This is a preliminary PDF of the author-produced manuscript that has been peer-reviewed and accepted for publication. Since it is being posted so soon after acceptance, it has not yet been copyedited, formatted, or processed by AMS Publications. This preliminary version of the manuscript may be downloaded, distributed, and cited, but please be aware that there will be visual differences and possibly some content differences between this version and the final published version.

The DOI for this manuscript is doi: 10.1175/2010JCLI3328.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.
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

*Corresponding author address: Jason E. Smerdon, Lamont-Doherty Earth Observatory of Columbia University, 61 Route 9W, P.O. Box 1000, Palisades, NY 10964. E-mail: jsmerdon@ldeo.columbia.edu

†LDEO contribution number XXXX
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 Atmospheric 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 multiple (3-13) 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, 2007a, 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 (Cook et al. 1994, 2004; Mann et al. 1998, 1999; Evans et al. 2002; Luterbacher et al. 1999; Rutherford et al. 2005; Casty et al. 2005; Pauling et al. 2006; Hegerl et al. 2007), but 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; Xoplaki et al. 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; Ammann and Wahl 2007; Mann et al. 2003, 2005, 2007a,b,c, 2008, 2009; 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; Bradley et al.
2003; Mann et al. 2003, 2005, 2007a,b,c, 2009). 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
reconstruction uncertainties. These questions are tied to understanding, for example, the
impact of proxy distributions and abundance, the connections between climate and proxy re-
ponses across different spectral domains, the response of proxies to multiple environmental
variables, and the role of teleconnections and noise in the calibration data - questions that
are ultimately fundamental to the success of efforts to reconstruct past climatic variability
(e.g. North et al. 2006; Jansen et al. 2007).

An additional element of uncertainty in climate reconstructions that has recently gained
more attention is the degree to which specific reconstruction methodologies impose their own
error and biases on derived reconstructions. Here we focus specifically on the uncertainties in
hemispheric-scale temperature reconstructions of the past millennium that arise principally
from the applied methodology. Reconstruction methods for this purpose generally can be
divided into two groups, one in which individual indices are targeted (see discussion in Mann
et al. (2005)) and climate field reconstruction (CFR) methods (Evans et al. 2001). Index
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, 2005; Crowley and Lowery 2000; Moberg et al. 2005; Hegerl et al. 2007; D’Arrigo et al. 2006; Mann et al. 2007a, 2008). Although index 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 index 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, 2009; 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-conditioned nature of the problem, are more dependent on the stability of climate-proxy connections and climate teleconnections, and require more extensive distributions of proxies than index reconstructions.

In spite of the differences between index 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. Similarly, Zhang et al. (2004) investigated two methods for drought reconstructions over the continental United States and also found their performance comparable. A more recent study
has compared the field skill of two temperature field reconstruction methods over the North
Atlantic and the European continent (Riedwyl et al. 2009). 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 researchers must use proxy
networks of opportunity and thus of variable composition in proxy type, location and tem-
poral extent. Uncertainty in any given reconstruction is therefore the combination of un-
certainties 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 Ruther-
ford 2002) has circumvented some of these challenges, however, by granting a consistent test
bed on which to 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; Kützel et al. 2007; Riedwyl et al. 2009; Christiansen
et al. 2009; Tingley and Huybers 2010a,b).

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 et al. 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 millen-
inium; (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; Kütteil et al. 2007; Hegerl et al. 2007; Riedwyl et al. 2009; Moberg et al. 2008; Smerdon et al. 2008a; Christiansen et al. 2009; Tingley and Huybers 2010a,b).

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 1999; 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 briefly in 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 is a recently favored method for temperature reconstructions (e.g. Rutherford et al. 2005; Mann et al. 2005, 2007a, 2008, 2009), 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. 2009; 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). Because the application of CCA requires selection of only three model dimensions it is straightforward to assess the skill of the method and computationally cheap to construct all possible models. This latter characteristic is in contrast to the iterative and more computationally expensive RegEM algorithm. Hence comparison of the two methods can help elucidate the strengths and weaknesses of each.

2. Data

We use pseudoproxies derived from the millennial simulation (850-1999 C.E.) of the National Center for Atmospheric Research (NCAR) Community Climate System Model (CCSM) 1.4, a coupled atmosphere-ocean GCM that has been driven with natural and anthropogenic forcings (Ammann et al. 2007). The simulated model field of annual surface
temperature means has been interpolated to a 5° longitude-latitude grid using bilinear in-
terpolation (Smerdon et al. 2008b; Rutherford et al. 2008). Multiple previous studies have
used this field with an incorrect geographic orientation and thus sampled pseudoproxies from
unintended locations (Smerdon et al. 2010). Here we use the correct field orientation and
the intended pseudoproxy locations specified in previous studies (Mann et al. 2005, 2007a,b;
Smerdon and Kaplan 2007; Smerdon et al. 2008a) as represented in Figure 1. These 104
pseudoproxies were sampled from the 5° grid-box locations that approximate the actual
proxy locations of the Mann et al. (1998) multiproxy network. White noise at four different
levels was added to model temperature time series from these selected locations to produce
pseudoproxy time series with signal-to-noise ratios (SNRs), by standard deviation, of infinity
(noise free), 1.0, 0.5 and 0.25.

We also used a subsampling of the CCSM mean annual temperature field to approximate
the availability of instrumental temperature data. Grid points missing more than 30% of
the annual data between 1856-1998 C.E. in the Jones et al. (1999) dataset were excluded
from use as target data as described by 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 instrumental data restrictions, the subsampled instrumental (calibration) data
are constrained to 1856-1980 C.E.; all annual temperature values within this period are
retained for each targeted temperature grid. Note that the above mentioned error regarding
the incorrect geographic orientation of the CCSM temperature field also affected the applied
instrumental data mask in previous studies (Smerdon et al. 2010); our experiments herein
have applied the correct mask as intended by these earlier studies.
3. Methods

a. Least-Squares CFRs as Multivariate Linear Regression

Multivariate linear regression is the underlying formalism of most CFR methods used to date. 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:

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

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,

$$T' = BP' + \varepsilon,$$  \hspace{1cm} (1)

where $B$ is a matrix of regression coefficients with dimensions $r \times m$, and $\varepsilon$ is the residual error. The error variances of all the elements of $\varepsilon$ in (1) are simultaneously minimized if $B$ is chosen as:
\[ B = (T'P'')(P'T')^{-1}, \]  
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:

\[ \hat{T} = M_t + S_t B S_p^{-1} (P - M_p), \]  
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 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 to the time dimension. In such cases, the cross-covariance matrix \( \langle T'P' \rangle \) and covariance matrix \( \langle P'P' \rangle \) cannot be well estimated. 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.
b. 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). This formalism as applied to the CFR problem is presented in detail in Appendix A and summarized below.

Two elements of the CCA application involve the eigenvalue decomposition and subsequent truncation of the proxy and temperature matrices. Both of these reductions are helpful in real-world applications where the temperature and proxy fields each contain noise or small scale features that cannot be reliably calibrated. Retaining a subset of EOFs in both fields can therefore guard against the possibility of calibrating modes dominated by noise or local variability. CCA in particular can be susceptible to random sampling fluctuations when the number of temporal observations is significantly less than the number of spatial locations in the target field (e.g. *Barnett and Preisendorfer* 1987; *Bretherton et al.* 1992), as is the case in the present CFR context. Over short time intervals, these random fluctuations can cause artificially high correlations between patterns that account for very little variance in either field, causing instabilities in the solutions due to sampling variability. Prefiltering the proxy and temperature matrices by retaining only a few leading modes therefore can improve the stability of CCA solutions. With regard to the reduction of the temperature field specifically, there are examples in the literature of CFR approaches that choose to either neglect or adopt a reduction of the field (e.g. *Luterbacher et al.* 2004; *Mann et al.* 2007a). Although we build the potential for reduction of the temperature field into the CCA formalism, the degree of reduction is determined from a cross-validation scheme that does not a priori require truncation. This scheme is discussed later in the manuscript and provides an objective means of
Determining whether or not reduction is warranted and by how much.

Decomposition of the standardized proxy matrix $P'$ during the calibration interval using Singular Value Decomposition (SVD; Golub and Van Loan 1996) is written:

$$P' = U_p \Sigma_p V_p^T. \quad (4)$$

where the columns of $U_p$ represent spatial patterns (empirical orthogonal functions or EOFs) and the principal components (PCs), $\Sigma_p V_p$, are orthonormal time series that combine with the EOF patterns to produce the original data set. The diagonal matrix $\Sigma_p$ contains the non-negative singular values, with squares proportional to the variance captured by the corresponding EOF-PC pairs. If the diagonal elements of $\Sigma_p$ 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:

$$P^r = U_p^r \Sigma_p^r V_p^r T. \quad (5)$$

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:

$$T^r = U_t^r \Sigma_t^r V_t^r T, \quad (6)$$

where $T^r$ only uses $d_t$ singular values and the corresponding number of singular vectors.

Note that $\text{rank}(P^r) = d_p$ and $\text{rank}(T^r) = d_t$, while $\text{rank}(P') = \min(m, n - 1)$ and $\text{rank}(T') = \min(r, n - 1)$. 

12
The above decompositions can be substituted into (2) and the corresponding matrix of regression coefficients written as

\[ B_{\text{cca}} = U_r^r \Sigma_r^r V_t^r T V_p^r (\Sigma_p^r)^{-1} U_p^r T = U_r^r \Sigma_r^r O_{\text{cca}}^r O_p^r T (\Sigma_p^r)^{-1} U_p^r T, \]

where \( O_{\text{cca}}^r O_p^r T \) is the truncated SVD of the cross-covariance matrix \( V_t^r T V_p^r \) in which \( d_{\text{cca}} \) leading canonical coefficients have been retained. From the formal derivation in Appendix A, the above expression for \( B_{\text{cca}} \), takes a simple form:

\[ B_{\text{cca}} = C_t \Sigma_{\text{cca}} W_p^T, \]

where \( C_t = U_t^r \Sigma_t^r O_t^r \) has the CCA temperature patterns in its columns and \( W_p = U_p^r (\Sigma_p^r)^{-1} O_p^r \) is the CCA proxy weighting matrix.

Applying \( B_{\text{cca}} \) 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 \( \Sigma_{\text{cca}}^r \), to produce the CCA timeseries for temperature:

\[ \hat{Q}_t^T = \Sigma_{\text{cca}}^r Q_p^T, \]

(iii) and use the \( C_t \) patterns to reconstruct a standardized version of the temperature fields:
\[ \hat{T}' = C_t \hat{Q}_t^T. \]

Note that in our formulated pseudoproxy experiments 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.b.1.

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

\[ \hat{T} = M_t + S_t B_{cca} S_p^{-1} (P - M_p). \] (8)

Performing this reconstruction thus 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 \( B_{cca} \). Under the assumption of stationarity between the mutual proxy and climate statistics, (8) 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 change to match the length of the given reconstruction period.

The operator \( B_{cca} \) is a reduced-rank \( \text{rank}(B_{cca}) = d_{cca} \) representation of the standard multivariate regression operator. Given calibration interval data sets \( T \) and \( P \), the matrix
$B_{\text{cca}}$ is completely determined upon the selection of three parameters for truncated ranks, $d_{\text{cca}}$, $d_p$, and $d_t$. Note that traditional applications of CCA did not involve rank reductions of the predictor and predictand matrices, and thus only depended on $d_{\text{cca}}$ (see the discussion in Bretherton et al. (1992)). Steps for reducing these matrix ranks by selecting $d_p$ 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. For some special subsets of dimensional selections, the CCA formalism reduces to other specialized forms of multivariate regression. For instance, CCA reduces to PC regression if the target field rank is not reduced and $d_{\text{cca}} = \min(d_p, d_t)$. PC regression has been applied in multiple paleoclimate contexts (e.g. Luterbacher et al. 2004) and further research on the relative performance of CCA and PC regression is warranted. Cogent discussions about the connection between CCA and multiple multivariate regression methods can be found in Barnett and Preisendorfer (1987), Bretherton et al. (1992), Cherry (1996), von Storch and Zwiers (1999) and Tippett et al. (2008).

c. CCA Model-Dimension Selection

Appropriate selections of the $d_{\text{cca}}$, $d_p$, 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 $d_p$ and $d_t$ by maintaining a specific level of retained variance in $T$ and $P$ and imposing the additional constraint that $d_{\text{cca}}$ be equal the minimum of $d_p$ and $d_t$. Barnett and Preisendorfer (1987) used principal component truncation rules
to determine $d_p$ and $d_t$ as proposed by Preisendorfer et al. (1981). The number of canonical coefficients ($d_{\text{cca}}$) 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. This procedure produces cross-validation statistics by calibrating independently on either the first or second halves of the target data and using the left-out half for validation. 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 instrumental period is split into two temporal halves: 1856-1917 and 1918-1980 C.E. We generate two sets of reconstructions using (8) and calibrate using each half of the target data to estimate the $B_{\text{cca}}$ 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 of the instrumental data. 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 from 1856-1980 C.E.

Using the above cross-validation scheme we compute the RMSE and NHMC for a range of $d_{\text{cca}}$, $d_p$ and $d_t$ combinations. The optimal selection of $d_{\text{cca}}$, $d_p$ 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 (8) are computed for the entire calibration interval (1856-1980 C.E.) and used for reconstructions in the preinstrumental period. Using the definitions

\[ B_f = S_t B_{\text{cca}} S_p^{-1}, \quad M_f = M_t - B_f M_p, \]  

(9)

the reconstruction in (8) can be rewritten in the final form of a linear transform with a constant:

\[ \hat{T} = M_f + B_f P. \]  

(10)

All columns of the matrix \( M_f \) 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 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 \( B_{\text{cca}} \) has rank \( d_{\text{cca}} \), and \( B_f \) is obtained in (9) by multiplying \( B_{\text{cca}} \) by non-singular diagonal matrices, \( B_f \) has the same size \( (r \times m) \) and rank \( (d_{\text{cca}}) \) as \( B_{\text{cca}} \). Such a matrix has \( d_{\text{cca}} \) 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 \( \Sigma \) 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

\[ B_f = U \Sigma V^T. \]  

(11)
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 $d_{\text{cca}}$ orthonormal vectors has

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

Similarly, the number of independent parameters in $V$ is

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

and $N(\Sigma) = d_{\text{cca}}$. In the general case, non-zero singular values of a matrix $B_f$ are different, the decomposition (11) is unique and, therefore,

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

Together with the constant offset parameters, the number of independent parameters that have to be determined in order to produce the reconstruction formula (10) is

$$N_{\text{tot}} = N(B_f) + N(M_f) = d_{\text{cca}}(r + m - d_{\text{cca}}) + r. \quad (12)$$

Substituting the values of $r$ and $m$ specific to the present pseudoproxy scenario ($r = 669$ and $m = 104$), the number of independent parameters in the CCA temperature field reconstructions are

$$N_{\text{tot}} = 669 + 773d_{\text{cca}} - d_{\text{cca}}^2. \quad (13)$$

The number of independent parameters in the CCA reconstructions therefore depends only on $d_{\text{cca}}$, 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 (10) 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 $d_{cca} \ll r + m = 773$, the $d_{cca}^2$ term in (12) and (13) is negligible compared to $(r + m)d_{cca} = 773d_{cca}$. 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 (10) 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.a we demonstrate a practical means of selecting the smallest $d_{cca}$ that produces a reconstruction with cross-validated RMSE practically indistinguishable from the absolute minimum of RMSE over all combinations of $d_{cca}$, $d_p$, and $d_t$. Thus the above arguments underlie the dimensional selection strategy that we employ throughout the remainder of the manuscript.

\textit{d. 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, 2010; Mann et al. 2005, 2007a,c, 2008; Smerdon and Kaplan 2007; Lee et al. 2007; Smerdon et al. 2008a; Riedwyl et al. 2009; Christiansen et al. 2009, 2010; Tingley and Huybers 2010a,b). While RegEM is an iterative method, the underlying formalism is based on a linear regression model that, for a
given time step, reconstructs missing data $X_m$ from available data $X_a$ and can be written as

$$X_m = M_m + S_m B S_a^{-1} (X_a - M_a).$$  \hspace{1cm} (14)

The notation here is analogous to (3), except the subindices $a$ and $m$ denote available and missing data, respectively, and are consistent with the notation adopted by Schneider (2001).

For the conventional expectation maximization (EM) 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 $X'_a$:

$$B = (X'_m X'_a^T) (X'_a X'_a^T)^{-1}. \hspace{1cm} (15)$$

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, 2010; Mann et al. 2005, 2007a,c; Smerdon and Kaplan 2007; Christiansen et al. 2009, 2010), 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 (15) is replaced by

$$(X'_a X'_a^T)^{-1} \longrightarrow (X'_a X'_a^T + h^2 I)^{-1} \hspace{1cm} (16)$$

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. In general, the RegEM algorithm allows the set
of missing data locations to vary arbitrarily with time, thus making all matrices appearing in (14) time variable. Nevertheless, in cases where the missing values in the data matrix take the form of a block submatrix (i.e. $X_m$ now denotes the estimate for this submatrix and $X_a$ is a submatrix of available data for the time period corresponding to $X_m$), the final RegEM reconstruction takes the form of (14) for a simple specific choice of $M_a$, $S_a$, $M_m$, $S_m$, and $B$ (Smerdon et al. 2008a); such is the case in the present millennial CFR context. Moreover, 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 (14) essentially becomes (3) and is comparable to the same form given for the CCA reconstruction in (8). In fact, both of these reconstruction formulas can be brought to the form in (10) that uses one offset and one matrix transformation.

The main difference between (8) and (14-16) is of course the form of regularization used for the regression matrix $B$, and the iteratively computed estimates of RegEM. Several relative advantages of the RegEM-Ridge method have been noted (Schneider 2001). In typical climatological applications where only a few principal components are retained based on often weak separations of the leading elements in the eigenvalue spectrum, the continuous filtering of the spectrum in ridge regression may provide advantages over regularizations, like CCA, that use finite eigenvalue truncation. The iterative EM procedure also allows the use of all data in the data matrix, as opposed to only the predictand and predictor data during their period of overlap in the calibration interval. In the specific type of paleoclimatic application considered herein, however, this advantage is limited principally to the pre-calibration period of the proxy matrix because the target data are completely missing prior to the mid-19th century (cf. Smerdon et al. 2008a).
4. Reconstruction Results

a. Selected model dimensions

We select $d_{cca}$, $d_p$, 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 $B_{cca}$ using equation (7); reconstructions during the validation period are performed using (8).

Following the approach described in Section 3.c, the CCA model was calibrated on each half of the the instrumental data and tested on the other half using all combinations of $d_{cca}$, $d_p$ and $d_t$ between 1 and 50 modes such that $d_{cca} \leq \min(d_p,d_t)$ (yielding $d_{cca}^2$ triplets $(d_{cca},d_p,d_t)$ for each $d_{cca}$ 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). The cross-validation statistics for early and late-calibration halves are given in Table 1. These 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.

Table 2 gives the minimum RMSE and maximum NHMC values among all $d_{cca}$, $d_p$, 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 2.

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. This suggests that RMSE is a more robust statistic for optimizing the CCA reconstructions than the NHMC. Furthermore, the colors of the circles in Figure 2 denote the values of $d_{\text{cca}}$, that is correspond to the number of independent parameters in the reconstruction model that is being validated. While particularly small $d_{\text{cca}}$ (less than 10) correspond to reconstructions that are both poor in RMSE and NHMC performance, high $d_{\text{cca}}$ (larger than 30) correspond to high NHMC but the full range of RMSE values. RMSE performance is especially poor for reconstructions with the largest $d_{\text{cca}}$ values. We therefore use RMSE as the principal basis for our selection criterion in subsequent dimensional selections. There are of course alternative cross-validation statistics that could be adopted. The coefficient of efficiency (CE) and reduction of error (RE) statistics are often used in paleoclimate literature as statistical validation measures. Advocates of these statistics point out that RE and CE measure the robustness of both the resolved variance and reconstructed mean in derived reconstructions (e.g. Wahl and Ammann 2007). This advantage is shared by the RMSE statistic adopted in this study, indicating that all three skill measures would be expected to produce similar results. Nevertheless, we adopt RMSE in the present application given its readily interpretable characteristics.

As mentioned earlier, the total number of combinations used to determine the optimized
dimensions given in Table 2 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 \( d_{\text{cca}} \)
without deviating significantly from the absolute minimum RMSE. Figure 3 plots RMSE
versus \( d_{\text{cca}} \) 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 \( d_{\text{cca}} \):

\[
\text{RMSE}^*(d_{\text{cca}}) = \min_{d_p,d_t} \text{RMSE}(d_{\text{cca}}, d_p, d_t).
\]

If \( d_p^*(d_{\text{cca}}) \) and \( d_t^*(d_{\text{cca}}) \) are the values of \( d_p \) and \( d_t \) that respectively minimize \( \text{RMSE}(d_{\text{cca}}, d_p, d_t) \)
for a given \( d_{\text{cca}} \), then the triplet \( (d_{\text{cca}}, d_p^*(d_{\text{cca}}), d_t^*(d_{\text{cca}})) \) defines the optimal (by the cross-
validated RMSE criterion) CCA reconstruction among all models with a fixed number of
independent parameters. Figure 3 demonstrates that \( \text{RMSE}^*(d_{\text{cca}}) \) decreases steeply for all
noise levels at small values of \( d_{\text{cca}} \). Beginning at a given \( d_{\text{cca}} \) 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 \( d_{\text{cca}} \) 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 \( d_{\text{cca}} \) at which an
increase by one does not reduce RMSE$^*$($d_{cca}$):

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

Optimal solutions ($d_{cca}^*$, $d_p^*(d_{cca}^*)$, $d_t^*(d_{cca}^*)$) are identified by stars in the panels of Figure 3 and are listed in Table 3 along with the corresponding values of RMSE and NHMC cross-validation statistics. At any noise level, RMSE$^*$($d_{cca}^*$) does not exceed $\min_{d_{cca}}$ (RMSE$^*$) by more than 1.5%. In subsequent presentations herein, we use these “beginning of the plateau” solutions ($d_{cca}^*$, $d_p^*(d_{cca}^*)$, $d_t^*(d_{cca}^*)$) as our preferred choices of the CCA dimensions (terming 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 ($d_{cca}$, $d_p$, $d_t$): Figure 4 presents two-dimensional fields of the RMSE minima with respect to the individual dimensions for a SNR = 0.5. 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 $d_{cca}$. Nevertheless, reductions in the dimensions of the temperature field are warranted. The yellow lines in Figure 3 plot the minimum RMSE values in the subset of solutions when $d_t$ is held constant at 50 (close to 62 or 63, the full-rank of the temperature
field in the two halves of the instrumental period). At all noise levels, the preferred solutions display reduced RMSE when the dimension of the temperature field is truncated. Moreover, in real-world reconstructions that involve predictand fields that contain noise or calibration intervals that can be shorter than the ones used herein, improvements in calibration errors due to the reduction of the predictand matrix would likely be more significant (Bretherton et al. 1992).

b. CCA Reconstructions

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 (see Section 3.b and Appendix A). 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 \((\Sigma_{cca}Q_T^T)\).

The three temperature covariance maps plotted in Figure 5 take on dynamically interpretable characteristics representing the variability in the CCSM versions of the global (hemispheric) mean, the North Atlantic Oscillation, and the El Niño/Southern Oscillation, respectively. This demonstration illustrates the physical interpretability of the derived covariance maps, which ultimately can be evaluated in terms of the reconstruction skill associated with individual dynamical patterns in the field.

As demonstrated in step (ii) of the three-step procedure in Section 3.b, 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 4 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 4). 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 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 temperature anomalies in the temperature covariance maps. These maps also indicate relatively balanced loadings of the pseudoproxies in which no single record is weighted heavily in a given pattern. Equivalent maps in real-world CFR applications would similarly be useful for evaluating the impact of specific proxies in the derived reconstructions.
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 $d_{\text{cca}}$ values, which were determined for the preferred solutions in Section 4.a to range from 13 in the no-noise case to 3 at an SNR of 0.25 (see Table 3). 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 5). These correlations are interestingly less than those determined for the first three canonical patterns at all noise levels given in Table 4. This is indicative of the fact that the leading individual patterns are reconstructed more skillfully than the mean of the combined field containing the full range of scaled 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.a 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 $r = 0.98$ or better during the reconstruction interval. The robustness of the
achieved results and the prevalence of the observed warm biases and variance losses in the NH means is thus illustrated by the strong similarities between Figures 6a and 6b. Local correlations also reflect a strong consistency between the absolute minimum and preferred reconstructions: the area-weighted mean field correlations from 850-1855 C.E. between the two reconstructions are 0.92, and 0.97 for SNR = infinity and 1.0, respectively (note that the dimensional selections for the SNR = 0.5 and 0.25 cases were the same for both the absolute minimum and preferred solutions, thus no correlation statistics are necessary for those noise levels). 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).

The box plots in Figures 6c and 6d are calculated from the distribution of the individual annual means in each NH time series during the reconstruction interval. The plots further demonstrate the warm biases and variance losses in the reconstructed NH time series, as well as 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.
3) RECONSTRUCTED FIELDS

Figure 7 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 5. Field correlations of course reduce with increased noise, but Figure 7 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 reconstructions 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 important 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 7). Standard deviation ratios (sample standard deviation of the reconstruction 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; note that although standard deviation ratios expressed in this manner cannot serve as a reasonable estimate of the system variance, they do serve as a convenient measure of the skill of the reconstruction (Smerdon et al. 2008a). 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 7, while the ratio drops to 0.29 at a
SNR of 0.25 (Table 5). 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 7).

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 parts of the Pacific and Atlantic Oceans). 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 5); high-noise reconstructions therefore are
dominated by warm-biased regions.

The bottom panels in Figure 7 show the RMSE of the fields, which 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.

c. Comparison of CCA and RegEM Reconstructions

We have used the same pseudoproxies from the CCA experiments above to compute cor-
responding non-hybrid (Rutherford et al. 2005) RegEM-Ridge reconstructions. The derived
reconstructions employ a standardization scheme realistically confined to the calibration interval (Smerdon and Kaplan 2007) during which no detrending has been applied; all reconstructions have used a stagnation tolerance of $1 \times 10^{-4}$. Figures 8a and 8b 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: reconstruction interval correlations between the CCA and RegEM-Ridge time series are 0.94, 0.98, 0.97, and 0.87 for SNR=\infty, 1.0, 0.5, 0.25, respectively. The reconstructed NH means also correlate with the true model mean at comparable levels (Table 5). There is, however, an indication that the RegEM-Ridge method performs slightly better given that the correlations increase by a few hundredths above those observed for CCA. The variance losses are larger in the RegEM-Ridge reconstructions, however, and can be clearly seen in the box plots in Figures 8c and 8d. These losses are also manifest in the failure to reconstruct extreme events, which is also slightly more apparent in the RegEM-Ridge reconstructions as illustrated in the latter panels of Figure 8.

The correlation fields between the CFRs derived from the two methods are plotted in Figure 9, 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 in the reconstruction interval are 0.89, 0.91, 0.84, and 0.61 for SNR=\infty, 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 experiments).

Validation fields for the RegEM-Ridge reconstructions are shown in Figure 10. These are
directly comparable to the CCA-validation fields shown in Figure 7. 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 5. 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 68% of those achieved for the CCA reconstructions at a SNR of infinity and fall to 52% 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 5) 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; and (2) the slightly higher correlations associated with the RegEM-Ridge reconstructions offset the errors associated with variance losses.
5. Discussion

Comparisons between CCA and RegEM-Ridge show that the methods produce very similar results, with the exception of the larger variance losses observed in the RegEM-Ridge reconstructions. 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 used 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 (cf. discussion surrounding eq. (19) in Schneider 2001). By contrast, the finite truncation of the CCA method yields leading modes that are unaffected by the truncation. To demonstrate this fact, Figure 11 plots the eigenspectra for the true model field and for the RegEM-Ridge and CCA CFRs at SNR levels of 1.0 and 0.5. The magnitudes of the RegEM-Ridge eigenvalues are strikingly reduced in comparison to those of CCA. Apart from their scaling, however, Figure 9 and the similarity of the correlation statistics for both methods shown in Figures 7 and 10 indicate that the two methods are reconstructing similar patterns, and differ primarily by the dampened variability of the leading modes in the RegEM-Ridge spectrum. Regarding the smaller differences between the CCA and RegEM-Ridge reconstructions, there are additional methodological choices that are likely contributors, e.g. differences in the cross-validation and parameter choices that
the two methods employ. Exploring further differences between the CCA and RegEM-Ridge methods would be insightful for understanding the small differences in the reconstructions that they generate and should be the subject of further research. Nevertheless, the results provided herein indicate that the methods produce broadly consistent reconstructions except for the larger variance losses observed for RegEM-Ridge.

The above discussion is relevant to the important and yet-to-be-explained difference between pseudoproxy CFRs derived using RegEM-Ridge and RegEM using truncated total least squares (hereinafter RegEM-TTLS; Mann et al. 2007a). This latter method has been shown to perform well in one pseudoproxy context (Mann et al. 2007a), particularly in terms of its ability to reproduce the NH mean index, while the former has not (Smerdon and Kaplan 2007). 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 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, making the Mann et al. (2007a,c) explanation implausible. The similarity between the CCA and RegEM-Ridge results presented herein further indicate that mean biases and variance losses in currently employed CFR methods are not tied to a specific methodological choice. 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 an improved understanding of why the
differences exist.

It is also important to highlight the observed concentration of the highest field correlations
(and preserved variance) in 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. Despite this emphasis on large-scale patterns, the observed
correlation distributions demonstrate that the methods perform best where dense sampling
exists. This suggests that addition of low-noise proxy data outside of existing highly-sampled
regions should be an important priority for improvement of regional-scale CFR skill.

Given the above observation, it is important to understand better the origin of the ob-
served 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. Previous experiments have indi-
cated that methodological performance is not strongly dependent on the employed model
simulation. Integrations from two different GCMs were used by von Storch et al. (2004, 2006)
to test the Mann et al. (1998) method and results were consistent across the simulations in
terms of the NH means. The authors also reported no significant dependence on the sam-
pling 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) used 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 data sets.

The use of pseudoproxy experiments as a research tool has proceeded under the assumption that modeled climates and pseudoproxies approximate well the conditions in real-world reconstruction problems. 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 with notable success (Evans et al. 2006; Anchukaitis et al. 2006) and can be used to simulate synthetic tree-ring chronologies for use in pseudoproxy studies. The seasonal dependencies of proxy records should also be considered in future work. Significant variations in field skill have been observed for multiproxy networks that target individual seasons (e.g. Pauling et al. 2003) and suggest that the annual pseudoproxy records used in most studies to date is another important idealization.

Incorporating these more complicated proxy characteristics in pseudoproxy studies will provide more realistic evaluations of CFR methods. Recent work also has shown the importance
of evaluating ensembles of reconstructions generated from multiple noise realizations in both the proxy and target datasets (Christiansen et al. 2009). Not all differences between methods tested on individual noise realizations may be statistically significant when uncertainties due to random errors are incorporated. Christiansen et al. (2009) have shown this is the case for NH mean estimates, but such ensemble work has not been done in the context of reconstruction performance in the field. Future work to evaluate field skill in ensembles of CFRs is therefore highly warranted.

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 robust and parsimonious dimensional combinations while guarding against artificial skill in the reconstruction. Our experiments demonstrate that the CCA method faithfully reconstructs between 3 and 13 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 patterns during the last several millennia.

Field correlations were 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 field errors will 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. Evaluation of the method using additional millennial simulations from AOGCMs or observational fields should be pursued to determine the robustness of the spatial skill dependencies that we have identified. More realistic pseudoproxy networks should also be considered that incorporate seasonal dependencies, multivariate climate responses and autoregressive noise structures. The impact of these more complicated pseudoproxy characteristics should be considered specifically with regard to the field characteristics as we have outlined in the present manuscript, as opposed
to the more widely evaluated performance of the NH mean. Their impact within different calibration scenarios is also important, particularly with regard to the length of the calibration interval and the range of climate variability represented in the calibration interval relative to the reconstruction interval (e.g. Jones et al. 2009).

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. These similarities are encouraging regarding the consistency of the two 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. These studies 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 present estimates of late-Holocene temperature variability.

The similarity 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. Resolving the origin of these differences is not only important for studies that have attempted to reconstruct temperatures over the last millennium (Rutherford et al. 2005; Mann et al. 2005, 2007a, 2008, 2009), but also for efforts that have applied RegEM in other contexts (e.g. Zhang et al. 2004; Steig et al. 2009). This necessity is further supported by the fact that pseudoproxy experiments have demonstrated differences between the performance of the two RegEM approaches, while real-world reconstructions of late-Holocene
temperatures derived from the two methods have not been notably different - at least in their representation of the NH mean (*Mann et al. 2007a*). Each of these observations indicates that the focus within the literature on only NH means is insufficient for evaluating CFR methods and their derived results. Explaining the performance differences between various CFR methods remains an open research question, but the persistence of similar problems in now multiple linear reconstruction methods suggests that caution must be exercised in the interpretation of published real-world CFR results.

*Acknowledgments.*

We gratefully acknowledge three anonymous reviewers for their constructive criticism and useful suggestions, all of which helped to improve the manuscript. This research was supported by the National Science Foundation by grants ATM04-07909, ATM-0902436, and ATM-0902715 and by the National Oceanic and Atmospheric Administration, U.S. Department of Commerce, by grant NA07OAR4310060 and NAOAR4320912, the latter under the Cooperative Institute for Climate Applications Research (CICAR). Part of this research was completed while DC was supported by a research internship from the Hughes Science Pipeline Project and JES was supported by a Mellon Postdoctoral Fellowship, both through the Department of Environmental Science at Barnard College. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the any of the above organizations or agencies.
Application of CCA to the Climate Field

Reconstruction Problem

Beginning with the SVDs of the proxy and temperature matrices written in Section 3.b, we use multivariate linear regression with a matrix $B'$

$$V^r_T = B'V^r_p + e_v$$

($e_v$ is the residual error) to predict the prewhitened PCs of temperature using the prewhitened proxy PCs:

$$\hat{V}^r_T = B'V^r_p.$$  

Because the prewhitened PCs are orthonormal, $V^r_p V^r_p = I$ (i.e. the identity matrix), the expression for $B'$ simplifies:

$$B' = (V^r_p V^r_p)(V^r_p V^r_p)^{-1} = V^r_p V^r.$$  

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

$$B' = V^r_p O_t \Sigma_{cca} O_p^T.$$  

(A1)
and can then be truncated by retaining only $d_{\text{cca}} \leq \min(d_p, d_t)$ leading singular values and corresponding patterns:

$$B' = O_t^r \Sigma_{\text{cca}}^r O_p^r T.$$  \hfill (A2)

Prediction of the prewhitened temperature PCs using $B'$ instead of $B'$, i.e.,

$$\hat{V}_t^r T = O_t^r \Sigma_{\text{cca}}^r O_p^r T V_p^r T$$ transforms into a simple form

$$\hat{Q}_t^T = \Sigma_{\text{cca}}^r Q_p^T$$ \hfill (A3)

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

$$Q_t = V_t^r O_t^r, \quad Q_p = V_p^r O_p^r.$$ \hfill (A4)

Similarly, the predicted $\hat{Q}_t$ corresponds to the predicted prewhitened 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 the weight matrices,

$$W_t = U_t^r (\Sigma_t^r)^{-1} O_t^r, \quad W_p = U_p^r (\Sigma_p^r)^{-1} O_p^r.$$ \hfill (A5)

so that

$$Q_t^T = W_t^T T', \quad Q_p^T = W_p^T P',$$ \hfill (A6)

where Eqs. (5) and (6) and the orthonormality of the truncated EOF sets of $U_t^r$ and $U_p^r$, i.e. columns, were used.
It follows from (A-4) that the columns of $Q_t$ and $Q_p$ are orthonormal sets. Moreover, inserting (A-4) into (A-1) yeilds,

$$Q_t^T Q_p = \Sigma^r_{cca},$$

hence the columns of $Q_t$ and $Q_p$ with different ordering are orthogonal, while those with the same ordering are positively correlated. The correlation coefficients of these latter columns are equal to the diagonal elements of $\Sigma^r_{cca}$, and are called canonical correlations. Because of the SVD decomposition in (A-1), these 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 (A-3) amount to a simple multiplication of the CCA time series of the proxies by the diagonal elements of $\Sigma^r_{cca}$. 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:

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

\[
C_p = (P'_p Q_p)(Q'_p Q_p)^{-1} = P'_p \Sigma_p V'_p V'_p O'_p = U'_p \Sigma'_p O'_p \tag{A7}
\]

and similarly,

\[
C_t = T'_t Q_t = U'_t \Sigma'_t O'_t. \tag{A8}
\]

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

\[
B_{cca} = U'_t \Sigma'_t V'_p V'_p (\Sigma'_p)^{-1} U'_p = U'_t \Sigma'_t O'_t \Sigma'_p O'_p (\Sigma'_p)^{-1} U'_p,
\]

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

\[
(P' P^T)^{-1} \longrightarrow (P' P^T)^+ = (P'_p P'_p)^+ = U'_p (\Sigma'_p)^{-2} U'_p.
\]

Given the definitions in Eqs. (A-5) and (A-8), \( B_{cca} \), takes a simple form:

\[
B_{cca} = C_t \Sigma'_p W'_p. \tag{A9}
\]
REFERENCES


Christiansen, B. T. Schmith, and P. Thejll, 2010: Reply to Comment on A surrogate ensemble study .. by Rutherford et al., J. Climate, in press.


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


Fritts, H. C., T. J. Blasing, B. P. Hayden, and J. E. Kutzbach, 1971: Multivariate techniques


Mann, M.E., S. Rutherford, E. Wahl, C. Ammann, 2007b: Reply to Comments on “Testing


North, G.R. et al., 2006: *Surface Temperature Reconstructions for the Last 2,000 Years*, The

Pauling, A., J. Luterbacher, and H. Wanner, 2003: Evaluation of proxies for European
and North Atlantic temperature field reconstructions, *Geophys. Res. Lett.*, 30(15), 1787,

Pauling, A., J. Luterbacher, C. Casty, and H. Wanner, 2006: 500 years of gridded high-
resolution precipitation reconstructions over Europe and the connection to large-scale


Riedwyl, N., M. KütteI, J. Luterbacher, H. and Wanner, 2009: Comparison of climate field
reconstruction techniques: Application to Europe. *Clim. Dynam.*, 32, 381-395

Rutherford, S., and M. E. Mann, 2004: Correction to Optimal surface temperature recon-
structions using terrestrial borehole data by Mann et al. *J. Geophys. Res.*, 109, D11107,

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 reconstruc-
tions: Sensitivity to methodology, predictor network, target season and target domain. *J.
Climate*, 18, 2308-2329.


### List of Tables

<table>
<thead>
<tr>
<th></th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Early (1856-1916 C.E.) and late-half (1917-1980 C.E.) cross-validation statistics for CCA; all statistics and dimensions represent those achieved for the minimum RMSE in the two respective cross-validation periods.</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>CCA reconstruction statistics using the absolute minimum RMSE or maximum NHMC criteria during the calibration interval (1856-1980 C.E.).</td>
<td>61</td>
</tr>
<tr>
<td>3</td>
<td>CCA reconstruction statistics for the preferred solutions in which parsimonious dimensional combinations have been chosen as the first local minimum of the RMSE statistic.</td>
<td>62</td>
</tr>
<tr>
<td>4</td>
<td>Correlation statistics between the true canonical temperature time series, $Q_t$, and those predicted by the proxy PCs, i.e. $\Sigma_{_{\text{cca}}}Q_p^T$. Statistics are shown for both the reconstruction and calibration intervals.</td>
<td>63</td>
</tr>
<tr>
<td>5</td>
<td>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.</td>
<td>64</td>
</tr>
</tbody>
</table>
Table 1. Early (1856-1916 C.E.) and late-half (1917-1980 C.E.) cross-validation statistics for CCA; all statistics and dimensions represent those achieved for the minimum RMSE in the two respective cross-validation periods.

<table>
<thead>
<tr>
<th>SNR</th>
<th>RMSE</th>
<th>(d_{cca})</th>
<th>(d_p)</th>
<th>(d_t)</th>
<th>NHMC</th>
<th>RMSE</th>
<th>(d_{cca})</th>
<th>(d_p)</th>
<th>(d_t)</th>
<th>NHMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inf</td>
<td>0.54</td>
<td>19</td>
<td>24</td>
<td>39</td>
<td>0.84</td>
<td>0.58</td>
<td>16</td>
<td>33</td>
<td>30</td>
<td>0.77</td>
</tr>
<tr>
<td>1.0</td>
<td>0.62</td>
<td>16</td>
<td>23</td>
<td>21</td>
<td>0.57</td>
<td>0.65</td>
<td>16</td>
<td>24</td>
<td>20</td>
<td>0.64</td>
</tr>
<tr>
<td>0.5</td>
<td>0.70</td>
<td>10</td>
<td>40</td>
<td>19</td>
<td>0.54</td>
<td>0.71</td>
<td>7</td>
<td>26</td>
<td>11</td>
<td>0.53</td>
</tr>
<tr>
<td>0.25</td>
<td>0.74</td>
<td>9</td>
<td>40</td>
<td>39</td>
<td>0.21</td>
<td>0.76</td>
<td>3</td>
<td>48</td>
<td>32</td>
<td>-0.02</td>
</tr>
</tbody>
</table>
Table 2. CCA reconstruction statistics using the absolute minimum RMSE or maximum NHMC criteria during the calibration interval (1856-1980 C.E.).

<table>
<thead>
<tr>
<th>SNR</th>
<th>RMSE</th>
<th>(d_{\text{cca}})</th>
<th>(d_p)</th>
<th>(d_t)</th>
<th>NHMC</th>
<th>RMSE</th>
<th>(d_{\text{cca}})</th>
<th>(d_p)</th>
<th>(d_t)</th>
<th>NHMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inf</td>
<td>0.57</td>
<td>22</td>
<td>35</td>
<td>46</td>
<td>0.77</td>
<td>0.59</td>
<td>25</td>
<td>29</td>
<td>50</td>
<td>0.83</td>
</tr>
<tr>
<td>1.0</td>
<td>0.64</td>
<td>15</td>
<td>23</td>
<td>20</td>
<td>0.59</td>
<td>0.69</td>
<td>16</td>
<td>37</td>
<td>47</td>
<td>0.67</td>
</tr>
<tr>
<td>0.5</td>
<td>0.71</td>
<td>6</td>
<td>22</td>
<td>10</td>
<td>0.52</td>
<td>0.74</td>
<td>5</td>
<td>21</td>
<td>26</td>
<td>0.61</td>
</tr>
<tr>
<td>0.25</td>
<td>0.76</td>
<td>3</td>
<td>30</td>
<td>4</td>
<td>0.26</td>
<td>0.79</td>
<td>6</td>
<td>50</td>
<td>48</td>
<td>0.33</td>
</tr>
</tbody>
</table>
Table 3. CCA reconstruction statistics for the preferred solutions in which parsimonious dimensional combinations have been chosen as the first local minimum of the RMSE statistic.

<table>
<thead>
<tr>
<th>SNR</th>
<th>RMSE</th>
<th>d_{cca}</th>
<th>d_p</th>
<th>d_t</th>
<th>NHMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inf.</td>
<td>0.57</td>
<td>13</td>
<td>34</td>
<td>26</td>
<td>0.74</td>
</tr>
<tr>
<td>1.0</td>
<td>0.65</td>
<td>10</td>
<td>23</td>
<td>12</td>
<td>0.60</td>
</tr>
<tr>
<td>0.5</td>
<td>0.71</td>
<td>6</td>
<td>22</td>
<td>10</td>
<td>0.52</td>
</tr>
<tr>
<td>0.25</td>
<td>0.76</td>
<td>3</td>
<td>30</td>
<td>4</td>
<td>0.26</td>
</tr>
<tr>
<td>CCA Rank</td>
<td>SNR Infinity Cal.</td>
<td>SNR Infinity Recon.</td>
<td>SNR 1.0 Cal.</td>
<td>SNR 1.0 Recon.</td>
<td>SNR 0.5 Cal.</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>-------------</td>
<td>---------------</td>
<td>-------------</td>
</tr>
<tr>
<td>1</td>
<td>0.997</td>
<td>0.987</td>
<td>0.956</td>
<td>0.909</td>
<td>0.901</td>
</tr>
<tr>
<td>2</td>
<td>0.996</td>
<td>0.991</td>
<td>0.938</td>
<td>0.915</td>
<td>0.849</td>
</tr>
<tr>
<td>3</td>
<td>0.994</td>
<td>0.981</td>
<td>0.914</td>
<td>0.868</td>
<td>0.801</td>
</tr>
<tr>
<td>4</td>
<td>0.992</td>
<td>0.963</td>
<td>0.908</td>
<td>0.835</td>
<td>0.764</td>
</tr>
<tr>
<td>5</td>
<td>0.986</td>
<td>0.953</td>
<td>0.828</td>
<td>0.706</td>
<td>0.678</td>
</tr>
<tr>
<td>6</td>
<td>0.969</td>
<td>0.891</td>
<td>0.797</td>
<td>0.421</td>
<td>0.659</td>
</tr>
<tr>
<td>7</td>
<td>0.959</td>
<td>0.904</td>
<td>0.768</td>
<td>0.635</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.953</td>
<td>0.919</td>
<td>0.651</td>
<td>0.573</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.948</td>
<td>0.888</td>
<td>0.619</td>
<td>0.519</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.933</td>
<td>0.777</td>
<td>0.473</td>
<td>0.273</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.915</td>
<td>0.822</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.905</td>
<td>0.792</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.853</td>
<td>0.654</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. 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.
### Table 5

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.

**CCA**

<table>
<thead>
<tr>
<th>SNR</th>
<th>NHMC</th>
<th>Mean Field Corr.</th>
<th>Mean STD Ratio</th>
<th>Mean Bias (K)</th>
<th>Mean RMSE (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inf.</td>
<td>0.91</td>
<td>0.70</td>
<td>0.76</td>
<td>0.03</td>
<td>0.49</td>
</tr>
<tr>
<td>1.0</td>
<td>0.84</td>
<td>0.58</td>
<td>0.58</td>
<td>0.09</td>
<td>0.59</td>
</tr>
<tr>
<td>0.5</td>
<td>0.76</td>
<td>0.44</td>
<td>0.44</td>
<td>0.17</td>
<td>0.67</td>
</tr>
<tr>
<td>0.25</td>
<td>0.43</td>
<td>0.23</td>
<td>0.29</td>
<td>0.27</td>
<td>0.77</td>
</tr>
</tbody>
</table>

**RegEM-Ridge**

<table>
<thead>
<tr>
<th>SNR</th>
<th>NHMC</th>
<th>Mean Field Corr.</th>
<th>Mean STD Ratio</th>
<th>Mean Bias (K)</th>
<th>Mean RMSE (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inf.</td>
<td>0.91</td>
<td>0.68</td>
<td>0.52</td>
<td>-0.01</td>
<td>0.52</td>
</tr>
<tr>
<td>1.0</td>
<td>0.85</td>
<td>0.60</td>
<td>0.38</td>
<td>0.11</td>
<td>0.60</td>
</tr>
<tr>
<td>0.5</td>
<td>0.78</td>
<td>0.47</td>
<td>0.24</td>
<td>0.20</td>
<td>0.69</td>
</tr>
<tr>
<td>0.25</td>
<td>0.49</td>
<td>0.24</td>
<td>0.15</td>
<td>0.28</td>
<td>0.76</td>
</tr>
</tbody>
</table>
List of Figures

1 Map of grid-cell locations for the pseudoproxy network chosen to approximate the Mann et al. (1998) proxy locations. 69

2 Cross-validation statistics during the calibration interval (1856-1980 C.E.) for the ensemble of CCA reconstructions at a SNR of 0.5. Colors in the figure indicate the value of $d_{\text{cca}}$, 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 (pink square), and preferred solution based on RMSE (black star), which overlays the black dot for this particular noise realization. 70

3 RMSE as a function of $d_{\text{cca}}$ 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. Yellow lines in each of the RMSE plots indicate the minimum RMSE achieved when $d_{t}$ is held constant at 50. Black dots correspond to the absolute minimum RMSE and the values of $d_{\text{cca}}, d_{p}$, and $d_{t}$ are given in the parentheses 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 parentheses. 71
Minimum RMSE values for each pairing of the $d_{\text{cca}}, d_p$ and $d_t$ dimensions at an SNR of 0.5. The absolute minimum RMSE value is plotted as a white dot; the preferred solution value is plotted as a white star (which overlays the white dot for this particular noise realization).

Temperature homogeneous covariance maps ($C_t$; left column), target and predicted time series of the temperature maps ($Q_t$ and $\Sigma_{\text{cca}} Q_p^t$; 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 circles in the proxy maps represent the loadings for each pattern, where blue and red circles correspond to positive and negative loadings, respectively, and the circle areas are proportional to the loading magnitude. 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. The percentage of variance explained in the target field during the calibration interval are 11.6, 4.5 and 9.8% by the first, second and third covariance maps, respectively.
Area-weighted NH time series for the CCA reconstructions using $d_{\text{cca}}$, $d_p$, and $d_t$ values associated with: (a) the preferred solution (Table 3); and (b) the absolute minimum RMSE values (Table 2). 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 $d_{\text{cca}}$, $d_p$, and $d_t$ values. These plots were calculated from the distribution of the individual annual means in each NH time series during the reconstruction interval.

Field comparisons between derived CCA reconstructions (using the preferred-solution values of $d_{\text{cca}}$, $d_p$, and $d_t$) and the known CCSM model field: 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 4. All statistics are computed over the reconstruction interval (850-1855 C.E.).

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 5.

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.).

Same as in Figure 7, but for the RegEM-Ridge reconstructions.
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.
Fig. 1. Map of grid-cell locations for the pseudoproxy network chosen to approximate the Mann et al. (1998) proxy locations.
Fig. 2. Cross-validation statistics during the calibration interval (1856-1980 C.E.) for the ensemble of CCA reconstructions at a SNR of 0.5. Colors in the figure indicate the value of $d_{\text{cca}}$, 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 (pink square), and preferred solution based on RMSE (black star), which overlays the black dot for this particular noise realization.
Fig. 3. RMSE as a function of $d_{\text{cca}}$ 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. Yellow lines in each of the RMSE plots indicate the minimum RMSE achieved when $d_t$ is held constant at 50. Black dots correspond to the absolute minimum RMSE and the values of $d_{\text{cca}}, d_p$, and $d_t$ are given in the parentheses 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 parentheses.
Fig. 4. Minimum RMSE values for each pairing of the $d_{\text{cca}}$, $d_p$, and $d_t$ dimensions at an SNR of 0.5. The absolute minimum RMSE value is plotted as a white dot; the preferred solution value is plotted as a white star (which overlays the white dot for this particular noise realization).
Fig. 5. Temperature homogeneous covariance maps ($C_t$; left column), target and predicted time series of the temperature maps ($Q_t$ and $\Sigma_{cca}Q_t^p$; 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 circles in the proxy maps represent the loadings for each pattern, where blue and red circles correspond to positive and negative loadings, respectively, and the circle areas are proportional to the loading magnitude. 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. The percentage of variance explained in the target field during the calibration interval are 11.6, 4.5 and 9.8% by the first, second and third covariance maps, respectively.
Fig. 6. Area-weighted NH time series for the CCA reconstructions using \( d_{\text{cca}} \), \( d_p \), and \( d_t \) values associated with: (a) the preferred solution (Table 3); and (b) the absolute minimum RMSE values (Table 2). 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 \( d_{\text{cca}}, d_p, \) and \( d_t \) values. These plots were calculated from the distribution of the individual annual means in each NH time series during the reconstruction interval.
Fig. 7. Field comparisons between derived CCA reconstructions (using the preferred-solution values of $d_{\text{cca}}$, $d_p$, and $d_t$) and the known CCSM model field: 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 4. All statistics are computed over the reconstruction interval (850-1855 C.E.).
Fig. 8. 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 5.
Fig. 9. 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.).
Fig. 10. Same as in Figure 7, but for the RegEM-Ridge reconstructions.
Fig. 11. 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.