Reviews of Geophysics

REVIEW ARTICLE
10.1029/2022RG000782

The Historical Development of Large-Scale Paleoclimate Field Reconstructions Over the Common Era

Jason E. Smerdon[1,2], Edward R. Cook[1,2], and Nathan J. Steiger[2,3]

[1]Columbia Climate School, Columbia University, New York, NY, USA, [2]Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY, USA, [3]Institute of Earth Sciences, The Hebrew University, Jerusalem, Israel

Key Points:

- Reconstructions of climate fields are valuable paleoclimatic products that have been developed since the 1970s
- Important similarities and differences exist across the approaches used to derive hydroclimate and temperature climate field reconstructions (CFRs) targeting the Common Era (CE)
- Controlled and systematic comparisons of CFRs should be a priority for the CE paleoclimate community

Correspondence to:
J. E. Smerdon,
jsmerdon@ldeo.columbia.edu

Citation:
Smerdon, J. E., Cook, E. R., & Steiger, N. J. (2023). The historical development of large-scale paleoclimate field reconstructions over the Common Era. *Reviews of Geophysics*, *61*, e2022RG000782. https://doi.org/10.1029/2022RG000782

Received 30 JUN 2022
Accepted 29 AUG 2023

Abstract Climate field reconstructions (CFRs) combine modern observational data with paleoclimatic proxies to estimate climate variables over spatiotemporal grids during time periods when widespread observations of climatic conditions do not exist. The Common Era (CE) has been a period over which many seasonally- and annually-resolved CFRs have been produced on regional to global scales. CFRs over the CE were first produced in the 1970s using dendroclimatic records and linear regression-based approaches. Since that time, many new CFRs have been produced using a wide range of proxy data sets and reconstruction techniques. We assess the early history of research on CFRs for the CE, which provides context for our review of advances in CFR research over the last two decades. We review efforts to derive gridded hydroclimatic CFRs over continental regions using networks of tree-ring proxies. We subsequently explore work to produce hemispheric- and global-scale CFRs of surface temperature using multi-proxy data sets, before specifically reviewing recently-developed data assimilation techniques and how they have been used to produce simultaneous reconstructions of multiple climatic fields globally. We then review efforts to develop standardized and digitized databases of proxy networks for use in CFR research, before concluding with some thoughts on important next steps for CFR development.

Plain Language Summary Few observational records of meteorological variables like surface temperature and precipitation extend beyond 100–150 years. The length of observational records is therefore insufficient for learning how climate varies over decades to centuries, or for estimating many climate extremes. In contrast, annually-resolved climate proxies, such as tree rings, ice cores, and corals, when used in concert with observational records, can provide information on how climate conditions have changed over decades to millennia. These proxies are also abundant enough over the last two millennia to create reconstructions in both space and time, or maps of climate conditions at seasonal or annual intervals. These kinds of reconstructions are called climate field reconstructions (CFRs) and we review their scientific history back to the 1970s when they were first attempted. We survey the underlying methods that are used to derive these reconstructions and how they have been applied, with a specific focus on the last two decades when the number of CFRs has greatly increased. We conclude with a survey of the data networks that are being cultivated for CFRs and some reflections on how the science can move forward.

1. Introduction

Fritts et al. (1971) began their seminal paper on the reconstruction of climate from tree rings with a sentence about the paucity of instrumental climate data, the sentiment of which has likely been expressed at the beginning of almost every paleoclimatic paper since that time:

> Most climatic and hydrologic records in western North America are relatively short and inadequate for assessing long-term climatic variations.

The authors went on to demonstrate the potential for tree-ring records to serve as a means of overcoming the limitations of short and sparse observational data. These now widely used and well-known paleoclimatic proxies (e.g., Cook & Kairiukstis, 1990; Fritts, 1976; Jones et al., 2009; PAGES Hydro2k Consortium, 2017) contain seasonal to annual signals related to past climatic conditions that are used to extend estimates of climate variability centuries to millennia into the past.

In many ways the Fritts et al. (1971) study was the birth of the modern field of dendroclimatology. It was an important demonstration of the paleoclimatic potential of tree-ring proxies and provided skillful estimates of past

© 2023 The Authors.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

atmospheric pressure maps over North America using a dendroclimatic data set. What is perhaps underappreciated about Fritts et al. (1971), however, is the fact that the authors presented pressure reconstructions in both space and time, which was likely the first regression-based, annually-resolved, paleoclimatic field reconstruction. The focus of this review is on the historical development of such spatiotemporal reconstruction techniques that are used to produce what are now called climate field reconstructions (CFRs). These approaches have been applied most widely to reconstruct the climate of the last 2000 years or Common Era (CE), which will be the primary timescale considered herein. In the remainder of this Introduction we provide a chronological accounting of the important CE CFR developments that followed Fritts et al. (1971). In the subsequent sections, we review CE CFR research over the last two decades, with a focus on large-scale hydroclimate and temperature reconstructions, and more recent work on paleo data assimilation (DA) products. Our goal is to provide a comprehensive assessment of work on CFRs over the last two decades. We will highlight important developments in how these products have been assessed and how their underlying methods share common challenges. We also characterize how uncertainties are quantified in CFRs, survey the data assessment efforts that are critical for producing CFRs, and finally reflect on the important next steps to develop CFRs as important tools to understand climates of the past and future.

A central challenge of many CFR problems is that they are ill posed regression problems. In other words, it is often the case that the number of spatial grid cells in the targeted climate field, which itself is characterized by a large number of spatial degrees of freedom, is much larger than the number of time steps available in the observational record that is used to relate proxy measurements to climate variables. Similarly, the number of proxies is often much smaller than the number of spatial grid cells. These factors create ill-posed conditions for traditional multivariate linear regression problems. Such cases require the addition of more information or a reduction in the amount of information that is to be estimated. Many of the early CFR approaches pursued the latter option through various forms of regularized multivariate regression, which are regression approaches that constrain the number and, in some cases, the size of the regression coefficients that are to be estimated. In the case of Fritts et al. (1971), the authors applied a regularized form of canonical correlation analysis (CCA; see Section 3 for a more detailed description) to successfully derive atmospheric pressure CFRs (although the veracity of these results would later be challenged by Kutzbach and Guetter (1980)).

Application of the CCA approach was an outgrowth of Fritts' earlier work using principal component analysis (PCA) to reduce tree-ring networks to their leading patterns of spatial variability (Fritts, 1965; Lamarche & Fritts, 1971). PCA techniques were receiving more attention at the time, given the increasing computing power that was drastically improving the ability to perform the large matrix operations associated with PCA on climate and paleoclimate data (Fritts, 1972). PCA had found earlier applications in meteorology (e.g., Holmstrom, 1963; Lorenz, 1956), but its exploration by Kutzbach (1967) was the effort that brought PCA to the broader climate research community (Bretherton et al., 1992). It was in his 1967 paper that Kutzbach also discussed the fact that PCA could be applied to multiple field variables, which could then be used to investigate spatiotemporal relationships between them—a critical conceptual element of many paleoclimatic CFR techniques.

In addition to proposing the multivariate applications of PCA, Kutzbach's work was particularly relevant for the history of dendroclimatic and other paleoclimatic work because Fritts visited the University of Wisconsin-Madison (where Kutzbach was on the faculty) in 1968–1969 as a John Simon Guggenheim Fellow to work on developing new spatiotemporal climate reconstruction methods (Fritts, 1972; Fritts et al., 1971; Leavitt et al., 2019). During their time together, Fritts and Kutzbach would specifically implement a version of CCA, which had recently been published by Glahn (1968), to perform the atmospheric pressure CFRs in Fritts et al. (1971). Fritts' visit, and the work that it engendered, was therefore the confluence of the methods and thinking that both Fritts and Kutzbach were pursuing in the mid and late 1960s.

Despite some subsequent follow up by Fritts and colleagues (Blasing & Fritts, 1976; Fritts et al., 1979, 1980) and the later publication of a book on the methods and results of the North American atmospheric pressure CFRs (Fritts, 1991), the problem of spatial reconstructions in dendroclimatic contexts was not revisited with much significance until the mid and late 1980s. Keith Briffa at the Climatic Research Unit of the University of East Anglia would explore the application of orthogonal spatial regression (OSR) as a CFR technique in his doctoral dissertation, which was ultimately published in Briffa et al. (1986). During this time, Briffa, Edward Cook, and Philip Jones would undertake a comparison of OSR and CCA in a report to the Scientific Affairs Division of the North American Treaty Organization (Cook et al., 1985), which was later published as Cook et al. (1994). OSR

and CCA share theoretical underpinnings, both of which are premised on the idea of reducing either or both of the target climate data and the tree-ring predictor data to leading spatiotemporal modes through matrix reduction before preforming multivariate regressions between the two. There are important differences between methods like OSR and CCA; the literature also contains differences in the details of methods termed CCA, principal component regression (PCR), or otherwise (e.g., Bretherton et al., 1992). For our purposes herein, however, we will simply use "reduced-space" to refer to those methods that apply some form of multivariate regression using matrix decompositions and subsequent rank reductions for either or both of the targeted climate field and the network of proxy predictors.

An important split from reduced-space approaches occurred in the early 1990s when Cook et al. (1992) laid the groundwork that would later form the basis of the point-by-point regression (PPR) CFR technique (Cook et al., 1996). The approach of this CFR method was to reconstruct a spatial climate field on a point-by-point basis, using tree-ring chronologies selected within a local search radius centered on each specific grid point in a gridded field. The motivating rationale for PPR was the observation that reduced-space multivariate regression methods were ill-suited for reconstructing the relatively small-scale patterns of spatial variability typically found in precipitation-based climate fields (Cook et al., 1999). The fundamental regression step is therefore to reconstruct a single time series at each grid point in a field, using a network of locally adjacent tree-ring chronologies that are reduced to a few leading time series using PCA. A CFR is therefore constructed by combining all of the individually reconstructed grid cells into a complete spatiotemporal grid that reflects the underlying spatial covariance of the climate field being reconstructed, as opposed to targeting some of the large-scale climate patterns in a matrix decomposition of the target field. PPR was extensively tested and used in Cook et al. (1999, 2004) to derive the first annually-resolved field reconstructions of CE hydroclimate (specifically the Palmer Drought Severity index or PDSI; W. C. Palmer, 1965) over the entirety of North America. This approach, with various methodological updates, has now been widely applied over many continental areas to estimate CFRs for growing season PDSI, including North America (Cook et al., 2004; Williams et al., 2020), South America (Morales et al., 2020), Monsoon Asia (Cook et al., 2010a), Europe (Cook et al., 2015), Europe and Russia (Cook et al., 2020) and Australia-New Zealand (J. G. Palmer et al., 2015). These large-scale hydroclimate reconstructions and their recent methodological advances are the subject of Section 2.

Parallel to the work on PPR-based hydroclimate CFRs, reduced-space methods also began to be more widely applied in the late 1990s and early 2000s. Mann et al. (1998) was the first application of reduced-space approaches to derive a global temperature CFR from a global network of multiple proxy sources. Evans et al. (2002) would similarly derive a sea surface temperature CFR over the Pacific Ocean using coral proxies (based on site-selection experiments that were performed earlier, Evans et al., 1998). Luterbacher et al. (2002, 2004) would similarly derive estimates of sea level pressure and surface temperature fields over Europe and the North Atlantic region from a multi-proxy data set using PCR. The subsequent history of these efforts, particularly regarding attempts to derive large-scale temperature reconstructions over the last millennium or longer, has been well documented (e.g., Anchukaitis & Smerdon, 2022; Christiansen & Ljungqvist, 2017; Frank et al., 2010; Jones & Mann, 2004; Jones et al., 2009; Mann, 2007; Mann & Jones, 2003; Smerdon, 2012; Smerdon & Pollack, 2016). We therefore focus in Section 3 on the most recent developments, roughly the last decade, in large-scale CFRs that have targeted temperature with reduced-space approaches.

The last 10 years have also seen rapid growth in the application of DA techniques to the CFR problem. The prospects of paleo DA were previewed by Fang and Li (2016), but the last decade of work has not been comprehensively reviewed, which is the focus of Section 4. Finally, in Section 5 we consider proxy and CFR database development. This effort has been critical for standardizing, digitizing, and making publicly available the proxy records that are used for CFRs, while subsequently making CFR products publicly available for further application and testing.

We finally note that while this review surveys the majority of methods that have been applied to the CFR problem, our methodological emphasis is roughly proportional to the extent that the various methods have been applied in the literature within the various subsections. Variants of the PPR method dominate the currently available collection of continental-scale hydroclimate CFRs, while most of the large-scale reduced-space temperature CFRs have applied various forms of multivariate regression. DA methods are now being widely applied to derive reconstructions of multiple climatic fields, including hydroclimate and temperature fields, which motivates our DA focus

later in the paper. We nevertheless also highlight, where appropriate, the source literature and findings of studies that have used other CFR methods that have been applied less commonly.

2. Hydroclimatic Field Reconstructions

The following discussion of hydroclimate CFRs (e.g., drought indices or precipitation) predominantly focuses on PPR (Cook et al., 1999), but we begin with a brief review of reduced-space methods that have targeted hydroclimate (these methods are discussed in detail in Section 3 with regard to temperature reconstructions, but we review some of their general features in this section as a comparison to the PPR approach). Reduced-space CFRs filter one or both of the spatiotemporal predictand (targeted climate field) and predictor (proxy) data sets, typically using orthogonal decompositions in the form of empirical orthogonal functions (EOFs; Lorenz, 1956), which are then rank reduced and used to perform multivariate regressions. The spatial resolution of the resulting reconstructions is therefore determined by how many EOFs are targeted for reconstruction. The reconstructed fields are spatially smoothed at a level commensurate with the number of retained target EOFs, which assumes that the small-scale spatial patterns accounting for relatively little variance in the target field are of little interest or should be discarded as noise. No universally accepted EOF truncation rule exists, although "rules of thumb" like the Kaiser-Guttman eigenvalue-1 cutoff (Guttman, 1954; Kaiser, 1960) are often used (see Preisendorfer et al. (1981) for an extensive review of EOF selection rules that can be used for objective truncation of reduced-space applications).

There has been some success reconstructing hydroclimate fields using reduced-space multivariate CFR methods. Stockton and Meko (1975) reconstructed drought at multiple locations over the western United States (US) using the Fritts et al. (1971) CCA method; Diaz and Wahl (2015) have more recently produced a precipitation CFR over the western US also using a truncated EOF approach. Pauling et al. (2006) reconstructed precipitation over Europe using a CFR method similar to OSR, and Neukom et al. (2010) reconstructed summer and winter precipitation over southern South America from an ensemble of reduced-space CFRs. Cook et al. (1994) directly compared CCA and OSR for reconstructing PDSI and found both methods to be equally skillful at the levels tested. Regularized Expectation Maximization (RegEM; T. Schneider, 2001) was also tested using the same data used in Cook et al. (1999) to produce PDSI reconstructions over the conterminous US (Zhang et al., 2004). The application of RegEM as the regression approach within PPR was found to provide modestly improved verification skill over some US areas, but not others, compared to PPR using PCR (verification skill in this context refers to statistical measures of similarity between the derived CFR and the data withheld from the calibration—see the discussion in Box 1). RegEM was only successfully applied, however, after the PDSI field itself was separated into eight relatively homogenous regions of summer drought through a cluster analysis.

BOX 1. CFR skill metrics

Throughout this review we discuss reconstruction skill as a measure of the robustness and accuracy of derived reconstructions. There are multiple ways in which skill is quantified within CFRs. The traditional approach uses observational data for calibration of a regression model, while withholding some data for validation after the reconstruction is performed. Historically this has been done by using the most recent half to two-thirds of the observational data for calibration, while using the remaining data for validation. More recent practices use leave-one-out cross-validation schemes, or leave-X-out schemes, in which X represents some multiple of the sampling interval of the observational data (e.g., Williams et al., 2020). Such schemes perform many reconstructions by employing iterative calibration-validation configurations and then report mean or median calibration and validation statistics for the ensemble of results. Whatever the approach, these intervals are the basis of calibration and validation statistics that are traditionally reported for reconstructions, each with critically different implications for the interpreted performance of a reconstruction. This performance is interpreted as skillful if each statistic calculated within the calibration and validation intervals exceeds a given skill threshold. It is also important to be mindful of the fact that data quality and availability change as a function of time within the observational data. This is particularly relevant for approaches that use static intervals for calibration

and validation, which may, for instance, use the sparsest data with the greatest uncertainties for the validation interval (perhaps with good reason because it is advantageous to calibrate reconstruction models on the highest quality data). Data quality also varies differently for different climate variables. For instance, the reduction of data quality back in time is more acute for hydroclimatic variables (e.g., Cook et al., 2010b; Stahle et al., 2020) than temperature fields that contain higher spatial covariance information, though declining instrumental data quality can limit spatial skill assessments for CFRs of all kinds.

Calibration statistics present a measure of the goodness of fit of a reconstruction model and are often reported as the coefficient of determination (R^2), which yields an estimate of the variance resolved in the target data, or as the root-mean-squared error, which assesses both the variance resolved and the mean bias of the reconstruction (see Table 1 and Figure 4 as an example of these statistics). Table 1 also reports a leave-one-out cross-validation statistic (CVRE). CVRE is the R^2 equivalent of the Prediction Error Sum of Squares (PRESS) statistic (Allen, 1974; Quan, 1988) and, thus, provides a less biased estimate of explained variance compared to R^2. In poorly calibrated cases, CVRE can actually be negative, which is impossible for R^2. Calibration measures provide an estimate of the robustness of the in-sample regression results, but are notoriously subject to overfitting. As such, validation statistics, or out-of-sample measures, are critical for characterizing the skill of reconstructions. Common validation statistics are again the R^2 value calculated over the validation interval, the reduction of error (RE; Cook et al., 1994, 1999), and the coefficient of efficiency (CE; Cook et al., 1994, 1999). When validation R^2, RE, and CE are positive, they are different measures of reconstruction skill expressed in units of fractional explained variance over the validation period. Negative values indicate no reconstruction skill as measured. Additionally, the formulation of these statistics requires that RSQ $\geq$ RE $\geq$ CE when calculated from the same data in the validation interval, thus making CE the hardest verification statistic to pass (see Cook et al., 1994, 1999 for detailed descriptions of these model verification statistics).

An additional skill score in the DA literature is the continuous ranked probability skill score (CRPSS), which is based on the continuous ranked probability score and is further discussed in Steiger and Smerdon (2017). This score functions like a correlation statistic but accounts for the skill of the entire posterior reconstruction distribution. This feature raises an important consideration for ensemble methods in which each posterior reconstruction is considered equally probable. Skill estimates that evaluate the success of the entire posterior distribution are therefore important beyond reporting, for instance, the skill of the ensemble mean reconstruction.

We finally address the difference between formal skill estimates for real-world reconstructions, as we have discussed above, and pseudoproxy experiments that have been important mechanisms for interpreting how CFRs perform, but are not used directly in formal skill evaluations of CFR results. For instance, pseudoproxy experiments have been used to determine that CFRs perform best in regions with dense pseudoproxy sampling or to characterize how the covariance structure of an underlying field can impact spatial CFR skill (Kutzbach & Guetter, 1980; Smerdon, 2012; Yun et al., 2021). It also has become customary to perform pseudoproxy experiments to test new CFR methods before they are applied to real-world data. These experiments therefore have come to be employed as useful complements to traditional skill assessments of CFRs, but are insufficient individual measures of reconstruction skill for real-world CFRs; nor have they been formally combined in any quantitative measure with traditional skill measures.

Another regression-like CFR method used to spatially reconstruct past drought is based on neural networks (D'Odorico et al., 2000; Ni et al., 2002; Nicault et al., 2008; Woodhouse, 1999). A recent study by Trevino et al. (2021) is the most relevant to this review, which used a nonlinear neural network as an explicit nonlinear alternative to PCR for the pointwise reconstruction of gridded PDSI across the contiguous US (in this sense the method is not reduced-space because it is a point-by-point approach, but we include it here as another method different from the more traditional PCR-based PPR method). These reconstructions were then compared to those produced by a version of PCR that closely followed the method described in Cook et al. (1999). In the majority of cases, the nonlinear effects were found to be minor or undetectable. Similar to RegEM, nonlinear neural network

verification skill modestly increased at some locations, but not at others, compared to PCR. In the southwestern US, however, there did appear to be meaningful nonlinear effects modeled by the nonlinear neural network, which are missed by linear PCR. While this suggests some utility in applying nonlinear neural networks as a pointwise CFR method, it is much harder to implement and interpret compared to PCR. Nonlinear relationships can also be modeled in the current version of PPR using an adaptive power transformation for reconstructing seasonal precipitation (Stahle et al., 2020).

In contrast to the reduced-space CFR methods, PPR explicitly acknowledges the importance of small-scale spatial variability that is commonly found in hydroclimate fields because of the relatively localized properties of precipitation (Cook et al., 1999). Large-scale patterns associated with PPR CFRs can be interpreted as emergent properties of the underlying field, as opposed to patterns that have been explicitly targeted by construction in reduced-space methods. The emergent PPR patterns are preserved because each PPR model is calibrated against the actual data at each grid point, which in turn indirectly preserves the spatial covariance of the actual field at scales larger than the grid-point resolution itself, assuming the available grid points sufficiently sample the large-scale covariance patterns. As a test, the reconstruction of summer PDSI across the contiguous US was made using PPR (Cook et al., 1999) and compared to the spatiotemporal analyses of gridded instrumental PDSI across the same domain by Karl and Koscielny (1982). In all conducted tests, the gridded PPR reconstructions based on tree rings matched the larger spatiotemporal properties of PDSI reported by Karl and Koscielny (1982), thus demonstrating the ability of PPR to preserve the spatiotemporal variability of the overall field along with acceptable levels of reconstruction skill over time. In a more recent analysis, Baek et al. (2017) demonstrate that the large-scale hydroclimatic teleconnection patterns associated with the El Niño-Southern Oscillation, Atlantic Multidecadal Variability, and the North Atlantic Oscillation phenomena were well represented in the collection of Northern Hemisphere (NH) drought atlases based on comparisons with the equivalent patterns in observational data.

The approach of PPR also does not require a regularization in the sense that the regression performed at each grid cell is well posed, that is, a single time series is reconstructed from a collection of principle components estimated from one or more tree-ring series identified within a search radius centered on a given grid point. The information is nevertheless smoothed in the sense that a heterogeneous network of tree-ring chronologies is used to estimate hydroclimate conditions in a spatial grid that is not evenly sampled by the locations of the tree-ring sites. Information in the tree-ring network is therefore recombined in space at a length scale that is determined by the size of the search radii and the degree to which the search radii exceed the grid spacing of the field being reconstructed. The latter imparts additional local smoothing to the field. Confirming the robustness of the large-scale patterns, as discussed above, is therefore important for PPR, given that any given grid point will be predicted by a recombination of the information sampled in a proxy network, with ambiguous implications for the representation of large-scale patterns.

2.1. The PPR Method

There are four general steps in the production of CFRs using PPR, which will be discussed below in the context of a PDSI reconstruction using dendroclimatic predictors extracted from Cook et al. (2004). Most PPR applications have used dendroclimatic proxy data as predictors, but there is nothing specific to the PPR methodology that excludes other proxies, as long as they have temporal resolutions equivalent to the calibration data and sufficient overlap with the instrumental interval to allow for robust statistical calibrations. The following steps also equally apply to PPR CFRs of other hydroclimate metrics, such as precipitation (Stahle et al., 2020).

Step 1. An initial search radius (e.g., 450 km; Cook et al., 1999) around each grid point is used to locate a specified minimum number of predictors for reconstructing hydroclimate (see Figure 1). A typical number of selected predictors ranges from 5 to 20, depending on the density of the tree-ring network, the resolution of the target grid, and the size of the search radius. If the specified minimum number of chronologies is not found due to the variable density of the tree-ring network, the search radius is dynamically expanded by 50-km increments until the minimum number is met or exceeded. This allows PDSI reconstructions to be made in areas with fewer than the minimum number of tree-ring chronologies found within the initial search radius.

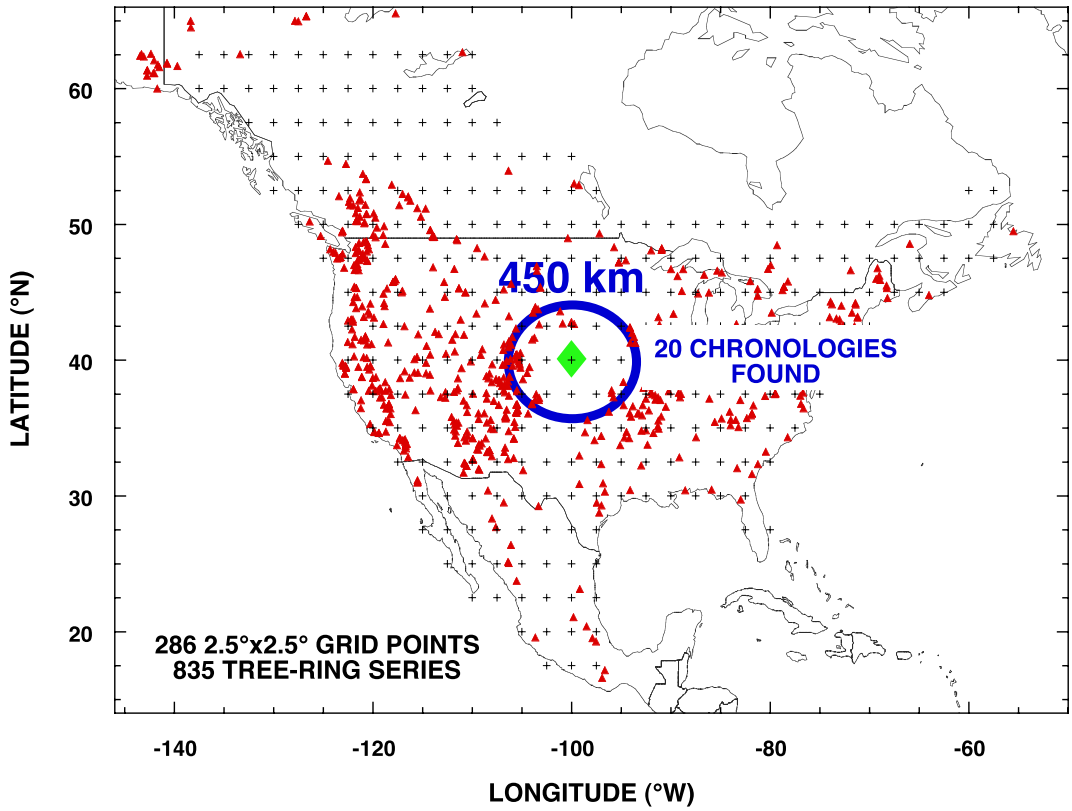

Figure 1. Map showing an example 450-km search radius applied to reconstructing the Palmer Drought Severity Index over North America using point-by-point regression (PPR) and the number of tree-ring chronologies found around the "green diamond" grid point (red triangles indicate locations of tree-ring chonologies). These "Level-1" chronologies are the initial candidate predictors used in PPR.

The search radius in PPR was initially designed to limit the tree-ring predictors used in the reconstructions to ones that were proximal to the grid point in question and thus likely to have sampled hydroclimate at the given grid location (Cook et al., 1999). The "optimal" search radius to use in PPR for hydroclimate reconstruction was initially not apparent, but Cook et al. (1999) found that the best search radius for reconstructing summer PDSI across the contiguous US was 450 km. Somewhat fortuitously, this optimal search radius was effectively the same as the correlation decay distance estimated globally for precipitation (I. Harris et al., 2020). In contrast, mean 2-m surface air temperature was estimated by I. Harris et al. (2020) to have a correlation decay distance of 1,200 km. Because air temperature plays an important role in the calculation of PDSI via estimates of evapotranspiration, the correlation decay distance for PDSI might therefore be greater than that for precipitation alone. This is exactly what was found by Cook et al. (2015) when estimating correlation decay distance for PDSI over Europe, where the average distance was estimated to be 800 km. But these estimates are only averages with considerable geographic variability. This variability is tied to multiple regional features, including latitude and orography, as well as large-scale features that may be dependent on coupled atmosphere-ocean modes like the El Niño-Southern Oscillation phenomenon or monsoon systems, or large-scale radiative forcings that drive precipitation or temperature changes (e.g., PAGES Hydro2k Consortium, 2017; Sarachik & Cane, 2010). There is consequently no single optimal search radius to use in PPR.

Step 2. Choose a calibration period for estimating the reconstruction regression models. Cook et al. (1999) used 1928–1978, but the chosen interval may vary according to the time period of overlap between the tree-ring chronologies and the PDSI field. The correlation between the tree-ring chronologies and PDSI (e.g., summer average) over the calibration period is then calculated and only those tree-ring chronologies that correlate at, for example, the 90% 2-tailed significance level are retained as candidate predictors (Cook et al., 1999). This "Level-2" screening step usually reduces the size of the predictor pool from that found by

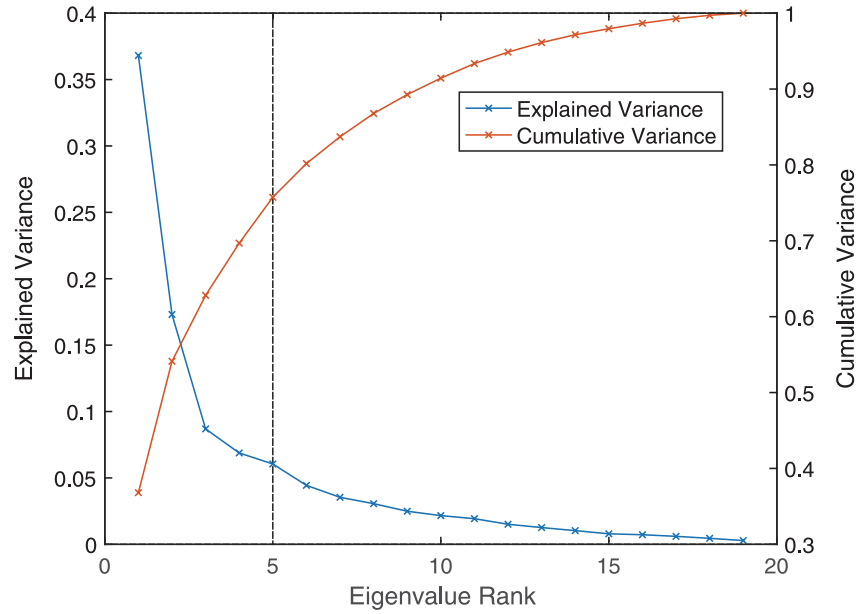

Figure 2. Results of principal component analysis applied to the retained 19 chronologies found by Level-2 screening over the 1928–1978 calibration period. The explained variance of each eigenvalue is shown (blue), as is the cumulative variance of the eigenvalues through a given rank (red). The eigenvalue-1 cutoff criterion was used to select the subset of principal components to use as candidate predictors in the regression. That number is 5 (dashed line in figure), which accounts for 75.7% of the cumulative variance. This is the "Level-3" screening of point-by-point regression.

the search radius alone (e.g., shown in Figure 1, 19 of the 20 chronologies in the search radius passed this level of screening). If fewer than the minimum number of chronologies, but more than zero chronologies are available in a given search radius, the radius is increased until the number of screened chronologies is equal to or greater than the minimum threshold. In rare cases in which there are no available chronologies in an initial search radius, the radius is increased until at least one chronology is available. This screening serves to diminish non-climate related noise in the predictors for the next step, but there is no a priori optimal screening probability and inadvertent loss of signal may occur for a chosen screening level (a matter addressed later in this section).

Step 3. PCA is applied to the retained chronologies that pass Step-2 screening. See Figure 2 for the PCA results for the 19 tree-ring predictors that passed Level-2 screening in the search-radius area shown in Figure 1. PCA is applied over the calibration period (here 1928–1978) to guarantee the orthogonality of the resulting principal component predictors, with their scores extended back in time over the time period common to all retained chronologies. Only those principal components with eigenvalues >1.0 (Kaiser-Guttman Rule) are retained as candidate predictors in PCR, which in the case of the example shown in Figure 2 was five, resulting in a retained cumulative variance of 75.7%.

Step 4. Enter the orthogonal tree-ring principal components from Step 3 into the linear regression model by order of explanatory variance. The orthogonality of the predictors produced by PCA over the calibration period greatly streamlines the stepwise process of adding principal components to the regression model by making it strictly additive. The number of predictors and, thus the final regression model, is estimated using the minimum Akaike Information Criterion (Akaike, 1974; Hurvich & Tsai, 1989) (other information theory criteria may also be used). Note that not all retained principal components are necessarily used as predictors—only those that have enough explanatory power to reduce the Akaike Information Criterion to offset its penalty function, and they are frequently not retained according to the amount of principal component variance accounted for in Figure 2. See Table 1 for full regression results based on the five retained principal components as candidate predictors of PDSI. The final regression model is then used to reconstruct hydroclimate back in time as far as the tree rings allow and the reconstructions are tested over the withheld data verification period (here 1900–1927) for reconstruction skill and uncertainty estimates (See Box 1).

Reviews of Geophysics

10.1029/2022RG000782

Table 1
The Principal Component Regression Model Results

Step-4 principal component regression model								
Calibration period: 1928–1978								
STEP	VAR	PCVAR	CORR	TSTAT	CRSQ	CUM	CVRE	AIC
1	1	36.8	0.725	7.37	0.526	0.526	0.521	−33.78
2	3	8.7	0.222	1.59	0.049	0.575	0.557	−37.11
3	2	17.3	0.222	1.59	0.049	0.624	0.596	−41.02
4	5	6.1	−0.216	−1.55	0.047	0.671	0.639	−45.32
5	4	6.9	0.200	1.43	0.040	0.711*	0.670*	−49.38*
Minimum AIC regression model order: 5% variance explained: 71.1%								
Verification period: 1900–1927								
STEP		VAR		RSQ		RE		CE
1		1		0.379		0.474		0.358
2		3		0.251		0.351		0.207
3		2		0.361		0.462		0.343
4		5		0.431		0.524		0.419
5		4		0.598*		0.642*		0.562*

Note. The principal components were ranked by order of explanatory variance and entered sequentially until the minimum Akaike Information Criterion * was satisfied. That order is five, explaining 71.1% of the Palmer Drought Severity Index variance. The verification period statistics indicate a high degree of skill in the reconstruction, for example, CE = 0.562.

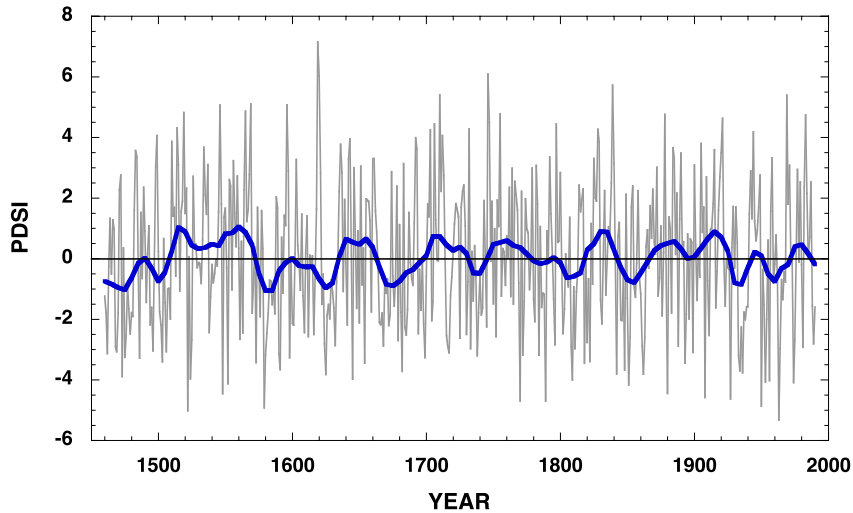

Figure 3. The annual (gray) and decadal lowpass filtered (blue) Palmer Drought Severity Index (PDSI) reconstruction of a single grid point (green diamond in Figure 1) example using point-by-point regression. The reconstruction covers the period 1460–1990, with the 1979–1990 data being instrumental PDSI to provide an update of PDSI variability beyond that available from tree rings.

The PPR PDSI reconstruction from Steps 1–4 extends back to 1,460 (Figure 3). Note that the reconstruction is longer than the shortest tree-ring series beginning in 1743. This reconstruction uses nested PPR in order to use the maximum lengths of the tree-ring chronologies for reconstruction (Cook et al., 2004; Meko, 1997). The process of nesting produces reconstructions in each interval, or nest, for which there are the same number of available proxies (the number decreases back in time), and then combines nests by using the reconstruction from the most populated nest as the estimate in any given year. Finally, what is shown thus far are the results for one grid point reconstruction. PPR then repeats the above four steps to reconstruct each PDSI grid point across the entire field.

PPR produces calibration and verification statistics (see Box 1) for all grid points in the reconstructed field, allowing the production of maps of the spatial skill distributions (Figure 4). It is apparent that the highest reconstruction skill is where the tree-ring network is most dense (cf. Figures 1 and 4), a characteristic also widely noted in reduced-space CFR and paleo DA approaches (Smerdon, 2012; Smerdon et al., 2011, 2016; Steiger & Smerdon, 2017; Steiger et al., 2014; Wang et al., 2014). Declining instrumental data quality in the early part of the 20th century in Canada and Mexico nevertheless also plays a role in areas where skill metrics are reduced (cf. Figure 2 in Stahle et al. (2020)). The same problem was found over Asia prior to 1950 in the construction of the Monsoon Asia Drought Atlas (MADA; Cook et al., 2010a). Thus, the indicated skill provided by the verification tests are at a certain level conditional on the quality of the withheld instrumental data used for that purpose. This issue was pointed out by Wahl and Morrill (2010) in their commentary on the MADA publication. They noted that the reconstructed patterns between historically documented droughts over eastern China and the MADA was much better than the verification statistics over the same region. Similar results are reported in Morales et al. (2020) for the South American Drought Atlas.

2.2. Recent PPR Advances

Since the development of PPR for producing drought atlases over North America, considerable advances in the methodology have been made. This was necessitated by the expansion of drought atlas development to other continents of the world with far different tree-ring networks and PDSI data quality than found over North America. The first modification of PPR was the incorporation of multiple search radii in the development of the MADA. Compared to North America, the tree-ring network over Monsoon Asia was relatively sparse with large gaps in its distribution. For this reason, experiments were conducted with four search radii: 500, 1,000, 2,000, and 3,000 km. The 500-km radius closely matched that used in North America and the estimated correlation decay distance for precipitation. The larger search radii accommodate the reality of working with the irregular and spatially incomplete MADA tree-ring network, relative to the regular PDSI grid used as a target. The larger search radii nevertheless are in contrast to the original design rationale of PPR, while introducing important evaluation requirements like assessments of teleconnection stability, which dictates the far-field signals between the proxies and the targeted grid points.

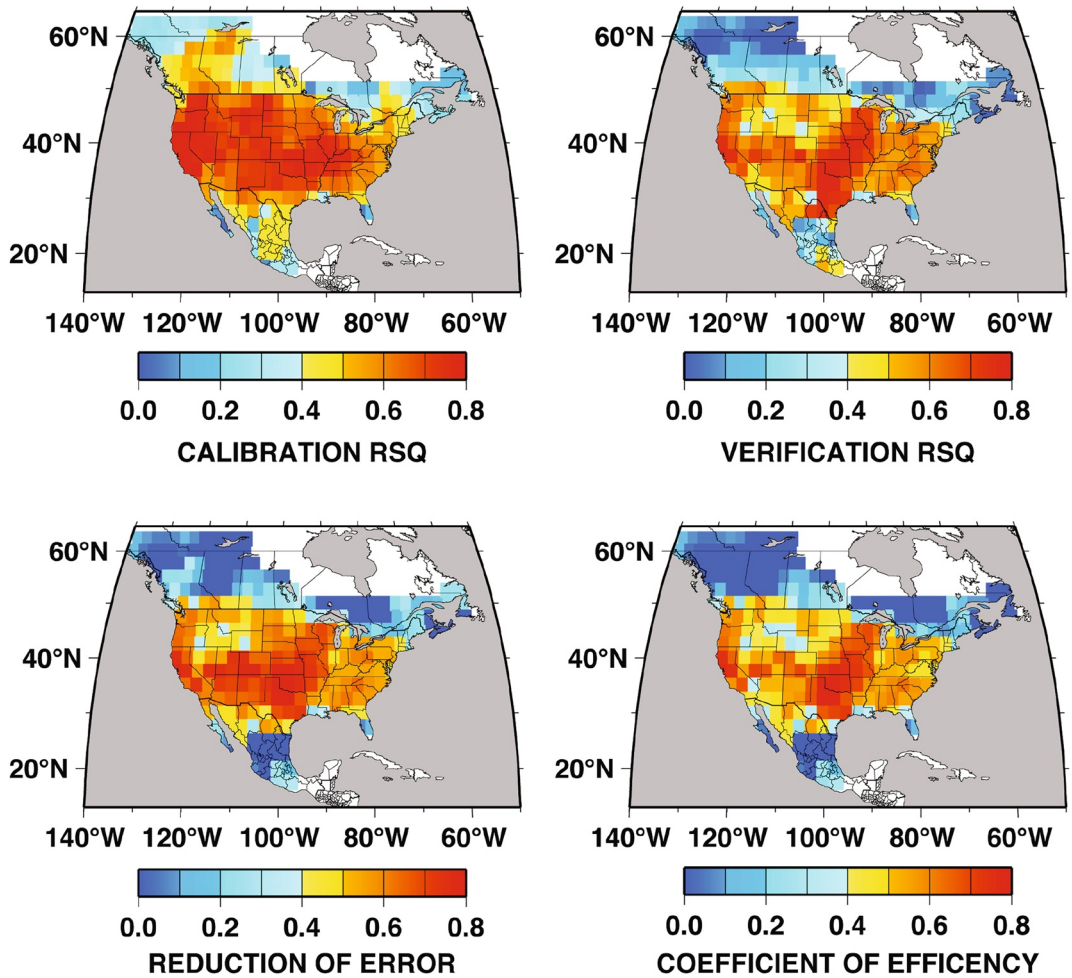

Figure 4. Maps of point-by-point regression calibration and verification skill for the North American Drought Atlas (Cook et al., 2004), as measured by the coefficient of determination (RSQ), the reduction of error, and coefficient of efficiency (see discussion in Box 1). Values of the latter two range from $-\infty$ to 1 and are considered skillful for values greater than zero. Note that the single grid-point reconstruction shown in Figure 3 is included in the above maps.

The second modification of PPR used in the MADA was the elimination of a fixed correlation threshold for screening tree-ring predictors to be used in regression modeling and reconstruction of PDSI. Over North America, screening at the 90% significance level was successfully used, but a fixed screening threshold potentially loses some useful tree-ring signal if, for example, a tree-ring chronology that correlated with PDSI at the 89% level was rejected. On the other hand, keeping all series within a given search radius as predictors without any screening might allow too much noise to enter into the PCR model. Thus, the decision was made to produce a suite of PPR field reconstructions using a correlation-weighted approach:

$$wTR = uTR * |r|^p$$

where uTR is the unweighted tree-ring chronology in normalized ($N[0, 1]$) form over the calibration period, $|r|$ is its absolute correlation with the climate variable being reconstructed over the calibration period, p is a power applied to r, and wTR is the resulting correlation-weighted chronology. This weighting method transforms the correlation matrix of tree-ring predictors used in PCR analysis into a covariance matrix (for all but $p = 0$), which emphasizes the more heavily weighted (better correlated with climate) tree-ring series in the regression model. There is nevertheless no a priori reason why any particular power weighting should be optimal. Thus, a range of powers are typically used: [0, 0.1, 0.25, 0.5, 0.67, 1.0, 1.5, 2.0] per search radius. The functional forms of these power transformations are shown in Figure 5, yielding results that either weight all chronologies equally ($p = 0$) or heavily weight chronologies with high correlations over weakly correlated series that play little role in the resulting reconstruction ($p = 2$).

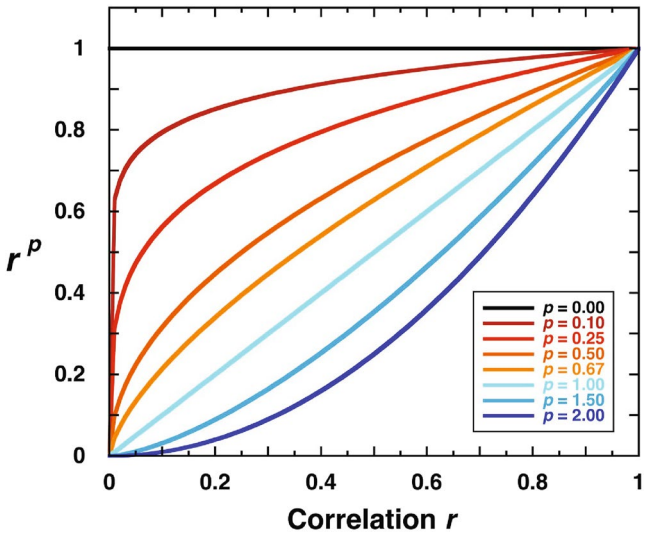

Figure 5. The functional forms of the power transformation option, applied to tree rings in point-by-point regression based on their correlations with climate over a range of correlation (r) and power (p) values.

The combination of multiple search radii and power weightings of chronologies establishes an ensemble PPR method that yields an ensemble of drought atlases estimated from the same data. Beginning with the MADA in 2010 and until recently in the South American Drought Atlas (Morales et al., 2020), the final ensemble PPR drought atlas has been the average of all ensemble members with full calibration and verification statistics re-estimated for the ensemble mean. The averaging process serves to dampen the noise associated with reconstructions based on different search radii and power weightings (e.g., Cook et al., 2010b, 2020; Morales et al., 2020; J. G. Palmer et al., 2015). This process does not necessarily result in better verification skill, but the mean of the ensemble PPR is likely to produce a more robust drought atlas.

The most recent modification of ensemble PPR is based on results from Williams et al. (2020), who found that the optimal search radius varied geographically based on differences in orography and other geographic variables, which was well expressed in the validation skill. For example, reconstructions in more mountainous areas or areas with strong rain shadows often had higher skill when a smaller search radius was used. Conversely, reconstructions in larger, flatter, more homogeneous areas generally performed better using a larger search radius. Additionally, different power weightings applied to the same tree-ring data found in a given search radius had variable effects on verification skill that were not predictable—sometimes better, sometimes worse. Thus, there is likewise no optimal power weighting, but the spatial patterns of the power weightings are expected to uniquely depend on least three things: (a) the underlying spatial structure of the predictor data set, (b) the spatial covariance of the target field, and (c) the selected predictors for the regression at each point. The dependencies of both the optimum search radius and power weighting are expressed in the ensemble PPR maps (Figure 6) for NH PDSI reconstructions, where the highest

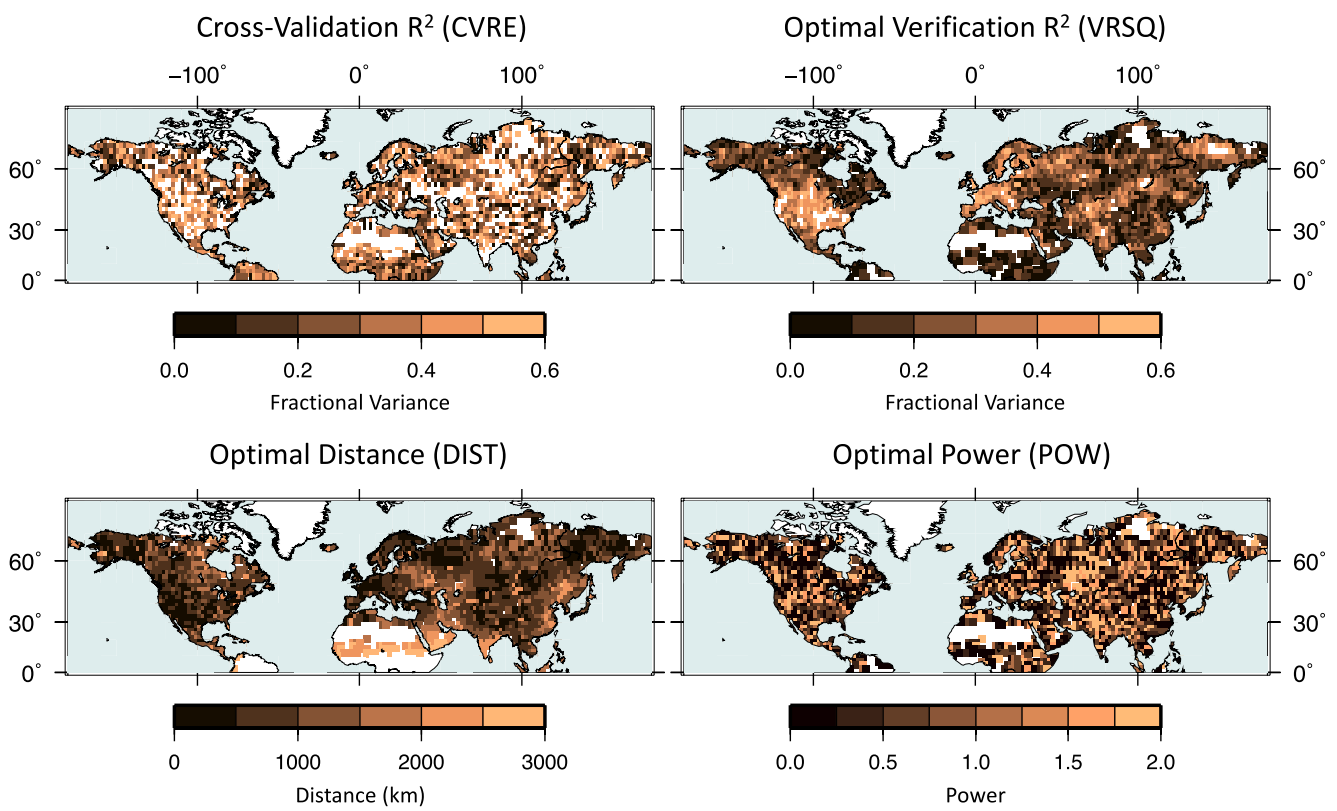

Figure 6. Maps of cross-validation R^2 (CVRE), optimal out-of-sample verification R^2 (VRSQ), optimal distance (DIST) and optimal power weighting (POW).

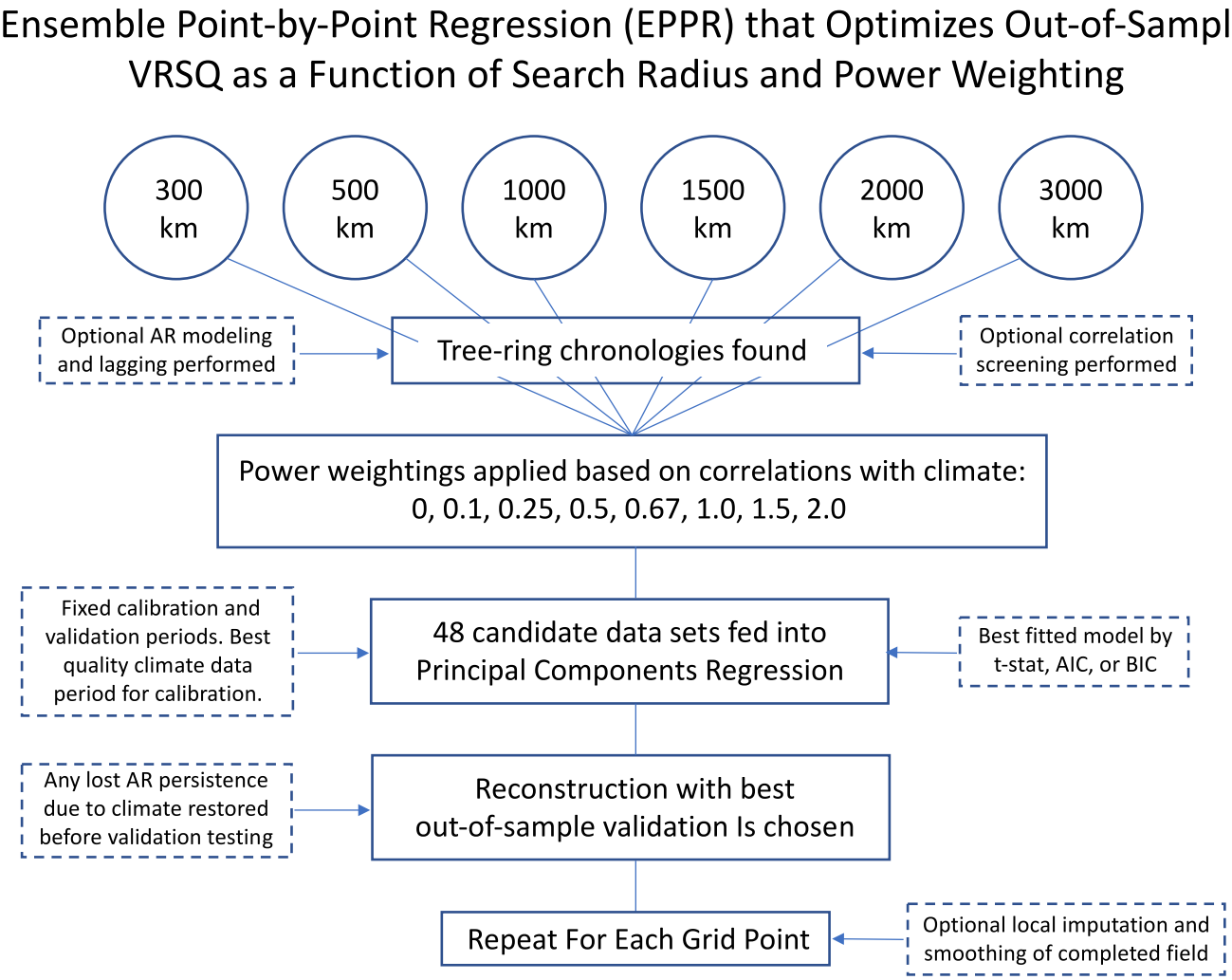

Figure 7. A flowchart of the ensemble point-by-point regression (PPR) method described in this paper (progression is from the top to the bottom of the flow chart). Note that the last optional local imputation and smoothing step is described in detail in Cook et al. (2020).

out-of-sample verification R^2 skill has been used to select the optimal distance and power values. The calibration period cross-validation R^2 is simply that associated with optimal verification R^2.

All of the above-described steps of the ensemble PPR method are represented in Figure 7. These are the most up-to-date collections of choices as described and implemented in Cook et al. (2020) and Williams et al. (2020, 2022). As demonstrated by how these choices were added over the course of the development of multiple drought atlases, the application of the decisions for each of the possible choices is dictated by the nature of the region and data that are incorporated by ensemble PPR.

2.3. PPR Uncertainty Estimation

A final note is necessary regarding PPR uncertainty estimation, which provides measures of both reconstruction *accuracy* and *precision*. Measures of reconstruction *accuracy* are provided by PPR in the form of correlation, reduction of error, and coefficient of efficiency statistics (Cook et al., 1994, 1999), which are estimated by comparing actual and reconstructed data not used for model calibration (Box 1). These statistics have a long history of use in dendroclimatology going back to Fritts (1976), but do not provide estimates of *precision*, that is, how much uncertainty exists around each reconstructed value. Rather, they are simply measures of the similarity between reconstructed values and the actual data not used for calibration. But we also wish to know how much uncertainty exists around each reconstructed value to determine the significance of changes in the reconstruction. This evaluation is thus associated with the precision of the reconstruction.

Because PPR is a regression-based method, estimates of reconstruction uncertainty (precision) are expressed in the form of "prediction intervals" (Olive, 2007; Seber & Lee, 2003) around the reconstructed value. These prediction intervals are based on formulations that typically incorporate both the lack of fit in the form of the mean-square error of the estimate and a "leverage" term that can quantify additional uncertainty associated with the degree to which the reconstructed value is an interpolation consistent with the data used for calibration or an extrapolation that falls outside the calibration data range (Olive, 2007). This measure of uncertainty is very useful, but it has important limitations. Note that the uncertainty only changes as much as the leverage estimate changes over time, but the calibration period mean-square error remains constant. If the calibration period mean-square error is an unbiased estimate of true model skill, it may not matter. But this cannot be assumed true because regression models almost always overfit the calibration period data. The uncertainty estimate is therefore very likely to underestimate the true uncertainty of the reconstruction outside the calibration interval.

To get around the above limitation, computationally intensive bootstrapping methods can be applied. One method that shows promise is the maximum entropy bootstrap (Vinod, 2006), which is unique among bootstrap methods in the way that it preserves the overall properties of time series data, including persistence and trend. This is very difficult to do using block-bootstrap methods. As applied in PPR, the maximum entropy bootstrap option provides more data-driven uncertainties that are wider than the prediction intervals described above, vary in width over the full reconstruction, and can also be asymmetric, which is not possible if based on traditional prediction intervals (Cook et al., 2013). Additionally, maximum entropy bootstrap prediction intervals have been shown to be comparable to Bayesian credible intervals (Rao et al., 2018). Like all resampling methods, however, it is not well suited for very large ensemble PPR reconstruction problems with many thousands of grid points because of its computational expense. The maximum entropy bootstrap therefore has not yet been used in drought atlas development, but it remains part of the overall PPR methodology nonetheless.

3. Temperature Field Reconstructions

Despite early attempts to reconstruct hemispheric mean surface temperatures over the last millennium (Groveman & Landsberg, 1979; Jacoby & Darrigo, 1989; Landsberg et al., 1978), the publications of Mann et al. (1998), Jones et al. (1998), and Mann et al. (1999) generated significant interest in the possibility of deriving hemispheric and global temperature reconstructions over the last several millennia; surface temperature reconstructions from inversions of temperature-depth profiles measured in terrestrial boreholes were also emerging in the late 1990s (Beltrami, 2002; R. N. Harris & Chapman, 2001; Huang et al., 1996, 2000; Pollack & Huang, 2000; Pollack et al., 1998) and similarly fed excitement about the possibility of robust estimates of last-millennium temperature changes (e.g., Broecker, 2001; Pollack & Chapman, 1993). The evolution of research on large-scale temperature reconstructions since these early publications has been well chronicled elsewhere (Anchukaitis & Smerdon, 2022; Christiansen & Ljungqvist, 2017; Frank et al., 2010; Jones & Mann, 2004; Jones et al., 2009; Mann, 2007; Mann & Jones, 2003; Smerdon, 2012; Smerdon & Pollack, 2016). For the purposes of this review, however, it is important to point out that Mann et al. (1998) was a CFR, while very few CFRs were actually published as part of the larger debate in the 2000s about global and hemispheric mean temperature reconstructions (Smerdon & Pollack, 2016). Rutherford et al. (2005) would publish a CFR based on the ridge regression formulation of RegEM (RegEM-Ridge) using multiple input data sets, including the proxy network from Mann et al. (1998). Mann et al. (2009) would later use an alternative version of RegEM (based on truncated total least squares or RegEM-TTLS) to produce an updated CFR based on an expanded multiproxy data set. But most of the research on CFRs during the 2000s was methodological in nature (see Tingley et al. (2012) for a review), and thus largely avoided the production of CFRs using real-world proxy data (Smerdon, 2012). It has only been over the last decade that more large-scale temperature CFR estimates have been produced. These developments will be reviewed in due course, but we first review in more detail the methodological basis of most approaches to reconstructing large-scale temperature CFRs, as they were pursued through the end of the first decade of the 21st century.

3.1. Reduced-Space CFR Methodology

Most of the CFR methods applied in the first decade of the 21st century were variations on multivariate linear regression (Christiansen et al., 2009; Tingley et al., 2012). In the simplest of terms, these methods relate a matrix of climate proxies to a matrix of climate data during a common time interval (generally termed the calibration

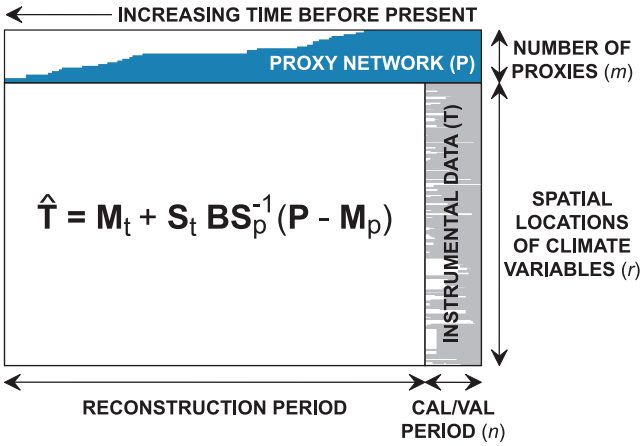

Figure 8. Schematic representation of the data matrix for large-scale temperature climate field reconstructions. See text for the variable definitions. White space in all sections of the data matrix represent missing data.

interval) using a linear model (Figure 8). If $\mathbf{P}$ is an $m \times n$ matrix of proxy values and $\mathbf{T}$ is 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 time of overlap between the proxy and instrumental data, the linear relationship is written,

$$\mathbf{T}' = \mathbf{B}\mathbf{P}' + \varepsilon,$$

where $\mathbf{B}$ is a matrix of regression coefficients with dimensions $r \times m$, ε is the residual error and the primes indicate that the $\mathbf{T}$ and $\mathbf{P}$ matrices have been centered and normalized over the time interval of n. According to standard linear regression theory, the error variances of all the elements in ε are simultaneously minimized if $\mathbf{B}$ is chosen as

$$\mathbf{B} = \left(\mathbf{T}'\mathbf{P}'^{\mathrm{T}}\right)\left(\mathbf{P}'\mathbf{P}'^{\mathrm{T}}\right)^{-1},$$

where the superscript T denotes a matrix transpose. Temperatures prior to measurement, $\hat{\mathbf{T}}$, thus can be inferred, or "reconstructed," using the regression matrix $\mathbf{B}$ during periods in which proxy data are available but observed temperatures are not:

$$\hat{\mathbf{T}} = \mathbf{M}_t + \mathbf{S}_t\mathbf{B}\mathbf{S}_p^{-1}\left(\mathbf{P} - \mathbf{M}_p\right),$$

where the $\mathbf{M}$ and $\mathbf{S}$ matrices correspond to the respective means and standard deviations of the temperature (subscript t) and proxy (subscript p) data that were removed to perform the regression, and the temporal dimension of $\mathbf{P}$ is the full duration of the proxy network (or nest), that is, greater than n. This formalism is represented in Figure 8 in schematic form.

While the above-described 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. The challenge for CFR methods involves the manner in which $\mathbf{B}$ is estimated in practical situations when this condition is not met. It is often the case in hemispheric or global CFRs that the number of target variables exceeds the time dimension, yielding an ill-posed estimation problem (this is to be contrasted with the PPR approach that does not suffer from this challenge; see discussion at the beginning of Section 2). In such cases, the cross-covariance and covariance matrices in $\mathbf{B}$ cannot be well estimated. Some form of regularization is therefore necessary to apply additional constraints on the problem, which is a well-established statistical method for solving ill-posed estimation problems and reducing the degree of over-fitting. Forms of regularization are widely discussed in the literature and represent a continuing area of statistical and computational research (Fierro et al., 1997; Hoerl & Kennard, 1970; Tikhonov & Arsenin, 1970; Van Huffel & Vandewalle, 1991).

A detailed technical review of these regression challenges specifically in the context of the paleoclimate reconstruction problem is provided by Tingley et al. (2012) and other cogent discussions have been provided by multiple authors (Christiansen et al., 2009; Cook et al., 1994; Guillot et al., 2015; Smerdon, Kaplan, Chang, & Evans, 2010; Wang et al., 2014, 2015). For the purposes of our discussion herein, it is sufficient to summarize many of the methodological debates about regression-based large-scale CFRs as hinging on three choices that are aimed at (a) aimed at limiting and/or constraining the information that is to be reconstructed, (b) enhancing the signal against the background noise in the proxy network, and (c) better estimating the covariance between the proxy and temperature data to derive $\mathbf{B}$ for the reconstruction. To represent these choices formally, consider the rank-reduced versions of the matrices $\mathbf{T}'$ and $\mathbf{P}'$ using singular value decomposition (SVD):

$$\mathbf{T}^r = \mathbf{U}_t^r\Sigma_t^r\mathbf{V}_t^{r\mathrm{T}},$$

and,

$$\mathbf{P}^r = \mathbf{U}_p^r\Sigma_p^r\mathbf{V}_p^{r\mathrm{T}},$$

where the superscript r denotes the reduced rank versions of $\mathbf{T}'$ and $\mathbf{P}'$. The selection of rank for each of these matrices underlies the first two choices, that is, whether the target temperature field should be reduced to leading

patterns of covariance and how many patterns should be retained, and whether the proxy matrix should similarly be reduced to leading principal components indicative of shared covariance in the proxy network. Regarding the first choice, it is critical to note that the derived CFRs will never exceed the amount of information that remains in the target field after truncation, and that the amount of recovered information may further be reduced by the final choice associated with regularization. To represent this last choice, consider the estimate of $\mathbf{B}$ using $\mathbf{T}^r$ and $\mathbf{P}^r$:

$$\mathbf{B} = \mathbf{U}_t^r \Sigma_t^r \mathbf{V}_t^{r\mathrm{T}} \mathbf{V}_p^r \left(\Sigma_p^r\right)^{-1} \mathbf{U}_p^{r\mathrm{T}}.$$

At this point, different regularization choices can be used to filter the correlation matrix $\mathbf{V}_t^{r\mathrm{T}} \mathbf{V}_p^r$. For our example herein, we consider the CCA formalism, which factors the matrix using SVD and in turn retains a rank associated with a determined number of canonical coefficients:

$$\mathbf{V}_t^{r\mathrm{T}} \mathbf{V}_p^r = \mathbf{O}_t^r \Sigma_{CCA}^r \mathbf{O}_p^{r\mathrm{T}}.$$

This example illustrates the regularization choice, which is first one of method, that is, whether to use SVD and rank reduction or some other method to filter the correlation estimate in $\mathbf{B}$. The second choice is how much to regularize, which in the case of CCA is associated with the number of canonical coefficients that are retained. Various means of objectively selecting the degree of regularization exist, including traditional truncation rules (Preisendorfer et al., 1981), optimization of CVREs (Smerdon, Kaplan, Chang, & Evans, 2010), or minimization of the generalized cross validation function (T. Schneider, 2001). Regardless of the details, however, these choices ultimately determine how the regression between the two matrices is regularized and by how much.

The basic choices listed above underlie the application of traditional methods like CCA or PCA (which uses SVD reductions of $\mathbf{T}$ and $\mathbf{P}$ but does not truncate the cross-covariance matrix between the two) to the paleo-CFR problem (see Cook et al. (1994) for examples). Many of the methodological debates about CFRs in the first decade of the 21st century ultimately boiled down to which of the three choices should be applied, how to objectively determine the various rank reductions, and which regularization should be chosen. The one new methodological advance on this theme that emerged in the early 2000s was the application of the iterative RegEM algorithm (T. Schneider, 2001) to derive hemispheric and global CFRs (Rutherford et al., 2003, 2005). The estimates within this framework are still based on regularized multivariate linear regressions, specifically ridge regression (RegEM-Ridge; Rutherford et al., 2003; Rutherford et al., 2005) and truncated total least squares (RegEM-TTLS; Mann et al., 2007; Mann et al., 2009), but the regression coefficients are non-linearly and iteratively estimated by casting CFRs as a missing-value or imputation problem. Within this formalism, the mean and covariance of an incomplete data set are initially infilled, updated, and ultimately selected based on the minimization of the expected mean-squared error of the infilled data within some specified threshold of convergence. T. Schneider (2001) demonstrated that RegEM-Ridge performed more skillfully than non-iterative linear regression methods within the context of synthetic experiments using observational gridded data sets and noted in the conclusion of his paper that the method also could be applied in the context of the CFR problem. This optimism was reflected in Rutherford et al. (2003), who were the first to test RegEM-Ridge as a CFR method. Rutherford et al. (2005) would subsequently use RegEM-Ridge to derive a CFR with real proxies, but later analyses demonstrated that the original tests of RegEM-Ridge were incorrectly applied (Smerdon & Kaplan, 2007; Smerdon et al., 2008) or based on erroneous data (Smerdon, Kaplan, & Amrhein, 2010). These developments ultimately led to the adoption of RegEM-TTLS as the RegEM CFR method of choice (Mann et al., 2007), which was subsequently used to derive a new CFR with a widely expanded proxy data set (Mann et al., 2009). One justification for the use of RegEM-TTLS was its use of an errors-in-variables, or inverse regression, formulation that was shown to better preserve variance in reconstructions that targeted NH indices (Hegerl et al., 2007). Consistent with the index reconstruction applications, RegEM-TTLS also better preserves variance in synthetic CFR reconstructions, albeit in a spatially heterogeneous manner, while other skill assessments indicate that the method does not perform in a universally better or more advantageous way than other reduced-space CFR methods (Smerdon et al., 2011, 2016).

Except for a modified version of RegEM based on Markov Random Fields (GraphEM; Guillot et al., 2015), RegEM has not been widely used as a CFR method over the last decade. Synthetic experiments using last-millennium climate model simulations, known as pseudoproxy experiments (Smerdon, 2012), have demonstrated that many of the regression-based large-scale CFR methods perform similarly, including RegEM, and that they are all subject to significant spatial uncertainties (Christiansen et al., 2009; Evans et al., 2014; B. Li & Smerdon, 2012;

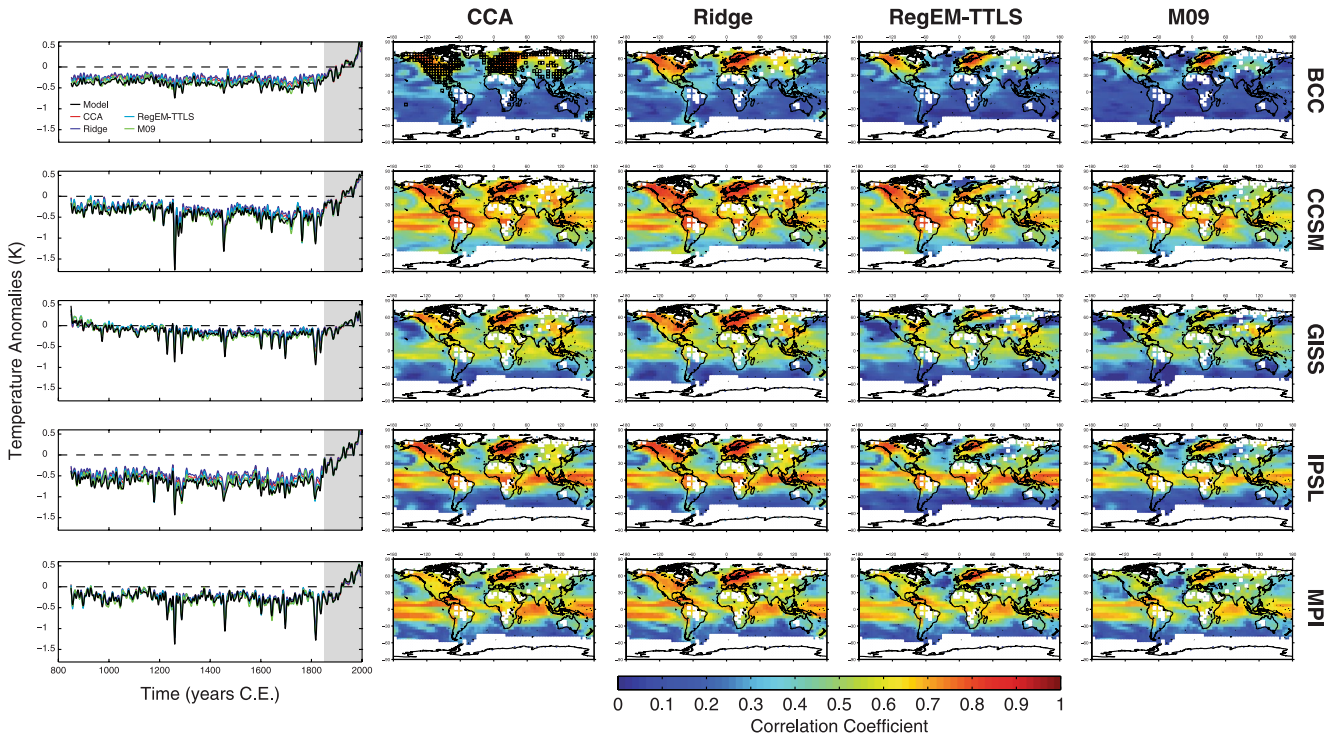

Figure 9. Skill assessments using four climate field reconstruction (CFR) methods and pseudoproxy experiments based on five different last-millennium climate simulations (BCC, CCSM, GISS, IPSL, and MPI; see (Smerdon et al., 2016) for more specific model descriptions). The first column on the left plots the area-weighted global mean surface air temperature time series derived from pseudoproxy experiments (SNR = 0.5) and the CCA, ridge regression and two implementations of the RegEM-TTLS CFR methods (RegEM-TTLS as originally formulated by RegEM-TTLS; T. Schneider, 2001 and as modified and applied by M09; Mann et al., 2009). These mean time series are compared against the global mean time series calculated for the true model fields. Time series have been smoothed using a decadal low-pass filter (ten-point Butterworth). Gray shaded areas represent the calibration interval (1850–1995 CE). The four remaining columns map the local correlation coefficients (Pearson's *r*) for the four CFR methods using pseudoproxies with SNRs of 0.5 and the five model fields. All methods use the same pseudoproxies, target field, and intervals for calibration (1850–1995 CE). Grid-point locations of the pseudoproxies used in all the pseudoproxy experiments are shown in the upper left panel as open grid squares, which approximate the distribution of the Mann et al. (2009) network. Correlation coefficients are 95% significant (two-sided p-value < 0.05) in pseudoproxy experiments for values of about 0.06 or higher (with many regions displaying correlation coefficients that are 10 times that level or more), where the significance is based on the degrees of freedom in the validation period of the pseudoproxy experiment from 850 to 1849 Common Era.

B. Li et al., 2016; Riedwyl et al., 2009; Smerdon et al., 2011, 2016; Wang et al., 2015; Wang et al., 2014; Yun et al., 2021). Figure 9 provides one example of pseudoproxy results from four CFR methods: CCA and ridge regression (Smerdon, Kaplan, Chang, & Evans, 2010) and RegEM-TTLS as applied in either T. Schneider (2001) or Mann et al. (2009); note that for a missing block matrix as represented in Figure 8, RegEM-Ridge converges to the traditional ridge regression result, which is why ridge regression is applied outside of the RegEM formalism in the example shown in Figure 9. Across all of the methods, there is widespread variability in the correlation between the derived CFR and the model target during the validation period, which varies even more widely across the pseudoproxy experiments based on different models.

3.2. Spatial Skill and Uncertainties in Reduced-Space CFRs

The importance of evaluating *spatial* skill has been long recognized in tree-ring based CFRs of hydroclimate (Cook et al., 1994, 1999) and limited spatial skill assessments were provided in Mann et al. (1998) for their derived temperature CFRs. Some later studies also reported summaries of field statistics or provided spatial plots of some skill metrics in either pseudoproxy experiments or real-world temperature CFRs (Mann et al., 2005, 2007; Rutherford et al., 2003). The results in Figure 9 are nevertheless characteristic of a growing awareness of the spatial uncertainties in large-scale temperature CFRs, an insight that was born out of significant efforts over the last decade (Anchukaitis et al., 2017; Evans et al., 2014; Hakim et al., 2016; King et al., 2021; B. Li & Smerdon, 2012; B. Li et al., 2016; Neukom et al., 2019; Riedwyl et al., 2009; Smerdon, Kaplan, Chang, & Evans, 2010; Smerdon et al., 2011, 2016; Steiger & Smerdon, 2017; Steiger et al., 2014, 2018; Wang

et al., 2014, 2015; Wilson et al., 2010; Yun et al., 2021) and represented a break from the typical skill assessments of earlier temperature CFRs that tended to focus primarily on the performance of large-scale means (Smerdon & Pollack, 2016). This focus was insufficient for fully assessing the spatial skill of CFRs (Smerdon, 2012), which is further demonstrated in Figure 9 based on a comparison of the NH mean performance and the global correlation patterns. It is worth noting that the skill assessments in Figure 9 are all based on comparisons with the full target field, despite the filtering of the reconstruction target due to rank reduction. Some recent pseudoproxy assessments have addressed this difference (Yun et al., 2021), but more work should assess how skill assessments are impacted by accounting for whether or not the target field is filtered in a given reconstruction method. This focus on the spatial skill of CFRs over the last decade has been an important development because the true value of large-scale CFRs is in their ability to provide useful spatiotemporal information for interpreting large-scale dynamics and for comparisons with climate model simulations (Anchukaitis et al., 2017; Coats et al., 2013, 2015, 2020; Herweijer et al., 2007; King et al., 2021; Mann et al., 2009; PAGES Hydro2k Consortium, 2017; Schmidt et al., 2014; Seager, 2007; Seager et al., 2007; Steiger et al., 2019; Tejedor et al., 2021a, 2021b).

In addition to spatial skill evaluations, uncertainty estimation in large-scale temperature CFRs has been an important part of their interpretation. Uncertainty estimation has typically followed two approaches: (a) analysis of regression residuals; and (b) CFR ensemble generation and subsequent confidence interval estimates using perturbations to regression parameters and choices. Early publications analyzed regression residuals by testing for heteroscedasticity and autoregressive properties to determine whether unresolved variance could be estimated as Gaussian and temporally invariant (Mann et al., 1998; Rutherford et al., 2005). In cases where these assumptions were justified, the unresolved variance of the residuals at each grid point was used to estimate confidence intervals, which were in turn propagated to determine spatial mean estimates. Despite these studies assuming time-invariant properties of the analyzed residuals, it is noted that uncertainties still increase back in time because new regressions for each proxy nest were performed, yielding nests with increasing uncertainties as the proxy networks diminished back in time.

Later studies also used regression residuals to estimate uncertainties, but applied more complicated bootstrapping methods to model the residuals nonparametrically and derive ensemble regression results from which confidence intervals were estimated (Li et al., 2007; Wahl & Smerdon, 2012). To the degree that these approaches perform many regressions using residual bootstrapping, they share similarities with other approaches that generated ensemble uncertainties by testing a range of regression choices. Similar to the screening and weighting schemes that were discussed in the preceding section for PPR, ensemble CFRs have been derived by varying regression choices tied to proxy sampling, calibration windows, and eigenvalue truncation thresholds tied to explained variance ranges (Neukom et al., 2011, 2019). Confidence intervals in these cases are in turn estimated based on distribution thresholds within the ensemble of CFRs about the mean or median reconstruction estimate. These two approaches, residual analyses and bootstrapping of regression choices, describe the general approaches that have been used to estimate the uncertainties in large-scale temperature CFRs. We also note that the two could be combined to generate one grand ensemble of CFRs as the most comprehensive estimate of ensemble uncertainty. Despite these advances, however, robust uncertainty estimates for reduced space CFRs require continued attention and research to decipher the differences between reconstruction accuracy and precision, as discussed in the case of PPR.

3.3. Expanding CFR Techniques

In addition to developing a more detailed understanding of the spatial skill and uncertainties in large-scale temperature CFRs, the last decade saw the development of a wider range of CFR techniques that deviated from the multivariate linear regression formalism that had been the hallmark of approaches over the previous several decades. Tingley and Huybers (2010a) developed and applied a hierarchical Bayesian model (named by the authors as BARCAST) to assimilate proxy and observational temperatures and, hence, probabilistic estimates of a gridded temperature CFR over North America; despite the new formalism, the relationship between proxies and temperature was still modeled as linear in this application, although the BARCAST framework theoretically allows any process-based model to be adopted. The resulting reconstruction was shown to perform skillfully; comparisons between BARCAST and results from RegEM-Ridge and -TTLS showed the former to perform more skillfully than the two RegEM methods (Tingley & Huybers, 2010b). BARCAST was later evaluated and compared to results from CCA over the European domain using pseudoproxy experiments (Werner et al., 2013)

and ultimately used to derive a temperature CFR over Europe (Luterbacher et al., 2016). Despite these advances, however, the BARCAST method is computationally expensive and except for the regional CFRs just listed, it has not been widely used as a CFR method and, in particular, has not been used to derive any CFRs on hemispheric or global scales. Similarly, various machine-learning techniques have also been considered in the context of the CE CFR problem or closely related applications (Kadow et al., 2020; Monteleoni et al., 2013; Salonen et al., 2019), but have yet to be applied widely in a CFR context. The beginning of the last decade also saw extensive work and development on paleo DA techniques (Annan & Hargreaves, 2012; Bhend et al., 2012; Goosse et al., 2010; Steiger et al., 2014; Widmann et al., 2010), which we will review in detail in Section 4. Analog methods also have been applied to the large-scale temperature CFR problem (Gómez-Navarro et al., 2017), but because of their similarity to some DA techniques they are also discussed in Section 4.

Several large-scale temperature CFRs have also been produced using more traditional regression-based techniques over the last decade. PCR was used to derive continental-scale temperature CFRs over South America (Neukom et al., 2011) and North America (Wahl & Smerdon, 2012; Wahl et al., 2014). Despite the more traditional casting of these PCR techniques and their regional focus, they are important because of their ensemble approaches to the reconstruction problem. These studies used a range of regression parameters to produce ensembles of reconstructions as a means of incorporating uncertainties associated with regression decisions (similar to what had been done earlier in the context of testing NH mean results based on regression choices (Burger & Cubasch, 2005; Burger et al., 2006) or through variations in the underlying climatic field properties (Christiansen et al., 2009). This ensemble approach has also been adopted in many hydroclimate reconstructions (see discussion in Section 2) and is an important approach to quantifying the data and methodological uncertainties in CFRs. More recently, it has been argued that ensembles should be used to produce CFRs in order to bracket the rather large uncertainties associated with the choice of proxies and CFR methods (Büntgen et al., 2021a; Neukom et al., 2019; PAGES2k Consortium et al., 2019). On the other hand, investigator decisions based on specific knowledge of proxy data or a priori reasons for why specific methods are most appropriate may provide important justifications for curated data sets or specifically applied CFR choices (e.g., Anchukaitis et al., 2017; King et al., 2021).

In addition to the earlier PCR products, Anchukaitis et al. (2017) used the PPR technique to derive a warm-season temperature CFR for the NH extratropics exclusively from tree rings. This was the first hemispheric-scale CFR produced with PPR, in contrast to the much broader application of PPR for continental-scale hydroclimate CFRs. The Anchukaitis et al. (2017) study also highlights an important aspect of CFRs that continues to be debated, namely the use of single versus multi-proxy databases or even the use of annually-resolved versus mixed resolution proxies. While multi-proxy databases provide more spatial coverage, single-proxy networks may allow the incorporation of specific insights and approaches associated with individual proxy systems (Anchukaitis et al., 2017; King et al., 2021).

Franke et al. (2020) investigated the importance of input data quality and quantity in producing a NH temperature CFR using DA and exclusively a tree-ring network. This study revealed certain tradeoffs between using the geographically restricted set of highly screened temperature-sensitive tree-ring chronologies used by Anchukaitis et al. (2017) and much larger, but less well screened, tree-ring data sets for more complete NH coverage (Breitenmoser et al., 2014; PAGES2k Consortium, 2017). Franke et al. (2020) confirmed the highly skillful results of Anchukaitis et al. (2017) for the NH extratropics, but found that reconstruction skill based on that network strongly declined southward from the high northern latitudes. A much larger, more geographically complete, tree-ring network produced the most spatially complete and skillful NH temperature CFR based on the Breitenmoser et al. (2014) data, but only after stringent statistical screening was applied to retain only those tree-ring chronologies with a statistically significant ($p < 0.05$) climate signal. The increase in overall geographic skill was primarily due to the inclusion of precipitation-sensitive tree-ring records with significant *negative* correlations with growing-season temperature resulting from evapotranspiration demand. This differs significantly from the Anchukaitis et al. (2017) CFR, which used tree-ring records that have strictly *positive* correlations with growing season temperature. There remains some controversy about the wisdom of using the inverse temperature signals from drought-stressed trees to reconstruct past temperatures because it is unclear that the spectral properties of those reconstructions are equivalent to those based on direct temperature effects on tree growth. The same concern relates to the evaluation of volcanic forcing in temperature reconstructions based on drought-stressed trees. Until these matters are thoroughly investigated, such reconstructions must be interpreted with an added element of caution.

Finally, proxies often sample specific seasonal windows (see Anchukaitis and Smerdon (2022) for a recent review). Tree-rings, for instance, are predominantly sensitive to climatic conditions during or prior to the growing season (e.g., Baek et al., 2017; St. George et al., 2010). For this reason, more recent CFRs have tended to explicitly reconstruct warm-season temperatures (Anchukaitis et al., 2017; Luterbacher et al., 2016), but this is not always the case (e.g., Neukom et al., 2019). These seasonal sensitivities are particularly complicated on global scales where the temperature-sensitive responses of proxies like tree growth (roughly June-August in the NH and December-February in the Southern Hemisphere) are effectively 6 months out of phase. Recent reconstructions of annual temperatures attempted to compensate for this anti-phasing problem by reconstructing April-March mean temperature (Neukom et al., 2019), thus spanning both growing seasons in an annual mean, or by employing proxy networks in subsets based on seasonal sensitivities (Steiger et al., 2018). It is nevertheless unclear how these approaches might bias the resulting field reconstructions and the subsequent evaluations of forced responses. Reconstructing extratropical temperature for the NH and SH separately would accommodate the out-of-phase character of the interhemispheric warm and cool seasons (see Mann and Jones (2003) for an early example). Such approaches may themselves reveal some interesting dynamical insights that would otherwise be obscured by reconstructing April-March annual temperatures. This approach was employed by J. G. Palmer et al. (2015) and Morales et al. (2020) to reconstruct DJF hydroclimate over Australia-New Zealand and South America, respectively, as opposed to reconstructing JJA conditions in drought atlases within the NH (e.g., Cook et al., 2004). Despite the remaining challenges, these seasonal considerations have become more widely recognized relative to early work on large-scale temperature CFRs that targeted annual calendar-year temperatures without addressing seasonal sensitivities (e.g., Mann et al., 1998; Mann et al., 2009; Rutherford et al., 2005).

4. Paleo Data Assimilation

We have to this point alluded to paleo DA approaches for producing CFRs, but the growth in this area has been rapid and expansive enough over the last decade that it deserves explicit attention. The first papers to explore forms of DA in the context of paleoclimate reconstructions were published more than a decade ago (Crespin et al., 2009; Goosse et al., 2010; Widmann et al., 2010). Since that time, paleo DA has dramatically grown in its application to derive and test CFRs over the CE (Amrhein et al., 2020; Bhend et al., 2012; Franke et al., 2017; Hakim et al., 2016; King et al., 2021; Matsikaris et al., 2015, 2016; Okazaki & Yoshimura, 2017; PAGES2k Consortium et al., 2019; Parsons et al., 2021; Perkins & Hakim, 2017, 2021; Samakinwa et al., 2021; Steiger & Smerdon, 2017; Steiger et al., 2014, 2017, 2018; Tardif et al., 2019; Valler et al., 2021), and, more recently, it has been used to derive reconstructions over timescales spanning the Holocene and Last Glacial Maximum (Badgeley et al., 2020; Erb et al., 2022; Osman et al., 2021; Tierney et al., 2020). It is also worth noting that in addition to the more recent DA efforts incorporating multiproxy networks and fully-coupled climate models, there is a well-established history of approaches that combine various model hierarchies and proxy information specifically from paleoceanographic data (Amrhein et al., 2018; Dail & Wunsch, 2014; Kurahashi-Nakamura et al., 2014, 2017; LeGrand & Wunsch, 1995).

The allure of DA is perhaps best summarized by two related features. The first is the ability of DA to provide additional information to the information-limited global CFR problem, namely the model-based physics that are used to estimate the covariance structure of a given climatic field. The second is the potential for DA to estimate additional climatic fields, for instance atmospheric pressure, that are physically related to a variable that imposes the principal constraint on proxy variability such as surface temperature or precipitation (note that regression-based approaches can also leverage covariances among variables to provide simultaneous reconstructions of fields, but these reconstructions would be based on statistical associations and not physically defined relationships). DA thus offers the prospect of skillfully reconstructing a broader climatic state space than the single variables that are typically targeted in other CFR methods. In some senses, this is a full-circle return to the idea advanced by Fritts et al. (1971), namely that trees limited by temperature or soil moisture can be used to reconstruct large-scale atmospheric pressure patterns. The important contrast with DA, however, is that Fritts et al. (1971) assumed a linear relationship between atmospheric pressure and the responses measured in tree-ring widths, while the connections between variables like atmospheric pressure, surface temperature and precipitation are all physically constrained to varying degrees by the applied model physics in a DA application. These physical constraints, along with the ease of reconstructing multiple variables and incorporating a wide range of paleoclimatic proxies, make the DA methodology attractive for CFR applications.

There are many different methodological flavors of DA. In general, however, the majority of approaches can be grouped into four categories: pattern nudging (von Storch et al., 2000), ensemble Kalman filters (Bhend et al., 2012; Dirren & Hakim, 2005; Huntley & Hakim, 2010; Pendergrass et al., 2012), forcing singular vectors (van der Schrier & Barkmeijer, 2005), and particle filters, which select simulated ensemble members that best match proxy data (Annan & Hargreaves, 2012; Franke et al., 2011; Goosse et al., 2010). The initial applications of DA in the context of paleoclimatic problems pursued forcing singular vectors and particle filters using proxy data and Earth system models of intermediate complexity (Crespin et al., 2009; Goosse et al., 2010). Pattern nudging has also been used to prescribe atmospheric circulation anomalies that in turn yield temperature anomalies consistent with proxy data (Widmann et al., 2010).

Several broad characteristics of these methods distinguish among them, with implications for computational expense and the fidelity of their association with model physics. The first distinction of note is the so-called offline versus online methods. Offline methods use an ensemble prior derived from an existing and completed forced-transient simulation from a fully-coupled climate model, typically a last-millennium simulation spanning one thousand years or more (e.g., Hakim et al., 2016). Online approaches use time-dependent ensemble priors in which a new ensemble of simulations is generated at each time step of the DA (e.g., Matsikaris et al., 2015; Okazaki & Yoshimura, 2017). A critical difference between these approaches is that online techniques are much more computationally expensive. Online approaches nevertheless incorporate the model physics as a constraint on the time evolution of a given DA reconstruction, while the time evolution of offline approaches is only dependent on the proxy data. This means that the time history and power spectra of an offline reconstruction may be inconsistent with the physics of the underlying model. Some of these considerations will be addressed in more detail in the following sections, but the details in general need to be evaluated in the context of how DA reconstructions are employed and the questions they are used to address.

4.1. Data Assimilation Using the Offline Ensemble Kalman Filter

Although various DA flavors have been applied to the CFR problem, the offline ensemble Kalman filter method has been the most widely used paleo DA method for deriving CE CFRs to date (Hakim et al., 2016; King et al., 2021; PAGES2k Consortium et al., 2019; Perkins & Hakim, 2017, 2021; Steiger et al., 2017, 2018; Tardif et al., 2019). While offline and online approaches each have their own merit, the offline approach has proven to be skillful in paleo DA contexts while requiring less computational expense. We will revisit this issue below, but for now it is useful to highlight several aspects of the offline approach and use it as the basis of expected advances moving forward.

It is customary to begin with the solution to the so-called update equation for the Kalman filter (Kalnay, 2003), assuming Gaussian-distributed errors:

$$\boldsymbol{x}_a = \boldsymbol{x}_b + \mathbf{K}[\boldsymbol{y} - \mathcal{H}(\boldsymbol{x}_b)],$$

where $\boldsymbol{x}_b$ is the prior state vector and $\boldsymbol{x}_a$ is the posterior-estimated state vector, and by state vectors we mean the collection of climate states that are being targeted in the DA application (these can be fields like surface temperature or indices like the Niño3.4 index). The vector $\boldsymbol{y}$ contains the network of proxy observations at a given time step and the prior-estimated values of these observations are derived from $\mathcal{H}(\boldsymbol{x}_b)$, which is a vector operator that maps $\boldsymbol{x}_b$ from the modeled climate state (e.g., temperature) to a proxy observation (e.g., the ring width of a tree). The so-called innovation term, $\boldsymbol{y} - \mathcal{H}(\boldsymbol{x}_b)$, is representative of the information that has been added to the prior, which is in turn weighted and transformed into the state space by the Kalman gain matrix, $\mathbf{K}$:

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1},$$

where $\mathbf{B}$ is the covariance of the prior state space, $\mathbf{R}$ is the error covariance matrix of the proxy data, and $\mathbf{H}$ is a linearization of $\mathcal{H}$ about the prior mean. $\mathbf{B}$ and $\mathbf{R}$ can be time dependent, but in offline approaches over the CE these estimates typically remain constant. The key takeaway from the above two equations is that the posterior is estimated based on the covariance between the model's estimate of the proxy observations at their specific locations and the state space spanned by the prior. Importantly, paleo DA requires a quantification of the error in the prior covariance, which is estimated through the ensemble of the prior state. It is also useful to note that this

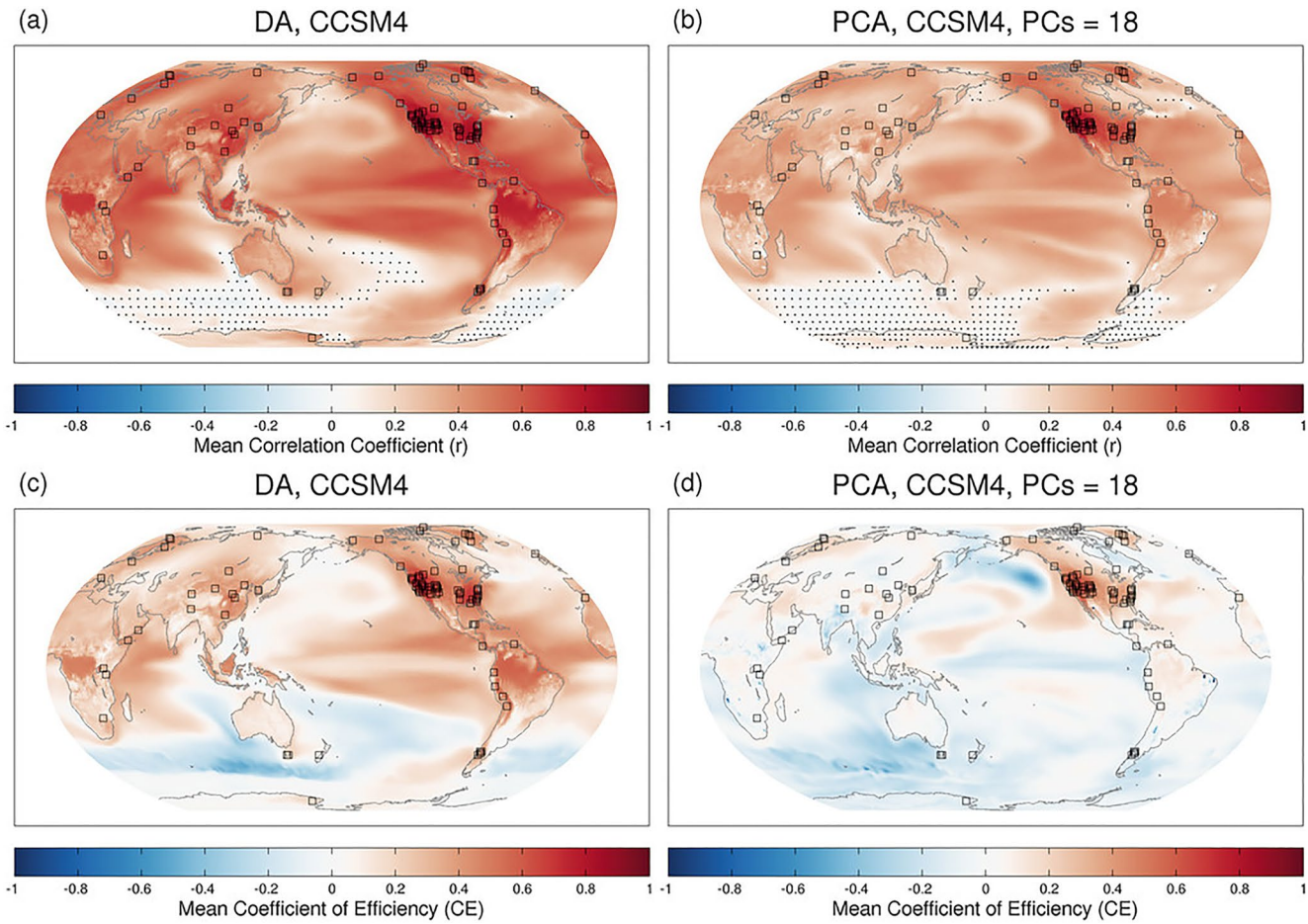

Figure 10. Spatial skill maps for pseudoproxy tests of two climate field reconstruction methods: the offline ensemble Kalman filter data assimilation method and principal component regression (PCR). Skill measured using correlation coefficients is shown in panels (a) and (b), while skill measured using the coefficient of efficiency is shown in panels (c) and (d). The calibration periods for the two methods is 1881–1980, while the reconstruction period is 1300–1880 (also the period over which the skill measures are computed). Empty black boxes are centered over pseudoproxy locations, and stippling indicates correlations that are not significant at the 95% level. Figure from Steiger et al. (2014).

DA formalism shares a similarity with what has come to be known as the analog method in the sense that the analog method uses a collection of climate states, or a prior, to estimate past state spaces that are statistically most consistent with proxy observations in a given year or season (Gómez-Navarro et al., 2017; Neukom et al., 2019). The principal difference between the analog method and DA is that the analog method simply selects the best matching climate state (or states), while DA changes the climate states themselves via the Kalman filter equation. The posterior climate state from DA is thus a combination of a purely simulated climate state with a correction driven by the proxy data. In contrast to the analog method, the posterior DA estimate therefore may not be explicitly consistent with the physics in the climate model prior.

4.2. Data Assimilation and Comparisons With Other CFR Methods

The above DA formalism has been discussed and developed extensively in prior publications (e.g., Hakim et al., 2016; Steiger & Smerdon, 2017; Steiger et al., 2014; Tardif et al., 2019). In the context of this review, it is useful to highlight a few examples in which the paleo DA results have been compared to results from other CFR methods. In one of the first such comparisons, Steiger et al. (2014) used pseudoproxy experiments to compare results from the offline ensemble Kalman filter DA method with those derived from a reduced-space approach using PCA. This comparison showed both the DA and reduced-space methods to perform skillfully in terms of significant correlations with the target field over most regions during the validation interval (Figure 10). The

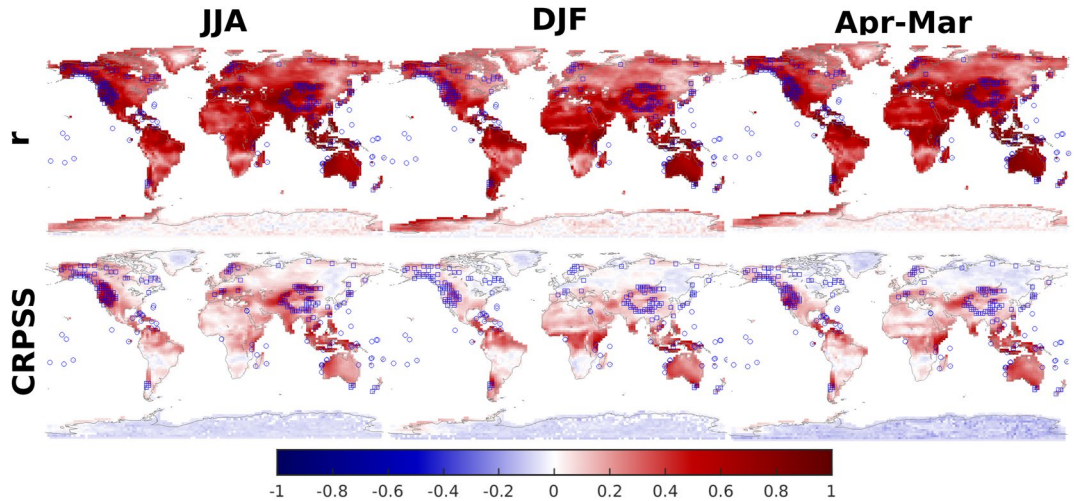

Figure 11. Reconstruction skill of pseudoproxy-based experiments: correlation (r), and the mean continuous ranked probability skill score for the Palmer drought severity index. Figure from Steiger and Smerdon (2017).

more stringent coefficient of efficiency measure indicated more mixed results, with some regions yielding skillful reconstructions (coefficient of efficiency > 0), while other regions did not pass the skill measure. While the patterns of skill are similar across the two tested methods, the DA reconstructions achieved higher values of skill, while also yielding more widespread regions of skill than the reduced-space approach; it also should be noted that both methods yield the highest skill in locations of dense pseudoproxy sampling, a characteristic that has been observed and noted for a wide range of CFR methods (Anchukaitis & Smerdon, 2022; Smerdon, 2012; Smerdon et al., 2011, 2016; Wang et al., 2014).

Similar to the pseudoproxy tests of temperature reconstructions described above, a version of the offline ensemble Kalman filter DA method was also tested in the context of pseudoproxy experiments targeting hydroclimate reconstructions (Steiger & Smerdon, 2017). Relative to the earlier Steiger et al. (2014) study, the tests involved a pseudoproxy sampling that reflected an updated proxy network, a more recent last-millennium model simulation as the basis of the experiment (Otto-Bliesner et al., 2016), and targeted June-August and December-February seasons, as well as annual water year (April-March). Similar to results for temperature, reconstructions were skillful for targeted hydroclimate variables (see results for PDSI in Figure 11), while indicating concentrated skill in densely sampled pseudoproxy regions. Figure 11 also provides results for the CRPSS, which is further discussed in Box 1 and is used for skill assessments across an ensemble of reconstructions like the one derived in the applied DA method.

The DA method tested in Figure 11 was latter applied by Steiger et al. (2018) to derive the Paleo Hydrodynamics Data Assimilation (PHYDA) product, an effort that included a comparison to the regional drought atlas products that have targeted growing season PDSI in multiple continental regions. The comparisons between PHYDA and the drought atlases yielded strong and significant correlations over most of the comparison domains (Figure 12), despite the vastly different applied methodologies and overlapping but different proxy data sets that were employed; the drought atlases use exclusively dendroclimatic records (with a small exception in the Australia and New Zealand drought atlas), while the PHYDA included some, but not all, of the dendroclimatic records included in the drought atlases and a broader range of multiproxy records derived from the PAGES2k database (PAGES2k Consortium et al., 2019). An interesting feature of the comparisons shown in Figure 12 is that the reconstructions agree with the highest correlation values over the regions where the proxy sampling in the PHYDA and drought atlases is most dense. This demonstration once again confirms that the most robust CFR results are likely over regions with dense proxy sampling. Moreover, the regions of disagreement tend to be in locations that are farthest from proxy sampling, where the covariance assumptions in each of the products will be most at odds. In the case of the drought atlases, the covariance assumption is dependent on the size of the search radius and the distribution of the proxy network, where the grids farthest

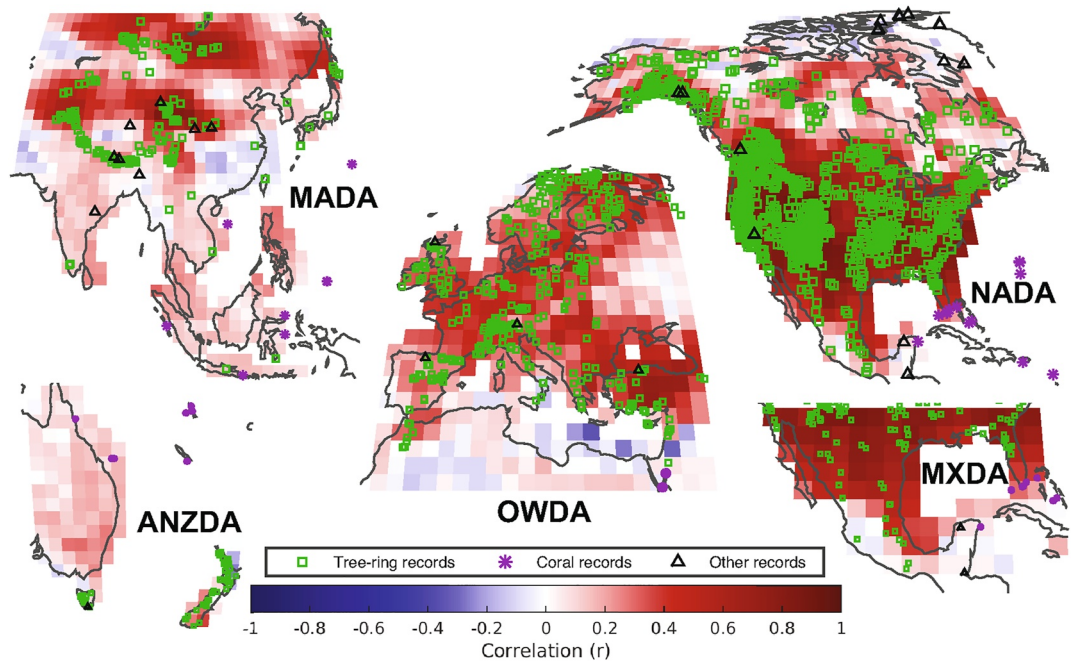

Figure 12. Correlation between a data assimilation (DA)-based Palmer Drought Severity Index (PDSI) reconstruction (PHYDA) and the PDSI reconstructions of the drought atlases from the years 1500–2000: the North American drought atlas (NADA), the Old World drought atlas (OWDA), the Monsoon Asia drought atlas (MADA), the Mexican drought atlas (MXDA), and the Australia and New Zealand drought atlas (ANZDA). Correlations are computed using the JJA reconstruction for NADA, OWDA, MADA, and MXDA and the DJF reconstruction for ANZDA. Prior to computing the correlations, the drought atlases were interpolated to the land surface grid of the CESM LME model simulation used in the DA-based reconstruction presented here. Figure from Steiger et al. (2018).

from proxy sampling will have assumed the farthest afield teleconnections to a given grid point. The paleo DA method, by contrast, assumes spatial covariances defined by the model prior. These covariance assumptions are therefore an important likely explanation for some of the regions where agreement between the PHYDA and the drought atlases yield small or negative correlations. Despite any differences, the agreement between the two products is important and highlights a critical need for more systematic comparisons between the rapidly growing number of CFR products derived from multiple methods and data sets (e.g., King et al., 2021; Tejedor et al., 2021a, 2021b).

The above-described methodological tests and comparisons highlight a fraction of the widespread testing and application of paleo DA methods. They are expanding in scope and application, while more DA products based on observed proxy data are being produced. These latter efforts are also increasing the degree to which DA-based CFRs are being intercompared, as well as compared to CFRs derived from other methods. Table 2 provides a list of currently available DA CFRs. For each of the source publications, the reconstructed variables, temporal resolution, and space and time range are listed.

4.3. Data Assimilation and Future Methodological Questions

There are several outstanding research questions that continue to be relevant as more DA CFRs are derived and the methodology is refined for paleoclimate applications. The first of these issues is the role of the prior, $\boldsymbol{x}_b$, and the various advantages and disadvantages of online and offline approaches. In most existing DA CFRs, the last-millennium simulations that have been used for the offline prior are derived from a single model. While DA CFRs appear to have relatively limited dependence on the nature of the offline prior selected from the collection of last-millennium simulations in the CMIP5 database (Hakim et al., 2016), more recent work has suggested that structural uncertainties in prior-estimated covariances contribute significant uncertainties (Amrhein et al., 2020; Parsons et al., 2021); bias corrections of the prior also appear to improve skill in DA-derived CFRs (Steiger et al., 2018). These prior dependencies are therefore critical to further evaluate,

Table 2
Available Paleo Data Assimilation Products and Their Summary Characteristics

Publication	Reconstructed variables	Temporal resolution	Space/Time range
Steiger et al. (2018)	Surface Air Temperature (2 m), PDSI, SPEI, Niño Indices (1 + 2, 3, 3.4, 4), Zonal Pacific SST Gradient, Latitude of ITCZ Position in 11 tropical boxes, AMV Index	Seasonal (DJF and JJA) and Annual (April-March)	Global/1–2000 CE
Tardif et al. (2019)	Surface Air Temperature (2 m), Sea Surface Temperature, Sea Level Pressure, Geopotential Height (500 hPa), Precipitation, Precipitable Water, PDSI, Indices: AMV, AO, PDO (index and pattern), NAO, Niño3.4, SAM, SOI, North Pacific Lat/Lon	Annual (January-December)	Global/1–2000 CE
Valler et al. (2021)	Surface Air Temperature (2 m), Total precipitation, Sea-level Pressure, Geopotential height (500 and 100 hPa), Eastward and Northward wind Components (850 and 200 hPa), Vertical velocity at 500 hPa	Monthly	Global/1600–2005 CE
Samakinwa et al. (2021)	Sea Surface Temperature, Sea Ice Concentration	Monthly	Global/1000–1849 CE
Perkins and Hakim (2021)	Surface Air Temperature (2 m), Precipitation, Sea Level Pressure, Geopotential Height (500 hPa), Outgoing Top-of-Atmosphere Longwave and Shortwave Radiation, Sea-Surface Temperature and Salinity, Dynamic Ocean Surface Height, and 0–700 m Ocean Heat Content	Annual (April-March) and Some Seasonal	Global/1000–2000 CE
King et al. (2021)	Surface Air Temperature (2 m)	Seasonal (March-August)	30°N–90°N/750–2011 CE

Note. Abbreviations in the table are Palmer Drought Severity Index (PDSI), Standardized Precipitation-Evaporation Index (SPEI), Atlantic Multidecadal Variability (AMV), Arctic Oscillation (AO), Pacific Decadal Oscillation (PDO), North Atlantic Oscillation (NAO), Southern Annular Mode (SAM), and Southern Oscillation Index (SOI). In cases where a second version of a product has been released, we include the more recent version (Tardif et al. (2019) is included as an update to Hakim et al. (2016), Valler et al. (2021) is included as an update to Franke et al. (2017)).

as are the method and means of bias corrections within the prior estimates. More computationally efficient online methods have also been investigated in which a time-dependent prior is generated using a linear inverse model to approximate temporal and spatial covariances in the state space (Perkins & Hakim, 2017, 2021). Offline approaches that approximate temporal dependencies with methods such as Kalman smoothers may also prove useful in the paleoclimatic context (Bruhwiler et al., 2005; Michalak, 2008). Such approaches may ultimately infuse offline paleo DA applications with additional temporal information based on physical time dependencies in the climate system; these time dependencies are currently only reflected in offline DA CFRs as manifest in the proxy observations themselves.

The second important area of future research relates to which proxies are used in the assimilation process and how those proxies are interpreted. The first consideration may be manifest as an approach that adopts a widescale incorporation of multiple proxies (e.g., Hakim et al., 2016; Sanchez et al., 2021; Steiger et al., 2018; Tardif et al., 2019) or one that curates a specific proxy type for assimilation (King et al., 2021). For those that employ large multi-proxy databases, different weightings or screenings of the proxy records, similar to those discussed for PPR, may also be adopted. For instance, Steiger et al. (2018) selected proxies as temperature or hydroclimate predictors in each reconstructed season based on the maximum correlation with the targeted state variables during the calibration interval. The second consideration is largely captured by how $\mathcal{H}$ is formulated, that is, the nature of the proxy system models (PSMs) that are employed in DA applications. The last decade has seen an increase in attempts to develop PSMs that are more complicated than the univariate linear models that have been traditionally adopted, and to in turn apply these PSMs in the context of CFRs (Evans et al., 2013). PSMs attempt to mathematically model the physical, chemical, biological and/or geological processes that lead to paleoclimate measurements. Investigations of tree-ring PSMs were published in the 2000s (Anchukaitis et al., 2006; Evans et al., 2006) and were presented as a means of more realistically capturing proxy-climate connections within tree-ring records. With the advent of DA and Bayesian methods in which non-linear PSMs can be seamlessly incorporated into the CFR methodology, PSM work has become more widespread and investigated in a range of proxy systems (Dee et al., 2015, 2016; Thompson et al., 2011). Despite the ongoing promise of PSM development, however, most DA CFRs have only applied univariate linear models as the principal PSM in practice. Exceptions include some bivariate linear models that are used for PSMs of corals and trees in some DA applications (Steiger et al., 2018; Tardif et al., 2019; Valler et al., 2021). PSM development therefore needs to progress in tandem with more applications that advance the use of PSMs in DA CFRs, while caution must be practiced to verify that more complicated PSMs improve reconstruction skill relative to the simpler models that have been adopted.

An additional issue in DA CFRs that originates with PSMs is that their parameters are usually trained on observational data. For example, in a simple univariate linear case, a proxy may be regressed against local temperature to determine the linear transfer function for that proxy at the given location. Once $\mathcal{H}$ is thus defined across all proxy locations, observational data are no longer incorporated into the DA framework, which will only involve the model-based prior and proxy measurements. But the use of the observational data to define $\mathcal{H}$ means that DA CFRs are not truly independent from the observational data over the period during which the proxy transfer functions were determined.

It nevertheless has been customary to evaluate the skill of DA CFRs by comparing them to observational data with little attention to what would be cast as in-sample (calibration interval) and out-of-sample (validation interval) periods in traditional regression-based approaches (see Box 1). More recent DA CFRs have nevertheless recognized the importance of validations with truly independent observational data (Steiger et al., 2018; Tardif et al., 2019). This awareness needs to be more widely applied for robust evaluations of DA reconstruction skill and more consistent skill comparisons against other regression-based CFRs that treat calibration and validation more operationally.

One additional advantage of DA CFR approaches (shared by their Bayesian counterparts) is that uncertainty estimates are provided naturally through the DA framework. In ensemble Kalman filter approaches, the posterior ensemble spread indicates the amount of uncertainty in the reconstruction. This posterior spread is determined by three factors: the initial prior ensemble spread, the spatiotemporal sampling of the data network, and the PSM error estimates in matrix **R**. This means that care must be taken in selecting the climate model for the prior ensemble and in accurately estimating how proxies are related to climate variables. Some DA methods rely only on these built-in methods for estimating uncertainties (Steiger et al., 2018), while other DA methods generate additional ensembles by resampling the proxy network and the climate model states chosen for the prior ensemble (e.g., Tardif et al., 2019).

5. Database Development

It is important to mention the crucial and ongoing efforts to collect, develop, and interpret more high-quality proxy records, the availability of which are ultimately the biggest determinant of CFR skill and utility. There remain large gaps in high-resolution multi-proxy networks over regions like the Southern Hemisphere, the tropics, and the global oceans, all of which change as networks extend back in time (for a detailed discussion of these issues in the context of hemispheric and global temperature reconstructions, see the recent review by Anchukaitis and Smerdon (2022)). More sampling in these regions is critical for improving CFRs, which ultimately and not surprisingly are most skillful in the regions where proxy networks have the highest sampling density (B. Li et al., 2016; Smerdon et al., 2016, 2011; Wang et al., 2014; Yun et al., 2021). Improvements in sampling, development, and interpretation advance at their own pace within each specific proxy community, but there have been multiple community-based advances over the last decade that have been very helpful for large-scale CFR development. The first has been the cultivation of multi-proxy databases by groups of experts working on specific proxy systems and in specific regions. The PAGES2k network (PAGES2k Consortium, 2013, 2017) has involved the most high-profile of these efforts, which has focused initially on developing a temperature-sensitive proxy database. This network has now served as the basis for multiple CFRs (Hakim et al., 2016; Neukom et al., 2019; Steiger et al., 2018; Tardif et al., 2019), while also spawning several other efforts to compile proxies and reconstruct sea surface temperatures over ocean regions (McGregor et al., 2016; Tierney et al., 2015) and to develop multiple kinds of hydroclimate databases (Konecky et al., 2020; PAGES Hydro2k Consortium, 2017). Similar community efforts to specifically develop temperature-sensitive tree-ring networks have also moved forward independent of the PAGES-based work (Anchukaitis et al., 2017; Wilson et al., 2016). With regard to tree-rings specifically, the last decade also has seen a massive expansion in the availability of tree-ring density and isotope records that have already had implications for interpretations of volcanic impacts on climate (Anchukaitis et al., 2017; Esper et al., 2015; L. Schneider et al., 2015; Tejedor et al., 2021a, 2021b; Wilson et al., 2016) and past hydroclimate (Büntgen et al., 2021b).

There have also been important advances in database standards and protocols. In many cases, each proxy community has developed unique protocols for archiving, describing, and distributing data. These standards can often be in conflict with each other, cumbersome, or insufficient for the wide range of applications that now seek to use proxy records, including the production of large-scale CFRs. Developing data archiving standards and better databases for digital data distribution has therefore been recognized as an important technological focus across paleoclimate communities (Emile-Geay & Eshleman, 2013; McKay & Emile-Geay, 2016), while digitally archiving derived CFRs has also become standard practice.

Finally, as mentioned in the previous section, PSMs represent an ongoing and important area of work. In addition to their application within CFR methodologies, PSMs are important in the context of proxy interpretation

and proxy-model comparisons outside of DA approaches. Additionally, PSMs have been used in the context of pseudoproxy experiments to generate more realistic pseudoproxy time series that are in turn tested in CFRs that mostly assume linear climate-proxy relationships (Evans et al., 2014; Steiger et al., 2017). These studies have been important for further characterizing the uncertainties that arise because proxies are non-linear and multivariate, which typically violates the underlying assumptions of many CFR methods. PSMs therefore have a host of applications, making them an important area of focus within the context of proxy record development and interpretation, in addition to their uses in CFRs.

6. Conclusions

It is worth concluding using Neukom et al. (2019) as a means of summarizing the state of the science in the comparison and production of CFRs. Although this work principally produced global temperature CFRs and addressed the question of spatial synchronicity in temperature anomalies across past climate epochs, the underlying methodological effort of the paper was based on the ensemble construction of a CFR using six different methods: (a) a point-by-point method using a composite plus scale approach that weights every proxy in the PAGES2k network based on the proxy correlations with temperature at each grid cell in the target field (Neukom et al., 2014), (b) PCR (Luterbacher et al., 2002), (c) CCA (Smerdon, Kaplan, Chang, & Evans, 2010; Wang et al., 2014), (d) GraphEM (Guillot et al., 2015), (e) a paleo DA technique (Steiger & Smerdon, 2017; Steiger et al., 2018), and (f) an analog method technique (Gómez-Navarro et al., 2017). Each of these reconstructions was derived from the same PAGES2k Consortium (2017) multi-proxy data set as predictors and the same April-March mean surface air temperature field as target. This collection of CFRs therefore provides the opportunity for a more controlled comparison across multiple state-of-the-art techniques, which is essential for further vetting these methods, understanding where substantial spatiotemporal uncertainties remain, understanding the impact of the difference in seasonal phasing of climate variability between the hemispheres and its effects on derived CFRs, and where additional sampling might universally improve CFR estimates. King et al. (2021) come to similar conclusions in a comparison between their DA product for NH temperatures and five other publicly available CFRs. In noting substantial differences across the CFR products, King et al. (2021) advocate for the "*development of a paleoclimate reconstruction intercomparison framework for systematically examining the consequences of proxy network composition and reconstruction methodology…*" This is indeed a needed effort as more CFRs are being produced using many different data sets, methods, and decisions within individual methodological frameworks. Controlled comparisons will help articulate the strengths and weaknesses of individual methods and whether specific regions, circumstances, or variables are more appropriate for specific approaches or data sets. Methodological comparisons therefore must also be coupled with ongoing attempts to develop standardized multi-proxy databases with appropriately structured digital access and metadata. Together these efforts will help further the ultimate goal of CE CFR research, in much the same spirit as originally proposed by Fritts et al. (1971), namely to extend our characterization of seasonal and annual estimates of climate over the past 2000 years as a means of improving our understanding of past, present, and future climatic change. We nevertheless conclude by noting that "*the reliability and utility of reconstructions should always be defined in terms of what purpose and questions they are being used to address*" (Anchukaitis & Smerdon, 2022). A clear definition of this intent is critical for deciding which CFRs are appropriate for a given purpose and how the uncertainties of a given reconstruction may impact the ability to address desired research questions. Moreover, if multiple CFRs are deemed appropriate, research findings should be tested across the expanding number of such CFR products, making evaluation of ensembles of CFR results a central approach for quantifying or making inferences about past climates or the behavior of the climate system at different temporal and spatial scales.

Data Availability Statement

Data in Figures 1–6 and Table 1 are from Gille et al. (2017).
Data in Figure 9 are from Smerdon et al. (2016).
Data in Figure 10 are from Steiger et al. (2014).
Data in Figure 11 are from Steiger and Smerdon (2017).
Data in Figure 12 are from Steiger et al. (2018).

Acknowledgments
This work was supported in part by NSF Grants OISE-1743738, AGS-1805490, and AGS-2101214 and ISF Grant 2654/20. We thank Gabi Hergerl and two anonymous reviewers for their helpful and insightful comments on the initially submitted version of this manuscript.

References

Akaike, H. (1974). New look at statistical-model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/tac.1974.1100705

Allen, D. M. (1974). Relationship between variable selection and data augmentation and a method for prediction. *Biometrics*, *30*(2), 374–375.

Amrhein, D. E., Hakim, G. J., & Parsons, L. A. (2020). Quantifying structural uncertainty in paleoclimate data assimilation with an application to the last millennium. *Geophysical Research Letters*, *47*(22), e2020GL090485. https://doi.org/10.1029/2020GL090485

Amrhein, D. E., Wunsch, C., Marchal, O., & Forget, G. (2018). A global glacial ocean state estimate constrained by upper-ocean temperature proxies. *Journal of Climate*, *31*(19), 8059–8079. https://doi.org/10.1175/jcli-d-17-0769.1

Anchukaitis, K. J., Evans, M. N., Kaplan, A., Vaganov, E. A., Hughes, M. K., Grissino-Mayer, H. D., & Cane, M. A. (2006). Forward modeling of regional scale tree-ring patterns in the southeastern United States and the recent influence of summer drought. *Geophysical Research Letters*, *33*(4), L04705. https://doi.org/10.1029/2005gl025050

Anchukaitis, K. J., & Smerdon, J. E. (2022). Progress and uncertainties in global and hemispheric temperature reconstructions of the Common Era. *Quaternary Science Reviews*, *286*, 107537. https://doi.org/10.1016/j.quascirev.2022.107537

Anchukaitis, K. J., Wilson, R., Briffa, K. R., Büntgen, U., Cook, E. R., D'Arrigo, R., et al. (2017). Last millennium northern hemisphere summer temperatures from tree rings: Part II, spatially resolved reconstructions. *Quaternary Science Reviews*, *163*, 1–22. https://doi.org/10.1016/j.quascirev.2017.02.020

Annan, J. D., & Hargreaves, J. C. (2012). Identification of climatic state with limited proxy data. *Climate of the Past*, *8*(4), 1141–1151. https://doi.org/10.5194/cp-8-1141-2012

Badgeley, J. A., Steig, E. J., Hakim, G. J., & Fudge, T. J. (2020). Greenland temperature and precipitation over the last 20000 years using data assimilation. *Climate of the Past*, *16*(4), 1325–1346. https://doi.org/10.5194/cp-16-1325-2020

Baek, S. H., Smerdon, J. E., Coats, S., Williams, A. P., Cook, B. I., Cook, E. R., & Seager, R. (2017). Precipitation, temperature, and teleconnection signals across the combined North American, monsoon Asia, and Old world drought atlases. *Journal of Climate*, *30*(18), 7141–7155. https://doi.org/10.1175/JCLI-D-16-0766.1

Beltrami, H. (2002). Climate from borehole data: Energy fluxes and temperatures since 1500. *Geophysical Research Letters*, *29*(23), 26-1–26-4. https://doi.org/10.1029/2002gl015702

Bhend, J., Franke, J., Folini, D., Wild, M., & Brönnimann, S. (2012). An ensemble-based approach to climate reconstructions. *Climate of the Past*, *8*(3), 963–976. https://doi.org/10.5194/cp-8-963-2012

Blasing, T. J., & Fritts, H. C. (1976). Reconstructing past climatic anomalies in north pacific and western north-America from tree-ring data. *Quaternary Research*, *6*(4), 563–579. https://doi.org/10.1016/0033-5894(76)90027-2

Breitenmoser, P., Broennimann, S., & Frank, D. (2014). Forward modelling of tree-ring width and comparison with a global network of tree-ring chronologies. *Climate of the Past*, *10*(2), 437–449. https://doi.org/10.5194/cp-10-437-2014

Bretherton, C., Smith, C., & Wallace, J. M. (1992). An intercomparison of methods for finding coupled patterns in climate data. *Journal of Climate*, *5*(6), 541–560. https://doi.org/10.1175/1520-0442(1992)005<0541:aiomff>2.0.co;2

Briffa, K. R., Jones, P. D., Wigley, T. M. L., Pilcher, J. R., & Baillie, M. G. L. (1986). Climate reconstruction from tree rings. 2. Spatial reconstruction of summer mean sea-level pressure patterns over Great-Britain. *Journal of Climatology*, *6*(1), 1–15. https://doi.org/10.1002/joc.3370060102

Broecker, W. S. (2001). Paleoclimate—Was the medieval warm period global? *Science*, *291*(5508), 1497–1499. https://doi.org/10.1126/science.291.5508.1497

Bruhwiler, L. M. P., Michalak, A. M., Peters, W., Baker, D. F., & Tans, P. (2005). An improved Kalman Smoother for atmospheric inversions. *Atmospheric Chemistry and Physics*, *5*(10), 2691–2702. https://doi.org/10.5194/acp-5-2691-2005

Büntgen, U., Allen, K., Anchukaitis, K. J., Arseneault, D., Boucher, É., Bräuning, A., et al. (2021a). The influence of decision-making in tree ring-based climate reconstructions. *Nature Communications*, *12*(1), 3411. https://doi.org/10.1038/s41467-021-23627-6

Büntgen, U., Urban, O., Krusic, P. J., Rybníček, M., Kolář, T., Kyncl, T., et al. (2021b). Recent European drought extremes beyond Common Era background variability. *Nature Geoscience*, *14*(4), 190–196. https://doi.org/10.1038/s41561-021-00698-0

Burger, G., & Cubasch, U. (2005). Are multiproxy climate reconstructions robust? *Geophysical Research Letters*, *32*(23), L23711. https://doi.org/10.1029/2005gl024155

Burger, G., Fast, I., & Cubasch, U. (2006). Climate reconstruction by regression—32 variations on a theme. *Tellus Series a-Dynamic Meteorology and Oceanography*, *58*(2), 227–235. https://doi.org/10.1111/j.1600-0870.2006.00164.x

Christiansen, B., & Ljungqvist, F. C. (2017). Challenges and perspectives for large-scale temperature reconstructions of the past two millennia. *Reviews of Geophysics*, *55*(1), 40–96. https://doi.org/10.1002/2016RG000521

Christiansen, B., Schmith, T., & Thejll, P. (2009). A surrogate ensemble study of climate reconstruction methods: Stochasticity and robustness. *Journal of Climate*, *22*(4), 951–976. https://doi.org/10.1175/2008jcli2301.1

Coats, S., Cook, B. I., Smerdon, J. E., & Seager, R. (2015). North American pancontinental droughts in model simulations of the last millennium. *Journal of Climate*, *28*(5), 2025–2043. https://doi.org/10.1175/jcli-d-14-00634.1

Coats, S., Smerdon, J. E., Seager, R., Cook, B. I., & González-Rouco, J. F. (2013). Megadroughts in southwestern North America in ECHO-G millennial simulations and their comparison to proxy drought reconstructions. *Journal of Climate*, *26*(19), 7635–7649. https://doi.org/10.1175/jcli-d-12-00603.1

Coats, S., Smerdon, J. E., Stevenson, S., Fasullo, J. T., Otto-Bliesner, B., & Ault, T. R. (2020). Paleoclimate constraints on the spatiotemporal character of past and future droughts. *Journal of Climate*, *33*(22), 9883–9903. https://doi.org/10.1175/jcli-d-20-0004.1

Cook, E. R., Anchukaitis, K. J., Buckley, B. M., D'Arrigo, R. D., Jacoby, G. C., & Wright, W. E. (2010a). Asian monsoon failure and megadrought during the last millennium. *Science*, *328*(5977), 486–489. https://doi.org/10.1126/science.1185188

Cook, E. R., Briffa, K. R., & Jones, P. D. (1985). Spatial regression methods for dendroclimatology: A review and comparison of two techniques; Part 1: The theory. A research report submitted to: The scientific Affairs division North Atlantic treaty organization, Brussels, Belgium (pp. 1–31).

Cook, E. R., Briffa, K. R., & Jones, P. D. (1994). Spatial regression methods in dendroclimatology—A review and comparison of 2 techniques. *International Journal of Climatology*, *14*(4), 379–402. https://doi.org/10.1002/joc.3370140404

Cook, E. R., & Kairiukstis, L. A. (1990). Methods of dendrochronology applications in the environmental sciences.

Cook, E. R., Meko, D. M., Stahle, D. W., & Cleaveland, M. K. (1996). *Tree-ring reconstructions of past drought across the coterminous United States: Tests of a regression method and calibration/verification results*. Radiocarbon.

Cook, E. R., Meko, D. M., Stahle, D. W., & Cleaveland, M. K. (1999). Drought reconstructions for the continental United States. *Journal of Climate*, *12*(4), 1145–1162. https://doi.org/10.1175/1520-0442(1999)012<1145:drftcu>2.0.co;2

Cook, E. R., Palmer, J. G., Ahmed, M., Woodhouse, C. A., Fenwick, P., Zafar, M. U., et al. (2013). Five centuries of Upper Indus River flow from tree rings. *Journal of Hydrology*, *486*, 365–375. https://doi.org/10.1016/j.jhydrol.2013.02.004

Cook, E. R., Seager, R., Heim, R. R., Jr., Vose, R. S., Herweijer, C., & Woodhouse, C. (2010b). Megadroughts in North America: Placing IPCC projections of hydroclimatic change in a long-term palaeoclimate context. *Journal of Quaternary Science*, *25*(1), 48–61. https://doi.org/10.1002/jqs.1303

Cook, E. R., Seager, R., Kushnir, Y., Briffa, K. R., Büntgen, U., Frank, D., et al. (2015). Old world megadroughts and pluvials during the Common Era. *Science Advances*, *1*(10), e1500561. https://doi.org/10.1126/sciadv.1500561

Cook, E. R., Solomina, O., Matskovsky, V., Cook, B. I., Agafonov, L., Berdnikova, A., et al. (2020). The European Russia drought atlas (1400–2016 CE). *Climate Dynamics*, *54*(3–4), 2317–2335. https://doi.org/10.1007/s00382-019-05115-2

Cook, E. R., Stahle, D. K., & Cleaveland, M. K. (1992). Dendroclimatic evidence from eastern North America. In *Climate since AD 1500* (1 ed., pp. 331–348). Taylor and Francis.

Cook, E. R., Woodhouse, C. A., Eakin, C. M., Meko, D. M., & Stahle, D. W. (2004). Long-term aridity changes in the western United States. *Science*, *306*(5698), 1015–1018. https://doi.org/10.1126/science.1102586

Crespin, E., Goosse, H., Fichefet, T., & Mann, M. E. (2009). The 15th century Arctic warming in coupled model simulations with data assimilation. *Climate of the Past*, *5*(3), 389–401. https://doi.org/10.5194/cp-5-389-2009

Dail, H., & Wunsch, C. (2014). Dynamical reconstruction of upper-ocean conditions in the last glacial maximum Atlantic. *Journal of Climate*, *27*(2), 807–823. https://doi.org/10.1175/jcli-d-13-00211.1

Dee, S. G., Emile-Geay, J., Evans, M. N., Allam, A., Steig, E. J., & Thompson, D. M. (2015). PRYSM: An open-source framework for PRoxY System Modeling, with applications to oxygen-isotope systems. *Journal of Advances in Modeling Earth Systems*, *7*(3), 1220–1247. https://doi.org/10.1002/2015MS000447

Dee, S. G., Steiger, N. J., Emile-Geay, J., & Hakim, G. J. (2016). On the utility of proxy system models for estimating climate states over the common era. *Journal of Advances in Modeling Earth Systems*, *8*(3), 1164–1179. https://doi.org/10.1002/2016MS000677

Diaz, H. F., & Wahl, E. R. (2015). Recent California water year precipitation deficits: A 440-year perspective. *Journal of Climate*, *28*(12), 4637–4652. https://doi.org/10.1175/jcli-d-14-00774.1

Dirren, S., & Hakim, G. J. (2005). Toward the assimilation of time-averaged observations. *Geophysical Research Letters*, *32*(4), L04804. https://doi.org/10.1029/2004GL021444

D'Odorico, P., Revelli, R., & Ridolfi, L. (2000). On the use of neural networks for dendroclimatic reconstructions. *Geophysical Research Letters*, *27*(6), 791–794. https://doi.org/10.1029/1999gl011049

Emile-Geay, J., & Eshleman, J. A. (2013). Toward a semantic web of paleoclimatology. *Geochemistry, Geophysics, Geosystems*, *14*(2), 457–469. https://doi.org/10.1002/ggge.20067

Emile-Geay, J., McKay, N. P., Kaufman, D. S., von Gunten, L., Wang, J., Anchukaitis, K. J., et al. (2017). A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data*, *4*(1), 170088. Data Descriptor. https://doi.org/10.1038/sdata.2017.88

Erb, M. P., McKay, N. P., Steiger, N., Dee, S., Hancock, C., Ivanovic, R. F., et al. (2022). Reconstructing Holocene temperatures in time and space using paleoclimate data assimilation. *Climate of the Past*, *18*(12), 2599–2629. https://doi.org/10.5194/cp-18-2599-2022

Esper, J., Schneider, L., Smerdon, J. E., Schoene, B. R., & Buentgen, U. (2015). Signals and memory in tree-ring width and density data. *Dendrochronologia*, *35*, 62–70. https://doi.org/10.1016/j.dendro.2015.07.001

Evans, M. N., Kaplan, A., & Cane, M. A. (1998). Optimal sites for coral-based reconstruction of global sea surface temperature. *Paleoceanography*, *13*(5), 502–516. https://doi.org/10.1029/98pa02132

Evans, M. N., Kaplan, A., & Cane, M. A. (2002). Pacific sea surface temperature field reconstruction from coral $\delta^{18}O$ data using reduced space objective analysis. *Paleoceanography*, *17*(1), 7-1–7-13. https://doi.org/10.1029/2000PA000590

Evans, M. N., Reichert, B. K., Kaplan, A., Anchukaitis, K. J., Vaganov, E. A., Hughes, M. K., & Cane, M. A. (2006). A forward modeling approach to paleoclimatic interpretation of tree-ring data. *Journal of Geophysical Research*, *111*(G3), G03008. https://doi.org/10.1029/2006jg000166

Evans, M. N., Smerdon, J. E., Kaplan, A., Tolwinski-Ward, S. E., & González-Rouco, J. F. (2014). Climate field reconstruction uncertainty arising from multivariate and nonlinear properties of predictors. *Geophysical Research Letters*, *41*(24), 9127–9134. https://doi.org/10.1002/2014gl062063

Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., & Anchukaitis, K. J. (2013). Applications of proxy system modeling in high resolution paleoclimatology. *Quaternary Science Reviews*, *76*, 16–28. https://doi.org/10.1016/j.quascirev.2013.05.024

Fang, M., & Li, X. (2016). Paleoclimate data assimilation: Its motivation, progress and prospects. *Science China Earth Sciences*, *59*(9), 1817–1826. https://doi.org/10.1007/s11430-015-5432-6

Fierro, R. D., Golub, G. H., Hansen, P. C., & Oleary, D. P. (1997). Regularization by truncated total least squares. *SIAM Journal on Scientific Computing*, *18*(4), 1223–1241. https://doi.org/10.1137/s1064827594263837

Frank, D., Esper, J., Zorita, E., & Wilson, R. (2010). A noodle, hockey stick, and spaghetti plate: A perspective on high-resolution paleoclimatology. *Wiley Interdisciplinary Reviews-Climate Change*, *1*(4), 507–516. https://doi.org/10.1002/wcc.53

Franke, J., Brönnimann, S., Bhend, J., & Brugnara, Y. (2017). A monthly global paleo-reanalysis of the atmosphere from 1600 to 2005 for studying past climatic variations. *Scientific Data*, *4*(1), 170076. https://doi.org/10.1038/sdata.2017.76

Franke, J., Gonzalez-Rouco, J. F., Frank, D., & Graham, N. E. (2011). 200 years of European temperature variability: Insights from and tests of the proxy surrogate reconstruction analog method. *Climate Dynamics*, *37*(1–2), 133–150. https://doi.org/10.1007/s00382-010-0802-6

Franke, J., Valler, V., Brönnimann, S., Neukom, R., & Jaume-Santero, F. (2020). The importance of input data quality and quantity in climate field reconstructions – Results from the assimilation of various tree-ring collections. *Climate of the Past*, *16*(3), 1061–1074. https://doi.org/10.5194/cp-16-1061-2020

Fritts, H. C. (1965). Tree-ring evidence for climatic changes in Western North America. *Monthly Weather Review*, *93*(7), 421–443. https://doi.org/10.1175/1520-0493(1965)093<0421:trefcc>2.3.co;2

Fritts, H. C. (1972). Tree rings and climate. *Scientific American*, *226*(5), 92–100. https://doi.org/10.1038/scientificamerican0572-92

Fritts, H. C. (1976). *Tree rings and climate*. Academic Press.

Fritts, H. C. (1991). *Reconstructing large-scale climatic patterns from tree-ring data: A diagnostic analysis* (p. 286). University of Arizona Press.

Fritts, H. C., Blasing, T. J., Hayden, B. P., & Kutzbach, J. E. (1971). Multivariate techniques for specifying tree-growth and climate relationships and for reconstructing anomalies in paleoclimate. *Journal of Applied Meteorology and Climatology*, *10*(5), 845–864. https://doi.org/10.1175/1520-0450(1971)010<0845:mtfstg>2.0.co;2

Fritts, H. C., Lofgren, G. R., & Gordon, G. A. (1979). Variations in climate since 1602 as reconstructed from tree rings. *Quaternary Research*, *12*(1), 18–46. https://doi.org/10.1016/0033-5894(79)90090-5

Fritts, H. C., Lofgren, G. R., & Gordon, G. A. (1980). Past climate reconstructed from tree rings. *Journal of Interdisciplinary History*, *10*(4), 773–793. https://doi.org/10.2307/203070

Gille, E. P., Wahl, E. R., Vose, R. S., & Cook, E. R. (2017). NOAA/WDS paleoclimatology—Living blended drought atlas (LBDA) version 2—Recalibrated reconstruction of United States summer PMDI over the last 2000 years [Dataset]. NOAA National Centers for Environmental Information. https://doi.org/10.25921/7xm8-jn36

Glahn, H. R. (1968). Canonical correlation and its relationship to discriminate analysis and multiple regression. *Bulletin of the American Meteorological Society*, *49*(3), 312.

Gómez-Navarro, J. J., Zorita, E., Raible, C. C., & Neukom, R. (2017). Pseudo-proxy tests of the analogue method to reconstruct spatially resolved global temperature during the Common Era. *Climate of the Past*, *13*(6), 629–648. https://doi.org/10.5194/cp-13-629-2017

Goosse, H., Crespin, E., de Montety, A., Mann, M. E., Renssen, H., & Timmermann, A. (2010). Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation. *Journal of Geophysical Research*, *115*(D9), D09108. https://doi.org/10.1029/2009JD012737

Groveman, B. S., & Landsberg, H. E. (1979). Simulated northern hemisphere temperature departures 1579–1880. *Geophysical Research Letters*, *6*(10), 767–769. https://doi.org/10.1029/gl006i010p00767

Guillot, D., Rajaratnam, B., & Emile-Geay, J. (2015). Statistical paleoclimate reconstructions via Markov random fields. *Annals of Applied Statistics*, *9*(1), 324–352. https://doi.org/10.1214/14-aoas794

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, *19*(2), 149–161. https://doi.org/10.1007/bf02289162

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., et al. (2016). The last millennium climate reanalysis project: Framework and first results. *Journal of Geophysical Research-Atmospheres*, *121*(12), 6745–6764. https://doi.org/10.1002/2016jd024751

Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate data set. *Scientific Data*, *7*(1), 109. https://doi.org/10.1038/s41597-020-0453-3

Harris, R. N., & Chapman, D. S. (2001). Mid-Latitude (30 degrees-60 degrees N) climatic warming inferred by combining borehole temperatures with surface air temperatures. *Geophysical Research Letters*, *28*(5), 747–750. https://doi.org/10.1029/2000gl012348

Hegerl, G. C., Crowley, T. J., Allen, M., Hyde, W. T., Pollack, H. N., Smerdon, J., & Zorita, E. (2007). Detection of human influence on a new, validated 1500-year temperature reconstruction. *Journal of Climate*, *20*(4), 650–666. https://doi.org/10.1175/jcli4011.1

Herweijer, C., Seager, R., Cook, E. R., & Emile-Geay, J. (2007). North American droughts of the last millennium from a gridded network of tree-ring data. *Journal of Climate*, *20*(7), 1353–1376. https://doi.org/10.1175/jcli4042.1

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression—Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Holmstrom, I. (1963). On a method for parametric representation of the state of the atmosphere. *Tellus*, *15*(2), 127–149. https://doi.org/10.1111/j.2153-3490.1963.tb01372.x

Huang, S. P., Pollack, H. N., & Shen, P. Y. (2000). Temperature trends ever the past five centuries reconstructed from borehole temperatures. *Nature*, *403*(6771), 756–758. https://doi.org/10.1038/35001556

Huang, S. P., Shen, P. Y., & Pollack, H. N. (1996). Deriving century-long trends of surface temperature change from borehole temperatures. *Geophysical Research Letters*, *23*(3), 257–260. https://doi.org/10.1029/96gl00020

Huntley, H. S., & Hakim, G. J. (2010). Assimilation of time-averaged observations in a quasi-geostrophic atmospheric jet model. *Climate Dynamics*, *35*(6), 995–1009. https://doi.org/10.1007/s00382-009-0714-5

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time-series model selection in small samples. *Biometrika*, *76*(2), 297–307. https://doi.org/10.1093/biomet/76.2.297

Jacoby, G. C., & Darrigo, R. (1989). Reconstructed northern hemisphere annual temperature since 1671 based on high-latitude tree-ring data from north-America. *Climatic Change*, *14*(1), 39–59. https://doi.org/10.1007/bf00140174

Jones, P. D., Briffa, K. R., Barnett, T. P., & Tett, S. F. B. (1998). High-resolution palaeoclimatic records for the last millennium: Interpretation, integration and comparison with general circulation model control-run temperatures. *The Holocene*, *8*(4), 455–471. https://doi.org/10.1191/095968398667194956

Jones, P. D., Briffa, K. R., Osborn, T. J., Lough, J. M., van Ommen, T. D., Vinther, B. M., et al. (2009). High-resolution palaeoclimatology of the last millennium: A review of current status and future prospects. *The Holocene*, *19*(1), 3–49. https://doi.org/10.1177/0959683608098952

Jones, P. D., & Mann, M. E. (2004). Climate over past millennia. *Reviews of Geophysics*, *42*(2), RG2002. https://doi.org/10.1029/2003rg000143

Kadow, C., Hall, D. M., & Ulbrich, U. (2020). Artificial intelligence reconstructs missing climate information. *Nature Geoscience*, *13*(6), 408–413. https://doi.org/10.1038/s41561-020-0582-5

Kaiser, H. F. (1960). The application of electronic-computers to factor-analysis. *Educational and Psychological Measurement*, *20*(1), 141–151. https://doi.org/10.1177/001316446002000116

Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press.

Karl, T. R., & Koscielny, A. J. (1982). Drought in the united-states—1895–1981. *Journal of Climatology*, *2*(4), 313–329. https://doi.org/10.1002/joc.3370020402

King, J. M., Anchukaitis, K. J., Tierney, J. E., Hakim, G. J., Emile-Geay, J., Zhu, F., & Wilson, R. (2021). A data assimilation approach to last millennium temperature field reconstruction using a limited high-sensitivity proxy network. *Journal of Climate*, *18*, 1–64. https://doi.org/10.1175/jcli-d-20-0661.1

Konecky, B. L., McKay, N. P., Churakova, O. V., Comas-Bru, L., Dassié, E. P., DeLong, K. L., et al. (2020). The Iso2k database: A global compilation of paleo-$\delta^{18}O$ and $\delta^{2}H$ records to aid understanding of Common Era climate. *Earth System Science Data*, *12*(3), 2261–2288. https://doi.org/10.5194/essd-12-2261-2020

Kurahashi-Nakamura, T., Losch, M., & Paul, A. (2014). Can sparse proxy data constrain the strength of the Atlantic meridional overturning circulation? *Geoscientific Model Development*, *7*(1), 419–432. https://doi.org/10.5194/gmd-7-419-2014

Kurahashi-Nakamura, T., Paul, A., & Losch, M. (2017). Dynamical reconstruction of the global ocean state during the Last Glacial Maximum. *Paleoceanography*, *32*(4), 326–350. https://doi.org/10.1002/2016PA003001

Kutzbach, J. E. (1967). Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North America. *Journal of Applied Meteorology and Climatology*, *6*(5), 791–802. https://doi.org/10.1175/1520-0450(1967)006<0791:eeoslp>2.0.co;2

Kutzbach, J. E., & Guetter, P. J. (1980). On the design of paleoenvironmental data-networks for estimating large-scale patterns of climate. *Quaternary Research*, *14*(2), 169–187. https://doi.org/10.1016/0033-5894(80)90046-0

Lamarche, V. C., & Fritts, H. C. (1971). Anomaly patterns of climate over western united-states, 1700–1930, derived from principal component analysis of tree-ring data. *Monthly Weather Review*, *99*(2), 138–142. https://doi.org/10.1175/1520-0493(1971)099<0138:apocot>2.3.co;2

Landsberg, H. E., Groveman, B. S., & Hakkarinen, I. M. (1978). Simple method for approximating annual temperature of northern hemisphere. *Geophysical Research Letters*, *5*(6), 505–506. https://doi.org/10.1029/gl005i006p00505

Leavitt, S. W., Cook, E. R., & Hughes, M. K. (2019). Harold Clark Fritts 1928–2019 in memoriam. *Tree-Ring Research*, *75*(2), 167–169. https://doi.org/10.3959/1536-1098-75.2.167

LeGrand, P., & Wunsch, C. (1995). Constraints from paleotracer data on the North Atlantic circulation during the last glacial maximum. *Paleoceanography*, *10*(6), 1011–1045. https://doi.org/10.1029/95PA01455
Li, B., Nychka, D. W., & Ammann, C. M. (2007). The 'hockey stick' and the 1990s: A statistical perspective on reconstructing hemispheric temperatures. *Tellus A: Dynamic Meteorology and Oceanography*, *59*(5), 591–598. https://doi.org/10.1111/j.1600-0870.2007.00270.x
Li, B., & Smerdon, J. E. (2012). Defining spatial comparison metrics for evaluation of paleoclimatic field reconstructions of the Common Era. *Environmetrics*, *23*(5), 394–406. https://doi.org/10.1002/env.2142
Li, B., Zhang, X., & Smerdon, J. E. (2016). Comparison between spatio-temporal random processes and application to climate model data. *Environmetrics*, *27*(5), 267–279. https://doi.org/10.1002/env.2395
Lorenz, E. N. (1956). Empirical orthogonal functions and statistical weather prediction. In *Statistical forecasting project, Department of meteorology, Massachusetts Institute of Technology, Science Report No. 1*.
Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M., & Wanner, H. (2004). European seasonal and annual temperature variability, trends, and extremes since 1500. *Science*, *303*(5663), 1499–1503. https://doi.org/10.1126/science.1093877
Luterbacher, J., Werner, J. P., Smerdon, J. E., Fernandez-Donado, L., González-Rouco, F. J., Barriopedro, D., et al. (2016). European summer temperatures since Roman times. *Environmental Research Letters*, *11*(2), 024001. https://doi.org/10.1088/1748-9326/11/2/024001
Luterbacher, J., Xoplaki, E., Dietrich, D., Rickli, R., Jacobeit, J., Beck, C., et al. (2002). Reconstruction of sea level pressure fields over the Eastern North Atlantic and Europe back to 1500. *Climate Dynamics*, *18*(7), 545–561. https://doi.org/10.1007/s00382-001-0196-6
Mann, M. E. (2007). Climate over the past two millennia. *Annual Review of Earth and Planetary Sciences*, *35*(1), 111–136. https://doi.org/10.1146/annurev.earth.35.031306.140042
Mann, M. E., Bradley, R. S., & Hughes, M. K. (1998). Global-scale temperature patterns and climate forcing over the past six centuries. *Nature*, *392*(6678), 779–787. https://doi.org/10.1038/33859
Mann, M. E., Bradley, R. S., & Hughes, M. K. (1999). Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophysical Research Letters*, *26*(6), 759–762. https://doi.org/10.1029/1999gl900070
Mann, M. E., & Jones, P. D. (2003). Global surface temperatures over the past two millennia. *Geophysical Research Letters*, *30*(15), 1820. https://doi.org/10.1029/2003gl017814
Mann, M. E., Rutherford, S., Wahl, E., & Ammann, C. (2005). Testing the fidelity of methods used in proxy-based reconstructions of past climate. *Journal of Climate*, *18*(20), 4097–4107. https://doi.org/10.1175/jcli3564.1
Mann, M. E., Rutherford, S., Wahl, E., & Ammann, C. (2007). Robustness of proxy-based climate field reconstruction methods. *Journal of Geophysical Research*, *112*(D12), D12109. https://doi.org/10.1029/2006jd008272
Mann, M. E., Zhang, Z., Rutherford, S., Bradley, R. S., Hughes, M. K., Shindell, D., et al. (2009). Global signatures and dynamical origins of the little ice age and medieval climate anomaly. *Science*, *326*(5957), 1256–1260. https://doi.org/10.1126/science.1177303
Matsikaris, A., Widmann, M., & Jungclaus, J. (2015). On-line and off-line data assimilation in palaeoclimatology: A case study. *Climate of the Past*, *11*(1), 81–93. https://doi.org/10.5194/cp-11-81-2015
Matsikaris, A., Widmann, M., & Jungclaus, J. (2016). Assimilating continental mean temperatures to reconstruct the climate of the late pre-industrial period. *Climate Dynamics*, *46*(11–12), 3547–3566. https://doi.org/10.1007/s00382-015-2785-9
McGregor, H. V., Martrat, B., Evans, M. N., Thompson, D., Reynolds, D., & Addison, J. A. (2016). Data, age uncertainties and ocean $\delta^{18}O$ under the spotlight for Ocean2k Phase 2. *Global Change*, *24*(1), 44. https://doi.org/10.22498/pages.24.1.44
McKay, N. P., & Emile-Geay, J. (2016). Technical note: The linked paleo data framework—A common tongue for paleoclimatology. *Climate of the Past*, *12*(4), 1093–1100. https://doi.org/10.5194/cp-12-1093-2016
Meko, D. M. (1997). Dendroclimatic reconstruction with time varying predictor subsets of tree indices. *Journal of Climate*, *10*(4), 687–696. https://doi.org/10.1175/1520-0442(1997)010<0687:drwtvp>2.0.co;2
Michalak, A. M. (2008). Technical Note: Adapting a fixed-lag Kalman smoother to a geostatistical atmospheric inversion framework. *Atmospheric Chemistry and Physics*, *8*(22), 6789–6799. https://doi.org/10.5194/acp-8-6789-2008
Monteleoni, C., Schmidt, G. A., Alexander, F., Niculescu-Mizil, A., Steinhaeuser, K., Tippett, M., et al. (2013). Climate informatics.
Morales, M. S., Cook, E. R., Barichivich, J., Christie, D. A., Villalba, R., LeQuesne, C., et al. (2020). Six hundred years of South American tree rings reveal an increase in severe hydroclimatic events since mid-20th century. *Proceedings of the National Academy of Sciences*, *117*(29), 16816–16823. https://doi.org/10.1073/pnas.2002411117
Neukom, R., Gergis, J., Karoly, D. J., Wanner, H., Curran, M., Elbert, J., et al. (2014). Inter-hemispheric temperature variability over the past millennium. *Nature Climate Change*, *4*(5), 362–367. https://doi.org/10.1038/nclimate2174
Neukom, R., Luterbacher, J., Villalba, R., Kuettel, M., Frank, D., Jones, P. D., et al. (2010). Multi-centennial summer and winter precipitation variability in southern South America. *Geophysical Research Letters*, *37*(14), L14708. https://doi.org/10.1029/2010gl043680
Neukom, R., Luterbacher, J., Villalba, R., Küttel, M., Frank, D., Jones, P. D., et al. (2011). Multiproxy summer and winter surface air temperature field reconstructions for southern South America covering the past centuries. *Climate Dynamics*, *37*(1), 35–51. https://doi.org/10.1007/s00382-010-0793-3
Neukom, R., Steiger, N., Gómez-Navarro, J. J., Wang, J., & Werner, J. P. (2019). No evidence for globally coherent warm and cold periods over the preindustrial Common Era. *Nature*, *571*(7766), 550–554. https://doi.org/10.1038/s41586-019-1401-2
Ni, F. B., Cavazos, T., Hughes, M. K., Comrie, A. C., & Funkhouser, G. (2002). Cool-season precipitation in the southwestern USA since AD 1000: Comparison of linear and nonlinear techniques for reconstruction. *International Journal of Climatology*, *22*(13), 1645–1662. https://doi.org/10.1002/joc.804
Nicault, A., Alleaume, S., Brewer, S., Carrer, M., Nola, P., & Guiot, J. (2008). Mediterranean drought fluctuation during the last 500 years based on tree-ring data. *Climate Dynamics*, *31*(2–3), 227–245. https://doi.org/10.1007/s00382-007-0349-3
Okazaki, A., & Yoshimura, K. (2017). Development and evaluation of a system of proxy data assimilation for paleoclimate reconstruction. *Climate of the Past*, *13*(4), 379–393. https://doi.org/10.5194/cp-13-379-2017
Olive, D. J. (2007). Prediction intervals for regression models. *Computational Statistics & Data Analysis*, *51*(6), 3115–3122. https://doi.org/10.1016/j.csda.2006.02.006
Osman, M. B., Tierney, J. E., Zhu, J., Tardif, R., Hakim, G. J., King, J., & Poulsen, C. J. (2021). Globally resolved surface temperatures since the last glacial maximum. *Nature*, *599*(7884), 239–244. https://doi.org/10.1038/s41586-021-03984-4
Otto-Bliesner, B. L., Brady, E. C., Fasullo, J., Jahn, A., Landrum, L., Stevenson, S., et al. (2016). Climate variability and change since 850 CE an ensemble approach with the community earth system model. *Bulletin of the American Meteorological Society*, *97*(5), 735–754. https://doi.org/10.1175/bams-d-14-00233.1
PAGES2k Consortium. (2013). Continental-scale temperature variability during the past two millennia. *Nature Geoscience*, *6*(5), 339–346. https://doi.org/10.1038/ngeo1797

PAGES2k Consortium, Neukom, R., Barboza, L. A., Erb, M. P., Shi, F., Emile-Geay, J., et al. (2019). Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era. *Nature Geoscience*, *12*(8), 643–649. https://doi.org/10.1038/s41561-019-0400-0

PAGES Hydro2k Consortium. (2017). Comparing proxy and model estimates of hydroclimate variability and change over the Common Era. *Climate of the Past*, *13*(12), 1851–1900. https://doi.org/10.5194/cp-13-1851-2017

Palmer, J. G., Cook, E. R., Turney, C. S. M., Allen, K., Fenwick, P., Cook, B. I., et al. (2015). Drought variability in the eastern Australia and New Zealand summer drought atlas (ANZDA, CE 1500–2012) modulated by the Interdecadal Pacific Oscillation. *Environmental Research Letters*, *10*(12), 124002. https://doi.org/10.1088/1748-9326/10/12/124002

Palmer, W. C. (1965). Meteorological drought. In *Research paper no. 45* (p. 58).

Parsons, L. A., Amrhein, D. E., Sanchez, S. C., Tardif, R., Brennan, M. K., & Hakim, G. J. (2021). Do multi-model ensembles improve reconstruction skill in paleoclimate data assimilation? *Earth and Space Science*, *8*(4), e2020EA001467. https://doi.org/10.1029/2020EA001467

Pauling, A., Luterbacher, J., Casty, C., & Wanner, H. (2006). Five hundred years of gridded high-resolution precipitation reconstructions over Europe and the connection to large-scale circulation. *Climate Dynamics*, *26*(4), 387–405. https://doi.org/10.1007/s00382-005-0090-8

Pendergrass, A. G., Hakim, G. J., Battisti, D. S., & Roe, G. (2012). Coupled air-mixed layer temperature predictability for climate reconstruction. *Journal of Climate*, *25*(2), 459–472. https://doi.org/10.1175/2011jcli4094.1

Perkins, W. A., & Hakim, G. J. (2017). Reconstructing paleoclimate fields using online data assimilation with a linear inverse model. *Climate of the Past*, *13*(5), 421–436. https://doi.org/10.5194/cp-13-421-2017

Perkins, W. A., & Hakim, G. J. (2021). Coupled atmosphere–ocean reconstruction of the last millennium using online data assimilation. *Paleoceanography and Paleoclimatology*, *36*(5), e2020PA003959. https://doi.org/10.1029/2020PA003959

Pollack, H. N., & Chapman, D. S. (1993). Underground records of changing climate. *Scientific American*, *268*(6), 44–50. https://doi.org/10.1038/scientificamerican0693-44

Pollack, H. N., & Huang, S. P. (2000). Climate reconstruction from subsurface temperatures. *Annual Review of Earth and Planetary Sciences*, *28*(1), 339–365. https://doi.org/10.1146/annurev.earth.28.1.339

Pollack, H. N., Huang, S. P., & Shen, P. Y. (1998). Climate change record in subsurface temperatures: A global perspective. *Science*, *282*(5387), 279–281. https://doi.org/10.1126/science.282.5387.279

Preisendorfer, R. W., Zwiers, F. W., & Barnett, T. P. (1981). *Foundations of principal component selection rules*. Scripps Institution of Oceanography.

Quan, N. T. (1988). The prediction sum of squares as a general measure for regression diagnostics. *Journal of Business & Economic Statistics*, *6*(4), 501–504. https://doi.org/10.2307/1391469

Rao, M. P., Cook, E. R., Cook, B. I., Palmer, J. G., Uriarte, M., Devineni, N., et al. (2018). Six centuries of upper indus basin streamflow variability and its climatic drivers. *Water Resources Research*, *54*(8), 5687–5701. https://doi.org/10.1029/2018wr023080

Riedwyl, N., Kuettel, M., Luterbacher, J., & Wanner, H. (2009). Comparison of climate field reconstruction techniques: Application to Europe. *Climate Dynamics*, *32*(2–3), 381–395. https://doi.org/10.1007/s00382-008-0395-5

Rutherford, S., Mann, M. E., Delworth, T. L., & Stouffer, R. J. (2003). Climate field reconstruction under stationary and nonstationary forcing. *Journal of Climate*, *16*(3), 462–479. https://doi.org/10.1175/1520-0442(2003)016<0462:cfrusa>2.0.co;2

Rutherford, S., Mann, M. E., Osborn, T. J., Bradley, R. S., Briffa, K. R., Hughes, M. K., & Jones, P. D. (2005). Proxy-based Northern Hemisphere surface temperature reconstructions: Sensitivity to method, predictor network, target season, and target domain. *Journal of Climate*, *18*(13), 2308–2329. https://doi.org/10.1175/jcli3351.1

Salonen, J. S., Korpela, M., Williams, J. W., & Luoto, M. (2019). Machine-learning based reconstructions of primary and secondary climate variables from North American and European fossil pollen data. *Scientific Reports*, *9*(1), 15805. https://doi.org/10.1038/s41598-019-52293-4

Samakinwa, E., Valler, V., Hand, R., Neukom, R., Gómez-Navarro, J. J., Kennedy, J., et al. (2021). An ensemble reconstruction of global monthly sea surface temperature and sea ice concentration 1000–1849. *Scientific Data*, *8*(1), 261. https://doi.org/10.1038/s41597-021-01043-1

Sanchez, S. C., Hakim, G. J., & Saenger, C. P. (2021). Climate model teleconnection patterns govern the Niño-3.4 response to early Nineteenth-century volcanism in coral-based data assimilation reconstructions. *Journal of Climate*, *34*(5), 1863–1880. https://doi.org/10.1175/jcli-d-20-0549.1

Sarachik, E. S., & Cane, M. A. (2010). *The El Niño-southern oscillation phenomenon*. Cambridge University Press.

Schmidt, G. A., Annan, J. D., Bartlein, P. J., Cook, B. I., Guilyardi, E., Hargreaves, J. C., et al. (2014). Using palaeo-climate comparisons to constrain future projections in CMIP5. *Climate of the Past*, *10*(1), 221–250. https://doi.org/10.5194/cp-10-221-2014

Schneider, L., Smerdon, J. E., Buentgen, U., Wilson, R. J. S., Myglan, V. S., Kirdyanov, A. V., & Esper, J. (2015). Revising midlatitude summer temperatures back to AD600 based on a wood density network. *Geophysical Research Letters*, *42*(11), 4556–4562. https://doi.org/10.1002/2015gl063956

Schneider, T. (2001). Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *Journal of Climate*, *14*(5), 853–871. https://doi.org/10.1175/1520-0442(2001)014<0853:aoicde>2.0.co;2

Seager, R. (2007). The turn of the century North American drought: Global context, dynamics, and past analogs. *Journal of Climate*, *20*(22), 5527–5552. https://doi.org/10.1175/2007jcli1529.1

Seager, R., Graham, N., Herweijer, C., Gordon, A. L., Kushnir, Y., & Cook, E. (2007). Blueprints for medieval hydroclimate. *Quaternary Science Reviews*, *26*(19–21), 2322–2336. https://doi.org/10.1016/j.quascirev.2007.04.020

Seber, G. A. F., & Lee, A. J. (2003). Confidence intervals and regions. In *Linear regression analysis* (pp. 119–137).

Smerdon, J. E. (2012). Climate models as a test bed for climate reconstruction methods: Pseudoproxy experiments. *Wiley Interdisciplinary Reviews-Climate Change*, *3*(1), 63–77. https://doi.org/10.1002/wcc.149

Smerdon, J. E., Coats, S., & Ault, T. R. (2016). Model-dependent spatial skill in pseudoproxy experiments testing climate field reconstruction methods for the Common Era. *Climate Dynamics*, *46*(5–6), 1921–1942. https://doi.org/10.1007/s00382-015-2684-0

Smerdon, J. E., & Kaplan, A. (2007). Comments on "testing the fidelity of methods used in proxy-based reconstructions of past climate": The role of the standardization interval. *Journal of Climate*, *20*(22), 5666–5670. https://doi.org/10.1175/2007jcli1794.1

Smerdon, J. E., Kaplan, A., & Amrhein, D. E. (2010). Erroneous model field representations in multiple pseudoproxy studies: Corrections and implications. *Journal of Climate*, *23*(20), 5548–5554. https://doi.org/10.1175/2010jcli3742.1

Smerdon, J. E., Kaplan, A., & Chang, D. (2008). On the origin of the standardization sensitivity in RegEM climate field reconstructions. *Journal of Climate*, *21*(24), 6710–6723. https://doi.org/10.1175/2008jcli2182.1

Smerdon, J. E., Kaplan, A., Chang, D., & Evans, M. N. (2010). A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millennium. *Journal of Climate*, *23*(18), 4856–4880. https://doi.org/10.1175/2010jcli3328.1

Smerdon, J. E., Kaplan, A., Zorita, E., González-Rouco, J. F., & Evans, M. N. (2011). Spatial performance of four climate field reconstruction methods targeting the Common Era. *Geophysical Research Letters*, *38*(11), L11705. https://doi.org/10.1029/2011gl047372

Smerdon, J. E., & Pollack, H. N. (2016). Reconstructing earth's surface temperature over the past 2000 years: The science behind the headlines. *Wiley Interdisciplinary Reviews: Climate Change*, *7*(5), 746–771. https://doi.org/10.1002/wcc.418

Stahle, D. W., Cook, E. R., Burnette, D. J., Torbenson, M. C. A., Howard, I. M., Griffin, D., et al. (2020). Dynamics, variability, and change in seasonal precipitation reconstructions for North America. *Journal of Climate*, *33*(8), 3173–3195. https://doi.org/10.1175/jcli-d-19-0270.1

Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S., & Roe, G. H. (2014). Assimilation of time-averaged pseudoproxies for climate reconstruction. *Journal of Climate*, *27*(1), 426–441. https://doi.org/10.1175/JCLI-D-12-00693.1

Steiger, N. J., & Smerdon, J. E. (2017). A pseudoproxy assessment of data assimilation for reconstructing the atmosphere-ocean dynamics of hydroclimate extremes. *Climate of the Past Discussions*, *13*(10), 1–22. https://doi.org/10.5194/cp-13-1435-2017

Steiger, N. J., Smerdon, J. E., Cook, B. I., Seager, R., Williams, A. P., & Cook, E. R. (2019). Oceanic and radiative forcing of medieval megadroughts in the American Southwest. *Science Advances*, *5*(7), eaax0087. https://doi.org/10.1126/sciadv.aax0087

Steiger, N. J., Smerdon, J. E., Cook, E. R., & Cook, B. I. (2018). A reconstruction of global hydroclimate and dynamical variables over the Common Era. *Scientific Data*, *5*(1), 180086. Data Descriptor. https://doi.org/10.1038/sdata.2018.86

Steiger, N. J., Steig, E. J., Dee, S. G., Roe, G. H., & Hakim, G. J. (2017). Climate reconstruction using data assimilation of water isotope ratios from ice cores. *Journal of Geophysical Research: Atmospheres*, *122*(3), 1545–1568. https://doi.org/10.1002/2016jd026011

St. George, S., Meko, D. M., & Cook, E. R. (2010). The seasonality of precipitation signals embedded within the North American Drought Atlas. *The Holocene*, *20*(6), 983–988. https://doi.org/10.1177/0959683610365937

Stockton, C. W., & Meko, D. M. (1975). A long-term history of drought occurrence in western United States as inferred from tree rings. *Weatherwise*, *28*(6), 244–249. https://doi.org/10.1080/00431672.1975.9931775

Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., et al. (2019). Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling. *Climate of the Past*, *15*(4), 1251–1273. https://doi.org/10.5194/cp-15-1251-2019

Tejedor, E., Steiger, N., Smerdon, J. E., Serrano-Notivoli, R., & Vuille, M. (2021a). Global temperature responses to large tropical volcanic eruptions in paleo data assimilation products and climate model simulations over the last millennium. *Paleoceanography and Paleoclimatology*, *36*(4), e2020PA004128. https://doi.org/10.1029/2020pa004128

Tejedor, E., Steiger, N. J., Smerdon, J. E., Serrano-Notivoli, R., & Vuille, M. (2021b). Global hydroclimatic response to tropical volcanic eruptions over the last millennium. *Proceedings of the National Academy of Sciences*, *118*(12), e2019145118. https://doi.org/10.1073/pnas.2019145118

Thompson, D. M., Ault, T. R., Evans, M. N., Cole, J. E., & Emile-Geay, J. (2011). Comparison of observed and simulated tropical climate trends using a forward model of coral $\delta^{18}O$. *Geophysical Research Letters*, *38*(14), L14706. https://doi.org/10.1029/2011gl048224

Tierney, J. E., Abram, N. J., Anchukaitis, K. J., Evans, M. N., Giry, C., Kilbourne, K. H., et al. (2015). Tropical sea surface temperatures for the past four centuries reconstructed from coral archives. *Paleoceanography*, *30*(3), 226–252. https://doi.org/10.1002/2014PA002717

Tierney, J. E., Zhu, J., King, J., Malevich, S. B., Hakim, G. J., & Poulsen, C. J. (2020). Glacial cooling and climate sensitivity revisited. *Nature*, *584*(7822), 569–573. https://doi.org/10.1038/s41586-020-2617-x

Tikhonov, A. N., & Arsenin, V. Y. (1970). *Solution of ill-posed problems*. Winston and Sons.

Tingley, M. P., Craigmile, P. F., Haran, M., Li, B., Mannshardt, E., & Rajaratnam, B. (2012). Piecing together the past: Statistical insights into paleoclimatic reconstructions. *Quaternary Science Reviews*, *35*, 1–22. https://doi.org/10.1016/j.quascirev.2012.01.012

Tingley, M. P., & Huybers, P. (2010a). A bayesian algorithm for reconstructing climate anomalies in space and time. Part I: Development and applications to paleoclimate reconstruction problems. *Journal of Climate*, *23*(10), 2759–2781. https://doi.org/10.1175/2009jcli3015.1

Tingley, M. P., & Huybers, P. (2010b). A bayesian algorithm for reconstructing climate anomalies in space and time. Part II: Comparison with the regularized expectation-maximization algorithm. *Journal of Climate*, *23*(10), 2782–2800. https://doi.org/10.1175/2009jcli3016.1

Trevino, A. M., Stine, A. R., & Huybers, P. (2021). Regional nonlinear relationships across the United States between drought and tree-ring width variability from a neural network. *Geophysical Research Letters*, *48*(14), e2020GL092090. https://doi.org/10.1029/2020gl092090

Valler, V., Franke, J., Brugnara, Y., & Brönnimann, S. (2021). An updated global atmospheric paleo-reanalysis covering the last 400 years. *Geoscience Data Journal*, *9*(1), 89–107. https://doi.org/10.1002/gdj3.121

van der Schrier, G., & Barkmeijer, J. (2005). Bjerknes' hypothesis on the coldness during AD 1790–1820 revisited. *Climate Dynamics*, *24*(4), 355–371. https://doi.org/10.1007/s00382-004-0506-x

Van Huffel, S., & Vandewalle, J. (1991). The total least squares problem: Computational aspects and analysis. In *Paper presented at the frontiers in applied Mathematics, Philadelphia, PA*.

Vinod, H. D. (2006). Maximum entropy ensembles for time series inference in economics. *Journal of Asian Economics*, *17*(6), 955–978. https://doi.org/10.1016/j.asieco.2006.09.001

von Storch, H., Cubasch, U., Gonzalez-Rouco, J. F., Jones, J. M., Voss, R., Widmann, M., & Zorita, E. (2000). Combining paleoclimatic evidence and GCMs by means of data assimilation through upscaling and nudging (DATUN). In *Proceedings of the 11th symposium on global change studies* (pp. 28–31).

Wahl, E. R., Diaz, H. F., Smerdon, J. E., & Ammann, C. M. (2014). Late winter temperature response to large tropical volcanic eruptions in temperate western North America: Relationship to ENSO phases. *Global and Planetary Change*, *122*, 238–250. https://doi.org/10.1016/j.gloplacha.2014.08.005

Wahl, E. R., & Morrill, C. (2010). Climate change toward understanding and predicting monsoon patterns. *Science*, *328*(5977), 437–438. https://doi.org/10.1126/science.1188926

Wahl, E. R., & Smerdon, J. E. (2012). Comparative performance of paleoclimate field and index reconstructions derived from climate proxies and noise-only predictors. *Geophysical Research Letters*, *39*(6), L06703. https://doi.org/10.1029/2012gl051086

Wang, J., Emile-Geay, J., Guillot, D., McKay, N. P., & Rajaratnam, B. (2015). Fragility of reconstructed temperature patterns over the Common Era: Implications for model evaluation. *Geophysical Research Letters*, *42*(17), 7162–7170. https://doi.org/10.1002/2015gl065265

Wang, J., Emile-Geay, J., Guillot, D., Smerdon, J. E., & Rajaratnam, B. (2014). Evaluating climate field reconstruction techniques using improved emulations of real-world conditions. *Climate of the Past*, *10*(1), 1–19. https://doi.org/10.5194/cp-10-1-2014

Werner, J. P., Luterbacher, J., & Smerdon, J. E. (2013). A pseudoproxy evaluation of bayesian hierarchical modeling and canonical correlation analysis for climate field reconstructions over Europe. *Journal of Climate*, *26*(3), 851–867. https://doi.org/10.1175/jcli-d-12-00016.1

Widmann, M., Goosse, H., van der Schrier, G., Schnur, R., & Barkmeijer, J. (2010). Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium. *Climate of the Past*, *6*(5), 627–644. https://doi.org/10.5194/cp-6-627-2010

Williams, A. P., Cook, B. I., & Smerdon, J. E. (2022). Rapid intensification of the emerging southwestern North American megadrought in 2020–2021. *Nature Climate Change*, *12*(3), 232–234. https://doi.org/10.1038/s41558-022-01290-z

Williams, A. P., Cook, E. R., Smerdon, J. E., Cook, B. I., Abatzoglou, J. T., Bolles, K., et al. (2020). Large contribution from anthropogenic warming to an emerging North American megadrought. *Science*, *368*(6488), 314–318. https://doi.org/10.1126/science.aaz9600

Wilson, R., Anchukaitis, K., Briffa, K. R., Buentgen, U., Cook, E., D'Arrigo, R., et al. (2016). Last millennium northern hemisphere summer temperatures from tree rings: Part I: The long term context. *Quaternary Science Reviews*, *134*, 1–18. https://doi.org/10.1016/j.quascirev.2015.12.005

Wilson, R., Cook, E., D'Arrigo, R., Riedwyl, N., Evans, M. N., Tudhope, A., & Allan, R. (2010). Reconstructing ENSO: The influence of method, proxy data, climate forcing and teleconnections. *Journal of Quaternary Science*, *25*(1), 62–78. https://doi.org/10.1002/jqs.1297

Woodhouse, C. A. (1999). Artificial neural networks and dendroclimatic reconstructions: An example from the Front Range, Colorado, USA. *The Holocene*, *9*(5), 521–529. https://doi.org/10.1191/095968399667128516

Yun, S., Smerdon, J. E., Li, B., & Zhang, X. (2021). A pseudoproxy assessment of why climate field reconstruction methods perform the way they do in time and space. *Climate of the Past*, *17*(6), 2583–2605. https://doi.org/10.5194/cp-17-2583-2021

Zhang, Z. H., Mann, M. E., & Cook, E. R. (2004). Alternative methods of proxy-based climate field reconstruction: Application to summer drought over the conterminous United States back to AD1700 from tree-ring data. *The Holocene*, *14*(4), 502–516. https://doi.org/10.1191/0959683604hl727rp