A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millennium

Jason E. Smerdon and Alexey Kaplan

Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY

Diana Chang

Tri-Institutional Training Program in Computational Biology and Medicine, Cornell University, Ithaca, NY

Michael N. Evans

Department of Geology and Earth Systems Science Interdisciplinary Center, University of Maryland, College Park, MD

Abstract

Canonical correlation analysis (CCA) is evaluated for paleoclimate field reconstructions in the context of pseudoproxy experiments assembled from the millennial integration (850-1999 C.E.) of the National Center for Atmopsheric Research Climate System Model 1.4. A parsimonious method for selecting the order of the CCA model is presented. Results suggest that the method is capable of resolving approximately 3-18 climatic patterns given the estimated proxy observational network and the amount of observational uncertainty. CCA reconstructions are compared to those derived from the regularized expectation maximization method using ridge regression regularization (RegEM-Ridge). CCA and RegEM-Ridge yield similar skill patterns that are characterized by high correlation regions collocated with dense pseudoproxy sampling areas in North America and Europe. Both methods also produce reconstructions characterized by spatially variable warm biases and variance losses, particularly at high pseudoproxy noise levels. RegEM-Ridge in particular is subject to significantly larger variance losses than CCA, even though the spatial correlation patterns of the two methods are comparable. Results collectively indicate the importance of evaluating the field performance of methods that target spatial

climate patterns during the last several millennia, and indicate that the results of currently available climate field reconstructions should be interpreted carefully.

1. Introduction

A concerted research effort over the last decade has focused on reconstructing global or hemispheric climate during the last millennium using networks of climate proxies (e.g. Folland et al. 2001; Jansen et al. 2007, North et al. 2006, Jones and Mann 2004; Jones et al. 2009). These efforts are in many ways an outgrowth of earlier studies that developed reconstructions on regional scales, particularly pioneering work in dendroclimatology that extends back to the 1960s and 70s (e.g. Fritts et al. 1971). Recent efforts have employed single-proxy (e.g. Cook et al. 1994, 2004; Briffa 2000; Briffa et al. 2001; Esper et al. 2002; Evans et al. 2002; D'Arrigo et al. 2006, 2009) or multi-proxy statistical approaches (Mann et al., 1998, 1999, 2005, 2007, 2008; Jones et al., 1998; Crowley and Lowery 2000; Rutherford et al. 2005; Moberg et al. 2005; Hegerl et al. 2007) to calibrate proxy records on observational data during their period of overlap and subsequently to reconstruct past climate variability using derived climate-proxy relationships. Various efforts have demonstrated the promise of these approaches (e.g. Cook et al. 1994, 2004; Mann et al. 1998, 1999; Evans et al. 2002; Luterbacher et al. 1999, 2004; Rutherford et al. 2005; Casty et al. 2005; Pauling et al. 2006), but it in some cases results and methodologies have been vigorously debated (Broecker 2001; Huang et al. 2000; Harris and Chapman 2001; Esper et al. 2002; Beltrami 2002; González-Rouco et al. 2003, 2006; von Storch et al. 2004, 2006; Pollack and Smerdon 2004; Rutherford and Mann 2004; McIntyre and McKitrick 2005; von Storch and Zorita 2005; Bürger and Cubasch 2005; Huybers 2005; Wahl et al. 2006; Bürger et al. 2006; Zorita et al. 2007; Lee et al. 2007; Smerdon and Kaplan 2007, Smerdon et al. 2008a; Wahl and Ammann 2007; Mann et al. 2003, 2005, 2007a, b, c, 2008; Moberg et al. 2005, 2008; Hegerl et al. 2007; Küttel et al. 2007; Christiansen et al. 2009). One of the principal issues of this debate surrounds the magnitude of reconstructed temperature variability during the last millennium on decadal and longer time scales, particularly as it relates to the magnitude, phasing and ubiquity of the putative Medieval Climatic Anomaly and Little Ice Age (e.g. Hughes and Diaz 1994; Broecker 2001; Mann 2002; Mann et al. 2003, 2005, 2007a, b, c). Although a great deal of progress has been made to understand how various reconstructions may or may not accurately represent the characteristics of these past epochs, there remain important unanswered questions about reconstructions uncertainties. These

questions are tied to understanding, for example, the impact of proxy distributions and abundance, the connections between climate and proxy responses across different spectral domains, the nature of multivariate responses in proxy records, and the role of teleconnections and noise in the calibration data - questions that are ultimately fundamental to the success of efforts that attempt to reconstruct past climatic variability (e.g. North et al. 2006; Jansen et al. 2007).

An additional element of uncertainty in climate reconstructions, which has recently gained more attention, is the degree to which specific reconstruction methodologies impose their own error and biases in derived reconstructions. Here we focus specifically on the uncertainties in regional and hemispheric scale temperature reconstructions of the past millennium that arise principally from methodology. Reconstruction methods for this purpose generally can be divided into two groups that have been termed composite-plus scale (CPS; Mann et al. 2005) and climate field reconstruction (CFR) methods (Evans et al. 2001). CPS methods target mean hemispheric or global temperature time series as predictand, therefore yielding reconstructions of only these individual indices (e.g. Groveman and Landsberg 1979, Esper et al. 2002, Crowley and Lowery 2000, Moberg et al. 2005, Hegerl et al. 2007, D'Arrigo et al. 2006). Although CPS methods have the disadvantage of offering no spatial information, they have the benefit of being more straightforward, robust and likely require no more than a few tens of predictors for skillful reconstructions of hemispheric or global temperature variability (e.g. Crowley and Lowery 2000; Hegerl et al. 2007). In contrast to CPS approaches, CFR methods attempt to reconstruct spatial patterns of temperature variability, which is the fundamental promise of these methods (e.g. Cook et al. 1994, Mann et al. 1998, 1999, 2005, 2007a, Rutherford et al. 2005, Evans et al. 2002; Luterbacher et al. 2004, Xoplaki et al. 2005). CFR methods can be complicated, however, by the ill-posed nature of the problem (rank deficiency), are more dependent on the stability of climate-proxy connections and climate teleconnections, and require more extensive distributions of proxies than their CPS counterparts.

In spite of the differences between CPS and CFR methods, the debate surrounding temperature reconstructions of the last millennium has almost exclusively been limited to comparisons between mean NH or global time series (e.g. Briffa and Osborn 2002; Jones and Mann 2004; North et al. 2006; Folland et al. 2001; Jansen et al. 2007); in the case of CFRs, these mean time series are computed from the underlying reconstructed fields. Consequently, there have been few assessments of the robustness of spatial patterns in the collection of available CFRs. Some field comparisons of CFRs have been done

on regional scales. Cook et al. (1994) compared two CFR techniques applied to dendroclimatic series in western Europe and eastern North America and found them to produce similar results. Zhang et al. (2004) similarly compared two methods over the continental United States and found their performance comparable. A more recent study has compared the field skill of two reconstruction methods over the North Atlantic and the European continent (Riedwyl et al. 2008). At global and hemispheric scales, however, proxy distributions are more diffuse, predictor networks comprise multiple proxies, and teleconnection patterns are likely more essential to the skill of the reconstruction. It therefore is crucial to evaluate not only the mean global or hemispheric characteristics of CFRs, but also the spatial skill of the fields derived from these methods.

A significant challenge for CFR comparisons is the fact that most researchers use proxy networks of opportunity and thus of variable composition in proxy type and location. Uncertaintiy in any given reconstruction is therefore the convolution of uncertainties in the method used, the spatial sampling of the proxy network, and the actual climate-proxy connection of each of the proxy series used in the network. If the objective is to isolate the impact of one of these factors, it is difficult to do so from comparisons between these real-world CFR results. The advent of pseudoproxy experiments (Mann and Rutherford 2002) has circumvented some of these challenges, however, by constructing a consistent test bed on which to objectively test reconstruction methodologies (González-Rouco et al. 2006; von Storch et al. 2004, 2006; Mann et al. 2005, 2007a; Hegerl et al. 2007; Smerdon and Kaplan 2007; Smerdon et al. 2008a; Lee et al. 2007; Küttel et al. 2007; Christiansen et al. 2009; Riedwyl et al. 2008).

Pseudoproxy experiments have typically employed millennial integrations from General Circulation Models (GCMs) that only recently have become available (González-Rouco et al. 2003, 2006; Ammann 2007). These experiments are generally performed in the following steps: (1) the complete GCM field is subsampled to mimic the availability of instrumental and proxy information in real-world climate reconstructions of the last millennium; (2) the time series that represent proxy information are perturbed to simulate the spatial and temporal noise characteristics present in real-world proxies; (3) reconstruction algorithms are applied to the model-sampled pseudo “instrumental data” and pseudoproxy series to derive a reconstruction of the climate simulated by the GCM; and (4) the derived reconstruction is compared to the known model target. There are indeed some open questions associated with these experiments, such as whether or not the adopted noise models in the pseudoproxy

network are realistic, and how well the model statistics represent real-world climate characteristics that affect reconstruction skill (e.g. teleconnections). Nevertheless, the utility of pseudoproxy experiments lies in their ability to provide an objective dataset on which to test reconstruction methods. While future improvements in the implementation of pseudoproxy tests will undoubtedly be made, much insight into the performance of multiple reconstruction methods has already been gained from this approach (von Storch et al. 2004, 2006; Mann et al. 2005, 2007a; Smerdon and Kaplan 2007; Lee et al. 2007; Küttel et al. 2007; Hegerl et al. 2007; Christiansen et al. 2009; Riedwyl et al. 2008; Moberg et al. 2008; Smerdon et al. 2008a).

Here we investigate skill and uncertainty in CFRs arising from application of a reconstruction algorithm using canonical correlation analysis (CCA). CCA is a well-established method within the climate sciences (e.g. Anderson 1984; Barnett and Preisendorfer 1987; Bretherton et al. 1992; Cook et al. 1994; Wilks 1995; von Storch and Zwiers 2000; Luterbacher et al. 2000; Tippett et al. 2003, 2008), but has not been widely applied for the purpose of deriving large-scale temperature CFRs (CCA is mentioned in passing within Mann et al. (1998) as being unsuitable for their purposes and has more recently been applied by Christiansen et al. (2009) as one of a number of methods tested in the context of reconstructed NH means). Our purposes herein are to evaluate in detail the application of CCA for reconstructing NH temperatures during the last millennium and to specifically focus on the field characteristics of the derived CFRs.

In addition to investigating the performance of CCA, we compare CCA-derived results to those obtained using the regularized expectation maximization (RegEM) method (Schneider 2001). RegEM has been a recently favored method for NH temperature reconstructions (e.g. Rutherford et al. 2005; Mann et al. 2005, 2007a, 2008), but pseudoproxy experiments also have shown some implementations of RegEM to be susceptible to warm biases and variance losses (Smerdon and Kaplan 2007; Smerdon et al. 2008a; Riedwyl et al. 2008; Christiansen et al. 2009). These findings are consistent with previous pseudoproxy experiments that have demonstrated similar behavior associated with the Mann et al. (1998, 1999) CFR method (von Storch et al. 2004, 2006). A benefit of comparing CCA and RegEM reconstructions is the relative transparency of the CCA method, application of which requires the selection of only three model dimensions to yield a final reconstruction by a linear transformation of rank equal to the smallest of these dimensions. Such transparency allows a straightforward interpretation of the source of skill in CCA results. This characteristic of CCA is in contrast to the more complicated

structure of the iterative RegEM algorithm, which also can have significant computational costs. The comparisons of these two methods therefore can help elucidate the underpinning behavior of both.

2. Data

We use pseudoproxies derived from the millennial simulation (850-1999 C.E.) of the National Center for Atmospheric Research (NCAR) Climate System Model (CSM) 1.4, a coupled atmosphere-ocean GCM that has been driven with natural and anthropogenic forcings (Ammann et al. 2007). The simulated model fields of annual surface temperature means have been interpolated to a 5° longitude-latitude grid (Smerdon et al. 2008b; Rutherford et al. 2008). For consistency in latter comparisons to RegEM-derived results, we use the same realizations of CSM pseudoproxies employed by Mann et al. (2005) with locations shown in Figure 1 (publicly available at http://fox.rwu.edu/ rutherfo/supplements/Pseudoproxy05/). These pseudoproxies were sampled from the 5° grid-box locations that approximate the actual proxy locations of the Mann et al. (1998) multiproxy network, totaling 104 sampled grid cells. Pseudoproxies at these selected locations contain white noise at four different levels to produce pseudoproxies with signal-to-noise ratios (SNRs), by standard deviation, of infinity (noise free), 1.0, 0.5 and 0.25.

To further facilitate comparisons with previous pseudoproxy work, we also use the same subsampled CSM field used by Mann et al. (2005, 2007) to approximate the availability of the instrumental temperature data. Grid points missing more than 30% of the annual data from 1856-1998 C.E. in the Jones et al. (1999) dataset were excluded from use as target data (Mann and Rutherford 2002). This restriction limits the total number of grid cells to 669 in the Eq-70° N region (the target region). Also in keeping with Mann et al. (2005, 2007a), the subsampled instrumental (calibration) data are constrained to the 1856-1980 C.E. period.

3. Methods

3.1 Least-Squares CFRs as Multivariate Linear Regression

Multivariate linear regression is the underlying formalism of most CFR methods. The fundamental approach relates a matrix of climate proxies to a matrix of climate data during a common time interval (generally termed the calibration interval) using a linear model. For instance, let P be an $m \times n$ matrix of proxy values and T be an $r \times n$ matrix of instrumental temperature records where m is the number of proxies, r is the number of spatial locations in the instrumental field, and n is the temporal

dimension corresponding to the period of overlap between the proxy and instrumental data. We write the regression of T columns on P columns for time-standardized matrices (T' and P') with rows that have means of zero and standard deviations of one:

$$T = M_t + S_t T', \qquad P = M_p + S_p P',$$

where M_t is a matrix of identical columns equal to the average of all columns of the matrix T, and S_t is a diagonal matrix with elements that are the standard deviations of the rows of matrix T; M_p and S_p are similarly defined for matrix P. In these terms,

$$T' = BP' + \varepsilon, \tag{1}$$

where B is a matrix of regression coefficients with dimensions $r \times m$, and ε is the residual error. The error variances of all the elements of ε in (1) are simultaneously minimized if B is chosen as:

$$B = (T'P'^T)(P'P'^T)^{-1}, \tag{2}$$

where the superscript T denotes the matrix transpose. Temperature thus can be predicted, or "reconstructed", using this regression matrix during periods in which proxy data are available:

$$\hat{T} = M_t + S_t B S_p^{-1}(P - M_p), \tag{3}$$

where $\hat{T}$ denotes a matrix of reconstructed temperature values.

While the above formalism is straightforward, it works best when the system is overdetermined; that is, the time dimension n is much larger than the spatial dimension m, because the covariances are more reliably estimated. The challenge for CFR methods involves the manner in which B is estimated in practical situations when this condition is not met. It is often the case in climate applications that the number of target variables exceeds the time dimension, yielding a rank-deficient (ill-posed) problem. For instance, in most global or NH CFRs, the number of grid cells in the climate field is typically on the order of many hundreds or a few thousands, while the observational record usually contains 150 annual fields or less. The number of proxies is typically on the order of a few tens to hundreds, which may exceed or at least be comparable with the time dimension. In such cases, the covariance matrices $\langle T'P'^T \rangle$ and $\langle P'P'^T \rangle$ cannot be well estimated, even to the extent that $\langle P'P'^T \rangle$ can be ill-conditioned.

The inversion in (2) therefore requires some form of regularization. Published linear methods for global temperature CFRs vary primarily in the form of this regularization. In the following subsections we discuss CCA and RegEM as the two regularization approaches considered in this manuscript.

3.2. Canonical Correlation Analysis

For the purposes described herein, we outline the Barnett and Preisendorfer (1987) version of CCA formalism as presented by Tippett et al. (2003, 2008). We begin by decomposing the standardized proxy matrix P' during the calibration interval using Singular Value Decomposition (SVD; Golub and Van Loan, 1996):

$$P' = U_p \Sigma_p V_p^T. \tag{4}$$

where the columns of U_p represent spatial patterns (empirical orthogonal functions or EOFs) and the principal components (PCs), V_p, are orthonormal timeseries of coefficients that combine with the EOF patterns to produce the original data set. The diagonal matrix Σ_p contains non-negative elements (singular values), arranged in descending order, with squares proportional to the variance captured by the corresponding EOF-PC pairs. If the diagonal elements of Σ_p decrease quickly, as is often the case in climatological data where leading climate patterns dominate over many more weakly expressed local patterns or noise, a reduced-rank representation of P' using only a few leading EOF-PC pairs is typically a good approximation of the full-rank version. Thus we employ a reduced rank representation of P' such that d_p EOF-PC pairs are retained:

$$P^r = U_p^r \Sigma_p^r {V_p^r}^T. \tag{5}$$

Here P^r denotes the reduced-rank representation of P', and matrices with the superscript r are the truncated versions of the SVD factors corresponding to the retained number of d_p singular values. Similarly, the reduced-rank version of T' is written:

$$T^r = U_t^r \Sigma_t^r {V_t^r}^T, \tag{6}$$

where T^r only uses d_t singular values and the corresponding number of singular vectors. Note that $\mathrm{rank}(P^r) = d_p$ and $\mathrm{rank}(T^r) = d_t$, while $\mathrm{rank}(P') = \min(m, n-1)$ and $\mathrm{rank}(T') = \min(r, n-1)$.

Now we use multivariate linear regression with a matrix B'

$$V_t^{r\,T} = B' V_p^{r\,T} + \varepsilon_v$$

(ε_v is the residual error) to predict the PCs of temperature using the proxy PCs:

$$\hat{V}_t^{r\,T} = B' V_p^{r\,T}.$$

Because the PCs are orthonormal, $V_p^{r\,T} V_p^r = I$ (i.e. the identity matrix), the expression for B' simplifies:

$$B' = (V_t^{r\,T} V_p^r)(V_p^{r\,T} V_p^r)^{-1} = V_t^{r\,T} V_p^r.$$

The last expression for B' can be decomposed using SVD:

$$B' = V_t^T V_p^r = O_t \Sigma_{\rm cca} O_p^T. \tag{7}$$

and can then be truncated by retaining only $d_{\rm cca} \leq \min(d_p, d_t)$ leading singular values and corresponding patterns:

$$B^r = O_t^r \Sigma_{\rm cca}^r O_p^{r\,T}. \tag{8}$$

Predicting temperature PCs using B^r, instead of B', i.e.

$$\hat{V}_t^{r\,T} = O_t^r \Sigma_{\rm cca}^r O_p^{r\,T} V_p^{r\,T}$$

,

transforms into a simple form:

$$\hat{Q}_t^T = \Sigma_{\rm cca}^r Q_p^T. \tag{9}$$

if written in terms of the CCA time series; these are projections of the PC vectors V_t^r and V_p^r onto the sets of patterns O_t^r and O_p^r, respectively:

$$Q_t = V_t^r O_t^r, \qquad Q_p = V_p^r O_p^r. \tag{10}$$

Similarly, the predicted $\hat{Q}_t$ corresponds to the predicted temperature PCs $\hat{V}_t^r$:

$$\hat{Q}_t = \hat{V}_t^r O_t^r.$$

To obtain the CCA timeseries, Q_t and Q_p, directly from the standardized data sets, it is convenient to define weight matrices,

$$W_t = U_t^r (\Sigma_t^r)^{-1} O_t^r, \qquad W_p = U_p^r (\Sigma_p^r)^{-1} O_p^r, \tag{11}$$

so that

$$Q_t^T = W_t^T T', \qquad Q_p^T = W_p^T P', \tag{12}$$

where the columns of Q_t and Q_p are orthonormal. Moreover, due to (7) and (10),

$$Q_t^T Q_p = \Sigma_{\rm cca}^r.$$

The columns of Q_t and Q_p with different ordering are therefore also orthogonal, while those with the same ordering are positively correlated. The correlation coefficients of these latter columns, also called canonical correlations, are equal to the diagonal elements of $\Sigma_{\rm cca}^r$ and because of the SVD decomposition in (7), are maximized in the following sense: the correlation coefficient between the first columns of Q_t and Q_p is the largest among the projections of V_t^r and V_p^r on any unit length vectors (patterns); these maximizing patterns are the first columns of O_t^r and O_p^r, respectively. The remaining correlation coefficients are arranged in descending order, i.e. the coefficient between the second columns of Q_t and Q_p is the largest among projections of V_t^r and V_p^r on unit length vectors orthogonal to the first columns of O_t^r and O_p^r, respectively, and the patterns that achieve the latter correlation are the second columns of O_t^r and O_p^r; the correlation coefficient between the third columns of Q_t and Q_p is the largest among projections of V_t^r and V_p^r on unit length vectors orthogonal to the first and second columns of O_t^r and O_p^r, and so on.

The predictions of the CCA temperature time series by (9) amount to a simple multiplication of the CCA time series of the proxies by the diagonal elements of $\Sigma_{\rm cca}^r$. To perform these predictions for the fields of temperature on the basis of the original proxy data, however, we require the spatial patterns of their regression on the CCA timeseries:

$$T' = C_t Q_t^T + \varepsilon_t, \qquad P' = C_p Q_p^T + \varepsilon_p.$$

To determine C_p and C_t (the CCA patterns) or the CCA homogeneous covariance maps, we use the orthonormality of the CCA timeseries and the decomposition in (4):

$$C_p = (P' Q_p)(Q_p^T Q_p)^{-1} = P' Q_p = U_p \Sigma_p V_p^T V_p^r O_p^r = U_p^r \Sigma_p^r O_p^r \tag{13}$$

and similarly,

$$C_t = T' Q_t = U_t^r \Sigma_t^r O_t^r. \tag{14}$$

Thus the use of the low-rank CCA approximations in (5), (6), and (8) in the regression matrix formula given in (2) results in,

$$B_{\rm cca} = U_t^r \Sigma_t^r {V_t^r}^T V_p^r (\Sigma_p^r)^{-1} {U_p^r}^T = U_t^r \Sigma_t^r O_t^r \Sigma_{\rm cca}^r {O_p^r}^T (\Sigma_p^r)^{-1} {U_p^r}^T,$$

if the inverse of the proxy covariance matrix is replaced by the pseudo-inverse (Golub and Van Loan, 1996):

$$(P' P'^T)^{-1} \longrightarrow (P' P'^T)^+ = (P^r {P^r}^T)^+ = U_p^r (\Sigma_p^r)^{-2} {U_p^r}^T.$$

Given the definitions in Eqs. (11) and (14), $B_{\rm cca}$, takes a simple form:

$$B_{\rm cca} = C_t \Sigma_{\rm cca}^r W_p^T. \tag{15}$$

Applying $B_{\rm cca}$ to P' in order to reconstruct T' is therefore equivalent to a three-step procedure:

(i) use the weighting patterns W_p to convert P' into the CCA time series

$$Q_p^T = W_p^T P',$$

(ii) scale these time series by the canonical correlations, i.e. the diagonal element of $\Sigma_{\rm cca}^r$, to produce the CCA timeseries for temperature:

$$\hat{Q}_t^T = \Sigma_{\rm cca}^r Q_p^T,$$

(iii) and use the C_t patterns to reconstruct a standardized version of the temperature fields:

$$\hat{T}' = C_t \hat{Q}_t^T.$$

Since the CCA patterns and canonical correlations are determined at the calibration stage and do not change thereafter, the CCA time series provide the main and simplified diagnostics of the reconstruction performance; i.e. the actual CCA temperature time series,

$$Q_t^T = W_t^T T',$$

during the reconstruction period can be directly compared with their prediction on the basis of the proxies in item (ii) above. The use of these statistics are illustrated further in Section 4.2.1.

For the non-standardized version of temperature fields and proxies given in (3), the CCA temperature CFR becomes

$$\hat{T} = M_t + S_t B_{\rm cca} S_p^{-1}(P - M_p). \tag{16}$$

Performing this reconstruction requires the determination of five matrices: two in which all columns contain the mean vectors for the temperature field and the proxies, M_t and M_p; the two diagonal matrices of the temperature and proxy standard deviations, S_t and S_p, and the CCA low-rank regression matrix $B_{\rm cca}$. Under the assumption of stationarity between the mutual proxy and climate statistics, (16) can be used to reconstruct temperatures in any temporal interval, including those outside of the calibration period. The only formal change is in the number of columns in matrices M_t and M_p, which of course changes to match the length of the given reconstruction period.

The operator $B_{\rm cca}$ is a reduced-rank ($\mathrm{rank}(B_{\rm cca}) = d_{\rm cca}$) representation of the standard multivariate regression operator. Given calibration interval datasets T and P, upon the selection of three parameters for truncated ranks, $d_{\rm cca}$, d_p, and d_t, the matrix $B_{\rm cca}$ is completely determined. Note that traditional applications of CCA did not involve rank reductions of the predictor and predictand matrices, and thus only depended on $d_{\rm cca}$ (see the discussion in Bretherton et al. 1992). Steps for reducing these matrix ranks by selecting the d_p and d_t parameters prior to estimating the CCA time series and maps were added by Barnett and Preisendorfer (1987) (termed the BP method by Bretherton et al. 1992). Tippett et al. (2003) and Christiansen et al. (2009) used and referred to this latter BP version as CCA, as do we hereinafter.

3.3. CCA Model-Dimension Selection

Appropriate selections of the $d_{\rm cca}$, d_p, and d_t dimensions are crucial for the application of the CCA method. Previous CCA applications have proposed various forms of model selection. Christiansen et al. (2009) set the number of dimensions by maintaining a specific level of retained variance and imposing the additional constraint that the dimensional reduction be equal for all three decompositions.

Barnett and Preisendorfer (1987) used principal component truncation rules for the reduction of the predictor and predictand data (d_p and d_t) as proposed by Preisendorfer et al. (1981). The number of canonical coefficients ($d_{\rm cca}$) was then estimated using jackknife cross-validation statistics computed for a set of withheld single time samples ("leave-one-out"). Tippett et al. (2003) employed a similar approach, but used a jackknife cross-validation scheme to optimize all three truncation dimensions. Our approach is similar to the latter application except we use a much cheaper "leave-half-out" approach to cross-validation to reduce computational costs. In an application using proxy data series with annual resolution, this approach is also more conservative with respect to validation of the reconstructed decadal-centennial timescale variations.

To perform the leave-half-out cross-validation procedure, the period when both instrumental and proxy data are considered available is split into two temporal halves: 1856-1917 and 1918-1980 C.E. We generate two sets of reconstructions using (16) and calibrate using each half of the target data to estimate the $B_{\rm cca}$ matrix, as well as the means and standard deviation fields for the proxy and temperature data ($M_p,\ S_p,\ M_t,\ S_t$). The reconstructions are verified on the left-out halves. Two cross-validation statistics are used: (1) the area-weighted Root Mean Square Error (RMSE) of the reconstructed field relative to the target; and (2) the correlation between the reconstructed and target area-weighted mean NH time series (hereinafter termed NH mean correlation (NHMC)). These validation statistics from both experiments are combined to determine the statistics for the entire instrumental data interval 1856-1980.

Using the above cross-validation scheme we compute the RMSE and NHMC for a range of $d_{\rm cca}$, d_p and d_t combinations. The optimal selection of $d_{\rm cca}$, d_p and d_t is based on the cross-validated reconstruction skill in terms of either small RMSE or large NHMC. After this selection, all the matrix parameters of (16) can be computed for the entire calibration interval (1856-1980 C.E.) and used for reconstructions in the preinstrumental period. Using the definitions

$$B_f = S_t B_{\rm cca} S_p^{-1}, \qquad M_f = M_t - B_f M_p, \tag{17}$$

the reconstruction in (16) can be rewritten in the final form of a linear transform with an offset:

$$\hat{T} = M_f + B_f P. \tag{18}$$

All columns of matrix the M_f are identical and specify offsets for all r locations of the predicted temperature fields, therefore M_f contains r independent parameters. The linear-transform matrix B_f

has the dimensions $r \times m$ and thus contains $rm = 669 \cdot 104 = 69,576$ elements. This number is about one third smaller than the number of of elements in the target temperature data during the calibration period ($rn = 669 \cdot 125 = 96,625$) from which the elements of B_f must be determined. Fortunately, not all elements in B_f are independent parameters because of the CCA rank reduction. Since $B_{\rm cca}$ has rank $d_{\rm cca}$, and B_f is obtained in (17) by multiplying $B_{\rm cca}$ by non-singular diagonal matrices, B_f has the same size ($r \times m$) and rank ($d_{\rm cca}$) as $B_{\rm cca}$. Such a matrix has $d_{\rm cca}$ non-zero singular values and as many left and right singular vectors corresponding to these values. Using the non-zero singular values of B_f in non-increasing order to form a diagonal matrix Σ and arranging the corresponding singular vectors as the columns of matrices U and V, we can uniquely (up to the reordering of the columns in U and V corresponding to identical singular values) present B_f as

$$B_f = U\Sigma V^T. \tag{19}$$

The first column of U, as a unit vector in the r-dimensional space, has $r-1$ degrees of freedom. The second column, subject to an additional constraint of orthogonality to the first column has $r-2$ degrees of freedom, etc. Therefore the entire matrix U, consisting of $d_{\rm cca}$ orthonormal vectors has

$$N(U) = \sum_{i=1}^{d_{\rm cca}} (r-i) = rd_{\rm cca} - \frac{d_{\rm cca}(d_{\rm cca}+1)}{2} = d_{\rm cca}\left(r - \frac{d_{\rm cca}+1}{2}\right).$$

Similarly, the number of independent parameters in V is

$$N(V) = d_{\rm cca}\left(m - \frac{d_{\rm cca}+1}{2}\right),$$

and $N(\Sigma) = d_{\rm cca}$. In the general case, non-zero singular values of a matrix B_f are different, the decomposition (19) is unique and, therefore,

$$N(B_f) = N(U) + N(\Sigma) + N(V) = d_{\rm cca}(r + m - d_{\rm cca}).$$

Together with the constant offset parameters, the number of independent parameters that have to be determined in order to produce the reconstruction formula (18) is

$$N_{\rm tot} = N(B_f) + N(M_f) = d_{\rm cca}(r + m - d_{\rm cca}) + r. \tag{20}$$

Substituting the values of r and m specific to the present pseudoproxy scenario ($r = 669$ and $m = 104$), we find that for the CCA reconstructions presented herein

$$N_{\mathrm{tot}} = 669 + 773 d_{\mathrm{cca}} - d_{\mathrm{cca}}^2. \tag{21}$$

The number of independent parameters in the CCA reconstructions therefore depends only on d_{cca}, the number of CCA modes retained. The number does not depend on d_p and d_t, i.e. the numbers of retained EOF modes for the proxy and temperature data, respectively. The actual values of B_f and M_f in (18) of course do depend on the d_p and d_t choices, but the underlying number of parameters that need to be specified in order to determine these values does not. Furthermore, when $d_{\mathrm{cca}} \ll r + m = 773$, the d_{cca}^2 term in (20) and (21) is negligible compared to $(r + m) d_{\mathrm{cca}} = 773 d_{\mathrm{cca}}$. Analyses we will present suggest that reasonable values of d_{cca} are well below 50. Therefore, N_{tot} grows nearly linearly with d_{cca}, and 773 additional parameters need to be specified in the coefficients of (18) when d_{cca} increments by 1. Considering the relative shortness of the data set available for calibration and cross-validation, choosing a reconstruction model that requires a smaller, rather than larger number of free parameters (i.e. value of d_{cca}) becomes especially important. In Section 4.1 we demonstrate a practical means of selecting the smallest d_{cca} that produces a reconstruction with cross-validated RMSE that is practically indistinguishable from the absolute minimum of RMSE over all combinations of d_{cca}, d_p, and d_t.

3.4. RegEM

Application of the RegEM method to the problem of NH CFRs has been discussed in detail within the literature (Schneider 2001; Rutherford et al. 2005; Mann et al. 2005, 2007a,c, 2008; Smerdon and Kaplan 2007; Lee et al. 2007; Smerdon et al. 2008a; Christiansen et al. 2009; Riedwyl et al. 2008). While RegEM is an iterative method, the underlying formalism is based on a linear regression model that reconstructs missing data X_m from available data X_a and can be written as

$$X_m = M_m + S_m B S_a^{-1} (X_a - M_a). \tag{22}$$

The notation here is analogous to (3), except the subindices a and m denote available and missing data respectively.

For the conventional expectation maximization algorithm, in which regularization is not employed, the estimate of the regression matrix B is given, in full analogy to (2), by the standard multivariate

regression formula for standardized data sets X'_m and X'_a:

$$B = (X'_m X'^T_a)(X'_a X'^T_a)^{-1}. \tag{23}$$

Similar to CCA, however, regularization is required for application to CFRs of the last millennium. Multiple regularization approaches for the expectation maximization algorithm have been discussed (Schneider 2001; Rutherford et al. 2005; Mann et al. 2005, 2007a,c; Smerdon and Kaplan 2007; Christainsen et al. 2009), but the differences between reconstructions derived from these approaches has not been sufficiently explored (Smerdon et al. 2008a). For our purposes herein we employ the more widely applied ridge regression regularization in which the inverse covariance matrix in (23) is replaced by

$$(X'_a X'^T_a)^{-1} \longrightarrow (X'_a X'^T_a + h^2 I)^{-1} \tag{24}$$

where h is a positive number called the ridge parameter (see Schneider (2001) for a detailed derivation and discussion of these equations). In keeping with the reconstructions performed by Rutherford et al. (2005) and Mann et al. (2005), h is chosen herein by minimization of the generalized cross validation (GCV) function. Although the further details of the RegEM method are extensive, it is important to note that the final RegEM reconstruction in this millennial CFR context is simply derived from a set of linear operators (Smerdon et al. 2008a). The final reconstruction is thus principally the product of the regularized regression coefficients acting on the proxy matrix, i.e. it takes the form of (22) for a specific choice of M_a, S_a, M_m, S_m, and B. If the proxy data in P are substituted for the "available" data X_a, and the "missing" data X_m are taken to be temperature T during the reconstruction interval, then the RegEM reconstruction in (22) essentially becomes (3) and is comparable to the same form given for the CCA reconstruction in (16). The main difference between these formulas is of course the form of regularization used for the regression matrix B, and the iteratively computed mean offset, scale matrices and regression coefficients in the RegEM formula (22). In fact, both of these reconstruction formulas can be brought to the form in (18) using one offset and one linear transformation.

4. Reconstruction Results

4.1 Selected model dimensions

Our motivation is to select $d_{\rm cca}$, d_p, and d_t values for the collection of CCA reconstructions that calibrate the 104 pseudoproxies on the instrumental period from 1856-1980 C.E. and compute CFRs during the interval 850-1855 C.E. These are the same experiments performed by Mann et al. (2005, 2007a) to test the RegEM method using white-noise pseudoproxies. In all cases, P and T are standardized over the calibration period prior to estimating the regression matrix $B_{\rm cca}$ using equation (15); reconstructions during the validation period are performed using (16).

Following the approach described in Section 3.3, CCA was calibrated on each half of the data during the instrumental period and tested on the other half using all combinations of $d_{\rm cca}$, d_p and d_t between 1 and 50 modes such that $d_{\rm cca} \leq \min(d_p, d_t)$ (yielding $d_{\rm cca}^2$ triplets $(d_{\rm cca}, d_p, d_t)$ for each $d_{\rm cca}$ value between 1 and 50 and thus a total of $1^2 + 2^2 + \cdots + 50^2 = 50 \cdot (50+1) \cdot (2 \cdot 50 + 1)/6 = 42,925$ reconstruction models). Test results for both halves of the instrumental period were combined to produce cross-validation statistics for the entire interval and a given set of dimensions. RMSE values were combined as the square root of the mean residual sum of squares in the two intervals and NHMCs were calculated as the average correlation coefficients for the two intervals weighted by the number of years in each interval.

Table 1 displays the minimum RMSE and maximum NHMC values among all $d_{\rm cca}$, d_p, and d_t combinations used, as well as the dimensional combinations that achieve these extrema. Results are tabulated for each pseudoproxy noise level. While the two statistics are optimized at somewhat similar dimensional combinations, the results are not identical; the alternative statistic for each optimization is also provided in Table 1.

The RMSE and the NHMC statistics are plotted in Figure 2 for an SNR of 0.5, showing that the former generally decreases as the latter increases. More importantly, the range of possible NHMCs decreases as the RMSE becomes smaller. The reciprocal constraint, however, is much weaker: increases in NHMCs are not accompanied by nearly as large a decrease in the range of RMSE. For instance, when confined to a range of RMSE values within 1% of the minimum, the range of possible NHMCs spans 12% their total range. By contrast, if confined to the range of NHMCs that are within 1% of the maximum, the range of possible RMSE values spans 34% of the total RMSE range. These observations suggest that RMSE is a more robust statistic for optimizing the CCA reconstructions than the NHMC. Furthermore, the color of the circles in Figure 2 represent $d_{\rm cca}$, i.e. corresponds to the number of independent parameters in the reconstruction model that is being validated. While particularly small

d_{cca} (less than 10) correspond to reconstructions that are both poor in RMSE and NHMC performance, high d_{cca} (larger than 30) correspond to high NHMC but the full range of RMSE values. RMSE performance is especially poor for reconstructions with the largest d_{cca} values. We therefore use RMSE as the principal basis for our selection criterion in subsequent dimensional selections.

As mentioned earlier, the total number of combinations used to determine the optimized dimensions given in Table 1 is 42,925. This collection of models was tested for their cross-validated performance on only 125 annual fields of target data, thus some combinations might correspond to low RMSE simply by chance and yield optimal reconstructions impacted by artificial skill. To guard against this likelihood we adopt a conservative selection strategy that seeks to find the most parsimonious of acceptable models by minimizing the number of free parameters in the final reconstruction model, which is equivalent to minimizing d_{cca} without deviating significantly from the absolute minimum RMSE. Figure 3 plots RMSE versus d_{cca} for all tested combinations of the CCA dimensions at each pseudoproxy noise level; the black dashed line connects the RMSE minima for each value of d_{cca}:

$$\mathrm{RMSE}^*(d_{\mathrm{cca}}) = \min_{d_p, d_t} \mathrm{RMSE}(d_{\mathrm{cca}}, d_p, d_t).$$

If $d_p^*(d_{\mathrm{cca}})$ and $d_t^*(d_{\mathrm{cca}})$ are values of d_p and d_t that respectively minimize $\mathrm{RMSE}(d_{\mathrm{cca}}, d_p, d_t)$ for a given d_{cca}, then the triplet $(d_{\mathrm{cca}}, d_p^*(d_{\mathrm{cca}}), d_t^*(d_{\mathrm{cca}}))$ defines the optimal (by the cross-validated RMSE criterion) CCA reconstruction among all models with a fixed number of independent parameters corresponding to a given d_{cca}. Figure 3 demonstrates that $\mathrm{RMSE}^*(d_{\mathrm{cca}})$ decreases steeply for all noise levels at small values of d_{cca}. Beginning at a given d_{cca} value, however, this drop is replaced by a rather flat plateau. For all noise levels except the highest one, the absolute minimum (identified by the closed circle) is rather far from the beginning of this plateau. Alternatively, using the d_{cca} value corresponding to the beginning of the plateau yields a solution with an RMSE performance that is similar to the absolute RMSE minimum but corresponds to a model with a much smaller number of independent parameters.

We identify the beginning of the plateau by selecting the minimum d_{cca} at which an increase by one does not reduce $\mathrm{RMSE}^*(d_{\mathrm{cca}})$:

$$d_{\mathrm{cca}}^* = \min\{d_{\mathrm{cca}} : \ \mathrm{RMSE}^*(d_{\mathrm{cca}}) \leq \mathrm{RMSE}^*(d_{\mathrm{cca}} + 1)\}.$$

Optimal solutions $(d_{\mathrm{cca}}^*, d_p^*(d_{\mathrm{cca}}^*), d_t^*(d_{\mathrm{cca}}^*))$ are identified by stars in the panels of Figure 3 and are

listed in Table 2 along with the corresponding values of RMSE and NHMC cross-validation statistics. At any noise level, $\mathrm{RMSE}^*(d^*_{\mathrm{cca}})$ does not exceed $\min_{d_{\mathrm{cca}}}(\mathrm{RMSE}^*)$ by even 0.5%. In subsequent presentations herein, we use these "beginning of the plateau" solutions $(d^*_{\mathrm{cca}}, d^*_p(d^*_{\mathrm{cca}}), d^*_t(d^*_{\mathrm{cca}}))$ as our preferred choices of the CCA dimensions (termed the preferred solutions hereinafter).

Note that in the preferred solutions, the values of d_p and d_t are chosen as those corresponding to the absolute minimum of RMSE for the preselected value of d_{cca}. Relatively fluid color transitions in the panels of Figure 3 suggest smooth but significant dependence of RMSE on d_p. This impression is borne out in a more detailed illustration of the RMSE dependence on the CCA parameters $(d_{\mathrm{cca}}, d_p, d_t)$: Figure 4 presents two-dimensional fields of the RMSE minima with respect to the individual dimensions. The area of the RMSE minimum is quite wide, therefore changes in d_p or d_t by a few units should not affect the reconstruction quality very much. The dependence of RMSE on d_t is particularly poorly constrained by the data: for all d_{cca} in the range between 5 and 30, a value of d_p could be selected so that RMSE is quite close to the absolute minimum for any value of d_t exceeding d_{cca}.

4.2 CCA Reconstructions

4.2.1 Assembly of the CCA Reconstructions

To demonstrate the individual elements of the CCA reconstruction we plot in Figure 5 the homogeneous covariance maps (C_t and C_p) and the associated time series (Q_t) for the first three canonical patterns of the no-noise reconstruction. In the case of Q_t, we plot both the true time series from the target data, as well as the estimated time series from the pseudoproxy matrix ($\Sigma_{cca} Q_p^T$).

The three temperature covariance maps plotted in Figure 5 take on dynamically interpretable characteristics, although the patterns are rotated from the original model EOFs. The three plotted maps combine features of global-warming-, El-Nino-Southern-Oscillation-, and North-Atlantic-Oscillation-like patterns. This demonstration illustrates the physical interpretability of the derived covariance maps, which ultimately can be evaluated in terms of the reconstructive skill associated with individual dynamical patterns in the field.

As demonstrated in step (ii) of the three-step procedure in Sec 3.2, the time series of the temperature covariance maps are estimated during the reconstruction interval by the product of the canonical coefficients and the time series of the proxy covariance maps. These time series are plotted in Figure 5 and compare closely to the true time series of the temperature covariance maps. Correlations between

the true and estimated time series for these first three patterns are all above 0.99 in the calibration interval and above 0.98 in the reconstruction interval (see Table 3 for these statistics at all noise levels). As dictated by the CCA formulation, correlations within the calibration interval progressively decrease from the maximum of the first pattern for all noise levels (Table 3). This is interestingly not the case in the reconstruction interval when some of the correlations for higher-order patterns exceed those of the lower-order patterns.

Figure 5 also plots the relative values of the proxy covariance maps for the first three canonical patterns. These maps scale location markers for the 104 pseudoproxies by their relative loadings and also designate where the loadings are positive or negative using the color of the markers. Upon inspecting the two sets of temperature and pseudoproxy covariance maps one can see the that the proxy maps effectively reflect local sampling from the temperature maps. For instance, the leading canonical pattern associated with predominant warming is reflected in the proxy map that contains universally positive loadings. In the other two patterns, the positive and negative loadings are roughly collocated with the areas of positive and negative temperatures in the temperature covariance maps.

4.2.2 Northern Hemisphere Means

The temperature covariance maps and proxy-estimated time series presented in Figure 5 are combined to yield a complete field reconstruction for each of the investigated noise levels. The total number of combined patterns is of course dictated by the number of retained d_{cca} values, which were determined for the preferred solutions in Section 4.1 to range from 18 in the no-noise case to 3 at an SNR of 0.25 (see Table 2). Complete CCA reconstructions are assembled from these collections of patterns and time series. We first plot the area-weighted mean NH time series associated with these complete reconstructions in Figure 6a.

The correlations between the reconstructed mean NH time series and the model target are all significant, even though they reduce with increasing noise levels (Table 4). These correlations are interestingly less than those determined for the first three canonical patterns at all noise levels given in Table 3. This is indicative of the fact that the leading individual patterns (Table 3) are reconstructed more skillfully than the combined field comprising the scaled full range of canonical patterns.

Although the determined correlations are all significant, the time series in Figure 6a suffer from warm biases and variance losses during the reconstruction interval, both of which increase with higher

noise levels. This behavior is not associated with the difference between the dimensions chosen for the preferred solutions in Section 4.1 and those for the absolute minimum RMSE: Figure 6b plots the mean time series from the reconstructions using the latter-derived dimensions and the results still suffer from the observed effects. These absolute-minimum time series correlate with the preferred-solution reconstructions at levels of r = 0.97 or better. In fact, it is virtually impossible to discern the differences between the two panels in Figure 6, pointing to the robustness of the achieved results and the prevalence of the observed warm biases and variance losses in the NH means.

Box plots in Figures 6c and 6d further demonstrate the warm biases and variance losses in the reconstructed NH time series. These plots also illustrate the reduced number of extreme events in the reconstructed time series relative to the known model target. These extrema are typically associated with volcanic events in the model simulated NH mean, and are manifest as cold outliers in both the model target and the reconstructed time series. The number and extent of the outliers is diminished in the reconstructed time series, however, and indicates that the reconstructions have the potential to miss the characterization of these important annual events in the model simulated climate.

4.2.3 Reconstructed Fields

We further confirm the robustness of the reconstructions derived for the preferred solutions against those for the absolute minimum RMSE dimensions in Figure 7 showing the local correlations between the two reconstructions for SNRs of 1.0 and 0.5. For all noise cases, the area-weighted mean field correlations are 0.97, 0.95, and 0.89 for SNR = infinity, 1.0, and 0.5, respectively (note that the dimensions for SNR = 0.25 were the same in both cases and all correlations were determined over the reconstruction interval from 850-1855 C.E.). These comparisons demonstrate a spatial consistency between the two dimensional choices and suggest that the large-scale features are well captured for different sets of CCA dimensions (assuming the RMSE is held close to the absolute minimum).

Figure 8 shows the spatial distributions of validation statistics for the preferred CCA reconstructions at SNRs of 1.0 and 0.5; statistics are computed during the reconstruction interval and summary statistics for all noise levels are given in Table 4. Field correlations of course reduce with increased noise, but Figure 8 illustrates the spatial variability of the local correlation coefficient. In all reconstructions, regions containing the largest correlations are over North America and Europe. These regions correspond to the areas with the largest density of pseudoproxies (see Figure 1), i.e. the reconstruc-

tions perform best where the field is sampled the most. Similarly, regions that are not sampled in the pseudoproxy network have comparatively low verification correlations. Correlations fall to particularly low values over some regions (e.g. subtropical and mid-latitude ocean basins or the Asian continent) at high-noise levels.

The warm biases and variance losses observed in the mean NH time series (Figure 6) are also manifest in the reconstructed fields, but their spatial patterns show important regional distinctions (Figure 8). Standard deviation ratios (reconstructed standard deviation divided by the corresponding model value) indicate that variance is most strongly preserved in areas where field correlations are high, whereas variance losses are largest over the ocean basins where the lowest field correlations are observed. Overall, significant variance losses are observed for all noise levels: the area-weighted mean standard deviation ratio is respectively 0.58 and 0.44 for the SNR cases of 1.0 and 0.5 shown in Figure 8, while the ratio drops to 0.37 at a SNR of 0.25 (Table 4). Additionally, large variance losses can accompany reconstructions with relatively high correlations in the field: standard deviation ratios drop below 0.5 in many regions of the reconstruction for an SNR of 1.0 (Figure 8).

Mean biases also display regional variations, although they appear more spatially uniform than observed for the local correlations or standard deviation ratios. While most regions of the reconstructions are warmer than the actual model field, means are colder in a few areas (e.g. North America and the North Atlantic). The proportion of colder to warmer regions is reduced with increasing noise levels and is reflected in the average mean biases calculated for the fields (see Table 4); high-noise reconstructions therefore are dominated by warm-biased regions.

The bottom panels in Figure 8 show the RMSE of the fields, which essentially combine errors associated with variance losses and mean biases. The RMSE patterns follow most closely the patterns in the mean biases, indicating that the error is dominated by differences between the reconstructed and actual means. Contrary to the correlation patterns, it is also important to note that the RMSE is in some cases largest over regions where the pseudoproxy network is densest. Mean biases, and therefore RMSE, do not appear to be as strongly tied to the distribution of the pseudoproxy network as the correlation and standard deviation ratios.

4.3. Comparison of CCA and RegEM Reconstructions

We have used the same pseudoproxies from the CCA experiments above to compute corresponding

non-hybrid (Rutherford et al. 2005) RegEM-Ridge reconstructions. The derived reconstructions are the same reconstructions presented by Smerdon et al. (2008a) and employ a standardization scheme realistically confined to the calibration interval (Smerdon and Kaplan 2007). Figures 9a and 9b compare the mean NH time series computed from the CCA and RegEM-Ridge reconstructed fields at SNRs of 1.0 and 0.5. The time series at all noise levels compare very closely: correlations between the CCA and RegEM-Ridge time series are 0.96, 0.97, 0.96, and 0.89 for SNR=infinity, 1.0, 0.5, 0.25, respectively. The reconstructed NH means also correlate with the true model mean at comparable levels (Table 4). There is, however, an indication that the RegEM-Ridge method performs slightly better at higher noise levels given that the correlations increase by a few hundredths above those observed for CCA. The mean biases and variance losses are larger in the RegEM-Ridge reconstructions, however, and can be clearly seen in the box plots in Figures 9c and 9d. The failure to reconstruct extreme events is also most strongly associated with the RegEM-Ridge reconstructions as illustrated in these latter panels of Figure 9.

The correlation fields between the CFRs derived from the two methods are plotted in Figure 10, again showing results for SNRs of 1.0 and 0.5. Correlations between the two reconstructions depend on location, but overall the area-weighted mean field correlations are 0.89, 0.92, 0.85, and 0.65 for SNR=infinity, 1.0, 0.5, 0.25, respectively. As discussed in Section 3, CCA and RegEM-Ridge select regression coefficients in two distinctly different ways, but the widespread high field correlations between the results from both methods indicate that they reconstruct similar patterns of variability in the target field (note that the exact same pseudoproxies have been used for each of these methods).

Validation fields for the RegEM-Ridge reconstructions are shown in Figure 11. These are directly comparable to the CCA-validation fields shown in Figure 8. The close correspondence between the two figures further attests to the similarities between the results derived from both methods. Summary statistics for the RegEM-Ridge field correlations, standard deviation ratios, mean biases, and RMSE are given in Table 4. The mean field correlations associated with the two methods are very similar, yet indicate RegEM-Ridge to have slightly more correlation skill at increased noise levels. The RegEM-Ridge mean biases also have spatial patterns very similar to CCA, but indicate that RegEM-Ridge produces larger biases at increased noise levels. The most notable difference between the two methods is associated with their standard deviation ratios. RegEM-Ridge standard deviation ratios have patterns similar to the CCA reconstructions and also maintain the most variance where the field correlations

are highest. The variance loss in RegEM-Ridge, however, is much more pronounced than in the CCA reconstructions: mean standard deviation ratios are only 62% of those achieved for the CCA reconstructions at a SNR of infinity and fall to almost 40% of the CCA counterpart at a SNR of 0.25. These variance losses are manifest in the higher RMSE values associated with the RegEM-Ridge fields, but result in only modest increases in the mean field errors (Table 4) relative to CCA. Two factors contribute to the similar RMSE fields in spite of the larger variance losses in the RegEM-Ridge CFRs: (1) the mean biases dominate the error fields, which are not significantly different in the reconstructions from the two methods; and (2) the slightly higher correlations associated with the RegEM-Ridge reconstructions offset the errors associated with variance losses.

5. Discussion

We have evaluated the performance of a widely used and traditional multivariate regression method, CCA, for the purpose of reconstructing climate fields over the last millennium. The advantages of the method in such a context reside in the interpretability of the reconstructed patterns and the relative efficiency of the algorithm. The method also operationalizes the three principal choices of a reconstruction method: the degree of spatial reduction in the target and proxy fields and the regularization of the least-squares regression. Derived skill statistics can be readily interpreted for their dependence on these three choices and therefore provide a measure of the uncertainties associated with the selection of the statistical model. Characterization of this uncertainty is further aided by the computational efficiency of the CCA method by allowing collections of reconstructions to be calculated. In the present work we have generated reconstructions from all possible combinations of the CCA dimensions between 1 and 50 modes, which has provided robust estimates of the error dependence on these selections. Further work will similarly allow calculations of reconstruction ensembles generated from multiple noise realizations in the proxy or target fields. The method of the dimensional selection is of course the biggest challenge for deriving robust predictions of past climate fields while also guarding against the likelihood of artificial skill. We have provided a heuristic means of selecting these dimensions, but alternative approaches undoubtedly exist. Further application of the CCA method may therefore dictate improved means of dimensional selection.

We have also compared the CCA results to those derived using RegEM-Ridge. Our results show that the methods produce very similar results, with the exception of the larger variance losses observed

in the RegEM-Ridge reconstructions. The smaller variance losses and mean biases associated with the CCA method might warrant its use over RegEM-Ridge, but the differences are minor as measured in the total RMSE. The source of variance losses is likely associated with the manner in which the eigenvalue spectra are truncated in the two methods. Ridge regression filters the eigenvalue spectrum using a continuous filter function, i.e. there is no abrupt eigenvalue truncation like that applied in CCA where modes that cannot be reliably calibrated are simply set to zero. This was indeed one reason why RegEM-Ridge was originally proposed as a potentially advantageous method in CFR contexts (Schneider 2001). A consequence of the continuous filtering function, however, is the fact that leading modes may be overly dampened if only a small number of them carry a large percentage of the total variance, as in the case of the CFR application presently considered. As a demonstration of this fact, Figure 12 plots the eigenspectra for the true model field and for the RegEM-Ridge and CCA reconstruction fields at SNR levels of 1.0 and 0.5. The figure demonstrates that the scaling of the lower-order modes is similar for the two reconstruction methods except for the reduced magnitudes of the RegEM-Ridge eigenvalues and the abrupt transitions to zero in the CCA spectra. The similarity of the spectra in the lower-order modes, and the similarity of the correlation fields in Figures 8 and 11, indicates that the two methods are likely reconstructing similar patterns, and differ primarily by the dampened variability of the leading modes in the RegEM-Ridge spectrum. By contrast, the finite truncation of the CCA method yields leading modes that are unaffected by the truncation. The result is effectively very similar reconstructions that differ primarily by the amount of variance preserved.

The above discussion raises the important and yet-to-be-explained difference between pseudoproxy CFRs derived using RegEM-Ridge and RegEM-TTLS. This latter method has been shown to perform well in one pseudoproxy context, particularly in terms of its ability to reproduce the NH mean index (Mann et al. 2007a), while the former has not. The original explanation for the differences between the performance of RegEM-Ridge and RegEM-TTLS was tied to the selection of the ridge parameter by means of generalized cross validation (GCV) in RegEM-Ridge (Mann et al. 2007a,c). Because GCV was not used within RegEM-TTLS, Mann et al. (2007a,c) concluded that the problem was specific to RegEM-Ridge. Smerdon et al. (2008a), however, demonstrated that the mean biases and variance losses in RegEM-Ridge were not associated with the GCV selection of the ridge parameter, leaving the differences between the two regularization procedures unexplained. Prior to the CCA results presented herein, one could have argued that the different approaches to eigenvalue truncation in the two RegEM

regularization methods still might explain their differences. This argument is unlikely, however, in light of the fact that CCA and RegEM-TTLS both truncate the eigenvalue spectrum and CCA still suffers from warm biases and variance losses. Moreover, the similar shortcomings observed for the Mann et al. (1998) CFR method noted by von Storch et al. (2004, 2006) supports the idea that the effects cannot be connected to something specific in RegEM-Ridge. It therefore is unlikely that differences in the reported performance of multiple CFRs can be specifically associated with the method of eigenvalue truncation or filtration, pointing to the need for alternative explanations.

Comparisons of the field performances of the different methods will ultimately help explain some of the above inconsistencies. While a complete investigation of the differences between the two RegEM methods is outside the scope of this paper, some preliminary observations are possible. Mann et al. (2007a) provide validation statistics for RegEM-TTLS using the same calibration (1856-1980 C.E.) and reconstruction (850-1855 C.E.) intervals used herein, and include the mean r^2 values for the NH field (labeled Multivariate r^2 in Table 1 of Mann et al. (2007a)). Their experiments d, e, f, and h correspond to the same white-noise pseudoproxy experiments performed herein for SNRs of infinity, 1.0, 0.5 and 0.25, respectively (although these pseudoproxies involve different noise realizations). The principal difference between these experiments is that Mann et al. (2007a) performed hybrid reconstructions that calibrate separately in high- and low-frequency domains (split at the 20-year period), before combining the two reconstructed domains in a final CFR; the authors report there to be little difference between hybrid and non-hybrid results.

The r^2 values reported in Mann et al. (2007a) for RegEM-TTLS are 0.30, 0.23, 0.19 and 0.06, for SNRs of infinity, 1.0, 0.5 and 0.25, respectively. These values are equivalently 0.51, 0.36, 0.20, and 0.05 for CCA, and 0.48, 0.37, 0.23, and 0.07 for RegEM-Ridge. Except for the highest noise level, for which all methods perform similarly poorly (and are likely within uncertainties imposed by different pseudoproxy noise realizations), these validation statistics indicate that CCA and RegEM-Ridge produce CFRs with more field skill than RegEM-TTLS. Mann et al. (2007a) also provide the r^2 values between the target and reconstructed NH means: 0.87, 0.86, 0.83, and 0.34 for SNRs of infinity, 1.0, 0.5 and 0.25, respectively, as compared to 0.86, 0.74, 0.52 and 0.17 for CCA and 0.83, 0.73, 0.55, and 0.24 for RegEM-Ridge. RegEM-TTLS thus appears to produce more skillful mean NH time series than CCA and RegEM-Ridge, whereas the latter two methods produce more skillful field reconstructions. The potential differences between hybrid and non-hybrid methods notwithstanding, it

is possible that the reduced field skill associated with the RegEM-TTLS reconstructions is the result of targeting only a few leading modes in the field, i.e. those tied most strongly to the field mean. Such an application of RegEM-TTLS is indeed discussed and employed in Mann et al. (2008), who target specifically a NH index reconstruction. An approach that exclusively targets the NH mean could therefore yield skillful NH mean results, but provide little field skill. Similarly, one would expect the field skills to be comparable at high noise levels when all methods are effectively targeting only a few leading modes, which is indeed observed in the above comparisons.

It is also important to highlight the observed concentration of the highest field correlations (and preserved variance) in those areas with high pseudoproxy concentrations, a feature of both the CCA and RegEM-Ridge CFRs. Although this result may seem intuitive, it is not necessarily an expected characteristic of either the CCA or RegEM-Ridge methods. Both of these techniques attempt to reconstruct large-scale climate patterns by discarding smaller-scale modes of variability and noise. In spite of this emphasis on large-scale patterns, the observed correlation distributions suggest that the methods perform best where dense sampling exists, indicating that low-noise proxies outside of the highly sampled regions would have the most potential to improve CFR field skill. It is important, however, to understand better the origin of the observed skill concentrations and their dependence on the underlying character of the target field. In the case of the reported pseudoproxy experiments, the skill patterns are dependent on the internal statistics of the model-simulated climate. Integrations from two different GCMs were used by von Storch et al. (2004, 2006) to test the Mann et al. (1998) method and results where consistent across the simulations in terms of the NH means. The authors also reported no significant dependence on the sampling distribution. Similarly, Mann et al. (2007a) indicated no significant sensitivity to the two GCM integrations or sampling distribution used to test RegEM-TTLS. Christiansen et al. (2009) use yet another model integration and method for generating ensemble statistics and observed mean biases and variance losses in NH means derived from multiple methods. It therefore is unlikely that differences in model integrations will affect the gross performance of reconstruction methods already reported. Nevertheless, the underlying field performance of CFRs is likely more sensitive to the spatial statistics of the model simulations and should be tested on multiple model integrations. More experiments using observational data (e.g. Evans et al. 2001, 2002) are also needed in order to determine whether the skill patterns of pseudoproxy experiments are similar to those estimated from real-world datasets.

The above discussion leads to one final point of discussion regarding the strengths and weakness of pseudoproxy experiments. Pseudoproxy research has proceeded under the assumption that modeled climates are a good approximation of real-world field characteristics. This assumption may require the most caution, however, when interpreting results dependent on the underlying spatial statistics of the field and the associated teleconnections. Furthermore, noise structures in real-world proxies are undoubtedly more complicated than the white noise models used in this study. While it is appropriate to approach the results contained herein as a best-case scenario, further work is necessary to more faithfully capture the nonlinear, multivariate and nonstationary noise characteristics that are likely present in many proxy series (e.g. North et al. 2006). For instance, tree-ring models have been developed to simulate dendroclimatic series and have been applied in several studies with notable success (Evans et al. 2006; Anchukaitis et al. 2006). Such work will ultimately help to incorporate more representative noise structures into pseudoproxy series and therefore provide more realistic evaluations of CFR methods.

6. Conclusions

Successful application of the CCA method to the problem of reconstructing NH temperature fields during the last millennium has been demonstrated and evaluated using pseudoproxies. An element of this application involved the development of a selection procedure for the three CCA dimensions. We have demonstrated a "leave-half-out" cross-validation procedure that selects a robust and parsimonious model selection while guarding against artificial skill in the reconstruction. Our experiments demonstrate that the CCA method faithfully reconstructs between 3 and 18 climatic patterns given a proxy distribution approximating the Mann et al. (1998) proxy network and a range of observational uncertainties from no noise to an SNR of 0.25 (the exact number of resolved patterns will of course vary with different noise realizations at a given SNR value and is idealized in the pseudoproxy framework). Subsequent application of the CCA method to real-world climate proxies is thus easily attainable in future work. The transparency of the CCA method and its well-developed theoretical basis in the literature is a strong motivation for its application. These characteristics provide straightforward evaluations of the CCA model selection and the source of skill in derived reconstructions. The results of our pseudoproxy experiments, however, suggest that CFRs derived using CCA, just like those derived from RegEM-Ridge, should be interpreted carefully when applied to the problem of reconstructing large-

scale climate patters during the Common Era. We note in particular that CCA CFRs have the potential to suffer from significant mean biases and variance losses across a range of noise levels spanning those of real-world proxies.

Field correlations were also shown to diminish significantly with increasing noise, particularly in regions with few or no pseudoproxies. Given that SNRs in real proxy records are estimated to be on the order of 0.4 (e.g. Mann et al. 2007a) and typically characterized by more complicated autoregressive and moving average structures than the white-noise models adopted herein, the observed skill reductions should be considered a best-case scenario. In real-world CFRs derived with CCA, the spatial patterns of the above characteristics would likely depend on at least five factors: (1) the spatial distribution of the proxies; (2) the magnitude and character of noise in the proxy network; (3) the spatial coherence of the target field, i.e. the strength and character of its teleconnections; (4) the true historical variability of the climate during the reconstruction interval; and (5) the length of the calibration period used for estimating proxy-climate correlations. The dependence of the spatial skill associated with the CCA method to these factors requires further testing using additional model-generated and observational fields. Nevertheless, our results are robust within the CSM model integration and are supported by their close correspondence with reconstructions derived using the RegEM-Ridge method.

Comparisons between reconstructions derived from CCA and RegEM-Ridge demonstrate strong similarities between the two methods, both in terms of the derived mean NH temperatures and the spatial characteristics of the reconstructed fields. It is therefore difficult to argue for the advantages of RegEM-Ridge over the more traditional CCA application in light of the similar skill performance observed for both methods and the higher conceptual complexity and computational costs of RegEM-Ridge. Furthermore, these similarities are encouraging regarding the consistency of the two linear methods, but are also an indication that there may be problems endemic to the present generation of CFR methods used to reconstruct large-scale temperature patterns during the last millennium. More research therefore is needed to characterize the performance of multiple CFR methods in terms of their field performance and to draw distinct conclusions about the similarities and differences across a range of methods. These results are particularly needed in the context of CFRs derived from real-world proxies as a means of deriving a better description of the uncertainties in these results.

The good agreement between the CCA and RegEM-Ridge results further points to the need to understand the differences in the performance of the RegEM-Ridge and RegEM-TTLS methods. Initial

779 comparisons explored herein indicate that RegEM-TTLS may produce more skillful NH mean indices,
780 while yielding CFRs that are less skillful than those produced by either CCA or RegEM-Ridge. Resolv-
781 ing the origin of these differences is not only important for studies that have attempted to reconstruct
782 temperatures over the last millennium (Rutherford et al. 2005, Mann et al. 2005, 2007a, 2008), but also
783 for efforts that have applied RegEM in other contexts (e.g. Zhang et al. 2004; Steig et al. 2009). This
784 necessity is further supported by the fact that pseudoproxy experiments have demonstrated differences
785 between the performance of the two RegEM regularization approaches, while real-world reconstruc-
786 tions of millennial temperatures have not been notably different - at least in their representation of the
787 NH mean (Mann et al. 2007a). Each of these observations indicates that the focus within the literature
788 on only NH means is insufficient for evaluating CFR methods and their derived results. Furthermore,
789 explaining the performance differences between various CFR methods remains an open research ques-
790 tion, but the persistence of similar problems in now multiple independent linear reconstruction methods
791 suggests that caution must be exercised in the interpretation of published real-world CFR results.
792
793 **Acknowledgments**
794 This research was supported in part by the National Science Foundation by grant ATM04-07909 to AK
795 and by the National Oceanic and Atmospheric Administration, U.S. Department of Commerce, by grant
796 NA07OAR4310060 to JES, AK and MNE and by grant NAOAR4320912 to JES and AK under the
797 Cooperative Institute for Climate Applications Research (CICAR). Part of this research was completed
798 while DC was supported by a research internship from the Hughes Science Pipeline Project and JES
799 was supported by a Mellon Postdoctoral Fellowship, both through the Department of Environmental
800 Science at Barnard College. The statements, findings, conclusions, and recommendations are those of
801 the authors and do not necessarily reflect the views of the any of the above organizations or agencies.
802 LDEO contribution XXXX.
803
804 **References**
805 Ammann, C. M., F. Joos, D. S. Schimel, B. L. Otto-Bliesner, and R. A. Tomas, 2007: Solar influence
806 on climate during the past millennium: Results from transient simulations with the NCAR Climate
807 System Model. *Proc. Nat. Acad. Sci. USA*, **104**, 3713-3718, doi:10.1073-pnas.0605064.103.
808

809 Anchukaitis, K. J., M. N. Evans, A. Kaplan, E. A. Vaganov, M. K. Hughes, H. D. Grissino-Mayer, and
810 M. A. Cane, 2006: Forward modeling of regional scale tree-ring patterns in the southeastern United
811 States and the recent influence of summer drought. *Geophys. Res. Lett.*, **33**, L04705, doi:10.1029/2005GL025050.
812
813 Anderson, T.W., 1984: *An Introduction to Multivariate Statistical Analysis*. 2nd ed. Wiley-Interscience,
814 704 pp.
815
816 Barnett, T., and R. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for
817 United States surface air temperature determined by canonical correlation analysis. *Monthly Weather*
818 *Review*, **115**, 1825-1850.
819
820 Bretherton, C., C. Smith, and J.M. Wallace, 1992: An intercomparison of methods for finding coupled
821 patterns in climate data. *J. Climate*, **5**, 541-560.
822
823 Beltrami, H., 2002: Climate from borehole data: Energy fluxes and temperatures since 1500. *Geophys.*
824 *Res. Lett.*, **29**, doi:10.1029/2002GL015702.
825
826 Briffa, K. R., 2000: Annual climate variability in the Holocene: interpreting the message of ancient
827 trees. *Quat. Sci. Rev.*, **19**, 87-105.
828
829 Briffa, K. R., and T. J. Osborn, 2002: Blowing hot and cold. *Science*, **295**, 2227-2228.
830
831 Briffa, K. R., T.J. Osborn, F.H. Schweingruber, I.C. Harris, P.D. Jones, S.G. Shiyatov, S.G. and E.A.
832 Vagano, 2001: Low-frequency temperature variations from a northern tree ring density network. *J.*
833 *Geophys. Res.*, **106**, 2929-2941.
834
835 Broecker, W.S., 2001: Paleoclimate - Was the medieval warm period global? *Science*, **291**, 5508,
836 1497-1499.
837
838 Bürger, G., and U. Cubasch, 2005: Are multiproxy climate reconstructions robust? *Geophys. Res.*

Lett., **32**, L23711, doi:10.1029/2005GL024155.

Bürger, G., I. Fast, and U. Cubasch, 2006: Climate reconstruction by regression. *Tellus*, Ser. A, **58**, 227235.

Casty, C., D. Handorf, and M. Sempf, 2005: Combined winter climate regimes over the North Atlantic/European sector 1766-2000. *Geophys. Res. Lett.*, **32**, L13801, doi:10.1029/2005GL022431.

Christiansen, B. T. Schmith, and P. Thejll, 2009: A surrogate ensemble study of climate reconstruction methods: Stochasticity and robustness, *J. Climate*, **22**(4), 951-976, DOI:10.1175/2008JCLI2301.1.

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

Cook, E., J. Esper, and R. D'Arrigo, 2004: Extra-tropical Northern Hemisphere land temperature variability over the past 1000 years. *Quat. Sci. Rev.*, **23**, 2063-2074.

Crowley, T. J., and T. S. Lowery, 2000: How warm was the Medieval Warm Period? *Ambio*, **29**, 5154.

D'Arrigo, R., R. Wilson, and G. Jacoby, 2006: On the long-term context for late twentieth century warming. *J. Geophys. Res.*, **111**, D03103, doi:10.1029/2005JD006352.

D'Arrigo, R., R. Wilson, and A. Tudhope, 2009: The impact of volcanic forcing on tropical temperatures during the past four centuries. *Nature Geosci.*, **2**, 51-56.

Esper, J., E. R. Cook, and F. H. Schweingruber, 2002: Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. *Science*, **295**, 2250-2253.

Evans M.N., A. Kaplan, M.A. Cane, and R. Villalba, 2001: Globality and optimality in climate field reconstructions from proxy data, in V. Markgraf (ed.) *Present and Past Inter-Hemispheric Linkages in*

the Americas and Their Societal Effects, Cambridge University Press, p. 53-72.

Evans, M. N., A. Kaplan, and M. A. Cane, 2002: Pacific sea surface temperature field reconstruction from coral δ^{18}O data using reduced space objective analysis. *Paleoceanography*, **17**(1), 1007, doi:10.1029/2000PA000590.

Evans, M.N., B.K. Reichert, A. Kaplan, K.J. Anchukaitis, E.A. Vaganov, M.K. Hughes and M.A. Cane, 2006: A forward modeling approach to paleoclimatic interpretation of tree-ring data. *J. Geophys. Res.*, **111**, G03008, doi:10.1029/2006JG000166

Folland, C.K., T.R. Karl, J.R. Christy, R.A. Clarke, G.V. Gruza, J. Jouzel, M.E. Mann, J. Oerlemans, M.J. Salinger, and S.-W. Wang, 2001: Observed Climate Variability and Change. pp. 99-181 In: *Climate Change 2001: The Scientific Basis*. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van der Linden, P.J., Dai, X., Maskell, K. and Johnson, C.A. Eds.). Cambridge University Press, Cambridge, UK, 881 pp.

Fritts, H. C., T. J. Blasing, B. P. Hayden, and J. E. Kutzbach, 1971: Multivariate techniques for specifying tree-growth and climate relationships and for reconstructing anomalies in paleoclimate. *J. Appl. Meteorol.*, **10**, 845-864

González-Rouco, F., H. von Storch, and E. Zorita, 2003: Deep soil temperature as proxy for surface air-temperature in a coupled model simulation of the last thousand years. *Geophys. Res. Lett.*, **30**(21), 2116, doi:10.1029/2003GL018264.

González-Rouco, J. F., H. Beltrami, E. Zorita, and H. von Storch, 2006: Simulation and inversion of borehole temperature profiles in surrogate climates: Spatial distribution and surface coupling. *Geophys. Res. Lett.*, **33**, L01,703,doi:10.1029/2005GL024,693.

Groveman, B. S., and H. E. Landsberg, 1979: Simulated northern hemisphere temperature departures

1579-1880. *Geophys. Res. Lett.* , **6**, 767-770.

Golub, G. H. and C. F. Van Loan, 1996: *Matrix Computations*, Johns Hopkins University Press, 3rd Ed.

Harris, R.N., and D.S. Chapman, 2001: Mid-latitude ($30°N - 60°N$) climatic warming inferred by combining borehole temperatures with surface air temperatures. *Geophys. Res. Lett.*, **28**, 747-750.

Hegerl, G. C., T. Crowley, M. Allen, W. T. Hyde, H. Pollack, J. Smerdon, and E. Zorita, 2007: Detection of human influence on a new 1500 yr climate reconstruction. *J. Climate*, **20**, 650-666, DOI:10.1175/JCLI4011.1.

Huang, S., H.N. Pollack, and P.Y. Shen, 2000: Temperature trends over the last five centuries reconstructed from borehole temperatures. *Nature*, **403**, 756-758.

Hughes, M.K. and H.F. Diaz, 1994: Was there a medieval warm period, and if so, where and when? *Clim. Ch.*, **26**, 2-3, 109-142.

Huybers, P., 2005: Comment on "Hockey sticks, principal components, and spurious significance" by S. McIntyre and R. McKitrick. *Geophys. Res. Lett.*, **32**, L20705, doi:10.1029/2005GL023395.

Jansen, E., J. Overpeck, K.R. Briffa, J.-C. Duplessy, F. Joos, V. Masson-Delmotte, D. Olago, B. Otto-Bliesner, W.R. Peltier, S. Rahmstorf, R. Ramesh, D. Raynaud, D. Rind, O. Solomina, R. Villalba and D. Zhang, 2007: Palaeoclimate. In: *Climate Change 2007: The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jones, P.D., K.R. Briffa, T.J. Osborn, J.M. Lough, T.D. van Ommen, B.M. Vinther, J. Luterbacher, E.R. Wahl, F.W. Zwiers, M.E. Mann, G.A. Schmidt, C.M. Ammann, B.M. Buckley, K.M. Cobb, J. Esper,

H. Goosse, N. Graham, E. Jansen, T. Kiefer, C. Kull, M. Küttel, E. Mosely-Thompson, J.T. Overpeck, N. Riedwyl, M. Schulz, A.W. Tudhope, R. Villalba, H. Wanner, E. Wolff, E. Xoplaki, 2009: High-resolution paleoclimatology of the last millennium: a review of current status and future prospects, *Holocene*, **19**, 3-49.

Jones, P. D., and M. E. Mann, 2004: Climate over past millennia. *Rev. Geophys.*, **42**, RG2002, doi:10.1029/2003RG000143.

Jones, P. D., M. New, D. E. Parker, S. Martin, and J. G. Rigor, 1999: Surface Air Temperature and its Changes over the Past 150 Years. *Rev. Geophys.*, **37**, 173-199.

Jones P.D., K.R. Briffa, T.P. Barnett, and S.F.B. Tett, 1998: High-resolution palaeoclimatic records for the last millennium: Interpretation, integration and comparison with general circulation model control-run temperatures. *Holocene*, **8**, 455-471.

Küttel, M., J. Luterbacher, E. Zorita, E. Xoplaki, N. Riedwyl, and H. Wanner, 2007: Testing a European winter surface temperature reconstruction in a surrogate climate. *Geophys. Res. Lett.*, **34**, L07710, doi:10.1029/2006GL027907.

Lee, T.C.K, F.W. Zwiers, and M. Tsao, 2007: Evaluation of proxy-based millennial reconstruction methods. *Clim. Dyn.*, DOI 10.1007/s00382-007-0351-9

Luterbacher, J., C. Schmutz, D. Gyalistras, E. Xoplaki, and H. Wanner, 1999: Reconstruction of monthly NAO and EU indices back to AD 1675. *Geophys. Res. Lett.*, **26**, 2745-2748.

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**, 149-1503.

Luterbacher, J., R. Rickli, C. Tinguely, E. Xoplaki, E. Schüpbach, D. Dietrich, J. Hsler, M. Ambhl, C. Pfister, P. Beeli, U. Dietrich, A. Dannecker, T. D. Davies, P. D. Jones, V. Slonosky, A. E. J. Ogilvie, P.

959 Maheras, F. Kolyva-Machera, J. Martin-Vide, M. Barriendos, M. J. Alcoforado, F. Nunez, T. Jónsson,
960 R. Glaser, J. Jacobeit, C. Beck, A. Philipp, U. Beyer, E. Kaas, T. Schmith, L. Bärring, P. Jönsson, L.
961 Racz, L. and H. Wanner, 2000: Reconstruction of monthly mean sea level pressure over Europe for the
962 Late Maunder Minimum period (1675-1715). *Int. J. Climatol.*, **20**, 1049-1066.
963
964 Mann, M.E., Ammann, C.M., Bradley, R.S., Briffa, K.R., Crowley, T.J., Hughes, M.K., Jones, P.D.,
965 Oppenheimer, M., Osborn, T.J., Overpeck, J.T., Rutherford, S., Trenberth, K.E., Wigley, T.M.L., 2003:
966 On past temperatures and anomalous late 20th-century warmth. *Eos*, **84**, 256-258.
967
968 Mann, M.E., 2002: The Value of Multiple Proxies. *Science*, **297**, 1481-1482.
969
970 Mann, M. E., and S. Rutherford, 2002: Climate reconstruction using pseudoproxies. *Geophys. Res.*
971 *Lett.*, **29**(10), 1501, doi:10.1029/2001GL014554.
972
973 Mann, M. E., R. S. Bradley, and M. K. Hughes, 1998: Global-scale temperature patterns and climate
974 forcing over the past six centuries. *Nature*, **392**, 779-787.
975
976 Mann, M. E., R. S. Bradley, and M. K. Hughes, 1999: Northern Hemisphere temperatures during the
977 past millennium: Inferences, uncertainties, and limitations. *Geophys. Res. Lett.*, **26**, 759-762.
978
979 Mann, M. E., S. Rutherford, E. Wahl, and C. Ammann, 2005: Testing the fidelity of methods used in
980 proxy-based reconstructions of past climate. *J. Climate*, **18**, 4097-4107.
981
982 Mann, M. E., S. Rutherford, E. Wahl, and C. Ammann, 2007a: Robustness of proxy-based climate
983 field reconstruction methods. *J. Geophys. Res.*, **112**, D12109, doi:10.1029/2006JD008272.
984 Mann, M.E., Rutherford, S., Wahl, E., Ammann, C., 2007b: Reply to Comments on "Testing the
985 Fidelity of Methods Used on Proxy-based Reconstructions of Past Climate" by Zorita et al. *J. Climate*,
986 **20**, 3699-3703.
987
988 Mann, M.E., Rutherford, S., Wahl, E., Ammann, C., 2007c: Reply to Comments on "Testing the

989 Fidelity of Methods Used on Proxy-based Reconstructions of Past Climate" by Smerdon and Kaplan.
990 *J. Climate*, **20**, 5671-5674.
991 Mann, M. E., Z. Zhang, M. K. Hughes, R. S. Bradley, S. K. Miller, S. Rutherford, and F. Ni, 2008:
992 Proxy-based reconstructions of hemispheric and global surface temperature variations over the past
993 two millennia. *Proc. Nat. Acad. Sci. USA*, **105**, 36, 13252-13257.
994
995 McIntyre, S., and R. McKitrick, 2005: Hockey sticks, principal components, and spurious significance.
996 *Geophys. Res. Lett.*, **32**, L03710, doi:10.1029/2004GL021750.
997
998 Moberg, A., R. Mohammad, and T. Mauritsen, 2008: Analysis of the Moberg et al. (2005) hemispheric
999 temperature reconstruction. *Clim. Dyn.*, **31**, 957-971, doi: 10.1007/s00382-008-0392-8.
1000
1001 Moberg, A, D.M. Sonechkin, K. Holmgren, N.M Datsenko, W. Karlen, 2005: Highly variable Northern
1002 Hemisphere temperature reconstructed from low- and high-resolution proxy data. *Nature*, **433**, 7026,
1003 613-617.
1004
1005 North, G.R. et al., 2006: *Surface Temperature Reconstructions for the Last 2,000 Years*, The National
1006 Academies Press, Washington D.C., 196 pp.
1007
1008 Pauling, A., J. Luterbacher, C. Casty, and H. Wanner, 2006: 500 years of gridded high-resolution
1009 precipitation reconstructions over Europe and the connection to large-scale circulation. *Clim. Dyn.*,
1010 **26**, 387-405.
1011
1012 Pollack, H. N., and J. E. Smerdon, 2004: Borehole climate reconstructions: Spatial structure and
1013 hemispheric averages. *J. Geophys. Res.*, **109**, D11106, doi:10.1029/2003JD004163.
1014
1015 Preisendorfer, R., F. Zwiers and T.P. Barnett, 1981: Foundations of principal component selection
1016 rules, SIO Ref. Ser. 81-4.
1017
1018 Riedwyl, N., M. Küttel, J. Luterbacher, H. and Wanner, 2009: Comparison of climate field reconstruc-

tion techniques: Application to Europe. *Clim. Dynam.*, **32**, 381-395

Rutherford, S., and M. E. Mann, 2004: Correction to Optimal surface temperature reconstructions using terrestrial borehole data by Mann et al. *J. Geophys. Res.*, **109**, D11107, doi:10.1029/2003JD004290.

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

Rutherford, S., Mann, M.E., Wahl, E., Ammann, C., 2008: Reply to comment by Jason E. Smerdon et al. on "Robustness of proxy-based climate field reconstruction methods." *J. Geophys. Res.*, **113**, D18107, doi:10.1029/2008JD009964.

Schneider, T., 2001: Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *J. Climate*, **14**, 853-887.

Smerdon, J.E., 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**, 22, 5666-5670.

Smerdon, J.E., A. Kaplan, and D. Chang 2008a: On the standardization sensitivity of RegEM climate field reconstructions, *J. Climate*, **21**(24), 6710-6723.

Smerdon, J.E., J.F. González-Rouco, and E. Zorita, 2008b: Comment on "Robustness of proxy-based climate field reconstruction methods" by Michael E. Mann et al. *J. Geophys. Res.*, **113**, D18106, doi:10.1029/2007JD009542.

Steig, E.J., D.P. Schneider, S.D. Rutherford, M.E. Mann, J.C. Comiso, D.T. Shindell, 2009: Warming of the Antarctic ice sheet surface since the 1957 International Geophysical Year. *Nature*, **1457**, 459-463.

Tippett, M. K., M. Barlow, and B. Lyon, 2003: Statistical correction of Central Southwest Asia winter precipitation simulations. *Int. J. Climatol.*, **23**, 1421-1433.

Tippett, M. K., T. DelSole, S. J. Mason, and A. G. Barnston, 2008: Regression-Based Methods for Finding Coupled Patterns. *J. Climate*, **21**(17), 4384-4398.

von Storch, H., and F.W. Zwiers, 2000: *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, United Kingdom, 484 pp.

von Storch, H., and E. Zorita, 2005: Comment on "Hockey sticks, principal components, and spurious significance" by S. McIntyre and R. McKitrick. *Geophys. Res. Lett.*, **32**, L20701, doi:10.1029/2005GL022753.

von Storch, H., E. Zorita, J. M. Jones, Y. Dimitriev, F. González-Rouco, and S. F. B. Tett, 2004: Reconstructing past climate from noisy data. *Science*, **306**, 679-682.

von Storch, H., E. Zorita, J. M. Jones, F. González-Rouco, and S. F. B. Tett, 2006: Response to comment on "Reconstructing past climate from noisy data." *Science*, **312**, 529c.

Wahl, E. R., and C. M. Ammann, 2007: Robustness of the Mann, Bradley, Hughes reconstruction of surface temperatures: Examination of criticisms based on the nature and processing of proxy climate evidence. *Clim. Ch.*, **85**, 1-2, 33-69.

Wahl, E. R., D. M. Ritson, and C. M. Ammann, 2006: Comment on "Reconstructing past climate from noisy data." *Science*, **312**, 529b.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press: New York., 467 pp.

Xoplaki, E., J. Luterbacher, H. Paeth, D. Dietrich, N. Steiner, M. Grosjean, and H. Wanner (2005), Eu-

ropean spring and autumn temperature variability and change of extremes over the last half millennium. *Geophys. Res. Lett.*, **32**, L15713, doi:10.1029/2005GL023424.

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

Zorita, E., J. F. González-Rouco, and H. von Storch, 2007: Comment on "Testing the fidelity of methods used in proxy-based reconstructions of past climate" by Mann et al. *J. Climate*, **20**, 14, 2693-3698.

	Absolute minimum RMSE					Absolute maximum NHMC				
SNR	RMSE	d_{cca}	d_p	d_t	NHMC	RMSE	d_{cca}	d_p	d_t	NHMC
Inf.	0.4838	21	26	50	0.8178	0.5157	21	41	35	0.8697
1.0	0.5608	24	27	45	0.7207	0.5670	20	27	35	0.7356
0.5	0.6545	15	25	47	0.6275	0.6744	25	34	32	0.6905
0.25	0.7077	3	44	4	0.2799	0.7294	4	12	44	0.4376

Table 1: CCA reconstruction statistics using the absolute minimum RMSE or maximum NHMC criteria during the calibration interval (1856-1980 C.E.).

Preferred Solutions

SNR	RMSE	d_{cca}	d_p	d_t	NHMC
Inf.	0.4839	18	25	36	0.8378
1.0	0.5636	13	27	21	0.7031
0.5	0.6562	7	28	12	0.6162
0.25	0.7077	3	44	4	0.2799

Table 2: CCA reconstruction statistics for the preferred solutions in which parsimonious dimensional combinations have been chosen as the first local minimum of the RMSE statistic.

	SNR Infinity		SNR 1.0		SNR 0.5		SNR 0.25	
CCA Rank	Cal.	Recon.	Cal.	Recon.	Cal.	Recon.	Cal.	Recon.
1	0.999	0.991	0.975	0.916	0.920	0.734	0.876	0.397
2	0.997	0.990	0.969	0.875	0.895	0.653	0.792	0.454
3	0.996	0.987	0.958	0.853	0.876	0.528	0.719	0.252
4	0.994	0.969	0.936	0.839	0.817	0.700	—	—
5	0.989	0.965	0.928	0.810	0.785	0.671	—	—
6	0.984	0.950	0.905	0.671	0.694	0.252	—	—
7	0.983	0.951	0.869	0.652	0.658	0.333	—	—
8	0.975	0.899	0.820	0.638	—	—	—	—
9	0.971	0.916	0.803	0.553	—	—	—	—
10	0.964	0.889	0.764	0.422	—	—	—	—
11	0.956	0.861	0.757	0.586	—	—	—	—
12	0.950	0.875	0.625	0.364	—	—	—	—
13	0.913	0.700	0.592	0.279	—	—	—	—
14	0.912	0.815	—	—	—	—	—	—
15	0.887	0.752	—	—	—	—	—	—
16	0.879	0.800	—	—	—	—	—	—
17	0.820	0.548	—	—	—	—	—	—
18	0.747	0.644	—	—	—	—	—	—

Table 3: Correlation statistics between the true canonical temperature time series, Q_t, and those predicted by the proxy PCs, i.e. $\Sigma_{cca} Q_p^T$. Statistics are shown for both the reconstruction and calibration intervals.

CCA

SNR	NHMC	Mean of Field Correlations	Mean STD Ratio	Mean Bias (K)	Local RMSE Mean (K)
Inf.	0.93	0.69	0.76	0.04	0.37
1.0	0.86	0.58	0.58	0.09	0.45
0.5	0.72	0.43	0.44	0.17	0.53
0.25	0.41	0.22	0.37	0.23	0.61

RegEM-Ridge

SNR	NHMC	Mean of Field Correlations	Mean STD Ratio	Mean Bias (K)	Local RMSE Mean (K)
Inf.	0.91	0.68	0.47	0.01	0.42
1.0	0.86	0.60	0.37	0.11	0.47
0.5	0.74	0.46	0.23	0.20	0.55
0.25	0.49	0.24	0.15	0.26	0.61

Table 4: Validation statistics computed during the reconstruction interval (850-1855 C.E.) for the CCA and RegEM-Ridge reconstructions. Reconstructions from each method were derived with the same set of pseudoproxies at all noise levels. All field statistics were weighted by the cosine of the mid-latitude for each grid cell.

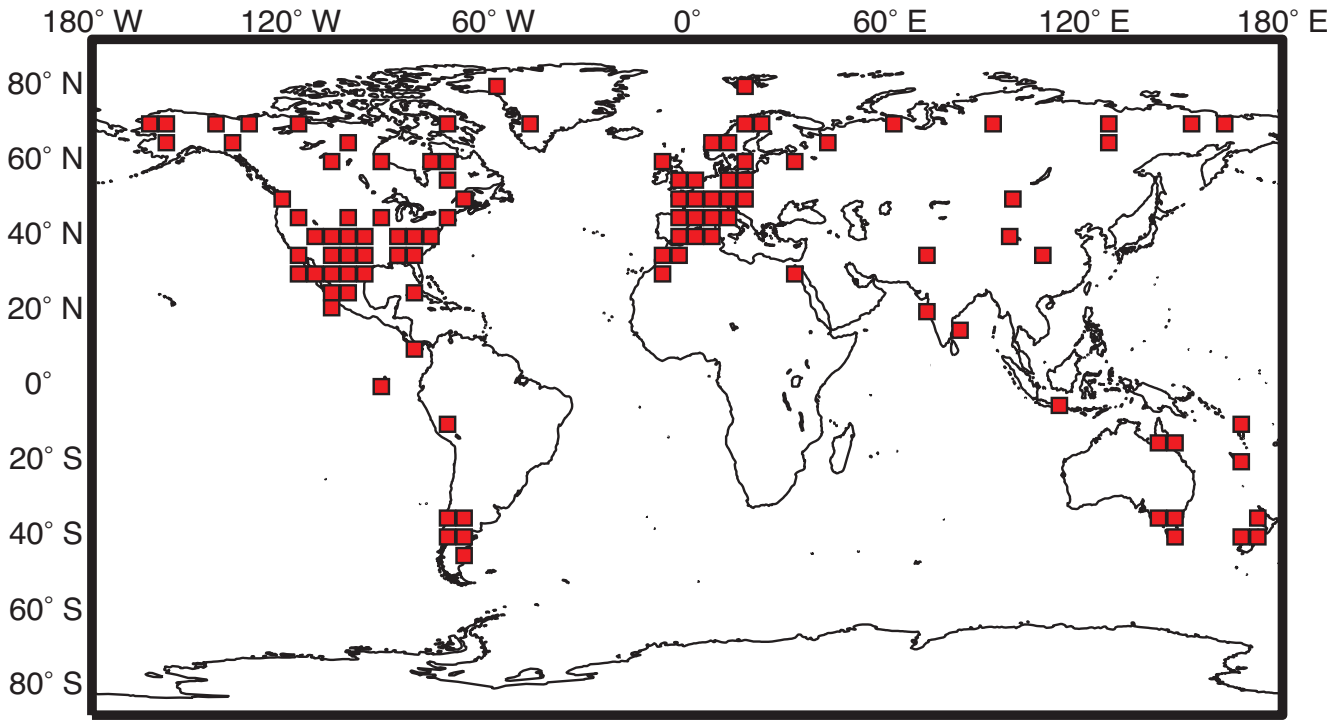

Figure 1: Map of grid-cell locations for the pseudoproxy network chosen to approximate the Mann et al. (1998) proxy locations.

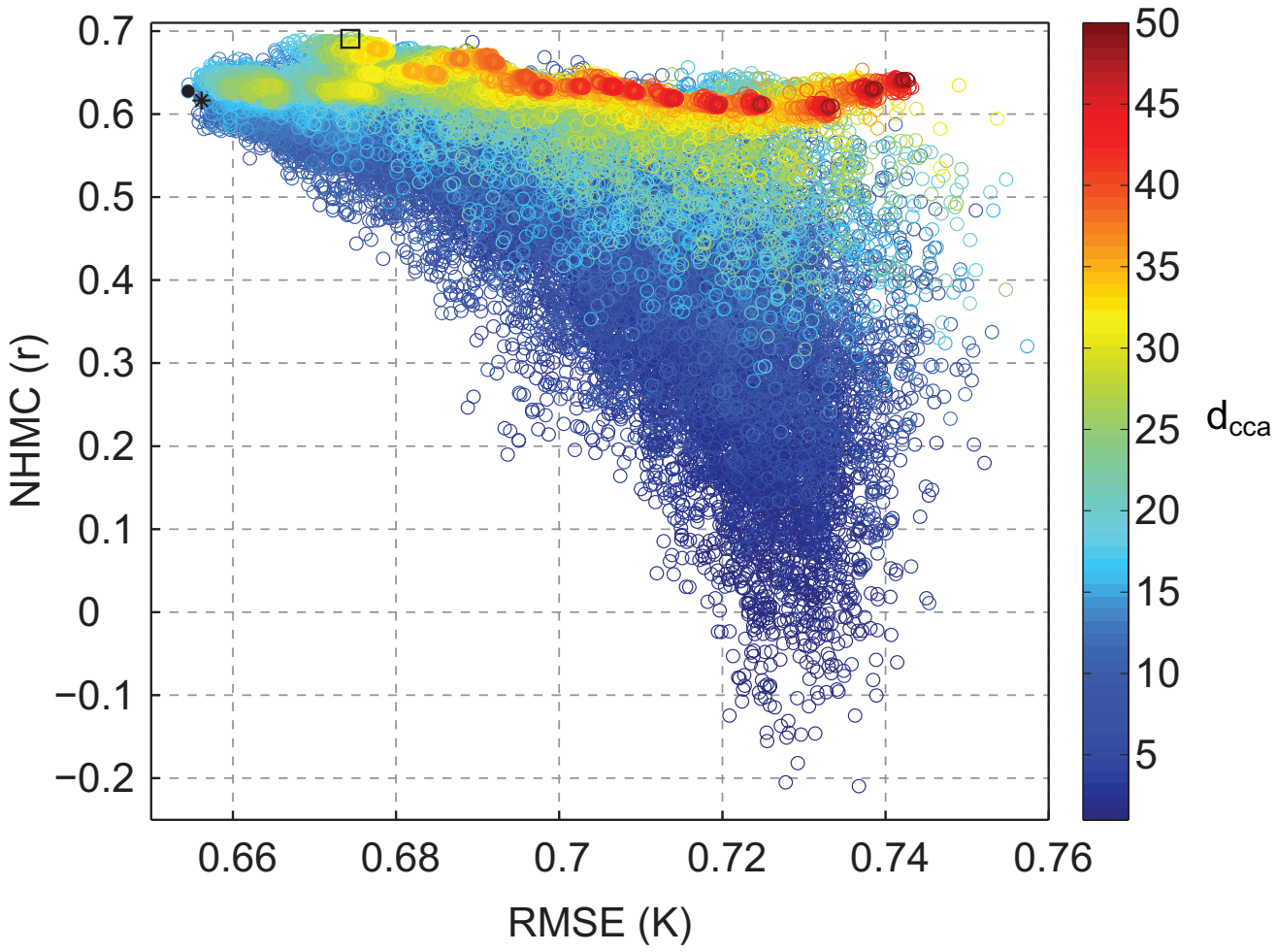

Figure 2: Cross-validation statistics during the calibration interval (1856-1980 C.E.) for the ensemble of CCA reconstructions at an SNR of 0.5. Colors in the figure indicate the value of d_{cca}, which ranges from 1-50. The symbols in the figure correspond to the following CCA solutions: the absolute minimum RMSE (black dot), the maximum NHMC (black square), and preferred solution based on RMSE (black star).

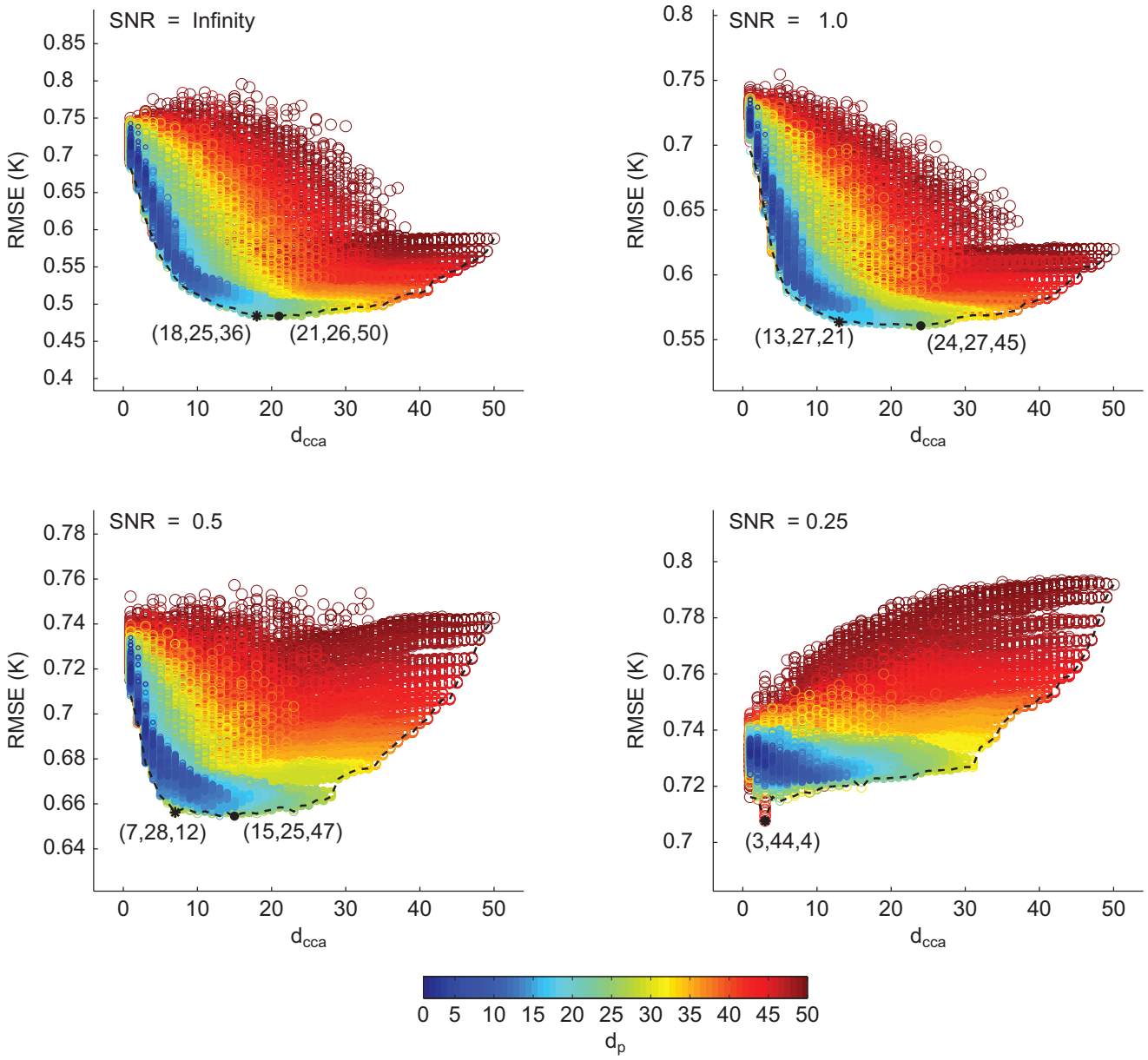

Figure 3: RMSE as a function of d_{cca} for all reconstructions spanning the collection of dimensional combinations between 1 and 50. Colors in the figure indicate the value of d_p chosen for the derived RMSE value. Black dots correspond to the absolute minimum RMSE and the values of d_{cca}, d_p, and d_t are given in the parenthesis next to each dot. The locations of the preferred solutions based on RMSE are also shown in each panel with a black star; the dimensional combinations for these values are also given in parenthesis.

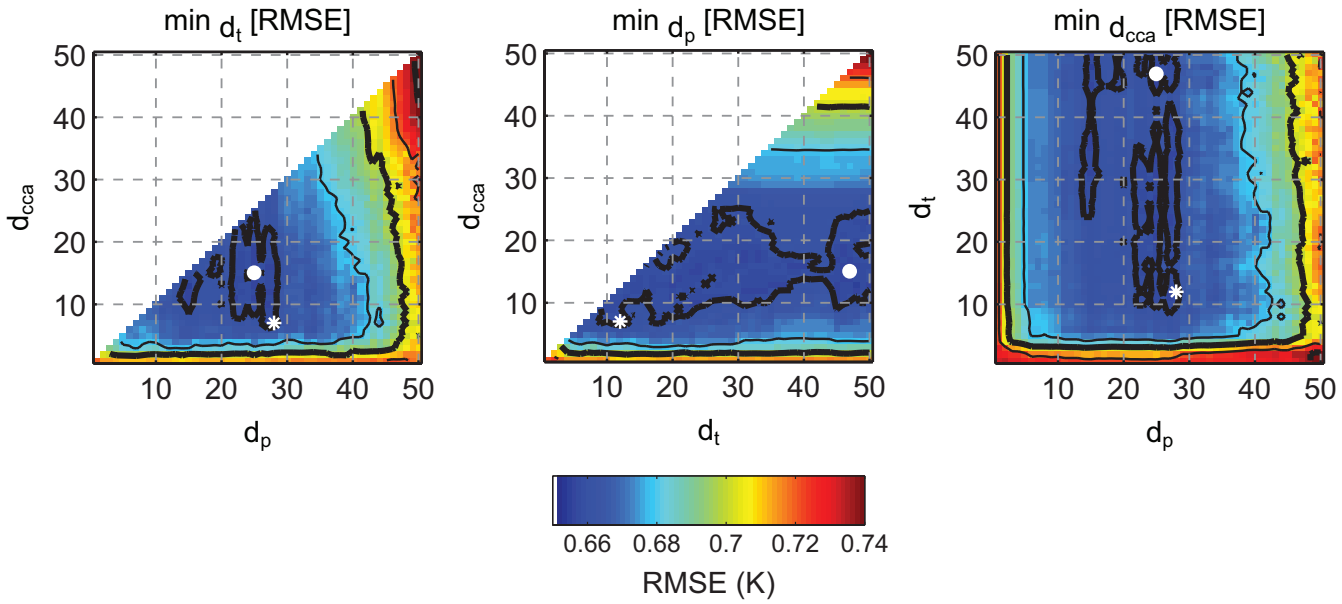

Figure 4: Minimum RMSE values for each pairing of the d_{cca}, d_p and d_t dimensions. The absolute minimum RMSE value is plotted as a white dot; the preferred solution value is plotted as a white star.

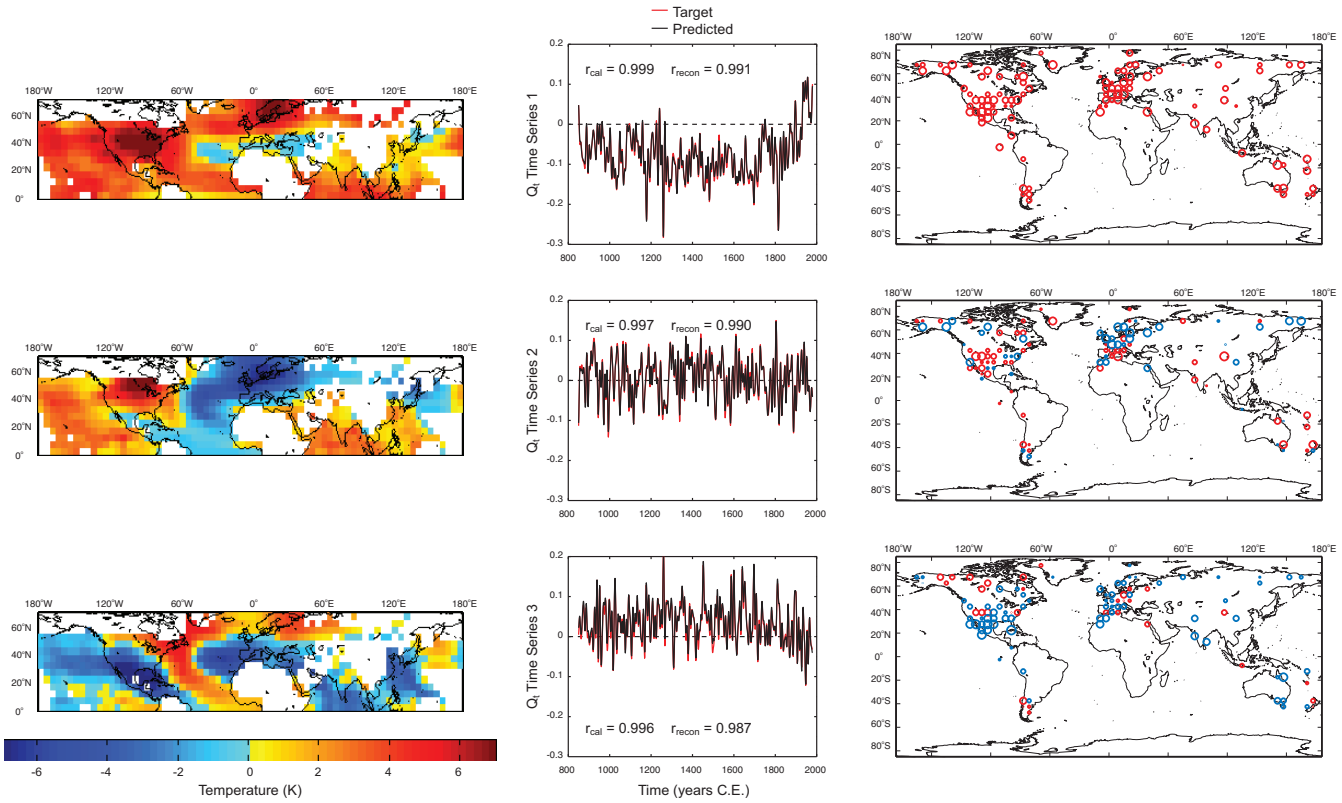

Figure 5: Temperature homogeneous covariance maps (C_t; left column), target and predicted time series of the temperature maps (Q_t and $\Sigma_{cca}Q_p^t$; middle column) and proxy homogeneous covariance maps (C_p; right column) for the first three canonical patterns of the no-noise reconstructions (rank increases from the top panels to the bottom). The markers in the proxy covariance maps reflect the loadings for each pattern, where blue and red markers are positive and negative loadings respectively. All elements are estimated over the calibration interval, but the time series are extended into the reconstruction interval by projecting the covariance maps onto the temperature and proxy matrices over the full temporal period. Correlations between the target and predicted time series during the calibration and reconstruction intervals are given in the middle column of panels.

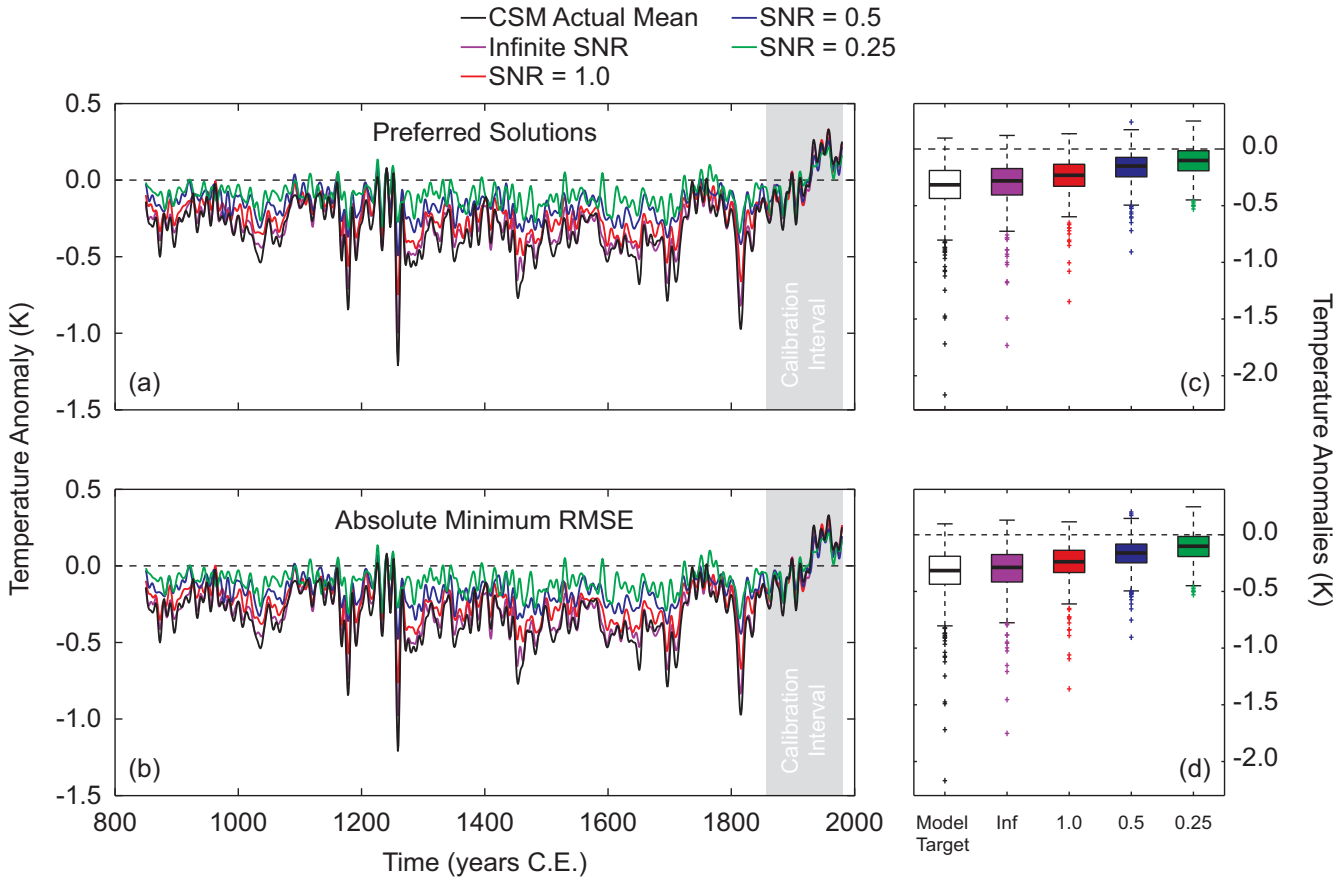

Figure 6: Area-weighted NH time series for the CCA reconstructions using d_{cca}, d_p, and d_t values associated with: (a) the preferred solution (Table 2); and (b) the absolute minimum RMSE values (Table 1). Time series have been smoothed using a decadal low-pass filter. Also shown in (c) and (d) are the box plots associated with the two combinations of the of d_{cca}, d_p, and d_t values, calculated from the annual (unfiltered) NH time series.

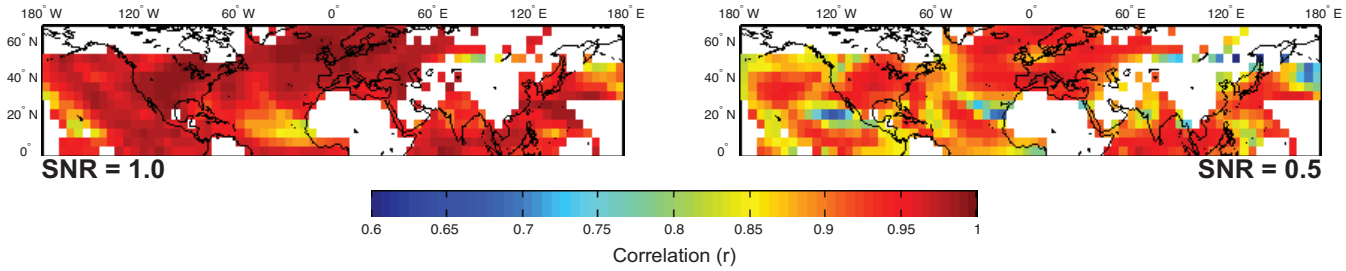

Figure 7: Correlation fields between the preferred (Table 2) and absolute-minimum RMSE reconstructions (Table 1). Results are shown for SNRs of 1.0 (left panel) and 0.5 (right panel) and are computed over the reconstruction interval (850-1855 C.E.).

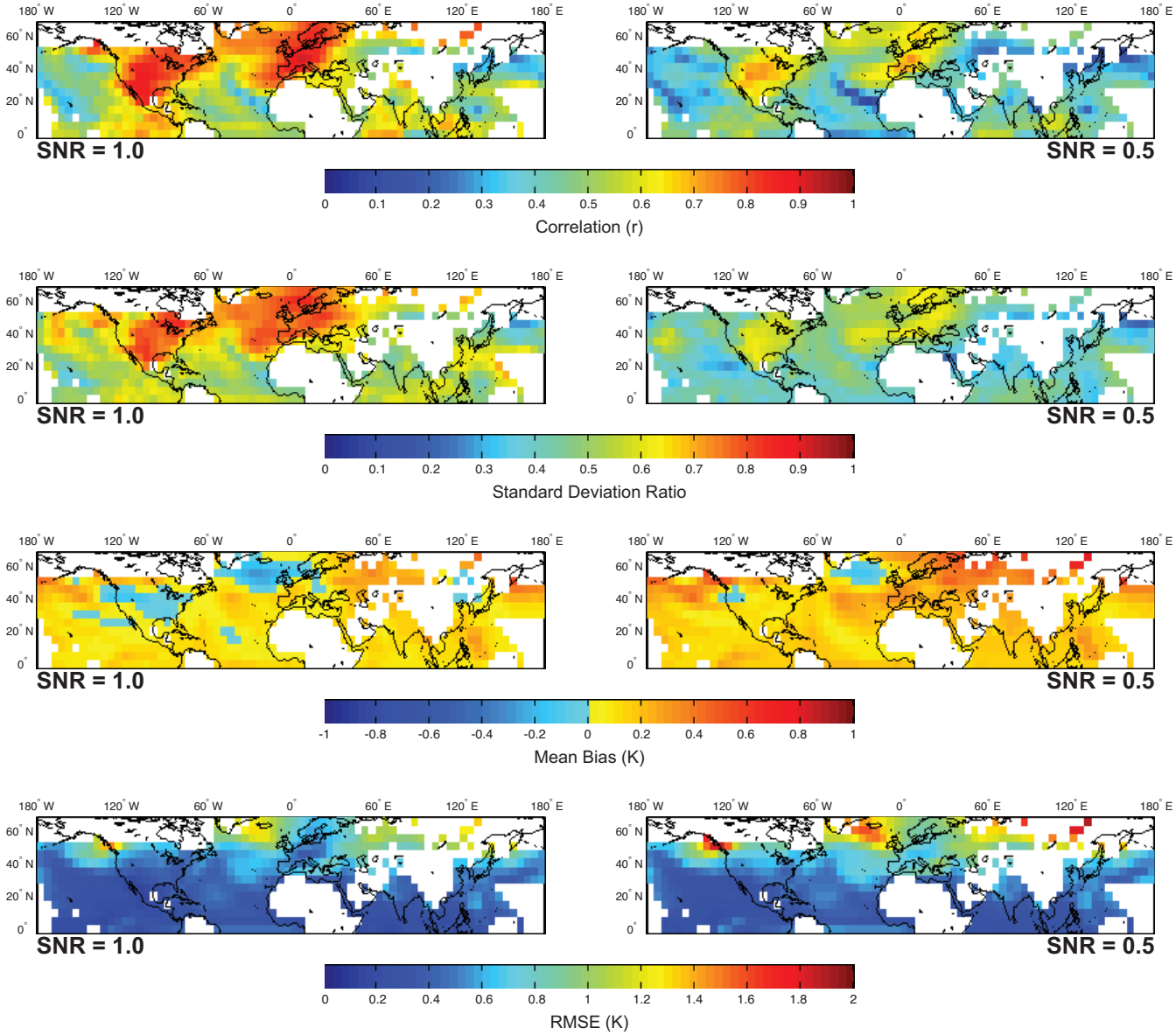

Figure 8: Field comparisons between derived CCA reconstructions (using the preferred-solution values of d_{cca}, d_p, and d_t) and the known CSM model fields: correlation (top row), standard deviation ratios (second row), mean biases (third row) and RMSE (last row). Standard deviation ratios are computed between the reconstruction and model and mean biases are computed as reconstruction minus model, i.e. negative (positive) biases indicate a colder (warmer) reconstruction mean. Results are shown for SNRs of 1.0 (left panels) and 0.5 (right panels); summary statistics for all noise levels are given in Table 3. All statistics are computed over the reconstruction interval (850-1855 C.E.).

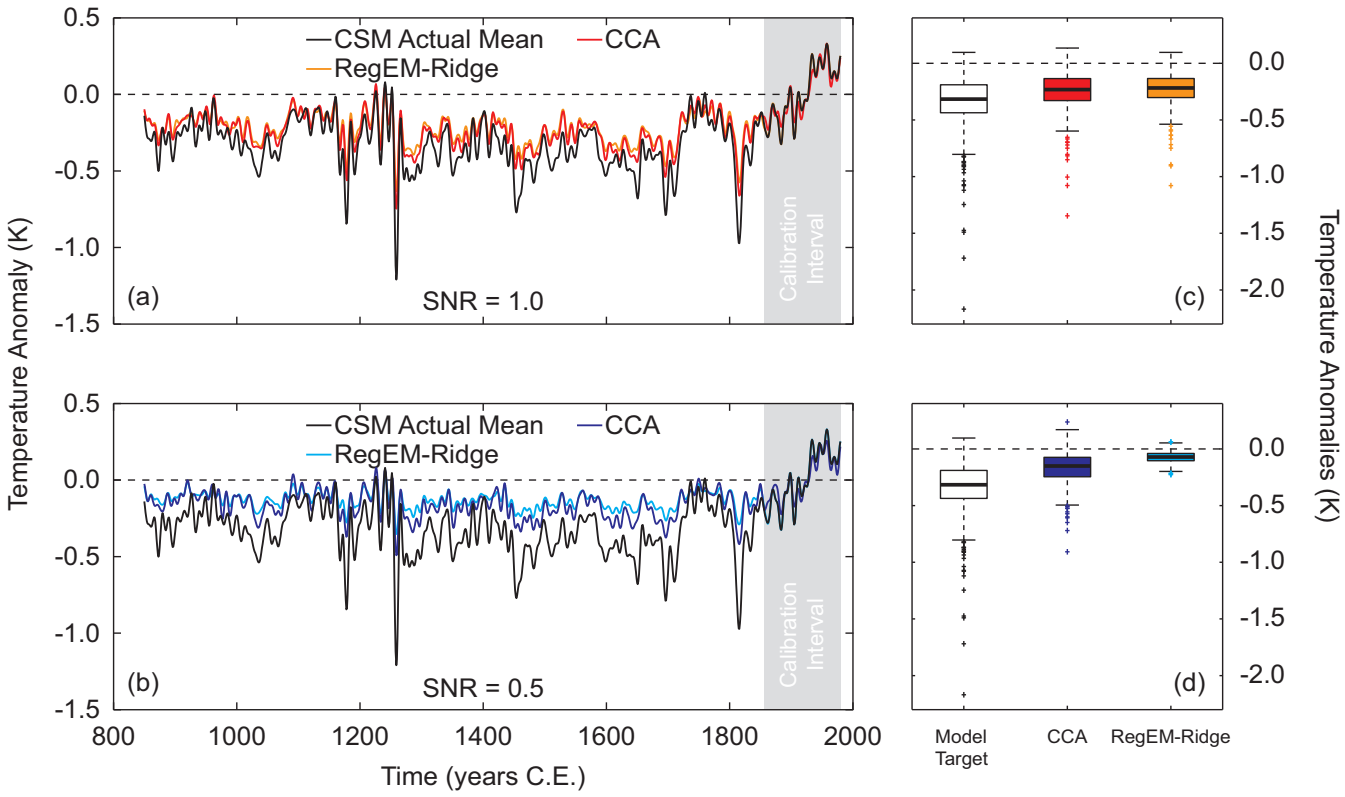

Figure 9: Same as in Figure 6, but for comparisons between the area-weighted NH time series for CCA and RegEM-Ridge reconstructions. Results are shown for SNRs of 1.0 and 0.5; summary statistics for all noise levels are given in Table 4.

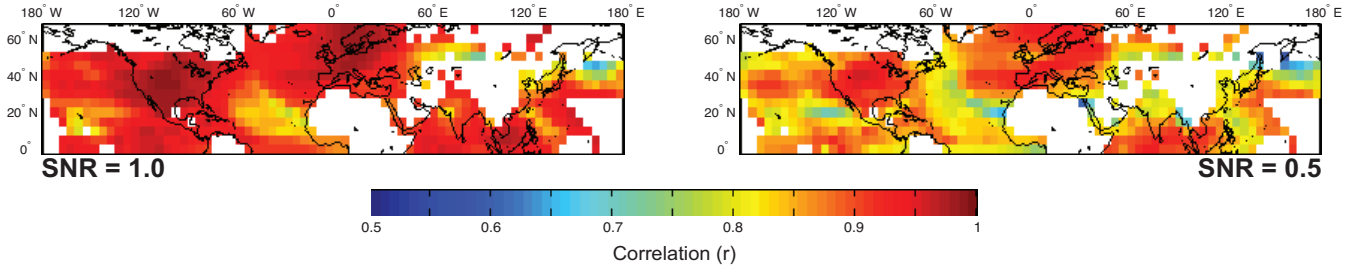

Figure 10: Correlation fields between the CCA and RegEM-Ridge reconstructions. Results are shown for SNRs of 1.0 (left panel) and 0.5 (right panel) and are computed over the reconstruction interval (850-1855 C.E.).

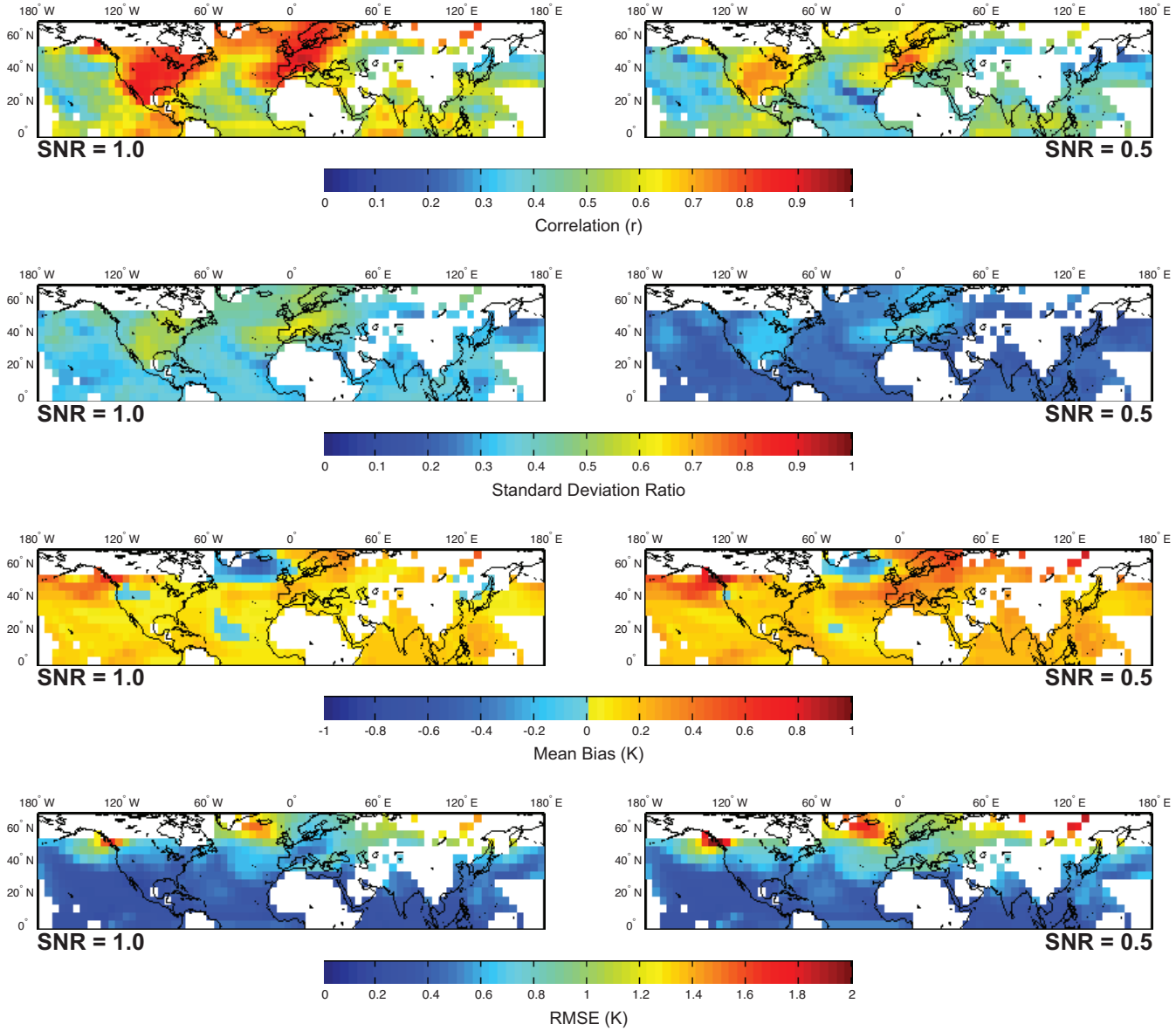

Figure 11: Same as in Figure 7, but for the RegEM-Ridge reconstructions.

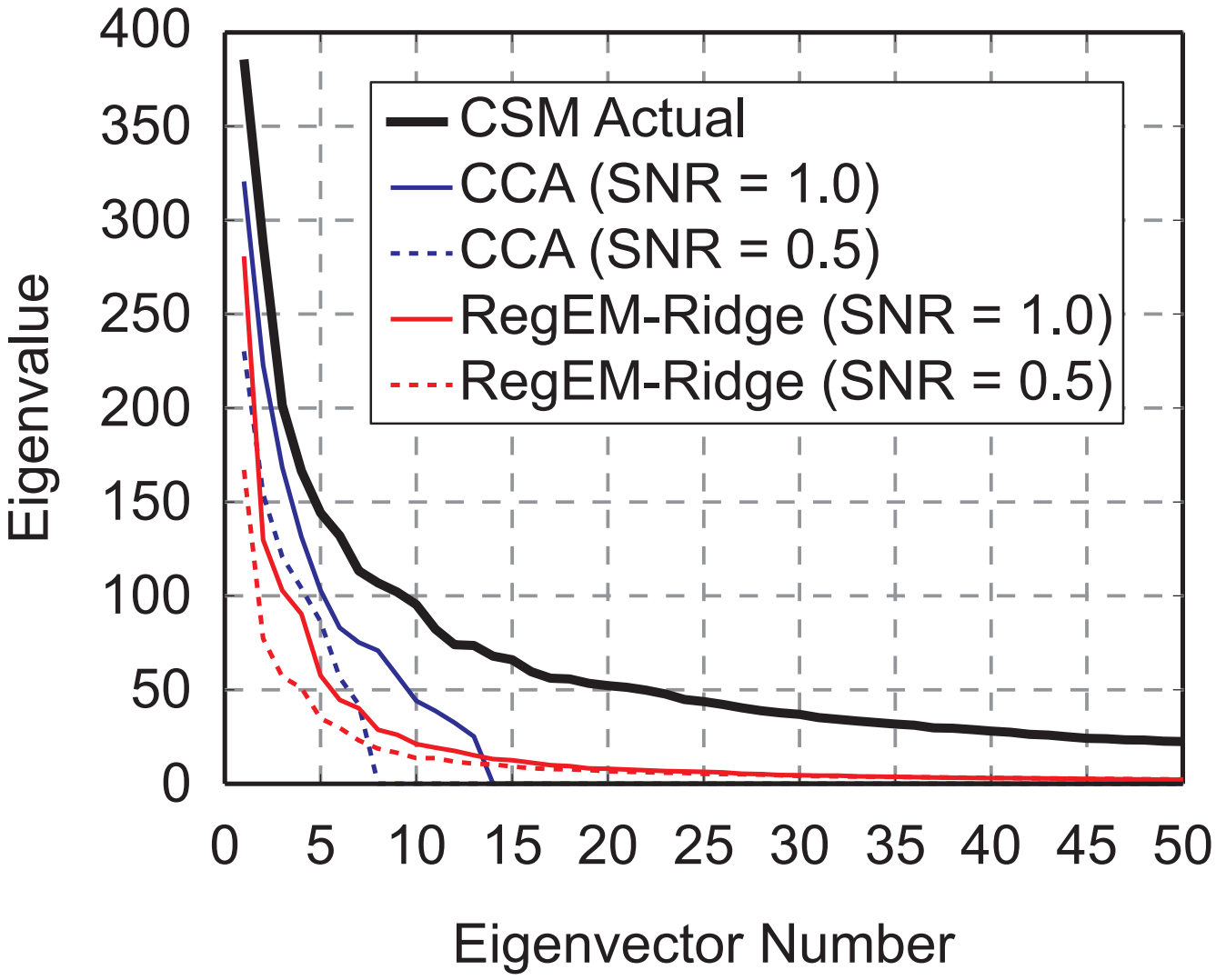

Figure 12: Eigenspectra computed from the true model temperature field and the CCA and RegEM-Ridge reconstructed temperature fields during the reconstructed interval (850-1855 C.E.). The CCA spectra have the characteristic truncation to zero at the selected rank, while the RegEM-Ridge spectra reflect the continuous filtration constraint applied in ridge regression.