Advanced Review

Reconstructing Earth's surface temperature over the past 2000 years: the science behind the headlines

Jason E. Smerdon[1]* and Henry N. Pollack[2]

Edited by Eduardo Zorita, Domain Editor, and Mike Hulme, Editor-in-Chief

The last quarter century spans the publication of the first assessment report of the Intergovernmental Panel on Climate Change in 1990 and the latest report published in 2013–2014. The five assessment reports appearing over that interval reveal a marked increase in the number of paleoclimate studies addressing the climate of the last 2000 years (the Common Era). An important focus of this work has been on reconstruction of hemispheric and global temperatures. Several early studies in this area generated considerable scientific and public interest, and were followed by high-profile and sometimes vitriolic debates about the magnitude of temperature changes over all or part of the Common Era and their comparison to 20th- and 21st-century global temperature increases due to increasing levels of atmospheric greenhouse gases. Behind the more public debates, however, several consistent themes of scientific inquiry have developed to better characterize climate variability and change over the Common Era. These include attempts to collect more climate proxy archives and understand the signals they contain, improve the statistical methods used to estimate past temperature variability from proxies and their associated uncertainties, and to compare reconstructed temperature variability and change with climate model simulations. All of these efforts are driving a new age of research on the climate of the Common Era that is developing more cohesive and collaborative investigations into the dynamics of climate on time scales of decades to centuries, and an understanding of the implications for modeled climate projections of the future. © 2016 Wiley Periodicals, Inc.

How to cite this article:
WIREs Clim Change 2016. doi: 10.1002/wcc.418

INTRODUCTION

> The farther backward you can look, the farther forward you are likely to see.
>
> —Winston Churchill

This well-known comment by Winston Churchill asserts that there are lessons of history that help us better understand the present and more clearly anticipate the future. Herein we review the application of this tenet to paleoclimatology, the scientific field that seeks to estimate, or reconstruct, Earth's past climatic states and interpret the causes of past variability. We principally address climate reconstructions of the past two millennia, a period of time known as the Common Era (CE), that use proxies for climate information prior to the instrumental record of the past two centuries or so. A climate proxy,

---

*Correspondence to: jsmerdon@ldeo.columbia.edu

[1] Division of Ocean and Climate Physics, Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY, USA

[2] Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI, USA

Conflict of interest: The authors have declared no conflicts of interest for this article.

by definition, is a surrogate for an instrumental observation such as temperature, precipitation, or solar irradiance. Proxies are selected principally on the basis of their sensitivity to changes in target quantities, their geographic distribution, and their temporal range and resolution. Each proxy is nevertheless an imperfect representation of past climate, a reality that is central to ongoing efforts to improve proxy acquisition, interpretation, and application for a better understanding of past and, consequently, future climate variability and change.

Our focus in this review is specifically on hemispheric and global temperature reconstructions of parts or all of the CE as they have developed and evolved over the past quarter century. This interval corresponds to the period of time in which the five assessment reports of the Intergovernmental Panel on Climate Change (IPCC) have appeared: 1990, 1995, 2001, 2007, and 2013–2014.[1–5] Earlier scientific reviews related to this topic[6–11] were all published before the 2013–2014 report. While these previous reviews provide insightful surveys and discussions about the state of the field at different points over the course of the IPCC publications, they are not structured in a chronological arc. A partial exception is the perspectives piece on high-resolution paleoclimatology offered by Frank et al.,[10] which does provide a short chronological description of the 1990, 2001, and 2007 IPCC reports. In contrast, we specifically structure this review chronologically in order to place developments in the context of an evolving scientific understanding, one often overshadowed by the sometimes vitriolic exchanges outside the pages of scientific journals.

EMERGING INTEREST AND UNDERSTANDING

The past quarter century has seen a dramatic growth of interest in paleoclimate generally, as questions about abrupt climate change, past climate responses to elevated levels of atmospheric CO_2, and internal climate variability, all have become relevant for understanding anthropogenic influences on the climate system. With regard to the CE specifically, climate scientists have grappled with the question of whether observed climate changes in the 20th and 21st centuries have been a significant departure from conditions of the last several millennia, or whether these changes were simply the most recent of perhaps many similar events in the past. The separation of forced versus internal climate variability therefore has been an important motivation within the field, with implications for how each contribute to decadal-to-centennial variability and ultimately to future climate change.

The growth of interest in paleoclimatic information is reflected in the five IPCC assessment reports. Table 1 summarizes the number of pages in each of the reports devoted to paleoclimate, and the number of literature citations found in those pages. While this growth has been significant, interest in paleoclimate of course precedes the first IPCC report. It can be traced back at least to the 18th century when naturalists, geographers, and geologists began to piece together the history of the ice ages in Europe.[12] A more relevant example for the CE is the study of dendrochronology, a field with scientific roots going back centuries, but with more formal development beginning in the early 20th century. Andrew Ellicott Douglass (1867–1962), an astronomer, established the pioneering Laboratory of Tree Ring Research (LTRR) at the University of Arizona in the first part of the 20th century, with the motivation of using tree-ring records to study sunspot cycles.[13] Near the middle of the 20th century, Harold Fritts, a botanist, worked as an LTRR scientist to develop the specific field of dendroclimatology more formally[13]; he pioneered the application of statistical methods for calibrating tree-ring networks that still serve as the basis of many modern techniques in

TABLE 1 | Indicators of Growing Representation of Paleoclimate in the IPCC Assessment Reports

Assessment Report	Number of Paleoclimate Pages (Numbers in Parentheses Are Exclusive of Reference Pages)	Number of Paleoclimate Citations	Percentage of Paleoclimate Pages (Including Reference Pages) in the Complete WG1 Section of the AR
AR1 (1990)	~5	~40	~1.4
AR2 (1995)	~7	~100	~1.4
AR3 (2001)	~12	~250	~1.4
AR4 (2007)	65 (51)	609	6.5
AR5 (2013)	82 (62)	1014	5.3

The approximate numbers shown for Assessment Reports 1, 2, and 3 arise because in those reports paleoclimate was embedded within a broader discussion of observational changes in the climate system over all timescales. Paleoclimate was addressed a in separate chapter in AR4 and AR5.

dendroclimatology, and in many multiproxy reconstruction efforts as well.[14,15] Similarly, the modern use of multiproxy methods that combine different types of proxies as estimates of past hemispheric and global temperature variability was rooted in studies prior to the IPCC reports. The first quantitative reconstructions spanning a portion of the last millennium were performed for Northern Hemisphere (NH) temperatures from a multiproxy network and published in the late 1970s.[16,17] Tree-rings would later be used in the late 1980s to further advance the quantitative effort to produce estimates of NH temperature variability over the CE using climate proxies.[18] These examples by no means provide a comprehensive list of paleoclimate efforts prior to the first IPCC report, but stand as a few important highlights from a field that is rooted in well over a century of scientific thought and research.

One conclusion that came from research prior to the first IPCC report was that significant spatiotemporal paleoclimatic 'events' over the CE had existed, including the Medieval Climate Anomaly (MCA, also known as the Medieval Warm Period; ca. 800–1300 years CE), the Little Ice Age (LIA; ca. 1300–1850 years CE), and the Roman Warm Period (ca. 250 BCE–400 CE). Each event was expressed in space and time by an array of historical temperature, precipitation, and documentary records, mostly from Europe, and various individual climate proxy records, such as annually resolved tree-rings and glacier records (for popular accounts of the LIA and MCA, see Refs 19 and 20). Estimates of the timing and geographic expression of these 'events' were quite variable, and led to debates on whether they were regional, hemispheric or global in extent.

Both the MCA and LIA appear in Figure 1(a), a semi-quantitative temperature reconstruction published in the first IPCC Assessment Report (AR1[a]) and focused principally on temporal rather than spatial variability. The caption of the figure as it appeared in AR1 stated that it was a *schematic* diagram of *global* temperature variation over the last thousand years, with the dashed line nominally representing mean conditions near the beginning of the 20th century. The vertical axis shows the magnitude of the climate anomaly, residing well within a 1.5°C range. Notably, there was no source identified for this figure, and various authors have since pursued possible provenances. It is now widely agreed[9] that this figure had its origins in a 1965 paper,[27] based in part on the long-term Central England temperature record,[28] and therefore it is not a global representation, contrary to the figure caption for the schematic in AR1. Note that we present the figure as drafted in AR1 (some text has been enlarged relative to the original figure), which included a temperature scale on the vertical axis, despite contrary reporting in previous reviews.[9,10]

Modest advances beyond the schematic representation in Figure 1(a) appeared in AR2 as seen in Figure 1(b). This six-century NH temperature reconstruction of decadal averages was carefully qualified as representing only summer averages over regions of Europe, Asia, and North America. The reconstruction was nevertheless a multiproxy effort employing historical documentary sources, tree-ring widths and densities, instrumental records, and glacial melt records. AR2 also included the initial recognition of geo-thermometry (subsurface temperatures measured in terrestrial boreholes) as a new source of information about the ground surface temperature (GST) history of a region. Reconstructed GST histories, by the time of AR3, would play a significant role in the debates that would develop around proxy spectral fidelity.

Between AR2 and AR3, an important transition in reconstruction methodology occurred: the development of more quantitative techniques for using multiproxy networks for reconstruction, and for estimating their accompanying uncertainty. Such reconstructions became fully quantitative, and recognized the temporal and spatial variability, and uncertainties, of the individual proxies. In the closing years of the 20th century a few 'new era' reconstructions appeared,[22–25] based principally on seasonally or annually resolved proxies, and all were represented in AR3 (Figure 1(c)). Mann et al.[23] (hereafter MBH98) presented an annually resolved and spatially gridded reconstruction of global temperature variability over the past 600 years, and also presented correlations of the NH annual mean temperature time series with possible forcing factors, both natural and anthropogenic, that revealed an increasing dominance of greenhouse gas forcing in the 20th century. Mann et al.[24] (hereafter MBH99) was an outgrowth of MBH98 by the same authors in the following year, which extended and analyzed reconstructed NH mean temperatures and their uncertainties over the full past millennium. It was this reconstruction that famously became known as the Hockey Stick, so named because the graph displayed 900 years of slowly declining surface temperatures over the last millennium (seen as the shaft of the hockey stick), followed by a sharp rise in the 20th century (the blade).

The methodology and reconstructions that appeared in MBH98 and MBH99 played a particularly influential role in AR3, including a prominent

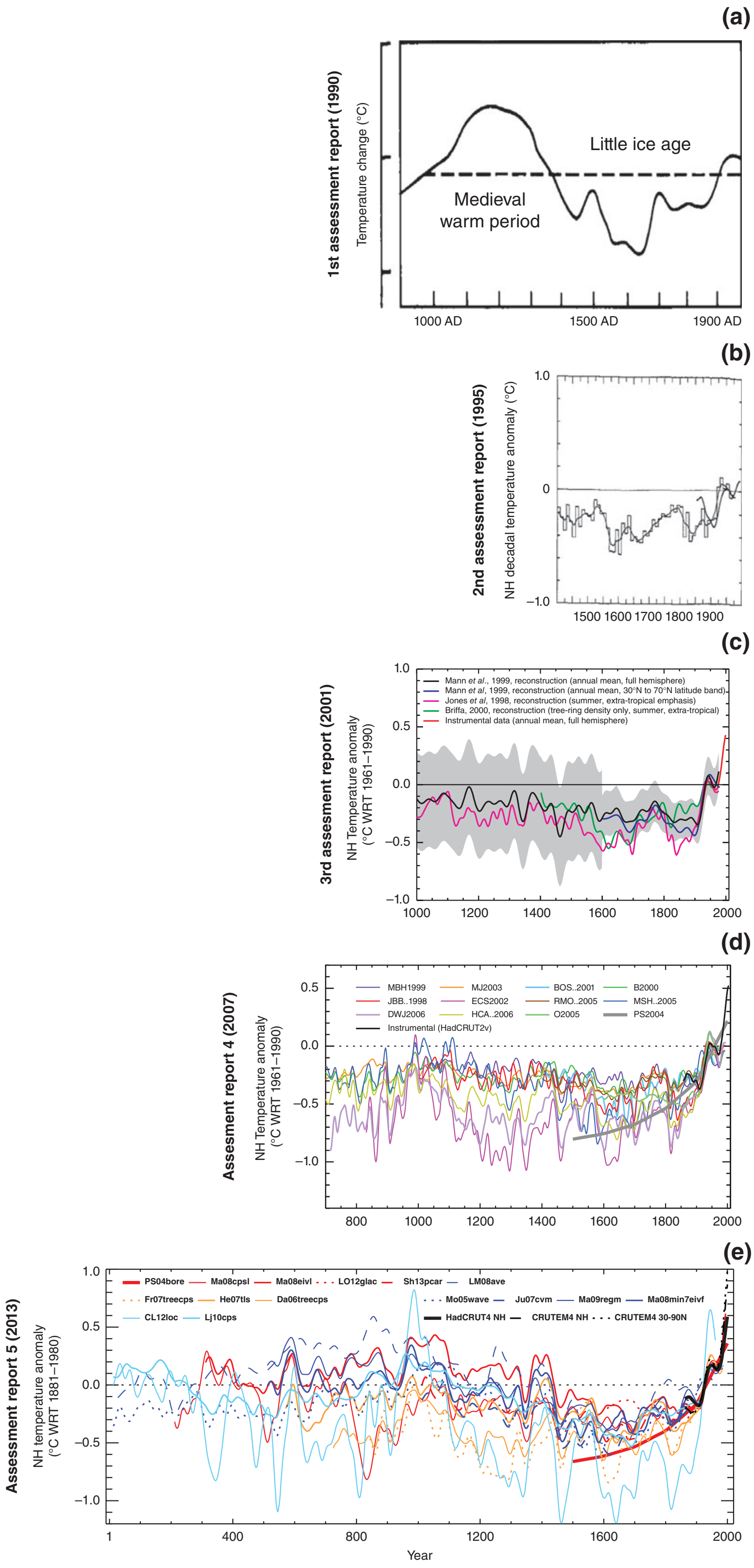

FIGURE 1 | Legend on next page.

appearance of the MBH99 mean NH temperatures in the AR3 Summary for Policy Makers. The high-profile appearance of the MBH99 reconstruction attracted not only scientific attention, but also widespread coverage in the popular press. Presumably because of this visibility, it excited strong reverberations in non-scientific communities as well. The reverberations arose because the sharp reversal in temperature trends in the 20th century left little room to avoid the conclusion that anthropogenic climate forcings had overtaken natural forcings, and had become the dominant factor driving warming, as concluded in MBH98. Nevertheless, the MBH99 publication itself did not quantify the causal mechanisms of the 20th-century trend relative to the preceding centuries; the only statement by MBH99 relevant to the forcing issue was that the 20th-century warming interrupted a long-term cooling trend associated with astronomical (i.e., natural) forcing.

For those entrenched in the position that all climate change was driven by natural forcings alone, the mere suggestion of an anthropogenic role in 20th-century warming was presumably sufficient to launch an attack on the data and methodology of the Hockey Stick reconstruction. An important focus of such critiques was on replication of the Hockey Stick analysis, and on access to data and computer codes necessary to facilitate replication efforts. These highly public confrontations eventually led to U.S. Congressional hearings, a U.S. National Academy of Sciences report,[7] unauthorized acquisition and public posting of private e-mail correspondence between scientists, and ultimately litigation. This more public side of the Hockey Stick has been well documented in many venues.[29–32] Often lost in the swirling dust of the public tussle, however, is an awareness of the scientific context in which the Hockey Stick evolved, and its role in the array of research on the CE that has appeared in the years since its publication. It is this 'science behind the headlines' that we wish to illuminate in the remainder of this paper.

IN SEARCH OF THE MCA AND LIA

One of the first scientific topics to capture attention around the time of the AR3 publication centered on

FIGURE 1 | Collection of diagrams plotting temperature reconstructions for the Common Era in each of the five IPCC assessment reports. Vertical scaling and horizontal alignment are exactly comparable for panels (c)–(e), while only approximate for panels (a) and (b). As noted in most of the panel descriptions below, the baseline or reference periods for the anomalies in each panel are different: in panel (a) the baseline is not specified, but is approximately the mean near the beginning of the 20th century; in panel (b) the baseline is not specified, but is approximately the mean of the four decades ranging from 1930 to 1960; in panels (c) and (d) the baseline is the mean between 1961 and 1990; in panel (e) the baseline is the mean between 1881 and 1980, approximately 0.2°C below the baseline for panels (c) and (d). For reference, the following panel captions are reproduced as presented in each of the IPCC assessment reports, although some minor editing has been included for clarity and referencing style. (a) Figure 7.1 from Climate Change 1990: The IPCC Scientific Assessment[1]: Schematic diagram of global temperature variations over the last thousand years. The dashed line nominally represents mean conditions near the beginning of the 20th century. (b) Figure 10 (TS) from Climate Change 1995—The Science of Climate Change[2]: Decadal summer temperature index for the Northern Hemisphere,[21] up to 1970–1979. The record is based on the average of 16 proxy summer temperature records from North America, Europe, and East Asia. The smooth line was created using an approximately 50-year Gaussian filter. Recent instrumental data for Northern Hemisphere summer temperature anomalies (over land and ocean) are also plotted (thick line). The instrumental record is probably biased high in the mid-19th century because of exposures differing from current techniques. (c) Figure 2.20 from Climate Change 2001: The Scientific Basis[3]: Comparison of warm-season[22] and annual mean[23,24] multiproxy-based and warm season tree-ring-based[25] millennial Northern Hemisphere temperature reconstructions. The recent instrumental annual mean Northern Hemisphere temperature record to 1999 is shown for comparison. Also shown is an extra-tropical sampling of the Ref 24 temperature pattern reconstructions more directly comparable in its latitudinal sampling of Ref 22. The self-consistently estimated two standard error limits (shaded region) for the smoothed Ref 24 series are shown. The horizontal zero line denotes the 1961–1990 reference period mean temperature. All series were smoothed with a 40-year Hamming-weights lowpass filter, with boundary constraints imposed by padding the series with its mean values during the first and last 25 years. (d) Figure 6.10 (b) from Climate Change 2007: The Physical Science Basis[4]: Records of NH temperature variation during the last 1.3 kyr. Reconstructions using multiple climate proxy records, including three records shown in AR3, and the instrumental temperature record in black. All series have been smoothed with a Gaussian-weighted filter to remove fluctuations on time scales less than 30 years; smoothed values are obtained up to both ends of each record by extending the records with the mean of the adjacent existing values. All temperatures represent anomalies (°C) from the 1961 to 1990 mean. (e) Figure 5.7 (a) from Climate Change 2013: The Physical Science Basis[5]: Reconstructed Northern Hemisphere annual temperatures during the last 2000 years. Individual reconstructions, grouped by color according to their spatial representation (red: land-only all latitudes; orange: land-only extratropical latitudes; light blue: land and sea extra-tropical latitudes; dark blue: land and sea all latitudes) and instrumental temperatures shown in black (Hadley Centre/ Climatic Research Unit (CRU) gridded surface temperature-4 data set (HadCRUT4) land and sea, and CRU Gridded Dataset of Global Historical Near-Surface Air TEMperature Anomalies Over Land version 4 (CRUTEM4) land-only[26]). All series represent anomalies (°C) from the 1881 to 1980 mean (horizontal dashed line) and have been smoothed with a filter that reduces variations on time scales less than about 50 years.

the spectral fidelity of various proxies, and on the degree to which the identification, extraction, and preservation of a proxy's full spectral content could be impacted by various data processing techniques. These discussions were motivated in part by questions surrounding the magnitude and spatial extent of the MCA and LIA, questions thoroughly examined in AR3. The high-profile attention focused on the MBH99 reconstruction after the AR3 publication initiated several important scientific debates that stemmed from the fact that the reconstruction did not display a well-defined expression of either the MCA or LIA. Defenders of the MBH99 reconstruction pointed out that these events had been described principally in Europe and parts of North America (in part because those regions had been the areas of most concentrated study) and that the MCA and LIA were likely only regional events with different seasonal expressions.[33] Such arguments noted that in sampling a broader spatial area and annual (as opposed to seasonal) responses in a multiproxy network, the MBH99 mean annual temperature reconstruction represented hemispheric climatic conditions and thus did not reconstruct enhanced MCA and LIA changes resulting from regional or seasonal biases.[8,33,34]

Criticisms of the absence of an MCA or LIA in the MBH99 reconstruction tended to focus on whether dendroclimatic records, which comprise the majority of proxies in the MBH99 network, capture decadal-to-centennial temperature variability with robust fidelity. In the context of how well the MBH99 reconstruction represented low-frequency variability, a 2001 commentary by Broecker[35] noted that 'tree-ring records are useful for measuring temperature fluctuations over short time periods but cannot pick up long-term trends because there is no way to establish the long-term evolution in ring thickness were temperatures to have remained constant.' The basis of this criticism is related to a processing step in dendroclimatology known as standardization[b], in which the long-term growth trends associated with individual trees are removed. The principal long-term trend in tree-ring thickness (from thicker to thinner as a tree ages) is associated with the growth of the tree circumference over the lifetime of the tree.[14] In order to focus on the climate-induced variability in a given tree-ring record, this long-term growth trend must be removed. The methods used to separate the trend, however, have the potential to remove information about multidecadal to centennial climate variations also present in the raw tree-ring width data, a potential problem that was well recognized by the end of the 20th century.[36] It was this standardization challenge to which Broecker[35] was referring; he went on to argue that the only proxies capable of sampling centennial-scale temperature variability on the order of 0.5°C were snowline elevations and terrestrial borehole temperature profiles.

Whether Broecker's[35] comments started the many subsequent lines of inquiry or were simply reflective of the scientific zeitgeist of the early 2000s is immaterial, but they did highlight a subject that would become the focus of significant debates and research. Dendroclimatic standardization methods that addressed low-frequency fidelity had been previously developed, and it was not long before they would be applied to derive a new NH temperature reconstruction. In particular, a method known as regional curve standardization (RCS)[37] was developed as a means of estimating growth curves from a regional collection of tree-ring records. The concept of RCS was based on the idea that all trees of a given species would contain similar physiologic growth curves. When stacked together by tree age, not calendar age, the growth curves would be aligned but the growth induced by climate variability would not. An average across an ensemble of such species-specific curves therefore would yield an estimated growth curve that could be used to standardize all trees of that species in the region. This RCS method was used to construct a tree-ring-only network across the NH that was in turn used to derive a NH temperature reconstruction[38] featuring increased amplitudes of decadal-to-centennial variability relative to MBH99. Regarding the implication of their results for arguments about low-frequency fidelity, Esper et al.[38] noted that their finding 'refutes the contention [of Broecker[35]] that long-term tree-ring records can not preserve multi-centennial fluctuations due to climate.'

TEMPERATURES BENEATH EARTH'S SOLID SURFACE: ISOLATING THE BASS NOTES

In addition to being important for interpretations of dendroclimatic records, Esper et al.[38] was significant because it estimated larger temperature contrasts between a cold LIA and warm MCA, using a tree-ring network that broadly sampled the extratropical NH. It was particularly notable that the increased magnitude of the negative temperature anomaly during the LIA was more consistent with the magnitude of temperature changes back to 1500 that were estimated by another proxy mentioned by Broecker,[35] namely reconstructions derived from terrestrial borehole temperature profiles.[39–43] This geothermal method of proxy reconstruction was developed in the interval between AR1 and AR2 (see Ref 44 for a review). New proxies, of course, are always appearing, but what made this entry perhaps more significant is that the Earth acts as a low-pass filter on downward-propagating surface temperatures. High-frequency temperature variations, particularly daily and annual temperature oscillations, are attenuated beyond detection within just a few meters below the ground surface. Reconstructions of GST histories from deeper subsurface temperature profiles therefore capture *only* decadal-to-centennial frequencies, the very spectral range that was the focus of the emerging discussions about the spectral fidelity of multiproxy reconstructions based largely on dendroclimatological data. Because of their relevance to the debates about proxy spectral fidelity that arose after the publication of AR3, here we provide additional background on borehole reconstructions.

Efforts to reconstruct GST histories from borehole temperature measurements were an outgrowth of a subsurface temperature database program begun in the early 1990s by the International Heat Flow Commission of the International Association of Seismology and Physics of the Earth's Interior. Earlier reconstructions from boreholes first appeared from measurements in Canada[45] and Cuba,[46] and later larger-scale work was presented in a paleoclimate session at the 1993 annual meeting of the American Association for the Advancement of Science in Boston, MA.

Initial borehole results were mentioned in AR2, and many more were reported in 1995 at the quadrennial assembly of the International Union of Geodesy and Geophysics where a special paleoclimate symposium addressed the climate of the past millennium as influenced by natural forcings and human activities. This symposium covered, *inter alia*, the means and techniques of climate and climate forcing reconstructions over this period, and their associated character in space and time. Shortly thereafter, the centennial timescale was further highlighted[47] in the inversion of subsurface temperatures by parameterizing borehole reconstructions to comprise only century-long trends (as they appeared in Refs 40 and 43, and AR3 and AR4, as shown in Figure 1(d) and (e)). By the end of the 20th century, aggregate results from hundreds of boreholes around the world had appeared[39–43] and the borehole reconstructions were presented in AR3, along with the prominent discussion of MBH99.

Prior to Esper et al.,[38] these borehole reconstructions stood as the principal conflicting estimate of hemispheric temperature change relative to the reconstructions from Refs 22–25 presented in AR3, based on the fact that they estimated a substantially colder LIA period extending back to at least 1500 CE. It therefore was significant that the new tree-ring based reconstruction[38] agreed more closely with the borehole estimate, which lent credence to the contention that the other AR3 reconstructions, containing tree-ring chronologies derived principally with non-RCS standardization techniques, had perhaps underestimated the magnitude of the LIA temperature excursion and low-frequency variability in general.

There were, however, questions about the borehole estimates as well. Perhaps most notably, the borehole method estimates changes in GST, not surface air temperatures (SAT), which is the target of most other terrestrial temperature proxies. Land-surface processes associated with surface moisture fluxes, snow cover, ground freezing and vegetation can cause differences between GST and SAT on diurnal and annual timescales. In the context of hemispheric and global temperature reconstructions, however, it was less clear whether these processes can cause differences between the *rates of change* in GST and SAT on decadal-to-centennial timescales, the periods that are relevant to the interpretation of GST histories derived from terrestrial borehole temperature profiles.

The above interpretations also relate to the seasonality of borehole records, namely whether they are representative of long-term changes in annual temperatures or specific seasonal windows. Borehole signals have traditionally been interpreted as representative of changes in annual means, because they are ultimately dependent on the year-round energy balance at the land-atmosphere boundary, as opposed to season-specific proxy responses such as biological growing seasons. Nevertheless, it is possible that seasonal changes in energy partitioning governed by cryogenic, hydrological or biological processes may influence the seasonal partitioning of sensible heat into the subsurface, which is ultimately the quantity that is measured in terrestrial boreholes. Various authors[48–50] have indeed appealed to long-term changes in land-surface processes to argue that the enhanced rates of temperature change from the LIA into the 20th century estimated from borehole reconstructions are due to seasonal biases inherent in the interpretation of GST as an estimate of annual SAT changes. These arguments were contested by borehole research groups,[43,51,52] and subsequent research on GST–SAT comparisons lent support to the interpretation that long-term rates of change in GST reliably estimate long-term rates of change in annual SAT.[43,53–61] A more straightforward criticism was that some NH mean temperature reconstructions derived from borehole profiles were not weighted by latitude to reduce spatial sampling biases.[49,62] Differences between weighted and unweighted mean results were nevertheless shown to be minimal, and the original estimate of the area-weighting effects in Mann et al.[49] was shown to be erroneous.[43,63]

Debates about the spectral fidelity and seasonality of borehole reconstructions and additional proxy systems have continued to the present day. We will return to some of these issues in the concluding sections of this review. Toward the middle of the 2000s, however, the debate shifted away from the underlying spectral fidelity of proxy records to focus more on reconstruction methodologies and how they alone may influence the spectral character of reconstructions. To understand how these debates unfolded it is necessary to review the emergence of another important area of research on the climate of the CE namely attempts to model this period with forced-transient simulations from fully coupled climate models.

SIMULATING THE COMMON ERA WITH CLIMATE MODELS

The first published forced-transient simulation of the last millennium using a fully coupled climate model was performed with the GKSS ECHO-g model.[60] Among the many relevant scientific questions that might have been addressed using this unprecedented new climate model simulation, it was a sign of the times that the first analysis of the simulation, published in 2003, was couched in the context of borehole climate reconstructions. González-Rouco et al.[60] compared the ECHO-g simulated annual SAT and subsurface (~10-m depth) temperatures, and found that while there were seasonal and annual differences between the two temperatures, they displayed strong decadal and lower-frequency coherence. This was in contrast to another modeling paper published in the same year that compared simulated GST and SAT from the GISS ModelE over the 1951–1998 CE period and found divergent trends in the two temperatures.[48] The results of this latter study were debated,[51,64] but regardless, it was not capable of characterizing decadal-to-centennial variability as robustly as the much longer ECHO-g last-millennium simulation that did demonstrate strong coupling between GST and SAT over decades and centuries[60] (note that this original study has been supported by subsequent work with multiple model simulations and analyses that take the added step of inverting simulated subsurface temperature profiles to show that derived synthetic reconstructions are consistent with simulated SAT timeseries[61,65]).

In addition to the original GKSS ECHO-g simulation, known as ERIK1, an ERIK2 simulation was published in 2006 using the identical ECHO-g model but with different model initialization in year 1000 CE.[61] This second run was produced in part because the original ERIK1 simulation was criticized for its initialization, which was representative of the mid-20th century and caused a drift in the first several centuries of the simulation.[66] The consequence was a warmer MCA than what emerged in the ERIK2 simulation, which began with colder preindustrial initial conditions.[61,66,67] Although the appropriate initialization for last-millennium simulations remains uncertain, many of the subsequent model simulations have adopted colder preindustrial initial conditions. The next widely used last-millennium simulation was published in 2007 using the NCAR CCSM1.4 model,[67,68] and several other modeling centers prior to the last-millennium experiments of AR5[69] also produced individual or last-millennium ensemble runs.[70,71]

Shortly after the initial focus on the borehole reconstructions,[60] the ERIK1 simulation,[72] and a millennium-length ECHO-g control simulation[73] were used to conduct pseudoproxy experiments (PPEs), an analysis approach that would ultimately have broad-ranging impact on the interpretation of the MBH99 reconstruction and of reconstruction methodologies more generally. PPEs are synthetic reconstruction experiments that use last-millennium simulations as test beds to evaluate the performance of a reconstruction method using controlled and systematic experiments (see Smerdon[11] for a review). von Storch et al.[72] was the first to use a forced-transient last-millennium simulation in a PPE context and showed that the MBH99 regression *methodology* was capable of reducing low-frequency variability such that the magnitudes of reconstructed MCA and LIA temperatures were muted relative to the known temperature history produced in the modeled climate. Subsequent discussions clarified the correct application of the MBH99 methodology[74–76] and demonstrated that the results for mean NH temperature reconstructions were not dependent on the employed last-millennium simulation.[77–83] These discussions nevertheless were the beginning of widespread debates about the methods used to reconstruct hemispheric and global temperatures from multiproxy networks.

THE METHOD EXPLOSION

The challenge at the heart of multiproxy reconstruction methods stems from the classical general problem of interpreting incomplete, inaccurate, and conflicting information. The incompleteness derives from the geographical and temporal sparseness of the proxy networks employed. The inaccuracies stem from the reality that proxies are seldom sensitive to only a single climatological quantity, are often subject to site-specific causes of noise, and can be measured only with a specific degree of precision. Examples of multivariate influences include the fact that isotopic ratios in corals are dependent on both ocean temperatures and salinities, or that the annual growth of trees is sensitive to multiple factors, including temperature, precipitation, and received solar radiation. Moreover, the range of climatic conditions over which a proxy may be sensitive potentially plays a role in their interpretation. Tree growth, for example, has been investigated in the context of the purported divergence issue in which some tree-ring chronologies diverge from late 20th-century temperature trends (see Ref 84 for a review) and some explanations hinge on the possibility that growth is subdued by heat stress at temperature levels that may differ by species.[85] Uncertainties in derived reconstructions of past climate can thus lead to conflicting reconstructions, due to differences in the methods used to extract a climate signal that is embedded in the noise of inaccurate and heterogeneous data.

Controlling for the above factors in methodological experiments with real proxy networks and observational data is not always straightforward. The advent of PPEs using last-millennium simulations were thus an important development for testing methods in systematic experiments, and began an explosive new research effort to understand and improve reconstruction methods and to characterize their potential biases and uncertainties. These developments would in many ways shift the discussion away from proxy spectral fidelity (although research in that area would continue) toward reconstruction techniques and their assessment, but much of the discussion was still framed in the context of how methods themselves impact the variance of reconstruction products. Consequently, many new methods have been developed over the last decade and have yielded vast improvements in our understanding of the challenges and uncertainties associated with reconstruction methods.

While PPEs would drive much of the new methodological research beginning in mid-2000s, a more traditional bootstrapping analysis using red noise series was used to explore the implication of specific steps in the MBH99 methodology,[86] with results that were also important in the early and evolving discussion of reconstruction techniques. The primary investigation focused on an unconventional centering decision applied in a principal component analysis (PCA) in MBH99. PCA was used to condense the North American tree-ring network, a component of the MBH99 multiproxy network, into leading PCs to mitigate the potentially excessive influence from the high-density of tree-ring records within the region. The original MBH99 approach centered all of the North American tree-ring records over the calibration interval (1902–1980) and then applied PCA over the longer multicentury duration of the network. This approach was used to ensure that all proxy records were centered in relation to a common period, given that MBH99 used multiple nests comprising different temporal lengths that reflected the reduced availability of proxies back in time. This decision nevertheless has the potential to produce a leading PC that displays the hockey-stick shape, which in turn can be preferentially weighted in the full MBH99 methodology.[86] A Monte Carlo analysis to benchmark the skill of the MBH99 reconstruction was additionally used to further call into question its robustness in the early centuries of the last millennium.[86]

Multiple subsequent studies contested the conclusions of McIntyre and McKitrick.[86] Direct comments on the paper included one using PPEs based on the ERIK1 simulation that indicated a minimal impact of the MBH99 centering decision on the resulting NH reconstruction.[87] A second comment[88] argued that using the correlation matrix instead of the covariance matrix to compute the tree-ring PCs diminished the biases reported by McIntyre and McKitrick[86] and that variance matching of the derived reconstruction to the calibration target improved the skill assessment of the reconstruction relative to that characterized by McIntyre and McKitrick.[86] Responses to these comments[89,90] contested the arguments on the grounds that the former assumed unrealistic noise characteristics in the constructed pseudoproxies and that the theoretical justification for the choices in the latter were unfounded. Further arguments would arise that indicated the impact of the centering convention used for the PC reduction of the North American tree-ring records was insignificant given the selection criteria applied to the reduction of the full predictor network for the final reconstruction.[91,92] In particular, Ammann and Wahl[92] showed that the impact of avoiding the MBH99 centering decision in favor of full-period centering was to spread the Hockey-Stick shape over the first two tree-ring PCs, rather than concentrating it in the first PC as occurred using the MBH99 centering convention. Because the actual MBH99 reconstruction used both the first and second PCs, the resulting reconstructions with either centering convention are nearly identical when the first two PCs are retained (cf. Supporting Information in Ref 92). Although the discussion would soon shift away from the original MBH99 methodology, these initial discussions were important for shedding light on the various methodological choices that were part of the reconstruction process. Indeed, the spirit of the moment was captured in a 2005 study[93] that demonstrated the wide range of NH mean results that could be achieved in an MBH99-type reconstruction when applying a collection of different methodological decisions with varying levels of theoretical and practical justifications. These discussions also further supported a widespread understanding of the importance of providing code and data necessary for replication of reconstruction results.[7,94,95]

It is important to note that these discussions in the early and mid part of the 2000s also motivated an evolving awareness and investigation of the methods by which reconstructions are statistically evaluated as estimates of past climate variability. PPEs were an important component of that process, but more specifically the skill metrics used for the verification of derived reconstructions became a subject of focus as a consequence of the discussions around McIntyre and McKitrick.[86] Traditional verification metrics have their roots in the dendroclimatological literature,[96] which themselves came out of forecasting and hydrological applications. The most commonly employed metrics include the coefficient of efficiency, reduction of error (RE), and coefficient of determination (r^2), the first two of which can range between negative infinity and one and the latter ranges between 0 and 1. These skill metrics are typically computed using withheld target data and reconstructions in coeval years. RE and the coefficient of efficiency are considered skillful if they are positive (yielding a prediction that is better than climatology), while the r^2 significance threshold is determined by the number of degrees of freedom in the verification interval and the nature of the applied null hypothesis. While these basic details of the verification process are straightforward, appreciation of multiple verification nuances have evolved within the large-scale temperature reconstruction literature. The first is an awareness of the fact that while the zero-threshold is a minimum target for RE and the coefficient of efficiency, benchmarking experiments reveal that positive values of the skill metrics can be achieved in some cases using noise-only predictors comprising different noise models.[11,86,97–99] In many cases, static calibration and verification intervals have also been eschewed for multiple verification intervals or hold-out blocks to avoid skill assessments that may be biased by the specifics of a given verification window.[97–101] Work to develop and apply alternative skill metrics has also been pursued,[102] some of which has focused on verification of spatial patterns in spatiotemporal reconstructions targeting more than single mean indices.[103,104] All of these advancements have increased the sophistication with which reconstructions are being validated, while addressing some of the early debates about the skill of MBH99. Perhaps the most critical conclusion of these findings is the importance of reporting multiple skill metrics for reconstruction verification,[98–100,105–108] while also including PPE assessments when possible.[99,106,109–111]

RECONSTRUCTING MAPS AND INDICES

In light of the earlier discussions about the magnitude of the MCA and LIA temperature excursions, and the associated low-frequency fidelity of proxy records, perhaps the most troubling aspect of the emerging understanding of the MBH99 reconstruction methodology was the potential that the method itself yielded variance losses. While the magnitude of such losses as demonstrated by von Storch et al.[72] was ultimately shown to be substantially reduced[76] for the MBH99 method, concerns about the effect began with the publication of von Storch et al.[72] and were in part connected to the discussion around McIntyre and McKitrick.[86] Variance losses therefore would become the challenge addressed by many subsequent methodological developments. For example, a 2005 reconstruction focusing on proxy spectral fidelity integrated annually resolved proxy records like tree rings and lower-resolution records such as ocean sediment cores using a wavelet spectral analysis that allowed the former to define annual and interannual variability and the latter to reconstruct decadal and longer periods of variability.[112] This method thus blended the various strengths of a multiproxy network to derive a normalized index, which was then scaled and shifted to match the variance and mean of a NH mean temperature index during a calibration interval. The study was an important methodological attempt to combine proxies of different resolutions into an estimate of NH mean temperatures and yielded a reconstruction that matched well the borehole and Esper et al.[38] estimates.

Despite the complexity of the manner in which Moberg et al.[112] blended multiproxy data, the final time series was simply scaled to have the same mean and variance as the targeted NH temperature index, a technique that is a common approach for deriving calibrated proxy-derived indices.[113] It is important to note, however, that these attempts to target the NH mean time series are in contrast to the MBH98/99 reconstructions that targeted spatiotemporal fields, or maps, of temperature change over time. These two types of reconstructions have been classified as index reconstructions and climate field reconstructions (CFRs), respectively. The vast majority of reconstructions that have targeted hemispheric and global temperatures have been index reconstructions. For example, of the reconstructions highlighted in the AR4 summary in Figure 1(d) only two are aggregated from CFRs and in the AR5 summary in Figure 1 (e) only one CFR result appears (as an update to those included in AR4).

Methodological developments specifically for index reconstructions following Esper et al.[38] and Moberg et al.[112] continued through the late 2000s. Hegerl et al.[110,114] used a total least squares (TLS) regression methodology to derive an index reconstruction of NH temperature using proxies with exclusively decadal resolution. The TLS method formulates the linear regression problem to include estimated errors in both the predictor estimate and the predictand temperatures. It was argued that this reduces the loss of low-frequency variability that had been demonstrated for other methods. The ultimate reconstruction was validated using out-of-sample verification statistics computed from withheld target data and a PPE using the ERIK1 last-millennium simulation.[110] The resulting reconstruction again included more low-frequency variability and was specifically demonstrated to be consistent with the Pollack and Smerdon[43] borehole reconstruction. Another reconstruction derived only from tree-rings was published during the same time that again derived a mean NH temperature estimate with increased low-frequency variability and larger contrasts between a warm MCA and cold LIA.[115] These reconstruction results were analyzed in terms of tree chronologies that were assembled from more traditional standardization techniques and those employing RCS, and provided further support that the latter yielded reconstructions with enhanced low-frequency variability. Overall, the derived reconstruction agreed favorably with the earlier tree-ring-only RCS reconstruction in Esper et al.,[38] while the multiproxy reconstruction from Moberg et al.[112] still remained the largest estimate of LIA cold among the growing collection of studies. Notably, D'Arrigo et al.[115] also argued for the value of deriving proxy-specific reconstructions, given the uncertainties infused in multiproxy studies that incorporate records with variable multivariate climate influences and seasonal sensitivities. The spirit of this suggestion was indeed reflected in a reconstruction that was published in 2005 based only on glacial-length measurements,[116] a proxy newly applied in the NH temperature reconstruction context. Interestingly, this single-proxy reconstruction more closely matched the MBH99 reconstruction and thus suggested a warmer LIA than the other reconstructions emerging during the same time.

The developments around the above index reconstructions were highlighted in the 2006 U.S. National Research Council (NRC) review[7] (carried out following a request from the U.S. House of Representatives). This report, commissioned to evaluate the state of the science (data and methods) being used to reconstruct hemispheric and global mean temperatures of the last millennium, was in part motivated by the political reactions to multiple studies.[72,86] Figure S1 of the NRC report showed the evolving picture of the NH temperature history over the last millennium (similar to the AR4 report, shown in Figure 1(d), that was published one year later), and displayed a greater range of temperature variation estimated from newer reconstruction techniques and proxies. The emergent LIA and MCA contrasted with the long monotonous temperature decline that preceded the 20th-century warming in the MBH99 reconstruction featured in AR3, although the NRC report also highlighted the increasing uncertainties associated with the reconstructions prior to about 1600.

Shortly after the publication of the NRC report and AR4, a new collection of index reconstructions[117] was published using techniques similar to the TLS method adopted by Hegerl et al.[110] and a multiproxy dataset expanded significantly from MBH99. This new result showed increased amplitudes of low-frequency variability that was more in line with the then growing body of estimates (similar index reconstructions would be added by AR5[118–124]). While many of these studies addressed specific methodological and proxy uncertainties, the ensemble result in AR5 (Figure 1(e)) is clearly one in which low-frequency variability is more strongly expressed, relative to the earlier assessments. While the 20th-century and early 21st-century temperature rise is clearly apparent, it is also clear that more than a decade of efforts to improve methods and interpretations of proxies yielded reconstructions with a larger contrast between a warm MCA and cold LIA than the earlier estimates indicated. While these developments might indicate a more broadly expressed MCA and LIA across the NH or globe, studies of glacial advances and retreats (the other evidence originally highlighted by Broecker[35]) provide an interesting counter example in this discussion. These efforts have suggested more regional warming and cooling signals associated with the MCA and LIA and/or temporal phasing of the events that were not coincident in the Northern and Southern Hemispheres,[125,126] providing support for arguments that hemispheric or global reconstructions of temperature might not have pronounced MCA and LIA events.

Although work on CFR methodologies progressed in parallel with the above index efforts, it was often the case during early methodological discussions that the NH or global mean representations of CFR results were stressed over the spatiotemporal information that they provided. This was perhaps a legacy of the MBH98/99 discussions, which were themselves CFRs, but the dominant focus and discussion of these results were centered on how they represented the NH mean temperature index.[91,92] The emphasis on the NH mean result is evidenced by the discussions of the regularized expectation maximization (RegEM) method, which was suggested[127] as an early alternative to the MBH99 CFR methodology[c].

The original evaluations of the RegEM method in the context of paleoclimate CFRs were performed with PPEs[67] and focused largely on the ability of the method to derive skillful NH mean temperatures. The reported success of the method in reproducing NH mean temperatures in PPEs was used as motivation to apply RegEM to the MBH99 multiproxy network,[135] the NH mean temperature of which was shown to compare favorably with that from MBH99. The PPE results from Mann et al.[67] were later called into question, however, due to the fact that they used information outside of the calibration interval that would never be available in real-world CFRs.[136] If RegEM was instead restricted to only information in the calibration interval, PPEs demonstrated that the method produced NH mean temperature indices that suffered from significant low-frequency variance losses,[136,137] similar to what had been shown for the MBH99 method.[72] The fact that both the MBH99 and RegEM methods were subject to variance losses in PPEs and that they produced similar reconstructions when applied to the MBH99 multiproxy network, lent further credence to the concerns that the MBH99 reconstruction may have underestimated the extent of low-frequency temperature variability.

These initial discussions of the RegEM method motivated a change in the way that it was implemented, after which the method was shown to perform skillfully in reproducing the NH mean temperature index.[77] This alternative approach was subsequently used to produce a new global temperature CFR using an expanded multiproxy dataset.[138] To appreciate the distinction between the original and later implementations of RegEM it is necessary to discuss the basic theoretical underpinnings of the CFR methodology.

REGULARIZING FOR SUCCESS

We provide a general discussion of the multivariate linear regression framework as it pertains to CFRs in the Box 1, which also provides references for further review. It nevertheless is sufficient to summarize many of the methodological debates about CFRs as hinging on three choices that are aimed at limiting and/or constraining the information that is to be reconstructed, enhancing the signal against the background noise in the proxy network (the inaccurate and conflicting information referred to earlier), and better estimating the covariance between the proxy and temperature data to derive the regression coefficients of the reconstruction (the **B** matrix in Box 1). These choices often boil down to: (1) whether the target temperature field should be reduced to leading patterns of covariance and how many patterns should be retained; (2) whether the proxy matrix should similarly be reduced to leading PCs indicative of shared covariance in the network; and (3) how the regression between the two matrices should be regularized (see Box 1 for more discussion on regularization). The differences between the two versions of the RegEM algorithm largely can be understood in these terms. The early applications of RegEM used ridge regression (RegEM-Ridge), a form of regularized regression, and the method was used to target the full temperature field. The later efforts, however, reduced the target temperature field to a few leading empirical orthogonal function (EOF) patterns and used a truncated TLS (RegEM-TTLS) regression method, which is both a different form of regularized regression and assumes errors in the target temperature field (similar to the TLS index method used by Hegerl et al.[95]). Although the errors in the temperature field are not well identified and must be estimated, the technique has been demonstrated to reduce the variance losses associated with other methods such as RegEM-Ridge that only assume errors in the temperatures predicted from the proxies (note that variance losses in multiple regression contexts are not universal and that inflation errors can also occur in settings that include

BOX 1

BASICS OF CFR METHODOLOGY

In the simplest of terms, CFR methods relate a matrix of climate proxies to a matrix of climate data during a common time interval (generally termed the calibration interval) using a linear model. If **P** is an $m \times n$ matrix of proxy values and **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 **B** is a matrix of regression coefficients with dimensions $r \times m$, ε is the residual error and the primes indicate that the **T** and **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 **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 the matrix transpose. Temperature thus can be estimated, or 'reconstructed,' using the regression matrix **B** during periods in which proxy data are available but observed temperatures are not.

While the above formalism is straightforward, it works best when the system is overdetermined, that is, the time dimension n is much larger than the spatial dimension m. The challenge for CFR methods involves the manner in which **B** is estimated in practical situations when this condition is not met. It is often the case in CFRs that the number of target variables exceeds the time dimension, yielding an ill-posed estimation problem. For instance, in most global or NH CFRs, the number of grid cells in the climate field is typically on the order of many hundreds or a few thousands, while the observational record usually contains 150 annual fields or less. The number of proxies is typically on the order of a few tens to hundreds, which may exceed or at least be comparable to the time dimension. In such cases, the cross-covariance and covariance matrices in the estimate of **B** cannot be well estimated. The estimate therefore requires some form of regularization to apply additional constraints on the problem, which is a well-established statistical method for solving ill-posed estimation problems and preventing over-fitting. Forms of regularization are widely discussed in the literature and represent a continuing area of statistical and computational research.[141–146] A detailed technical review of these regression challenges specifically in the context of the paleoclimate reconstruction problem is provided by Ref 147 and other cogent discussions have been provided by multiple authors.[105–107,140]

multiple predictor variables[139]). The reduction of the targeted temperature information in the application of RegEM-TTLS is also important, as is the finite truncation of the eigenvalues in TTLS. This latter point arises in contrast to the smooth filtering of ridge regression that was shown to diminish variance in the leading eigenvalues of the regression in the CFR context because of the few eigenvalues that are ultimately retained.[140]

Despite the demonstrated success of RegEM-TTLS in PPEs,[77] there remained several outstanding issues that were the subject of subsequent discussion in the late 2000s. First, both RegEM methods applied a split calibration approach in which the reconstructions were performed separately in high and low frequency domains before combination. Although the spectral separation was done using more traditional spectral filters, the approach was similar in spirit to that of Moberg et al.[112] The efficacy of this filtering approach was challenged,[148,149] motivated in part by an earlier study[150] that showed variance losses for RegEM-TTLS in a collection of alternative PPEs using an ensemble framework. Several exchanges[137,151–153] also illuminated multiple technical problems with the PPE framework that was used by Mann et al.[77] It is also worth noting that RegEM-TTLS was used to derive a real-world NH CFR using the MBH99 multiproxy network.[77] In an as yet unresolved contradiction of results, the RegEM-Ridge and TTLS reconstructions yield very similar real-world reconstructions from the MBH99 network, despite the demonstrated tendency of the former method to lose significant variance in PPEs while the latter does not. All of these discussions ultimately indicate that the ability of RegEM and other methods to produce skillful CFRs became an emergent subject of debate in the late 2000s, which was made further relevant by the fact that RegEM-TTLS was used to derive an updated global temperature CFR in 2009.[138]

MAPPING ERRORS AND SKILL

The true value of CFRs is in their ability to provide spatiotemporal information that can be used to interpret large-scale dynamics and more completely compare reconstructions to last-millennium climate model simulations. Early appreciation of the value of spatial information was demonstrated in various studies,[23,24,33,154] as was the potential for comparing model simulations to either reconstructed indices[155,156] or patterns from CFRs.[138,157] Particularly, these latter studies are evidence of the importance of evaluating CFR skill not only in terms of hemispheric and global mean temperature indices, but also in terms of their reconstructed spatial patterns. The importance of evaluating spatial skill has been long recognized in tree-ring based CFRs of hydroclimate[130,158,159] and limited spatial skill assessments were provided for early temperature CFRs.[23,24] Some later studies also reported summaries of field statistics or provided spatial plots of some skill metrics in either PPEs or real-world CFRs.[67,77,135] Overwhelmingly, however, the early evaluations of CFR methods focused on their ability to derive skillful NH or global mean indices, but such evaluations are insufficient for assessing the spatial performance of CFRs (see, e.g., the review in Smerdon[11]).

The first few studies that directly assessed spatial skill of CFR methods reported significant variations in spatial performance. These demonstrations were done in regional studies,[160–162] but were investigated at hemispheric and global scales as well.[137,138,140] The first global-scale PPE analysis was published in 2011 and evaluated the spatial skill of four CFR methods using two model-based PPEs and showed significant spatial variations in the skill performance of each method.[81] A growing number of PPE studies have since appeared and evaluate the spatial performance of hemispheric or global CFRs in PPE contexts.[103–107,163–167] Some of these studies have noted that the construction of more realistic pseudoproxies can further degrade the spatial skill of CFRs, and that the character of the model-simulated fields can influence PPE conclusions about CFR spatial skill.[105,166]

Among the various conclusions of PPEs that have focused on spatial skill is the realization that important spatial errors can exist in CFRs derived from a range of state-of-the-art methods and these errors are expressed relatively consistently across all techniques. The magnitude of error has been shown to be dependent on the character and level of noise in pseudoproxy networks, the pseudoproxy distributions and availability back in time, whether the pseudoproxies sample univariate or multivariate climate characteristics, and method-specific parameter choices such as the degree of regularization in multivariate regression formulations. An emerging example of the implications of these results surrounds the interpretation of the La Niña-like pattern in the difference between the MCA and LIA periods in the Mann et al.[138] CFR. This pattern is compelling because it supports a long-standing hypothesis that increased radiative forcings during the MCA may have caused a centuries-long mean shift in the tropical Pacific due to a hypothesized ocean thermostat mechanism.[168] Among other implications of such a prolonged shift in the mean state of the El Niño-Southern Oscillation phenomenon is the possibility that it could explain the infamous megadroughts that occurred throughout the North American West during the MCA (see Refs 169 and 170 for reviews). Despite these interesting possible explanations and the demonstration of the pattern in one CFR,[138] subsequent PPEs have shown the capacity for some CFR methods to produce La Niña-like patterns on centennial timescales that are not characteristic of the original model fields.[167] More directly, Wang et al.[106] has shown that the MCA–LIA pattern is not robust when using two multiproxy networks and multiple methods and proxy selection choices. While these findings will be further refined, they speak to the importance of more fully vetting the field skill of contemporary methods using multiple models and PPE designs, while more directly connecting PPE results to the specific characteristics of real proxies and climate fields.

BACK TO THE FUTURE

Over the course of developments in temperature reconstructions since AR3, as reflected in the progression of estimates presented in Figure 1, the hockey metaphor itself has evolved. The clearer emergence of the MCA and LIA on the shaft of the hockey stick, but still leading to a strong 20th-century warming, might warrant a comment that the hockey stick is bent, but not broken. Another commonly used phrasing has suggested that the growing number of reconstructions that have appeared since MBH99 have progressed from a hockey stick to a full hockey team with the implication being that the collective ensemble of reconstructions supports the conclusions derived from the original Hockey Stick reconstruction. It would be a mistake, however, to assume that similarities among reconstructions in Figure 1(d) and (e) are indicative of a cohesive group that broadly agrees about the means and ways of climate reconstruction and interpretation. To the contrary, it has been the *differences* between the reconstructions represented in Figure 1(d) and (e) that have motivated much of the research over the past 15 years or more on CE temperature variability. These differences have driven an evolving understanding about the importance of developing proxies in undersampled regions, the spectral fidelity of proxy records, the statistical methods and uncertainties of climate reconstruction techniques, the similarities and differences between reconstructed and modeled estimates of CE climate, and the dynamics of forced and internal climate variability.

In the spirit of reconciling the differences between the various reconstruction estimates, perhaps the most important step has been, and will continue to be, the immense effort to expand proxy networks, understand the climate signals and noise that they contain, and to characterize and improve the methods used to homogenize proxies into index or field reconstructions. The PAGES2k Consortium[95,113] effort has been an important initiative to organize the proxy compilation task, which was founded on the premise that groups of domain and region-specific proxy experts are best suited to determine the appropriate proxy records for inclusion in multiproxy networks used for hemispheric and global reconstructions. This effort has been transparent in its efforts and selection criteria and has made public access to employed proxy records a priority.[95] Record collection has also continued apace as proxy experts work to expand observations in poorly sampled areas, such as the tropics, the oceans and the Southern Hemisphere.[113,171–173]

With the increase in the availability and diversity of proxy records, the prospect of screening proxy networks has also become a practice within the field. This process can include any number of desirable *a priori* proxy screening criteria such as their temporal extent, resolution, public availability, temperature sensitivity, site selection, or their method of processing (e.g., the standardization method of a tree-ring chronology).[95,100,113,117,174,175] Statistical screening of proxies based on their association with temperature has also been explored,[105,106,117] but the implications of statistical selection bias must also be further addressed in these contexts.[176] In general, the ability to consider different proxy screening strategies is a positive development stemming from the increasing number of available proxies. Among the many benefits of these capabilities, networks can and should be constructed with updated proxy records processed with state-of-the-art methods and understanding[99–101] while allowing independent reconstruction verifications from withheld proxies that may have different spectral resolutions.[100]

In addition to proxy record acquisition and network construction, many research communities also have worked to understand and model proxy systems from a process-based perspective.[177–184] These efforts are not only promising as a means of understanding the more fundamental multivariate and non-linear responses of proxy systems,[166,179] but also may allow more realistic process-based models describing climate-proxy connections to be incorporated into reconstruction methodologies.[179] Efforts to understand the spectral fidelity of proxies have also continued. For lower frequencies (decadal and longer) this process will always be limited by the difficulty of ground truthing low-frequency variability against an instrumental record that is only about 150 years long and of course growing only very slowly (year by year!) in real time. A recent contradiction in the literature highlights this fundamental challenge. An analysis of multiple proxy records in North America and model data concludes that climate models underestimate low-frequency precipitation variability and must be adjusted to better reflect the spectral character of climate estimates from proxies that have redder spectral densities.[185] A contemporary study[186] similarly compares model, tree-ring and observational data (for both temperature and precipitation globally) and also notes that tree-ring records are redder than the observational and model data. The authors nevertheless conclude exactly the opposite of the former study, by indicating that the differences suggest that tree-ring records are biased red.

In contrast to the challenges of assessments on decadal and centennial timescales, a more positive recent example highlights successes evaluating spectral fidelity on interannual timescales. In attempts to explain differences between large-scale temperature responses to volcanic eruptions estimated from tree rings and those simulated in climate models, Mann et al.[187] proposed that these differences were the consequence of missing rings in the tree-ring record due to extreme cold during the growing season subsequent to an eruption. This hypotheses has been rejected,[188–191] but one awareness developed around its evaluation has been how volcanic responses are represented in maximum latewood density and ring width records.[99,101,190,191] The latter quantity is subject to strong biological persistence that is thought to impose higher temporal autocorrelation in derived chronologies, while density records display lower autocorrelation (see Wilson et al.[101] for a cogent review of these issues). The consequence for estimates of climatic cooling subsequent to volcanic eruptions are that density records indicate enhanced cooling spikes in the year directly following the eruption and a quicker return to mean climate, relative to width records that tend to smooth the signal and yield smaller deviations that persist longer.[99,101,191,192] This awareness is spurring new attempts to better characterize climatic responses to volcanic eruptions, but it remains uncertain how the relative contributions of ring width and density may influence the decadal and centennial variability of large-scale reconstructions. Regardless, the spectral fidelity of proxies remains a critical topic of research that is vital to our ability to characterize climate variability over a wide range of timescales.

All of the above efforts can be classified as a more general attempt to better characterize the signals, noise and uncertainties in proxy records. It is beyond the scope of this review to discuss the classification of these constituent elements for individual proxy systems (a proxy-specific discussion is provided in the review by Jones et al.[9]), but it is important to note that from the perspective of large-scale reconstruction efforts the characterization of noise and uncertainties is in part contingent on the statistical models that are embedded in calibration or reconstruction techniques. For example, calibrations that assume univariate linear relationships between proxies and temperature by definition leave any proxy variance defined by other climate covariates or nonlinear proxy responses as noise. Attempts to characterize signal-to-noise ratios (SNRs) under these univariate linear assumptions have demonstrated a wide range (~0.1–1.3) of SNRs in contemporary multiproxy networks.[105] Moreover, the impact of SNR on the spatial uncertainties of CFRs can have complicated spatial patterns,[81,105,167] including the spatial expression of spectral biases in high and low frequencies.[167] Continued efforts to better characterize signal, noise and uncertainties in proxy systems is therefore important and must ultimately be connected to methodological efforts to derive large-scale reconstructions and their uncertainties. Moreover, one person's noise may be another person's signal, and further understanding the often multivariate climate-proxy relationship is important for better mining each proxy record for the complete climate signal that it contains. As we have already described, proxy system models use approximations of the nonlinear and multivariate responses of proxy systems in attempts to improve proxy interpretation and evaluation. Application of these models in reconstruction contexts may ultimately prove more effective than the univariate linear models that have been primarily used to date. Alternatively, more guided network selections may also allow tailored networks that better match the underlying assumptions of reconstruction methodologies.[101]

Research into reconstruction methods also has brought new insights and best practices to the field. Various forms of index reconstruction methods are now available, many of which appear to be skillful at reconstructing regional to global temperature indices given assumed levels and character of noise in the proxy records.[98,110,113,117,193] However, work to understand CFR skill reveals that uncertainties remain. Perhaps most fundamentally, these methods unsurprisingly do best where data are dense,[81,98,105,106,111,167] pointing to the obvious need for more proxy records in regions of limited or no sampling. Despite the fact that the spatiotemporal information of a given CFR is its greatest utility, spatial patterns in these reconstructions are also still subject to considerable uncertainties and should be interpreted cautiously. Emerging techniques based on formulations that go beyond multivariate regression models show promise as new CFR techniques, including Bayesian hierarchical modeling,[100,111,161,162,194] Markov random fields[107] and data assimilation.[155,156,165,195] Even for traditional CFR methods, more is being done to vet reconstruction products[106] and determine how to compare their characteristics beyond aggregated spatial means.[103,104] Increased proxy densities may also allow more locally constrained approaches[130–134,159] and enhanced attention on the seasonal window during which the proxies are sensitive.[113,196] This latter issue is particularly critical and deserves further attention in large-scale reconstruction efforts. While seasonal sensitivities have long been discussed in the respective proxy communities, and to some extent in the large-scale reconstruction literature,[196] multiproxy networks are often used to target annual climate fields despite comprising proxies with many different seasonal windows of sensitivity.[5,106,117,138] Some large-scale reconstruction efforts have confined the season of temperature reconstruction,[100,113,134] but further tests on the impacts of seasonal sampling windows is warranted.

Reconstruction and new modeling efforts also have coalesced to form important insights in the study of the CE. Phase 3 of the Paleoclimate Modelling Intercomparison Project and Phase 5 of the

Coupled Model Intercomparison Project (PMIP3/CMIP5), produced in conjunction with AR5, adopted the last-millennium experiment into the PMIP3/CMIP5 framework.[69] An important aspect of the experimental design was that all of the simulations were created with the same configurations and resolutions as the equivalent model experiments for the historical, future projections and preindustrial control runs. This has revolutionized the ability of paleoclimatic model-data comparisons to have direct relevance for 21st-century climate projections. Analyses of last-millennium simulations consequently have more broadly focused on interpretations of last-millennium dynamics[197–202] and on benchmarking future projections against the last millennium.[203] In a similar vein, more modeling centers are beginning to produce last-millennium ensemble simulations that include single forcing runs,[71,204,205] which allow for robust separation of forced and internal variability and assessment of climate responses to individual radiative forcings. These efforts have made comparisons between reconstructions and model simulations a powerful means of gaining insight into the climate of the CE that goes beyond hemispheric or global means, and has similarly provided evaluations of model performance with direct quantitative relevance to future projections. Methods of model-data comparison are also fundamental, and recent research is exploring how best to perform them[197–200,206,207] and how they might be used to constrain future projections.[208]

AN EMERGING ORCHESTRA

All of the above developments are driving a new age of research on the climate of the CE and contributing to a detailed historical account of how and why climate evolved over the last several millennia. These advances are encouraging more cohesive and collaborative investigations into the dynamics of climate variability and change over decades to centuries. Understanding and characterizing how climate varies on these time scales is fundamental to assessing the full range of risks associated with the combined impact of anthropogenic climate change and internal climate variability – the near-term future will be significantly impacted by both.[209,210]

It may indeed be time to abandon the hockey metaphor altogether, in favor of one that portrays the research evolution in this field as a collection of musicians in an orchestra who are warming up and tuning their instruments prior to a performance. The cacophonous mixture of sound during such moments comprises contributions from many instruments, each with their own tonal range of frequencies and each contributing to the collective sound in the chamber. As the musicians continue to warm up, however, the sound becomes more cohesive and instead of disharmony, the ensemble of instruments begins to coalesce around common meters and melodies. The continuing attempts to reconstruct and understand the climate of the CE are not unlike this orchestral analogy. Different proxies, methods and models passed through a period of dissonance, but from the early cacophony some harmony has arisen. Subsequent progress in several important areas has revealed common tones in an emerging scientific composition, one that provides an improved understanding of how Earth's climate has varied over the CE[d]. Our understanding of climate over the last 2000 years will never be defined with the precision of a symphonic composition, but progress in many important areas has discovered common tones that now contribute to an improved understanding. Prospects for the future include the addition of new instruments to our scientific orchestra; some will enhance our coverage of temporal frequencies, others will provide greater geographic coverage, and still others will illuminate not only temperature, but also hydrological variables, salinity, ice cover and more. Continued development of reconstruction techniques will improve the synthesis of these new and growing proxy networks while better characterizing the strengths and uncertainties of the spatial and temporal information in reconstruction products.

Perhaps the ultimate prospect will be to incorporate the empirical knowledge gleaned from proxy reconstructions into the evaluation and improvement of climate model simulations. These efforts will be rooted in data-model comparisons over the CE that evaluate model performance and constrain projections of future climate based on these comparisons. In this truly quantitative way, the spirit expressed in the Churchill quote that opened this review can actually be realized as a means of peering into the future with stronger constraints from the past.

NOTES

[a] Earlier nomenclature referenced the First and Second Assessment Reports as the FAR and SAR respectively, but we eschew this designation because of the eventual ambiguity among the First, Fourth, and Fifth Assessment Reports. We reference the five IPCC Assessment Reports as AR1 through AR5.

[b] Despite its wide use in the dendroclimatic literature, *standardization* can be ambiguous in this context because it is used differently in other branches of data analysis and

statistics and even in the context of climate reconstructions where it can mean the normalization and centering of a timeseries. *Detrending* is therefore often used as an alternative and synonymous term in the literature. We nevertheless continue to use the term standardization herein, given its widespread usage in the dendroclimatological literature.

[c] Although not applied to reconstruct the global temperature field, various authors[128,129] used the reduced space optimal interpolation method to reconstruct Pacific Ocean sea surface temperatures from 1607–1990. Similarly, the point-by-point regression scheme[130] was developed to produce continental-scale CFRs targeting hydroclimate[130–133] and temperature.[134] Alternative CFR methods were therefore being discussed in the literature contemporaneous to the hemispheric and global temperature debates, but they were being applied in different contexts.

[d] From the perspective of this climate analogy, it is ironic that the instrument (the oboe) that is commonly used for the tonal calibration of the other instruments during the warm-up period is selected because its voice, among all the instruments, is least sensitive to the variable conditions of temperature, humidity, and air circulation in the orchestra hall.

ACKNOWLEDGMENTS

We thank two anonymous reviewers for their helpful suggestions and comments, and Dave Chapman for comments on an early draft of this manuscript. Lamont contribution #8016.

REFERENCES

1. Folland CK, Karl TR, Vinnikov KY. Observed climate variations and change. In: Houghton JT, Jenkins GJ, Ephraums JJ, eds. *Climate Change, the IPCC Scientific Assessment*. WMO/UNEP/IPCC. Cambridge and New York: Cambridge University Press; 1990.
2. Nicholls N, Grtuza GV, Jouzel J, Karl TR, Ogallo LA, Parker DA. Observed climate variability and change. In: Houghton JT, Meira Filho LG, Callander BA, Harris N, Kattenberg A, Maskell K, eds. *Climate Change 1995, The Science of Climate Change, in the Second Assessment Report of the IPCC*. Cambridge and New York: Cambridge University Press; 1995.
3. Folland CK, Karl TR, Christy JR, Clarke RA, Gruza GV, Jouzel J, Mann ME, Oerlemans J, Salinger MJ, Wang SW. Observed climate variability and change. In: Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA, eds. *Climate Change 2001, The Scientific Basis, in the Third Assessment Report of the IPCC*. Cambridge, and New York: Cambridge University Press; 2001.
4. Jansen E, Overpeck J, Briffa KR, Duplessy JC, Joos F, Masson-Delmotte V, Olago D, Otto-Bliesner B, Peltier WR, Rahmstorf S, et al. Palaeoclimate. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL, eds. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge and New York: Cambridge University Press; 2007.
5. Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM, eds. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge and New York: Cambridge University Press; 2013.
6. Jones PD, Mann ME. Climate over past millennia. *Rev Geophys* 2004, 42:RG2002. doi:10.1029/2003RG000143.
7. North GR, Biondi F, Bloomfield P, Christy JR, Cuffey KM, Dickinson RE, Druffel ERM, Nychka D, Otto-Bliesner B, Roberts N, et al. *Surface Temperature Reconstructions for the Last 2,000 Years*. Washington, DC: The National Academies Press; 2006.
8. Mann ME. Climate over the past two millennia. *Annu Rev Earth Planet Sci* 2007, 35:111–136.
9. Jones PD, Briffa KR, Osborn TJ, Lough JM, Van Ommen TD, Vinther BM, Luterbacher J, Wahl ER, Zwiers FW, Mann ME, et al. High-resolution palaeoclimatology of the last millennium: a review of current status and future prospects. *Holocene* 2009, 19:3. doi:10.1177/0959683608098952.
10. Frank D, Esper J, Zorita E, Wilson R. A noodle, hockey stick, and spaghetti plate: a perspective on high-resolution paleoclimatology. *WIREs Clim Change* 2010, 1:507–516. doi:10.1002/wcc.53.
11. Smerdon JE. Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments. *WIREs Clim Change* 2012, 3:63–77. doi:10.1002/wcc.149.
12. Imbrie F, Imbrie KP. *Ice Ages: Solving the Mystery*. 2nd ed. Cambridge, MA: Harvard University Press; 1986.

13. Cook ER. *Dendrochronology and Dendroclimatology: Encyclopedia of Environmetrics*. John Wiley & Sons, Ltd; 2006. Available at: http://onlinelibrary.wiley.com/doi/10.1002/9780470057339.vad013/abstract.
14. Fritts HC. Dendroclimatology and dendroecology. *Quat Res* 1971, 1:419–449.
15. Fritts HC. *Tree Rings and Climate*. London: Academic Press; 1976.
16. Groveman B, Landsberg H. Simulated northern hemisphere temperature departures 1579–1880. *Geophys Res Lett* 1979, 6:767–769.
17. Landsberg H, Groveman B, Hakkarinen I. A simple method for approximating the annual temperature of the Northern Hemisphere. *Geophys Res Lett* 1978, 5:505–506.
18. Jacoby G, D'Arrigo R. Reconstructed Northern Hemisphere annual temperature since 1671 based on high-latitude tree-ring data from North America. *Clim Change* 1989, 14:39–59.
19. Fagan B. *The Little Ice Age: How Climate Made History 1300–1850*. New York: Basic Books; 2000.
20. Fagan B. *The Great Warming: Climate Change and the Rise and Fall of Civilizations*. New York: Bloomsbury Press; 2008.
21. Bradley RS, Jones PD. Little Ice Age summer temperature variations; their nature and relevance to recent global warming trends. *Holocene* 1993, 3:367–376.
22. Jones PD, Briffa KR, Barnett TP, Tett SFB. High-resolution palaeoclimatic records for the last millennium: interpretation, integration and comparison with General Circulation Model control run temperatures. *Holocene* 1998, 8:455–471.
23. Mann ME, Bradley RS, Hughes MK. Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* 1998, 392:779–787.
24. Mann ME, Bradley RS, Hughes MK. Northern hemisphere temperatures during the past millennium: inferences, uncertainties, and limitations. *Geophys Res Lett* 1999, 26:759–762.
25. Briffa KR. Annual climate variability in the Holocene: interpreting the message of ancient trees. *Quat Sci Rev* 2000, 19:87–105.
26. Morice CP, Kennedy JJ, Rayner NA, Jones PD. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 data set. *J Geophys Res* 2012, 117:D08101.
27. Lamb H. The early medieval warm epoch and its sequel. *Palaeogeogr Palaeoclimatol Palaeoecol* 1965, 1:13–37. doi:10.1016/0031-0182(65)90004-0.
28. Manley G. The mean temperature of central England, 1698–1952. *Q J R Meteorol Soc* 1953, 79:242–261. doi:10.1002/qj.49707934006.
29. Oreskes N, Conway EM. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. New York: Bloomsbury Press; 2010.
30. Bradley RS. *Global Warming and Political Intimidation*. Amherst, MA: University of Massachusetts Press; 2011.
31. Powell JL. *The Inquisition of Climate Science*. New York: Columbia University Press; 2011.
32. Mann ME. *The Hockey Stick and the Climate Wars: Dispatches from the Front Lines*. New York: Columbia University Press; 2012.
33. Shindell DT, Schmidt GA, Miller RL, Mann ME. Volcanic and solar forcing of climate change during the preindustrial era. *J Clim* 2003, 16:4094–4107.
34. Mann ME, Hughes MK. Tree-ring chronologies and climate variability. *Science* 2002, 296:848.
35. Broecker WS. Was the medieval warm period global? *Science* 2001, 291:1497–1499.
36. Cook ER, Briffa KR, Meko DM, Graybill DA, Funkhouser G. The 'segment length curse' in long tree-ring chronology development for palaeoclimatic studies. *Holocene* 1995, 5:229–237.
37. Briffa KR, Jones PD, Bartholin TS, Eckstein D, Schweingruber FH, Karlén W, Zetterberg P, Eronen M. Fennoscandian summers from ad 500: temperature changes on short and long timescales. *Clim Dyn* 1992, 7:111–119.
38. Esper J, Cook ER, Schweingruber FH. Low-frequency signals in long tree-ring chronologies for reconstruction of past temperature variability. *Science* 2002, 295:2250–2253.
39. Pollack HN, Huang S, Shen PY. Climate change record in subsurface temperatures: a global perspective. *Science* 1998, 282:279–281. doi:10.1126/science.282.5387.279.
40. Huang S, Pollack HN, Shen PY. Temperature trends over the past five centuries reconstructed from borehole temperatures. *Nature* 2000, 403:756–758.
41. Harris RN, Chapman DS. Mid-latitude (30°–60°N) climatic warming inferred by combining borehole temperatures with surface air temperatures. *Geophys Res Lett* 2001, 28:747–750.
42. Beltrami H. Climate from borehole data: energy fluxes and temperatures since 1500. *Geophys Res Lett* 2002, 29:2111. doi:10.1029/2002GL015702.
43. Pollack HN, Smerdon JE. Borehole climate reconstructions: spatial structure and hemispheric averages. *J Geophys Res* 2004, 109:D11106. doi:10.1029/2003JD004163.
44. Pollack HN, Huang S. Climate reconstruction from subsurface temperatures. *Annu Rev Earth Planet Sci* 2000, 28:339–365.

45. Cermak V. Underground temperature and inferred climatic temperature of the past millennium. *Palaeogeogr Palaeoclimatol Palaeoecol* 1971, 10:1–19.
46. Cermak V, Bodri L, Safanda J. Underground temperature fields and changing climate: evidence from Cuba. *Palaeogeogr Palaeoclimatol Palaeoecol* 1992, 97:325–337.
47. Huang S, Shen PY, Pollack HN. Deriving century-long trends of surface temperature change from borehole temperatures. *Geophys Res Lett* 1996, 23:257–260.
48. Mann ME, Schmidt GA. Ground vs. surface air temperature trends: implications for borehole surface temperature reconstructions. *Geophys Res Lett* 2003, 30:1607. doi:10.1029/2003GL017170.
49. Mann ME, Rutherford S, Bradley RS, Hughes MK, Keiming FT. Optimal surface temperature reconstructions using terrestrial borehole data. *J Geophys Res* 2003, 108(D7):4203. doi:10.1029/2002JD002532.
50. Stieglitz M, Dery SJ, Romanovsky VE, Osterkamp TE. The role of snow cover in the warming of arctic permafrost. *Geophys Res Lett* 2003, 30:1721. doi:10.1029/2003GL017337.
51. Chapman DS, Bartlett MG, Harris RN. Comment on “Ground vs. surface air temperature trends: Implications for borehole surface temperature reconstructions” by M. E. Mann and G. Schmidt. *Geophys Res Lett* 2004, 31:L07205. doi:10.1029/2003GL019054.
52. Huang S. Merging information from different resources for new insights into climate change in the past and future. *Geophys Res Lett* 2004, 31:L13205. doi:10.1029/2004GL019781.
53. Lin X, Smerdon JE, England AW, Pollack HN. A model study of the effects of climatic precipitation changes on ground temperatures. *J Geophys Res* 2003, 108(D7):4230. doi:10.1029/2002JD002878.
54. Smerdon JE, Pollack HN, Enz JW, Lewis MJ. Conduction-dominated heat transport of the annual temperature signal in soil. *J Geophys Res* 2003, 108(B9):2431. doi:10.1029/2002JB002351.
55. Smerdon JE, Pollack HN, Cermak V, Enz JW, Kresl M, Safanda J, Wehmiller JF. Air-ground temperature coupling and subsurface propagation of annual temperature signals. *J Geophys Res* 2004, 109:D21107. doi:10.1029/2004JD005056.
56. Smerdon JE, Pollack HN, Cermak V, Enz JW, Kresl M, Safanda J, Wehmiller JF. Daily, seasonal, and annual relationships between air and subsurface temperatures. *J Geophys Res* 2006, 111:D07101. doi:10.1029/2004JD005578.
57. Bartlett MG, Chapman DS, Harris RN. Snow and the ground temperature record of climate change. *J Geophys Res* 2004, 109:F04008. doi:10.1029/2004JF000224.
58. Bartlett MG, Chapman DS, Harris RN. Snow effect on North American ground temperatures, 1950–2002. *J Geophys Res* 2005, 110:F03008. doi:10.1029/2005JF000293.
59. Beltrami H, Kellman L. An examination of short- and long-term air-ground temperature coupling. *Glob Planet Change* 2003, 38:291–303.
60. González-Rouco F, von Storch H, Zorita E. Deep soil temperature as proxy for surface air-temperature in a coupled model simulation of the last thousand years. *Geophys Res Lett* 2003, 30:2116. doi:10.1029/2003GL018264.
61. González-Rouco JF, Beltrami H, Zorita E, von Storch H. Simulation and inversion of borehole temperature profiles in surrogate climates: spatial distribution and surface coupling. *Geophys Res Lett* 2006, 33:L01703. doi:10.1029/2005GL024693.
62. Briffa KR, Osborn TJ. Blowing hot and cold. *Science* 2002, 295:2227–2228.
63. Rutherford S, Mann ME. Correction to “Optimal surface temperature reconstructions using terrestrial borehole data.” *J Geophys Res* 2004, 109:D11107. doi:10.1029/2003JD004290.
64. Schmidt GA, Mann ME. Reply to comment on “Ground vs. surface air temperature trends: Implications for borehole surface temperature reconstructions” by D. Chapman et al. *Geophys Res Lett* 2004, 31:L07206. doi:10.1029/2003GL019144.
65. García-García A, Cuesta-Valero FJ, Beltrami H, Smerdon JE. Simulation of air and ground temperatures in PMIP3/CMIP5 last millennium simulations: implications for climate reconstructions from borehole temperature profiles. *Env Res Lett* 2016, 11:044022. doi:10.1088/1748-9326/11/4/044022.
66. Osborn TJ, Raper SCB, Briffa KR. Simulated climate change during the last 1,000 years: comparing the ECHO-G general circulation model with the MAGICC simple climate model. *Clim Dyn* 2006, 27:185–197.
67. Mann ME, Rutherford S, Wahl E, Ammann C. Testing the fidelity of methods used in proxy-based reconstructions of past climate. *J Clim* 2005, 18:4097–4107.
68. Ammann CM, Joos F, Schimel DS, Otto-Bliesner BL, Tomas RA. Solar influence on climate during the past millennium: results from transient simulations with the NCAR Climate System Model. *Proc Natl Acad Sci USA* 2007, 104:3713–3718. doi:10.1073-pnas.0605064.103.
69. Taylor KE, Stouffer RJ, Meehl GA. An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 2012, 93:485–498. doi:10.1175/BAMS-D-11-00094.1.
70. Fernández-Donado L, González-Rouco JF, Raible CC, Ammann CM, Barriopedro D, García-Bustamante E, Jungclaus JH, Lorenz SJ,

Luterbacher J, Phipps SJ, et al. Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium. *Clim Past* 2013, 9:393–421. doi:10.5194/cp-9-393-2013.

71. Jungclaus JH, Lorenz SJ, Timmreck C, Reick CH, Brovkin V, Six K, Segschneider J, Giorgetta MA, Crowley TJ, Pongratz J, et al. Climate and carbon-cycle variability over the last millennium. *Clim Past* 2010, 6:723–737. doi:10.5194/cp-6-723-2010.
72. von Storch H, Zorita E, Jones JM, Dimitriev Y, González-Rouco F, Tett SFB. Reconstructing past climate from noisy data. *Science* 2004, 306:679–682.
73. Zorita E, Gonzalez-Rouco F, Legutke S. Testing the Mann et al. (1998) approach to paleoclimate reconstructions in the context of a 1000-yr control simulation with the ECHO-G coupled climate model. *J Clim* 2003, 20:1378–1390
74. von Storch H, Zorita E, Jones JM, Dimitriev Y, González-Rouco F, Tett SFB. Response to comment on “Reconstructing past climate from noisy data”. *Science* 2006, 312:529c.
75. Wahl ER, Ritson DM, Ammann CM. Comment on “Reconstructing past climate from noisy data”. *Science* 2006, 312:529b.
76. von Storch H, Zorita E, González-Rouco F. Assessment of three temperature reconstruction methods in the virtual reality of a climate simulation. *Int J Earth Sci* 2009, 98:67–82.
77. Mann ME, Rutherford S, Wahl E, Ammann C. Robustness of proxy-based climate field reconstruction methods. *J Geophys Res* 2007, 112:D12109. doi:10.1029/2006JD008272.
78. Mann ME, Rutherford S, Wahl E, Ammann C. Reply to comments on “Testing the fidelity of methods used on proxy-based reconstructions of past climate” by Zorita et al. *J Clim* 2007, 20:3699–3703.
79. Rutherford S, Mann ME, Wahl E, Ammann C. Reply to comment by Jason E. Smerdon et al. on “Robustness of proxy-based climate field reconstruction methods”. *J Geophys Res* 2008, 113:D18107. doi:10.1029/2008JD009964.
80. Smerdon JE, González -Rouco JF, Zorita E. Comment on “Robustness of proxy-based climate field reconstruction methods” by Michael E. Mann et al. *J Geophys Res* 2008, 113:–D18106. doi:10.1029/2007JD009542.
81. Smerdon JE, Kaplan A, Zorita E, González-Rouco JF, Evans MN. Spatial performance of four climate field reconstruction methods targeting the Common Era. *Geophys Res Lett* 2011, 38:L11705. doi:10.1029/2011GL047372.
82. Zorita E, González-Rouco JF, von Storch H. Comment on “Testing the fidelity of methods used in proxy-based reconstructions of past climate” by Mann et al. *J Clim* 2007, 20:2693–3698.
83. Rahmstorf S. Testing climate reconstructions. *Science* 2006, 312:1872.
84. D'Arrigo R, Wilson R, Liepert B, Cherubini P. On the ‘divergence problem’ in Northern forests: a review of the tree-ring evidence and possible causes. *Glob Planet Change* 2008, 60:289–305.
85. Briffa KR, Schweingruber FH, Jones PD, Osborn TJ, Shiyatoff SG, Vaganov EA. Reduced sensitivity of recent tree-growth to temperature at high northern latitudes. *Nature* 1998, 391:678–682.
86. McIntyre S, McKitrick R. Hockey sticks, principal components, and spurious significance. *Geophys Res Lett* 2005, 32:L03710. doi:10.1029/2004GL021750.
87. von Storch H, Zorita E. Comment on “Hockey sticks, principal components, and spurious significance” by S. McIntyre and R. McKitrick. *Geophys Res Lett* 2005, 32:L20701. doi:10.1029/2005GL022753.
88. Huybers P. Comment on “Hockey sticks, principal components, and spurious significance” by S. McIntyre and R. McKitrick. *Geophys Res Lett* 2005, 32:L20705. doi:10.1029/2005GL023395.
89. McIntyre S, McKitrick R. Reply to comment by von Storch and Zorita on “Hockey sticks, principal components, and spurious significance”. *Geophys Res Lett* 2005, 32:L20714. doi:10.1029/2005GL023089.
90. McIntyre S, McKitrick R. Reply to comment by Huybers on “Hockey sticks, principal components,and spurious significance”. *Geophys Res Lett* 2005, 32: L20713. doi:10.1029/2005GL023586.
91. Wahl E, Ammann CM. Robustness of the Mann, Bradley, Hughes reconstruction of Northern Hemisphere surface temperatures: examination of criticisms based on the nature and processing of proxy climate evidence. *Clim Change* 2007, 85:33–69. doi:10.1007/s10584-006-9105-7.
92. Ammann C, Wahl E. The importance of the geophysical context in statistical evaluations of climate reconstruction procedures. *Clim Change* 2007, 85:71–88. doi:10.1007/s10584-007-9276-x.
93. Bürger G, Cubasch U. Are multiproxy climate reconstructions robust? *Geophys Res Lett* 2005, 32: L23711. doi:10.1029/2005GL024155.
94. Emile-Geay J, Eshleman JA. Toward a semantic web of paleoclimatology. *Goechem Geophys Geosyst* 2013, 14:457–469. doi:10.1002/ggge.20067.
95. PAGES 2k Consortium. A community-driven framework for climate reconstructions. *Eos Trans Am Geophys Union* 2014, 95:361.
96. Cook E, Briffa KR, Jones PD. Spatial regression methods in dendroclimatology: a review and comparison of two techniques. *Int J Clim* 1994, 14:379–402.
97. McShane BB, Wyner AJ. A statistical analysis of multiple temperature proxies: are reconstructions of surface temperatures over the last 1000 years reliable? *Ann Appl Stat* 2011, 5:5–44.

98. Wahl ER, Smerdon JE. Comparative performance of paleoclimatic field and index reconstructions derived from climate proxies and noise-only predictors. *Geophys Res Lett* 2012, 39:L06703. doi:10.1029/2012GL051086.

99. Schneider L, Smerdon JE, Büntgen U, Wilson RJS, Myglan VS, Kirdyanov AV, Esper J. Revising midlatitude summer temperatures back to AD 600 based on a wood density network. *Geophys Res Lett* 2015, 42:4556–4562. doi:10.1002/2015GL063956.

100. Luterbacher J, Werner JP, Smerdon JE, Fernández-Donado L, Gonzalez-Rouco JF, Barriopedro D, Ljungqvist F, Büntgen U, Zorita E, Wagner S, et al. - European summer temperatures since Roman times. *Environ Res Lett* 2016, 11:024001. doi:10.1088/1748-9326/11/2/024001.

101. Wilson R, Anchukaitis K, Briffa KR, Büntgen U, Cook E, D'Arrigo R, Davi N, Esper J, Franke D, Gunnarsonk B, et al. Last millennium northern hemisphere summer temperatures from tree rings. Part I: the long term context. *Quat Sci Rev* 2016, 134:1–18.

102. Gómez-Navarro JJ, Werner J, Wagner S, Luterbacher J, Zorita E. Establishing the skill of climate field reconstruction techniques for precipitation with pseudoproxy experiments. *Clim Dyn* 2015, 45:1395–1413.

103. Li B, Smerdon JE. Defining spatial assessment metrics for evaluation of paleoclimatic field reconstructions of the Common Era. *Environmetrics* 2012, 23:394–406.

104. Li B, Zhang X, Smerdon JE. Comparison between spatio-temporal random processes and application to climate model data, *Environmetrics*. In press.

105. Wang J, Emile-Geay J, Guillot D, Smerdon JE, Rajaratnam B. Evaluating climate field reconstruction techniques using improved emulations of real-world conditions. *Clim Past* 2014, 10:1–19. doi:10.5194/cp-10-1-2014.

106. Wang J, Emile-Geay J, Guillot D, McKay NP, Rajaratnam B. Fragility of reconstructed temperature patterns over the Common Era: implications for model evaluation. *Geophys Res Lett* 2015, 42:7162–7170. doi:10.1002/2015GL065265.

107. Guillot D, Rajaratnam B, Emile-Geay J. Statistical paleoclimate reconstructions via Markov random fields. *Ann Appl Stat* 2015, 9:324–352.

108. Neukom R, Luterbacher J, Villalba R, Küttel M, Frank D, Jones PD, Grosjean M, Esper J, Lopez L, Wanner H. Multi-centennial summer and winter precipitation variability in southern South America. *Geophys Res Lett* 2010, 37:L14708. doi:10.1029/2010GL043680.

109. Anchukaitis KJ, D'Arrigo RD, Andreu-Hayles L, Frank D, Verstege A, Curtis A, Buckley BM, Jacoby GC, Cook ER. Tree-ring-reconstructed summer temperatures from northwestern North America during the last nine centuries. *J Clim* 2013, 26:3001–3012. doi:10.1175/JCLI-D-11-00139.1.

110. Hegerl GC, Crowley T, Allen M, Hyde WT, Pollack H, Smerdon J, Zorita E. Detection of human influence on a new 1500-yr climate reconstruction. *J Clim* 2007, 20:650–666. doi:10.1175/JCLI4011.1.

111. Werner JP, Luterbacher J, Smerdon JE. A pseudoproxy evaluation of bayesian hierarchical modeling and canonical correlation analysis for climate field reconstructions over Europe. *J Clim* 2013, 26:851–867.

112. Moberg A, Sonechkin DM, Holmgren K, Datsenko NM, Karlen W. Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature* 2005, 433:613–617.

113. PAGES 2k Consortium. Continental-scale temperature variability over the last two millennia. *Nat Geosci* 2013, 6:339–346. doi:10.1038/ngeo1797.

114. Hegerl GC, Crowley TJ, Hyde WT, Frame D. Climate sensitivity constrained by temperature reconstructions of the last seven centuries. *Nature* 2006, 440:1029–1032.

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

116. Oerlemanns J. Extracting a climate signal from 169 glacier records. *Science* 2005, 308:675–677.

117. Mann ME, Zhang Z, Hughes MK, Bradley RS, Miller SK, Rutherford S, Ni F. Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc Natl Acad Sci USA* 2008, 105:13252–13257.

118. Christiansen B, Ljungqvist FC. The extra-tropical Northern Hemisphere temperature in the last two millennia: reconstructions of low-frequency variability. *Clim Past* 2012, 8:765–786.

119. Ljungqvist FC. A new reconstruction of temperature variability in the extratropical northern hemisphere during the last two millennia. *Geogr Ann Ser A-Phys Geogr* 2010, 92:339–351.

120. Shi F, Yang B, Mairesse A, von Gunten L, Li J, Bräuning A, Yang F, Xiao X. Northern hemisphere temperature reconstruction during the last millennium using multiple annual proxies. *Clim Res* 2013, 56:231–244.

121. Frank D, Esper J, Cook ER. Adjustment for proxy number and coherence in a large-scale temperature reconstruction. *Geophys Res Lett* 2007, 34:L16709.

122. Juckes MN, Allen MR, Briffa KR, Esper J, Hegerl GC, Moberg A, Osborn TJ, Weber SL. Millennial temperature reconstruction intercomparison and evaluation. *Clim Past* 2007, 3:591–609.

123. Loehle C, McCulloch JH. Correction to: a 2000-year global temperature reconstruction based on non-tree ring proxies. *Energy Environ* 2008, 19:93–100.

124. Leclercq PW, Oerlemans J. Global and hemispheric temperature reconstruction from glacier length fluctuations. *Clim Dyn* 2012, 38:1065–1079.

125. Schaefer JM, Denton GH, Kaplan M, Putnam A, Finkel RC, Barrell DJA, Andersen BG, Schwartz R, Mackintosh A, Chinn T, et al. High frequency Holocene glacier fluctuations in New Zealand differ from the northern signature. *Science* 2009, 324:622–625. doi:10.1126/science.1169312.

126. Young NE, Schweinsberg AD, Briner JP, Schaefer JM. Glacier maxima in Baffin Bay during the medieval warm period coeval with Norse settlement. *Sci Adv* 2015, 1:e1500806. doi:10.1126/sciadv.1500806.

127. Schneider T. Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. *J Clim* 2001, 14:853–887.

128. Evans MN, Kaplan A, Cane MA. Pacific sea surface temperature field reconstruction from coral d18O data using reduced space objective analysis. *Paleoceanography* 2002, 17:10-1–10-8. doi:10.1029/2000PA000590.

129. Evans MN, Kaplan A, Cane MA, Villalba R. Globality and optimality in climate field reconstructions from proxy data. In: Markgraf V, ed. *Present and Past Inter-Hemispheric Linkages in the Americas and Their Societal Effects*. Cambridge and New York: Cambridge University Press; 2001.

130. Cook ER, Meko DM, Stahle DW, Cleaveland MK. Drought reconstructions for the continental United States. *J Clim* 1999, 12:1145–1162.

131. Cook BI, Smerdon JE, Seager R, Cook ER. Pan-continental droughts in North America over the last millennium. *J Clim* 2014, 27:383–397. doi:10.1175/JCLI-D-13-00100.1.

132. Cook ER, Anchukaitis KJ, Buckley BM, D'Arrigo RD, Jacoby GC, Wright WE. Asian monsoon failure and megadrought during the last millennium. *Science* 2010, 328:486–489.

133. Cook ER, Seager R, Kushnir Y, Briffa KR, Büntgen U, Frank D, Krusic PJ, Tegel W, van der Schrier G, Andreu-Hayles L, et al. Old World megadroughts and pluvials during the Common Era. *Sci Adv* 2015, 1:e1500561. doi:10.1126/sciadv.1500561.

134. Cook ER, Krusic PJ, Anchukaitis KJ, Buckley BM, Nakatsuka T, Sano M. PAGES Asia2k Members. Tree-ring reconstructed summer temperature anomalies for temperate East Asia since 800 C.E. *Clim Dyn* 2013, 41:2957–2972. doi:10.1007/s00382-012-1611-x.

135. Rutherford S, Mann ME, Delworth TL, Stouffer RJ. Climate field reconstruction under stationary and nonstationary forcing. *J Clim* 2003, 16:462–479.

136. Smerdon JE, Kaplan A. Comments on "Testing the fidelity of methods used in proxy-based reconstructions of past climate": the role of the standardization interval. *J Clim* 2007, 20:5666–5670.

137. Smerdon JE, Kaplan A, Chang D. On the standardization sensitivity of RegEM climate field reconstructions. *J Clim* 2008, 21:6710–6723.

138. Mann ME, Zhang Z, Rutherford S, Bradley RS, Hughes MK, Shindell D, Ammann C, Faluvegi G, Ni F. Global signatures and dynamical origins of the little ice age and the medieval climate anomaly. *Science* 2009, 326:1256–1260. doi:10.1126/science.1177303.

139. Cook RD, Weisberg S. *Applied regression including computing and graphics*. New York: John Wiley & Sons; 1999.

140. Smerdon JE, Kaplan A, Chang D, Evans MN. A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millennium. *J Clim* 2011, 24:1284–1309. doi:10.1175/2010JCLI4110.1.

141. Hoerl AE, Kennard RW. Ridge regression: biased estimation for non-orthogonal problems. *Technometrics* 1970, 12:55–67.

142. Tikhonov AN, Arsenin VY. *Solution of Ill-Posed Problems*. Scripta Series in Mathematics. Washington, DC: V. H. Winston and Sons; 1970.

143. Van Huffel S, Vandewalle J. *The Total Least Squares Problem: Computational Aspects and Analysis*. Frontiers in Applied Mathematics 9. Philadelphia, PA: SIAM; 1991.

144. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol* 1996, 58:267–288.

145. Fierro RD, Golub GH, Hansen PC, O'Leary DP. Regularization by truncated total least squares. *SIAM J Sci Comput* 1997, 18:1223–1241.

146. Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 2008, 9:432–441. doi:10.1093/biostatistics/kxm045.

147. Tingley MP, Craigmile PF, Haran M, Li B, Mannshardt-Shamseldin E, Rajaratnam B. Piecing together the past: statistical insights into paleoclimatic reconstructions. *Quat Sci Rev* 2012, 345:1–22.

148. Christiansen B, Schmith T, Thejll P. Reply. *J Clim* 2010, 23:2839–2844.

149. Rutherford SD, Mann ME, Ammann CM, Wahl ER. Comments on: "A surrogate ensemble study of climate reconstruction methods: stochasticity and robustness". *J Clim* 2010, 23:2832–2838.

150. Christiansen B, Schmith T, Thejll P. A surrogate ensemble study of climate reconstruction methods: stochasticity and robustness. *J Clim* 2009, 22:951–976. doi:10.1175/2008JCLI2301.1.
151. Smerdon JE, Kaplan A, Amrhein DE. Erroneous model field representations in multiple pseudoproxy studies: corrections and implications. *J Clim* 2010, 23:5548–5554.
152. Smerdon JE, Kaplan A, Amrhein DE. Reply to comment by Rutherford et al. on "Erroneous model field representations in multiple pseudoproxy studies: corrections and implications". *J Clim* 2013, 26:3485–3486.
153. Rutherford SD, Mann ME, Wahl E, Ammann C. Comments on "Erroneous model field representations in multiple pseudoproxy studies: corrections and implications.". *J Clim* 2013, 26:3482–3484. doi:10.1175/JCLI-D-12-00065.1.
154. Delworth TL, Mann ME. Observed and simulated multidecadal variability in the Northern Hemisphere. *Clim Dyn* 2000, 16:661–676.
155. Goosse H, Guiot J, Mann ME, Dubinkina S, Sallaz-Damaz Y. The medieval climate anomaly in Europe: comparison of the summer and annual mean signals in two reconstructions and in simulations with data assimilation. *Glob Planet Change* 2012, 84–85:35–47.
156. Goosse H, Crespin E, Dubinkina S, Loutre M-F, Mann ME, Renssen H, Sallaz-Damaz Y, Shindell D. The role of forcing and internal dynamics in explaining the "Medieval Climate Anomaly". *Clim Dyn* 2012, 39:2847–2866.
157. Anchukaitis KJ, Buckley BM, Cook ER, Cook BI, D'Arrigo RD, Ammann CM. The influence of volcanic eruptions on the climate of the Asian monsoon region. *Geophys Res Lett* 2010, 37:L22703. doi:10.1029/2010GL044843.
158. Zhang Z, Mann ME, Cook ER. Alternative methods of proxy-based climate field reconstruction: application to summer drought over the conterminous United States back to AD 1700 from tree-ring data. *Holocene* 2004, 14:502–516.
159. Diaz H, Wahl E. Recent California water year precipitation deficits: a 440-year perspective. *J Clim* 2015, 28:4637–4652. doi:10.1175/JCLI-D-14-00774.1.
160. Riedwyl N, Kuttel M, Luterbacher J, Wanner H. Comparison of climate field reconstruction techniques: application to Europe. *Clim Dyn* 2009, 32:381–395.
161. Tingley MP, Huybers P. A Bayesian algorithm for reconstructing climate anomalies in space and time. Part I: development and applications to paleoclimate reconstruction problems. *J Clim* 2010, 23:2759–2781.
162. Tingley MP, Huybers P. A Bayesian algorithm for reconstructing climate anomalies in space and time. Part II: comparison with the regularized expectation maximization algorithm. *J Clim* 2010, 23:2782–2800.
163. Annan JD, Hargreaves JC. Identification of climatic state with limited proxy data. *Clim Past* 2012, 8:1141–1151. doi:10.5194/ cp-8-1141-2012.
164. Dannenberg MP, Wise EK. Performance of climate field reconstruction methods over multiple seasons and climate variables. *J Geophys Res Atmos* 2013, 118:9595–9610. doi:10.1002/jgrd.50765.
165. Steiger NJ, Hakim GJ, Steig EJ, Battisti DS, Roe GH. Assimilation of time-averaged pseudoproxies for climate reconstruction. *J Clim* 2014, 27:426–441.
166. Evans MN, Smerdon JE, Kaplan A, Tolwinski-Ward SE, González-Rouco JF. Climate field reconstruction uncertainty arising from multivariate and nonlinear properties of predictors. *Geophys Res Lett* 2014, 41:9127–9134. doi:10.1002/2014GL062063.
167. Smerdon JE, Coats S. *Ault TR*. Clim Dyn: Model-dependent spatial skill in pseudoproxy experiments testing climate field reconstruction methods for the Common Era; 2015, 2684. doi:10.1007/s00382-015-2684-0.
168. Clement AC, Seager R, Cane MA, Zebiak SE. An ocean dynamical thermostat. *J Clim* 1996, 9:2190–2196.
169. Cook ER, Seager R, Cane MA, Stahle DW. North American drought: reconstructions, causes, and consequences. *Earth Sci Rev* 2007, 81:93–134. doi:10.1016/j.earscirev.2006.12.002.
170. Cook BI, Cook ER, Smerdon JE, Seager R, Williams AP, Coats S, Stahle DW, Villanueva Diaz J. North American megadroughts in the Common Era: reconstructions and simulations, *WIRES Clim Change* 7:411–432.
171. Tierney JE, Abram NJ, Anchukaitis KJ, Evans MN, Giry C, Kilbourne KH, Saenger CP, Wu HC, Zinke J. Tropical sea-surface temperatures for the past four centuries reconstructed from coral archives. *Paleoceanography* 2015, 30:226–252. doi:10.1002/ 2014PA002717.
172. McGregor HV, Evans MN, Goosse H, Leduc G, Martrat B, Addison JA, Mortyn PG, Oppo DW, Seidenkrantz MS, Sicre MA, et al. Robust global ocean cooling trend for the pre-industrial Common Era. *Nat Geosci* 2015, 8:671–677. doi:10.1038/ ngeo2510.
173. Neukom R, Gergis J, Karoly DJ, Wanner H, Curran M, Elbert J, González-Rouco F, Linsley BK, Moy AD, Mundo I, et al. Inter-hemispheric temperature variability over the past millennium. *Nat Clim Change* 2014, 4:362–367. doi:10.1038/nclimate2174.
174. Büntgen U, Frank D, Wilson R, Carrer M, Urbinati C, Esper J. Testing for tree-ring divergence in the European Alps. *Glob Change Biol* 2008, 14:2443–2453.

175. Esper J, Frank D, Büntgen U, Verstege A, Hantemirov RM, Kirdyanov AV. Trends and uncertainties in Siberian indicators of 20th century warming. *Glob Change Biol* 2010, 16:386–398.

176. Bürger G. Comment on "The spatial extent of 20th-century warmth in the context of the past 1200 years". *Science* 2007, 316:1844. doi:10.1126/science.1140982.

177. Stevens MB, González-Rouco JF, Beltrami H. North American climate of the last millennium: underground temperatures and model comparison. *J Geophys Res* 2008, 113:F01008. doi:10.1029/2006JF000705.

178. Evans MN, Reichert BK, Kaplan A, Anchukaitis KJ, Vaganov EA, Hughes MK, Cane MA. A forward modeling approach to paleoclimatic interpretation of tree-ring data. *J Geophys Res* 2006, 111:G03008. doi:10.1029/2006JG000166.

179. Evans MN, Tolwinski-Ward SE, Thompson DM, Anchukaitis KJ. Applications of proxy system modeling in high resolution paleoclimatology. *Quat Sci Rev* 2013, 76:16–28.

180. Anchukaitis KJ, Evans MN, Kaplan A, Vaganov EA, Hughes MK, Grissino-Mayer HD, Cane MA. Forward modeling of regional scale tree-ring patterns in the southeastern United States and the recent influence of summer drought. *Geophys Res Lett* 2006, 36:2419–2439. doi:10.1029/2005GL025050.

181. Tolwinski-Ward SE, Evans MN, Hughes MK, Anchukaitis KJ. An efficient forward model of the climate controls on interannual variation in tree-ring width. *Clim Dyn* 2010. doi:10.1007/s00382-010-0945-5.

182. Thompson DM, Ault TR, Evans MN, Cole JE, Emile-Geay J. Comparison of observed and simulated tropical trends using a forward model of coral d18O. *Geophys Res Lett* 2011, 38:L14706. doi:10.1029/2011GL048224.

183. Schmidt GA, LeGrande AN, Hoffmann G. Water isotope expressions of intrinsic and forced variability in a coupled ocean–atmosphere model. *J Geophys Res* 2007, 112:D10103. doi:10.1029/2006JD007781.

184. Fairchild IJ, Tuckwell GW, Baker A, Tooth AF. Modelling of dripwater hydrology and hydrogeochemistry in a weakly karstified aquifer (Bath, UK): implications for climate change studies. *J Hydrol* 2006, 321:213–231.

185. Ault TR, Cole JE, Overpeck JT, Pederson GT, Meko DM. Assessing the risk of persistent drought using climate model simulations and paleoclimate data. *J Clim* 2013, 26:5863–5878. doi:10.1175/JCLI-D-11-00732.1.

186. Franke J, Frank D, Raible CC, Esper J, Brönnimann S. Spectral biases in tree-ring climate proxies. *Nat Clim Change* 2013, 3:360–364.

187. Mann M, Fuentes J, Rutherford S. Underestimation of volcanic cooling in tree-ring-based reconstructions of hemispheric temperatures. *Nat Geosci* 2012, 5:202–205.

188. Anchukaitis K, Breitenmoser P, Briffa K, Buchwal A, Büntgen U, Cook E, D'Arrigo R, Esper J, Evans M, Frank D, et al. Tree rings and volcanic cooling. *Nat Geosci* 2012, 5:836–837. doi:10.1038/ngeo1645.

189. D'Arrigo R, Wilson R, Anchukaitis K. Volcanic cooling signal in tree-ring temperature records for the past millennium. *J Geophys Res-Atmos* 2013, 118:9000–9010. doi:10.1002/jgrd.50692.

190. Esper J, Schneider L, Smerdon J, Schöne B, Büntgen U. Signals and memory in tree-ring width and density data. *Dendrochronologia* 2015, 35:62–70.

191. Stoffel M, Khodri M, Corona C, Guillet S, Poulain V, Bekki S, Guiot J, Luckman B, Oppenheimer C, Lebas N, et al. Estimates of volcanic-induced cooling in the Northern Hemisphere over the past 1,500 years. *Nat Geosci* 2015, 8:784–788. doi:10.1038/ngeo2526.

192. Pretis, F, Schneider L, Smerdon JE, Hendry D. Detecting volcanic eruptions in temperature reconstructions by designed break-indicator saturation, *J Econ Surv*. In press, doi: 10.1111/joes.12148.

193. Stein ML. Editorial. *Ann Appl Stat* 2011, 5:1–4.

194. Li B, Nychka DW, Ammann CM. The value of multiproxy reconstruction of past climate. *J Am Stat Assoc* 2010, 105:883–895. doi:10.1198/jasa.2010.ap09379.

195. Widmann M, Goosse H, van der Schrier G, Schnur R, Barkmeijer J. Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium. *Clim Past* 2010, 6:627–644.

196. Pauling A, Luterbacher J, Wanner H. Evaluation of proxies for European and North Atlantic temperature field reconstructions. *Geophys Res Lett* 2003, 30:1787. doi:10.1029/2003GL017589.

197. Coats S, Smerdon JE, Seager R, Cook BI, González-Rouco JF. Megadroughts in Southwest North America in millennium-length ECHO-G simulations and their comparison to proxy drought reconstructions. *J Clim* 2013, 26:7635–7649. doi:10.1175/JCLI-D-12-00603.1.

198. Coats S, Smerdon JE, Cook BI, Seager R. Are simulated megadroughts in the North American southwest forced? *J Clim* 2015, 28:124–142. doi:10.1175/JCLI-D-14-00071.

199. Coats S, Cook BI, Smerdon JE, Seager R. North American pan-continental droughts in model simulations of the last millennium. *J Clim* 2015, 28:2025–2043.

200. Coats S, Smerdon JE, Seager R, Griffin D, Cook BI. Winter-to-summer precipitation phasing in southwestern North America: a multi-century perspective

from paleoclimatic model-data comparisons. *J Geophys Res Atmos* 2015, 120:8052–8064. doi:10.1002/2015JD023085.

201. Stevenson S, Timmermann A, Chikamoto Y, Langford S, DiNezio P. Stochastically generated North American megadroughts. *J Clim* 2015, 28:1865–1880.

202. Landrum L, Otto-Bliesner BL, Wahl ER, Conley A, Lawrence PJ, Rosenbloom N, Teng H. Last millennium climate and its variability in CCSM4. *J Clim* 2013, 26:1085–1111. doi:10.1175/JCLI-D-11-00326.1.

203. Cook BI, Ault TR, Smerdon JE. Unprecedented 21st-century drought risk in the American Southwest and Central Plains. *Sci Adv* 2015, 1:e1400082.

204. Colose CM, LeGrande AN, Vuille M. The influence of tropical volcanic eruptions on the climate of South America during the last millennium. *Clim Past Discuss* 2015, 11:3375–3424. doi:10.5194/cpd-11-3375-2015.

205. Otto-Bliesner B, Brady E, Fasullo J, Jahn A, Landrum L, Stevenson S, Rosenbloom N, Mai A, Strand G. Climate variability and change since 850 C.E.: an ensemble approach with the Community Earth System Model (CESM). *Bull Am Meteorol Soc*, doi:10.1175/BAMS-D-14-00233.1. In press.

206. PAGES2k-PMIP3 Group. Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium. *Clim Past* 2015, 11:1673–1699. doi:10.5194/cp-11-1673-2015.

207. Smerdon JE, Cook BI, Cook ER, Seager R. Bridging past and future climate across paleoclimatic reconstructions, observations, and models: a hydroclimate case study. *J Clim* 2015, 28:3212–3231.

208. Schmidt GA, Annan JD, Bartlein PJ, Cook BI, Guilyardi E, Hargreaves JC, Harrison SP, Kageyama M, LeGrande AN, Konecky B, et al. Using paleo-climate comparisons to constrain future projections in CMIP5. *Clim Past* 2014, 10:221–250. doi:10.5194/cp-10-221-2014.

209. Deser C, Knutti R, Solomon S, Phillips AS. Communication of the role of natural variability in future North American climate. *Nat Clim Change* 2012, 2:775–779. doi:10.1038/nclimate1562.

210. Deser C, Phillips A, Bourdette V, Teng H. Uncertainty in climate change projections: the role of internal variability. *Clim Dyn* 2012, 38:527–547. doi:10.1007/s00382-010-0977-x.