The spatial extent of 20th-century warmth in the context of the past 1200 years. *Science,* **311,** 841–844.

Rubin, D., 1976: Inference and missing data. *Biometrika,* **63,** 581–592.

Rutherford, S., M. Mann, T. Delworth, and R. Stouffer, 2003: Climate field reconstruction under stationary and nonstationary forcing. *J. Climate,* **16,** 462–479.

——, ——, T. Osborn, R. Bradley, K. Briffa, M. Hughes, and P. Jones, 2005: Proxy-based Northern Hemisphere surface temperature reconstructions: Sensitivity to method, predictor network, target season, and target domain. *J. Climate,* **18,** 2308–2329.

Schneider, T., 2001: Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *J. Climate,* **14,** 853–871.

Smerdon, J., and A. Kaplan, 2007: Comments on ''Testing the fidelity of methods used in proxy-based reconstructions of past climate: The role of the standardization interval.'' *J. Climate,* **20,** 5666–5670.

——, ——, and D. Chang, 2008: On the origin of the standardization sensitivity in RegEM climate field reconstructions. *J. Climate,* **21,** 6710–6723.

Steig, E., D. Schneider, S. Rutherford, M. Mann, J. Comiso, and D. Shindell, 2009: Warming of the Antarctic ice-sheet surface since the 1957 International Geophysical Year. *Nature,* **457,** 459–462.

Tikhonov, A., and V. Arsenin, 1977: *Solutions of Ill-Posed Problems.* Winston, 258 pp.

Tingley, M., 2009: A Bayesian approach to reconstructing space–time climate fields from proxy and instrumental time series, applied to 600 years of Northern Hemisphere surface temperature data. Ph.D. dissertation, Harvard University, 230 pp.

——, 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.

van Huffel, S., and J. Vandewalle, 1991: *The Total Least Squares Problem: Computational Aspects and Analysis.* Society for Industrial Mathematics, 314 pp.

von Storch, H., E. Zorita, and F. González-Rouco, 2009: Assessment of three temperature reconstruction methods in the virtual reality of a climate simulation. *Int. J. Earth Sci.,* **98,** 67–82.

Zar, J. H., 1999: *Biostatistical Analysis.* 4th ed. Pearson Education, 929 pp.

Zhang, Z., M. Mann, and E. Cook, 2004: Alternative methods of proxy-based climate field reconstruction: Application to summer drought over the conterminous United States back to AD 1700 from tree-ring data. *Holocene,* **14,** 502–516.