Clim Dyn (2013) 41:2039–2060
DOI 10.1007/s00382-013-1701-4

Pairwise comparisons to reconstruct mean temperature in the Arctic Atlantic Region over the last 2,000 years

Sami Hanhijärvi · Martin P. Tingley · Atte Korhola

Received: 6 July 2012 / Accepted: 10 February 2013 / Published online: 6 April 2013

Abstract Existing multi-proxy climate reconstruction methods assume the suitably transformed proxy time series are linearly related to the target climate variable, which is likely a simplifying assumption for many proxy records. Furthermore, with a single exception, these methods face problems with varying temporal resolutions of the proxy data. Here we introduce a new reconstruction method that uses the ordering of all pairs of proxy observations within each record to arrive at a consensus time series that best agrees with all proxy records. The resulting unitless composite time series is subsequently calibrated to the instrumental record to provide an estimate of past climate. By considering only pairwise comparisons, this method, which we call PaiCo, facilitates the inclusion of records with differing temporal resolutions, and relaxes the assumption of linearity to the more general assumption of a monotonically increasing relationship between each proxy series and the target climate variable. We apply PaiCo to a newly assembled collection of high-quality proxy data to reconstruct the mean temperature of the Northernmost Atlantic region, which we call Arctic Atlantic, over the last 2,000 years. The Arctic Atlantic is a dynamically important region known to feature substantial temperature variability over recent millennia, and PaiCo allows for a more thorough investigation of the Arctic Atlantic regional climate as we include a diverse array of terrestrial and marine proxies with annual to multidecadal temporal resolutions. Comparisons of the PaiCo reconstruction to recent reconstructions covering larger areas indicate greater climatic variability in the Arctic Atlantic than for the Arctic as a whole. The Arctic Atlantic reconstruction features temperatures during the Roman Warm Period and Medieval Climate Anomaly that are comparable or even warmer than those of the twentieth century, and coldest temperatures in the middle of the nineteenth century, just prior to the onset of the recent warming trend.

Keywords Multiproxy reconstruction · Pairwise comparisons · Non-linear method · North Atlantic · Temperature

1 Introduction

Climate reconstructions aim to elucidate the variability of the climate in the past by inferring climate variables during the era of interest (Bradley et al. 2003). The reconstructed variables help in painting a picture of the past and providing a context for interpreting recent observations and future predictions. To this end, accurate reconstructions are vital, which necessitates that the assumptions made by a particular reconstruction method are appropriate for the included proxy observations.

All existing multi-proxy reconstruction methods assume that the relation between a proxy and the target climate variable is linear (Tingley et al. 2012). The assumption is often reasonable: many proxy records are known to be linearly related to, for example, temperature and thus reconstruction methods that make the assumption perform

Electronic supplementary material The online version of this article (doi:10.1007/s00382-013-1701-4) contains supplementary material, which is available to authorized users.

S. Hanhijärvi (✉) · A. Korhola
Department of Environmental Sciences,
University of Helsinki, Helsinki, Finland
e-mail: sami.hanhijarvi@helsinki.fi

M. P. Tingley
Department of Earth and Planetary Sciences,
Harvard University, Cambridge, MA, USA

well in many cases. The linearity assumption is also convenient: much is known about linear methods (Bingham and Fry 2010) and it is simple to test if the assumption holds even approximately. However, it has been argued that some proxies, while appearing to be linear during the calibration period, exhibit non-linear behavior outside of the range of values in the calibration period (Anchukaitis et al. 2012; Mann et al. 2012a, b). There are also cases where linear correlation is weak (Ojala and Alenius 2005) or cannot be directly tested (Tiljander et al. 2003) even if the proxy is known to be strongly influenced by the target climate variable.

Linear methods can be used with non-linear proxies if it is possible to invert the known non-linear relation and transform the original record to a new record that is linearly related to the target climate variable. Non-linear methods have been developed for specific proxy types to reconstruct the target climate variable (Evans et al. 2006; McKay et al. 2008; Pflaumann et al. 1996; Tolwinski-Ward et al. 2011), effectively making the mentioned inverse transformation. However, if proxies are used from multiple sources, finding suitable inverse transformations for each may be difficult or even impossible. For example, hydrological proxies (Maidment 1993) often display non-linear or even thresholding type relations with precipitation, as runoff and sediment deposition only occur for sufficiently high streamflow, which is in turn a non-linear function of precipitation. Proxies that exhibit such threshholding behavior cannot be completely inverse transformed to linearity, since the exact value of a sample above (or below) the threshold is unknown. However, such samples still carry information since their values are known to be higher (or lower) than any other value that is below (or above) the threshold. If the linearity of a record cannot be tested, assumed or obtained by a transformation, the record would generally be excluded from a multi-proxy reconstruction with current reconstruction methods. Furthermore, even if a proxy seems linear and is assumed as such, it may exhibit a non-linear response to values of the climate variable outside of those contained within the calibration interval.

Another common assumption for all but one (the exception is Li et al. 2010) of the existing multi-proxy reconstruction methods is that they assume the same temporal resolution for all proxy records, while in practice different proxy records generally have different temporal resolutions. For example, tree ring records are annually resolved while lake and marine sediments often have a much lower resolution. Some reconstructions have used interpolation to achieve a common resolution (Christiansen and Ljungqvist 2011; Mann et al. 2008; Moberg et al. 2005; Sundqvist et al. 2010). However, interpolation artificially increases the influence of the interpolated records, changes their interpretation, invalidates the assumptions made by the reconstruction methods, and should thus be avoided. Another approach is to only include records that have the same resolution, but then many high-quality records are excluded.

We present here a novel Composite-Plus-Scale (CPS) method, PaiCo, so called as it relies on Pairwise Comparisons. PaiCo is based on comparing all pairs of sample values within each proxy record and producing a time series that best matches all of the pairwise comparisons. The reconstructed time series is unitless, so must subsequently scaled to match the units of the target climate variable, and represents the spatial average of the target climate variable over the geographical area the proxy records cover.

The pairwise comparisons framework allows us to assume that the transfer functions between proxy records and the climate variable are monotonically increasing, which is a weaker assumption than assuming they are linear. Specifically, any linear transfer function is also monotonic, but the opposite does not hold. PaiCo has other favorable properties in that proxy records need not share common temporal resolution (Li et al. 2010); standardization of records is not required (Tingley 2012); and missing values are handled intrinsically. While PaiCo discards information about the magnitudes of the proxy values, it also allows more information to enter into a reconstruction as a greater variety of proxy records can be included. Furthermore, we show in a quantitative analysis that PaiCo performs at par with the best linear methods in many linear cases and supersedes all existing methods in non-linear cases.

We apply PaiCo to a multi proxy data set from the northernmost region of the Atlantic. The study area is bounded to 60°–90°N and 50°W–30°E, and we call this area the Arctic Atlantic. Studies of North Atlantic climate variability have become a central focal point of climate research during recent decades. North Atlantic climate variability arises from diverse sources over broad spatial and temporal scales; several natural processes are working in parallel, sometimes enhancing and sometimes counteracting each other (Marshall et al. 2001). Climate at high northern latitudes is greatly affected by atmospheric pressure patterns, while oceanic processes of the North Atlantic also affect the regional as well as the global climate. Furthermore, a number of papers have argued that the capacity of sea ice to affect climate both through albedo and air-sea heat exchange and also the ability of sea ice to rapidly change its distribution make this a good candidate mechanism for driving abrupt climate changes in the North Atlantic and perhaps worldwide (Denton et al. 2005; Gildor and Tziperman 2003; Li et al. 2005; Kaspi et al. 2004; Timmermann et al. 2005; Wunsch 2006).

During the past few millennia, climate development in the North Atlantic sector of the Arctic was punctuated by centennial-scale warmer and colder episodes, of which the most well-known are the Little Ice Age (LIA about

AD1500–1900), the Medieval Climate Anomaly (MCA about AD900–1200), the Dark Ages Cold Period (DACP about BC100 to AD700) and the Roman Warm Period (RWP about BC900–100; Bradley et al. 2003; Korhola et al. 2000; Lamb 1995; McDermott et al. 2001; Moberg et al. 2005; O'Brien et al. 1995; Wang et al. 2005). Although there has been much discussion about the magnitude and geographical extent of these events, it nevertheless seems that the most consistent records of these classic climate periods come just from the North Atlantic sector of the Arctic (Bradley et al. 2003; Miller et al. 2010). The climate of the Arctic Atlantic is highly variable, but current reconstructions do not agree on the relative magnitudes of different cold and warm spells over the last two millennia. It is therefore a region well suited to further analysis using PaiCo, as it allows for the inclusion of a broader array of proxy types and thus provides a more complete description of climate variability in this region. The analysis using PaiCo focuses on the regional mean temperature, and we defer to future work the study of spatial patterns of the variability within the Arctic Atlantic.

We present the PaiCo method and associated theoretical properties in Sect. 2. The qualitative and quantitative properties of PaiCo are compared to existing methods in Sect. 3. Section 4 introduces the proxy data from the Arctic Atlantic and Sect. 5 presents the results of applying PaiCo to this multiproxy data set. Section 6 closes the paper with final discussion of both the methodology and our applied results.

2 Method

2.1 Model

Let $\mathbf{f}$ be a column vector that represents the time series of the target climate variable to be reconstructed. The values in $\mathbf{f}$ will represent the average of the climate variable over the geographical area the proxy records cover with equal weight for all proxy records. Hence, the reconstruction will not have a spatial component. For simplicity, we will focus on reconstructing the annual average temperature, but any other climate variable could be reconstructed as long as the proxy records are causally affected by that climate variable.

Let $\mathbf{P}^k$ be the vector of random variables corresponding to the kth proxy record. Each proxy record is a single time series, such as the $d^{18}O$ measurement from an ice core or a tree ring width chronology. We model the relation between the target climate variable and the proxy record values by

$$\mathbf{P}_i^k = g^k(\mathbf{X}_{i\cdot}^k(\mathbf{f} + \mathbf{e}^k)), \tag{1}$$

where $\mathbf{P}_i^k$ represents the random variable corresponding to the ith sample value of proxy record k, $\mathbf{e}$ is a vector of noise terms, and g^k is the transfer function for proxy k. We use the convention that the subscript indexes a variable and $\cdot$ denotes all elements of the corresponding index. For example, $\mathbf{X}_{ij}$ is the value of the matrix $\mathbf{X}$ at row i and column j. The row vector $\mathbf{X}_{i\cdot}^k$ represents the linear combination of the noisy terms $\mathbf{f} + \mathbf{e}$ that the proxy sample i supports, i.e., $\mathbf{X}_{i\cdot}^k$ accounts for the averaging behavior of the proxy record. The selection and averaging matrices $\mathbf{X}^k$ are discussed in Sect. 2.3. The noise terms $\mathbf{e}^k$ are assumed to follow a mean-zero multivariate normal distribution with covariance matrix Σ. While other noise structures are possible, multivariate normal variables result in a tractable maximization problem (see Sect. 2.4), and are likely a reasonable assumption for most proxies.

The transfer function g^k models the unit transformation from the linearly combined noisy climate variable to the proxy record. Existing reconstruction methods assume that the transfer function g^k is linear, i.e., $g^k(x) = \alpha^k(x) + \beta^k(x)$ for some $\alpha^k, \beta^k \in \mathbb{R}$, with appropriate scaling for the noise covariance structure Σ, and many additionally assuming $\alpha^k > 0$. For example, the linear model $\mathbf{P}_i^k = \alpha^k \mathbf{f}_i + \beta^k + \mathbf{e}_i$ can be expressed in Eq. (1) with $g^k(x) = \alpha^k(x) + \beta^k(x)$ and noise covariance Σ/α^k.

The major assumption underpinning PaiCo is that there is a monotonically increasing function relating the proxies to the temporally averaged climate variables. That is, for any x and y, it holds that $x < y \Leftrightarrow g^k(x) < g^k(y)$. Assuming only monotonically increasing transfer functions corresponds to saying that an increase in the value of a proxy record corresponds to an increase in the climate variable, but the amount of change in the climate variable is unknown. If the transfer function was assumed linear, a unit increase in the proxy record would necessarily correspond to an increase of α^k in the climate variable. Therefore, we do not assume as much about the relation between a proxy record and the target climate variable as existing methods do, but still assume the proxies are informative about the underlying climate variable. PaiCo thus allows for the inclusion of proxies that are linear in the climate over the calibration interval, but that may feature a different response to climate values that are outside of the range of the calibration interval values.

Figure 1 illustrates examples of both linear and monotonically increasing transfer functions. A linear transfer function is a straight line while a monotonically increasing transfer function is any function that increases when the input value increases. All linear transfer functions are also monotonically increasing assuming the coefficient α^k is positive. Note that if the transfer function is monotonically decreasing, the proxy record should be multiplied by -1, i.e., flipped upside down. In the remainder of the paper, we

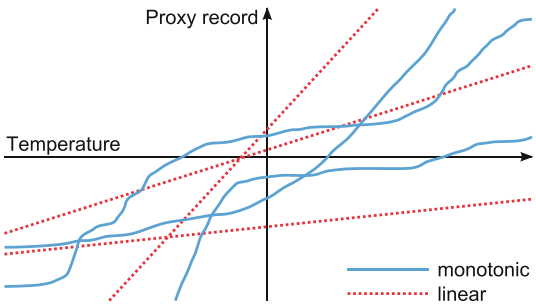

Fig. 1 Examples of linear (*red*) and monotonic (*blue*) functions. Note that the linear functions are also monotonic, but not vice versa

use only the word monotonic when we mean monotonically increasing, unless otherwise stated.

Let $\mathbf{p}_i^k$ be the realization of $\mathbf{P}_i^k$ in Eq. (1), i.e., $\mathbf{p}_i^k$ is the ith observed value of the proxy record k. Instead of using the actual values of the realizations $\mathbf{p}_i^k$, we only compare which is greater of $\mathbf{p}_i^k$ and $\mathbf{p}_j^k$ for all distinct pairs of i and j. These pairwise comparisons disregard the magnitudes of the values in a proxy record by only considering their pairwise orderings. The effect is that we can ignore a monotonic transfer function g^k, since

$$\begin{aligned} &\mathbf{P}_i^k < \mathbf{P}_j^k \\ &\Leftrightarrow g^k(\mathbf{X}_{i\cdot}^k(\mathbf{f}+\mathbf{e}^k)) < g^k(\mathbf{X}_{j\cdot}^k(\mathbf{f}+\mathbf{e}^k)) \\ &\Leftrightarrow \mathbf{X}_{i\cdot}^k(\mathbf{f}+\mathbf{e}^k) < \mathbf{X}_{j\cdot}^k(\mathbf{f}+\mathbf{e}^k). \end{aligned} \tag{2}$$

Note that comparisons are made only within a proxy record, and not between proxies, thus allowing heterogeneous proxies with different forward models g^k and temporal properties $\mathbf{X}^k$ to be combined. The error term $\mathbf{e}^k$ makes this comparison a Bernoulli random variable, with the difference $\mathbf{X}_{i\cdot}^k\mathbf{f} - \mathbf{X}_{j\cdot}^k\mathbf{f}$ determining the probability of $\mathbf{P}_i^k < \mathbf{P}_j^k$ and, likewise, the probability of $\mathbf{P}_i^k > \mathbf{P}_j^k$.

Using the model of Eq. (1) in Eq. (2) allows us to disregard the details of how the climate variable is transferred to a proxy, as long as we can assume the transfer function is both stationary and monotonically increasing. If a proxy truly tracks the changes of the climate variable, increases and decreases in the variable are reflected in similar behavior in the proxy. This allows us to use proxies that are known to be strongly related to the target climate variable but for which linearity can not be assumed or tested. There are trade-offs in making the less restrictive assumption of monotonicity, which we explore in Sect. 3.

Stationarity, in some shape or form, is also a necessary requirement, as otherwise we could not infer anything about the past. Stationarity in this model means that the relation between the climate variable and proxy remains unchanged throughout time.

Another effect of pairwise comparisons is that we do not require the proxies to be standardized before use as any scaling with a positive coefficient and any shifting can be considered to be contained in the transfer function g^k. Therefore, such operations make no difference with respect to the model.

Thurstone (1927) was the first to describe a method based on pairwise comparisons for applications in psychonometrics, and more recent developments can be found in Stern (1990).

2.2 Intuition behind pairwise comparisons

We develop the intuition behind pairwise comparisons by first considering a special case of the general formalism of Eq. (2). Let each proxy record and the target $\mathbf{f}$ have the same temporal structure and let the transfer functions be the identity, i.e., $g^k(x) = x$ for all k. In this case, each value of a proxy record is simply a noisy version of the corresponding target value. For a single pair of time points, the agreement between the proxy records over the ordering of the two corresponding observations within each proxy series is directly related to the difference between the target values. The more proxy time series feature a higher value for the first time point, the greater the value of the target at the first time point as compared to the second. Similarly, if the proxy series are about equally split between featuring the larger value at the first and second time points, then there is likely not much difference between the corresponding target values. Assuming Gaussian noise, we can use the amount of proxy agreement to calculate the relative difference between the pair of target values.

We now develop the mathematical formalism for the simple case, with g the identity and the proxies and climate having the same resolution. The only information we use is the collection of pairwise comparisons $\mathbf{p}_i^k < \mathbf{p}_j^k$ for all distinct pairs (i, j) of every proxy record k. In the simplistic setting above, $\mathbf{P}_i^k = \mathbf{P}_i = \mathbf{f}_i + \mathbf{e}_i$. As the values $\mathbf{p}_i^k$ are realizations of $\mathbf{P}_i^k$, the pairwise comparisons $\mathbf{p}_i^k < \mathbf{p}_j^k$ are realizations of $\mathbf{P}_i^k < \mathbf{P}_j^k$, and in this simple case, they are also realizations of $\mathbf{P}_i < \mathbf{P}_j$. Because of this θ, the probability that $\mathbf{P}_i < \mathbf{P}_j$, can be estimated with $\hat{\theta}_{ij} = \frac{1}{M}\sum_{k=1}^{M} \delta(\mathbf{p}_i^k < \mathbf{p}_j^k)$, where $\delta(\cdot)$ is the indicator function that returns 1 when the inner test is true and 0 otherwise. The value $\hat{\theta}_{ij}$ is the fraction of proxy records for which the proxy value i was less than the proxy value j, i.e., it is a measure of agreement between proxies. As $Pr(\mathbf{P}_i < \mathbf{P}_j) = Pr(\mathbf{f}_i + \mathbf{e}_i < \mathbf{f}_j + \mathbf{e}_j)$, the value $\hat{\theta}_{ij}$ is also an estimate of the probability $\theta = Pr(\mathbf{f}_i + \mathbf{e}_i < \mathbf{f}_j + \mathbf{e}_j)$. We have already assumed $\mathbf{e}$ is multivariate normal with zero mean and covariance Σ. Let us further assume the covariance matrix is diagonal, $\Sigma = \sigma^2\mathbf{I}$. Then,

$$\begin{aligned}\theta_{ij} &= Pr(\mathbf{f}_i + \mathbf{e}_i < \mathbf{f}_j + \mathbf{e}_j)\\ &= Pr(\mathbf{e}_i - \mathbf{e}_j < \mathbf{f}_j - \mathbf{f}_i)\\ &= \Phi(\mathbf{f}_j - \mathbf{f}_i, 0, 2\sigma^2)\end{aligned}$$

where $\Phi(\cdot, \mu, \sigma^2)$ is the Gaussian cumulative distribution function with mean μ and variance σ^2. Inverting the last equality gives a relative value for the difference $\mathbf{f}_j - \mathbf{f}_i$,

$$\frac{1}{\sqrt{2\sigma^2}}(\mathbf{f}_j - \mathbf{f}_i) = \Phi^{-1}(\theta_{ij}, 0, 1).$$

Therefore, the value of θ_{ij} is a measure of the difference between $\mathbf{f}_j$ and $\mathbf{f}_i$. The larger the difference $\mathbf{f}_j - \mathbf{f}_i$, the closer θ_{ij} is to 1 and the more we expect to see $\mathbf{p}_i^k < \mathbf{p}_j^k$. If the difference is small, $\hat{\theta}_{ij} \approx .5$ and the fraction of $\mathbf{p}_i^k < \mathbf{p}_j^k$ is roughly equal to the fraction of $\mathbf{p}_j^k < \mathbf{p}_i^k$. In other words, the fraction of proxies with $\mathbf{p}_i^k < \mathbf{p}_j^k$ gives information about the difference $\mathbf{f}_j - \mathbf{f}_i$. This result is the main idea underpinning PaiCo.

The intuition behind using pairwise comparisons on the proxy scale to infer difference in the target climate is illustrated in Fig. 2. The more general case, presented in Sect. 2.4, is more involved due to the averaging of the climate inherent to many proxies that is modeled using the matrices $\mathbf{X}^k$.

2.3 Temporal resolution and change of support

Proxy records often have different temporal characteristics. Ice core records are often annually resolved, i.e., there is a sample for every year, and have annual support, i.e., only the climate of the respective year affects the corresponding proxy observation. Lake sediment records, on the other hand, are often variably resolved since the time between samples can be anything from a few years to even several centuries, while the support of each sample often spans several years. Furthermore, quite often the target $\mathbf{f}$ is expected to be annually resolved with an annual support. This problem of inferring values for $\mathbf{f}$ from data that have different temporal characteristics is an example of a *change-of-support problem* (Tingley et al. 2012).

The linear combinations $\mathbf{X}_{i\cdot}^k$ indicate the values of $\mathbf{f}$ that affect the value of sample i of proxy k. The vectors $\mathbf{X}_{i\cdot}^k$ are assumed known with the requirement that the elements have to sum to 1, $\mathbf{X}_{i\cdot}^k[1, \cdots, 1]^\top = 1$. To give an example of $\mathbf{X}_{i\cdot}^k$, let us assume $\mathbf{f}$ is to be annually resolved with annual support. Consider a proxy record, $k = 1$, with annual resolution and annual support. Each vector $\mathbf{X}_{i\cdot}^1$ then consist of zeros, with a single entry of 1 such that $\mathbf{X}_{i\cdot}^1(\mathbf{f} + \mathbf{e}^1)$picks out the noisy value $\mathbf{f}_l + \mathbf{e}_l^1$ for the year l corresponding to the proxy observation. Consider next another proxy record, $k = 2$, with biannual support and 5-year resolution. Then

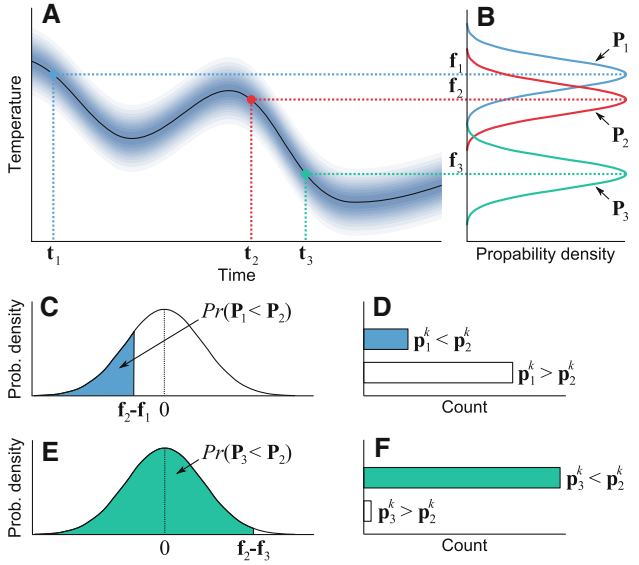

Fig. 2 Illustration of pairwise comparisons, where proxy values are simply noisy versions of the target, i.e., proxies have the same temporal structure as the target and the transfer functions are $g^k(x) = x$ for all k so that $\mathbf{P}_i^k = \mathbf{P}_i = \mathbf{f}_i + \mathbf{e}_i$. **A** Probability density of the target plus noise through time, with *darker colors* indicating regions of higher probability. The *black line* is the target without noise. The time points $\mathbf{t}_i$ correspond to the values target values $\mathbf{f}_i$. **B** Probability densities for proxy values, represented by random variables $\mathbf{P}_1 = \mathbf{f}_1 + \mathbf{e}_1$, $\mathbf{P}_2 = \mathbf{f}_2 + \mathbf{e}_2$ and $\mathbf{P}_3 = \mathbf{f}_3 + \mathbf{e}_3$. Notice that the target values determine how much the densities of the proxy values overlap. For a pair of time points, the overlap causes disagreement when comparing observed proxy values, and the amount of disagreement over all proxies is directly related to the distance of the target values. **C** Probability of $\mathbf{P}_1$ being smaller than $\mathbf{P}_2$. Since $\mathbf{f}_2$ and $\mathbf{f}_1$ are close, the probability $Pr(\mathbf{P}_1 < \mathbf{P}_2)$ is close to .5, which means the disagreement among proxies will be large. **D** The observed proxy values $\mathbf{p}_i^k$ are realizations of $\mathbf{P}_i^k = \mathbf{P}_1$, and therefore, the pairwise comparisons $\mathbf{p}_1^k < \mathbf{p}_2^k$ are realizations of comparisons $\mathbf{P}_1 < \mathbf{P}_2 \Leftrightarrow \mathbf{f}_1 + \mathbf{e}_1 < \mathbf{f}_2 + \mathbf{e}_2$. The fraction of $\mathbf{p}_1^k < \mathbf{p}_2^k$ among M proxy records gives an estimate of $Pr(\mathbf{P}_1 < \mathbf{P}_2) = Pr(\mathbf{f}_1 + \mathbf{e}_1 < \mathbf{f}_2 + \mathbf{e}_2)$. **E** The probability of $\mathbf{P}_3$ being smaller than $\mathbf{P}_2$. Now that $\mathbf{f}_3$ and $\mathbf{f}_2$ are far apart, the probability $Pr(\mathbf{P}_3 < \mathbf{P}_2)$ is close to 1 and the proxies will mostly agree on the pairwise order of the values. **F** Similarly to **D**, fraction of $\mathbf{p}_3^k < \mathbf{p}_2^k$ gives an estimate of $Pr(\mathbf{P}_3 < \mathbf{P}_2) = Pr(\mathbf{f}_3 + \mathbf{e}_3 < \mathbf{f}_2 + \mathbf{e}_2)$

$\mathbf{X}_{1\cdot}^2 = [\frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0, 0, 0 \cdots, 0]$ and $\mathbf{X}_{2\cdot}^2 = [0, 0, 0, 0, 0, \frac{1}{2}, \frac{1}{2}, 0, \cdots, 0]$, and hence, $\mathbf{X}_{1\cdot}^2(\mathbf{f} + \mathbf{e}^2) = \frac{1}{2}(\mathbf{f}_1 + \mathbf{e}_1^2) + \frac{1}{2}(\mathbf{f}_2 + \mathbf{e}_2^2)$ and $\mathbf{X}_{2\cdot}^2(\mathbf{f} + \mathbf{e}^2) = \frac{1}{2}(\mathbf{f}_5 + \mathbf{e}_5^2) + \frac{1}{2}(\mathbf{f}_6 + \mathbf{e}_6^2)$. In other words, the term $\mathbf{X}_{i\cdot}^k(\mathbf{f} + \mathbf{e}^k)$ in Eq. (1) calculates the weighted average of the noisy values $\mathbf{f} + \mathbf{e}$ over the time (e.g., years) the sample i of proxy k supports. Resolution and support are both encoded in $\mathbf{X}_{i\cdot}^k$ with, roughly speaking, sample time determining the position of the non-zero values in $\mathbf{X}_{i\cdot}^k$ and support determining how many consecutive values there are. These linear combinations are similar to those discussed by Li et al. (2010) and Tingley et al. (2012).

There are some special cases we need to consider. Some proxy records, including many tree ring chronologies, may only have seasonal support, while the target climate variable is often required to be annually resolved with annual support. In such case, the model in Eq. (1) is incorrect, since $\mathbf{X}_{i\cdot}^k(\mathbf{f}+\mathbf{e}^k)$, for any definition of $\mathbf{X}_{i\cdot}^k$, is an annual average over years while $\mathbf{p}_i^k$ only reflects the seasonal climate. To fix this, there are several possibilities. One option is to make the target $\mathbf{f}$ have the same minimum support as the proxy records. For example, $\mathbf{f}$ could have seasonal or monthly resolution and support, and $\mathbf{X}_{i\cdot}^k$ can then be defined as before. The second solution is to only include records that have the same seasonal support, and then inferring the target at this common same seasonal support. This option has been used before in climate reconstructions that target a specific season (e.g., Kaufman et al. 2009). A final option is to ignore seasonality and assume that all proxy observations have at least annual support. In other words, the growth season support of a tree ring chronology is ignored and it is assumed to be sensitive to annual climate variations (e.g., Ljungqvist 2010; Moberg et al. 2005).

Another special case to consider is the averaging behavior of some proxy records. For example, let us consider using an ice core record of 5-year averages of $d^{18}O$ to reconstruct annually resolved and supported temperature. In this case, the model

$$\mathbf{P}_i^k = \mathbf{X}_{i\cdot}^k g^k(\mathbf{f}+\mathbf{e}^k) \tag{3}$$

might be more reasonable than the one in Eq. (1), since $\mathbf{P}_i^k$ now expresses, with a suitable $\mathbf{X}_{i\cdot}^k$, the 5-year average of the transformed noisy temperature. The term $g^k(\mathbf{f}+\mathbf{e}^k)$ applies the function g^k separately to each element of the input vector. With Eq. (3), the transfer function is implicitly assumed linear, since only if g^k is linear does it hold that

$$\begin{aligned} &\mathbf{X}_{i\cdot}^k g^k(\mathbf{f}+\mathbf{e}^k) < \mathbf{X}_{j\cdot}^k g^k(\mathbf{f}+\mathbf{e}^k) \\ &g^k(\mathbf{X}_{i\cdot}^k(\mathbf{f}+\mathbf{e}^k)) < g^k(\mathbf{X}_{j\cdot}^k(\mathbf{f}+\mathbf{e}^k)) \\ &\mathbf{X}_{i\cdot}^k(\mathbf{f}+\mathbf{e}^k) < \mathbf{X}_{j\cdot}^k(\mathbf{f}+\mathbf{e}^k), \end{aligned} \tag{4}$$

and PaiCo requires the inequalities in Eq. (4) to hold. This does not violate the assumptions of the method, but g^k is assumed linear instead of monotonic for this proxy.

Finally, we discuss pollen records, but the discussion applies to other sediment proxies, such as chironomids. First note that we will not use the observed abundances of such a proxy but the published reconstruction calculated from the abundances. For example, we will later use the reconstructed temperature from Luoto et al. (2009) and not the observed chironomid abundances.

Each year, pollen is deposited in the sediment and these annual depositions of pollen reflect the climate over a span of several preceding years. If we were able to sample a sediment core precisely at annual scale, we would have an annually resolved proxy and each sample would support several years. Let $\mathbf{Z}_{j\cdot}^k$ express the support of each year of deposition, such that $\mathbf{Z}_{j\cdot}^k(\mathbf{f}+\mathbf{e}^k)$ expresses the noisy climate averaged over the years the single year j of deposition supports. However, sediment cores can often only be sampled at a much coarser scale, and each sample is an average over the deposition included in the sample. An appropriate model for a sediment proxy value would thus be

$$\mathbf{P}_i^k = \mathbf{Y}_{i\cdot}^k g^k(\mathbf{Z}^k(\mathbf{f}+\mathbf{e}^k)), \tag{5}$$

where $\mathbf{Y}_{i\cdot}^k$ expresses how the deposition of each year was averaged to produce the sample i of proxy k. Because of the averaging $\mathbf{Y}_{i\cdot}^k$, the transfer function g^k is again implicitly assumed linear, since only then do the inequalities in Eq. (4) hold. Fortunately, many of the published pollen reconstructions are already in degrees Celsius, which means g^k is linear by definition. With linear g^k, the model in Eq. (5) is equivalent to

$$\mathbf{P}_i^k = g^k(\mathbf{Y}_{i\cdot}^k \mathbf{Z}^k(\mathbf{f}+\mathbf{e}^k))$$

with respect to the inequalities Eq. (4). To use such a record, both $\mathbf{Y}_{i\cdot}^k$ and $\mathbf{Z}^k$ need to be determined to find out the linear combinations $\mathbf{X}_{i\cdot}^k = \mathbf{Y}_{i\cdot}^k \mathbf{Z}^k$. In reality, the averaging linear combinations $\mathbf{Y}_{i\cdot}^k$ can be derived from the sample limits, which in turn can be calculated fairly easily as shown in Sect. 2.8. However, determining the support matrices $\mathbf{Z}^k$ is outside of the scope of this work and, therefore, we will assume them to be the identity matrices in the analysis of Arctic Atlantic records.

2.4 Likelihood of pairwise comparisons

We now develop the mathematical formalism in the general case of Eq. (2), and describe our approach to maximizing the likelihood. Using the model in Eq. (1), we define the likelihood by comparing all pairs of observed proxy values $\mathbf{p}_i^k$ for each proxy record:

$$\begin{aligned} &Pr(\{\mathbf{p}_i^k\}|\mathbf{f}) \\ &= \prod_{k=1}^{M} \prod_{i,j:\mathbf{p}_i^k < \mathbf{p}_j^k} Pr(\mathbf{P}_i^k < \mathbf{P}_j^k|\mathbf{f}) \\ &= \prod_{k=1}^{M} \prod_{i,j:\mathbf{p}_i^k < \mathbf{p}_j^k} Pr(\mathbf{X}_{i\cdot}^k(\mathbf{f}+\mathbf{e}^k) < \mathbf{X}_{j\cdot}^k(\mathbf{f}+\mathbf{e}^k)) \\ &= \prod_{k=1}^{M} \prod_{i,j:\mathbf{p}_i^k < \mathbf{p}_j^k} \Phi((\mathbf{X}_{j\cdot}^k - \mathbf{X}_{i\cdot}^k)\mathbf{f}, 0, \sigma_{ji}^2), \end{aligned} \tag{6}$$

where M is the number proxy records, $\Phi(x, 0, \sigma^2)$ is the value of the cumulative distribution function of a Gaussian

distribution with 0 mean and σ^2 variance, and $\sigma_{ji}^2 = (\mathbf{X}_{j\cdot}^k - \mathbf{X}_{i\cdot}^k)\Sigma(\mathbf{X}_{j\cdot}^k - \mathbf{X}_{i\cdot}^k)^\top$, where Σ is the covariance matrix of the noise terms. Note that Σ is common to all proxies, which means that the noise variance is equal for each proxy. Therefore, we assume that each proxy, prior to lossy averaging and one-to-one transform of g^k, is equally informative of the climate. We explore the effects of departures from this assumption in Sect. 3.2.

The likelihood in Eq. (6) is a variant of Thurstonian scaling (Lipovetsky and Conklin 2004). We present here the iteration formulas to find the maximum likelihood solution to Eq. (6). First, let us simplify the likelihood by defining a matrix $\mathbf{A}$, which stacks the vectors $((\mathbf{X}_{j\cdot}^k - \mathbf{X}_{i\cdot}^k)\Sigma(\mathbf{X}_{j\cdot}^k - \mathbf{X}_{i\cdot}^k)^\top)^{-\frac{1}{2}}(\mathbf{X}_{j\cdot}^k - \mathbf{X}_{i\cdot}^k)$ for all k and $\{(i,j) \mid \mathbf{p}_i^k < \mathbf{p}_j^k\}$ in to a single matrix with height $\sum_{k=1}^M |\{(i,j) \mid \mathbf{p}_i^k < \mathbf{p}_j^k\}|$ and width $|\mathbf{f}|$. Notice that the multiplier before the last $(\mathbf{X}_{j\cdot}^k - \mathbf{X}_{i\cdot}^k)$ is a scalar. We also add a regularization term for $\mathbf{f}$, since the columns of $\mathbf{A}$ are linearly dependent. We regularize the calculations with the multivariate Gaussian distribution with zero mean and covariance matrix Σ_0. The objective function to maximize can then be expressed as

$$F(\mathbf{f}) = \prod_s \Phi(\mathbf{A}_{s\cdot}\mathbf{f}, 0, 1)\phi(\mathbf{f}, 0, \Sigma_0), \tag{7}$$

where s is the row index of $\mathbf{A}$ that runs through all of the rows of $\mathbf{A}$.

The idea in the optimization is to start from a close enough solution and use Newton–Raphson method (Bonnans 2006) to find the location where the gradient of the log-objective is zero on all dimensions. The iteration from solution $\mathbf{f}^a$ to $\mathbf{f}^{a+1}$ is carried out by solving the linear system

$$\mathbf{H}_F(\mathbf{f}^a)(\mathbf{f}^{a+1} - \mathbf{f}^a) = -\nabla_{\mathbf{f}} F(\mathbf{f}^a),$$

where $\mathbf{H}_F(\mathbf{f})$ is the Hessian matrix of F. The gradient of F has the form

$$\nabla_f F(\mathbf{f}^a) = \sum_s \frac{\phi(\mathbf{A}_{s\cdot}\mathbf{f}^a)}{\Phi(\mathbf{A}_{s\cdot}\mathbf{f}^a)}\mathbf{A}_{s\cdot} - \Sigma_0^{-1}\mathbf{f}^a$$

and the Hessian is

$$\mathbf{H}_F(\mathbf{f}^a) = -\sum_s \left(\frac{\phi(\mathbf{A}_{s\cdot}\mathbf{f}^a)}{\Phi(\mathbf{A}_{s\cdot}\mathbf{f}^a)}\mathbf{A}_{s\cdot}\mathbf{f}^a + \frac{\phi(\mathbf{A}_{s\cdot}\mathbf{f}^a)^2}{\Phi(\mathbf{A}_{s\cdot}\mathbf{f}^a)^2}\right)\mathbf{A}_{s\cdot}^\top\mathbf{A}_{s\cdot} - \Sigma_0^{-1},$$

where $\phi(x)$ and $\Phi(x)$ are the probability density and cumulative distribution function, respectively, of a Gaussian random variable with zero mean and unit variance. However, if the Hessian matrix $\mathbf{H}_F(\mathbf{f}^a)$ is singular and the regularization Σ_0 is weak, the iteration may fail to find a reasonable solution. To solve this, we use the Levenberg–Marguardt-style damping (Marquardt 1963)

$$(\mathbf{H}_F(\mathbf{f}^a) + \lambda\mathrm{diag}((\mathbf{H}_F(\mathbf{f}^a)))(\mathbf{f}^{a+1} - \mathbf{f}^a) = -\nabla_{\mathbf{f}} F(\mathbf{f}^a),$$

where $\mathrm{diag}(\cdot)$ is the diagonal of the given matrix with other elements set to 0, and λ is the damping factor. With small λ, the iteration is close to the Newton–Raphson method, and with large λ, the method becomes close to the gradient descent method.

As initial step, we use the heuristic starting point $\mathbf{f}^0 = \sum_s \mathbf{A}_{s\cdot}^\top$, which produces a time series roughly similar to the final result. We carry out the iteration until the change in $F(\mathbf{f}^a)$ between successive iterations is sufficiently small or a maximum number of iterations is achieved. An implementation of the method is available in the Electronic Supplementary Material.

When the iteration has converged, the vector $\mathbf{f}^a$ contains the maximum likelihood estimate $\hat{\mathbf{f}}_{ml}$ of $\mathbf{f}$. The estimate $\hat{\mathbf{f}}_{ml}$ does not have units and needs to be calibrated to instrumental data. Any of the calibration methods used by other CPS methods could be used. Even the more involved Bayesian methods (Li et al. 2010; Tingley and Huybers 2010a) could be used with $\hat{\mathbf{f}}_{ml}$ as the single proxy record. For simplicity, we will use the method-of-moments for calibration, where the estimate $\hat{\mathbf{f}}_{ml}$ is shifted and scaled such that it will have the same mean and variance as the instrumental data over the calibration period. The calibrated time series is later denoted as $\hat{\mathbf{f}}_{cal}$. Note that the estimate $\hat{\mathbf{f}}_{ml}$ is noisy since the number of proxy records is limited. The noise in instrumental data together with the noisy estimate creates the regression dilution problem (Ammann et al. 2010) as faced by other reconstruction methods. We chose to calibrate with the method-of-moments instead of regressing $\hat{\mathbf{f}}_{ml}$ on the instrumental data, since the former is slightly better at dealing with the regression dilution problem even though the latter is, in theory, optimal with respect to prediction error. Assuming i.i.d noise in the instrumental data, the dilution problem could be alleviated for regression by smoothing both $\hat{\mathbf{f}}_{ml}$ and the instrumental data, but it would have been difficult to choose the type and amount of smoothing without resorting to heuristics. Hence, we chose method-of-moments.

2.5 Estimating noise variance

The choice of distribution for noise $\mathbf{e}^k$ in Eq. (1) is due to convenience: the linear combination of multivariate Gaussian variables is also a Gaussian variable. If other distributions are used, the iteration formulas for the maximum likelihood solution need to be updated accordingly.

We have defined the mean of the noise to be zero in Eq. (6). This requirement is not necessary, as any shift of all elements of $\mathbf{f}$ with a common scalar would not affect the likelihood function, and therefore, we chose to use 0 for simplicity. However, the covariance structure Σ in Eq. (6) can be defined freely. One could, for instance, make the noise terms independent with a common variance, which would correspond to a diagonal covariance matrix with a constant value on the diagonal. Furthermore, auto-correlated noise can also be represented in the covariance matrix, but the amount of auto-correlation has to be defined by the user.

The sample variance of the maximum likelihood estimate $\hat{\mathbf{f}}_{ml}$ is related to the sample variance of the target values $\mathbf{f}$ and the magnitude of noise, Σ. Let us decompose the noise covariance matrix to noise variance $\sigma_{\mathbf{e}}^2$ and correlation $\mathbf{R}$, i.e., $\Sigma = \sigma_{\mathbf{e}}^2\mathbf{R}$. The CDFs in Eq. (6) can then be written as

$$\begin{aligned}&\Phi((\mathbf{X}_{j\cdot}^k - \mathbf{X}_{i\cdot}^k)\mathbf{f}, 0, \sigma_{ji}^2)\\ &= \Phi\left(((\mathbf{X}_{j\cdot}^k - \mathbf{X}_{i\cdot}^k)\mathbf{R}(_{j\cdot}^k - \mathbf{X}_{i\cdot}^k)^\top)^{-\frac{1}{2}}(\mathbf{X}_{j\cdot}^k - \mathbf{X}_{i\cdot}^k)\frac{\mathbf{f}}{\sigma_{\mathbf{e}}}, 0, 1\right).\end{aligned}$$

When we calculate the solution to Eq. (6) with the linear combinations $((\mathbf{X}_{j\cdot}^k - \mathbf{X}_{i\cdot}^k)\mathbf{R}(\mathbf{X}_{j\cdot}^k - \mathbf{X}_{i\cdot}^k)^\top)^{-\frac{1}{2}}(\mathbf{X}_{j\cdot}^k - \mathbf{X}_{i\cdot}^k)$, we can see that the resulting maximum likelihood estimate $\hat{\mathbf{f}}_{ml}$ is actually an estimate of $\frac{\mathbf{f}}{\sigma_{\mathbf{e}}}$. The sample variance of $\hat{\mathbf{f}}_{ml}$, $\mathrm{var}[\hat{\mathbf{f}}_{ml}] = \frac{1}{n}\sum_{i=1}^n((\hat{\mathbf{f}}_{ml})_i - \frac{1}{n}\sum_{i=1}^n(\hat{\mathbf{f}}_{ml})_i)^2$, will express the relation in magnitudes of the target signal $\mathbf{f}$ and noise $\mathbf{e}$, i.e., the signal-to-noise ratio (SNR). When the maximum likelihood estimate $\hat{\mathbf{f}}_{ml}$ is calibrated to instrumental data to produce $\hat{\mathbf{f}}_{cal}$, the noise variance can be estimated by

$$\hat{\sigma}_{\mathbf{e}}^2 = \frac{\mathrm{var}[\hat{\mathbf{f}}_{\mathrm{cal}}]}{\mathrm{var}[\hat{\mathbf{f}}_{ml}]}. \tag{8}$$

Therefore, we can very easily estimate the noise variance from the uncalibrated and calibrated estimators. However, very strong regularization may decrease $\mathrm{var}[\hat{\mathbf{f}}_{ml}]$ and, therefore, increase the estimate $\hat{\sigma}_{\mathbf{e}}^2$.

2.6 Estimating uncertainty

The estimated noise variance in the previous section is not a direct measure of uncertainty of the calibrated maximum likelihood estimate $\hat{\mathbf{f}}_{cal}$. We use bootstrapping to estimate the uncertainty by resampling all the data that is input to PaiCo. The proxy records are resampled with replacement, which, according to the model in Eq. (1), takes into consideration the effect of different noise realizations, temporal structures and transfer functions. The instrumental data is more difficult, since it is an uncertain representation of the target climate variable and we only have a single realization of it, hence we cannot carry out similar resampling with replacement. To account for this uncertainty, we assume the instrumental data is simply the target climate variable added with zero mean Gaussian noise, i.e., $\mathbf{f}^I \sim \mathcal{N}(\mathbf{f}, \sigma_I^2\mathbf{I})$, as is commonly done (Tingley et al. 2012). However, since we do not have a non-noisy version of the target climate variable during the instrumental period, we use the instrumental data we have as an estimate of it. The instrumental noise variance σ_I^2 needs to be given by the user, since we have no reliable way of estimating it from a single time series of instrumental data. Putting everything together, a bootstrap sample is generated by resampling the proxy records with replacement and adding to the instrumental data independent zero mean Gaussian noise with noise variance σ_I^2, and using all this with PaiCo. The bootstrap samples can then be used to calculate uncertainties for the various statements made about the results.

The resulting ensemble is a collection of reconstructions based on resampling the predictor network and, for each of these bootstrapped data sets, scaling the pairwise MLE to match the mean and variance of the instruments plus noise over a calibration interval. The spread in the resulting ensemble is a measure of the uncertainty in our best estimate of the past climate. The ensemble produced in this manner has a different interpretation than Bayesian posterior sampling (Li et al. 2010; Tingley and Huybers 2010a), which produces a collection of equally likely realizations of the climate conditional on the data and modeling assumptions. Our ensemble is useful for characterizing the confidence in our best estimate of past climate that arises from the particular collection of proxy time series used to estimate model parameters. However, the ensemble cannot be used to calculate probabilities for events in the past climate, since the ensemble is based on a single realization of the climate and not all possible realizations.

Note that if σ_I^2 is estimated, the accuracy of the estimate affects the accuracy of the uncertainty estimates. If σ_I^2 is underestimated, then the uncertainty is also underestimated. As a special case, when the instrumental data is noisy but the noise is ignored, then σ_I^2 can be seen to be estimated as 0 and, therefore, it is severely underestimated. An accurate instrumental noise variance is important for the accuracy of the uncertainty estimates. Further discussing the estimation of the instrumental noise variance is out of the scope of this paper.

2.7 Temporal fidelity

It has been suggested (Cook et al. 1995) that proxies have different fidelities to represent variations of different

temporal extent. For example, tree rings have been considered to only accurately represent variations in at most decadal scale, while pollen records represent variations in multidecadal and centennial scale (Moberg et al. 2005). The likelihood function in Eq. (6) allows taking this information into consideration. If, for example, one would only trust a proxy record to accurately represent variations in a decadal or smaller scale, it is possible to remove any pairwise comparison for that record from Eq. (6) where samples are more than a decade apart. This has the effect that the comparison of a pair of proxy values is only considered if the temporal distance is at most a decade. As an example, if two proxy observations are separated in time by 100 years, then the pairwise comparison between them would not be included in the likelihood. PaiCo thus has the flexibility to incorporate information from the proxies on selective timescales, so can reflect the scientific understanding of proxies that are not faithful recorders of climate at all timescales.

2.8 Inferring sample boundaries

Calculating the linear transformations $\mathbf{X}_{i\cdot}^{k}$ requires knowledge about the temporal boundaries of each sample of a proxy record. However, this information is often not available for records in the online databases nor even calculated by the original investigators. Optimally, if the age-depth model is available, the sample boundaries in depth could be used with the age-depth model to infer the temporal boundaries of each sample. However, the age-depth model is not always available, but samples from it are often tabulated with the proxy records as sample depth and sample middle age. These data can be used to approximate the age-depth model and to infer the temporal bounds of each sample from the depth boundaries.

Let $\mathbf{t}_i^k$ be the middle age of sample i of proxy record k, and let $\mathbf{d}_i^k$ be the corresponding depth. The age does not need to be the middle age, but can be the top or bottom age as long as $\mathbf{d}_i^k$ corresponds to the same position in the sample. Let $\mathbf{d}_{i1}^k$ and $\mathbf{d}_{i2}^k$ be the top and bottom depths of sample i of proxy record k, respectively. The idea is then to use an appropriate interpolation method with $\mathbf{t}^k$ and $\mathbf{d}^k$ to calculate the ages $\mathbf{t}_{i1}^k$ and $\mathbf{t}_{i2}^k$ at depths $\mathbf{d}_{i1}^k$ and $\mathbf{d}_{i2}^k$. The calculated top and bottom ages $\mathbf{t}_{i1}^k$ and $\mathbf{t}_{i2}^k$ then define the temporal boundaries for each sample i that can subsequently be used to define $\mathbf{X}_{i\cdot}^{k}$.

There are no restrictions to the interpolation method used. However, there is a need for extrapolation at either end, since, for example, middle ages and respective middle depths do not cover the top nor bottom-most boundary of a core. We use piecewise-linear interpolation since the age-depth models are often not very complex functions and thus a linear interpolation between samples is expected to produce adequate results.

3 Comparisons between PaiCo and related methods

Several methods for climate reconstructions from multiple proxy sources have been previously proposed. See Tingley et al. (2012) for latest review. We compare PaiCo to method-of-moments (MoM), also known as variance matching (Lee et al. 2008); ordinary least squares (OLS); principal component regression (PCReg, Tingley et al. 2012); RegEM (Schneider 2001); the method by Li et al. (2010) (LNA) and BARCAST (Tingley and Huybers 2010a).

3.1 Qualitative comparisons

This section lists and discusses the most important properties of PaiCo and compares them to the selected climate reconstruction methods. Table 1 summarizes the qualitative comparison of different climate reconstruction methods. The comparison is based on the current state of the methods. Many of the methods could possibly be extended to improve their properties with respect to this comparison.

3.1.1 Transfer function

Existing climate reconstruction methods all assume the transfer functions between proxy records and temperature are linear. However, it is known that some proxies are not linearly related to, for example, temperature (Anchukaitis et al. 2012; Maidment 1993; Mann et al. 2012a, b; McKay et al. 2008). When a non-linear transfer function is known or adequately approximated, it may be possible to invert the transfer function and bring the proxy record close to linear. Furthermore, many of the published high-quality proxy records are known to be linearly related to, for example, temperature or are even reconstructions of the temperature. Assuming linearity is often reasonable, and many raw proxy observations are processed to reflect a linear relationship with the target climate variable. If linearity can be assumed for all of the proxy records in a multi-proxy reconstruction, then linear methods should be used, as the more general assumptions of PaiCo are not required.

There may be cases where it is not reasonable to assume linearity. For example, the values of the proxy in the prediction interval may be different from those in the calibration interval, so that the predictions based on linearity require extrapolation of the linear relationship outside of

Table 1 Summary of qualitative comparison of climate reconstruction methods

	PaiCo	MoM	OLS	PCReg	RegEM	LNA	BARCAST
Transfer function	Monotonic	Linear	Linear	Linear	Linear	Linear	Linear
Auto-correlated noise	No	No	No	Yes	No	Yes	Yes
Multiresolution	Yes	No	No	No	No	Yes	No
Standardization	No	Yes	Yes	Yes	No	No	No

the calibration range. As the scientific understanding of many proxies is imperfect, the less restrictive assumption of monotonicity may be preferred. The weaker assumption may allow for the inclusion of more data sources than only those proxies that are well-known to have linear relationships with the target climate. There is a trade-off in doing so: on the one hand, PaiCo can include more information in the form of additional proxies, while on the other, PaiCo loses information from those truly linear proxies by using the weaker assumption of a monotonic transfer function. Quantitative experiments (Sect. 3.2) suggest that the loss of information is small, as PaiCo results in inferences that are competitive with linear methods when applied to linear proxies.

3.1.2 Auto-correlated noise

In LNA, noise is assumed to follow an AR(2) model and the parameters of the AR-process are estimated via Bayesian inference. In BARCAST, the latent temperature field is modeled as an AR(1)-process. It is possible to utilize information about auto-correlated noise in both PaiCo and RegEM, possibly with slight modification of the current implementations of the methods. However, obtaining this information may be difficult in practice, and therefore, independent noise is often assumed for RegEM and the same assumption is made for PaiCo in the rest of this paper. In MoM and OLS, noise is assumed independent (Tingley et al. 2012). While least squares regression models can be built for correlated noise, we do not consider the resulting weighted regression models as equivalent to OLS as used in the paleoclimate community. In PCReg, auto-correlated noise will likely appear as one or more principal components, since, assuming the model of PCReg is correct, the auto-correlated noise of each proxy is independent between proxies and independent of the target. If the principal components for regression are chosen appropriately, the auto-correlated noise should be ignored.

3.1.3 Multiresolution

While many of the proxy records used to this day have annual resolution, there are still numerous records that have subannual resolution. Only LNA and PaiCo can handle proxy records of various resolutions. If neither of those methods are used, the only viable option is to only use proxy records with the same resolution and exclude any other records. Note that BARCAST could be quite easily extended to handle multiple resolutions with, for example, ideas from LNA. However, this comparison is based on the current state of these methods.

3.1.4 Standardization

The assumption of linearity for the transfer function is often accompanied with the requirement to standardize the proxy records to a common scale before reconstruction. MoM, OLS and PCReg require the proxy records to be standardized to the same units, or otherwise the proxies with a higher sample variance may dominate the result. The problems of standardization boil down to how accurately can the scale of a proxy record be altered. Standardization is generally carried out by shifting and scaling each proxy record to have zero mean and unit variance over a specific standardization interval, using standard estimators for mean and variance. As standardization is carried out using sample, rather than population, values of the mean and standard deviations, the standardization itself is sensitive to the noise in these estimators. Finally, as the proxy time series are generally of different lengths, the standardization interval is generally short relative to the time interval spanned by the data set as a whole. In such cases, standardization can introduce spurious structure into the time series of the standard deviations across the records; see Tingley (2012) for further discussion and suggested solutions. When standardization is not required, the accuracy of these estimator is irrelevant and, therefore, the problems of standardization do not carry to the methods not requiring standardization.

The original definition of RegEM by Schneider (2001) includes mean and variance terms for each proxy record, and therefore, standardization may not be required. LNA has different mean and variance parameters for each proxy record. However, the parameters for noise are common to a proxy type and the authors suggest that the proxy records are standardized for the model of common noise to be reasonable. However, assigning each proxy as its own type removes the need to standardize. Similar to LNA, BARCAST has a set of coefficients for each proxy type and

defining each proxy record as its own proxy type renders standardization unnecessary.

PaiCo does not require standardization as any shift and scale operation can be considered to be contained in the transfer function g^k as long as the record is not flipped in sign. The mean and variance of a record are not directly used by PaiCo, and the method is invariant with respect to arbitrary monotonic transformations of the proxy series.

3.2 Quantitative comparisons

In this section we compare the methods quantitatively in two different pseudo-proxy scenarios.

For MoM, which is a CPS method, the uncalibrated composite is produced by, for each year, calculating the mean of the available standardized proxy values. For PCReg, we estimate the number of significant principal components from the eigenvalues (Mann et al. 2007). We use total truncated least squares for RegEM and estimate the truncation parameter with the same method as for PCReg. For BARCAST, we set all the spatial locations to the same coordinate, effectively removing the spatial component from the model. We also experimented with the state-space method by Lee et al (2008) with our implementation and with an implementation received from the authors, but failed to get either of them to converge. The method was thus excluded. Implementations of all of the used methods and Matlab source code for the experiments are available in the Electronic Supplementary Material.

3.2.1 Demonstrating properties of PaiCo

We first demonstrate the properties of PaiCo in comparison to other reconstruction methods using a set of idealized experiments. A target time series is constructed from four different parts concatenated in order: 100 point plateau at height 10, 100 point linear trend from 10 to 0, 100 point plateau at 0 and 200 point sinusoid with $\frac{2\pi}{100}$ angular frequency and unit amplitude. Figure 3A illustrates an example target time series. A pseudo-proxy record is generated by adding independent Gaussian noise with standard deviation $\sigma_{\mathbf{e}}$ to the target time series and normalizing them to zero mean and unit variance. The calibration period is set to the last 100 points, covering one complete cycle of the sinusoid and leaving the rest as the reconstruction period. In the experiments, the default setting has $M = 20$, where M is the number of proxy records, $\sigma_{\mathbf{e}} = 1$ for all proxies, and the proxies are linear. The default is varied to show the similarities and differences in the methods. The experiments are run 40 times for each setting to show the mean behavior of each reconstruction method. Figure 3 depicts all of the results. Details of the experiment, such as parameter values for each of the methods, are available in the Supplementary Online Material.

Figure 3A depicts results with $M = 20$ and $\sigma_{\mathbf{e}} = .5$. All methods are very capable of reconstructing the target time series with no significant differences. Figure 3B depicts the relative error in recovering the plateau at 10 when the number of proxy records is varied from 5 to 100. Many of the linear methods behave similarly with PaiCo performing equally well. OLS has trouble when the number of proxies grows large, since the number of predicands is equal to the number of samples. This could be remedied by applying some kind of regularization. LNA performs best when there are only few proxies. The relative performance of both Bayesian methods decreases when the number of proxies increases, but we believe this is a convergence issue, rather than a weakness of the underlying methods. As both have parameters for each proxy, adding proxies makes convergence slower. During the iteration, both methods came quite rapidly close to the plateu, but convergence to it became increasingly slower with decreasing distance. Therefore, we think the combination of the large number of parameters to learn and the shape of the target led to these convergence issues. We did moderate optimization and parameter tuning to try to speed up the convergence. However, we kept the number of draws from the posterior distribution fixed throughout the experiment due to computational costs, which did not ensure convergence in all cases. In a practical scenario, this issue could be solved by adding more draws, constraining the sampling and/or further tuning the initial parameter values.

Figure 3C depicts results with $M = 20$ and the common $\sigma_{\mathbf{e}}$ was varied from .1 to 3. For moderate to large $\sigma_{\mathbf{e}}$, LNA performs better than the other methods, with MoM, PaiCo and BARCAST performing slightly above average in these experiments. PCReg, RegEM, and OLS are below average in this case. However, with small $\sigma_{\mathbf{e}}$, PaiCo has a very large error. PaiCo is based on the fact that the proxy records are noisy and the noise is the essential ingredient that makes the proxies disagree in the pairwise comparisons. If the noise is very small or completely removed, the proxies completely agree and PaiCo reduces to a ranking of the target climate variables. The errors of LNA and RegEM similarly increased in low noise scenarios.

To explore the sensitivity of the methods to the assumption of common noise variance made by PaiCo, we compare results under varying levels of departure from this assumption. In these experiments, the noise variance for each proxy is calculated from

$$\sigma_{\mathbf{e}}^i = \exp\left(d\left(2\frac{i-1}{M-1} - 1\right)\right),$$

where i is the proxy index and d is the parameter that controls the dispersion of noise variance among proxies.

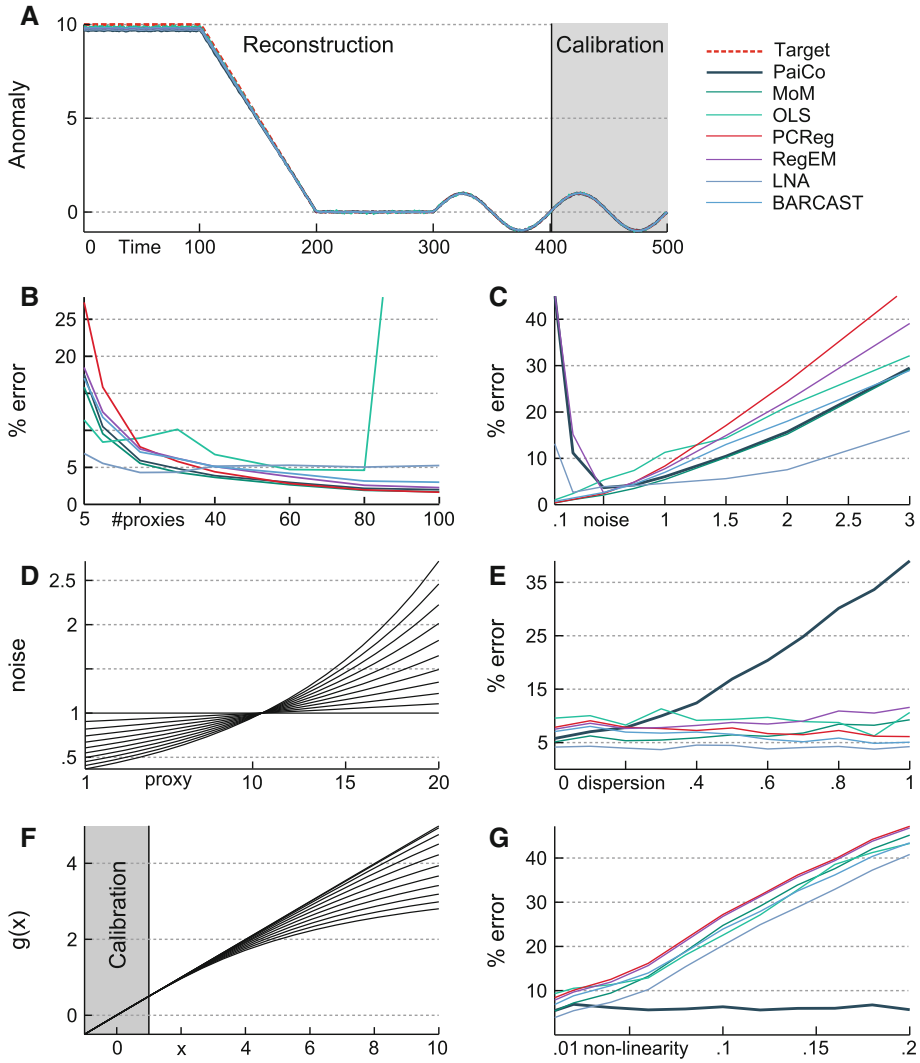

Fig. 3 Results of experiments demonstrating the different characteristics of reconstruction methods. Example target time series is depicted in **A** as *red dashed line* with initial plateau. A pseudo-proxy is generated by adding independent Gaussian noise with standard deviation σ_e. Unless otherwise stated, there are 20 linear pseudo-proxies with $\sigma_e = 1$. **A** Mean result of 40 runs for each method with $\sigma_e = .5$. **B** Percentage error for each reconstruction method, averaged over the 100 point plateau at the beginning of the time series and over the 40 runs. **C** As in **B**, but varying common noise standard deviation. **D** Noise standard deviation for each proxy for varying levels of dispersion. Each *line* depicts single level of dispersion. **E** As in **B**, but with varying levels of dispersion of noise standard deviation according to **D**. **F** Transfer functions of varying levels of non-linearity. Calibration shows the values during the calibration period. **G** As in **B**, but with varying levels of non-linearity according to **F**

With $d = 0$, all proxies have the same noise variance, $\sigma_e^i = 1$, and the dispersion in the noise variances increase with increasing d. Figure 3D illustrates the amount of noise for each proxy as d varies from 0, i.e., flat, to 1, i.e., the steepest curve. Figure 3E depicts the corresponding performances of the reconstruction methods. The performance of PaiCo degrades with increasing noise dispersion, and all other methods perform better than PaiCo for $d \geq .4$. This experiment was explicitly designed to highlight the limitations of PaiCo, and in practice we would expect the range of noise variances to be smaller than the $d = .4$ case. In fact, we tested this by estimating the noise standard deviation for each proxy in Kaufman et al. (2009) from the residuals and calculating the coefficient of variation for these estimated standard deviations. The coefficient of variation is defined as the standard deviation divided by the mean of a group of numbers, with larger values expressing higher variation. The coefficient of variation for Kaufman et al. (2009) is .13 while the coefficient of variation with $d = .4$ is .25, and hence, the dispersion of noise variances in Kaufman et al. (2009) is much smaller than in the case $d = .4$. Therefore, PaiCo would most likely perform well in that scenario. Notice that the decadal binning in Kaufman et al. (2009) has virtually no affect in this result, since the mean and standard deviation of the estimated proxy noise standard deviations are affected equally by the averaging of the binning.

Last, we explore the sensitivity of the methods to the linearity of the proxies. To produce non-linear proxies, we use an arctan(lx) transfer function, where x is the input value and l is the parameter controlling the amount of non-linearity. Arctan is close to linear around zero and progressively more non-linear at values larger in magnitude (Fig. 3F); we thus expect the performance of the linear methods to suffer as l increases from .01 to .2. We apply the transfer function individually to each proxy after adding the noise but before they are normalized to zero mean and unit variance. Figure 3G depicts the results for varying levels of non-linearity. As expected, PaiCo is the only method that does not suffer in performance as the proxies

become increasingly non-linear. The poor performance of the linear methods is to be expected, as the transfer function is explicitly designed to be nearly linear over the range of target values included in the calibration, but strongly non-linear for larger values of the target. If the entire range of the target variable were included in the calibration interval, we would expect the performance of the linear methods to improve, but to still not equal the performance of PaiCo.

3.2.2 *Auto-regressive pseudo-proxies*

To test the performance of the methods in a more realistic setting, we use the framework described in Tingley and Huybers (2010b), in which a 500-point target time series is generated as Gaussian noise with lag one auto-regressive coefficient of .5. To increase the low frequency variability, we add to this target time series a second time series with .9 lag one AR-coefficient. A pseudo-proxy is generated from the target by adding varying amounts of white noise to it to achieve a desired signal-to-noise (SNR) ratio. Depending on the case, the pseudo-proxy is further transformed with a random monotonic function. A number of these pseudo-proxies are constructed to form a dataset of pseudo-proxy records. Each of the methods are individually used with the dataset and the root mean square error (RMSE), Pearson's product-moment correlation coefficient, Reduction in Error (RE) and Coefficient of Efficiency (CE) are calculated. RMSE is always divided by the standard deviation of the target signal to arrive at comparable results.

We vary the number of proxy records (10, 20, 50, or 100), the signal-to-noise ratio (.4, .8, 1.2, 1.6, 2.0, 2.4, 2.8, or 3.2) and the linearity of the proxies. For each non-linear proxy record, a monotonic transfer function is generated by first sampling two values, y_1 and y_2, uniformly at random from [0,1]. We then space the values $\{0, y_1, y_2, 1\}$, assuming $y_1 \leq y_2$, evenly over the range of the noisy target values. We use cubic interpolation to fit a function through these points to transfer the noisy target values to the range [0,1]. Cubic interpolation is guaranteed to create a function that runs through the evenly spaced points $\{0, y_1, y_2, 1\}$, which means that the transfer functions will be monotonic. We run the experiment 100 times for each combination of parameter values to get a sense of the average behavior of each of the methods. Figure 4 depicts all of the results.

In Fig. 4A, the average RMSE over 100 runs is shown for each method and varying SNR values with 10 non-linear pseudo-proxy records. Only PaiCo is able to come close to the expected RMSE value, calculated as the standard deviation of the sample mean estimator, with all the linear methods having much higher values, particularly for larger values of the SNR. When the noise dominates the proxy data, i.e., SNR < 1, the methods behave somewhat

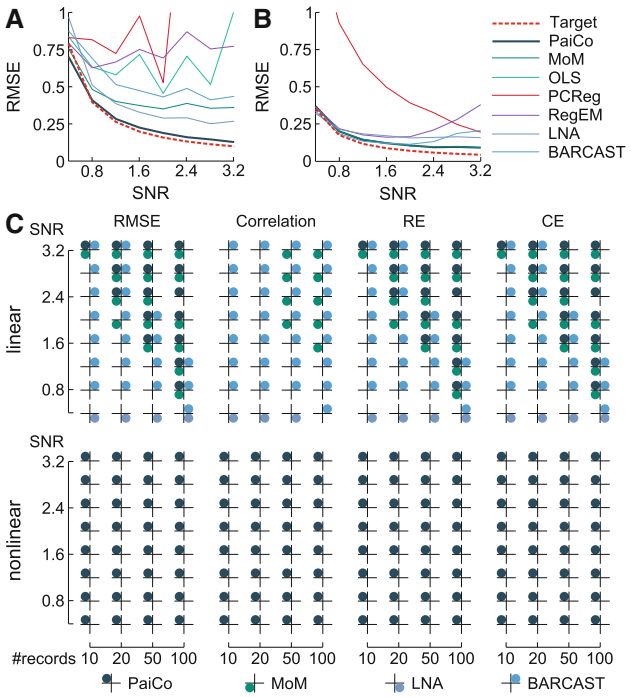

Fig. 4 Results of experiments with autoregressive time series. **A** Average root mean square (RMSE) values for each reconstruction method using 10 non-linear pseudo-proxies. The estimated RMSE is the standard deviation of the sample mean estimator without non-linearities. RMSE's are scaled with target standard deviation for comparability. **B** Average RMSE values for each reconstruction method using 50 linear pseudo-proxies. **C** Best methods with 5 % confidence in different settings

similar with PaiCo still being slightly better. Figure 4B has the same setting except that there are now 50 linear pseudo-proxies. All of the linear methods achieve better performance than before, to a varying degree. PaiCo, MoM and BARCAST perform equally well when $1.2 \leq \text{SNR} \leq 2.2$, and better than the other methods. For small values of SNR, all methods perform well, with BARCAST or LNA having slightly smaller RMSE than the other methods. For larger values of the SNR, the performance of BARCAST decreases relative to the other methods—most likely due to the convergence issues discussed above.

To give a more general sense of the the relative performance of the methods, Fig. 4C plots the best method according to four statistics for each combination of parameter values. For a single set of parameters and a single test statistic, the 100 runs result in 100 test statistic values for each method. The method with the best average test statistic value was chosen as the initial best method. Then, each other method was compared to the initial best method with a single-tail paired t test. A method was included in the group of best methods if the averages of the test statistics were indistinguishable at 5 % confidence level. This way we can find which methods are better than

all the other methods, and among which supremacy can not be statistically significantly determined. Note that multiplicity corrections need not be carried out here since we are using the t-tests only as a tool to summarize the performance of the methods in different settings.

Figure 4 shows a clear separation between the linear and non-linear settings. PaiCo performs better than any of the other methods when proxy records are non-linear, as expected. When the pseudo-proxies are linear, BARCAST is always among the best performing methods, with the exception of low SNR cases for which LNA performs better. However, PaiCo was unexpectedly also among the best performing methods as measured by RMSE, RE and CE when there are sufficiently many pseudo-proxy records or the records have little noise. The performance of MoM and PaiCo is often indistinguishable in the linear case. The relative performance of the methods as measured by correlation is similar to that measures by the RMSE, with the exception that PaiCo is no longer one of the best methods. BARCAST, which is by construction correct in all settings where the proxies are linear, performs well in settings when the proxies are noisy and few in number—provided they are linear. In summary, PaiCo performs on par with MoM when all proxy records are linear, and outperforms all linear methods when the proxies records are non-linear.

3.3 Further quantitative properties of PaiCo

We calculate the average execution times of each of the methods over all combinations of parameter values and 100 runs. PaiCo, RegEM, LNA and BARCAST are optimized in several ways to reduce the computation time. BARCAST was further modified to sample the complete spatiotemporal field at once instead of sampling the spatial field of each year separately in order to speed up its convergence. The results are depicted in Fig. 5A. RegEM and BARCAST have similar execution times, while LNA is several times slower. The most costly part of LNA is the sampling of the AR(2)-coefficients, which is done using Metropolis-Hastings and requires several calculations of the normal CDF at each iteration. However, this could be alleviated by using the uniform distribution as a proposal density.

We analyze the correctness of the uncertainty estimation of Sect. 2.6 by calculating the coverage rates as described by Li et al. (2010). A coverage rate is calculated by first deciding on a nominal level. Then, for each time point, the uncertainty estimation is used to find the upper and lower bounds corresponding to the quantiles of the nominal level. The corresponding actual coverage rate is then the fraction of time points for which the target is within the upper and lower bounds, i.e., is covered by the confidence intervals. We calculate the coverage rates in the same settings as

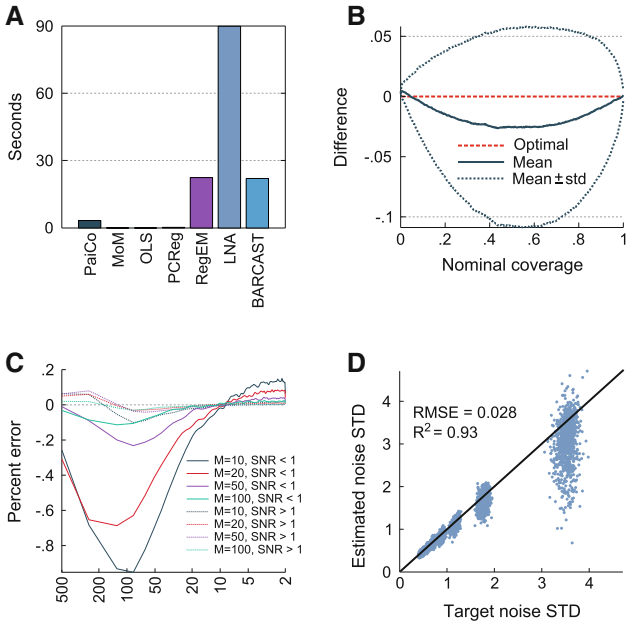

Fig. 5 Statistics from the auto-regressive pseudo-proxy experiments using PaiCo. **A** Average execution time for each method over all settings. **B** The difference between actual coverage and nominal coverage of uncertainty. Note that the uncertainties are point-wise. **C** Average difference of the power spectrums of PaiCo and the target. **D** Scatterplot of the target (true) standard deviation of the noise and the standard deviation estimated by PaiCo

before, but with 10 runs for each combination of parameter values and we obtain 100 resamples in each case. Figure 5B depicts the difference between the actual coverage and the nominal coverages. The uncertainty is slightly underestimated with around 2.5 % error. However, the errors are fairly small and, thus, the uncertainty estimation seems to be quite accurate.

We also compare the power spectrum of PaiCo against the power spectrum of the target using the method by Welch (1967). Figure 5C depicts the average difference of the power spectrums over different settings. When SNR is small, the noise dominates the pseudo-proxy records and, therefore, the low-frequency variability is underestimated while high-frequency variability is overestimated. This is expected from any method since, in high-noise cases, the noise "overwrites" the information about the target in the pseudo-proxies and, therefore, no method can recover the low-frequency variability. When there is little noise, the difference in power spectrums is much smaller. It seems that the error of PaiCo can be decomposed to a slight overestimation of the millennial to centennial scale variability and a slight underestimation of the centennial to decadal scale variability. However, as shown in Fig. 4, the errors in the reconstructions are among the smallest of any reconstruction method in many of the tested settings.

As a final comparison, we also test the correctness of the estimate of the standard deviation of the noise as shown in

Eq. (8). Figure 5D depicts as a scatterplot the relation between the true and estimated standard deviations of noise. The noise standard deviation estimates are generally quite good, but feature a slight bias towards underestimating the true values. However, as the one-to-one fit has $R^2 = .93$ and RMSE = .028, it is clear that the estimated standard deviation is quite accurate, especially when the noise is fairly small.

4 Proxy data

The study region is the area north of 60°N and between 50°W and 30°E. This area encompasses the northernmost part of the Atlantic Ocean and the surrounding land areas, which are also assumed to be influenced by the Atlantic climate. We focus on the temperature of the Arctic Atlantic during the past 2,000 years. The multiproxy data set is composed of all (to our knowledge) publicly available records that meet four predetermined and objective criteria.

First, each included proxy series must be used as a temperature sensitive proxy in the peer reviewed literature. This assures that each proxy record was objectively assessed by experts in the field and its relation to temperature was also justified. Furthermore, we used the reconstructed temperature whenever available. For example, we only used reconstructed temperatures for marine sediments, as PaiCo is unable in its current state to handle assemblages as proxy records. We also applied the processing steps suggested by the original authors. For example, if it was suggested that the record is dominated by human influence during some time period, we excluded such parts of the record. We, therefore, rely on the original authors to provide the most accurate proxy records.

Second, each included record must extend back to at least 1500 AD. There are far more short, modern records than long records, but by including only the longer records we maintain the focus of the analysis on the last 2,000 years as a whole.

Third, we included only those records with at least one well-dated age control every 500 years. As the method considers the ages in records to be correct, controlling the temporal quality becomes extremely important. This fairly strict criterion ensures the chronology control is of high quality and that we can be fairly certain about the age-depth model. Many lake sediment records were not included because they did not meet this criterion. Notice that varved (i.e., annually laminated) records are like trees, so they essentially have a date for every year.

Fourth, all included records have on average at least one observation every 50 years. We measure the resolution of a record as the mean difference of sample ages, including only those samples with ages between 0 and 2000 AD. We required the average distance between mid-ages of samples to be at most 50 years, so that those record covering the entire 2,000 year feature at least 40 observations. The resolution criterion was set to exclude low-resolution proxies that have little effect on the overall result. Borehole records and many lake sediment records were excluded on the basis of this resolution criterion.

Figure 6 shows the locations of the collected proxy records and their temporal structure. Details of each record are listed in Table 2. The study area is likely richer in long, high quality proxy records than any other region in the Arctic. Furthermore, the whole study area is fairly evenly covered without any significant clusters. The dataset is also heterogeneous in proxy sources. The ice core records constitute two fifths of the database, while the tree, marine and lake proxies share the remaining portion quite equally. We have also included a documentary record of sea ice history in Iceland, and a single speleothem record from Norway. While many of the records have previously been

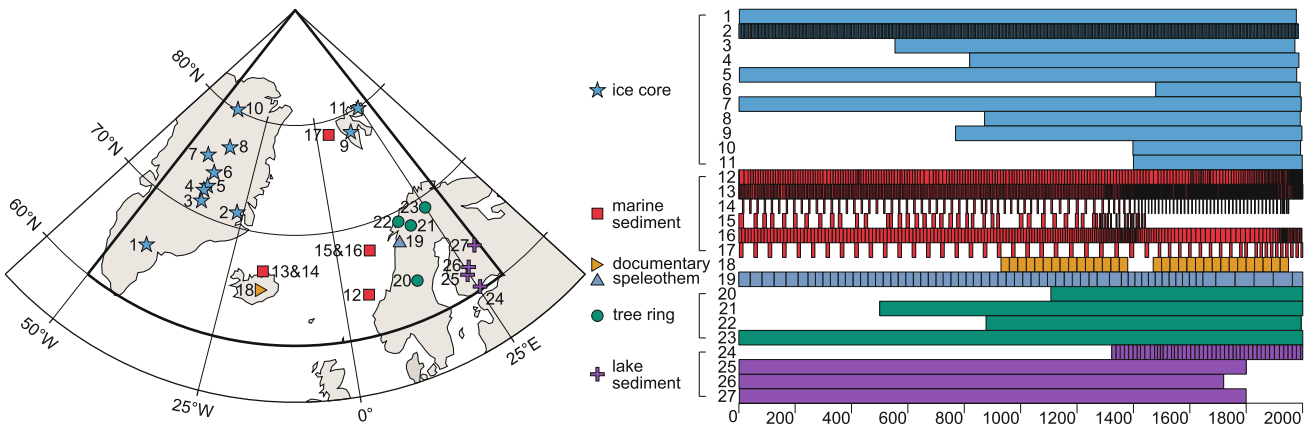

Fig. 6 The locations of the proxy records and their temporal structure from 0 to 2000 AD. A *continuous box on the right* represents annual resolution. Otherwise, each *box* represents the temporal boundaries of a single sample

interpreted as linearly related to temperature, we note that the the experiments in Sect. 3.2 suggest PaiCo is often competitive with linear methods when applied to linear proxies. Further, it is unlikely that assumption of linearity holds for all proxy records used here. The dataset is available in the Electronic Supplementary Material.

5 Results

Our Arctic Atlantic composite PaiCo reconstruction for the last 200 years is shown as temperature anomalies (reference period 1961–1990) in Fig. 7A, including noise estimation from Eq. (8) and uncertainty from Sect. 2.6 with 1,000 resampled time series. The annually resolved temperature reconstruction is significantly correlated ($r = .30$ with $p < .001$) with the instrumental surface air temperature (SAT) time series, derived as an area-weighted mean from the HadCRUT3 global temperature database (Brohan et al. 2006) from the meteorological stations of the corresponding region. Statistical significance is calculated by resampling 10^4 random time series with the same AR(1) correlation coefficient (.45) as the reconstruction, and calculating the fraction of random time series that had equal or larger correlation with the instrumental target than the reconstruction. The instrumental noise standard deviation for uncertainty estimation is obtained by first calculating the variance of the mean estimator for each year, and then calculating the square root of the average of these variances. The uncertainties in the following statements are reported in parenthesis as either a percentage of ensemble members for which the statement holds, or the mean and standard deviation of a statistic calculated from the ensemble. The ensemble is a collection of 10^4 bootstrapped reconstructions of the temperature, calculated according to Sect. 2.6 on the annual timescale.

Three distinct periods can be discovered in the observational record: strong regional warming over 1900–1940 and 1970–2000 separated by equally strong cooling from 1940 to 1970. The reconstruction generally reproduces these same features, although annual inspection reveals some deviations or even inverse correlations. The significant correlation between reconstructed and measured air temperature anomalies holds back with the onset of the instrumental record, around 1850. Within the last 200 years, North Atlantic temperatures were clearly below average between 1800 and 1900 (for 100 %, average of nineteenth century <0), and mostly above the average during the last century (for 90 %, average of 1920–2000 >0).

In Fig. 7B, we show the outcome of the PaiCo reconstruction for the entire 2,000 years, smoothed by calculating the average temperature for each decade. The decadal temperature reconstruction is significantly correlated ($r = .84$ with $p < .001$) to the decadal instrumental data.

Among the strongest overall features in our composite temperature reconstruction is a cooling from 1 to 2000 AD, with least squares linear regression yielding a cooling trend of $-.33$ °C/1,000 years ($-.30 \pm .09$). Superimposed on this cooling trend, another striking feature of the record is the pronounced low-frequency climate variability. On the basis of these clearly distinguishable climate fluctuations, the last 2 ka of climate evolution in the studied Arctic Atlantic domain can be subdivided into the Roman Warm Period (RWP, until ca. 600 AD), the Dark Ages Cold Period (DACP, ca 600 to 900 AD), the Medieval Climate Anomaly (MCA, ca. 900 to 1200 AD), the Little Ice Age (LIA, ca. 1500 to 1900 AD), and the Recent Warming (RW, 1900 to present). The overall temperature variability in our 2,000 year long record is 2.3 °C (95 % of values lie within this range; ensemble has range $2.5 \pm .4$ °C), with lowest values in the first part of the nineteenth century, during the LIA, and peak maxima during the RWP around 400 AD, the early MCA at 1000 to 1150 AD, and in the modern industrial period. The LIA shows clear evidence of multidecadal climatic variability, such as the three successive warm phases between 1400 and 1600 AD. In all, the end of LIA (1600 to 1900 AD) appears to be climatically more uniform than its initiation.

We compared the PaiCo Arctic Atlantic reconstruction (Fig. 7C) with another previously published Arctic (Kaufman et al. 2009) and some widely used Northern Hemisphere (Moberg et al. 2005; Mann et al. 2008) reconstructions based on various combinations of proxy data and differing statistical approaches. This comparison demonstrates rather striking agreement between alternative estimates over the past eight centuries. Prior to 1200 AD, the reconstructions still all agree rather well in respect to the overall temporal patterns, yet our new Arctic Atlantic record shows much greater variability over time. Our reconstruction is .29 °C warmer than the other records during the MCA ($.26 \pm .19$) and .56 °C warmer during the RWP ($.56 \pm .23$). The episode around 400 AD in our reconstruction is .63 °C warmer than in any other reconstructions ($.55 \pm .28$).

Figure 8 illustrates the most pronounced warming and cooling trends, calculated by finding the window with the steepest linear trend for each window size from 10 to 400, with overlapping windows merged together. A persistent and rapid multidecadal warming trend of 4.5 °C/100 years ($3.9 \pm .9$) took place in the onset of the modern (industrial) times, in particular, from 1900 to 1940 AD, which is consistent with other studies (see Sect. 6) Another systematic, rapid warming of 1.0 °C/100 years ($.94 \pm .23$) occurred from 700 to 800 AD. Conversely, the most systematic, fast and multidecadal cooling rate of -6.2 °C/100 years (-5.4 ± 1.5) occurred in the termination of the

Table 2 Details of used proxy records. Extent indicates the temporal span of a record in AD. The 'Resolution' is the average difference between sample mid-ages within 0-2000 AD

ID	Site	°N	°W	Proxy type	Measurement	Extent	Resolution	Reference
1	Dye-3	65.18	−43.83	Ice core	$\delta^{18}O$	1–1978	Annual	Vinther et al. (2010)
2	Renland	71.27	−26.73	Ice core	$\delta^{18}O$	2.5–1983	5.00	Vinther et al. (2008)
3	Crête	71.12	−37.32	Ice core	$\delta^{18}O$	553–1973	Annual	Vinther et al. (2010)
4	GISP2	72.10	−38.08	Ice core	$\delta^{18}O$	818–1987	Annual	Grootes and Stuiver (1997)
5	GRIP	72.58	−37.64	Ice core	$\delta^{18}O$	1–1979	Annual	Vinther et al. (2010)
6	B16	73.94	−37.63	Ice core	$\delta^{18}O$	1,478–1992	Annual	Schwager (1999)
7	NGRIP1	75.10	−42.32	Ice core	$\delta^{18}O$	0–1995	Annual	Vinther et al. (2006)
8	B18	76.62	−36.40	Ice core	$\delta^{18}O$	871–1992	Annual	Schwager (1999)
9	Longyearbyen	78.25	15.50	Ice core	$\delta^{18}O$	769–1997	Annual	Divine et al. (2011)
10	B21	80.00	−41.14	Ice core	$\delta^{18}O$	1,397–1993	Annual	Schwager (1999)
11	Austfonna	79.83	24.02	Ice core	$\delta^{18}O$	1,400–1998	Annual	Isaksson et al. (2005)
12	P1003	63.76	5.26	Marine sediment	$\delta^{18}O$	−5,931–1998	8.44	Sejrup et al. (2011)
13	MD99-2275	66.55	−17.37	Marine sediment	Diatoms	−2,549–2001	3.89	Jiang et al. (2005)
14	MD99-2275	66.55	−17.37	Marine sediment	Alkenones	−36–1949	19.53	Sicre et al. (2011)
15	MD95-2011	66.97	7.64	Marine sediment	Diatoms	−6,540–1440	28.04	Berner et al. (2011)
16	MD95-2011	66.97	7.64	Marine sediment	Alkenones	−4,076–1995	9.86	Calvo et al. (2002)
17	MSM5/5-712	78.92	6.77	Marine sediment	Planktic foraminifers	−94–2007	41.46	Spielhagen et al. (2011)
18	Iceland	64.77	−18.37	Documentary	Ice cover	945–1935	30.00	Bergthórsson (1969)
19	Okshola cave	67.00	15.00	Speleothem	$\delta^{18}O$	−5,565–1997	31.92	Linge et al. (2009)
20	Jämtland	63.50	15.50	Tree ring	Maximum density	1,107–2007	Annual	Gunnarson et al. (2010)
21	Torneträsk	68.26	19.60	Tree ring	Maximum density	500–2004	Annual	Grudd (2008)
22	Forfjorddalen 2	69.08	17.22	Tree ring	Ring width	877–1994	Annual	Kirchhefer (2001)
23	Finnish Lapland	69.00	25.00	Tree ring	Ring width	0–2005	Annual	Helama et al. (2010)
24	Lake Hampträsk	60.28	25.42	Lake sediment	Chironomids	1,330–2000	14.57	Luoto et al. (2009)
25	Lake Nautajärvi	61.81	24.68	Lake sediment	Organic matter	0–1800	Annual	Ojala and Alenius (2005)
26	Lake Korttajärvi	62.33	25.68	Lake sediment	X-ray density	0–1720	Annual	Tiljander et al. (2003)
27	Lake Lehmilampi	63.62	29.10	Lake sediment	Varve thickness	1–1800	Annual	Haltia-Hovi et al. (2007)

RWP from 590 to 620 AD. Another steep cooling trend of −1.4 °C/100 years (−1.2 ± .4) occurred between 400 and 490 AD. The longest cooling trend of −.31 °C/100 years (−.28 ± .08) was seen between 990 and 1410 AD. The millennial-scale average cooling trend is −.36 °C/1,000 years (average slope of each 1,000 year window; ensemble has −.32 ± .11).

6 Discussions and conclusions

We have presented a new method for reconstructing spatially averaged climate variables from multiple proxy sources. The key advantage of the method is the very general assumption about transfer functions. In contrast to all existing reconstruction methods, the presented method only assumes the transfer functions are monotonic. This allows the use of proxy records that display non-linear behavior but for which the transform function is unknown or for which linearity can not be assumed or tested. The monotonicity assumption does not exclude records, since all positive linear transfer functions are also monotonically increasing and, therefore, any linear proxy record can also be used.

The assumption of monotonicity instead of linearity comes at a cost. Considering only the order of the values for each pair of proxy values means that the information about the magnitudes in a proxy record are ignored. If the magnitudes in a proxy record are trustworthy information, losing such information is not desired where high-quality proxy data is scarce to begin with. If it is certain that the transfer functions between proxy records and target climate variable are linear, then linear methods should be used as their power is greater in such cases as they utilize all of the available information. However, the qualitative and quantitative comparisons to existing reconstruction methods suggest that reconstruction resulting from PaiCo should be comparable, according to standard metrics, with those resulting from existing linear methods, even if the linear assumption is adequate for all proxies.

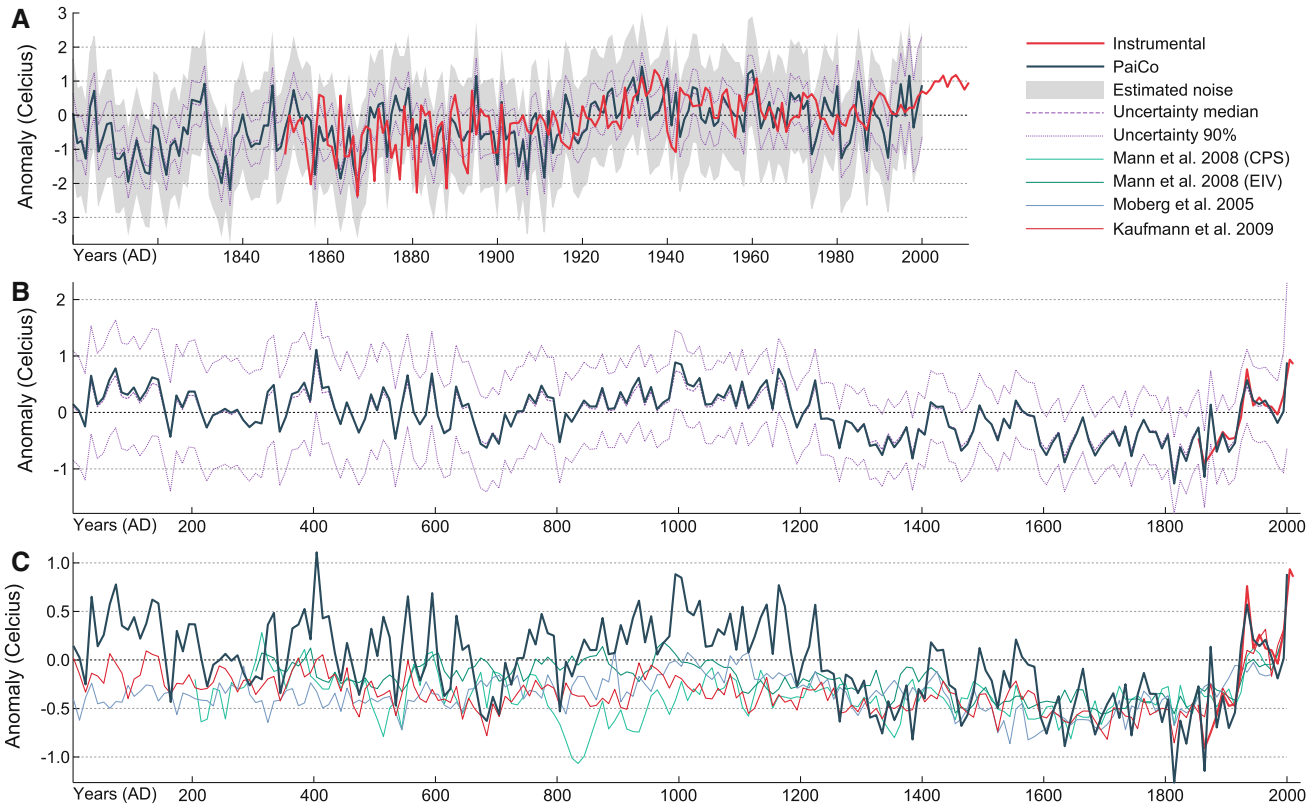

Fig. 7 Reconstructed temperature of the Arctic Atlantic. **A** Instrumental comparison of annual values between 1800 and 2000 AD with noise and uncertainty estimates. Uncertainty 90 % is derived pointwise from the percentiles of the ensemble of bootstrapped reconstructions. **B** Decadal averages of maximum likelihood and uncertainty estimates (as in **A**). **C** Comparison of the decadally averaged PaiCo reconstruction with previous reconstructions of the Arctic and Northern Hemisphere.

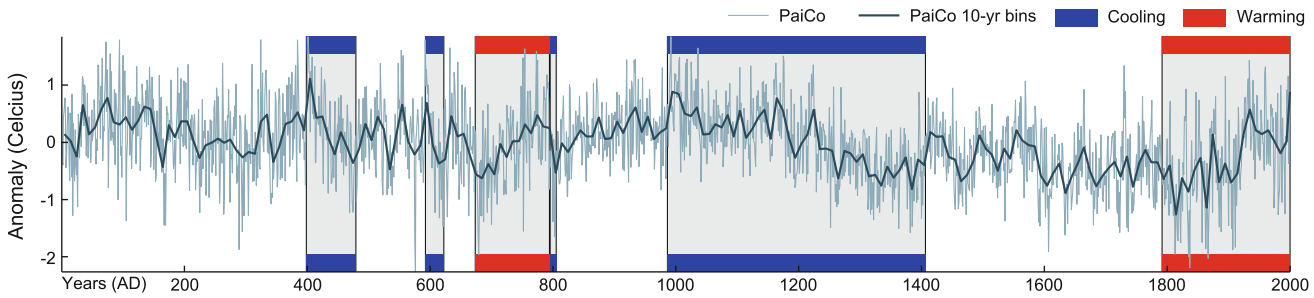

Fig. 8 Most pronounced warming and cooling trends, calculated by finding the window with the steepest linear trend for each window size from 10 to 400. We report the union of overlapping windows

Another advantage of PaiCo is that it allows for the inclusion of all proxy records on their original resolutions. While not unique to PaiCo, the only other currently available method that can do this is the linear method by Li et al. (2010). The formulation of LNA could easily be used to extend BARCAST to also have proxies of various resolutions. The task is more difficult for the other tested methods.

These Bayesian techniques produce an ensemble of equally likely climate histories conditional on the observations. The Bayesian ensemble has a different interpretation than the ensemble of bootstrapped reconstructions used here, as it provides a distribution of equally likely climate histories conditional on a fixed data collection, whereas the ensemble used here is based on bootstrapping the predictors. We leave for future work a Bayesian implementation of PaiCo.

PaiCo makes the assumption that the noise variance is the same for all proxy records, meaning that each proxy is assumed equally informative of the target climate variable. The quantitative comparisons suggest that PaiCo allows for small amount of dispersion of noise variance among

proxy records, which is to be expected in a real scenario. Indeed, PaiCo performed at par with other methods at the level of dispersion that was encountered by Kaufman et al. (2009).

The millennial-scale average cooling trend in our reconstruction is consistent with other proxy evidence showing that summer temperatures across the Arctic reached their maximum during the early Holocene, after which the climate progressively cooled towards the present (Wanner et al. 2008; Miller et al. 2010). The cooling can be associated with the monotonic reduction in summer insolation at high northern latitudes, driven by orbital configurations. In a millennial perspective, a pattern of the long term cooling has been revealed from the western Arctic ice cores (Kotlyakov et al. 2004) as well as the multi-proxy reconstruction of Arctic summer SAT (Kaufman et al. 2009). Even though the cooling trend magnitude estimated for circum-Arctic summer SAT, on the order of $-.22° \pm .06$ °C/1,000 years (Kaufman et al. 2009), is lower than the $-.36$ °C estimate for the Arctic Atlantic region, there is a qualitative similarity between the two reconstructions (Fig. 7). The insolation-driven cooling in the North Atlantic realm was presumably enhanced by numerous positive feedbacks (e.g., albedo, sea ice) that amplified the forcing more strongly in the area than elsewhere in the Arctic. In particular, a substantial sensitivity to variations in sea ice extent provide a reasonable explanation for this discrepancy.

Our study clearly demonstrates that the late-Holocene climate was more variable and extreme in the North Atlantic region than Arctic as a whole. This finding is in accordance with Mann et al. (2009) who found Medieval warmth and LIA cooling to be particularly pronounced in the high-latitude North Atlantic. One way to measure the relative MCA warmth is through the difference with the LIA. However, in such comparison it is crucial which time interval to choose to represent MCA and LIA (Goosse et al. 2012). Frank et al. (2010) used the intervals 1601–1630 and 1071–1100 for LIA and MCA, respectively, and found a relative hemispheric-mean MCA warmth of .38 °C, whereas Goosse et al. (2012) found in their simulations slightly smaller warmth of .33 °C using the same periods, but greater warmth of .37 °C when using a slightly different definition (i.e., 1620–1650) for the LIA. Our study suggest that the relative warmth in the Arctic Atlantic is more than double that of the previous studies: .86 °C ($.76 \pm .28$) when using the time periods of Frank et al. (2010) and .91 °C ($.81 \pm .27$) when using the time period of Goosse et al. (2012) for LIA.

In our reconstruction, the MCA in total lasted approximately from 800 to 1200 AD (peak warming around 1000 AD), while LIA was exceptionally long in duration, from 1250 to 1900 AD, with the coolest phase between 1600 and 1900 AD. In literature, the duration of the MCA varies greatly: 950–1100 AD (Mann et al. 2008), 950–1250 AD (Mann et al. 2009; Goosse et al. 2012), 850–1350 AD (Patterson et al. 2010), 800–1300 AD (Jungclaus 2009), 900–1350 AD (Graham et al. 2011), 950–1200 AD (Miller et al. 2010), and 950–1400 AD (Diaz et al. 2011). According to current understanding, the MCA was not a universally warm epoch (see, however, Graham et al. 2011), although much of the Arctic was warm during the Medieval times. In contrast, the LIA, which most studies date between 1500 and 1850 AD, was most probably a global phenomenon (Mann et al. 2009). In our proxy data, the coldest temperatures of the LIA were experienced in first part of the nineteenth century, in accordance with Overpeck et al. (1997).

The most pronounced warming phase in our reconstruction occurred between 1900 and 1940, which is clearly seen in the measured meteorological records as well. In the instrumental record, positive SAT anomalies were largest in the Arctic Atlantic region during this period (Wood and Overland 2010). This early twentieth-century warming (ETCW) has been subject to many studies, yet its reasons still defies full explanation. Natural and anthropogenic (land-use, aerosols) forcings are believed to have contributed to the ETCW (e.g., Delworth and Knutson 2000; Bengtsson et al. 2004; Brönnimann 2009). According to Chylek et al. (2009), the Arctic warming from 1900 to 1940 proceeded at a significantly faster rate than the warming during the more recent decades and was highly correlated with the Atlantic Multi-decadal Oscillation (AMO) suggesting that the Arctic temperature variability is highly linked to the Atlantic Ocean thermohaline circulation at various temporal scales.

Acknowledgments This study is part of the IGBP PAGES Arctic2k project. We thank IGBP PAGES for supporting the Arctic2k workshops. We thank the reviewers for their suggestions and comments. We thank Darrell Kaufman, Dmitry Divine, Fredrik Charpentier Ljungqvist, Håkan Grudd, and Samuli Helama for helping to collect proxy records. We also thank all the individual authors who provided us with records.

Authors' contribution SH and AK designed the overall research. SH designed and implemented PaiCo; assembled data sets and carried out all analyzes; and was primarily responsible for preparation of the manuscript. MPT contributed to the design of the quantitative comparisons and interpretation of results, and to the preparation of the manuscript. AK contributed to the interpretation of results and to the preparation of the manuscript.

References

Ammann CM, Genton MG, Li B (2010) Technical note: correcting for signal attenuation from noisy proxy data in climate reconstructions. Clim Past 6(2):273–279. doi:10.5194/cp-6-273-2010

Anchukaitis KJ, Breitenmoser P, Briffa KR, Buchwal A, Büntgen U, Cook ER, D'Arrigo RD, Esper J, Evans MN, Frank D, Grudd H, Gunnarson BE, Hughes MK, Kirdyanov AV, Körner C, Krusic PJ, Luckman B, Melvin TM, Salzer MW, Shashkin AV, Timmreck C, Vaganov EA, Wilson RJS (2012) Tree rings and volcanic cooling. Nat Geosci 5(12):836–837. doi:10.1038/ngeo1645

Bengtsson L, Semenov VA, Johannessen OM (2004) The early twentieth-century warming in the arctic—a possible mechanism. J Clim 17(20):4045–4057. doi:10.1175/1520-0442(2004017<4045:TETWIT>2.0.CO;2)

Bergthórsson P (1969) An estimate of drift ice and temperature in Iceland in 1,000 years. Jökull 19:94–101

Berner KS, Koç N, Godtliebsen F, Divine D (2011) Holocene climate variability of the Norwegian Atlantic current during high and low solar insolation forcing. Paleoceanography 26. doi:10.1029/2010PA002002

Bingham NH, Fry JM (2010) Regression: linear models in statistics. Springer, London

Bonnans JF (2006) Numerical optimization theoretical and practical aspects. Springer, Berlin

Bradley RS, Briffa KR, Cole J, Hughes MK, Osborn TJ (2003) The climate of the last millennium. In: Alverson K, Bradley RS, Pedersen TF (eds) Paleoclimate, global change, and the future. Springer, Berlin, pp 105–141

Brohan P, Kennedy JJ, Harris I, Tett SFB, Jones PD (2006) Uncertainty estimates in regional and global observed temperature changes: a new data set from 1850. J Geophys Res 111(D12). doi:10.1029/2005JD006548

Brönnimann S (2009) Early twentieth-century warming. Nat Geosci 2(11):735–736. doi:10.1038/ngeo670

Calvo E, Grimalt J, Jansen E (2002) High resolution U37K sea surface temperature reconstruction in the Norwegian sea during the Holocene. Quat Sci Rev 21(12–13):1385–1394. doi:10.1016/S0277-3791(01)00096-8

Christiansen B, Ljungqvist FC (2011) Reconstruction of the extra-tropical NH mean temperature over the last millennium with a method that preserves low-frequency variability. J Clim 24(23):6013–6034. doi:10.1175/2011JCLI4145.1

Chylek P, Folland CK, Lesins G, Dubey MK, Wang M (2009) Arctic air temperature change amplification and the Atlantic multidecadal oscillation. Geophys Res Lett 36:5. doi:200910.1029/2009GL038777

Cook ER, Briffa KR, Meko DM, Graybill DA, Funkhouser G (1995) The 'segment length curse' in long tree-ring chronology development for palaeoclimatic studies. Holocene 5(2):229–237. doi:10.1177/095968369500500211

Delworth TL, Knutson TR (2000) Simulation of early 20th century global warming. Science 287(5461):2246–2250. doi:10.1126/science.287.5461.2246

Denton G, Alley R, Comer G, Broecker W (2005) The role of seasonality in abrupt climate change. Quat Sci Rev 24(10–11):1159–1182. doi:10.1016/j.quascirev.2004.12.002

Diaz HF, Trigo R, Hughes MK, Mann ME, Xoplaki E, Barriopedro D (2011) Spatial and temporal characteristics of climate in medieval times revisited. Bull Am Meteorol Soc 92(11):1487–1500. doi:10.1175/BAMS-D-10-05003.1

Divine D, Isaksson E, Martma T, Meijer HA, Moore J, Pohjola V, van de Wal RS, Godtliebsen F (2011) Thousand years of winter surface air temperature variations in Svalbard and northern Norway reconstructed from ice-core data. Polar Res 30. doi:10.3402/polar.v30i0.7379

Evans MN, Reichert BK, Kaplan A, Anchukaitis KJ, Vaganov EA, Hughes MK, Cane MA (2006) A forward modeling approach to paleoclimatic interpretation of tree-ring data. J Geophys Res 111(G3):G03,008. doi:10.1029/2006JG000166

Frank DC, Esper J, Raible CC, Büntgen U, Trouet V, Stocker B, Joos F (2010) Ensemble reconstruction constraints on the global carbon cycle sensitivity to climate. Nature 463(7280):527–530. doi:10.1038/nature08769

Gildor H, Tziperman E (2003) Sea-ice switches and abrupt climate change. Philos Trans R Soc A 361(1810):1935–1944. doi:10.1098/rsta.2003.1244

Goosse H, Crespin E, Dubinkina S, Loutre MF, Mann M, Renssen H, Sallaz-Damaz Y, Shindell D (2012) The role of forcing and internal dynamics in explaining the "Medieval climate anomaly". Clim Dyn:1–20. doi:10.1007/s00382-012-1297-0

Graham N, Ammann C, Fleitmann D, Cobb K, Luterbacher J (2011) Support for global climate reorganization during the "Medieval climate anomaly". Clim Dyn 37(5):1217–1245. doi:10.1007/s00382-010-0914-z

Grootes PM, Stuiver M (1997) Oxygen 18/16 variability in Greenland snow and ice with 10^3- to 10^5-year time resolution. J Geophys Res 102(C12):26455–26470. doi:199710.1029/97JC00880

Grudd H (2008) Torneträsk tree-ring width and density ad 500–2004: a test of climatic sensitivity and a new 1500-year reconstruction of north Fennoscandian summers. Clim Dyn 31:843–857. doi:10.1007/s00382-007-0358-2

Gunnarson BE, Linderholm HW, Moberg A (2010) Improving a tree-ring reconstruction from west-central Scandinavia: 900 years of warm-season temperatures. Clim Dyn 36:97–108. doi:10.1007/s00382-010-0783-5

Haltia-Hovi E, Saarinen T, Kukkonen M (2007) A 2000-year record of solar forcing on varved lake sediment in eastern Finland. Quat Sci Rev 26(5–6):678–689. doi:10.1016/j.quascirev.2006.11.005

Helama S, Fauria MM, Mielikäinen K, Timonen M, Eronen M (2010) Sub-milankovitch solar forcing of past climates: mid and late Holocene perspectives. Geol Soc Am Bull 122(11–12):1981–1988. doi:10.1130/B30088.1

Isaksson E, Divine D, Kohler J, Martma T, Pohjola V, Motoyama H, Watanabe O (2005) Climate oscillations as recorded in Svalbard ice core omega18O records between AD 1200 and 1997. Geografiska Annaler Ser A 87:203–214. doi:10.1111/j.0435-3676.2005.00253.x

Jiang H, Eiriksson J, Schulz M, Knudsen KL, Seidenkrantz MS (2005) Evidence for solar forcing of sea-surface temperature on the north Icelandic shelf during the late Holocene. Geology 33(1):73–76. doi:10.1130/G21130.1

Jungclaus JH (2009) Palaeoclimate: lessons from the past millennium. Nat Geosci 2(7):468–470. doi:10.1038/ngeo559

Kaspi Y, Sayag R, Tziperman E (2004) A "triple sea-ice state" mechanism for the abrupt warming and synchronous ice sheet collapses during Heinrich events. Paleoceanography 19:12. doi:200410.1029/2004PA001009

Kaufman DS, Schneider DP, McKay NP, Ammann CM, Bradley RS, Briffa KR, Miller GH, Otto-Bliesner BL, Overpeck JT, Vinther BM, Abbott M, Axford Y, Bird B, Birks HJB, Bjune AE, Briner J, Cook T, Chipman M, Francus P, Gajewski K, Geirsdottir A, Hu FS, Kutchko B, Lamoureux S, Loso M, MacDonald G, Peros M, Porinchu D, Schiff C, Seppa H, Thomas E (2009) Recent warming reverses long-term Arctic cooling. Science 325(5945):1236–1239. doi:10.1126/science.1173983

Kirchhefer AJ (2001) Reconstruction of summer temperatures from tree-rings of scots pine (pinus sylvestris l.) in coastal northern Norway. Holocene 11(1):41–52. doi:10.1191/095968301670181592

Korhola A, Weckström J, Holmström L, Erästö P (2000) A quantitative Holocene climatic record from diatoms in northern Fennoscandia. Quat Res 54(2):284–294. doi:10.1006/qres.2000.2153

Kotlyakov V, Arkhipov S, Henderson K, Nagornov O (2004) Deep drilling of glaciers in Eurasian Arctic as a source of

paleoclimatic records. Quat Sci Rev 23(11–13):1371–1390. doi:10.1016/j.quascirev.2003.12.013

Lamb HH (1995) Climate, history and the modern world. Routledge, London

Lee TCK, Zwiers FW, Tsao M (2008) Evaluation of proxy-based millennial reconstruction methods. Clim Dyn 31(2–3):263–281. doi:10.1007/s00382-007-0351-9

Li B, Nychka DW, Ammann CM (2010) The value of multiproxy reconstruction of past climate. J Am Stat Assoc 105(491):883–895. doi:10.1198/jasa.2010.ap09379

Li C, Battisti DS, Schrag DP, Tziperman E (2005) Abrupt climate shifts in Greenland due to displacements of the sea ice edge. Geophys Res Lett 32:4. doi:200510.1029/2005GL023492

Linge H, Lauritzen SE, Andersson C, Hansen JK, Skoglund RO, Sundqvist HS (2009) Stable isotope records for the last 10,000 years from Okshola cave (Fauske, northern Norway) and regional comparisons. Clim Past 5:667–682. doi:10.5194/cp-5-667-2009

Lipovetsky S, Conklin WM (2004) Thurstone scaling via binary response regression. Stat Methodol 1(1–2):93–104. doi:10.1016/j.statmet.2004.04.001

Ljungqvist FC (2010) A new reconstruction of the temperature variability in the extra-tropical northern hemisphere during the last two millenia. Geografiska Ann Ser A Phys Geogr 92(3):339–351. doi:10.1111/j.1468-0459.2010.00399.x

Luoto TP, Sarmaja-Korjonen K, Nevalainen L, Kauppila T (2009) A 700 year record of temperature and nutrient changes in a small eutrophied lake in southern Finland. Holocene 19(7):1063–1072. doi:10.1177/0959683609341002

Maidment D (1993) Handbook of hydrology, 1st edn. McGraw-Hill Professional, New York

Mann ME, Rutherford S, Wahl E, Ammann C (2007) Robustness of proxy-based climate field reconstruction methods. J Geophys Res 112(D12). doi:10.1029/2006JD008272

Mann ME, Zhang Z, Hughes MK, Bradley RS, Miller SK, Rutherford S, Ni F (2008) Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. Proc Natl Acad Sci 105(36):13252–13257. doi:10.1073/pnas.0805721105

Mann ME, Zhang Z, Rutherford S, Bradley RS, Hughes MK, Shindell D, Ammann C, Faluvegi G, Ni F (2009) Global signatures and dynamical origins of the little ice age and medieval climate anomaly. Science 326(5957):1256–1260. doi:10.1126/science.1177303

Mann ME, Fuentes JD, Rutherford S (2012a) Reply to 'Tree rings and volcanic cooling'. Nature Geosci 5(12):837–838. doi:10.1038/ngeo1646

Mann ME, Fuentes JD, Rutherford S (2012b) Underestimation of volcanic cooling in tree-ring-based reconstructions of hemispheric temperatures. Nat Geosci. doi:10.1038/ngeo1394

Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. SIAM J Appl Math 11(2):431. doi:10.1137/0111030

Marshall J, Kushnir Y, Battisti D, Chang P, Czaja A, Dickson R, Hurrell J, McCartney M, Saravanan R, Visbeck M (2001) North Atlantic climate variability: phenomena, impacts and mechanisms. Int J Climatol 21(15):1863–1898. doi:10.1002/joc.693

McDermott F, Mattey DP, Hawkesworth C (2001) Centennial-scale Holocene climate variability revealed by a high-resolution speleothem δ18O record from SW Ireland. Science 294(5545):1328–1331. doi:10.1126/science.1063678

McKay NP, Kaufman DS, Michelutti N (2008) Biogenic silica concentration as a high-resolution, quantitative temperature proxy at Hallet lake, south-central Alaska. Geophys Res Lett 35:6. doi:200810.1029/2007GL032876

Miller G, Brigham-Grette J, Alley R, Anderson L, Bauch H, Douglas M, Edwards M, Elias S, Finney B, Fitzpatrick J, Funder S, Herbert T, Hinzman L, Kaufman D, MacDonald G, Polyak L, Robock A, Serreze M, Smol J, Spielhagen R, White J, Wolfe A, Wolff E (2010) Temperature and precipitation history of the Arctic. Quat Sci Rev 29(15–16):1679–1715. doi:10.1016/j.quascirev.2010.03.001

Moberg A, Sonechkin DM, Holmgren K, Datsenko NM, Karlen W (2005) Highly variable northern hemisphere temperatures reconstructed from low- and high-resolution proxy data. Nature 433(7026):613–617. doi:10.1038/nature03265

O'Brien SR, Mayewski PA, Meeker LD, Meese DA, Twickler MS, Whitlow SI (1995) Complexity of Holocene climate as reconstructed from a Greenland ice core. Science 270(5244):1962–1964. doi:10.1126/science.270.5244.1962

Ojala AE, Alenius T (2005) 10000 years of interannual sedimentation recorded in the lake Nautajärvi (finland) clastic-organic varves. Palaeogeogr Palaeoclimatol Palaeoecol 219:285–302. doi:10.1016/j.palaeo.2005.01.002

Overpeck J, Hughen K, Hardy D, Bradley R, Case R, Douglas M, Finney B, Gajewski K, Jacoby G, Jennings A, Lamoureux S, Lasca A, MacDonald G, Moore J, Retelle M, Smith S, Wolfe A, Zielinski G (1997) Arctic environmental change of the last four centuries. Science 278(5341):1251–1256. doi:10.1126/science.278.5341.1251

Patterson WP, Dietrich KA, Holmden C, Andrews JT (2010) Two millennia of north Atlantic seasonality and implications for Norse colonies. Proc Nat Acad Sci 107(12):5306–5310. doi:10.1073/pnas.0902522107

Pflaumann U, Duprat J, Pujol C, Labeyrie LD (1996) SIMMAX: a modern analog technique to deduce atlantic sea surface temperatures from planktonic foraminifera in deep-sea sediments. Paleoceanography 11(1):15. doi:10.1029/95PA01743

Schneider T (2001) Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. J Clim 14:853–871

Schwager M (1999) Ice core analysis on the spatial and temporal variability of temperature and precipitation during the late Holocene in north Greenland. PhD thesis, Alfred-Wegener-Institut für Polar- und Meeresforschung

Sejrup H, Haflidason H, Andrews J (2011) A Holocene north Atlantic SST record and regional climate variability. Quat Sci Rev 30(21–22):3181–3195. doi:10.1016/j.quascirev.2011.07.025

Sicre MA, Hall IR, Mignot J, Khodri M, Ezat U, Truong MX, Eiríksson J, Knudsen KL (2011) Sea surface temperature variability in the subpolar Atlantic over the last two millennia. Paleoceanography. doi:10.1029/2011PA002169

Spielhagen RF, Werner K, Sørensen SA, Zamelczyk K, Kandiano E, Budeus G, Husum K, Marchitto TM, Hald M (2011) Enhanced modern heat transfer to the Arctic by warm Atlantic water. Science 331(6016):450–453. doi:10.1126/science.1197397

Stern H (1990) A continuum of paired comparisons models. Biometrika 77(2):265–273. doi:10.1093/biomet/77.2.265

Sundqvist HS, Zhang Q, Moberg A, Holmgren K, Körnich H, Nilsson J, Brattström G (2010) Climate change between the mid and late Holocene in northern high latitudes—part 1: survey of temperature and precipitation proxy data. Clim Past 6(5):591–608. doi:10.5194/cp-6-591-2010

Thurstone LL (1927) A law of comparative judgment. Psychol Rev 34(4):273–286. doi:10.1037/h0070288

Tiljander M, Saarnisto M, Ojala AEK, Saarinen T (2003) A 3000-year palaeoenvironmental record from annually laminated sediment of lake Korttajärvi, Central Finland. Boreas 32(4):566–577. doi:10.1111/j.1502-3885.2003.tb01236.x

Timmermann A, An SI, Krebs U, Goosse H (2005) ENSO suppression due to weakening of the North Atlantic thermohaline circulation. J Clim 18(16):3122–3139. doi:10.1175/JCLI3495.1

Tingley MP (2012) A Bayesian ANOVA scheme for calculating climate anomalies, with applications to the instrumental temperature record. J Clim 25:777–791. Code is available at ftp://ftp.ncdc.noaa.gov/pub/data/paleo/softlib/anova

Tingley MP, Huybers P (2010a) A Bayesian algorithm for reconstructing climate anomalies in space and time. Part I: development and applications to paleoclimate reconstruction problems. J Clim 23(10):2759–2781. doi:10.1175/2009JCLI3015.1

Tingley MP, Huybers P (2010b) A Bayesian algorithm for reconstructing climate anomalies in space and time. Part II: comparison with the regularized expectation–maximization algorithm. J Clim 23(10):2782–2800. doi:10.1175/2009JCLI3016.1

Tingley MP, Craigmile PF, Haran M, Li B, Mannshardt E, Rajaratnam B (2012) Piecing together the past: statistical insights into paleoclimatic reconstructions. Quat Sci Rev 35:1–22. doi:10.1016/j.quascirev.2012.01.012

Tolwinski-Ward SE, Evans MN, Hughes MK, Anchukaitis KJ (2011) An efficient forward model of the climate controls on interannual variation in tree-ring width. Clim Dyn 36(11–12):2419–2439. doi:10.1007/s00382-010-0945-5

Vinther B, Jones P, Briffa K, Clausen H, Andersen K, Dahl-Jensen D, Johnsen S (2010) Climatic signals in multiple highly resolved stable isotope records from Greenland. Quat Sci Rev 29(3–4):522–538. doi:10.1016/j.quascirev.2009.11.002

Vinther BM, Clausen HB, Johnsen SJ, Rasmussen SO, Andersen KK, Buchardt SL, Dahl-Jensen D, Seierstad IK, Siggaard-Andersen ML, Steffensen JP, Svensson A, Olsen J, Heinemeier J (2006) A synchronized dating of three Greenland ice cores throughout the Holocene. J Geophys Res 111:11. doi:200610.1029/2005JD006921

Vinther BM, Clausen HB, Fisher DA, Koerner RM, Johnsen SJ, Andersen KK, Dahl-Jensen D, Rasmussen SO, Steffensen JP, Svensson AM (2008) Synchronizing ice cores from the Renland and Agassiz ice caps to the Greenland ice core chronology. J Geophys Res 113. doi:10.1029/2007JD009143

Wang Y, Cheng H, Edwards RL, He Y, Kong X, An Z, Wu J, Kelly MJ, Dykoski CA, Li X (2005) The Holocene asian monsoon: links to solar changes and north atlantic climate. Science 308(5723):854–857. doi:10.1126/science.1106296

Wanner H, Beer J, Bütikofer J, Crowley TJ, Cubasch U, Flückiger J, Goosse H, Grosjean M, Joos F, Kaplan JO, Küttel M, Müller SA, Prentice IC, Solomina O, Stocker TF, Tarasov P, Wagner M, Widmann M (2008) Mid- to late holocene climate change: an overview. Quat Sci Rev 27(19–20):1791–1828. doi:10.1016/j.quascirev.2008.06.013

Welch P (1967) The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans Audio Electroacoust 15(2):70–73. doi:10.1109/TAU.1967.1161901

Wood KR, Overland JE (2010) Early 20th century arctic warming in retrospect. Int J Climatol 30(9):1269–1279. doi:10.1002/joc.1973

Wunsch C (2006) Abrupt climate change: an alternative view. Quat Res 65(2):191–203. doi:10.1016/j.yqres.2005.10.006