Climate
of the Past

Spatio-temporal variability of Arctic summer temperatures over the past 2 millennia

Johannes P. Werner[1], **Dmitry V. Divine**[2,3], **Fredrik Charpentier Ljungqvist**[4,5], **Tine Nilsen**[3], and **Pierre Francus**[6,7]

[1]Bjerknes Centre for Climate Research and Department for Earth Science, University of Bergen,
P.O. Box 7803, 5020 Bergen, Norway
[2]Norwegian Polar Institute, Fram Centre, 9296 Tromsø, Norway
[3]Department of Mathematics and Statistics, University of Tromsø – The Arctic University of Norway, 9037, Norway
[4]Department of History, Stockholm University, 106 91 Stockholm, Sweden
[5]Bolin Centre for Climate Research, Stockholm University, 106 91 Stockholm, Sweden
[6]Centre – Eau Terre Environnement, Institut National de la Recherche Scientifique, 490 rue de la couronne,
Québec, QC G1K 9A9, Canada
[7]GEOTOP Research Center, Montréal, H3C 3P8, Canada

Correspondence: Johannes P. Werner (jopewerner@googlemail.com)

Abstract. In this article, the first spatially resolved and millennium-length summer (June–August) temperature reconstruction over the Arctic and sub-Arctic domain (north of 60° N) is presented. It is based on a set of 44 annually dated temperature-sensitive proxy archives of various types from the revised PAGES2k database supplemented with six new recently updated proxy records. As a major advance, an extension of the Bayesian BARCAST climate field (CF) reconstruction technique provides a means to treat climate archives with dating uncertainties. This results not only in a more precise reconstruction but additionally enables joint probabilistic constraints to be imposed on the chronologies of the used archives. The new seasonal CF reconstruction for the Arctic region can be shown to be skilful for the majority of the terrestrial nodes. The decrease in the proxy data density back in time, however, limits the analyses in the spatial domain to the period after 750 CE, while the spatially averaged reconstruction covers the entire time interval of 1–2002 CE.

The centennial to millennial evolution of the reconstructed temperature is in good agreement with a general pattern that was inferred in recent studies for the Arctic and its subregions. In particular, the reconstruction shows a pronounced Medieval Climate Anomaly (MCA; here ca. 920–1060 CE), which was characterised by a sequence of extremely warm decades over the whole domain. The medieval warming was followed by a gradual cooling into the Little Ice Age (LIA), with 1766–1865 CE as the longest centennial-scale cold period, culminating around 1811–1820 CE for most of the target region.

In total over 600 independent realisations of the temperature CF were generated. As showcased for local and regional trends and temperature anomalies, operating in a probabilistic framework directly results in comprehensive uncertainty estimates, even for complex analyses. For the presented multi-scale trend analysis, for example, the spread in different paths across the reconstruction ensemble prevents a robust analysis of features at timescales shorter than ca. 30 years. For the spatial reconstruction, the benefit of using the spatially resolved reconstruction ensemble is demonstrated by focusing on the regional expression of the recent warming and the MCA. While our analysis shows that the peak MCA summer temperatures were as high as in the late 20th and early 21st centuries, the spatial coherence of extreme years over the last decades of the reconstruction (1980s onwards) seems unprecedented at least back until 750 CE. However, statistical testing could not provide conclusive support of the contemporary warming to exceed the peak of the MCA in terms of the pan-Arctic mean summer temperatures: the reconstruction cannot be extended reliably past 2002 CE

due to lack of proxy data and thus the most recent warming is not captured.

1 Introduction

During the past decades, the Arctic has experienced a more rapid and pronounced temperature increase than most other parts of the world. The dramatically shrinking extent of Arctic sea ice in recent years – with a decline in both minimum extent in summer and maximum area in winter – accompanied by a transition to a younger and thinner sea ice cover, is often interpreted as the clearest and most unambiguous evidence of anthropogenic global warming (Comiso et al., 2008; Perovich et al., 2008; Serreze et al., 2007; Maslanik et al., 2011; Meier et al., 2014). Additionally, the Arctic region is of utmost importance in the context of global climate and global climate change. Reduction in perennial sea ice cover leads to increased heat transport northward (Müller et al., 2012; Smedsrud et al., 2008) as well as changes the Arctic energy balance due to positive albedo feedbacks (Curry et al., 1995; Miller et al., 2010; Perovich et al., 2002, 2011). Melting of permafrost can release methane (CH_4), a more efficient greenhouse gas than carbon dioxide (CO_2), and likewise gives a positive feedback that may further amplify the temperature increase (O'Connor et al., 2010; Shakhova et al., 2010). Even partial melting of the Greenland inland ice cap and/or the numerous smaller high-latitude glaciers would significantly raise the global sea level and threaten to flood low-laying coastal regions around the world (Grinsted et al., 2010; Vermeer and Rahmstorf, 2009).

The instrumental temperature record is too short and spatially sparse to assess whether this recent warming and the accompanying sea ice reduction experienced in the Arctic region so far fall outside the range of natural variability on centennial to millennial timescales. Moreover, general circulation models have limited capabilities in reliably simulating Arctic climate change on the centennial timescale and beyond (IPCC, 2013). The simplified parametrisation of dynamic and thermodynamic sea ice processes, and the limited skills in describing ocean – sea ice – atmosphere energy exchange, in particular in modelling polar clouds and oceanic heat flux, is especially evident from the lack of skill in reproducing the present-day rapid loss of Arctic sea ice (e.g. Hunke et al., 2010). Hence both the possible unprecedented nature of the ongoing Arctic warming during the Common Era (CE, the last 2000 years) and the relative role of anthropogenic and natural forcings driving the process are difficult to fully assess without a longer perspective from palaeoclimate proxy-based temperature reconstructions. Thus palaeoclimate data that can be used for understanding the range of natural climate variability in the Arctic region over long timescales are needed, together with methods that integrate different types of information from a variety of palaeoclimate archives.

Since the 1990s, several multi-proxy reconstructions of Arctic and sub-Arctic (usually 90–60° N) temperatures have been published. The first one of those was the multi-proxy reconstruction by Overpeck et al. (1997), who compiled 29 proxy records from lake sediment, tree ring, glacier, and marine sediment records to present a decadally resolved uncalibrated index of temperature variability since 1600 CE. They found that the highest temperatures in the Arctic region since 1600 CE occurred after 1920 CE. Kaufman et al. (2009) published the first quantitative multi-proxy reconstruction of summer temperature variability in the Arctic (90–60° N) during the past 2000 years at decadal resolution using the composite-plus-scaling method. This study concluded that the 20th century warming reverses a long-term orbitally driven summer cooling and that the mid- and late 20th century temperatures were the highest in the past 2 millennia.

Shi et al. (2012) published the first annually resolved multi-proxy summer (June–August) temperature reconstruction for the Arctic region, extending back to 600 CE, based on a set of 22 proxy records with annual resolution. They utilised a novel ensemble reconstruction method that combined the traditional composite-plus-scale method – known to underestimate low-frequency variability (e.g. von Storch et al., 2004) – and the LOC (local regression) method of Christiansen (2011) that exaggerates the high-frequency variability (see e.g. Christiansen and Ljungqvist, 2017). The reconstructed amplitude of the centennial-scale summer temperature variability was rather dampened and found to be less than 0.5 °C but with large year-to-year and decade-to-decade variability. Shi et al. (2012) found a clear cold anomaly from 630 to 770 CE, a peak warming from ca. 950 to 1050 CE, and overall relatively cold conditions around ca. 1200–1900 CE. However, three distinctly warmer periods during the Little Ice Age (LIA) were reconstructed around ca. 1470–1510, 1550–1570, and 1750–1770 CE. Contrary to Kaufman et al. (2009), Shi et al. (2012) found peak medieval Arctic summer temperatures in the 10th century to be approximately equal to recent Arctic summer temperatures.

Tingley and Huybers (2013) used BARCAST (Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time; Tingley and Huybers, 2010a), a method based on Bayesian inference of hierarchical models (see also Sect. 3), to reconstruct surface air temperatures of the last 600 years over land north of 60° N. Their reconstruction is mostly based on the proxy data set collected by the PAGES 2k Consortium (2013). They found that while the recent decades were the warmest over the last 600 years, the actual interannual variability has remained effectively constant. Much of the data (most of the ice core and lake sediment records) used therein are common with the work presented here, with a few updated records (see Sect. 2.2 and PAGES 2k Consortium, 2017) and a few additional proxies.

Hanhijärvi et al. (2013) presented a 2000-year-long annual mean temperature reconstruction for the North Atlantic sec-

tor of the Arctic (north of 60° N and between 50° W and 30° E) using 27 proxy records of various types, resolution, and length employing the novel pairwise comparison (PaiCo) method. Their reconstruction reveals centennial-scale temperature variations of an amplitude of over 1 °C, with a distinct Roman Warm Period, warm Medieval Climate Anomaly (MCA), and 20th century warming. A somewhat indistinct Dark Age Cold Period is found in the middle of the first millennium CE, whereas a very clear and persistently cold LIA extends from the mid-13th century until the turn of the 20th century, with the lowest temperatures in the 19th century. Peak temperatures during the Roman Warm Period and the MCA were found to equal recent temperatures in the North Atlantic sector of the Arctic. The PAGES 2k Consortium (2013) extended the PaiCo reconstruction to cover the whole Arctic (60–90° N), using 67 proxy records of various types, resolution, and length to reconstruct annual mean temperature variations over the past 2 millennia. They reconstructed a generally relatively warm first millennium CE, followed by a relatively indistinct MCA, and a relatively cold LIA from ca. 1250 CE to 1900 CE. The amplitude of the reconstructed low-frequency temperature variability in the whole Arctic by the PAGES 2k Consortium (2013) is smaller than that reconstructed for only the North Atlantic sector of the Arctic by Hanhijärvi et al. (2013). A revised Arctic2k reconstruction was subsequently published by McKay and Kaufman (2014), using an updated and corrected proxy database containing 59 records, showing a larger long-term cooling trend and being on average ca. 0.5 °C warmer prior to ca. 1250 CE than reported by PAGES 2k Consortium (2013). Peak temperatures during the Roman Warm Period and the MCA thus approximately equal recent temperatures in McKay and Kaufman (2014) as in Shi et al. (2012) and Hanhijärvi et al. (2013), instead of being much lower as in the Arctic2k reconstruction by the PAGES 2k Consortium (2013).

This study is mostly comparable with that of Tingley and Huybers (2013): our method is an update of theirs (Tingley and Huybers, 2010a; Werner and Tingley, 2015), and the proxy network is an update of the PAGES2k database (PAGES 2k Consortium, 2017). There are, however, a few notable differences: (i) the climate field (CF) reconstruction is performed on an equal area grid (land only), which should be more suitable for a spatially homogeneous process – especially at high latitudes. (ii) This target gridded instrumental data set is directly derived from meteorological observation data without any interpolation over grid cell boundaries. (iii) The gridded reconstruction is reliable back to 750 CE, and in principle goes back into the first millennium CE. (iv) The proxy data set is larger and more extensively screened, and (v) the age uncertainties of the proxies used are respected. Thus, the propagation of uncertainties from proxy data to the final reconstruction product is more complete. (vi) Additionally, while Tingley and Huybers (2013) use a single set of parameters for all proxies of one type, these are estimated here for each individual record.

This potentially removes spurious precision at proxy sites responding less strongly to the seasonal temperature anomalies and should increase the precision at locations with a stronger climate response.

2 Instrumental data and proxy data

The following section provides a short overview of the instrumental and palaeoclimate proxy data used. The quality of the input data and their distribution in space and time play a strong role in the reconstruction process and for the reconstruction reliability (see e.g. Wang et al., 2015).

2.1 Instrumental data

Several different gridded data sets for Earth surface air temperatures (SATs) are available from different research groups, derived from different subsets of instrumental data and presented on different types of grids. Most data sets, like CRUTEM4 (Jones et al., 2012) or CRU TS3 (Harris et al., 2014), for example, are presented on a regular equilateral grid, such as a 5° × 5° grid. Such a regular grid exhibits severe shortcomings when analysing data close to the poles, as the grid cells become very narrow in the meridional direction and almost triangular shaped. One data set, the Berkeley Earth Surface Temperature (BEST) (Rohde et al., 2013), is offered on a 1° × 1° grid as well as an equal area grid. While the latter would be a good fit for the process level model of BARCAST (see Sect. 3), an analysis revealed that this version of the data set shows long-distance correlations over our region of interest that might be deemed unphysical: the correlation length is on the order of the region size and the cross correlation between grid cells contains oscillatory parts with respect to the distance between the cells. The latter, especially, might be artefacts of the regridding and interpolation process.

Thus a new gridded instrumental data set is generated for this study. The instrumental data for the CRU TS3.24 (Harris et al., 2014) data set were downloaded from the CRU website (http://browse.ceda.ac.uk/browse/badc/cru/data/cru_ts/cru_ts_3.24.01, last access: 7 March 2017). First, the data were converted into anomalies for the period 1801–2016 CE, using the method of Tingley (2012). The equal area target grid is taken from Leopardi (2006). To construct the gridded data, the instrumental data within each grid cell were averaged, using the variance adjustment scheme described by Frank et al. (2006). We aimed at retaining the variability that a single instrumental record in the grid cell would exhibit. This is a compromise between an actual grid-cell-wide average and the limited spatio-temporal availability of instrumental data in high latitudes. In contrast to other regridding methods, no data were shared across grid cells by a prescribed spatial covariance structure or spatial interpolation algorithm. In contrast to Tingley and Huybers (2010a, b) and Werner et al. (2013) the instrumental data were not nor-
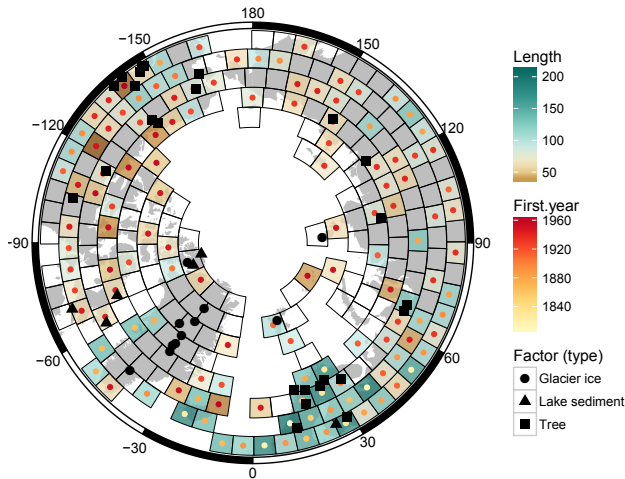
Figure 1. Distribution of input data. Length (fill of quadrilaterals) and first year (coloured circles) of the regridded instrumental data. Symbols show the locations and type of proxy data used (PAGES 2k Consortium, 2017). The reconstruction target area is all grid cells marked with wire frames.

malised to have unit standard deviation prior to running the BARCAST sampler.

As can be seen in Fig. 1, the resulting instrumental data set is very sparse in space and time. While ordinary reconstruction methods would indeed struggle with such input data, the advantage of the BARCAST method and the extension used here is that presence and absence of observations are explicitly modelled. The reconstruction target region is grid cells (wire frames in Fig. 1) that only contain land mass (continent and islands). This is necessary due to the constraints of the chosen reconstruction method (Tingley and Huybers, 2010a; Werner and Tingley, 2015), more specifically due the homogeneous process level model, which describes the temperature evolution on the grid cell level.

2.2 Proxy data

The proxy records (black symbols in Fig. 1) mostly come from the current version (2.0.0) of the PAGES 2k Consortium (2017) temperature database, with six recently updated ice core records from Greenland with revised and synchronised chronologies. The data set contains several types of natural archives (tree rings, ice cores, and marine or terrestrial sediments) and proxy measurements (such as ring width and stable isotopes). Thus the data are sensitive to different seasons, and on different timescales – partly due to different resolutions and the evaluation procedures but also owed to the processes generating the archives. All data north of 60° N contained in the database were selected, with an a priori aim of including all annually resolved records.

As the PAGES 2k Consortium (2017) set out to generate a very inclusive data set, the need arose to again scrutinise the

data. A few records were excluded (see Table D1), as they did not meet the required response characteristics on actual annual timescales. Additionally, data were divided into two classes: absolutely and precisely dated tree ring chronologies, and layer-counted proxies with age uncertainties. The latter comprises varved lacustrine sediments and ice core data. In contrast to the procedure outlined by Luterbacher et al. (2016), tree ring width measurements are not treated differently from maximum latewood density data, although the spectral properties should in principle warrant this separation (Zhang et al., 2015; Esper et al., 2015; Büntgen et al., 2015).

All of the proxy records used in this study are derived from annually banded archives. While tree ring records are compiled by cross-referencing a number of cores for each period, there is usually very limited replication of ice cores or varved lake sediments. Thus, these archives can (and usually do) contain age uncertainties (see Sigl et al., 2015) which need to be taken into account. Fortunately, the chosen method (Werner and Tingley, 2015) is able to deal with this issue, provided an ensemble of age models is given for each proxy. Appendix D details how these age models are generated. As the majority of the proxy data are more sensitive to summer or growing season temperatures, the target season for the reconstruction is the climatological summer season (the months from June to August, JJA) rather than the annual mean temperature.

3 Reconstruction method: BARCAST+AMS

Werner and Tingley (2015) published an extension to the BARCAST method. It extends the work of Tingley and Huybers (2010a), providing a means to treat climate archives with dating uncertainties. The original method has been used in a collection of pseudo-proxy experiments (Tingley and Huybers, 2010b; Werner et al., 2013; Gómez-Navarro et al., 2015), as well as CF reconstructions over the Arctic (Tingley and Huybers, 2013), Europe (Luterbacher et al., 2016), and Asia (Zhang et al., 2018).

The method uses a hierarchy of stochastic models to describe the spatio-temporal evolution of the target CF (here: temperatures) $C_t \in \mathbb{R}^N$ at N different locations throughout time t, and the dependence of the observations $O_t \in \mathbb{R}^N$ (proxy data as well as instrumental data) on it:

$$C_{t+1} - \mu = \alpha (C_t - \mu) + \epsilon_t$$
$$\epsilon_t \sim \mathcal{N}(\mathbf{0}, \Sigma) \quad \text{(independent)} \tag{1a}$$
$$\Sigma_{i,j} = \sigma^2 \exp\left(-\phi |x_i - x_j|\right).$$

The process level is thus AR(1) (first-order autoregressive) in time, with an overall mean μ and the coefficient α modelling the temporal persistence. The year-to-year (or rather summer-to-summer) innovations have an exponentially (with distance between locations x_i and x_j) decreasing spatial persistence that is homogeneous in space. The spatial e-folding

distance is $1/\phi$. The climate is thus persistent in space and time, and information is shared across these dimensions. This is critical in constraining age models (see discussion in Werner and Tingley, 2015). The climate process $\boldsymbol{C}$ is never directly observed without error (latent process). The observations are modelled as a noisy linear response function:

$$\boldsymbol{O}_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \odot (\mathbf{H}_t \cdot \boldsymbol{C}_t) + \boldsymbol{e}_{O,t}$$

$$\boldsymbol{e}_{O,t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\tau}^2 \odot \mathbf{I}) \quad \text{(independent),}$$

(1b)

where the element-wise product (Hadamard product) is denoted by $\odot$. The parameters $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\tau}^2)$ are vectors and thus different for each location with observations, while in the past one set of parameters was assigned to each proxy type (e.g. tree ring widths, ice layer thickness, or isotopic values) (Tingley and Huybers, 2013). The instrumental observations are assumed to be unbiased and on the correct scale, so that, for this type of observation $\boldsymbol{\beta}_0 = 0$ and $\boldsymbol{\beta}_1 = 1$. The selection matrix $\mathbf{H}_t$ is composed of zeros and ones and selects at time step t the locations for which there are proxy observations of a given type. That is, each proxy observation is assumed to be linear in the corresponding local, in time and space, value of the climate. While interannual temperatures roughly follow a normal distribution in our target region (Tingley and Huybers, 2013), a variable-like varve thickness is positive only. These variables are transformed using inverse quantile transformation (e.g. Emile-Geay and Tingley, 2016) to include them easily in the reconstruction.

This data-level model is then refined to include dating uncertainties. To this end, Werner and Tingley (2015) consider the dependence of the local observations $\boldsymbol{O}_s$ on the local climate:

$$\boldsymbol{O}_s | \mathcal{T}, \boldsymbol{C}_s = \beta_0 + \beta_1 \cdot \boldsymbol{\Lambda}_s^{\mathcal{T}} \cdot \boldsymbol{C}_s + \boldsymbol{e}_s$$

$$\boldsymbol{e}_s \sim \mathcal{N}(\boldsymbol{0}, \tau^2 \cdot \mathbf{I}) \quad \text{(independent).}$$

(1c)

The vector $\boldsymbol{e}_s$ is a time series of independent normal errors at location s (see $\boldsymbol{e}_t$ from Eq. (1b)). In analogy to $\mathbf{H}_t$ in Eq. (1b), $\boldsymbol{\Lambda}_s^{\mathcal{T}}$ is a selection matrix of zeros and ones that picks out the elements of the vector $\boldsymbol{C}_s$ corresponding to elements of $\boldsymbol{O}_s$ and is dependent on the age depth model (ADM) $\mathcal{T}$.

From these model equations, conditional posteriors for the CF and all of the parameters (CF and instrumental or proxy observations) are calculated. Then, a Metropolis-coupled Markov chain Monte Carlo (MC)[3] sampler (Altekar et al., 2004; Earl and Deem, 2005; Li et al., 2009) is used to iteratively draw solutions from these posteriors; see Tingley and Huybers (2010a) and Werner and Tingley (2015) for details and implementations. In the version implemented here, BARCAST is slightly modified. While Tingley and Huybers (2013) used a single set of response parameters $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\tau}^2)$ for all data of one type, and Luterbacher et al. (2016) actually set up a separate observation matrix with a set of parameters for each single proxy, the code is updated for this study. Here, the response parameters of Eq. (1b) are now vectors.

While this slows down the computations and also the convergence there is no good reason to assume that all proxies of one type respond in the same way across the whole domain and with no differences in proxy quality. This has been discussed already by Luterbacher et al. (2016), where two proxies that were initially in the PAGES2k database (PAGES 2k Consortium, 2013) proved to contain no clear temperature signal (see also changes in the updated database PAGES 2k Consortium, 2017) and were thus removed.

The reconstruction code is run in four chains for 8000 iterations without the age model selection code enabled. By then, the chains have clearly settled to a stable state, and the potential scale reduction factor (Gelman's $\hat{R}$) indicates convergence of the parameters ($|\hat{R} - 1| < 0.1$). Then, the (MC)[3] code of Werner and Tingley (2015) is enabled, and the age models are varied. While this was not necessary in the work of Werner and Tingley (2015), the real world data are much sparser and noisier and do not follow the exact prescribed stochastic model Eq. (1a–1c). While this additional step helps speed up convergence it can cause the algorithm to strongly favour one set of age models. This can be checked by analysing the mixing properties over the age models in the heated chains (see discussion in Werner and Tingley, 2015). As noted therein, there is a tradeoff between the switching efficiency and the number of chains. With the infrastructure used, four chains using two cores each (for parallel linear algebra using the OpenBLAS library http://www.openblas.net) were deemed a reasonable compromise.

Reconstruction quality

The reconstruction calibration and validation statistics are shown in Appendix A. It has been shown that some of the commonly used measures, like the coefficient of efficiency and the reduction of error (Cook et al., 1994) are not proper scoring rules and should be avoided in such an ensemble-based probabilistic framework (Gneiting and Raftery, 2007). Thus, reconstruction skill is assessed using the $\overline{\text{CRPS}}_{\text{pot}}$ (potential average continuously ranked probability score, which is akin to the mean absolute error of a deterministic forecast; see Gneiting and Raftery, 2007) as well as the reliability score (the validity of the uncertainty bands; Hersbach, 2000). Additionally, a probabilistic ensemble-based version of the coefficient of efficiency and the reduction of error are constructed from these (see Appendix A).

Both the $\overline{\text{CRPS}}_{\text{pot}}$ as well as the reliability score show a decent reconstruction quality (Fig. A1 top row). The $\overline{\text{CRPS}}_{\text{pot}}$ on average shows a mismatch of $0.2\,^{\circ}\text{C}$ ($0.4\,^{\circ}\text{C}$) in the calibration (validation) interval and the Reliability is mostly better than $0.2\,^{\circ}\text{C}$. This is on the order of the noise strength that the reconstruction code attributed to the instrumental observations. Additionally, the probabilistic ensemble-based version of the coefficient of efficiency and the reduction of error show a skilful reconstruction in most grid cells containing instrumental temperature data – at least in regions where proxy

and instrumental data are present over most of the validation period. Note that the quality of the instrumental data, or rather the representativeness of (often) a single meteorological station record can be debated. In fact, in contrast to other BARCAST-based reconstructions the one presented here shows a substantial ($\tau_I^2 \approx 0.25 \,^\circ\text{C}^2$) noise level for the instrumental data. As other gridded instrumental data sets employ spatial interpolation processes, these are generally smoother in space than the gridded instrumental data set generated for this study. Thus these gridded products are by design closer to the spatial characteristics of the process model in Eq. (1a).

Another means of assessing the reconstruction quality is to check the variability in or spread of the different ensemble members in space and time (see Appendix B). The effect of the spatially and temporally sparse data can easily be seen in Figs. B1 and B2, clearly indicating the increased uncertainties back in time and in space in the absence of proxy data. This analysis hints that while there could still be skill left in the mean Arctic summer temperature reconstruction in the first centuries CE, the precision of the spatial reconstruction rapidly decreases in areas that become more data sparse. While the reconstruction over the regions with local proxy data present – such as Fennoscandia – remains reliable, a time-varying reconstruction domain (or rather, domain over which the reconstruction is analysed) would be beyond the scope of this paper. Thus the gridded reconstruction is only shown back to 750 CE. However, for analyses over data-rich regions such as the North Atlantic sector, the full reconstruction period (1–2002 CE) can in principle be used. For a more uniform reconstruction skill back in time, more high-quality proxy data would be needed. As a first estimate, the distance between neighbouring proxy locations should be less than the spatial correlation length of the system, which is estimated to be around 1500 km.

Additionally, the spectral properties of both the reconstruction and the proxy input data are analysed (see Appendix C). Not all proxies contain signal on centennial or longer timescales, and the reconstruction method explicitly describes year-to-year summer temperatures as an AR(1) process. Despite the reconstruction showing properties of an AR(1) process over most of the reconstruction domain (see also Nilsen et al., 2018) the area-averaged temperature reconstruction exhibits similar variability on centennial and longer timescales as other multi-proxy reconstructions over the Arctic (see Fig. 3).

4   Results

In the following sections the resulting reconstruction is presented. First, the Arctic average is analysed and compared against other studies from the same region. Two periods of interest are identified in the reconstruction before the instrumental period: the warm MCA around 920–1060 CE and the

following LIA, which in this reconstruction culminated in the early 19th century. These then provide a context for the current warming of the Arctic. A detailed analysis of an earlier extended warm period in the fourth and fifth centuries CE is omitted due to a higher uncertainty of the derived reconstruction prior to 750 CE. Yet it is acknowledged that the scales of the detected warming could be comparable to the following episodes that occurred later during the MCA. Finally, the spatial variability in the reconstructed temperature field is explored, with a focus on the most extreme periods.

4.1   Mean Arctic results

The ensemble mean of the area-averaged summer temperature reconstruction is shown in the bottom half of Fig. 2 as the pointwise (year-to-year summers) ensemble mean (heavy blue line). The first millennium CE shows a mean reconstruction that exhibits an apparent change in variability. This is caused by the increased variability between the different ensemble members and thus by the reduction in proxy data coverage back in time. The effect of the spatial proxy data coverage on the reconstruction intra-ensemble variance is further discussed in Appendix B.

The new spatially averaged SAT reconstruction shows a pronounced variability on a broad range of timescales. The longer-term, centennial to millennial, evolution of the reconstructed SAT demonstrates a reasonably good agreement with a general pattern that was inferred in previous temperature reconstructions for the Arctic and its subregions (Fig. 3). Throughout most of the reconstruction period, the Arctic SAT anomaly shows an overall orbital-forcing-driven cooling trend.

Superimposed on the trend are three major centennial to multi-centennial scale anomalies: a warm period in the fourth and fifth centuries CE, the MCA – a warm period with a diverse spatial expression in the Northern Hemisphere during the 9th to 12th centuries CE, and the two phases of the cold LIA between ca. 1100–1450 CE and 1600–1900 CE. Figure 4 further demonstrates that the three aforementioned major climate anomalies together with the most recent period are associated with the likely warmest and coldest centuries in the Arctic over the last 2000 years. In particular, the 17th and 19th centuries CE with (within the uncertainty) similar ranged mean SAT anomalies of $-0.9 \pm 0.1 \,^\circ\text{C}$ appear coldest in the ensemble average (Fig. 4a) and are also ranked coldest in 52 and 25 % of the ensemble members, respectively. While the fifth century with a SAT anomaly of $0.1 \pm 0.2 \,^\circ\text{C}$ appears warmest in 48 % of reconstruction members over the 2000 years, this inference should be considered with caution due to a higher reconstruction uncertainty for this data-sparse pre-750 CE period. For the later periods with better proxy coverage, the 10th century CE, accommodating the MCA with the ensemble average mean SAT anomaly of $0.0 \pm 0.1 \,^\circ\text{C}$, along with the 20th century SAT anomaly of $0.0 \pm 0.05 \,^\circ\text{C}$ share the rank of the two warmest centuries
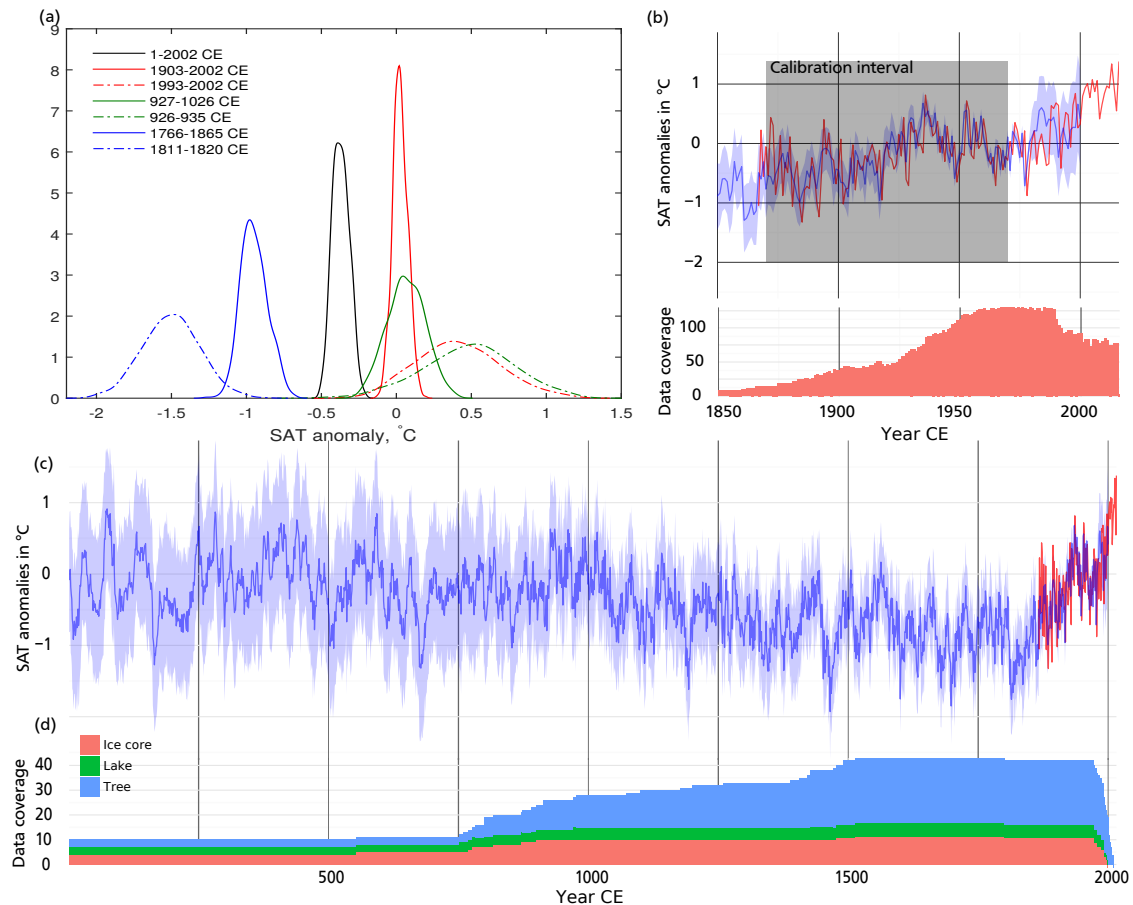
Figure 2. (a) Ensemble-based mean Arctic summer (June–August) SAT (land only) anomaly probability density for the three selected centennial-scale and three-decade-scale time periods. The periods presented correspond to the coldest and warmest century- and decade-long periods of the LIA and MCA together with the Contemporary Warm Period. The actual probability densities are estimated using the Gaussian kernel density function. **(b)** Arctic (land only) average summer temperature anomalies over the instrumental period and number of instrumental observations available. The grey box denotes the calibration interval. **(c)** Ensemble-based spatially averaged time variability in the seasonal SAT probability distribution over the reconstruction period. Blue line: ensemble mean; shading: (pointwise) 95 % posterior; red line: instrumental data. Note that before 1870 CE the number of instrumental observations rapidly decreases. **(d)** Number of proxies by archive type over time.

over the last 1200 years in the Arctic, which is in line with other studies (e.g. Ljungqvist et al., 2012, 2016).

The slow millennial-scale cooling is finally terminated by the contemporary warming, which is clearly identifiable since the middle of the 19th century. Figure 3 suggests that the LIA cooling is less pronounced in the new reconstruction compared with the same period reconstructed by McKay and Kaufman (2014), though the uncertainty intervals mostly overlap with their mean results. A likely explanation of this difference is the effect of targeting the summer season (as in our study) compared to annual mean in the reconstruction of McKay and Kaufman (2014). Throughout the LIA, sea ice cover has most likely experienced a pan-Arctic expansion as evidenced by proxy studies (e.g. Belt et al., 2010; Kinnard et al., 2011; Berben et al., 2014; Miettinen et al., 2015) and also supported by documentary evidence for the last phase of

the LIA (Divine and Dick, 2006; Walsh et al., 2017). Such sea ice expansion would lead to an increased continentality of the climate in most of the study domain, implying larger summer to winter SAT contrasts (see e.g. Grinsted et al., 2006, for Svalbard). This has potential effects on differently targeted reconstructions and the inferred magnitude of LIA cooling. The new reconstruction, however, shows larger low-frequency temperature variability than those reconstructed by Shi et al. (2012) or Tingley and Huybers (2013).

The transient features in the spatial mean reconstruction ensemble are analysed with the modified scale space method SiZer (Significance of Zero Crossings of the Derivative) (Chaudhuri and Marron, 1997). The original technique uses a local linear regression kernel-based estimator to produce a family of non-parametric smooth curves for the target data series for a range of kernel bandwidths (h). Assessment of
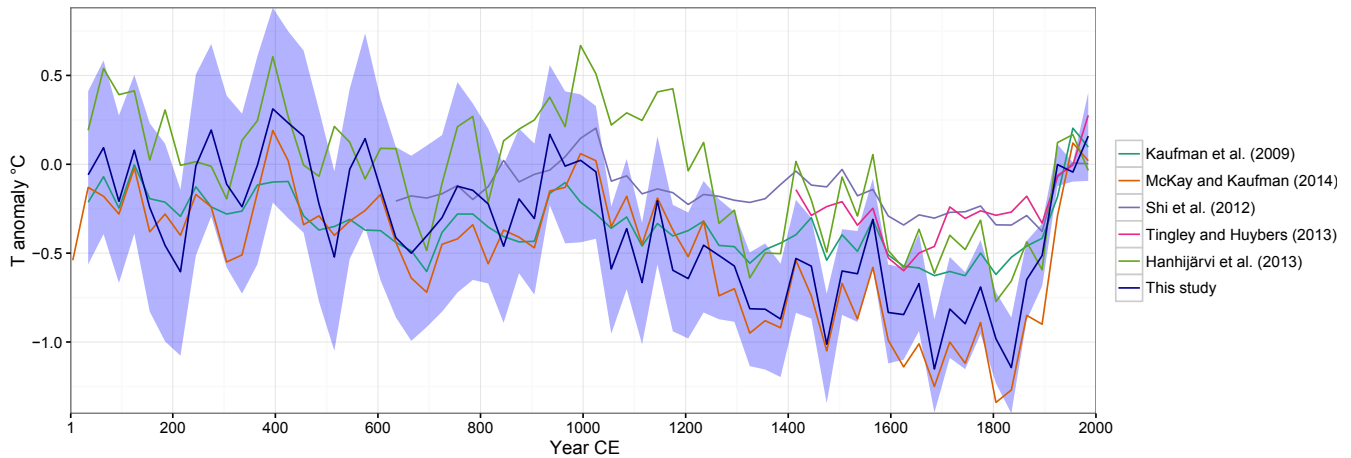
Figure 3. Comparison of this reconstruction (with 95 % confidence band) with other reconstructions; 30-year block averages. Note that McKay and Kaufman (2014) target annual temperatures and Hanhijärvi et al. (2013) reconstruct the North Atlantic sector of the Arctic (50° W–30° E).
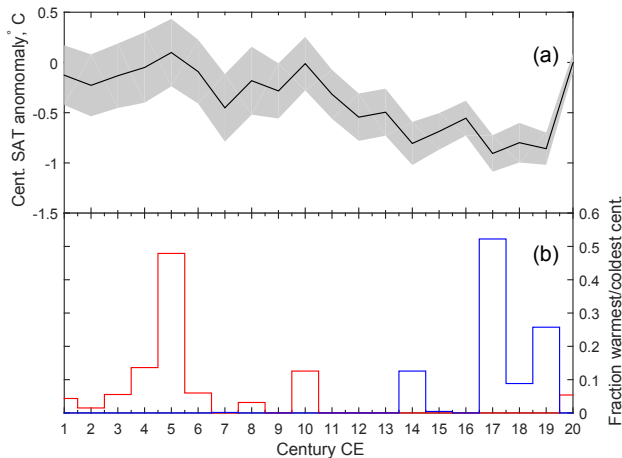

Figure 4. (a) Ensemble-averaged century and spatial mean Arctic SAT (solid black) with the respective 95 % CI highlighted in grey. **(b)** Mean Arctic SAT ensemble-based fraction warmest (red) and coldest (blue) centuries.

the statistical significance of the scale-dependent features in the observed data, such as the local linear trend estimate, is then provided based on the inferred variability in the data and the quantile specified.

The original SiZer summarises the data analysis results in a map which highlights the locations in scale (here the variability timescale) and space (here the point in time), where the slope of a smoothed version of the unknown true underlying curve is significantly positive or negative. The modification of SiZer used in this paper utilises the additional amount of information that is available via the ensemble of reconstructions. As the analysis is repeated for all individual members of the reconstruction ensemble, both the variability

in the estimated slope of the smoothed curve and the spread in slope significance for a certain scale and point in time can be tested. This approach therefore enables the assessment of the robustness of features detected as significant to be made across the entire range of independent and equally likely reconstructions.

Figure 5a presents results of the analysis, highlighting the timescales and periods for which at least 90 % of the ensemble members exhibit statistically significant changes. Results suggest that given the proxy network configuration and BARCAST settings used, the overall millennial-scale cooling trend as well as the MCA, LIA, and Contemporary Warm Period (CWP), appear as statistically significant features in the majority of the ensemble members. The MCA-to-LIA transition together with the onset of the CWP are the two coherent changes apparent on the broad range of timescales considered, down to a multi-decadal scale. The initial phase of the LIA-related cooling is flagged significant at a range of centennial to nearly millennial timescales. It is relatively early, centred already on 1030 CE. We note also that for the centennial timescale, Fig. 5a points to the onset of a statistically significant warming during the 1840s CE. This would justify using 1850 CE as the cut-off year for inferring the longer-term tendencies in the reconstruction prior to the CWP. Later, the period of 1917–1928 CE marks an ensemble-coherent warming trend in the terrestrial Arctic on the scale of about 30 years, which clearly links it to the early 20th century warming.

The statistically significant changes that are coherent across the reconstruction ensemble are four cooling and one warming episode revealed at the timescales of 30–100 years and centred at 1450 CE, 1591 CE, 1669 CE, 1810 CE (cooling), and 1477 CE (warming). In order to assess the magnitude and timings of the most rapid changes for the two se-
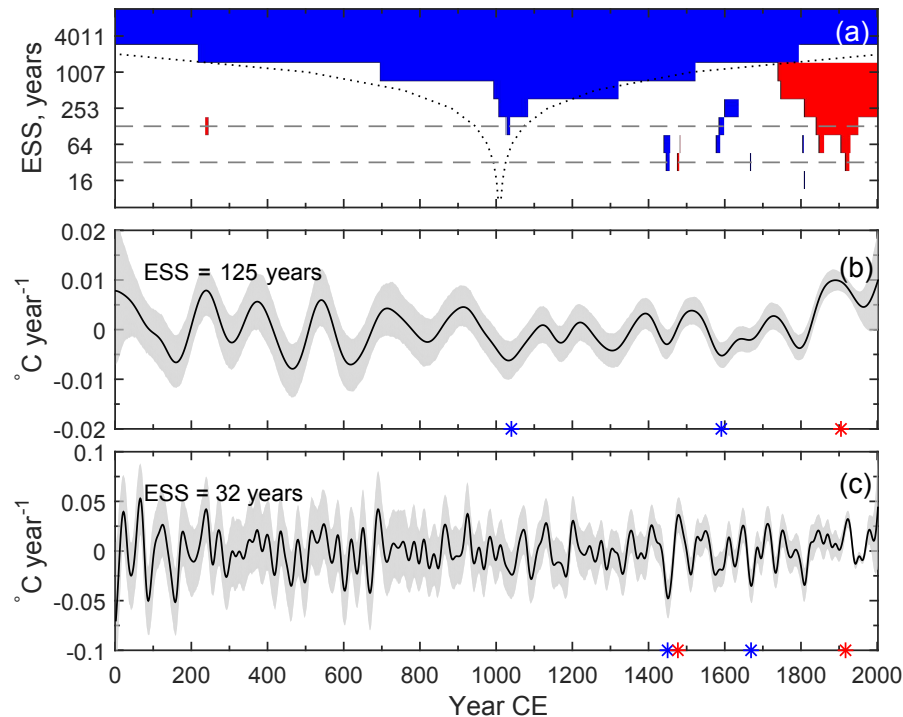
Figure 5. (a) Modified SiZer map of the spatial mean reconstructed SAT with colours red and blue marking the locations in scale (variability timescale) and space (time in this particular case) where the fraction of ensemble members exhibiting statistically significant warming and cooling exceeds 90 %; the parallel distance between the dotted lines indicates the effective size of the smoothing kernel used for a particular scale and hence gives an idea of the corresponding timescale involved at that level of smoothing. Solid lines in **(b)** and **(c)** show derivatives of local linear smooth lines together with the respective double standard deviation ranges (highlighted grey) for the two selected kernel bandwidths $\log_{10}(h) = \{0.9, 1.5\}$ with effective sample sizes of about 125 and 32 years, also marked as dashes in panel **(a)**. Blue or red asterisks in **(b, c)** mark the timings of the maxima in the rates of cooling or warming discussed in the text.

lected kernel bandwidths of $\log_{10}(h) = \{0.9, 1.5\}$, the derivatives of the respective kernel smooths for each ensemble member are calculated. The bandwidths selected correspond to the effective samples sizes of about 30 and 120 years and are hence representative of the three-decade- and centennial-scale variations. Figure 5b, c show the associated rates of changes as the ensemble mean together with the respective 95 % confidence interval (CI).

The two largest statistically significant cooling rates in the entire ensemble with average temperature changes of $-1.5 \pm 0.4\,°C$ and $-1.1 \pm 0.4\,°C$ over 3 decades are registered at 1450 CE and 1669 CE, respectively, while a recovery after the first cooling centred at 1477 CE featured a warming of $1.2 \pm 0.4\,°C$ over a similar 30-year time period. In terms of the rate of changes attained, the first cooling–warming episode appears unique over the 2000-year-long reconstruction, including one of the coldest decades in the reconstruction ensemble. At the highlighted centennial timescale, the most rapid changes are the MCA-to-LIA transition with a cooling of $-0.8 \pm 0.3\,°C$ centred at 1040 CE in line with glacier evidence on Svalbard (van der Bilt et al., 2015; Bakke et al., 2017), the cooling towards one of the LIA SAT minima at 1577 CE with $-0.7 \pm 0.2\,°C$, followed by the transi-

tion to the CWP centred at 1905 CE with an average warming of $1.2 \pm 0.2\,°C$ over ca. 120 years, which is also the largest centennial-scale warming rate detected in the entire ensemble. Note that the intra-ensemble variations hinder a robust detection of statistically significant changes common for the majority of the spatial mean reconstruction ensemble members at the timescales shorter than 3 decades, with the cooling towards the absolute decadal minimum of the record at 1811–1820 CE being the only remarkable exception. The same applies to the pre-750 CE period that appears highly variable on a range of scales when the reconstructions are considered individually, but shows no single episode that is localised in time across all ensemble members. The latter is related to a much reduced density of the multi-proxy network for the considered period (see discussion in Werner et al., 2013), and due to the age model selection code, which would delocalise events in time (see Werner and Tingley, 2015, for details). Given the sparse proxy network before 750 CE, and a correlation length on the order of 1500 km, this clearly highlights the need for proxy data to be extended back in time.
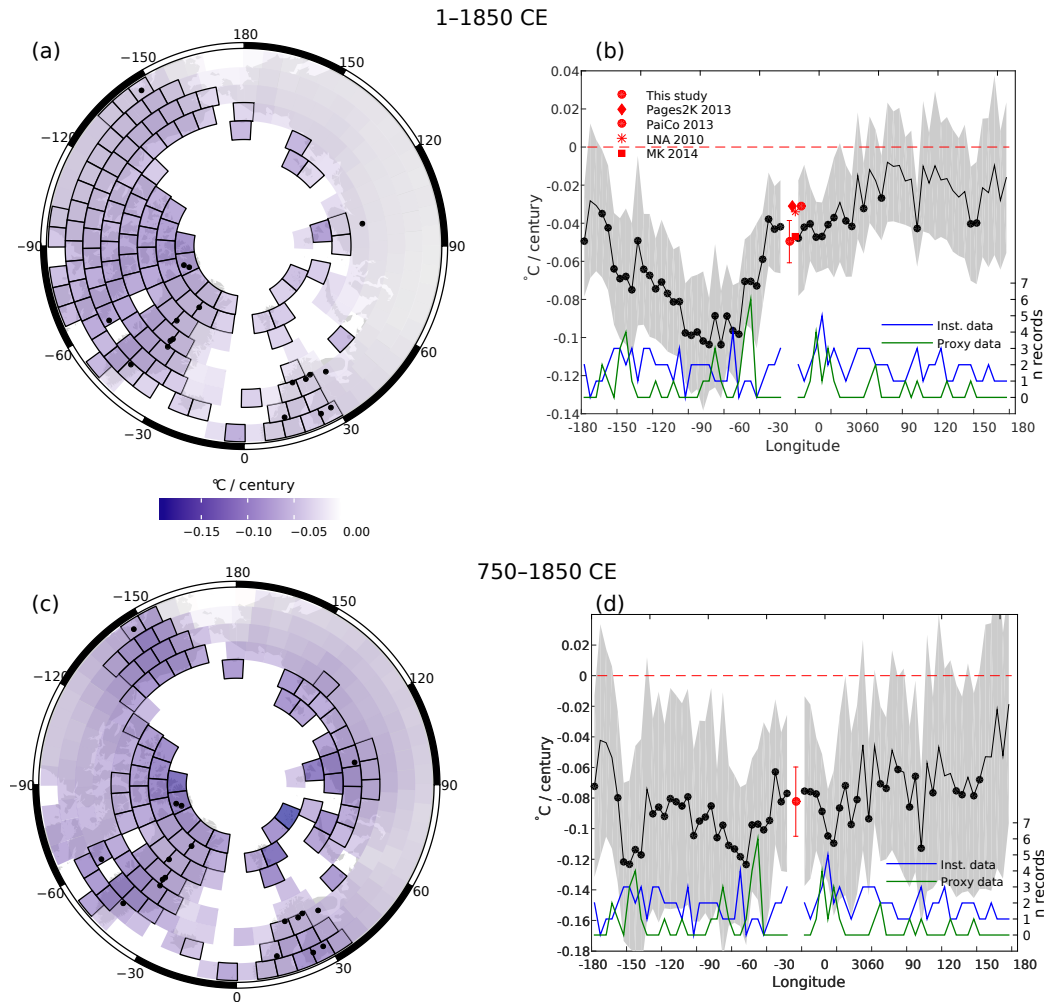
Figure 6. (a) Ensemble-averaged spatial linear trends over the period 1–1850 CE in °C per century; black dots mark proxy locations (back to 100 CE); quadrilaterals mark locations where the trend is statistically significant for more than 95 % of the ensemble members. **(b)** Ensemble-averaged meridional trends in the latitude-averaged reconstruction over the period 1–1850 CE (solid black line); meridional averaging over the 5° segments and zonal averaging over the terrestrial nodes is applied to each reconstruction ensemble member. Grey shading highlights the 95 % uncertainty interval on the estimated trend magnitude derived from the ensemble of spatially averaged reconstructions. Black circles indicate the meridional sectors where the trend is statistically significant. Solid green and blue lines show the number of proxy and instrumental records in each 5° longitudinal sector. For comparison, the pan-Arctic trend estimates for the same period are shown in red: PAGES 2k Consortium (2013), PaiCo 2013 and LNA 2010 (both: Hanhijärvi et al., 2013), MK 2014 (McKay and Kaufman, 2014), and for this study. Panels **(c)** and **(d)** show the analysis repeated for the period 750–1850 CE, with the proxies reaching to 750 CE marked in **(c)**.

4.2 Spatial trends over the Arctic

The analysis of the gridded reconstruction in the spatial domain is, in general, limited to proxy-rich periods and regions, especially on shorter timescales. Thus, the gridded reconstruction is mostly analysed back to 750 CE for the whole Arctic region. For a spatial subset that has a better proxy coverage back in time, such as the Atlantic sector, the whole reconstruction could be used (see Appendix B).

Since most other studies have analysed the temperature trends over the Arctic for the full period of 1–1850 CE, the spatial pattern of millennial-scale trends in the reconstructed Arctic SAT is depicted in Fig. 6, both back to 1 CE, and limited back to 750 CE. Note that there are large gaps in the proxy coverage for the early period, which result in non-significant and likely reduced trends over those parts of the target region. The results for the magnitude of the millennial-scale cooling in the spatial mean reconstruction are in line with the previous studies, although the new reconstruction tends to agree best with McKay and Kaufman (2014) (see Fig. 6b). While for the ensemble and the reconstruction domain average the rate of cooling attains -0.05 ± 0.01 °C per century, which results in an overall temperature decrease of about -0.9 °C during 1–1850 CE, the analysis reveals that

this long-term cooling trend seems spatially inhomogeneous. In particular, the largest magnitude of the millennial-scale cooling of up to $-0.13 \pm 0.02\,°C$ per century yielding a temperature decrease of $-2.4\,°C$ over the period of 1–1850 CE is registered in the region between 0–30° E and 10–170° W, and only this domain actually contains proxies covering the full CE. Averaging over the longitudes similarly suggests that the largest cooling over the period preceding the contemporary warming has likely occurred in the region encompassing Greenland and the Canadian Arctic between 30 and 120° W. At the same time, much less pronounced negative trends with an overall cooling of less than $-0.4\,°C$ over 1–1850 CE are detected in most of the Eurasian region within 30–180° E. This is statistically significant only in a few locations. The results outside the proxy-data-rich regions are mostly reflective of the overall mean cooling trend of the remaining proxies; any in-depth analysis needs (by design of the reconstruction method) to be limited to locations closer than about one or at most two e-folding lengths (ca. 1500 km) to the proxy data.

Since the analysis of the gridded reconstruction is limited to the time after 750 CE, the results above need to be interpreted carefully, especially more than about 1500 km away from any proxy data. Thus, the bottom half of Fig. 6 presents a similar analysis for the time span of 750–1850 CE. The revealed pattern suggests a more even cooling throughout the reconstruction domain with circum-Arctic trend magnitudes similar within the uncertainty estimate.

4.3 Contemporary Arctic warming in the context of MCA and LIA climate anomalies

Comparing the magnitude and spatial extent of past warm periods featuring similar settings in external forcing with the present-day warming is of major importance since it provides possible limits for the scales of naturally forced climate fluctuations. Figures 2 and 3 suggest that in the new reconstruction the period of 900–1050 CE, typically associated with the peak of the MCA, shows up at least similarly warm as the reconstruction for the late 20th and early 21st centuries, although the instrumental data suggest much warmer temperatures in the last decade (2006–2015 CE). This is in accordance with the conclusions reached previously in Shi et al. (2012), Hanhijärvi et al. (2013), and McKay and Kaufman (2014). The Arctic mean SAT reconstruction before about 750 CE has much higher uncertainties, and robustly identifying warm periods becomes more difficult. Especially in contrast to the reconstruction by Hanhijärvi et al. (2013) the Roman times around the first and second centuries CE do not show up as particularly warm in the circum-Arctic mean, which is also reflected in the analyses presented in the previous section. Note that their reconstruction was limited to the North Atlantic sector of the Arctic, and thus a direct comparison is difficult. Additionally the spatial skill of the reconstruction decreases back in time as the proxy data become sparser (see Appendix B), and spatial averages thus result in

higher uncertainties and the ensemble average will be closer to the overall mean. Taking these uncertainties into consideration, the focus will thus be on comparing the more tightly constrained MCA and LIA anomalies with the contemporary warm period.

The warmest century-long period of the mean SAT reconstruction after 750 CE associated with the MCA occupies most of the 10th century CE (927–1026 CE). The peak decade-long warmth of the MCA occurred during 926–935 CE, when the reconstructed spatial mean SAT anomaly attains $0.48 \pm 0.31\,°C$. The timing of the coldest centennial-scale period of the LIA, specifically 1766–1865 CE, broadly associates it with a Dalton grand solar minimum. This period also contains the coldest decadal-long event in the reconstruction detected during 1811–1820 CE with the mean Arctic ensemble-based temperature anomaly of $-1.5 \pm 0.2\,°C$. This cold decade also coincides with the period of increased volcanic activity, with two major tropical eruptions of 1808/1809 CE and Tambora in 1815 CE. The second coldest decade of the LIA with the SAT anomaly of $-1.4 \pm 0.2\,°C$ likely occurred during 1463–1472 CE, also following strong volcanic forcing.

Figure 2a presents the ensemble-based probability density functions (PDFs) of the mean SAT anomalies spatially averaged across the reconstruction domain for the six selected reconstruction sub-periods. These are the three selected century-long periods of 927–1026 CE, 1766–1865 CE, and 1903–2002 CE, representing both the aforementioned warmest and the coldest century-long periods of the record after 750 CE as well as the last century-long period, the second warmest of the reconstruction, which includes the CWP (in this study since 1978 CE onwards). For comparison, the same PDF for the entire reconstruction period is also presented. To further highlight the contrasts between the mean and extreme climate states, PDFs for the three shorter decadal-scale intervals corresponding to the anomalously warm and cold periods of 926–935 CE, 1811–1820 CE, and 1993–2002 CE (see also Sect. 4.4 for details) are displayed. The chosen decade of the CWP is the second warmest on average in the record after the MCA in the considered reconstruction period with a SAT anomaly of $0.41 \pm 0.28\,°C$, followed in rank by a warm decade of the early 20th century warming (from 1930 to 1939 CE; not shown here). The maps of spatial mean SAT anomalies for these periods follow in Fig. 7.

Comparing the coldest phase of the LIA with a mean centennial-scale SAT anomaly of $-0.94 \pm 0.09\,°C$ vs. the MCA $0.07 \pm 0.13\,°C$ and the last century of the reconstruction (SAT anomaly: $0.01 \pm 0.05\,°C$) emphasise the difference between the extreme warm and cold century-long periods in terms of the pan-Arctic summer temperature probability density. Figure 2a suggests that the centennial-scale maximum of the MCA could be at least as warm as the period 1903–2002 CE, although a reduction of proxy data after the
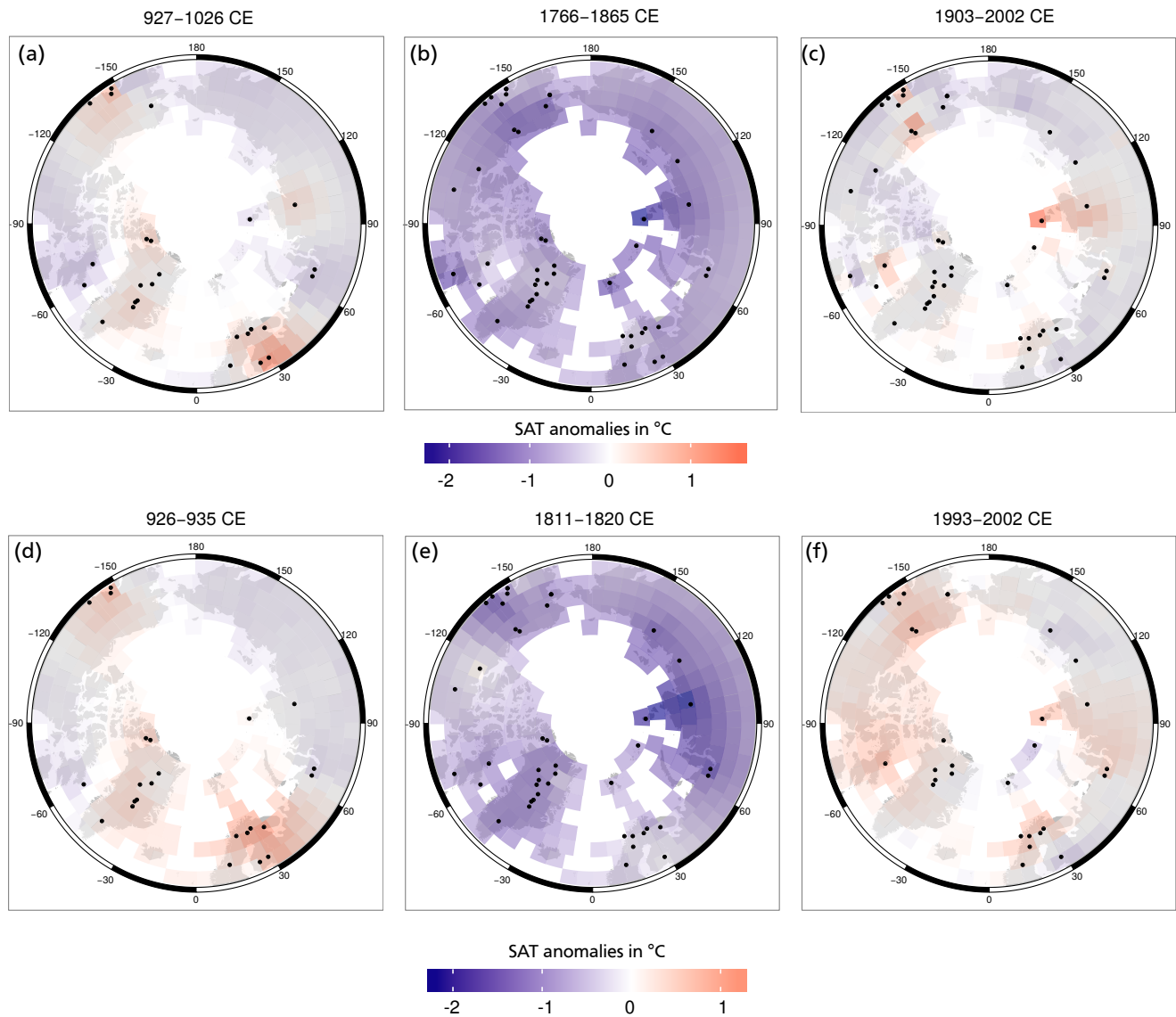
Figure 7. Ensemble average of the reconstructed Arctic2k SAT anomalies over the century- and decade-long periods of **(a)** 927–1026 CE, **(b)** 1766–1865 CE, and **(c)** 1903–2002 CE; **(d)** Ensemble-average SAT anomaly over the period of 926–935 CE, a potentially warmest decade since 750 CE with 7 of 10 years top ranked as potentially warmest. **(e)** Ensemble-average SAT over the period of 1811–1820 CE, a potentially coldest decade since 750 CE with 7 out of 10 years ranked potentially coldest. **(f)** Ensemble-average SAT anomaly over the period of 1993–2002 CE, a potentially warmest decade after the MCA with 5 out of 10 years ranked as potentially warmest. Colours show the temperature anomalies. Proxies marked by black dots.

1990s likely introduces a cold bias when estimating present-day warming in the reconstruction.

In order to quantitatively test the significance of the observed reconstructed differences in SAT anomalies between the selected periods, the two-sample t test is used on the samples of the derived distributions. During the testing procedure the realisations from different ensemble members of the Arctic SAT annual means are not pooled. Rather, the respective PDFs for the selected periods are derived for every individual ensemble member of the reconstructed SAT. The procedure uses bootstrap estimates of the PDF for the period (MCA and CWP) averages derived from 100 independent draws. The two-sample t test with separate variances is applied to test the null hypothesis of the two samples associated with the two different warm periods to originate from two normal distributions with equal means and unknown and non-equal variances. Using a one-tailed t test should then provide information on whether the MCA was on average warmer or colder than the last 100 years. The test statistics for each ensemble member are then collected and analysed.
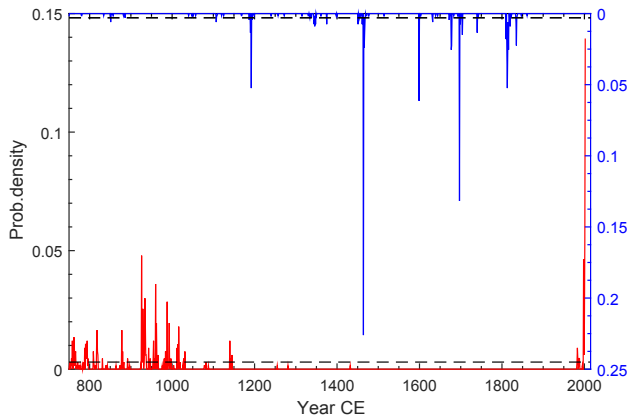
Figure 8. Probability density of statistically significant, potentially warmest (red) and coldest (blue) years over the period 750–2002 CE. Dashed lines indicate the 95 % significance level estimated from the ensemble of bootstrap surrogates.

Figure 9. **(a)** Fraction of potentially warmest or coldest years per decade with respect to time. **(b)** Occurrence of statistically significant, potentially warmest (red) and coldest (blue) years over the period 750–2002 CE in different sectors of the Arctic domain.

The testing results for a two-tailed test with unequal variances rejected H_0 of equal Arctic mean SAT anomalies between 927–1026 CE and 1903–2002 CE for 93 % of ensemble members. However, when considering hypotheses with a one-tailed test, no conclusive answer can be reached. Although the MCA appears slightly warmer on a centennial timescale compared with the last 100 years as shown in Fig. 2a, testing rejects H_0 for 64 % of the ensemble members only, whereas for the opposite alternative hypothesis (i.e. 1903–2002 CE warmer than MCA on average) the H_0 rejection rate is as high as 29 %.

Though this result somewhat favours the alternative hypothesis of $\overline{\text{SAT}}_{\text{MCA}} > \overline{\text{SAT}}_{\text{CWP}}$, the difference in the rejection rates appears negligible. We therefore conclude that given the collection of the proxy and instrumental data, and the reconstruction technique used, it is not possible to infer whether the Arctic summers of the last 100 years of the reconstruction (i.e. before 2002 CE) were unprecedentedly warm when compared with the previous major warm climate anomaly back to 750 CE. We also note that higher variability in the derived ensemble of realisations for the mean Arctic SAT anomaly during the warmest decade-long intervals of the MCA and CWP similarly prevents reaching any firm inference on the relative magnitudes of the two decade-long anomalously warm periods of the new reconstruction.
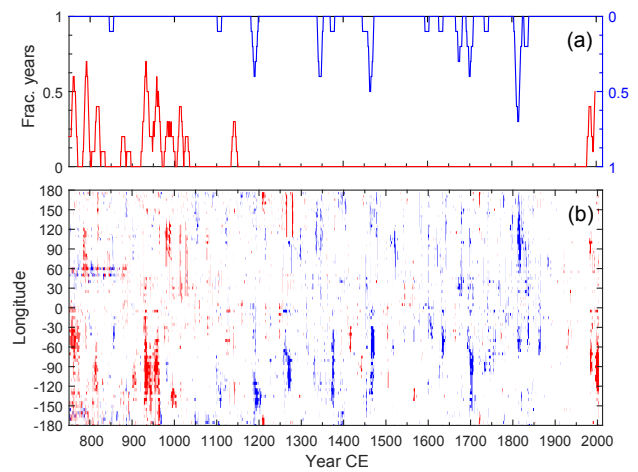
4.4 Spatial signature of past and recent extreme temperature anomalies

The distribution of extremely warm and cold years in both space and time is analysed by ranking the years according to their seasonal temperature for each ensemble member and the reconstruction node. Due to insufficient proxy data density and hence the inflated intra-ensemble variance (see Fig. B1) in the early part of the reconstruction period, the analysis is limited to the time after 750 CE. For the Arctic average the probability density for each year to be ranked as warmest or coldest is calculated across the entire ensemble using the spatial mean SAT. To check the statistical significance of the derived probability densities, the analysis is replicated on an ensemble of surrogates derived from the original reconstruction ensemble using block bootstrapping of the spatial mean reconstructions along the time axis. The block size of 10 years was assigned using an ensemble-average first-order autocorrelation coefficient of 0.8 and Mitchell et al. (1966) formula with adjustment of Nychka et al. (2000), yielding the efficient number of degrees of freedom in the data of about 125. The derived time-average 0.975 percentile of the probability density for the bootstrap surrogates is then used as the respective quantile for marking the years as potentially coldest or warmest during the analysis period (Fig. 8). In order to highlight a decadal-scale variability in occurrence of warm and cold extremes, the fractions of potentially warmest or coldest years per decade are calculated in sliding 10-year-long windows. In order to reduce the effects of the reconstruction uncertainties, the reconstruction is averaged over 5° longitudinal bins. To take the spatio-temporal autocorrelation into account during significance testing, the bootstrap replicates are drawn as 10-year-long time slices from individual reconstruction ensem-

ble members. The analysis results are presented as a time–longitude colour map in Fig. 9b.

The results of the analyses are reflective of the longer-term (millennial and secular) pan-Arctic tendencies in the seasonal SAT, yet the inter-regional differences are made clear as well. Of the series of past and present exceptional warmings, compared with the part of the present-day warm period before 2002 CE, the peak of the MCA features the two phases of a pronounced pan-Arctic warming with a consecutive series of spatially coherent warm extremes between ca. 920 and 970 CE (Fig. 9b). On a decadal timescale (Fig. 9a) the MCA is marked over the whole region by anomalies having a persistently high fraction of likely warmest decades with no decades containing a year ranked as coldest.

In particular, 7 out of 10 years during the decade of 926–935 CE resulting in the first MCA sequence of warm extremes and 6 out of 10 years of 954–965 CE during the second maximum were ranked as statistically significant warm extremes among the ensemble members. Note that sequences of potentially warmest years and hence decades with a higher fraction of extremes are also detected before 800 CE, though the eighth century does not appear in the reconstruction as particularly warm on average.

Figure 9b also highlights a difference in time evolution of the regional expression of the MCA via the spatial incoherence of extremely warm years or decades in this overall warm period. A somewhat earlier onset of warming in the European to Asian domain is evident from an increased frequency of warm extremes east of the zero meridian around 920 CE, followed by a coherent warming in the Greenland and North Atlantic sector of the study domain. Figure 9b also suggests that a second phase of the MCA could mainly be localised west of the prime meridian. Figure 7d exemplifies a picture of a pan-Arctic warming during the first warm decade (926–935 CE) of the first phase of the MCA with the largest reconstructed positive anomalies attained within the 170° W–30° E domain and 7 of 10 years ranked as potentially warmest in the reconstruction ensemble. A sequence of less pronounced MCA warm extremes occurred between 980 and 1040 CE localised primarily within the Atlantic sector (Greenland and Europe) of the study domain and do not exhibit as clear temporal coherence as the two main phases of the MCA.

Figures 8 and B1 demonstrate that the period after the MCA termination features a variable climate as manifested by an alternating sequence of potentially warmest and coldest years detected on the regional scale. Yet there is a pronounced lack of the pan-Arctic warm extremes, with a short exception of a 15-year-long warmth centred at 1142 CE. The following transition into the cold LIA is clearly marked by a drop in the frequency of potentially warmest years or decades to zero. During the LIA the cold extremes dominate on both the regional (Fig. 9b) and pan-Arctic scales (Fig. B1) until the onset of the contemporary warming after ca. 1880 CE. The first potentially warmest year after the termination of the MCA is detected only in 1983 CE, also indi-cating that a summer manifestation of the early 20th century warming in the terrestrial Arctic (e.g. Yamanouchi, 2011) was not pronounced enough to compete with significance of other annual- to decadal-scale warm extremes of the last 1250 years.

One can discern five major clusters of cold extremes during the LIA with the years of 1192, 1464, 1599, 1697, and 1813 CE ranked coldest in the majority (52 %) of reconstruction ensemble members. Figure 9 suggests that 1464 CE is most likely to be the coldest year after 750 CE, while the coldest decade of the reconstruction is represented by a sequence of spatially coherent potentially coldest years within 65° W–180° E (Fig. 9b). Figure 7e shows the spatial pattern of cooling for this decade of the LIA with 7 out of 10 years over the period of 1811–1820 CE ranked potentially coldest across the entire reconstruction ensemble.

Contemporary warming is manifested as a sequence of potentially warmest years starting in 1980 CE within 45–100° E and 60–110° W and since 1995 propagating to almost the entire reconstruction domain. Figure 7f shows a spatial map of temperature anomalies for the period 1993–2002 CE that features 5 out of 10 statistically significant warm extremes on the pan-Arctic scale. When compared with the probability density marginalised over the spatial domain displayed in Fig. 8, contemporary warming clearly reveals a coherence both in the spatial domain and agreement over the range of ensemble members that is at least as strong as the estimates made for the MCA. Additionally, about 30 % of the potentially warmest years in the entire reconstruction ensemble were registered in the time interval of 1993–2002 CE. The year 2002 is ranked warmest in 14 % of the reconstruction members, which is almost 3 times as many as any other potentially warmest year detected in the analysis. One should emphasise that this statistically significant sequence of warm extremes was detected outside the calibration period, which provides another indirect proof for a skill of our new Arctic reconstruction. This reconstruction, however, does not extend into the very last 15 years, over which warming in the Arctic has been continuing. With these years included in the analysis, the signature of the CWP would much likely become more prominent (see discussion in Sect. 4.1). The warmest periods in the reconstruction shown in Fig. 7 share similar features in the higher latitudes. The circum-Arctic warm anomalies at the shorelines are linked in the current period to the receding sea ice margin. This is indicative of a possible minimum of sea ice extent during the MCA similar to the one observed now.

5   Discussion and conclusions

This paper presented a new circum Arctic CF reconstruction of summer season temperatures back to 1 CE. Due to the sparse proxy network and thus large uncertainties, the spatial reconstruction is evaluated only back to 750 CE with the Arc-

tic average SAT anomaly evaluated back to 1 CE. The reconstruction uses a subset of 44 annually resolved temperature-sensitive terrestrial proxy archives of various types mainly from an updated and corrected PAGES2k database supplemented with six new recently published Greenland ice core series.

The technique applied is a recent extension of the Bayesian BARCAST, which explicitly treats climate archives with dating uncertainties (Werner and Tingley, 2015), which previously would be used on their "best guess" chronologies. The generated ensemble of 670 equally likely, independent realisations of past CF evolution in the Arctic together with a corresponding ensemble of synchronised chronologies represent the two major data outcomes of this study. As highlighted in Sect. 4.1, the probabilistic nature of the reconstruction results in straightforward uncertainty estimates even for complex analyses. As quantiles for a particular type of analysis are evaluated for individual ensemble members, the overall intra-ensemble coherence determines the spread and hence uncertainty of these quantities. The resulting ensemble of reconstructions including the ensemble of likely chronologies thus provides a convenient data set for further studies.

The quality of the reconstruction in the spatial and temporal domains was tested using a suite of metrics such as continuously ranked probability score ($\overline{\mathrm{CRPS}}_{\mathrm{pot}}$) and the reliability score, which are more appropriate for the Bayesian framework than the coefficient of efficiency and reduction of error typically used in palaeoclimate research. Judging from these scores it could be demonstrated that the new reconstruction is skilful for the majority of the terrestrial nodes in the reconstruction domain, making it a useful product for studying the late Holocene Arctic temperature variability at regional scales. However, from the analysis of intra-ensemble variability, but also from analyses on the extreme years and the calculated confidence intervals the reduction of skill back in time is apparent. This is mostly caused by the proxy network, which is getting sparser when going back in time, and should be taken into account when the new reconstruction is used for making any quantitative inferences.

In addition to presenting the new reconstruction and assessing its quality, the derived ensemble is used to uncover the potential of the new product and consider the results in light of previous studies on the subject. The major findings from the analysis of the new reconstruction are as follows.

The area-averaged Arctic2k reconstruction features similar major cold and warm periods throughout the last 2 millennia and thus compares favourably with earlier studies targeting a similar season and region. In particular, there is a pronounced orbital-scale cooling trend over the CE – a period over which the summer insolation has mostly been decreasing, although the spatial pattern cannot be reliably reconstructed over the full CE due to the sparse proxy network before ca. 750 CE. Since the proxy data set from Greenland is dominated by oxygen isotope series from ice cores, these can

be subject to a possible warm bias during the LIA bias caused by increased storm activity (Fischer et al., 1998) and/or be influenced by the site and source temperature-compensating effects (Hoffmann et al., 2001; Masson-Delmotte et al., 2005). The ice cores from northern Greenland are also expected to have a higher fraction of summer precipitation than those from the south due to the effect of continentality on the annual accumulation and hence exhibit a higher sensitivity to summer conditions. While site and source temperature compensating effects for the individual series can be accounted for by using the records of deuterium excess (Masson-Delmotte et al., 2005), other potential biases are difficult to resolve without additional support, e.g. from general circulation models.

The analysis of the reconstruction reveals the spatial signatures of the two major climate anomalies back to 750 CE, the MCA and LIA, as well as the beginning of the CWP in the circum-Arctic region. The MCA expression in the circum-Arctic region can be associated with a century-scale period between ca. 920 and 1060 CE showing an area-average SAT anomaly of $0.1 \pm 0.1\,°C$ during the warmest centennial period of 927–1026 CE. The MCA features two decadal-scale temperature maxima, both showing similar spatial extent of the regional SAT anomalies with the largest expression in the North American segment of the Arctic realm. A coherent warming of the period 927–936 CE during the first maximum of the MCA is associated with a potentially warmest decade of the reconstruction with the area-average summer temperature anomaly of $0.48 \pm 0.31\,°C$. While the most recent warming shows an even stronger regional coherence than the MCA, even across continents (PAGES 2k Consortium, 2013; Ljungqvist et al., 2016), the MCA was still an unusual and extremely warm period in the context of the past 2 millennia. Note that despite the evidence for prominent and lasting temperature fluctuations in the pre-750 CE period as well, these results should be interpreted cautiously due to the drastic reduction in proxy data density in the early part of the reconstruction period.

The new reconstruction suggests a relatively long, though interrupted by abrupt decadal-scale warmings, transition to the LIA after the second of the two MCA maxima ends at around 1060 CE. The coldest century-long period of 1766–1865 CE shows an almost spatially coherent circum-Arctic summer cooling. The cooling over the LIA, from essentially around the late 11th century went on until the mid 19th century CE. Most of the Arctic was coldest during the decade of 1811–1820 CE following the 1809 (unknown) and 1815 (Tambora) eruptions, which caused the "Year without a Summer" in 1816 over most of Europe and yielded a circum-Arctic SAT anomaly of $-0.8 \pm 0.2\,°C$.

The last decade of the reconstruction of 1993–2002 CE, being outside the calibration interval, accommodates some 30 % of the potentially warmest years across the ensemble since 750 CE, with half of them associated with the year 2002 CE alone. Yet given the input data available and the re-

construction method used, it still cannot be decided with sufficient confidence whether the warmest century and decade of the MCA or the CWP were warmest in the reconstruction. We speculate, however, that having the reconstruction extended into the very last 15 years, over which warming in the Arctic has been continuing, might have confirmed the summer SAT anomaly in the terrestrial Arctic to exceed the previous anomalous warm period of the Common Era.

The spectral characteristics of both proxies and reconstruction show that work is still needed on generating more and longer high-quality proxy series in parallel with a reanalysis of the existing data. Especially updating many of the only half-century-long North American tree ring series towards the present, but also possibly extending some of them into the first millennium CE seems to us like worthwhile effort (Babst et al., 2017). Additionally, a relative "flatness" of spectra on sub-decadal to multi-decadal timescales contrasting with an inflated variance of the multi-decadal to millennial variability (Appendix C) for some of the tree ring chronologies, suggests that a reassessment and potentially a revision of the raw data processing techniques used for these chronologies would be highly desirable.

BARCAST as a CF reconstruction technique still offers a large potential for future development and use in new, improved reconstructions. In addition to explicitly including the annually dated proxies with the chronological uncertainties in the scheme, which is a major innovation of the presented reconstruction, the next natural step will be a development of a theoretical and numerical framework to extend the technique to non-annually resolved proxy archives with chronological uncertainties. This will enable a substantial extension in the proxy coverage both in the spatial and time domains, including the marine realm dominated by non-annually resolved marine sediment proxy archives, potentially promoting an improved performance of the reconstructions at the low-frequency (centennial) timescales. While relatively flexible, the BARCAST framework would however still need major modifications that allow proxy response functions that are sensitive over different frequency bands. Additionally, these frequency bands need to be either proposed and fixed a priori, with possibly insufficient information available, or determined by the algorithm itself, potentially leading either to overfitting or convergence problems.

Code and data availability. The input proxy data are available through the individual publications (see tables in the Appendix). The majority are also available in the recent temperature database of PAGES 2k Consortium (2017), except for the NGT ice cores.

The base instrumental data (see Harris et al., 2014) can be downloaded from the BADC. The most recent version can be reached from the CRU home page http://browse.ceda.ac.uk/browse/badc/cru/data/cru_ts/cru_ts_3.24.01 (last access: 7 March 2017) under "observations".

The treated input data and the R script files used for the treatment of the input data as well as the reconstruction results (ensemble reconstruction, gridded ensemble mean, and area mean), together with the program code, are made available through NOAA (https://www.ncdc.noaa.gov/paleo/study/23031). The BARCAST code is © Werner and Tingley (2015).

Appendix A: Calibration and validation statistics

In order to estimate the skill of the reconstruction two different measures are used: the average potential continuously ranked probability score ($\overline{\text{CRPS}}_{\text{pot}}$) and the reliability (Reli) score (Hersbach, 2000; Gneiting and Raftery, 2007; Werner and Tingley, 2015; Tipton et al., 2016). The reliability analyses the accuracy of the uncertainty estimates. In principle it compares the empirical coverage rates of uncertainty bands with their respective nominal coverage rate, e.g. a 95 % confidence band should contain the target truth 95 % of the time. The $\overline{\text{CRPS}}_{\text{pot}}$ measures the accuracy of the reconstruction itself, i.e. the mismatch between the best estimate and the target. In a deterministic reconstruction it is equal to the mean absolute error. Both measures retain the original units of the data, and both signal a better result the lower they are. The results are shown in Fig. A1 (top row). For the calibration (validation) interval, the $\overline{\text{CRPS}}_{\text{pot}}$ is mostly below 0.2 °C (0.5 °C), and the reliability is sharper than 0.1 °C. This in principle indicates a relatively low reconstruction error, with uncertainty bands that (within reason) reflect the correct uncertainties.

Additionally the skill of the reconstruction beyond forecasting the calibration or validation period climatology is evaluated. In palaeoclimate reconstructions this is often assessed by the coefficient of efficiency and the reduction of error statistics (Cook et al., 1994). These analyse whether the reconstruction is closer to the validation target than the climatological mean of the calibration or validation period, respectively. However, these are not proper scoring rules (Gneiting and Raftery, 2007) and should thus not be used to analyse the results of a probabilistic reconstruction method. In essence, these two skill measures compare the reconstruction over the validation period to the mean climatology of the calibration (RE) and validation (CE) period (Lorenz, 1956; Briffa et al., 1988).

As introduced by Tipton et al. (2016), in order to generate a similar statistic, the mean and standard deviation over the validation and calibration intervals for each location with instrumental data are calculated. These are then used to generate an ensemble of time series. They act as simple surrogates for the calibration and validation interval climatologies, which are then compared against the target instrumental data of the validation period, using the $\overline{\text{CRPS}}_{\text{pot}}$. Should this value be lower than the $\overline{\text{CRPS}}_{\text{pot}}$ comparing the actual reconstruction ensemble against the instrumental data, the reconstruction does not add skill over the climatology. Thus, subtracting the $\overline{\text{CRPS}}_{\text{pot}}$ of the reconstruction from the $\overline{\text{CRPS}}_{\text{pot}}$ of the surrogates results in measures that indicate a skilful reconstruction if they are positive, i.e. a reconstruction that performs better than the climatology over the calibration (validation) interval. We denote these two scores as CRPS_{RE} and CRPS_{CE}. The bottom row in Fig. A1 shows that about half of the grid cells with instrumental data have a CRPS_{RE} and CRPS_{CE} that is above zero – and these grid cells are mostly those that have the longest instrumental time series (inside and outside the calibration interval). Thus, these results not only reflect a possibly weak reconstruction but more likely the lack of actual instrumental data to construct any meaningful comparison statistics over the validation period.
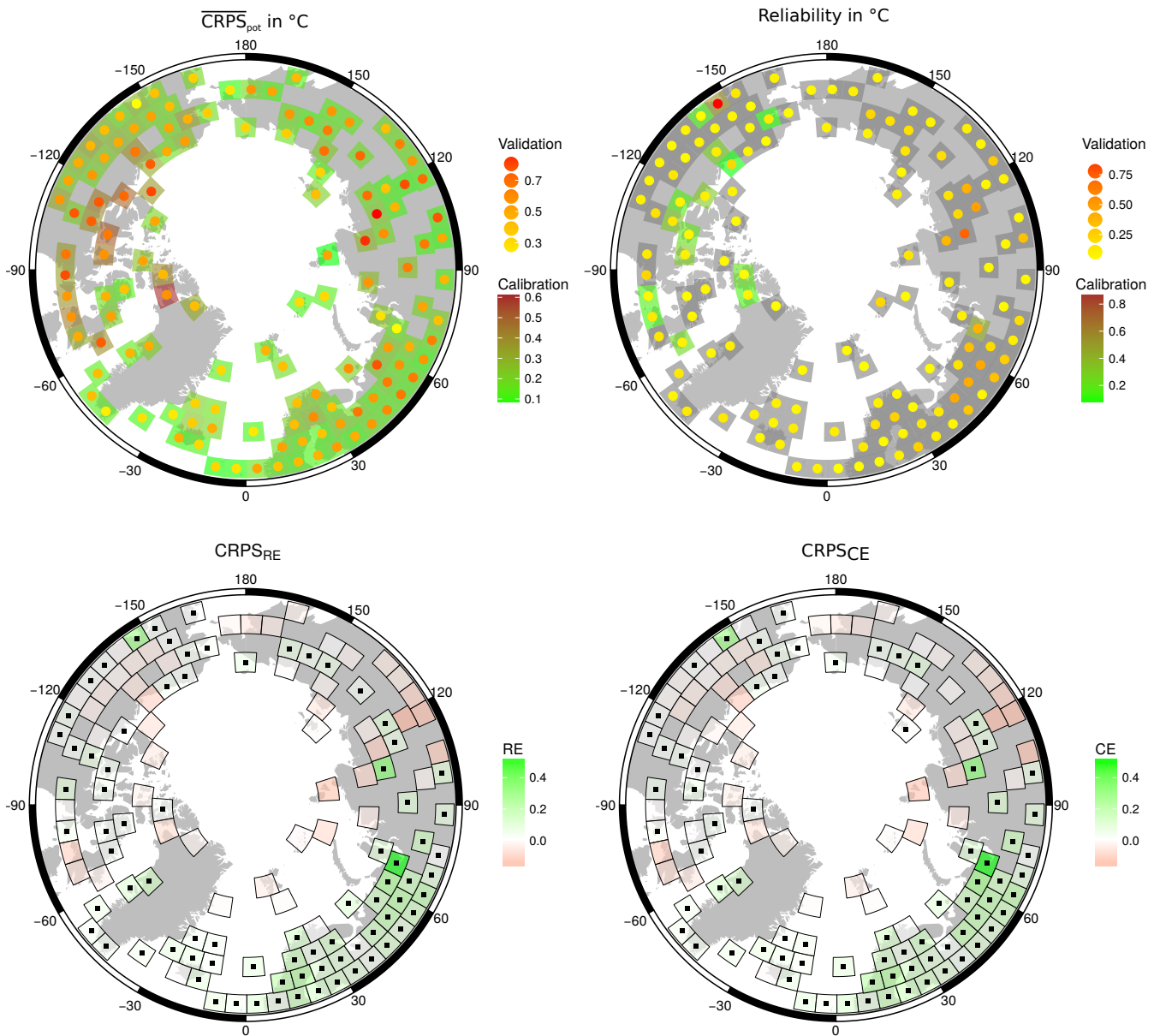
Figure A1. Calibration and validation results. The top row shows the $\overline{\mathrm{CRPS}}_{\mathrm{pot}}$ and reliability score for the calibration (quadrilaterals) and validation (points) periods. The bottom row shows the CRPS scores corresponding to an ensemble-based version of the reduction of error (RE) and coefficient of efficiency (CE) estimates. Squares denote grid cells with positive CRPS-RE or CRPS-CE, indicating a skilful reconstruction in the validation period. Grid cells with few data in the validation period show a lack of skill, which might be an artefact.

Appendix B: Intra-ensemble variance of the reconstruction

Figure B1 presents the time changes in the spatially averaged intra-ensemble variance as a measure of the spread across the ensemble members. The variance shows a progressive decline over the pre-industrial reconstruction more pronounced in the confidence intervals (CIs) for the period 800–1000 CE (which is linked with the time of an expansion of the multiproxy network). Along with the intra-ensemble variability, a progressive increase in the proxy data density over time contributes to the observed decrease in the ensemble spread. The introduction of the instrumental data into the scheme (corresponding to a calibration period in the regular climate reconstruction language) causes a sharp drop in the spread after 1850 CE that reaches a minimum around 1950 CE, a period of the maximal instrumental data coverage. Figure B2 further illustrates the effects of the spatial changes in input data density on the reconstruction intra-ensemble spread. The figure presents intra-ensemble spatial variances averaged over four time periods. The selected time slices are associated with periods of distinctly different proxy and calibration data density: part of the Roman Warm Period (200–300 CE) with a climate field reconstruction based on eight proxy records only, one of the coldest periods of the LIA (1600–1700 CE) with a complete multi-proxy network, and parts of the calibration period of 1850–1900 CE and 1950–1980 CE, representative of the low and high instrumental data coverages, respectively.
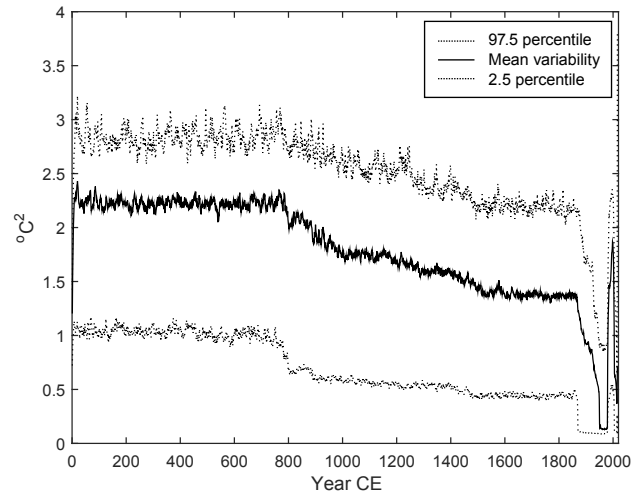

Figure B1. Time variability in spatially averaged intra-ensemble variance of the Arctic2k reconstruction together with the respective ensemble-based 95 % CIs.
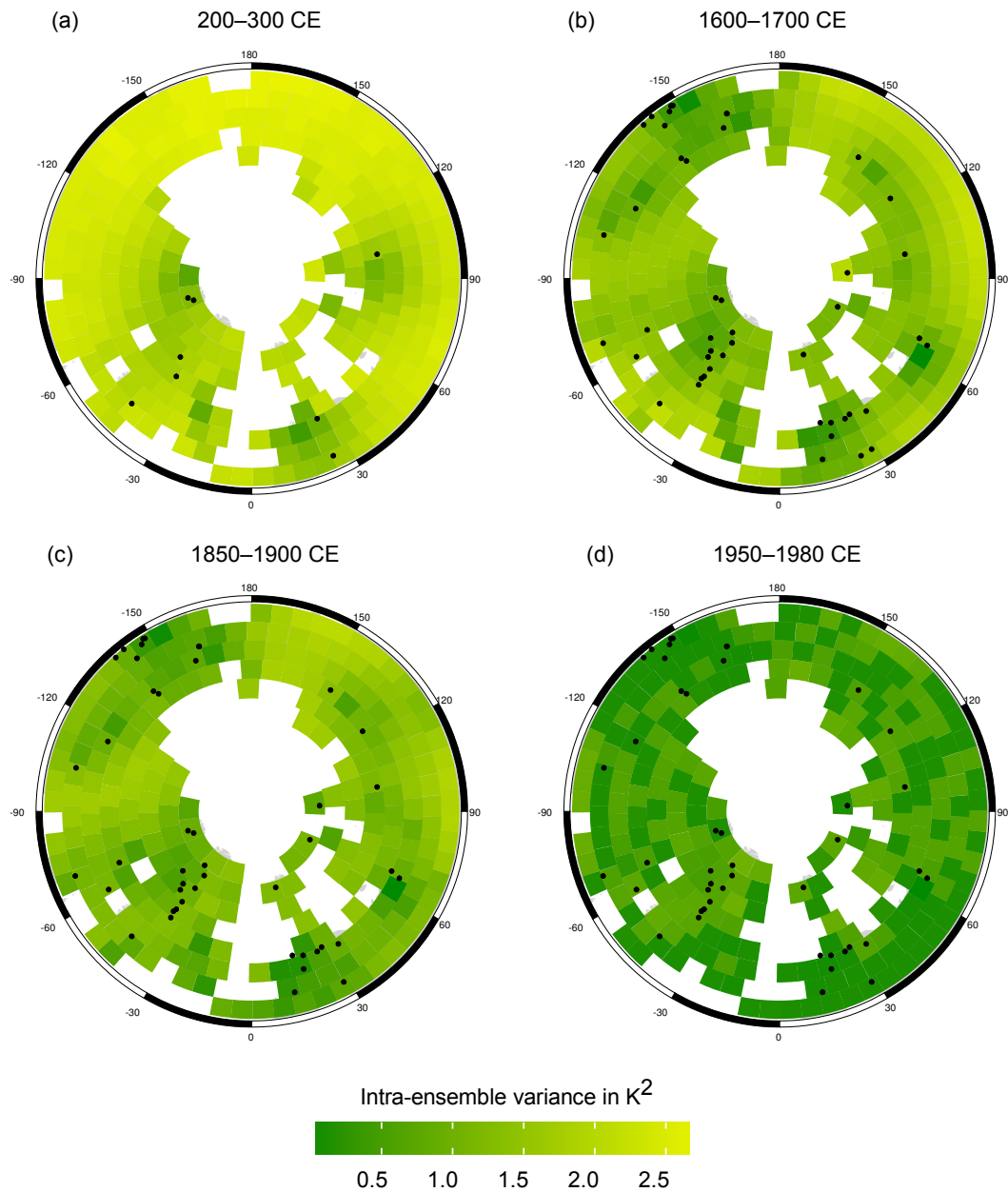
Figure B2. Time-averaged intra-ensemble variance of the Arctic2k reconstruction shown for the four sub-periods with a distinct difference in proxy data density (200–300 CE vs. 1600–1700 CE, panels **a** and **b**) and calibration sub-periods with different instrumental data coverage (1850–1900 vs. 1950–1980 CE, panels **c** and **d**). Black dots show the proxy locations with at least one data point over the period of averaging.

Appendix C: Statistical properties of the reconstruction and input data

As an additional test for the reliability of the proxy series and the validity of the climate field reconstruction, the temporal persistence of both need to be analysed. The Kolmogorov–Smirnov test is first used to test the normality of the proxy records and the climate field reconstruction, with a significance level $p = 0.05$, and additionally Q–Q plots are checked. Then, the power spectral density (PSD) is used to study the variability at different frequencies for the records, using the periodogram as an estimator of the PSD. The periodogram is defined here in terms of the discrete Fourier transform H_m as $S(f_m) = (2/N)|H_m|^2$, $m = 1, 2, \ldots, N/2$. The sampling time is the time unit (here: years), and the frequency is measured in cycles per time unit: $f_m = m/N$. $\Delta f = 1/N$ is the frequency resolution and the smallest frequency which can be represented in the spectrum, while $f_{N/2} = 1/2$ is the Nyquist frequency.

The characteristic shape of the spectrum provides useful information about the temporal persistence or memory of the underlying process. If the data are close to Gaussian and monofractal, the second-order statistics are sufficient to describe the statistical properties of the data. The spectral shape can then be associated with well-known stochastic processes. If the spectrum has a power-law shape, the process exhibits long-range memory (LRM). The strength of memory in an LRM stochastic process is described by the spectral exponent β, which can be estimated by a linear fit to the power spectrum; $\log S(f) = -\beta \log f + c$. If the spectrum is Lorentzian (power law on high frequencies, flat on low frequencies), the underlying process is closer to an AR(1) process. In all spectral analyses, the fitting is applied to log-binned periodograms to ensure that all timescales are weighted equally. If the Gaussianity and monofractality criteria are not met, there could be underlying structures such as intermittency that are not captured by the analyses. In the temperature time series considered here, deviations from normality are due to nonlinear dynamics associated with volcanic eruptions, for example.

C1 Spectral analyses of the proxy records

The six proxy records originating from lake sediments deviate substantially from a Gaussian distribution and thus had to be transformed before analysis. Afterwards, around 60 % of the individual proxy records are Gaussian according to the Q–Q plots and the p values from the Kolmogorov–Smirnov test.

The characteristic shape of the spectra for all of the proxy records are classified into three spectral categories: (1) AR(1) processes, (2) persistent power-law processes with spectral exponent $0 < \beta < 1$, and (3) records exhibiting weak persistence on high frequencies and increased levels of variability in frequencies corresponding to timescales longer than
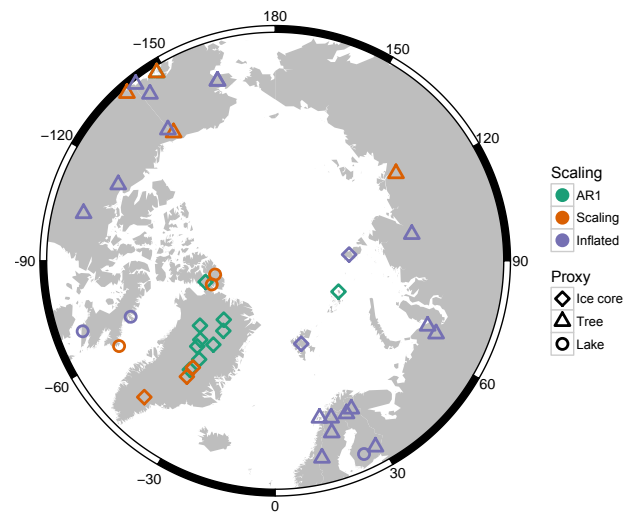

Figure C1. Proxy type (triangles: ice cores; diamonds: tree rings; squares: lake sediments) and persistence properties are colour coded.
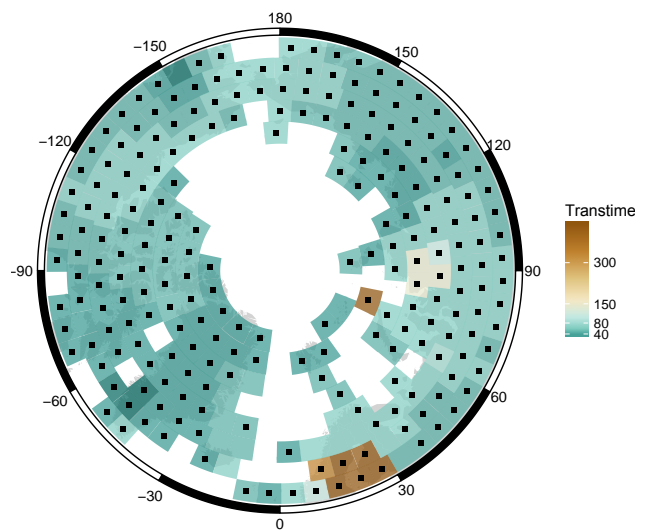

Figure C2. Analysis of scaling properties of the reconstruction. The transition timescale (colour coded) and strength of change (black dot: exponent change > 1). Most of the reconstruction domain resembles an AR(1) process.

decadal–centennial. Figure C1 illustrates the spatial distribution of the proxy records with proxy type indicated by shape, and categories with colours. The Greenland records are similar to either an AR(1) or an LRM process. The Greenland LRM records are in fact only weakly persistent, with a spectral exponent $0 < \beta < 0.3$. There is thus little evidence of long-term cooling due to orbital forcing from these records. Along with a few tree ring records, the Greenland ice core records are the longest records used for the present reconstruction. As the low-frequency variability in these records

dominates the reconstructed long-term variability, the resulting reconstructions do exhibit similarly low variability at long timescales.

The proxies of category 3 are mainly tree ring records, widely distributed along the reconstruction region. These records may require additional attention in future studies, as the level of high- versus low-frequency variability is unusual compared to other proxy records and also instrumental measurements. Similar spectral characteristics were obtained for other tree ring chronologies in Franke et al. (2013), Zhang et al. (2015), Esper et al. (2015) and Büntgen et al. (2015). The persistence in a number of millennium-long climate model simulations and proxy-based temperature reconstructions has been studied in Østvand et al. (2014) and Nilsen et al. (2016) using the power spectrum along with selected other techniques. In these studies, LRM was detected in all records up to centennial–millennial timescales.

C2 Spectral analyses of climate field reconstruction

The resulting p values from the Kolmogorov–Smirnov test indicate that for individual locations of the field reconstruction, about 60–80 % of the ensemble members are Gaussian. For each ensemble member of the reconstruction, and each location, a spectral analysis is performed. Then, the persistence is analysed for ensemble-averaged spectra of each location. The analyses indicate that the reconstructed temperature is best described as an AR(1) process in time at almost all grid cells. This is not surprising, as the longest proxy records exhibit low levels of long-term variability, and the BARCAST reconstruction technique assumes an AR(1) model for the temporal evolution of the temperature. Further details about the characteristic transition times are obtained by making a least-squares fit of a bilinear continuous function for the spectrum. The detected break is located where the two lines intersect. The coloured map in Fig. C2 shows the spatial distribution of the found transition timescales; black dots indicate that the difference between the spectral exponents for low and high frequencies is more than 1. The spatial coherence indicates that BARCAST performs well when extrapolating temperatures to locations where observations are unavailable. For most of the area we find a marked transition in the spectral slope (black symbols). Only the east coast of Greenland and the Scandinavian sector have slightly less difference between the high- and low-frequency variability; that is, the spectral exponent does not change much between the two identified scaling regimes. This indicates more similarity to an LRM process. Additionally, the transition timescale is above 100 years for a number of single locations. There, the reconstruction is indeed closer to an LRM process than an AR(1) process.
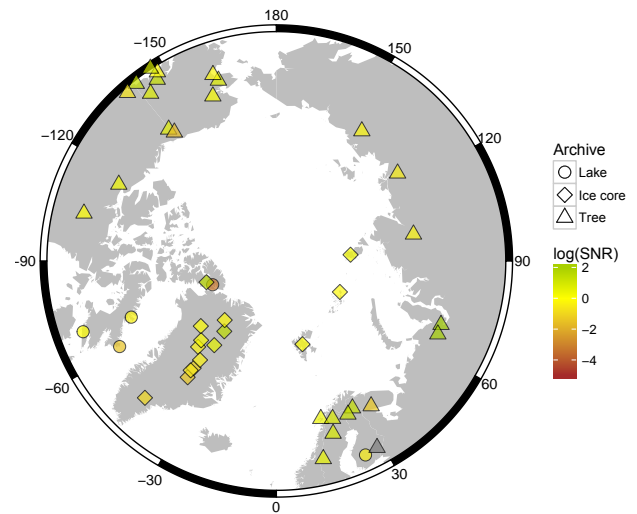

Figure C3. The mean of the posterior SNR for all proxy series. Note the logarithmic scale. There is one series with a negative β_1 (grey).

C3 Proxy response

The BARCAST output parameters contain information on the proxy signal strength (β_1 in Eq. 1b) and proxy noise level (τ_P^2, Eq. 1b). Under the assumption of a unit standard deviation climate variable, the ratio of β_1/τ_P returns an estimate of the signal-to-noise ratio (SNR, in amplitudes) of the individual proxy series. The mean of all ensemble draws is shown in Fig. C3. Note that one proxy series (Finnish Lakeland) has a negative (inverted) response ($\beta_1 \approx -0.38 \pm 0.7$).

In general, most tree ring series have a SNR > 1, and they seem to be the highest-quality proxies on average, followed by the ice core data. The lake sediments add only some skill, but are important especially in regions with no other data present, such as eastern North America. This might be caused by the in general higher dating uncertainties (see discussion in Werner and Tingley, 2015) or a response on different timescales. This underlines the necessity to really use multiple proxies and to further improve the reconstruction methods to make use of information on different timescales.

Appendix D: Input proxy data

Timescale modelling

In order to account for possible chronological uncertainties in the annually resolved proxy records, the technique of Comboul et al. (2014) is applied to the proxies with layer-counted timescales for the generation of ensembles of chronologies. BAM (banded archive modelling) simulates the timescale counting procedure as a superposition of two cumulative Poisson processes with age perturbations associated with two categories of errors: either miss (type 1) or double-count (type 2) of an annual layer. More specifically, for each measurement x_i assigned a time t_i with $i \in \{1, \ldots, n\}$, and a neighbouring x_{i+1} with t_{i+1}, $i \in \{2, \ldots, m\}$, the vector of time increments $\boldsymbol{\delta}$, $t_{i+1} - t_i = \delta_i$ comprises two independent stochastic processes P^{Θ_1} and P^{Θ_1}, with parameters Θ_1 and Θ_2, representing the rates of missing and doubly counted annual layers, respectively. We note that the approach implicitly assumes the independence of the two stochastic processes and depth (time) invariance of the error rates.

For the proxy series with chronologies constructed using a combination of annual layer counting and time markers (tie points) t_k, $k \in \{1, \ldots, K\}$, such as volcanic sulfate peaks or tephras with ages known to a specific precision (σ_k), a two-step procedure was implemented. The first step involved an Markov chain Monte Carlo simulation of M perturbed sets of tie points $\left[\widetilde{t_k^m}\right]$ following Divine et al. (2012), where $[\bullet]$ stands for rounding the argument to the nearest integer. For each particular set m of perturbed tie points and a time interval $\left[\widetilde{t_k^m}, \widetilde{t_{k+1}^m}\right]$, $k \in \{1, \ldots, K-1\}$ between the perturbed pairs of tie points timescale modelling was applied, and only those that satisfied a criterion of $\sum \delta_i = \widetilde{t_{k+1}^m} - \widetilde{t_k^m}$ were retained for further analysis. For ages older than $\widetilde{t_K^m}$ a model with a free boundary was used instead. In total $M = 1000$ timescales $\widetilde{t_i^m}$ per proxy archive were generated. Using interpolation, the proxy series x_i were further projected on the generated timescales $\widetilde{t_i^m}$ to yield the ensemble of proxy series with perturbed chronologies.

The error rates $\{\Theta_1, \Theta_2\}$ were estimated for each particular proxy archive. In the framework the counting procedure is defined, for each point t_i of the true unknown timescale the uncertainty of the modelled timescale follows the Skellam distribution with parameters $\{\lambda_1, \lambda_2\} = \{(t_s - t_i)\Theta_1, (t_s - t_i)\Theta_2\}$, where $(t_s - t_i)$ denotes the time lapse between t_i and a counting start point t_s (Comboul et al., 2014). For a symmetric error rate $\Theta_1 = \Theta_2$ and $(t_s - t_i)$ large enough, it converges to a normal distribution $N(0, \lambda_1 + \lambda_2)$. The error rates can therefore be estimated as

$$\left\{\hat{\Theta}_1, \hat{\Theta}_2\right\} = \mathrm{argmin}_{\Theta_1, \Theta_2}(\sqrt{\delta t_{\max} \cdot (\Theta_1 + \Theta_2)} - \Delta_t/4), \quad \text{(D1)}$$

where for a particular proxy archive $\delta t_{\max} = \mathrm{argmax}_k (t_{k+1} - t_k)$, $k \in \{1, \ldots, K-1\}$, or the entire length of the chronology, and Δ_t denotes an estimated largest offset of the reported timescale from the unknown true timescale. For the majority of records we estimated the type 1 and type 2 error rates using the authors reports on the tie point used and uncertainty of the constructed chronologies. For the few archives in which the chronological uncertainties were not reported, a conservative estimate of $[\Theta_1, \Theta_2][0.05, 0.05]$ was assigned.

Table D1 shows the list of proxy series together with parameters of the model used to simulate the annual layer-counting process. In total ensembles of timescales for 13 annually dated records of the Arctic2k network, six ice core and seven lake sediment records, plus seven annually dated ice cores from the North Greenland Traverse from 1993 to 1995 (recently reanalysed by Weißbach et al., 2016), are generated.

Table D1. List of the proxy records including the proxies of the Arctic2k network (PAGES 2k Consortium, 2017) with layer-counted timescales used in the present study together with parameters of the probabilistic model used in Monte Carlo simulations of the layer-counting process. The archives lacking information on the dating uncertainty are marked "*"; a conservative estimate of $[\Theta_1, \Theta_2][0.05, 0.05]$ was used in timescale modelling.

Site ID	Pages2k site name	Location lat, long	Elev., m	Proxy type	Season (month)	Time period (yr CE)	(Θ_1, Θ_2)	Reference
Arc_025	Lake Nautajärvi	61.81, 24.68	104	ls	3, 4, 5	−555–1800	0.055	Ojala and Alenius (2005)
Arc_076	Soper Lake	62.92, −69.88	14	ls	6	1514–1992	0.01	Hughen et al. (2000)
Arc_024	Donard Lake	66.73, −61.35	500	ls	6, 7, 8	752–1992	0.05*	Moore et al. (2001)
Arc_029	Big Round Lake	69.87, −68.83	180	ls	7, 8, 9	971–2000	0.025	Thomas and Briner (2009)
Arc_020	Lake C2	82.13, −77.15	1.5	ls	6, 7, 8	−1–1987	[0.053 0.007]	Lamoureux and Bradley (1996)
Arc_004	Lower Murray Lake	81.35, −69.53	106	ls	7	−1–2000	0.045	Cook et al. (2009)
Arc_065	Lomonosovfonna	78.87, 17.43	1250	ic	12, 1, 2	1598–1997	0.05	Divine et al. (2011)
Arc_064	ANIK	80.52, 94.82	750	ic	ann	900–1998	0.05	Opel et al. (2013)
Arc_031	NGRIP1	75.1, −42.32	2917	ic	ann	−1–1995	0.0002	Vinther et al. (2010)
Arc_034	Dye-3	65.18, −43.83	2480	ic	ann	1–1978	0.0002	Vinther et al. (2010)
Arc_035	GRIP	72.58, −37.64	3238	ic	ann	1–1979	0.0002	Vinther et al. (2010)
Arc_032	Agassiz Ice Cap	80.7, −73.1	1700	ic	ann	−1–1972	0.0002	Vinther et al. (2010)
Arc_033	Crête	71.12, −37.32	3172	ic	ann	553–1973	0.0002	Vinther et al. (2010)
Arc_011	GISP2	72.10, −38.08	3200	ic	ann	818–1987	0.01	Grootes and Stuiver (1997)
Arc_078	Windy Dome	81.0, 64.0	509	ic	ann	1225–1995	0.02	Kinnard et al. (2011)
16	B16	73.94, −37.63	3040	ic	ann	1469–1992	0.01	Weißbach et al. (2016)
17	B18	76.62, −36.40	2508	ic	ann	874–1992	0.002	Weißbach et al. (2016)
18	B20	78.83, −36.50	2147	ic	ann	777–1993	0.01	Weißbach et al. (2016)
19	B21	80.00, −41.14	2185	ic	ann	1373–1993	0.01	Weißbach et al. (2016)
20	B26	77.25, −49.22	2598	ic	ann	1505–1994	0.005	Weißbach et al. (2016)
21	B29	76.00, −43.49	2874	ic	ann	1471–1994	0.005	Weißbach et al. (2016)

Table D2. List of the tree ring records. Only records north of 60° N that went back at least to 1500 CE were included. This removes a number of short records, mostly from North America.

PAGES2k.ID	PAGES2k.Sitename	Lat (°N)	Long (°E)	Elev (m)	Time period CE	Reference	Data url
NAm_091	Almond Butter Upper	65.2	−162.2	213	1406–2002	D'Arrigo et al. (2005)	https://www.ncdc.noaa.gov/paleo/study/3044
Arc_065	Arjeplog	66.3	18.2	800	1200–2010	Björklund et al. (2014)	https://www.ncdc.noaa.gov/paleo/study/3264
Arc_002	Avam–Taimyr	72	101	250	0–2000	Briffa et al. (2008)	https://www.ncdc.noaa.gov/paleo/study/14188
NAm_126	Coppermine River	67.2	−115.9	213	1428–1977	Jacoby and D'Arrigo (1989)	https://www.ncdc.noaa.gov/paleo/study/3592
Eur_013	Finnish Lakeland	62	28.325	130	760–2000	Helama et al. (2014)	http://www.ncdc.noaa.gov/paleo/study/16790
NAm_104	Firth River 1236	68.7	−141.6	790	1073–2002	Anchukaitis et al. (2013)	https://www.ncdc.noaa.gov/paleo/study/14790
Arc_074	Forfjorddalen	68.73	15.73	200	1100–2007	McCarroll et al. (2013)	https://www.ncdc.noaa.gov/paleo/study/19943
Arc_007	Gulf of Alaska	61.03	−146.59	230	800–2010	Wiles et al. (2014)	https://www.ncdc.noaa.gov/paleo/study/19743
NAm_083	Herring Alpine	60.4	−147.8	275	1422–1972	PAGES 2k Consortium (2017)	https://www.ncdc.noaa.gov/paleo/study/3264
NAm_127	Hornby Cabin	64	−103.9	160	1491–1984	Jacoby and D'Arrigo (1989)	https://www.ncdc.noaa.gov/paleo/study/3598
Arc_016	Indigirka	69.5	147	80	1259–1994	Hughes et al. (1999)	https://www.ncdc.noaa.gov/paleo/study/14188
Arc_063	Jämtland, Sweden	63.24	13.34	650	783–2011	Zhang et al. (2016)	https://www.ncdc.noaa.gov/paleo/study/19743
		67.45	33.14	320	800–2005	McCarroll et al. (2013)	https://www.ncdc.noaa.gov/paleo/study/19943
NAm_002	Kobuk/Noatak	67.1	−159.6	100	978–1992	PAGES 2k Consortium (2017)	https://www.ncdc.noaa.gov/paleo/study/3706
Arc_071	Laanila	68.49	27.33	265	800–2005	McCarroll et al. (2013)	https://www.ncdc.noaa.gov/paleo/study/19943
NAm_032	Landslide	60.2	−138.5	800	913–2001	Clague et al. (2006)	https://www.ncdc.noaa.gov/paleo/study/13785
Arc_024	Lena River	70.67	125.87	180	1490–1994	MacDonald et al. (1998)	https://www.ncdc.noaa.gov/paleo/study/3902
NAm_088	Miners Well	60	−141.7	650	1428–1995	PAGES 2k Consortium (2017)	https://www.ncdc.noaa.gov/paleo/study/5244
NAm_094	Nabesna Mine	62.4	−143.1	1167	1471–1997	Davi (2003)	https://www.ncdc.noaa.gov/paleo/study/3605
Eur_003	Northern Scandinavia	68	25	300	−138–2006	Esper et al. (2012)	http://www.ncdc.noaa.gov/paleo/study/1003406
Arc_061	Polar Urals	66.9	65.6	250	891–2006	Schneider et al. (2015)	https://www.ncdc.noaa.gov/paleo/study/19743
NAm_003	Prince William Sound	60.5	−148.3	100	873–1991	Barclay et al. (1999)	https://www.ncdc.noaa.gov/paleo/study/14274
NAm_100	Seward Composite	65.2	−162.3	100	1389–2001	D'Arrigo et al. (2006)	https://www.ncdc.noaa.gov/paleo/study/13708
Arc_062	Torneträsk	68.26	19.6	320	−39–2010	Melvin et al. (2013)	https://crudata.uea.ac.uk/cru/papers/melvin2012holocene/
Arc_079	Yamalia	66.8	68	30	914–2003	Briffa et al. (2013)	https://crudata.uea.ac.uk/cru/papers/briffa2013qsr/
Arc_008	Yukon	67.9	−140.7	300	1177–2000	D'Arrigo et al. (2006)	https://www.ncdc.noaa.gov/paleo/study/13758

Table D3. List of the lake and ice core data from the PAGES2k database that were not used in this study.

Site ID	Pages2k site name	Location lat, long	Proxy type	Time period (yr CE)	Resolution	Reason
Arc_037	Iceland	64.77, −18.37	doc	945–1935	30	not annual
Arc_005	Camp Century	77.17, −61.13	ic	1242–1967	1	20 yr averages
Arc_018	Austfonna	79.83,24.02	ic	1400–1998	1	interpolated onto annual scale
Arc_075	Prince of Wales, Ellesmere Isl.	78.4, −80.4	ic	151–1995	1	not annual as per original article description
Arc_044	Devon Ice Cap	75.33, −82.5	ic	1–1971	5	not annual
Arc_059	Renland	71.27, −26.73	ic	3–1983	5	not annual
Arc_045	Penny Ice Cap P96	67.25, −66.75	ic	5–1980	25	not annual
Arc_001	Blue Lake	68.09, −150.47	ls	730–2000	1	very nonlinear response, short overlap with instrumental, unclear interpretation
Arc_014	Lake Lehmilampi	63.62,29.1	ls	1–1800	1	exact interpretation unclear from original article
Arc_022	Hvítárvatn	64.6, −19.8	ls	−1–2000	1	annual and centennial signal inconsistent
Arc_069	Kongressvatnet	78.0217,13.9311	ls	232–2008	10	not annual
Arc_067	Hallet Lake	61.5, −146.2	ls	116–2005	11	not annual
Arc_050	Lake Hampträsk	60.28, 25.42	ls	1359–1994	14	not annual
Arc_070	Lake E	67, −50.7	ls	−3642–1876	19	not annual
Arc_043	Braya Sø	67, −50.7	ls	−998–1999	29	not annual
Arc_040	Moose Lake	61.35, −143.6	ls	−718–1963	36	not annual
Arc_051	Lake Pieni-Kauro	64.28, 30.12	ls	462–1979	44	not annual
Arc_041	Hudson Lake	61.9, −145.66	ls	−837–1997	47	not annual
Arc_054	Lake 4	65.1, −83.79	ls	634–1997	50	not annual
Arc_042	Screaming Lynx Lake	66.07, −145.4	ls	−1067–1988	51	not annual

Table D4. Summary of some of the major features of the new Arctic2K reconstruction ensemble. Anomalies are given relative to the period of 1850–2000.

Feature name	Year/period considered, years CE	Anomaly value (STD)°C	Note
Warmest century	901–1000	0.00(0.13)	after 750 CE
Warmest century-long period	927–1026	0.07(0.13)	MCA
Second warmest century-long period	1903–2002	0.01(0.05)	after MCA
Warmest decade	926–935	0.48(0.31)	
Second warmest decade	1993–2002	0.41(0.28)	outside MCA, 750 CE
Coldest century	1601–1700	−0.9(0.1)	
Coldest century-long period	1766–1865	−0.94(0.09)	
Coldest decade	1811–1820	−1.5(0.2)	
Second coldest decade	1463–1472	−1.4(0.2)	
Millennial-scale trend	1–1850	−0.05(0.01)	per century, before onset of the CWP
Largest warming trend magnitude	centred at 1905	0.01(0.001)	per year, centennial scale
Largest warming trend magnitude	centred at 1477	0.04(0.01)	per year, ca. 30-year scale
Largest cooling trend magnitude	centred at 1040	−0.006(0.002)	per year, centennial scale
Largest cooling trend magnitude	centred at 1450	−0.05(0.01)	per year, ca. 30-year scale

References

Altekar, G., Dwarkadas, S., Huelsenbeck, J. P., and Ronquist, F.: Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference, Bioinformatics, 20, 407–415, https://doi.org/10.1093/bioinformatics/btg427, 2004.

Anchukaitis, K. J., D'Arrigo, R. D., Andreu-Hayles, L., Frank, D., Verstege, A., Curtis, A., Buckley, B. M., Jacoby, G. C., and Cook, E. R.: Tree-Ring-Reconstructed Summer Temperatures from Northwestern North America during the Last Nine Centuries, J. Climate, 26, 3001–3012, https://doi.org/10.1175/jcli-d-11-00139.1, 2013 (data available at: https://www.ncdc.noaa.gov/paleo/study/14790, last access: 7 March 2017).

Babst, F., Poulter, B., Bodesheim, P., Mahecha, M. D., and Frank, D. C.: Improved tree-ring archives will support earth-system science, Nature Ecol. Evolut., 1, 1–2, 2017.

Bakke, J., Balascio, N., van der Bilt, W., Bradley, R., D'Andrea, W., Gjerde, M., Ólafsdóttir, S., Røthe, T., and De Wet, G.: The Island of Amsterdamøya: a key site for studying past climate in the Arctic Archipelago of Svalbard, Quaternary Sci. Rev., 183, 157–163, https://doi.org/10.1016/j.quascirev.2017.11.005, 2017.

Barclay, D. J., Wiles, G. C., and Calkin, P. E.: A 1119-year tree-ring-width chronology from western Prince William Sound, southern Alaska, Holocene, 9, 79–84, https://doi.org/10.1191/095968399672825976, 1999 (data available at: https://www.ncdc.noaa.gov/paleo/study/14274, last access: 7 March 2017).

Belt, S. T., Vare, L. L., Massé, G., Manners, H. R., Price, J. C., MacLachlan, S. E., Andrews, J. T., and Schmidt, S.: Striking similarities in temporal changes to spring sea ice occurrence across the central Canadian Arctic Archipelago over the last 7000 years, Quaternary Sci. Rev., 29, 3489–3504, https://doi.org/10.1016/j.quascirev.2010.06.041, 2010.

Berben, S. M. P., Husum, K., Cabedo-Sanz, P., and Belt, S. T.: Holocene sub-centennial evolution of Atlantic water inflow and sea ice distribution in the western Barents Sea, Clim. Past, 10, 181–198, https://doi.org/10.5194/cp-10-181-2014, 2014.

Björklund, J. A., Gunnarson, B. E., Seftigen, K., Esper, J., and Linderholm, H. W.: Blue intensity and density from northern Fennoscandian tree rings, exploring the potential to improve summer temperature reconstructions with earlywood information, Clim. Past, 10, 877–885, https://doi.org/10.5194/cp-10-877-2014, 2014 (data available at: https://www.ncdc.noaa.gov/paleo/study/3264, last access: 7 March 2017).

Briffa, K. R., Jones, P. D., and Schweingruber, F. H.: Summer temperature patterns over Europe: a reconstruction from 1750 AD based on maximum latewood density indices of conifers, Quaternary Res., 30, 36–52, 1988.

Briffa, K. R., Shishov, V. V., Melvin, T. M., Vaganov, E. A., Grudd, H., Hantemirov, R. M., Eronen, M., and Naurzbaev, M. M.: Trends in recent temperature and radial tree growth spanning 2000 years across northwest Eurasia, Phil. Trans. R. Soc. B, 363, 2271–2284, 2008 (data available at: https://www.ncdc.noaa.gov/paleo/study/14188, last access: 7 March 2017).

Briffa, K. R., Melvin, T. M., Osborn, T. J., Hantemirov, R. M., Kirdyanov, A. V., Mazepa, V. S., Shiyatov, S. G., and Esper, J.: Reassessing the evidence for tree-growth and inferred temperature change during the Common Era in Yamalia, northwest Siberia, Quaternary Sci. Rev., 72, 83–107, https://doi.org/10.1016/j.quascirev.2013.04.008, 2013 (data available at: https://crudata.uea.ac.uk/cru/papers/briffa2013qsr/, last access: 7 March 2017).

Büntgen, U., Trnka, M., Krusic, P. J., Kyncl, T., Kyncl, J., Luterbacher, J., Zorita, E., Ljungqvist, F. C., Auer, I., Konter, O., Scneider, L., Tegel, W., Štepánek, P., Brönniman, S., Hellmann, L., Nievergelt, D., and Esper, J.: Tree-Ring Amplification of the Early Nineteenth-Century Summer Cooling in Central Europe, J. Climate, 28, 5272–5288, 2015.

Chaudhuri, P. and Marron, J. S.: SiZer for exploration of structures in curves, J. Am. Stat. Assoc., 94, 807–823, 1997.

Christiansen, B.: Reconstructing the NH mean temperature: Can underestimation of trends and variability be avoided?, J. Climate, 24, 674–692, 2011.

Christiansen, B. and Ljungqvist, F. C.: Challenges and perspectives for large-scale temperature reconstructions of the past two millennia, Rev. Geophys., 55, 40–96, https://doi.org/10.1002/2016RG000521, 2017.

Clague, J. J., Luckman, B. H., Van Dorp, R. D., Gilbert, R., Froese, D., Jensen, B. J., and Reyes, A. V.: Rapid changes in the level of Kluane Lake in Yukon Territory over the last millennium, Quaternary Res., 66, 342–355, https://doi.org/10.1016/j.yqres.2006.06.005, 2006 (data available at: https://www.ncdc.noaa.gov/paleo/study/13785, last access: 7 March 2017).

Comboul, M., Emile-Geay, J., Evans, M. N., Mirnateghi, N., Cobb, K. M., and Thompson, D. M.: A probabilistic model of chronological errors in layer-counted climate proxies: applications to annually banded coral archives, Clim. Past, 10, 825–841, https://doi.org/10.5194/cp-10-825-2014, 2014.

Comiso, J. C., Parkinson, C. L., Gersten, R., and Stock, L.: Accelerated decline in the Arctic sea ice cover, Geophys. Res. Lett., 35, L01703, https://doi.org/10.1029/2007GL031972, 2008.

Cook, E. R., Briffa, K. R., and Jones, P. D.: Spatial regression methods in dendroclimatology – A review and comparison of 2 techniques, Int. J. Climatol., 14, 379–402, 1994.

Cook, T. L., Bradley, R. S., Stoner, J. S., and Francus, P.: Five thousand years of sediment transfer in a high arctic watershed recorded in annually laminated sediments from Lower Murray Lake, Ellesmere Island, Nunavut, Canada, J. Paleolimnol., 41, 77–94, https://doi.org/10.1007/s10933-008-9252-0, 2009.

Curry, J. A., Schramm, J. L., and Ebert, E. E.: Sea ice-albedo climate feedback mechanism, J. Climate, 8, 240–247, 1995.

D'Arrigo, R. D., Cook, E. R., Wilson, R. J., Allan, R., and Mann, M. E.: On the variability of ENSO over the past six centuries, Geophys. Res. Lett., 32, L03711, https://doi.org/10.1029/2004GL022055, 2005 (data available at: https://www.ncdc.noaa.gov/paleo/study/3044, last access: 7 March 2017).

D'Arrigo, R., Wilson, R., and Jacoby, G.: On the long-term context for late twentieth century warming, J. Geophys. Res., 111, D03103, https://doi.org/10.1029/2005jd006352, 2006 (data available at: https://www.ncdc.noaa.gov/paleo/study/13708, https://www.ncdc.noaa.gov/paleo/study/13758, last access: 7 March 2017).

Davi, N.: Boreal temperature variability inferred from maximum latewood density and tree-ring width data, Wrangell Mountain region, Alaska, Quaternary Res., 60, 252–262, https://doi.org/10.1016/s0033-5894(03)00115-7, 2003 (data available at: https://www.ncdc.noaa.gov/paleo/study/3605, last access: 7 March 2017).

Divine, D. and Dick, C.: Historical variability of sea ice edge position in the Nordic Seas, J. Geophys. Res., 111, C01001, https://doi.org/10.1029/2004JC002851, 2006.

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

Divine, D. V., Godtliebsen, F., and Rue, H.: A modelling approach to assessing the timescale uncertainties in proxy series with chronological errors, Clim. Past Discuss., 8, 31–61, https://doi.org/10.5194/cpd-8-31-2012, 2012.

Earl, D. J. and Deem, M. W.: Parallel tempering: Theory, applications, and new perspectives, Phys. Chem. Chem. Phys., 7, 3910–3916, https://doi.org/10.1039/B509983H, 2005.

Emile-Geay, J. and Tingley, M.: Inferring climate variability from nonlinear proxies: application to palaeo-ENSO studies, Clim. Past, 12, 31–50, https://doi.org/10.5194/cp-12-31-2016, 2016.

Esper, J., Frank, D., Timonen, M., Zorita, E., Wilson, R., Luterbacher, J., Holzkämper, S., Fischer, N., Wagner, S., Nievergelt, D., Verstege, A., and Büntgen, U.: Orbital forcing of tree-ring data, Nat. Clim. Change, 2, 862–866, https://doi.org/10.1038/NCLIMATE1589, 2012 (data available at: http://www.ncdc.noaa.gov/paleo/study/1003406, last access: 7 March 2017).

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

Fischer, H., Werner, M., Wagenbach, D., Schwager, M., Thorsteinnson, T., Wilhelms, F., Kipfstuhl, J., and Sommer, S.: Little ice age clearly recorded in northern Greenland ice cores, Geophys. Res. Lett., 25, 1749–1752, https://doi.org/10.1029/98GL01177, 1998.

Frank, D., Esper, J., and Cook, E.: On variance adjustments in tree-ring chronology development, TRACE, 4, 56–66, 2006.

Franke, J., Frank, D., Raible, C. C., Esper, J., and Brönnimann, S.: Spectral biases in tree-ring climate proxies, Nat. Clim. Change, 3, 360–364, https://doi.org/10.1038/NCLIMATE1816, 2013.

Gneiting, T. and Raftery, A. E.: Strictly proper scoring rules, prediction, and estimation, J. Am. Stat. Assoc., 102, 359–378, 2007.

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

Grinsted, A., Moore, J. C., Pohjola, V., Martma, T., and Isaksson, E.: Svalbard summer melting, continentality, and sea ice extent from the Lomonosovfonna ice core, J. Geophys. Res., 111, D07110, https://doi.org/10.1029/2005JD006494, 2006.

Grinsted, A., Moore, J. C., and Jevrejeva, S.: Reconstructing sea level from paleo and projected temperatures 200 to 2100 AD, Clim. Dynam., 34, 461–472, 2010.

Grootes, P. and Stuiver, M.: Oxygen 18/16 variability in Greenland snow and ice with $10^{-3}-10^{-5}$-year time resolution, J. Geophys. Res., 102, 26455–26470, https://doi.org/10.1029/97JC00880, 1997.

Hanhijärvi, S., Tingley, M. P., and Korhola, A.: Pairwise comparisons to reconstruct mean temperature in the Arctic Atlantic Region over the last 2000 years, Clim. Dynam., 41, 2039–2060, 2013.

Harris, I., Jones, P., Osborn, T., and Lister, D.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, Int. J. Climatol., 34, 623–642, https://doi.org/10.1002/joc.3711, 2014.

Helama, S., Vartiainen, M., Holopainen, J., Mäkelä, H., Kolström, T., and Meriläinen, J.: A palaeotemperature record for the Finnish Lakeland based on microdensitometric variations in tree rings, Geochronometria, 41, 265–277, https://doi.org/10.2478/s13386-013-0163-0, 2014 (data available at: http://www.ncdc.noaa.gov/paleo/study/16790, last access: 7 March 2017).

Hersbach, H.: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, Weather Forecast., 15, 559–570, 2000.

Hoffmann, G., Jouzel, J., and Johnsen, S.: Deuterium excess record from central Greenland over the last millennium: Hints of a North Atlantic signal during the Little Ice Age, J. Geophys. Res., 106, 14265–14274, https://doi.org/10.1029/2000JD900585, 2001.

Hughen, K. A., Overpeck, J. T., and Anderson, R. F.: Recent warming in a 500-year palaeotemperature record from varved sediments, Upper Soper Lake, Baffin Island, Canada, Holocene, 10, 9–19, https://doi.org/10.1191/095968300676746202, 2000.

Hughes, M. K., Vaganov, E. A., Shiyatov, S., Touchan, R., and Funkhouser, G.: Twentieth-century summer warmth in northern Yakutia in a 600-year context, Holocene, 9, 629–634, 1999 (data available at: https://www.ncdc.noaa.gov/paleo/study/14188, last access: 7 March 2017).

Hunke, E. C., Lipscomb, W. H., and Turner, A. K.: Sea-ice models for climate study: retrospective and new directions, J. Glaciol.,

56, 1162–1172, https://doi.org/10.3189/002214311796406095, 2010.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/CBO9781107415324, 2013.

Jacoby, G. C. and D'Arrigo, R.: Reconstructed Northern Hemisphere annual temperature since 1671 based on high-latitude tree-ring data from North America, Clim. Change, 14, 39–59, https://doi.org/10.1007/bf00140174, 1989 (data available at: https://www.ncdc.noaa.gov/paleo/study/3592, https://www.ncdc.noaa.gov/paleo/study/3598, last access: 7 March 2017).

Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., and Morice, C. P.: Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010, J. Geophys. Res., 117, D05127, https://doi.org/10.1029/2011JD017139, 2012.

Kaufman, D. S., Schneider, D. P., McKay, N. P., Ammann, C. M., Bradley, R. S., Briffa, K. R., Miller, G. H., Otto-Bliesner, B. L., Overpeck, J. T., Vinther, B. M., and 2k Project Members, A. L.: Recent Warming Reverses Long-Term Arctic Cooling, Science, 325, 1236–1239, 2009.

Kinnard, C., Zdanowicz, C. M., Fisher, D. A., Isaksson, E., de Vernal, A., and Thompson, L. G.: Reconstructed changes in Arctic sea ice over the past 1,450 years, Nature, 479, 509–512, https://doi.org/10.1038/nature10581, 2011.

Lamoureux, S. F. and Bradley, R. S.: A late Holocene varved sediment record of environmental change from northern Ellesmere Island, Canada, J. Paleolimnol., 16, 239–255, https://doi.org/10.1007/BF00176939, 1996.

Leopardi, P.: A partition of the unit sphere into regions of equal area and small diameter, Electron. T. Numer. Ana., 25, 309–327, 2006.

Li, Z., Protopopescu, V. A., Arnold, N., Zhang, X., and Gorin, A.: Hybrid parallel tempering and simulated annealing method, Appl. Math. Comput., 212, 216–228, https://doi.org/10.1016/j.amc.2009.02.023, 2009.

Ljungqvist, F. C., Krusic, P. J., Brattström, G., and Sundqvist, H. S.: Northern Hemisphere temperature patterns in the last 12 centuries, Clim. Past, 8, 227–249, https://doi.org/10.5194/cp-8-227-2012, 2012.

Ljungqvist, F. C., Krusic, P. J., Sundqvist, H. S., Zorita, E., Brattström, G., and Frank, D.: Northern Hemisphere hydroclimate variability over the past twelve centuries, Nature, 532, 94–98, 2016.

Lorenz, E. N.: Empirical orthogonal functions and statistical weather prediction, Tech. Rep. 1, Massachusetts Institute of Technology, Boston MA, USA, 1956.

Luterbacher, J., Werner, J. P., Smerdon, J. E., Fernández-Donado, L., González-Rouco, F. J., Barriopedro, D., Ljungqvist, F. C., Büntgen, U., Zorita, E., Wagner, S., Esper, J., McCarroll, D., Toreti, A., Frank, D., Jungclaus, J. H., Barriendos, M., Bertolin, C., Bothe, O., Brázdil, R., Camuffo, D., Dobrovolný, P., Gagen, M., García-Bustamante, E., Ge, Q., Gómez-Navarro, J. J., Guiot, J., Hao, Z., Hegerl, G. C., Holmgren, K., Klimenko, V. V., Martín-Chivelet, J., Pfister, C., Roberts, N., Schindler, A., Schurer, A., Solomina, O., von Gunten, L., Wahl, E., Wanner, H., Wetter, O., Xoplaki, E., Yuan, N., Zanchettin, D., Zhang, H., and

Zerefos, C.: European summer temperatures since Roman times, Environ. Res. Lett., 11, 024001, https://doi.org/10.1088/1748-9326/11/2/024001, 2016.

MacDonald, G. M., Case, R. A., and Szeicz, J. M.: A 538-Year Record of Climate and Treeline Dynamics from the Lower Lena River Region of Northern Siberia, Russia, Arctic Alpine Res., 30, 334, https://doi.org/10.2307/1552005, 1998 (data available at: https://www.ncdc.noaa.gov/paleo/study/3902, last access: 7 March 2017).

Maslanik, J., Stroeve, J., Fowler, C., and Emery, W.: Distribution and trends in Arctic sea ice age through spring 2011, Geophys. Res. Lett., 38, L13502, https://doi.org/10.1029/2011GL047735, 2011.

Masson-Delmotte, V., Landais, A., Stievenard, M., Cattani, O., Falourd, S., Jouzel, J., Johnsen, S. J., Dahl-Jensen, D., Sveinsbjornsdottir, A., White, J. W. C., Popp, T., and Fischer, H.: Holocene climatic changes in Greenland: Different deuterium excess signals at Greenland Ice Core Project (GRIP) and NorthGRIP, J. Geophys. Res., 110, 14102, https://doi.org/10.1029/2004JD005575, 2005.

McCarroll, D., Loader, N. J., Jalkanen, R., Gagen, M. H., Grudd, H., Gunnarson, B. E., Kirchhefer, A. J., Friedrich, M., Linderholm, H. W., Lindholm, M., Boettger, T., Los, S. O., Remmele, S., Kononov, Y. M., Yamazaki, Y. H., Young, G. H., and Zorita, E.: A 1200-year multiproxy record of tree growth and summer temperature at the northern pine forest limit of Europe, Holocene, 23, 471–484, https://doi.org/10.1177/0959683612467483, 2013 (data available at: https://www.ncdc.noaa.gov/paleo/study/19943, last access: 7 March 2017).

McKay, N. P. and Kaufman, D. S.: An extended Arctic proxy temperature database for the past 2,000 years, Scientific Data, 1, 140026, https://doi.org/10.1038/sdata.2014.26, 2014.

Meier, W. N., Hovelsrud, G. K., van Oort, B. E., Key, J. R., Kovacs, K. M., Michel, C., Haas, C., Granskog, M. A., Gerland, S., Perovich, D. K., Makshtas, A., and Reist, J. D.: Arctic sea ice in transformation: A review of recent observed changes and impacts on biology and human activity, Rev. Geophys., 52, 185–217, https://doi.org/10.1002/2013RG000431, 2014.

Melvin, T. M., Grudd, H., and Briffa, K. R.: Potential bias in "updating" tree-ring chronologies using regional curve standardisation: Re-processing 1500 years of Torneträsk density and ring-width data, Holocene, 23, 364–373, https://doi.org/10.1177/0959683612460791, 2013 (data available at: https://crudata.uea.ac.uk/cru/papers/melvin2012holocene/, last access: 7 March 2017).

Miettinen, A., Divine, D. V., Husum, K., Koç, N., and Jennings, A.: Exceptional ocean surface conditions on the SE Greenland shelf during the Medieval Climate Anomaly, Paleoceanography, 30, 1657–1674, https://doi.org/10.1002/2015PA002849, 2015.

Miller, G. H., Alley, R. B., Brigham-Grette, J., Fitzpatrick, J. J., Polyak, L., Serreze, M. C., and White, J. W.: Arctic amplification: can the past constrain the future?, Quaternary Sci. Rev., 29, 1779–1790, 2010.

Mitchell, J., Dzerdzeevskii, B., and Flohn, H.: Climatic change: report of a working group of the Commission for Climatology, Technical Note 79, World Meteorological Organisation, World Meteorological Organisation, Geneva, Switzerland, 1966.

Moore, J., Hughen, K., Miller, G., and Overpeck, J.: Little Ice Age recorded in summer temperature reconstruction from vared sediments of Donard Lake, Baffin Island, Canada, J. Paleolimnol., 25, 503–517, https://doi.org/10.1023/A:1011181301514, 2001.

Müller, J., Werner, K., Stein, R., Fahl, K., Moros, M., and Jansen, E.: Holocene cooling culminates in sea ice oscillations in Fram Strait, Quaternary Sci. Rev., 47, 1–14, 2012.

Nilsen, T., Rypdal, K., and Fredriksen, H.-B.: Are there multiple scaling regimes in Holocene temperature records?, Earth Syst. Dynam., 7, 419–439, https://doi.org/10.5194/esd-7-419-2016, 2016.

Nilsen, T., Werner, J. P., and Divine, D. V.: How wrong are climate field reconstruction techniques in reconstructing a climate with long-range memory?, Clim. Past Discuss., https://doi.org/10.5194/cp-2018-17, in review, 2018.

Nychka, D., Buchberger, R., Wigley, T., Santer, B. D., Taylor, K. E., and Jones, R. H.: Confidence Intervals for Trend Estimates With Autocorrelated Observations, unpublished manuscript, available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.33.6828&rep=rep1&type=pdf (last access: 2 April 2018), 2000.

O'Connor, F. M., Boucher, O., Gedney, N., Jones, C. D., Folberth, G. A., Coppell, R., Friedlingstein, P., Collins, W. J., Chappellaz, J., Ridley, J., and Johnson, C. E.: Possible role of wetlands, permafrost, and methane hydrates in the methane cycle under future climate change: a review, Rev. Geophys., 48, 2010RG000326, https://doi.org/10.1029/2010RG000326, 2010.

Ojala, A. E. and Alenius, T.: 10 000 years of interannual sedimentation recorded in the Lake Nautajärvi (Finland) clastic–organic varves, Palaeogeogr. Palaeocl., 219, 285–302, https://doi.org/10.1016/j.palaeo.2005.01.002, 2005.

Opel, T., Fritzsche, D., and Meyer, H.: Eurasian Arctic climate over the past millennium as recorded in the Akademii Nauk ice core (Severnaya Zemlya), Clim. Past, 9, 2379–2389, https://doi.org/10.5194/cp-9-2379-2013, 2013.

Østvand, L., Nilsen, T., Rypdal, K., Divine, D., and Rypdal, M.: Long-range memory in internal and forced dynamics of millennium-long climate model simulations, Earth Syst. Dynam., 5, 295–308, https://doi.org/10.5194/esd-5-295-2014, 2014.

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., and Zielinski, G.: Arctic environmental change of the last four centuries, Science, 278, 1251–1256, 1997.

PAGES 2k Consortium: Continental-scale temperature variability during the past two millennia, Nat. Geosci., 6, 339–346, 2013.

PAGES 2k Consortium: A global multiproxy database for temperature reconstructions of the Common Era, Scientific Data, 4, 170088, https://doi.org/10.1038/sdata.2017.88, 2017 (data available at: https://www.ncdc.noaa.gov/paleo/study/3264, https://www.ncdc.noaa.gov/paleo/study/3706, https://www.ncdc.noaa.gov/paleo/study/5244, last access: 7 March 2017).

Perovich, D., Grenfell, T., Light, B., and Hobbs, P.: Seasonal evolution of the albedo of multiyear Arctic sea ice, J. Geophys. Res.-Oceans, 107, 8044, https://doi.org/10.1029/2000JC000438, 2002.

Perovich, D., Jones, K., Light, B., Eicken, H., Markus, T., Stroeve, J., and Lindsay, R.: Solar partitioning in a changing Arctic sea-ice cover, Ann. Glaciol., 52, 192–196, 2011.

Perovich, D. K., Richter-Menge, J. A., Jones, K. F., and Light, B.: Sunlight, water, and ice: Extreme Arctic sea ice melt during the summer of 2007, Geophys. Res. Lett., 35, L11501, https://doi.org/10.1029/2008GL034007, 2008.

Rohde, R., Muller, R., Jacobsen, R., Muller, E., Perlmutter, S., Rosenfeld, A., Wurtele, J., Groom, D., and Wickham, C.: A new estimate of the average Earth surface land temperature spanning 1753 to 2011, Geoinfor. Geostat. Overview, 1, 1–7, 2013.

Schneider, L., Smerdon, J. E., Büntgen, U., Wilson, R. J. S., Myglan, V. S., Kirdyanov, A. V., and Esper, J.: Revising midlatitude summer temperatures back to A.D. 600 based on a wood density network, Geophys. Res. Lett., 42, 4556–4562, https://doi.org/10.1002/2015gl063956, 2015 (data available at: https://www.ncdc.noaa.gov/paleo/study/19743, last access: 7 March 2017).

Serreze, M. C., Holland, M. M., and Stroeve, J.: Perspectives on the Arctic's shrinking sea-ice cover, Science, 315, 1533–1536, 2007.

Shakhova, N., Semiletov, I., Salyuk, A., Yusupov, V., Kosmach, D., and Gustafsson, Ö.: Extensive methane venting to the atmosphere from sediments of the East Siberian Arctic Shelf, Science, 327, 1246–1250, 2010.

Shi, F., Yang, B., Ljungqvist, F. C., and Yang, F.: Multi-proxy reconstruction of Arctic summer temperatures over the past 1400 years, Climate Res., 54, 113–128, 2012.

Sigl, M., Winstrup, M., McConnell, J. R., Welten, K. C., Plunkett, G., Ludlow, F., Büntgen, U., Caffee, M., Chellman, N., Dahl-Jensen, D., Fischer, H., Kipfstuhl, S., Kostick, C., Maselli, O. J., Mekhaldi, F., Mulvaney, R., Muscheler, R., Pasteris, D. R., Pilcher, J. R., Salzer, M., Schüpbach, S., Steffensen, J. P., Vinther, B. M., and Woodruff, T. E.: Timing and climate forcing of volcanic eruptions for the past 2,500 years, Nature, 523, 543–549, 2015.

Smedsrud, L. H., Sorteberg, A., and Kloster, K.: Recent and future changes of the Arctic sea-ice cover, Geophys. Res. Lett., 35, L20503, https://doi.org/10.1029/2008GL034813, 2008.

Thomas, E. K. and Briner, J. P.: Climate of the past millennium inferred from varved proglacial lake sediments on northeast Baffin Island, Arctic Canada, J. Paleolimnol., 41, 209–224, https://doi.org/10.1007/s10933-008-9258-7, 2009.

Tingley, M. P.: A Bayesian ANOVA scheme for calculating climate anomalies, with applications to the instrumental temperature record, J. Climate, 25, 777–791, https://doi.org/10.1175/JCLI-D-11-00008.1, 2012.

Tingley, M. P. and Huybers, P.: A Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time. Part I: Development and Applications to Paleoclimate Reconstruction Problems, J. Climate, 23, 2759–2781, https://doi.org/10.1175/2009JCLI3015.1, 2010a.

Tingley, M. P. and Huybers, P.: A Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time. Part II: Comparison with the Regularized Expectation-Maximization Algorithm, J. Climate, 23, 2782–2800, https://doi.org/10.1175/2009JCLI3016.1, 2010b.

Tingley, M. P. and Huybers, P.: Recent temperature extremes at high northern latitudes unprecedented in the past 600 years, Nature, 496, 201–205, https://doi.org/10.1038/nature11969, 2013.

Tipton, J., Hooten, M., Pederson, N., Tingley, M., and Bishop, D.: Reconstruction of late Holocene climate based on tree growth

and mechanistic hierarchical models, Environmetrics, 27, 42–54, https://doi.org/10.1002/env.2368, 2016.

van der Bilt, W., Bakke, J., Vasskog, K., D'Andrea, W., Bradley, R., and Ólafsdóttir, S.: Reconstruction of glacier variability from lake sediments reveals dynamic Holocene climate in Svalbard, Quaternary Sci. Rev., 126, 201–218, 2015.

Vermeer, M. and Rahmstorf, S.: Global sea level linked to global temperature, P. Natl. Acad. Sci. USA, 106, 21527–21532, 2009.

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

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

Walsh, J. E., Fetterer, F., Scott Stewart, J., and Chapman, W. L.: A database for depicting Arctic sea ice variations back to 1850, Geogr. Rev., 107, 89–107, 2017.

Wang, J., Emile-Geay, J., Guillot, D., McKay, N. P., and Rajaratnam, B.: Fragility of reconstructed temperature patterns over the Common Era: Implications for model evaluation, Geophys. Res. Lett., 42, 7162–7170, https://doi.org/10.1002/2015GL065265, 2015.

Weißbach, S., Wegner, A., Opel, T., Oerter, H., Vinther, B. M., and Kipfstuhl, S.: Spatial and temporal oxygen isotope variability in northern Greenland – implications for a new climate record over the past millennium, Clim. Past, 12, 171–188, https://doi.org/10.5194/cp-12-171-2016, 2016.

Werner, J. P. and Tingley, M. P.: Technical Note: Probabilistically constraining proxy age–depth models within a Bayesian hierarchical reconstruction model, Clim. Past, 11, 533–545, https://doi.org/10.5194/cp-11-533-2015, 2015.

Werner, J., Smerdon, J., and Luterbacher, J.: A Pseudoproxy Evaluation of Bayesian Hierarchical Modelling and Canonical Correlation Analysis for Climate Field Reconstructions over Europe, J. Climate, 26, 851–867, https://doi.org/10.1175/JCLI-D-12-00016.1, 2013.

Wiles, G. C., D'Arrigo, R. D., Barclay, D., Wilson, R. S., Jarvis, S. K., Vargo, L., and Frank, D.: Surface air temperature variability reconstructed with tree rings for the Gulf of Alaska over the past 1200 years, Holocene, 24, 198–208, https://doi.org/10.1177/0959683613516815, 2014 (data available at: https://www.ncdc.noaa.gov/paleo/study/19743, last access: 7 March 2017).

Yamanouchi, T.: Early 20th century warming in the Arctic: A review, Polar Science, 5, 53–71, https://doi.org/10.1016/j.polar.2010.10.002, 2011.

Zhang, H., Yuan, N., Esper, J., Werner, J. P., Xoplaki, E., Büntgen, U., Treydte, K., and Luterbacher, J.: Modified climate with long term memory in tree ring proxies, Environ. Res. Lett., 10, 084020, https://doi.org/10.1088/1748-9326/10/8/084020, 2015.

Zhang, H., Werner, J. P. García-Bustamante, E., González-Rouco, F., Wagner, S., Zorita, E., Fraedrich, K., Ljungqvist, F. C., Zhu, X., Xoplaki, E., Chen, F., Duan, J., Ge, Q., Hao, Z., Ivanov, M., Schneider, L., Talento, S., Wang, J., Yang, B., and Luterbacher, J.: East Asian warm season temperature variations over the past two millennia, Scientific Reports, accepted, 2018.

Zhang, P., Linderholm, H. W., Gunnarson, B. E., Björklund, J., and Chen, D.: 1200 years of warm-season temperature variability in central Scandinavia inferred from tree-ring density, Clim. Past, 12, 1297–1312, https://doi.org/10.5194/cp-12-1297-2016, 2016 (data available at: https://www.ncdc.noaa.gov/paleo/study/19743, last access: 7 March 2017).