Figure 1: Correlation between two grid cells vs. their distance in the BEST data

We have extensively rewritten the article: The chapter about the Arctic Amplification was removed. Instead we have now included a more in-depth discussion (and analysis) of the trends over the Common Era.

As noted in one of our Discussion replies, there was an error in how the lake and ice core data was imported into BARCAST. This resulted in the strange trend pattern over Greenland. We identified the issue now, and the reconstruction was updated accordingly. Thus, many changes in the article were neccesary, on top of those pointed out by the reviewers.

We are grateful for the patience of both the editors in charge and the reviewers, and hope that this revised manuscript can now be considered for publication. Below follows the point-by-point reply to the issues raised during the review.

1 Reply to Reviewer 1

General Comment

(paraphrased) No scientific question explicitly posed, article lacking focus.

Reply: We will remove the Arctic Amplification chapter, which will result already in a more coherent manuscript, making connections between the different chapters much clearer. After revising the reconstruction we will also try and make stronger statements about e.g. the spatial consistency of the warming and cooling episodes.

Other comments

- Abstract, lines 1-2: I think you need to be a little more clear about what is actually unique here. I understand that it is BOTH spatially resolved AND millennial in length, though there are several reconstructions that are one or the other.

R: we changed that statement in the final version of the paper and stress that it is both the spatial character and temporal aspect (as guessed correctly)

- Paragraph including lines 53-60: This paragraph comes across as a kind of special attack on the glacier advances work in a tone that Im not sure the authors intended. This summer temperature reconstruction (with skill primarily over Europe) is really different than a glacial reconstruction given the memory of glaciers, the different seasons and climate factors a glacier is responding to, etc. So I dont think a clear declaration against that work is necessarily warranted.

R: We apologize for the perceived tone. We did not mean to criticize glacial reconstructions from moraines in general and cosmogenic dating in particular. The sentence will be removed as it is also based on a misinterpretation of the results by Young et al. (pers. comm. from Young set this straight).

- define LOC

R: it stands for "local regression"

- Lines 124-125: Need a more specific criticism here or not discuss the issue at all. What constitutes a strange correlation? And on what firm basis can you reject the use of the BE product?

R: the issue is that the correlation between grid cells as a function of the distance (both chordal and orthodromic) is very long (10 000 km) and contains oscillatory parts (not as obvious in the attached figure as in other evaluations). Without analysing any details of the regridding method used in the BEST data, this looks too much like an artefact of an expansion in spherical functions. Truncating after a certain order can lead to spurious oscillations on the sphere. See the attached figure.

- Lines 147, 444: Question mark issues.

R: missing reference in the bibTeX file. fixed

- Lines 199-202: How are the response parameters being determined? Do your reconstructions happen to take account of the specific choice of parameter values?

R: The parameters are estimated using the described Gibbs sampler. The reconstructions are conditional on the estimated joint distribution of all parameters (proxy response and climate field). That is, it explicitly takes the uncertainty in the parameters into account.

- Lines 211-222: I think its important to note that only Europe has spatially coherent skill, otherwise its fairly patchy skill (at least in my reading of Fig. A1).

R: The skill shown in Figure A1 depends on the length (and quality) of the instrumental data in the specific grid cell. The data coverage over Europe is highest (space

and time), other regions (absence of colour in Fig. A1) are less well covered (thus the patchy appearance). Any estimates relying on short time series are thus to be interpreted carefully. This is what we mean by "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." (l. 550-551)

- Figure A1: It wasnt clear to me what is meant by CRPS CE and CRPS RE? As a related issue, CRPS is challenging to interpret because it doesnt have a reference. Perhaps use the skill score version of CRPS that takes account of a reference (e.g., your prior)?

R: We did indeed skim over this issue. We modified the last two paragraphs accordingly, and (hopefully) made the reasoning behind this much clearer.

- Fig 2b: Why is there so much precision right up to the end of the calibration interval, but a complete loss of annual precision from 1980 to the present? Are important proxies dropping out here?

R: The precision appears indeed relatively low, and Fig.2b greatly emphasises this over Fig.2c, especially since we have the "calibration" interval so prominently in there. One issue is that "calibration" is not the entirely correct term, as the reconstruction over this period is in fact mostly determined by the instrumental data (though not entirely, not the relatively high instrumental uncertainty and the spatially sparse coverage). This issue can be addressed by doing what is called a predictive run (see Luterbacher et al. 2016, or Tingley and Huybers 2013), which would in turn give another means of evaluating the reconstruction quality. However, as can be seen from 2d, this is also very likely a proxy availability effect.

- Figure 5: no color on the color labels

R: These were present in the initially uploaded pdf, we will look into the technical issues behind that. Most likely once the original artwork is uploaded and the LaTeX process takes place at Copernicus this issue will disappear.

Typos / grammatical issues:

- Lines 458-459 which gets sparser going back in time

- Line 5:14 used for these chronologies

Thanks for catching these!

We are greatful for the very detailed reveiw and the extremely thorough commenting in the uploaded pdf! In the following we will try to first answer the most important (as we understood them) comments and take the rest in shorter form later.

2 Warming trend

1. The trend map from 1–1850 CE. The gridded reconstruction itself is (repeatedly mentioned!) only going back to 750 CE.

2. The warming trend over Greenland is not supported by the raw proxy data

1. True, although the assessment this is based on relies on the spread of the annual reconstruction in the past. The trend should in principle be more robust against this. We did additionally plot a trend analysis going back only to 750 CE, though of course this could be influenced by the onset of the MCA.

2. Even more important is it to show a map with an estimated SNR (or the plain β_1, the scaling of the individual proxies) – as also commented by the reviewer. See Fig. A6 for the current estimates.

3 Arctic Amplification

The comparability of reconstructions is noted as questionable (or at least difficult) in the text, and thus the exercise is criticised (rightly so). There also seemed to be a misunderstanding regarding the analysis, in the sense that we limit ourselves to the European sector of the arctic – the circum-polar behaviour being indeed quite different timing and amplitude wise. We have decided to remove this chapter entirely, and wait for comprehensive reconstructions based on the North American and the Asian data.

4 Lake data

We made indeed a linear response assumption for the lake data, which might be defended by noting that the interpretation of some of the used records is based on the cross correlation to instrumental data, i.e. the linear assumption. However, we have now decided to transform the data using the inverse quantile transformation.

5 Other comments

The reviewer caught several typos and a few strange (i.e. wrong) grammatical constructs, these will (of course) be fixed.

- Fig 7, flip axes? add data coverage

R: We will add the data coverage. We will also try to toy with the axes orientation, however we feel that the current orientation with the longitude in the "natural" orientation is superior to a time axis going left to right. We also updated the estimate of what constitutes a "significant" warm or cold event.

- C1: how is this screenign used?

R: This is a misunderstanding, we mostly did this to see how the reconstruction changes the LRM properties. It has been commented by others (also at meetings and conferences) that basing the reconstruciton on an explicit spatial and temporal model is bound to change the correlation (in space and time) behaviour of the resulting reconstruction (see e.g. Raible et al. Clim.Chage 2006 for the effect on the spatial correlations in EOF based reconstructions). In principle, all spatio-temporal reconstruction mehtods impose an explicit model. In the classical world (PCA, CCA, . . .), the spatial patterns are truncated and the temporal process is assumed to be i.i.d. In other methods (Ed Cook's PPR), the data is pre-whitened to remove auto-correlations first, and then do a regression, while basing the spatial correlations on the instrumental period and imposing a convex spatial structure through the search radius.

Spatio-temporal variability of Arctic summer temperatures over the past two millennia: ~~an overview of the last major climate anomalies~~

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, PO Box 7803, N-5020 Bergen, Norway
[2]Norwegian Polar Institute, FRAM Centre, N-9296 Tromsø, Norway.
[3]Department of Mathematics and Statistics, University of Tromsø – The Arctic University of Norway, N-9037, Norway
[4]Department of History, Stockholm University, SE-106 91 Stockholm, Sweden
[5]Bolin Centre for Climate Research, Stockholm University, SE-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 to: J.P. Werner (johannes.werner@geo.uib.no)

Abstract.

In this article, the first spatially resolved ~~millennium-long~~ and millennium-length summer (June–August) temperature reconstruction over the Arctic and Subarctic domain (north of 60° N) is presented. It is based on a set of ~~54~~ 44 annually dated temperature sensitive proxy archives of various types, mainly from the updated and revised PAGES2k database supplemented with 6 new recently ~~published~~ updated proxy records. As a major ~~novelty~~ advance, an extension of the Bayesian BARCAST climate field (CF) reconstruction technique provides a means to treat climate archives with dating uncertainties. In total ~~1400~~ over 600 independent realisations of the temperature CF were generated, enabling further analyses to be carried out in a probabilistic framework. 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 analysis of basic features of the reconstructed seasonal CF focuses on the regional expression of past major climate anomalies in order to uncover the potential of the new product for studying Common Era temperature variability in the region.

The long-term, 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 sub-regions. The reconstruction shows a pronounced Medieval Climate Anomaly (MCA, here, ca. ~~960–1060~~ 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 ~~1580–1680~~ 1766–1865 CE as the longest centennial-scale cold period, culminating around ~~1812–1822~~ 1811–1820 CE for most of the target region. ~~At the same time there is evidence for a drastic reduction in sea-ice on the Greenland shelf, which is reflected by rather high summer temperatures over Greenland and Baffin Island during that decade.~~

20 ~~During the MCA, the contrast between reconstructed summer temperatures over mid- and high-latitudes in Europe and the European/North Atlantic sector of the Arctic shows a very dynamic expression of the Arctic amplification, with leads and lags between continental and more marine and extreme latitude settings.~~ While our analysis shows that the peak MCA summer temperatures were as high as in the late 20th and early 21st century, 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

25 provide conclusive support of the contemporary warming to ~~supersede~~ exceed the peak of the MCA in terms of the pan-Arctic mean summer temperatures: neither can the reconstruction be extended reliably past 2002 CE due to lack of proxy data and thus the most recent warming is not captured, nor is it (from a statistical viewpoint) advisable to directly compare the reconstruction and instrumental data.

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 time-scales. Moreover, general circulation models have limited capabilities in reliably simulating Arctic climate change on centennial time-scale and beyond (IPCC, 2013). The simplified parametrisations 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 ~~uniqueness~~ unprecedented nature of the on-going Arctic warming during the Common Era (CE~~) of the on-going Arctic warming~~, 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 time-scales are needed, ~~and an effort needs to be made to~~ together with methods that integrate different types of information from a variety of palaeoclimate archives.

~~Especially the spatial signature of past Arctic temperature variability remains poorly understood and has recently been a contested issue. Young et al. (2015), based on results from relatively poorly age-constrained moraine dates of glacier advances, questioned the long-standing notion within palaeoclimatology that Greenland was exceptionally warm around the time of Norse settlement (in the 980s CE). They argue that the medieval warming in the Arctic did not extend to Greenland, thus challenging our long-held understanding of the settlement and later abandonment of Norse Greenland. This is in direct contradiction to the spatial temperature reconstructions of Mann et al. (2009) and Ljungqvist et al. (2012, 2016), which instead point to very warm conditions in Greenland at that time. As we will see, our results show that the Greenland sector of the Arctic indeed was warm during the 10th and especially the early 11th centuries CE.~~

Since the 1990s, several multi-proxy reconstructions of Arctic and Subarctic (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 2,000-year 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 two 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 (c.f. 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 decadal-to-decadal variability. Shi et al. (2012) found a clear cold anomaly 630 to 770 CE, a peak warming ca. 950 to 1050 CE, and overall relatively cold conditions ca. 1200–1900 CE. However, three distinctly warmer periods during the Little Ice Age were reconstructed 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 been 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 sec. 3), to reconstruct surface-air temperatures of the last 600 years over land north of 60° N. The Their reconstruction is mostly based on the proxy dataset collected by the PAGES 2k Consortium (2013). They found that while the recent decades were the warmest over the last 600 years, the actual inter-annual variability has remained effectively constant. Much of the data used therein is (most of the ice core and lake sediment records) used therein are common with the work presented here, with a few updated records (see section 2.2, and PAGES 2k Consortium, 2017) and a few additional proxies (such as the tree ring width series).

Hanhijärvi et al. (2013) presented a 2000-year long annual mean temperature reconstruction for the North Atlantic sector 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 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 Little Ice Age 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 Medieval Climate Anomaly were found to equal recent temperatures in the 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 two millennia. They reconstructed a generally relatively warm first millennium CE, ~~and~~ followed by a relatively indistinct Medieval Climate Anomaly, and a relatively cold Little Ice Age 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 is on average ca. 0.5°C warmer prior to ca. 1250 CE than ~~in~~ reported by PAGES 2k Consortium (2013). Peak temperatures during the Roman Warm Period and the Medieval Climate Anomaly 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 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 dataset is directly derived from ~~available data from meteorological stations. ii~~meteorological observation data without any interpolation over grid cell boundaries. iii) The gridded reconstruction goes back into the first millennium CE. ~~iii~~iii) The proxy dataset is larger and more extensively screened, and ~~iv~~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. ~~v~~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 strong to the seasonal temperature anomalies, and should increase the precision at locations with stronger climate response.

2 Instrumental data and proxy data

The following section provides a short overview of the used instrumental and palaeoclimate proxy data. 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 ~~(c.f. e.g. Wang et al., 2015)~~(see e.g. Wang et al., 2015).

2.1 Instrumental data

Several different gridded data sets for ~~earth~~ Earth surface air temperatures (SAT) are available from different research groups, derived from different subsets of instrumental data and presented on different types of grids. Most datasets, like e.g. CRUTEM4 (Jones et al., 2012) or CRU TS3 (Harris et al., 2014) are presented on a regular equilateral grid, such as a $5°\times5°$ grid. Such a regular grid exhibits severe shortcomings when analysing data close to the poles, as the grid cells become very narrow in meridional direction and almost triangular shaped. One data set, the Berkeley Earth Surface Temperature (BEST) (Rohde et al., 2013), is offered on a $1°\times1°$ grid as well as an equal area grid. While the latter would be a good fit for the process level model
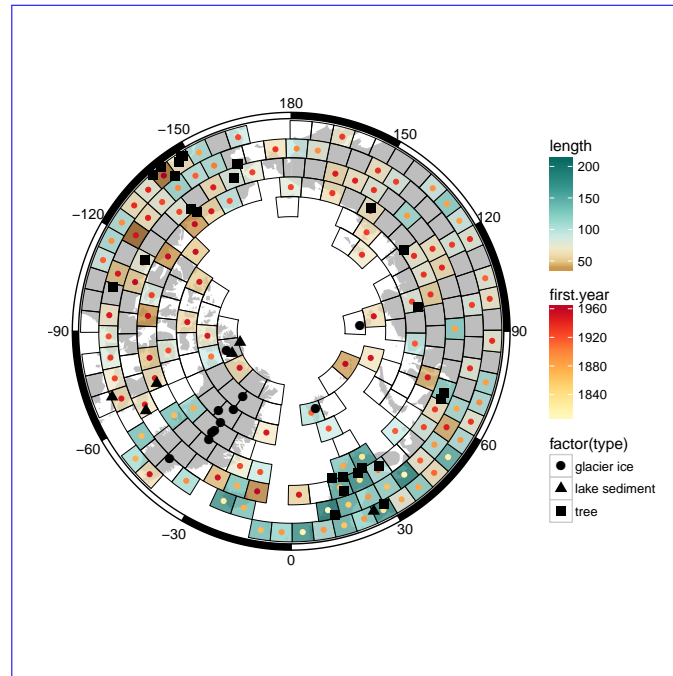
Figure 1. Distribution of input data. Length (fill of quadrilaterals) and first year (coloured circles) of the regridded instrumental data. Symbols show the locations and type of used proxy data (PAGES 2k Consortium, 2017). The reconstruction target area are all grid cells marked with wire frames.

of BARCAST (c.f. sec. 3), an analysis revealed that this version of the dataset shows ~~rather strange~~ long distance correlations

130 over our region of interest ~~. These~~ that might be deemed unphysical: the correlation length is in 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) dataset were downloaded from the CRU website http://browse.ceda.ac.uk/browse/badc/cru/data/cru_ts/cru_ts_3.24.01.

135 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)~~ 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). ~~In contrast to other methods, no data was shared across grid cells by a prescribed spatial covariance structure or spatial interpolation algorithm.~~ We aimed at retaining the variability that a single instrumental record in the grid cell would exhibit. This is a compromise between an

140 actual grid-cell wide average and the limited spatio-temporal availability of instrumental data in high latitudes. ~~Additionally while~~ In contrast to other regridding methods, no data was shared across grid cells by a prescribed spatial covariance structure or spatial interpolation algorithm. While some proxies (e.g. the tree-ring maximum latewood density record from Northern Scandinavia by Esper et al., 2012) represent larger, grid-cell sized regions rather than point estimates (see the discussion

in Luterbacher et al., 2016), none are large-scale regional averages ~~(such as the Central European documentary record of Dobrovolný et al. (2010) used by Luterbacher et al. (2016))~~(such as the Central European documentary reconstruction of Dobrovolný et al.. In contrast to (Tingley and Huybers, 2010a, b; Werner et al., 2013) the data was not normalised to have unit standard deviation prior to running the BARCAST sampler.

As can be seen in Figure 1, the resulting instrumental dataset is very sparse in space and time. While ordinary reconstruction methods would indeed struggle with such ~~an input grid~~input data, the advantage of the ~~(extended) BARCAST method~~ BARCAST method and the extension used here is that presence and absence of observations ~~is~~ are explicitly modelled. The reconstruction target region are land mass (continent and islands) containing grid cells only (wire frames in ~~Figure~~Fig. 1). 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 ~~Figure~~Fig. 1) mostly come from the current version (2.0.0~~of the (PAGES 2k Consortium, 2017)~~) of the PAGES 2k Consortium (2017) temperature data base, with 6 recently ~~published~~ 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 is sensitive to different seasons, and on different time-scales – 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 ~~was~~ 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 (c.f. table A1), as they did not meet the required response characteristics on actual annual time-scales. Additionally, data was 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 ~~would~~ 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 ~~no~~ very limited replication of ice-cores or varved lake sediments. Thus, these archives can (and usually do) contain age uncertainties (c.f. 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 D1 details how these age models are generated. As the majority of the proxy data is more sensitive to summer or growing season temperatures, the target season for the reconstruction is ~~summer (here:~~ the climatological summer season (the months from June to August, JJA) rather than the annual mean temperature.

3 Reconstruction method: BARCAST+AMS

~~In a recent article,~~ 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 climate field reconstructions over the Arctic (Tingley and Huybers, 2013), Europe (Luterbacher et al., 2016) and Asia (Zhang et al.).

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

$$
\begin{aligned}
\boldsymbol{C}_{t+1} - \mu &= \alpha \left(\boldsymbol{C}_t - \mu\right) + \epsilon_t \\
\epsilon_t &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}) \quad \text{(independent)} \\
\Sigma_{i,j} &= \sigma^2 \exp\left(-\phi |x_i - x_j|\right),
\end{aligned}
\tag{1a}
$$

The process level is thus AR(1) (1$^{\text{st}}$ order auto-regressive) 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:

$$
\begin{aligned}
\boldsymbol{O}_t &= \beta_0 + \beta_1 \cdot \mathbf{H}_t \cdot \boldsymbol{C}_t + \boldsymbol{e}_t \\
\epsilon_t &\sim \mathcal{N}(\boldsymbol{0}, \tau^2 \cdot \mathbf{I}) \quad \text{(independent)}.
\end{aligned}
\tag{1b}
$$

The parameters $(\beta_0, \beta_1, \tau^2, \mathbf{H}_t)$ are assumed to be different for each ~~observation~~ 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 $\beta_0 = 0$ and $\beta_1 = 1$. The selection matrix $\mathbf{H}_t$ is composed of zeros and ones, and selects out 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 into 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:

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

$$\boldsymbol{e}_s \sim \mathcal{N}(\boldsymbol{0}, \tau^2 \cdot \mathbf{I}) \quad \text{(independent)}. \tag{1c}$$

The vector $\boldsymbol{e}_s$ is a time series of independent normal errors at location s (c.f. $\boldsymbol{e}_t$ 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 climate field and all of the parameters (climate field and instrumental / 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; Werner and Tingley, 2015) for details and implementations. In the version implemented here, ~~we modify BARCAST slightly~~BARCAST is slightly modified. While Tingley and Huybers (2013) used a single set of response parameters $(\beta_0, \beta_1, \tau^2)$ for all data of one type, ~~and~~ while Luterbacher et al. (2016) actually set up a separate observation matrix with a set of parameters for each single proxy, ~~we choose to update the code~~ the code is updated for this study. The response parameters 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 know 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 Consort and were thus removed.

The reconstruction code is run in 4 chains for ~~4000~~ 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 is much sparser, noisier and does not follow the exact prescribed stochastic model (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 used infrastructure, 4 chains using 2 cores each (for parallel linear algebra using the OpenBLAS library http://www.openblas.net) were deemed a reasonable compromise.

3.1 Reconstruction Quality

The reconstruction calibration and validation statistics are shown in appendix A. ~~Both the~~ 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 Continuous Ranked Probability Score, which is akin to the

235 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 $\overline{\mathrm{CRPS}}_{\mathrm{pot}}$ as well as the Reliability score show a decent reconstruction quality (Figure A1 top row). The $\overline{\mathrm{CRPS}}_{\mathrm{pot}}$ on average shows a mismatch of ~~0.1~~0.2°C(~~0.3~~0.4°C) in the calibration (validation) interval and the Reliability is mostly better than

240 ~~0.1~~0.2°C. ~~Additionally, a~~ This is in 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 ~~(CE)~~ and the reduction of error ~~(RE), (Cook et al., 1994) are generated. These~~ show a skillful 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

245 debated. In fact, in contrast to other BARCAST based reconstructions the one presented here shows a substantial (~~$\tau_I^2 \approx 0.15$ in standardised units~~$\tau_I^2 \approx 0.25$°C^2) noise level for the instrumental data. As other gridded instrumental datasets employ spatial interpolation processes ~~they~~ these are generally smoother in space than the gridded instrumental dataset generated for this study. Thus these gridded products are by design closer to the spatial ~~part~~ characteristics of the process model in Eq. (1a).

Another means of assessing the reconstruction quality is to check the variability or spread of the different ensemble members

250 in space and time (see appendix B). The effect of the spatially and temporally sparse data can easily be seen in Figures A2 and A3, 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

255 rather, domain over which the reconstruction is analysed) would ~~unnecessarily complicate things~~be beyond the scope of this paper. Thus the gridded reconstruction is only shown back to 750 CE. However, for single analyses over data rich regions the full reconstruction period (1–2002 CE) can in principle be used.

Additionally, the spectral properties of both the reconstruction and the proxy input data are analysed (c.f. appendix C). Not all proxies contain signal on centennial or longer time scales, and the reconstruction method explicitly describes year-to-year

260 summer temperatures as an AR(1) process. Thus, the reconstruction shows properties of an AR(1) process over most of the reconstruction domain (cf. also Nilsen et al.). However, when comparing the area averaged temperature reconstructions against other multiproxy reconstruction data one can still see similar variability on centennial and longer time-scales (see Figure 3).

4 Results

In the following sections the resulting reconstruction is presented. First, the Arctic average is analysed and compared against

265 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 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 that can be associated with the Roman Warm Period (RWP) 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 of the reconstructed temperature field is explored, with a focus on the most extreme periods.

4.1 Mean Arctic Results

Ensemble-averaged spatial linear trends over the period 1–1850 CE in /century; black dots mark locations where the trend is statistically significant for more than 95% ensemble members.

~~The~~ The ensemble mean of the area averaged summer temperature reconstruction is shown in the bottom half of Figure 2 as the ~~point-wise (annual summer~~pointwise (year-to-year summers) ensemble mean (heavy blue line). The first millennium CE shows a mean reconstruction that ~~is relatively flat, with~~ 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 time-scales. 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 sub-regions (Figure 3). Throughout most of the reconstruction period, the Arctic SAT anomaly shows an overall orbital forcing-driven cooling trend. ~~Figure 6 demonstrates that the trend is however spatially inhomogeneous. The largest magnitude of the millennial-scale cooling of $-0.5 \pm 0.1°C$ over the period 1-1850 CE is registered in the comparatively proxy data-rich regions between 0–100 E and 110–170 W. The Greenland region and the Canadian Arctic, in contrast, show a warming with a mean temperature rise of $+0.5 \pm 0.1°C$ before 1850 CE. We note that the cooling trend reversal occurs already in the first half of the 19th century as a recovery after the last LIA minimum. The attribution of the actual timing depends on the considered time-scale and the use of 1850 CE as an upper limit for the trend magnitude estimate is therefore somewhat arbitrary.~~

~~The proxy dataset from Greenland is dominated by oxygen isotopes series from ice cores. These are not entirely coherent in the magnitude and sign of the millennial trend, and in some cases reflect more annual rather than summer season SAT. As a result the detected warming could stem from the different weights of the proxy records in the reconstruction procedure. The oxygen isotope series are also known to be subject to a possible warm bias due to increased storm activity which might accompany the LIA (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) that could actually mask the LIA cooling. Meanwhile, the north of Greenland is less influenced by precipitation associated with North Atlantic storms, though lasting seasonal accumulation changes could be another source of uncertainties important especially on longer time-scales: ice cores from northern Greenland are 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. The overall LIA cooling might have amplified this effect, thus promoting a higher percentage of istotopically warmer summer moisture in the annual accumulation, this way diminishing or even reversing the cooling trend. While site and source temperature compensating effects for the~~
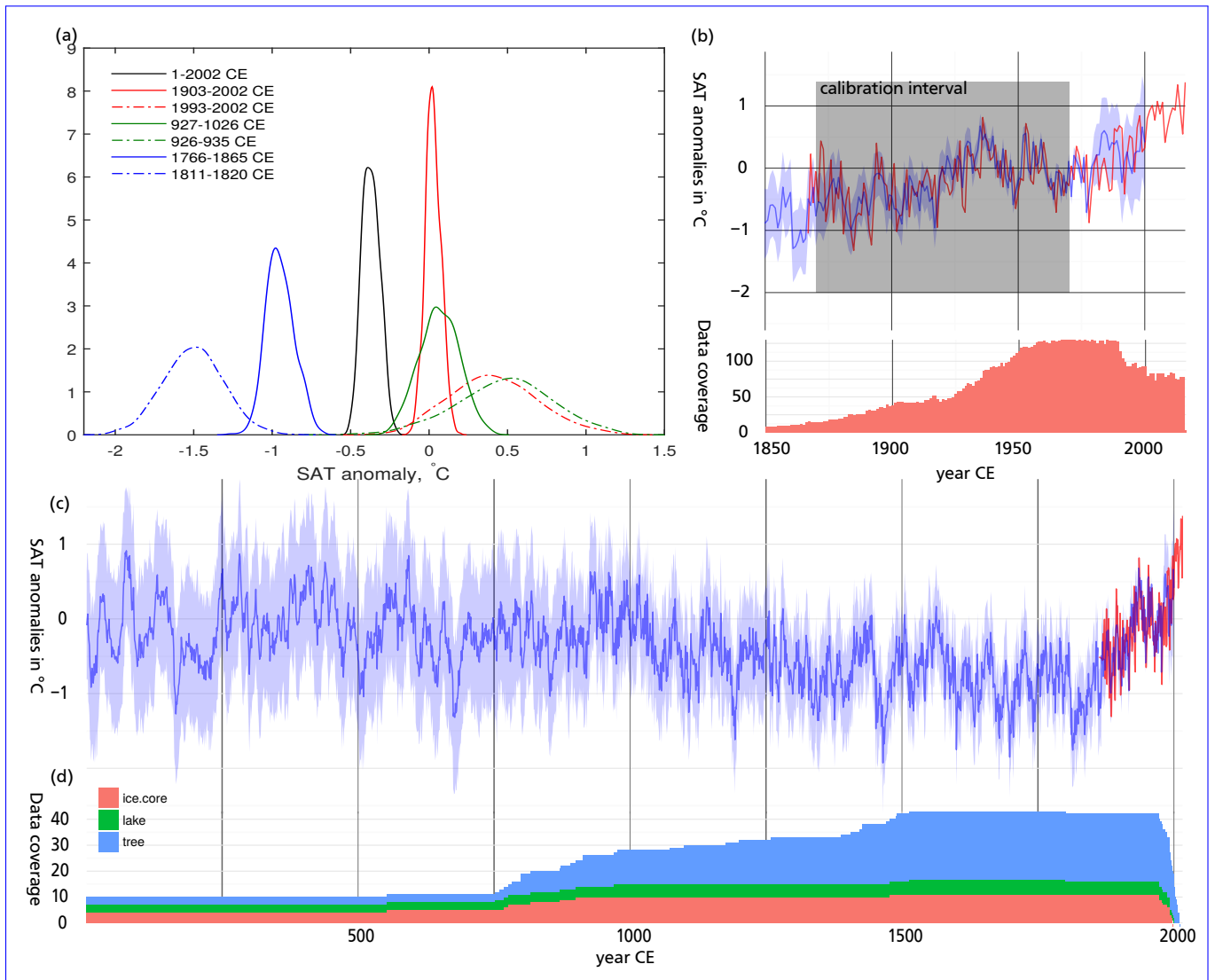
Figure 2. a) Ensemble-based mean Arctic summer (June–August) SAT (land only) anomaly probability density for the three selected centennial-scale and three decadal-scale time periods. ~~For comparison~~ The periods presented correspond to the ~~mean Arctic SAT anomaly for~~ coldest and warmest century- and decade-long periods of the ~~entire reconstruction period is also presented~~LIA and MCA together with CWP. ~~Note that~~ The actual probability densities are estimated using the ~~mean for the last decade (2006–2015 CE) is based on the infilled instrumental data only~~Gaussian kernel density function. b) Arctic (land only) average summer temperature anomalies over the instrumental period and number of instrumental observations available. Grey box denotes the calibration interval. c) Ensemble-based spatially averaged time variability of the seasonal SAT probability distribution over the reconstruction period~~, blue~~. Blue line: ensemble mean, shading: (point-wise) 95% posterior~~. Red~~, 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.
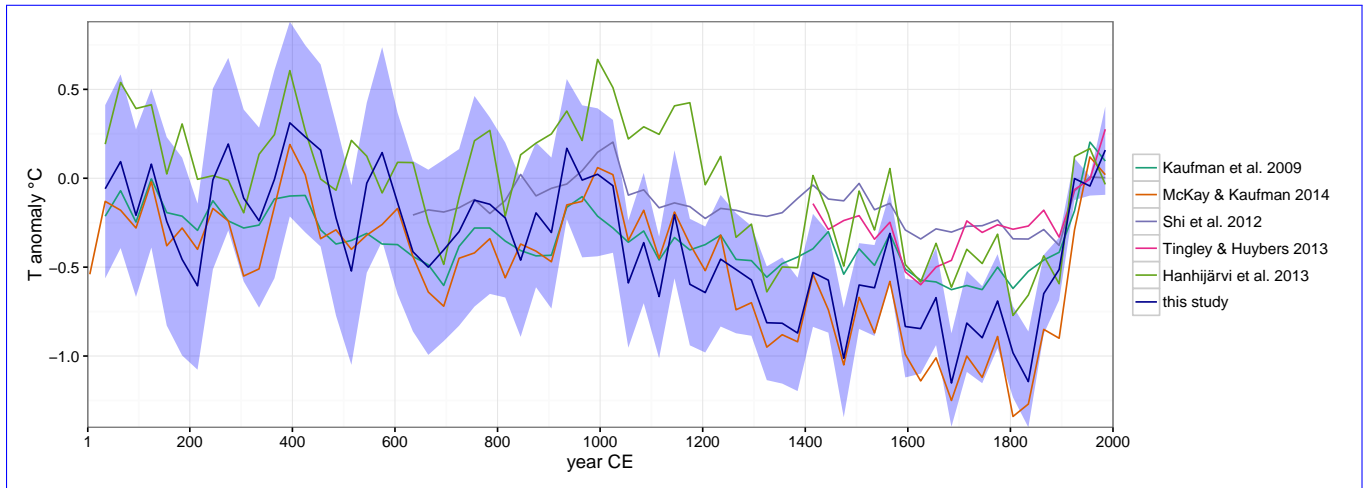
Figure 3. Comparison of this reconstruction (with 95% confidence band) with other reconstructions. 30 year averages. Note that McKay and Kaufman (2014) target annual temperatures and Hanhijärvi et al. (2013) reconstruct the North Atlantic sector of the Arctic (50° ~~W—30~~W–30° E).
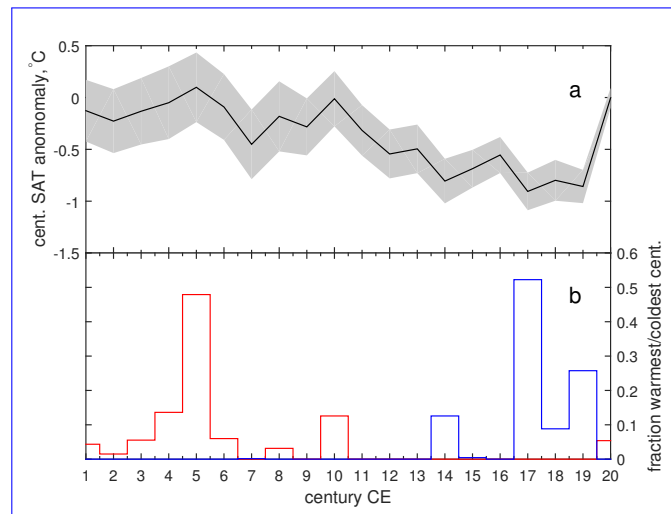
Figure 4. (a) Ensemble-averaged century and spatial mean Arctic SAT (solid black) with the respective ±2 standard deviations highlighted gray; (b) Mean Arctic SAT ensemble-based fraction warmest (red) and coldest (blue) centuries.

~~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.~~

~~Superimposed on the trend are the two major centennial to~~ Superimposed on the trend are three major centennial to
305   multi-centennial ~~scale anomalies: the~~ scale anomalies: a warm period in the 4th and 5th centuries CE broadly associated with a relatively late phase of the Roman Warm Period (RWP), the Medieval Climate Anomaly – a warm period during

13

the 10th and 11th centuries CE, and the two phases of the cold Little Ice Age between ca. ~~1100–1500~~1100–1450 CE and ~~1550–1900~~1600–1900 CE. ~~Our reconstruction suggests the coldest centennial-scale period of the LIA was during the 15th and 17~~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 ~~. The coldest decadal-scale event in the reconstruction, however, occurred in the early 19th during the Dalton Minimum of solar activity and a period of increased volcanic activity with two major tropical eruptions of 1808/1809~~with (within the uncertainty) similar ranged mean SAT anomalies of $-0.9 \pm 0.1^\circ$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 5th century with a SAT anomaly of $0.1 \pm 0.2^\circ$C appears warmest in 48% 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 and Tambora 1815.~~CE period. For the later period with better proxy coverage, the 10th century CE, accommodating the MCA with the ensemble average mean SAT anomaly of $0.0 \pm 0.1^\circ$C, along with the 20th century (SAT anomaly of $0.0 \pm 0.05^\circ$C) share the rank of the two warmest centuries 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 ~~an abrupt~~ the contemporary warming which is clearly identifiable since the ~~beginning of the 20~~middle of the 19th century.

Figure 3 suggests that the LIA cooling is less pronounced in the new reconstruction compared with the same period ~~in McKay and Kaufman (2014)~~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)~~, with the correspondent~~. 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).

We further explore the transient features in the spatial mean reconstruction ensemble using 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 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 summarizes the data analysis results in a map which highlights the locations in "scale" (*here:* the variability time scale) 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
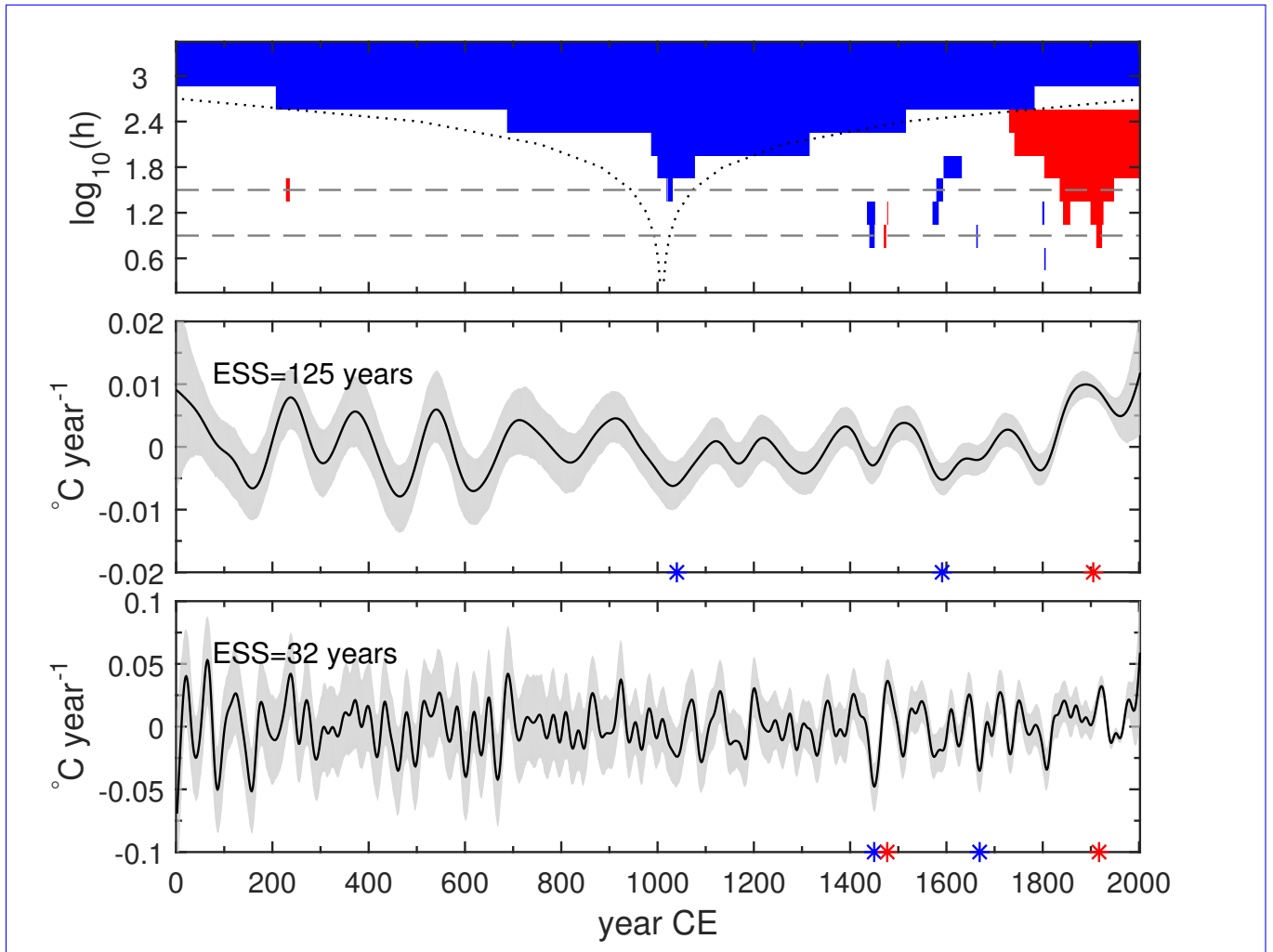
Figure 5. (a) Modified SiZer map of the spatial mean reconstructed SAT with colors red/blue marking the locations in scale (variability time scale) and space (time in this particular case) where the fraction of ensemble members exhibiting statistically significant warming/cooling exceeds 90%; the parallel distance between the dotted lines indicate the effective size of the smoothing kernel used for a particular scale and hence gives an idea of the corresponding time scale 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 gray) 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/red asterisks in (b,c) mark the timings of the maxima in the rates of cooling/warming discussed in the text.

reconstruction ensemble, both the variability of 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.

345   Figure 5a presents results of the analysis highlighting the time scales and periods where 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 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 multidecadal scale. In particular, the
350 initial phase of the LIA-related cooling is centered already at ca. 1030 CE and flagged significant at a range of timescales from centennial to nearly millennial. We note also that for the centennial timescale, Figure 5a points to the onset of a statistically significant warming during the 1840s CE. This would justify using 1850 CE as the cutoff 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.

355   The statistically significant changes that are coherent across the reconstruction ensemble are four cooling and one warming episodes revealed at the timescales of 30-100 years and centered 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 selected 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
360 three-decadal and centennial scale variations. Figure 5b,c shows the associated rates of changes as the ensemble mean together with the respective 95% 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 three decades are registered at 1450 CE and 1669 CE, respectively, while a recovery after the first cooling centered 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
365 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 centered at 1040 CE, the cooling towards one of the LIA SAT minima at 1577 CE with $-0.7 \pm 0.2°$C, followed by the transition to the CWP centered 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.
370 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 three 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 show no single episode that is localised in time across all ensemble members. The latter is related to a much reduced density
375 of the multiproxy 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 in the order of 1500 km, this clearly highlights the need for proxy data to be extended back in time.

The spatial pattern of millennial scale trends in the reconstructed Arctic SAT is further explored in Figure 6. The period of 380 1–1850 CE is considered as a reference period to enable comparison with the earlier studies. 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 Figure 6b). While for the ensemble- and the reconstruction domain average the rate of cooling attains $-0.05 \pm 0.01°$C/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 385 largest magnitude of the millennial-scale cooling of up to $-0.13 \pm 0.02°$C/century yielding an up to a $-2.4°$C temperature decrease 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 Common Era. 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 30–120° W. At the same time, much less pronounced negative trends with an overall cooling of 390 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 gridded reconstruction is limited to the time after 750 CE, the results above need to be interpreted carefully, 395 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.

4.2 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-400 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 around 1000the period of 900-1050 CE, typically associated with the peak of the MCA, show shows up at least similarly warm as the reconstruction for the late 20th and early 21st century, 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). At the 405 same time, 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 other reconstructions (notably Hanhijärvi et al., 2013) the the reconstruction by Hanhijärvi et al. (2013) the Roman times around the first and second century CE do not show up as particularly warm in the circum-Arctic mean, which is also reflected in the analyses presented in section 4.3the previous section. Note that their reconstruction was actually limited to the North Atlantic sector of the Arctic, and thus a direct comparison 410 is difficult. Additionally the spatial skill of the reconstruction decreases back in time as the proxy data becomes sparser (see
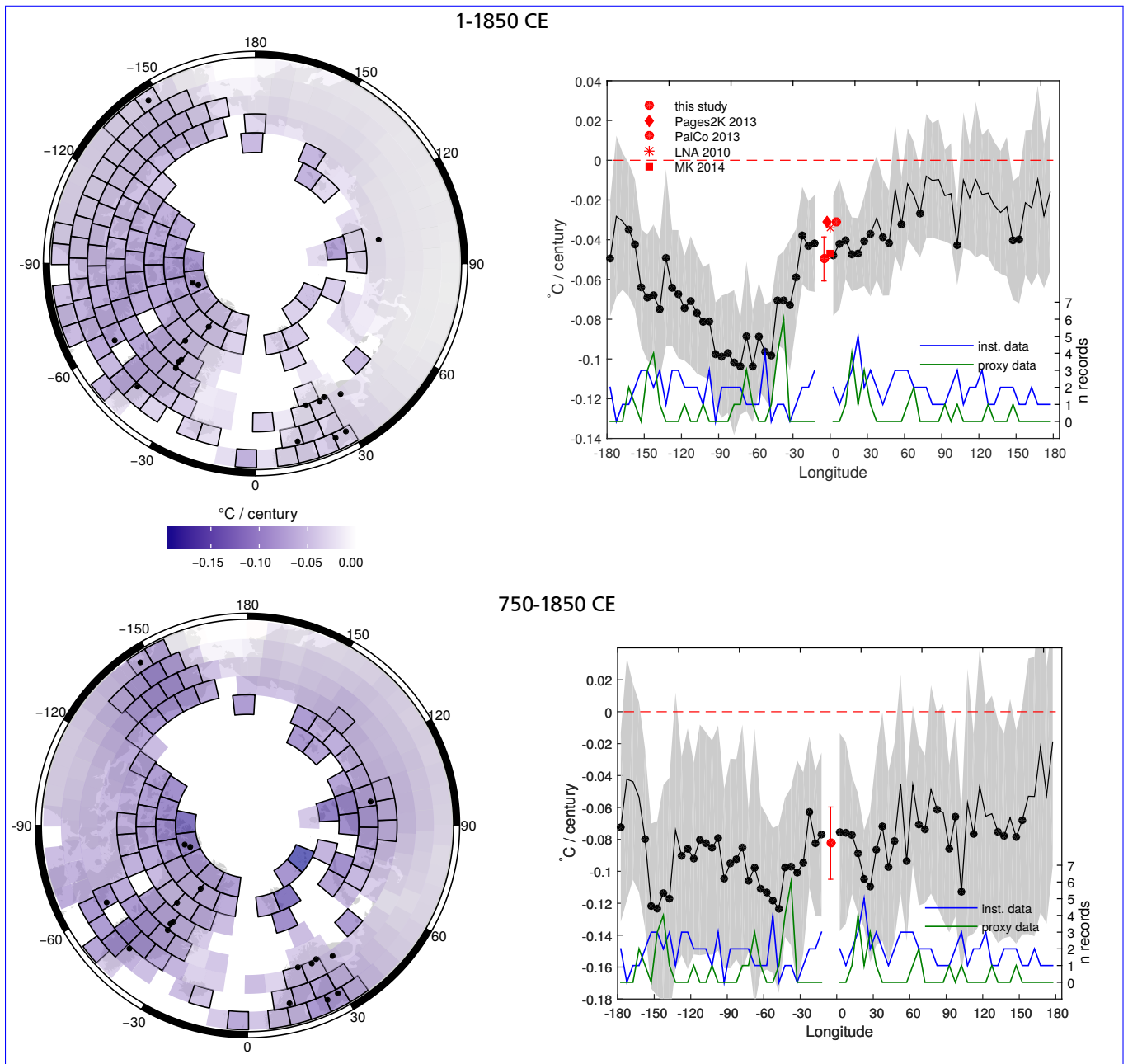
Figure 6. (a) Ensemble-averaged spatial linear trends over the period 1–1850 CE in °C/century; black dots mark proxy locations, quadrilaterals mark locations where the trend is statistically significant for more than 95% ensemble members. (b) Ensemble-averaged meridional trends in the latitude-averaged reconstruction over the period 1–1850 CE (solid black line); meridional averaging over the 5° segments and zonal over the terrestrial nodes is applied to each reconstruction ensemble member. Grey shading highlights the ±2*std 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 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.

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 attaines 0.48 ± 0.31°C. The timing of the coldest centennial-scale period of the LIA, specifically the 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 ± 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 1815 CE. The second coldest decade of the LIA with the SAT anomaly of -1.4 ± 0.2°C has likely occurred during 1463–1472 CE, also following strong volcanic forcing.

Figure 2a presents the ensemble-based probability densities (pdf) of the spatially averaged across the reconstruction domain mean SAT anomalies for the six selected reconstruction sub-periods. These are the three selected century-long periods of ~~960–1060~~927–1026 CE, ~~1580–1680~~1766–1865 CE and 1903–2002 CE, representing both the aforementioned warmest and the ~~two warmest and~~ coldest century-long periods of the record ~~and~~ after 750 CE as well as the last century-long period, the second warmest of the reconstruction~~that includes~~, which includes the Contemporary Warm Period (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 ~~four~~ three shorter decadal-scale intervals corresponding to the anomalously warm and cold periods ~~(see~~ of 926-935 CE, 1811–1820 CE and 1993–2002 CE (see also Subsection 4.3 for details) are displayed~~, including the most recent~~. 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 ± 0.28°C, followed in rank by a warm decade of ~~2006–2015~~the early 20th century warming 1930–1939 CE~~. For the latter we used an instrumental grid infilled by BARCAST.~~, not shown here). The maps of spatial mean SAT anomalies for these periods follow in Figure 7.

~~The~~ Comparing the coldest phase of the LIA with a mean centennial-scale SAT anomaly of ~~-0.5 ± 0.1~~-0.94 ± 0.09°C vs. ~~MCA 0.1 ± 0.1~~the MCA 0.07 ± 0.13°C and ~~CWP 0.03 ± 0.05~~the last century of the reconstruction (SAT anomaly: 0.01 ± 0.05°C~~emphasise a~~) emphasises 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 ~~CWP~~period 1903–2002 CE, although a reduction of proxy data after the 1990s likely ~~introduce~~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 is then collected and analysed.

The testing results for a two-tailed test with unequal variances rejected H_0 of equal Arctic mean SAT anomalies between ~~960–1060~~927–1026 CE and 1903–2002 CE for ~~98~~93% ensemble members. ~~No~~ However, when considering hypotheses with a one-tailed test no conclusive answer can ~~however be reached~~for the sign of the bias in the means between the centennial periods~~be reached. Although the MCA appears slightly warmer on a centennial time-scale compared with the last 100 years as shown in ~~2a), testing fails to reject the null-hypothesis for 7~~Figure 2a, testing rejects H_0 for 64% of the ensemble members only~~, whereas for the opposite alternative hypothesis (i.e. ~~CWP~~ 1903-2002 CE warmer than MCA on average) the ~~hypothesis~~ H_0 rejection rate is as high as ~~6%. Taking the multiple testing into account, we~~ 29%.

Though this result somewhat favors the alternative hypothesis of $\overline{SAT}_{MCA} > \overline{SAT}_{CWP}$ the difference in the rejection rates appears negligible. We conclude therefore 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 note also that higher variability in the derived ensemble of realisations for the mean Arctic SAT anomaly during the warmest decade-long intervals of the MCA ~~of $0.5 \pm 0.2$and CWP of 0.2 ± 0.3~~and CWP similarly prevents from reaching any firm inference on the relative magnitudes of the two decade-long anomalously warm periods of the new reconstruction.

~~The scale of the warming north of 60that occurred over the last decade (2006–2015 CE) is hinted at by the pdf of the (infilled, gridded) temperature data. It suggests a positive SAT anomaly of 1.0 ± 0.05, which appears significantly warmer than the reconstructed decadal temperature maximum of the MCA. However, first of all the comparison of the two independently infilled processes is not advisable. The underlying stationarity assumptions about systematic biases from the infilling are violated in this case, and additionally there is a tendency to underestimate extremes in the past due to sparse data, but also due to the assumed AR(1) process model. Thus, while hinting that the last decade could have been unprecedented over the last 1100 years, the exact amplitude of this current warming in the instrumental data with respect to that in the maximum of the MCA remains more uncertain than the pdfs would suggest at face value.~~

4.3 Spatial signature of past and recent extreme temperature anomalies

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 Figure A2) in the early part of the reconstruction period, the analysis is limited to the time after 750 CE. ~~The~~ For the Arctic average the probability density for each year to be ranked as warmest or coldest is calculated across the entire ensemble ~~for each individual location. Further marginalisation over the spatial domain yields a probability density of a particular year to be associated with an Arctic-wide warm or cold extreme. In order to~~ using
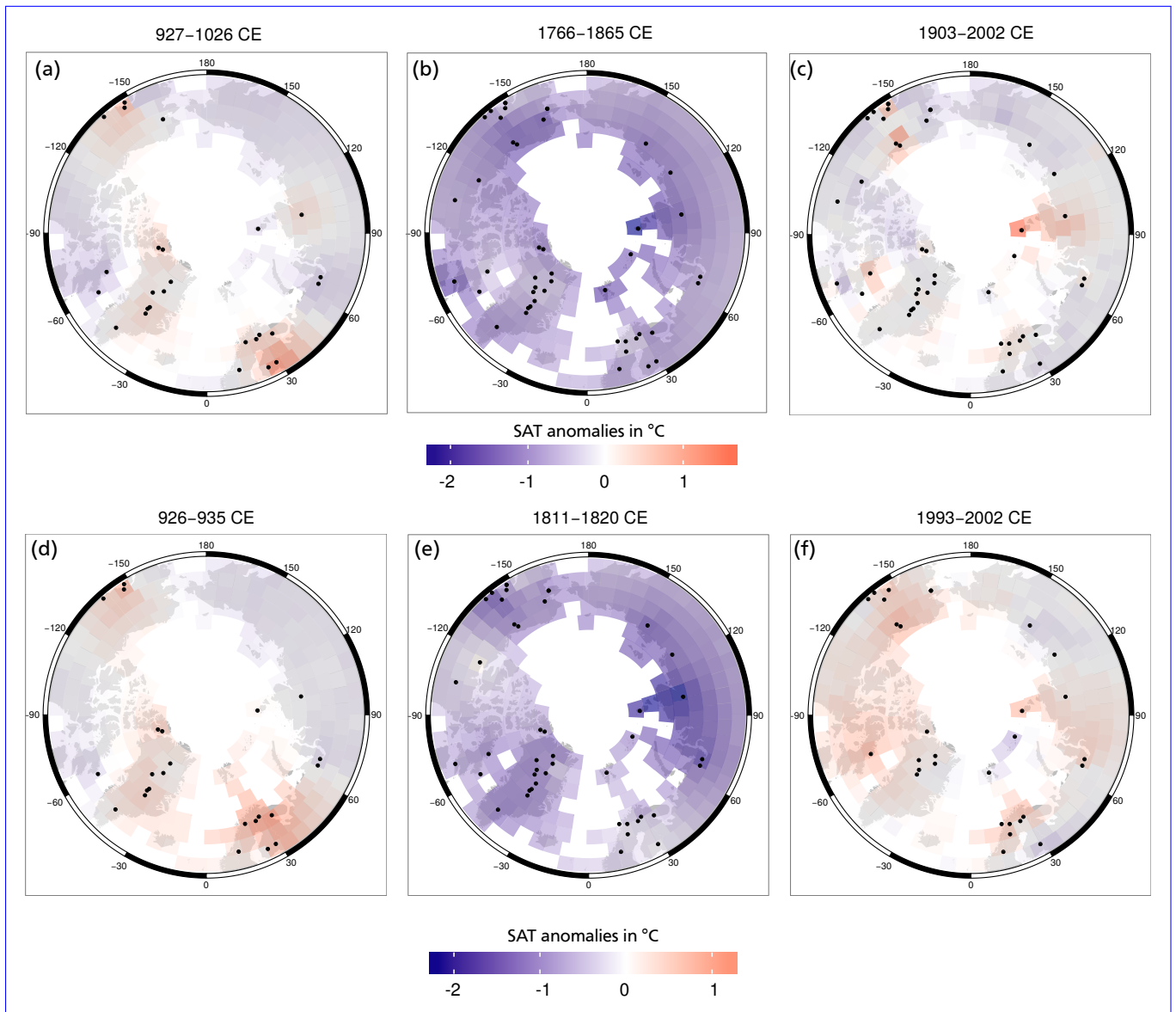
Figure 7. Ensemble average of the reconstructed Arctic2k SAT anomalies over the century-long periods of (a) ~~960–1060~~927–1026 CE, (b) ~~1580–1680~~1766–1865 CE and (c) 1903–2002 CE; (d) Ensemble average SAT anomaly over the period of ~~982–991~~926–935 CE, a potentially warmest decade since 750 CE with ~~nine~~ seven of ten years top ranked as potentially warmest; (e) Ensemble average SAT over the period of ~~1812–1821~~1811–1820 CE~~within six~~, a potentially coldest decade since 750 CE with seven out of ten years ~~are~~ ranked potentially coldest; (f) Ensemble average SAT anomaly over the period of 1993–2002 CE, a potentially warmest decade after the MCA with ~~12~~5 out of ~~16~~10 years ranked as potentially warmest. Colours show the temperature anomalies. Proxies marked by black dots.

21
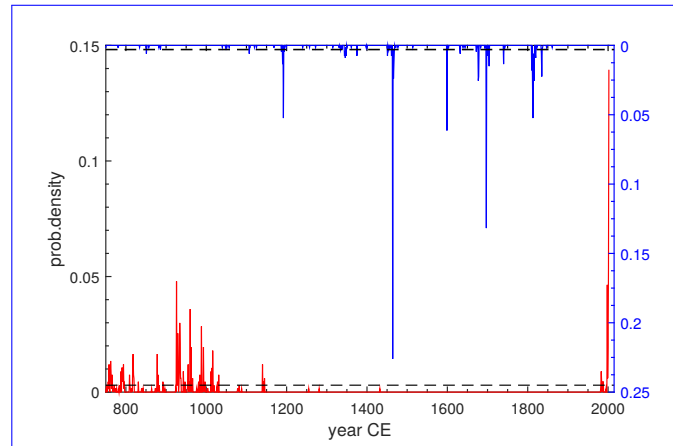
Figure 8. Probability density of statistically significant, potentially warmest (red) and coldest (blue) years over the ~~2000-year long period~~ ~~marginalised over~~ the ~~reconstruction domain~~period 750-2002 CE. Dashed lines indicate the 95% significance level estimated from the ensemble of bootstrap surrogates.

the spatial mean SAT. To check the statistical significance of the derived probability densities, the analysis is replicated on
480   an ensemble of surrogates derived from the original reconstruction ensemble using ~~bootstrapping~~ block bootstrapping of the spatial mean reconstructions along the time axis. The ~~surrogates were constructed in a way to mimic the statistical properties of the original reconstruction, such as a local serial correlation, variance and a spatial coherence.~~ 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
485   time-average ~~[0.025 0.975] percentiles of the spatial probability densities~~ 0.975 percentile of the probability density for the bootstrap surrogates ~~are~~ is then used as the respective ~~quantiles~~ quantile for marking the years as potentially coldest or warmest during the analysis period (Figure 8). ~~For convenience, the probability densities are further marginalised over the 5 degree longitude bins~~and In order to highlight a decadal scale variability in occurrence of warm and cold extremes, the fractions of potentially warmest/coldest years per decade are calculated in sliding 10-year long windows. In order to reduce the effects of
490   the reconstruction uncertainties, the reconstruction is averaged over five degree 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 ensemble members. The analysis results are presented as a time–longitude colour map in Figure 9a. ~~Additionally, potentially warmest and coldest decades using the spatially marginalised probability density functions are evaluated (Figure 9b). The significance testing procedure for the decadal extremes is based on a similar analysis.~~
495   The results of the analysis 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 ~~the same~~ a pronounced pan-Arctic warming with a consecutive series of spatially coherent warm extremes between ca. ~~935–1035~~920–970 CE (Figure
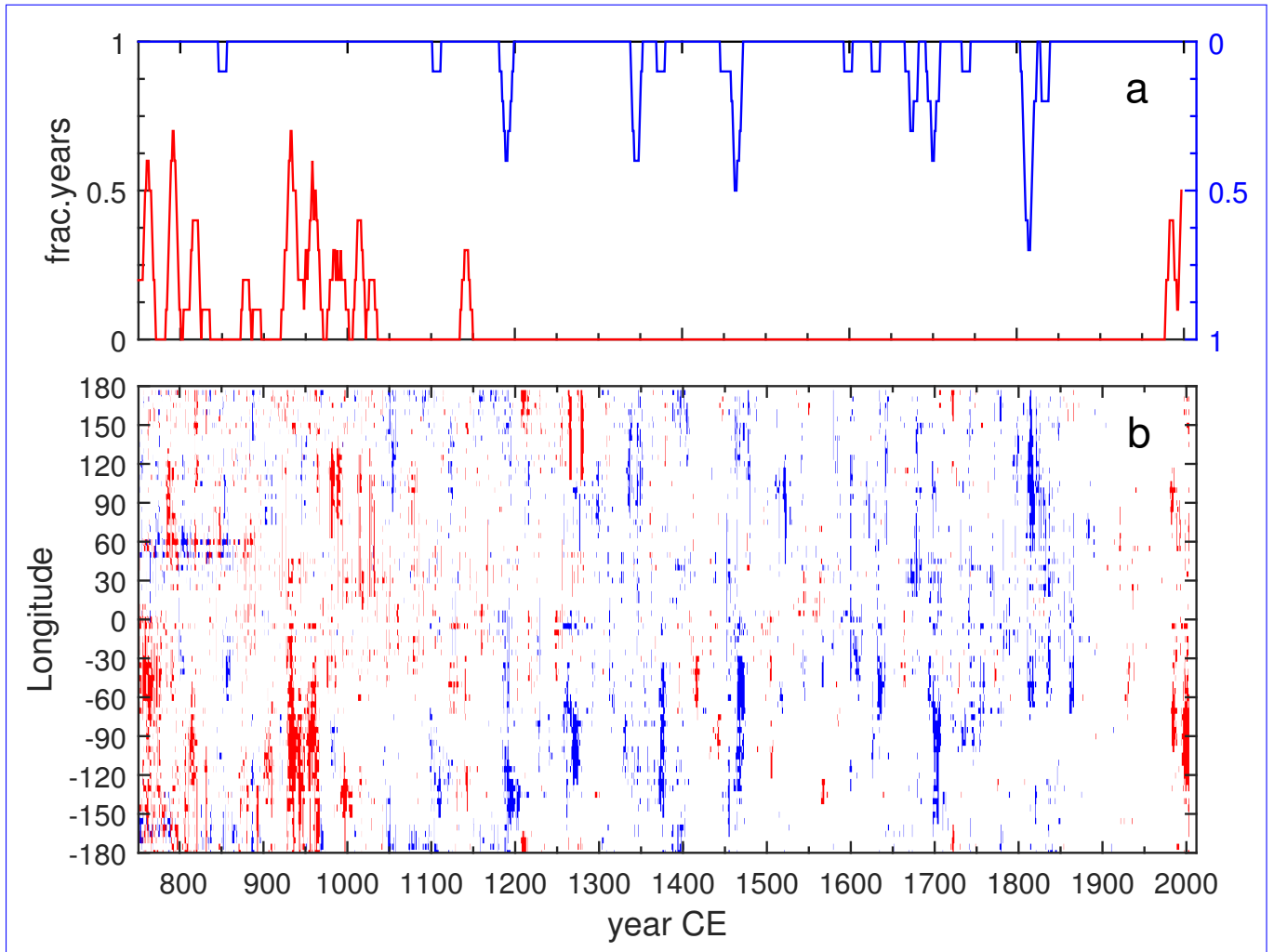
Figure 9. ~~Left:~~ (a) Fraction of potentially warmest/coldest years per decade with respect to time. (b) Occurrence of statistically significant, potentially warmest (red) and coldest (blue) years over the ~~2000-year long~~ period 750-2002 CE in different sectors of the Arctic domain. ~~Right: Fraction of potentially warmest/coldest decades with respect to time. Both MCA and LIA clearly shows up.~~

23

9a~~b~~). On decadal time-scale (Figure 9~~b~~a) the MCA is marked over the whole region by anomalies having a persistent high fraction of likely warmest decades with ~~a negligibly low fraction of coldest decades.~~ no decades containing a year ranked as coldest.

In particular, ~~nineteen of twenty~~ seven out of ten years during the ~~two decades of 981–1000~~decade of 926–935 CE ~~and all~~ making the first MCA sequence of warm extremes and six out of ten years of ~~the decade 1026–1035~~954–965 CE making the second maximum were ranked as statistically significant warm extremes among the ensemble members~~, which makes these two periods potentially warmest in the reconstruction~~. This agrees with a previously found reconstruction decadal scale temperature maximum since at least 750 CE ~~.~~ for the first of the two periods. Note that sequences of potentially warmest years and hence decades with a higher fraction of extremes are also detected before 800 CE, though the 8th century does not appear in the reconstruction as particularly warm on average.

Figure 9~~a~~ b also highlights a difference in time evolution of the regional expression of the MCA via the spatial incoherence of extremely warm years/decades ~~during the period associated with this climate anomaly~~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 ~~between 45~~east of zero meridian around 920~~E–160 E~~CE, followed by a coherent warming in the Greenland and North Atlantic (NA) sector of the study domain. Figure 9b also suggests that a second phase of the MCA could mainly be localized west of the prime meridian. Figure 7d exemplifies a picture of a pan-Arctic warming during the first warm decade ~~982–991~~926–935 CE of the first phase of the MCA with the largest reconstructed positive anomalies attained within the ~~80–130~~170° E domain. ~~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 980-1040 CE localized primarily within the Atlantic sector (Greenland/Europe) of the study domain and do not exhibit as clear temporal coherence as the two main phases of the MCA.

~~Figure 9b demonstrates~~ Figures 8 and A2 and demonstrate that the period after the MCA termination features a variable climate as manifested by an alternating sequence of potentially warmest and coldest ~~decades lasting until approximately 1560~~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 centered at 1142~~CE. Despite that this period is generally associated with the colder LIA, remarkable warm decadal-scale periods are detected between 1273-1289 CE, 1375-1393~~CE~~and 1528-1540 CE~~. The following transition into the ~~coldest phase of the LIA after 1560 CE~~cold LIA is clearly marked by a drop in the frequency of potentially warmest years/decades to zero~~, whereas the cold decades dominate~~. During the LIA the cold extremes dominate on both the regional (Figure 9b) and pan-Arctic scales (Figure A2) until the onset of the contemporary warming after ca.~~1980. Figure 7e shows the spatial ensemble average SAT anomaly for the potentially coldest decade of the LIA of 1812–1821~~1880 CE. The first potentially warmest year after the termination of the MCA is detected only in 1983 CE.

One can discern five major clusters of cold extremes during the LIA with one of the years of 1192, 1464, 1599, 1697 and 1813 CE ~~, with all ten years ranked potentially coldest across the entire reconstruction ensemble~~. ~~Note that Greenland and Baffin Island and the surrounding areas during this decade remain relatively warm in contrast to the rest~~ 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 ~~domain. This local anomaly agrees well with the abrupt minimum~~

of spring sea-ice extent on South-East Greenland shelf during the early 19th century, reconstructed from the high-resolution sediment core MD99-2322 Miettinen et al. (2015). We suggest this can be attributed to the effect of a strong negative North Atlantic Oscillation (NAO) anomaly during this period , possible in combination with a change in the Subpolar gyre circulation (e.g. Schleussner et al., 2015; Otterå et al., 2010) triggered by a sequence of volcanic eruption in the early 1800s including the unknown 1809 event and the major Tambora 1815 eruption is represented by a sequence of spatially coherent potentially coldest years within 65° W–180° E (Figure 9b). Figure 7e shows the spatial pattern of cooling for this decade of the LIA with seven out of ten 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 1990 1980 CE within 45–100° E and 60–110° W and since 1995 propagating to almost the entire reconstruction domain between the longitudes of 180 W to 100 E. Figure 7f shows a spatial map of temperature anomalies for the period 1993–2002 CE, that features 6 5 out 10 statistically significant warm extremes on the pan-Arctic scale. When compared with the probability density marginalised over the spatial domain displayed in Figure 8, contemporary warming clearly reveals a similar coherence both in the spatial domain and agreement over the range of ensemble members , compared with that is at least as strong as the estimates made for the MCA. One should emphasis 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 three 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 Arctic2k Arctic reconstruction. This reconstruction, however, does not extend into the very last 15 years, over which warming in teh the Arctic has been continuing (c. f. Figure 2). . With these years included in the analysis, the signature of the CWP would much likely become more prominent (see discussion in Section 4.1).

4.4 Arctic amplification in the European sector of the Arctic during the MCA and contemporary warming

Centennial-scale signature of Arctic amplification during the MCA and contemporary warming demonstrated via comparison of temperature differences between the cold 17th century and warmer periods of 10th–12th and 20th centuries across the latitudes. The probability densities are calculated as ensembles of centennial mean anomalies over the four latitudinal bands of 30–45 N (green), 45–60 N (blue), 60–70 N (black) and >60 N (red). For the first three bands averaged over 0–50 E lower latitude PAGES2k reconstruction over Europe (Luterbacher et al., 2016). Note that our Arctic2k reconstruction uses different calibration interval what hinders a direct comparison of the respective SAT anomalies distributions derived from other reconstructions.

Time evolution of ensemble averaged centennial mean SAT anomaly over the period of 700–1300 CE for the four latitudinal bands in the European sector 0–50 E of the Northern hemisphere. Latitude average SAT anomaly for the three bands south of 60 N were calculated using the PAGES2k reconstructions over Europe (Luterbacher et al., 2016), while our Arctic2k reconstruction is used for the Arctic higher latitudes (>60 N). Error bars highlight the ensemble-based CIs for the timings corresponding to the warmest year of the MCA at the respective range of latitudes. The anomalies were calculated relative to the respective centennial means for the coldest centennial-long period of the LIA, 1580–1680 CE.

The concept of "polar amplification" (of which Arctic amplification is the Northern Hemisphere expression) predicts a larger magnitude of warming in the polar regions during lasting warm periods than the one observed on a global or hemispheric scale (Hind et al., 2016). As the proxy data is sparse in space and time, we here focus on the regional expression of the Arctic amplification by analysing the previous warm anomaly, the MCA, over the European sector. While there are gridded temperature reconstructions over North America (Wahl and Smerdon, 2012) and Asia (Zhang et al.), the North American one goes back only to 1200 CE, and there is only a low number of proxy records in the Asian sector of the Arctic.

The analysis is based on the EuroMed2k reconstruction, the ensemble–based gridded summer temperature reconstruction by Luterbacher et al. (2016), who use a similar reconstruction method. While both gridded reconstructions (the one presented here and that by are skilful on the grid-cell level only back to around 750, averages over latitudinal bands will retain skill prior to that. The reconstructions are averaged over the three latitudinal bands 30–45 N, 45–60 N and 60–70 N within the longitudinal sector of 0–50 E. Then, the ensembles of centennial anomalies both for specific periods and in a sliding window in the time domain are calculated, using a coldest period in the Arctic2k reconstruction of 1580–1680 CE (Figure 7b) as a reference for deriving the respective regional centennial changes. In order to test the hypothesis of the two samples to have statistically significant means a paired-sample t-test was used, implicitly assuming the samples are nearly Gaussian distributed. Note that a direct comparison of the Arctic2k results with the EuroMed2k reconstructions are problematic due to different calibration periods used in BARCAST for these reconstructions; the inference is therefore mainly based on the European reconstruction alone. Results for the Arctic2k reconstruction, however, provide an outlook to the intra-ensemble spread on centennial time-scale and the time evolution in the respective centennial estimates relative to the reference period .

Figure **??** shows the spatially averaged probability densities of century mean SAT anomalies for the three centuries overlapping with the MCA. For comparison, similar results are also presented for the contemporary warming using the last possible centennial period of 1903–2002 CE. Although in the European sector the effect of the Arctic amplification is evident for all four century-long periods, the 11th century clearly stands out, featuring a seasonal SAT anomaly relative to the defined baseline in the 17th century of 0.8 ± 0.1 at 60–70 N. For comparison, the observed centennial warming at 30–45 N had a magnitude of 0.2 ± 0.1, yielding a 0.6 ± 0.2 statistically significant warming amplification across the considered range of latitudes.

While the Arctic amplification is apparent already at the peak of the MCA warming (anomaly of 1 at 60–70 N vs. 0.7 at 30–45 N), the analysis of the time evolution of centennial mean SAT anomaly shown in Figure **??** demonstrates the rather dynamical behaviour of Common Era temperature variability. This underlines the necessity to look beyond the usual analysis of equilibrium states (c.f. Hind et al., 2016). While the latitudinally banded temperature maximum was reached almost concurrently, with a hard to quantify lag at high latitudes, the cooling at higher latitudes lags behind the lower latitudes. Thus, there is an apparent warm amplification, which should rather be attributed to a lagged and slower cooling at higher latitudes of the European Arctic. This cooling commenced after the centennial average temperature maximum of the MCA (950–1000 CE). The multi-decadal time-scale of the observed lag points to a particular role of oceanic mechanisms in creating and maintaining the positive temperature anomaly after the LIA temperature maximum. Notably the cooling following the next less pronounced MCA-associated maximum around 1200 CE demonstrates the opposite pattern with the temperatures at the higher north

~~showing a more rapid decline during the 13th than at the lower latitudes, although the magnitudes of warming at higher 0.6 ± 0.1vs lower 0.4 ± 0.1latitudes remain statistically different.~~

~~Figure ??d shows that the modern period features a statistically significant offset of 0.4 ± 0.2between the higher (1.1 ± 0.1) and lower (0.7 ± 0.1) latitudes in the European sector, which is still less than the one registered in the reconstruction~~ 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 similar to the one observed now during the MCA~~maximum. The Arctic2k reconstruction however shows a more pronounced warm bias specifically for the last century (see the respective probability densities on panels (b) and (d) in Figure ??). Whether this is an artifact of time variability in the reconstruction quality or an indication/confirmation of an unprecedented decline in sea-ice extent during the recent decades still needs to be clarified.~~

5 Discussion and conclusions

This paper presented a new circum Arctic CF reconstruction of summer season temperatures back to 750 CE with the Arctic average SAT anomaly extended back to 1 CE. The reconstruction uses a subset of 54 annually ~~dated~~ resolved temperature sensitive terrestrial proxy archives of various types mainly from an updated and corrected PAGES2k database supplemented with 6 new recently published Greenland ice core series.

The technique applied is a recent extension of the Bayesian BARCAST which ~~provides a means to explicitly treat~~ explicitly treats climate archives with dating uncertainties which previously would be used on their "best guess" chronologies. Another added value of using the Bayesian technique was a generation of the ensemble of ~~1400~~ 670 equally likely, independent re-alisations of past CF evolution that enabled us considering the past climate regional variability in a probabilistic framework. Therefore this new Arctic2k reconstruction is essentially an ensemble of possible realisations of the past Arctic summer ~~climate~~ temperatures given the data and the reconstruction technique used.

The quality of the reconstruction in the spatial and temporal domains was tested using a suite of metrics such as continuously ranked probability score ($\overline{\text{CRPS}}_{\text{pot}}$) and the ~~reliability (Reli)~~ Reliability score which are more appropriate for the Bayesian framework than the "Coefficient of Efficiency" and "Reduction of Error", which are typically used in palaeoclimate research. Judging from these scores it could be demonstrated that the new reconstruction is skillful for the majority of the terrestrial nodes in the reconstruction domain, making it a useful product for studying the late Holocene Arctic ~~climate variability at the region scale~~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.

Although this study is mainly focused on presenting the new reconstruction and assessing its quality, we also ran some basic quantitative data analysis in order to uncover the potential of the new product and consider the results in light of previous studies on the subject. In particular, it is demonstrated that the area averaged Arctic2k reconstruction features similar major

cold and warm periods throughout the last two millennia and thus compares favourably with earlier studies targeting a similar season and region.

The major findings from the analysis of the new reconstruction are as follows:

There is a ~~notable lack of~~ pronounced orbital scale cooling trend over the Common Era – a period over which the summer insolation has mostly been decreasing~~. In agreement with findings of Nicolle et al. (2017), our results demonstrate this orbital~~
~~cooling trend is not necessarily observed in all sub-regions, although Nicolle et al. (2017) combine the data from around the North Atlantic realm into a single composite series. The magnitudes for the local cooling does not exceed -0.5 ± 0.1°C over the period 1-1850 CE and this change is found statistically significant only in the comparatively proxy data-rich regions between 0–100~~, although the spatial pattern cannot be reliably reconstructed over the full Common Era due to the sparse proxy network before ca. 750~~ E and 110–170 W. In contrast to this, the reconstruction shows that the Greenland region and~~
~~the Canadian Arctic even exhibit a warming with a mean temperature rise of $+0.5 \pm 0.1$°C before 1850~~ CE. ~~While this can in part be explained by controls on the stable water isotopes in the annual accumulation other than mean ambient temperature ,~~ ~~there is also the possibility for contrasting decadal to century long anomalies in oceanic and atmospheric temperatures between the eastern and western subpolar North Atlantic due to variations in the relative strength of~~ Since the proxy dataset from Greenland is dominated by oxygen isotopes 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 ~~eastern and western branches of the AMOC (e.g. Schulz et al., 2007; Hofer et al., 2011). In particular, Seidenkrantz et al. (2007) report an oceanic warming south of Greenland during the northeastern North Atlantic cooling episodes such as the LIA, indicating an increased advection~~
~~of warm Atlantic water by the West Greenland current. Miettinen et al. (2012) and Thornalley et al. (2009) show continuous warming trends in the central subpolar North Atlantic over the late Holocene driven by lasting oceanic circulation changes. Though the evidence of persistent changes in their strength during late Holocene are still rather inconsistent, this can still be considered as a possible contributing mechanism behind the observed lack of millennial cooling in the reconstruction for the Greenland region.~~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 ~~of the last two millennia, the LIA and MCA~~ back to 750 CE, the MCA and LIA, as well as the beginning of the CWP in the circum-Arctic region.
Although there is evidence for prominent and lasting ~~climate anomalies~~ temperature fluctuations in the pre-750 CE period ~~too~~as well, these results should be interpreted ~~cauthiously~~ cautiously due to the drastic reduction in proxy data density in the early part of the reconstruction period. The MCA in the circum-Arctic region can be associated with a ~~century-long~~ century-scale period between ca. ~~960–1060~~920–1060 CE showing an area average SAT anomaly of 0.1 ± 0.1°C. The MCA features two temperature maxima~~and different~~, both showing similar spatial extent of the regional ~~anomalies that progress~~

670 ~~from the Eurasian to the North American segments~~ SAT anomalies with largest expression in the North American segment of the Arctic realm. A coherent warming of the period ~~982–991~~927–936 CE ~~yields the~~ during the first maximum of the MCA is associated with a potentially warmest decade of the ~~MCA~~ reconstruction with the area average summer temperature anomaly of ~~0.5 ± 0.2~~0.48 ± 0.31°C. While the most recent warming shows an even stronger regional coherence than the MCA, even across continents ~~(PAGES 2k Consortium, 2013)~~(PAGES 2k Consortium, 2013; Ljungqvist et al., 2016), the MCA was still an

675 unusual and extremely warm period in the context of the past two millennia. However, given the input data available and the reconstruction method used it cannot be decided with any statistical significance whether the MCA or the CWP was warmest in the reconstruction.

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 ~~1580–1680~~1766–1865 CE

680 shows an almost spatially coherent circum-Arctic summer cooling~~with a remarkable exception in north Greenland. While the coldest century in the LIA(and in the reconstruction as a whole) was in the 17ᵗʰ century CE, the 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 ~~1812-1821~~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 yielding a circum-Arctic SAT anomaly of $-0.8 \pm$

685 0.2°C. ~~The cooling was, however, not pan-Arctic: warmer than normal conditions were reconstructed for Greenland. From independent proxy data it is known that during that time, the sea ice extent could be lower than average (Miettinen et al., 2015) which, again, could be a signature of the ocean circulation response triggered by negative anomalies in the radiative forcing (Schleussner et al., 2015; Otterå et al., 2010).~~

The ~~comparison of the time-variable behaviour of summer temperature anomalies between high latitudes and lower latitudes~~

690 ~~shows the very dynamic characteristics of the Arctic amplification. In addition to the amplified temperature anomaly in the high north compared with lower latitudes, during MCA this phenomenon also shows up as a clear multi-decadal lag in summer temperature maxima in the highest latitudes when comparing against mid latitude European summer temperature reconstructions.~~
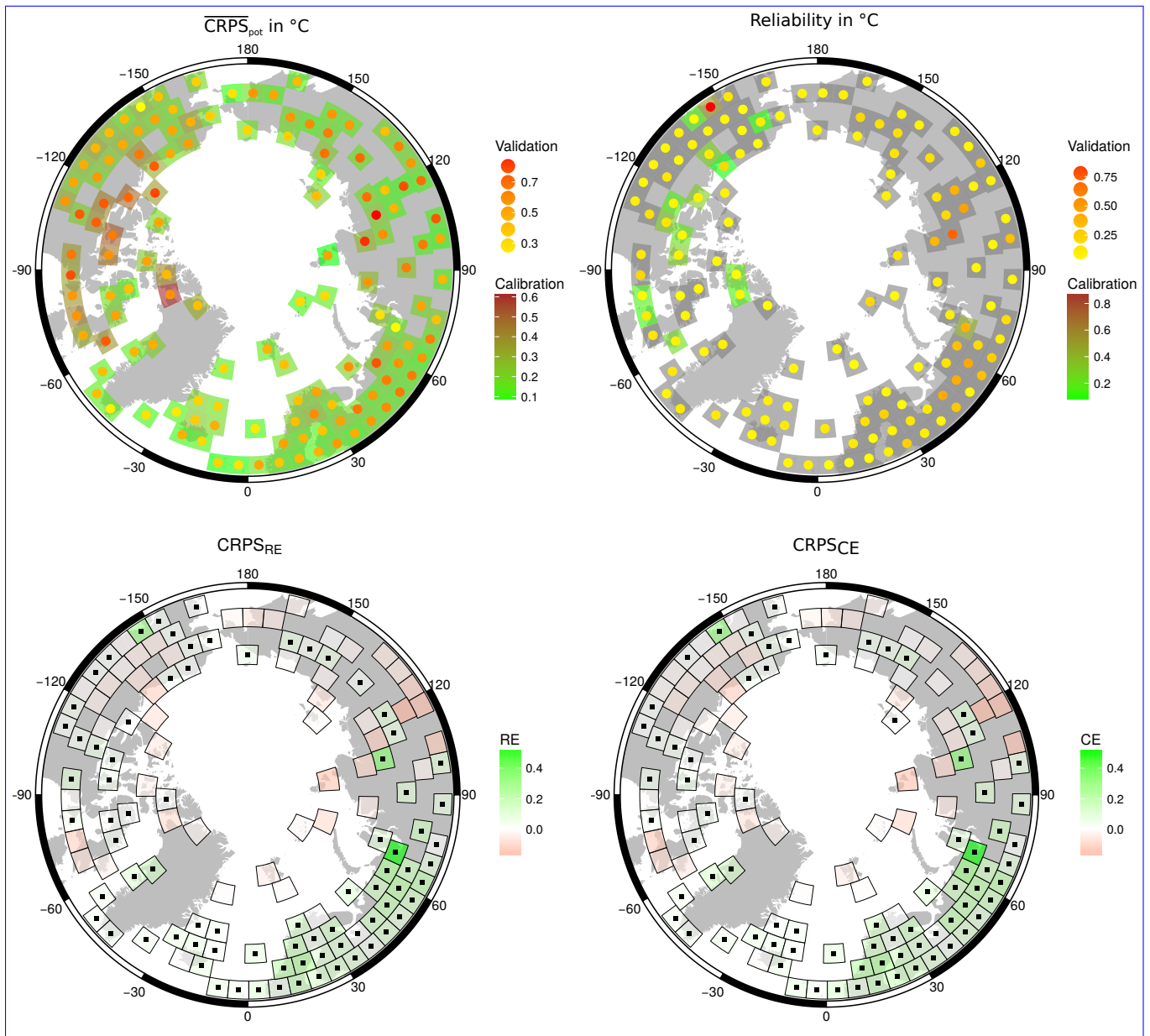
~~The lack of a pronounced orbital cooling trend throughout the Common Era, and the~~ spectral characteristics of both proxies

695 and reconstruction show that ~~still work is~~ work is still needed on generating more and longer ~~high quality~~ high-quality proxy series in parallel with a reanalysis of the existing data. Especially updating ~~the mostly~~ many of the only half a century long North American tree ring series towards the present, but also possibly extending some of them into the first millennium CE seem to us like worthwhile efforts (Babst et al., 2017). ~~Moreover~~Additionally, a relative ~~"flatness"~~ "flatness" of spectra on sub-decadal to multi-decadal time-scales contrasting with an inflated variance of the multi-decadal to millennial variability

700 (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 this 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 including explicitly the annually dated proxies with the chronological uncertainties into the scheme, what became 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 dated proxy archives with chronological un-
certainties. 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) time-scales. 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.

Figure A1. Calibration and Validation results. Top row: $\overline{CRPS}_{pot}$ and Reliability score for the calibration (quadrilaterals) and validation (points) period. Bottom row: CRPS scores corresponding to ~~a~~ 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 artifact.

Appendix A: Calibration and Validations Statistics

In order to estimate the skill of the reconstruction two different measures are used, the average potential continuously ranked probability score ($\overline{\text{CRPS}}_{pot}$) 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 in 95% of the time. The $\overline{\text{CRPS}}_{pot}$ 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 Figure A1 (top row). For the calibration (validation) interval, the $\overline{\text{CRPS}}_{pot}$ is mostly below ~~0.1~~0.2°C (~~0.3~~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 ~~mean~~ 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 analysing the results of a probabilistic reconstruction method. ~~To~~ 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 ~~"reconstruction", mimicking~~ of timeseries. These act as simple surrogates for the calibration and validation ~~intervals. The~~ interval climatology. These are then compared against the target instrumetal data of the validation period, using the $\overline{\text{CRPS}}_{pot}$ ~~over the validation period for these is calculated, and then compared it against the one calculated from the actual ensemble of reconstructions.(~~. Should this value be lower than the $\overline{\text{CRPS}}_{pot}$ comparing the actual reconstruction ensemble against the instrumental data, the reconstruction does not add skill over the climatology. Thus, subtracting the $\overline{\text{CRPS}}_{pot}$ of the reconstruction from the $\overline{\text{CRPS}}_{pot}$ 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}. Figure A1 bottom row ~~). About~~ shows that about half of the grid cells with instrumental data have a ~~CE and RE~~ CRPS_{RE} and CRPS_{CE} that is above zero – and these grid cells are ~~actually also~~ 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.
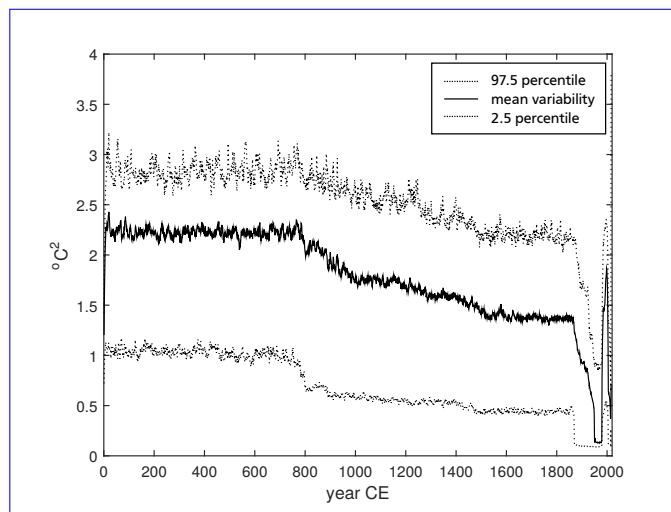
Figure A2. Time variability of spatially averaged intra-ensemble variance of the Arctic2k reconstruction together with the respective ensemble-based 95% CIs.

Appendix B: Intra-ensemble variance of the reconstruction

Figure A2 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 multi-proxy 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 A3 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 CF reconstruction based on 8 proxy records only, one of the coldest period 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 coverage, respectively.

## 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, ~~we analyse the scaling properties of both . The used methods require the analysed time series data to be normally distributed~~ 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 $p = 0.05$~~. Additionally, QQ-plots are analysed~~, and additionally Q-Q
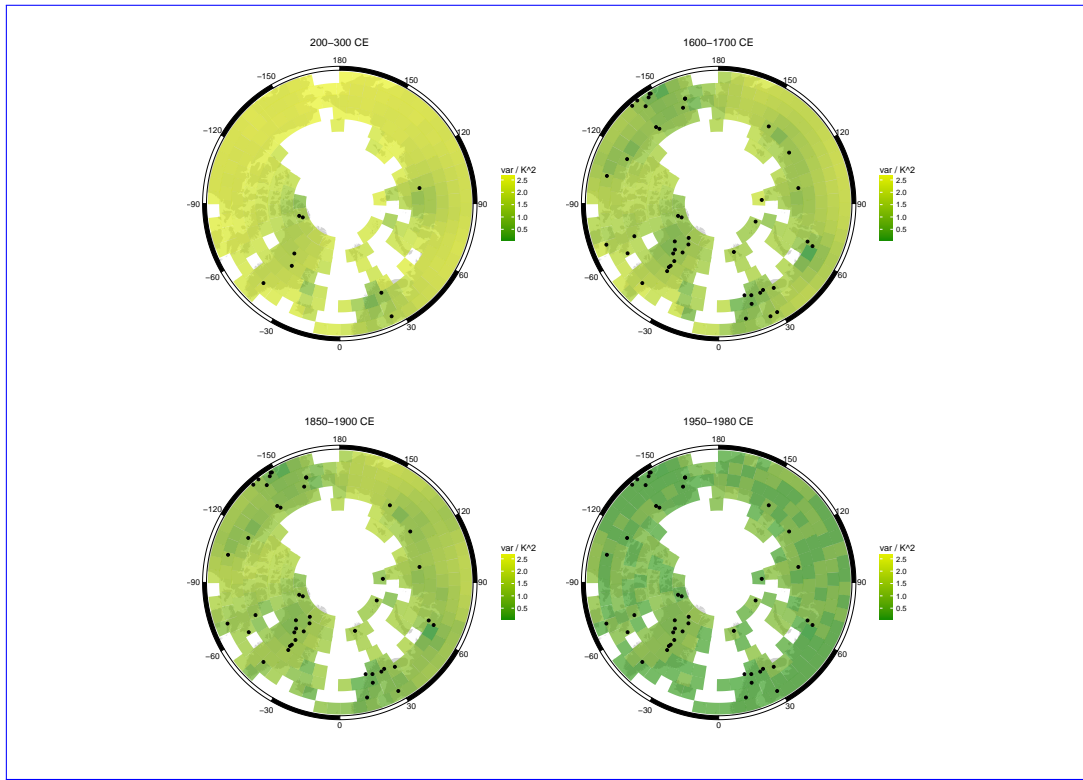
Figure A3. Time averaged intra-ensemble variance of the Arctic2k reconstruction shown for the four subperiods with a distinct difference in proxy data density (200–300 CE vs. 1600–1700 CE, panels a and b) and calibration subperiods with different instrumental data coverage (1850–1900 vs. 1950–1980 CE, panels c and d). Black dots shows the proxy locations with a least one data point over the period of averaging.

770   plots are checked. Then, the power spectral density (PSD) is used to study the variability on different ~~time-scales~~ frequencies for the records~~that were not deviating substantially from a normal distribution~~, using the periodogram as an estimator of the ~~PDS~~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

775   $f_{N/2} = 1/2$ is the Nyquist frequency~~(the highest frequency that can be resolved)~~.

The characteristic shape of the spectrum provides useful information about the temporal persistence or memory of the underlying process~~if~~. If the data is 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

780   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 underly-
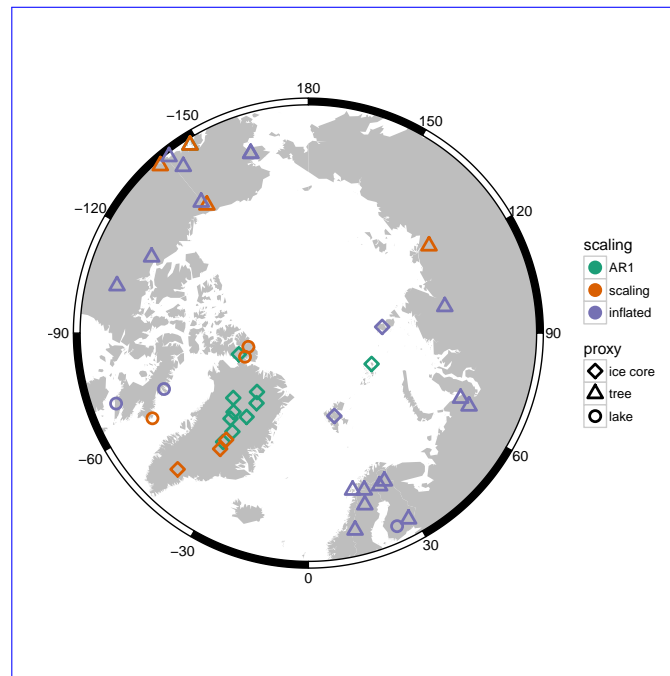
Figure A4. Proxy type (~~circle~~triangles: ice cores, ~~triangles~~diamonds: tree rings, squares: lake sediments,) and ~~scaling~~persistence properties colour coded.

ing process is closer to an AR(1) process. In all spectral analyses, the fitting is applied to log-binned periodograms to ensure that all time scales 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
785 from normality are due to nonlinear dynamics associated with e.g. volcanic eruptions.

C1 ~~Analyses~~ 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 ~~excluded from further spectral analyses, as second order statistics are insufficient to describe these types of records~~normalized before analysis. After normalization, around 60% of the individual proxy records are Gaussian according to the Q-Q plots and
790 the p-values from the Kolmogorov-Smirnov's test. The characteristic shape of the spectra for ~~the remaining~~ all of the proxy records are ~~then~~ classified into three spectral categories~~according to their characteristics~~: (1) AR(1) processes, (2) persistent ~~power law~~ power-law processes with spectral exponent $0 < \beta < 1$, and (3) records exhibiting weak persistence on high frequencies, and increased levels of variability on frequencies corresponding to time scales longer than decadal–centennial. Figure A4 illustrates the spatial distribution of the proxy records with proxy type indicated by shape, and categories with colours. The
795 Greenland records are similar to either an AR(1) or ~~a LRM processes~~an LRM process. The Greenland LRM records are in fact only weakly persistent, with a ~~scaling~~spectral exponent $0 < \beta < 0.3$. There is thus little evidence of long-term cooling due to
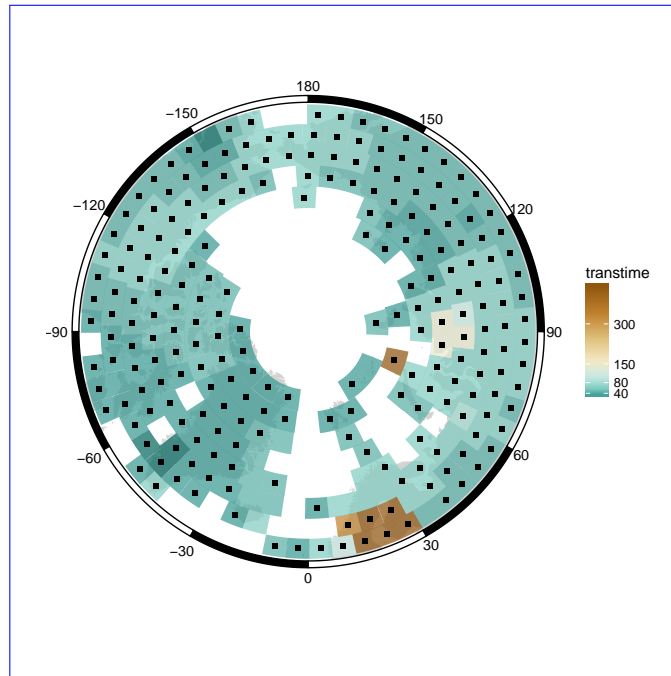
Figure A5. Analysis of scaling properties of the reconstruction. The transition time-scale (colour coded) and strength of change (black dot: exponent change > 1). Most of the reconstruction domain resembles an AR 1 process.

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 of these records dominates the reconstructed long-term variability, the resulting reconstructions does exhibit similarly low variability at long time scales.

800 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~~ 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; Büntgen et al., 2015). The ~~memory properties~~ persistence in a number of millennium-long climate model simulations and proxy-based temperature reconstructions

805 have been studied in (Østvand et al., 2014; 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 time scales.

C2 ~~Analyses~~ Spectral analyses of climate field reconstruction

The resulting p-values from Kolmogorov-Smirnov's test indicate that for individual locations of the field reconstruction, about ~~80~~60-80% of the ensemble members are Gaussian. For each ensemble member of the reconstruction, and each location, a

810 spectral analysis is performed. Then, the persistence is analysed for ensemble-averaged spectra of each location ~~are analysed for their scaling properties~~. The analyses ~~indicates~~ indicate that the reconstructed temperature ~~in all grid cells except one~~ is

36
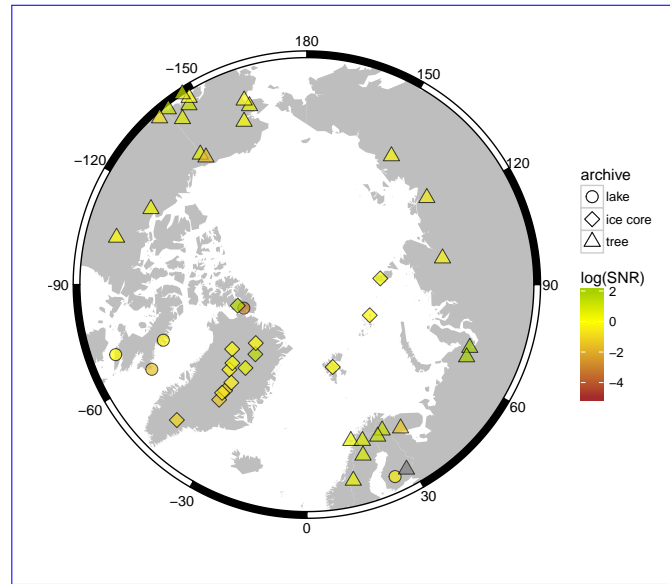
Figure A6. The mean of the posterior SNR for all proxy series. Note the logarithmic scale. There is one series with a negative β_1 (grey).

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
815 of a bilinear continuous function for the spectrum. The detected break is located where the two lines intersect.

The coloured map in Figure ~~A5~~ A5 shows the spatial distribution of the found transition time scales, black dots indicate that the difference between the ~~scaling~~ spectral exponents for low and high frequencies is more than one. 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 ~~break in the scaling~~ transition in the spectral slope (black symbols). Only the East coast
820 of Greenland and the Scandinavian sector has slightly less difference between the high- and low-frequency variability, that is ~~the scaling~~, the spectral exponent does not change much between the two identified scaling regimes. This indicates more similarity to ~~a~~ an LRM process. Additionally, the transition time scale is ~~close to century scale for some regions, and even~~ above a hundred years for a ~~single location in Arctic Russia~~number of single locations. There, the reconstruction is indeed closer to ~~a~~ an LRM process than an AR(1) process.

825 **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. A6. Note that one proxy series (Finnish Lakelands) has a negative (inverted) response ($\beta_1 \approx -0.38 \pm 0.7$).

Appendix D: Input proxy data

D1 Time-scale 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 time-scales in for the generation of ensembles of chronologies. BAM (Banded Archive Modelling) simulates the time-scale 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 $i \in \{1, ..., n\}$, and a neighbouring x_{i+1} with $t_{i+1}, i \in \{2, ..., m\}$, 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, $k \in \{1, ..., K\}$, such as volcanic sulphate peaks or tephras with ages known to a specific precision (σ_k), a two-step procedure was implemented. The first step involved an MCMC simulation of M perturbed sets of tie points $\left[t_k^{\tilde{m}}\right]$ 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 $\left[t_k^{\tilde{m}}, t_{k+1}^{\tilde{m}}\right]$, $k \in \{1, ..., K-1\}$ between the perturbed pairs of tie points time-scale modelling was applied, and only those that satisfied a criterion of $\sum \delta_i = t_k^{\tilde{m}} + 1 - t_k^{\tilde{m}}$ were retained for further analysis. For ages older than $t_K^{\tilde{m}}$ a model with a free boundary was used instead. In total $M = 1000$ time-scales $\tilde{t}_i^m$ per proxy archive were generated. Using interpolation, the proxy series x_i were further projected on the generated time-scales $\tilde{t}_i^m$ 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 time-scale the uncertainty of the modelled time-scale 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

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

where for a particular proxy archive $\delta t_{\max} = \mathrm{argmax}_k \left(t_{k+1} - t_k \right)$, $k \in \{1, ..., K-1\}$, or the entire length of the chronology, and Δ_t denotes an estimated largest offset of the reported time-scale from the unknown true time-scale. 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 where the chronological uncertainties were not reported, a conservative estimate of $[\Theta_1, \Theta_2]$ $[0.05, 0.05]$ was assigned.

Table A1 shows the list of proxy series together with parameters of the model used to simulate the annual layer counting process. In total ensembles of time-scales for 13 annually dated records of the Arctic2k network, 6 ice-core and 7 lake sediment records, plus seven annually dated ice cores from the Greenland German traverse from ~~1985~~ 1993–1995 (recently reanalysed by Weißbach et al., 2016), are generated.

Table A1. List of the proxy records including the proxies of the Arctic2k network (PAGES 2k Consortium, 2017) with layer counted time-scales used in the present study together with parameters of the probabilistic model used in MC simulations of the layer counting process. The archives with a lacking information on the dating uncertainty are marked '*', a conservative estimate of $[\Theta_1, \Theta_2]$ $[0.05, 0.05]$ was used in time-scale modelling.

Site ID	Pages2k site name	Location Lat, Lon	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 A2. 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.

	~~Site~~
PAGES2k.ID	~~site Name~~ PAGES2k.Sitename
~~NAm-ak057~~ NAm_091	Almond Butter ~~Lower D'Arrigo et al. (2005)65.2-162.216816072002summerNAm-ak058 Almond Butter~~ Upper
~~Arc-Arjeplog.Bjorklund.2014~~ Arc_065	Arjeplog
~~Arc-Avam-Taimyr.Briffa.2008~~ Arc_002	Avam-Taimyr
~~NAm-ak060~~ NAm_126	~~Burnt Over D'Arrigo et al. (2005)65.2-162.325916212002summerNAm-ak035 Canyon Creek PAGES 2k Consortium (2017)63.3-147.888416421997~~
~~NAm-ak010~~ Eur_013	~~Eureka Summit~~ Finnish Lakelands
~~NAm-ak132~~ NAm_104	Firth River 1236
~~Arc-GulfofAlaska.Wiles.2014~~ Arc_074	Forfjorddalen
Arc_007	Gulf of Alaska
~~NAm-ak001~~ NAm_083	Herring Alpine
~~NAm-cana155~~ NAm_127	Hornby Cabin
~~Arc-Indigirka.Hughes.1999~~ Arc_016	Indigirka
~~Arc-Zhang.Jamtland.2016~~ Arc_063	Jamtland, Sweden
~~Arc-Kittelfjall.Bjorklund.2013~~	~~Kittelfjäll~~ Khibiny
~~NAm-ak046~~ NAm_002	Kobuk/Noatak
~~NAm-cana354~~ Arc_071	~~Landslide~~ Laanila
NAm_032	Landslide
~~Arc-Porter2013~~ Arc_024	~~Mackenzie Delta~~ Lena River
~~NAm-ak021~~ NAm_088	Miners Well
~~NAm-ak074~~ NAm_094	Nabesna Mine
~~NAm-ak117~~ Eur_003	Northern ~~Alaska Composite~~ Scandinavia
~~Eur-NScan.Esper.2012~~ Arc_061	~~Northern Scandinavia~~ Polar Urals
~~NAm-ak094~~ NAm_003	Prince William Sound
~~NAm-cana378~~ NAm_100	~~Rock Glacier Yukon~~ Seward Composite
Arc_062	~~annual~~Tornetrask
~~NAm-ak115~~ Arc_079	~~Seward Composite~~ Yamalia
Arc_008	~~summer~~Yukon
~~NAm-cana029~~ height	

870 The two largest statistically significant cooling rates in the entire ensemble with average temperature changes of -0.05 ± 0.01°C/year and -0.04 ± 0.01°C/year over three decades are registered at 1450 CE and 1669 CE, respectively, while a recovery after the first cooling centered at 1477 CE featured a warming rate of 0.04 ± 0.01°C/year over the same time period. In terms of the rate of changes attained, the first cooling/warming episode appears unique over the 2000-year long reconstruction, embracing one of the coldest decades in the reconstruction ensemble. At the highlighted centennial timescale, the most rapid changes are

875 the MCA to LIA transition with a cooling of -0.006 ± 0.002°C/year centered at 1040 CE, cooling towards one of the LIA SAT minima at 1577 CE with -0.04 ± 0.02°C/year, and the transition to CWP centered at 1905 CE with an average warming rate of 0.01 ± 0.001°C over about 30 years.

Appendix: Data availability

~~The~~

Table A3. List of the lake and ice core data from the PAGES2k database that was not used in this study.

Site ID	Pages2k site name	Location Lat, Lon	Proxy type	Time period (yr CE)	Resolution	Reason
Arc_005	Iceland	Björklund et al. (2012) 64.77,-18.37	doc	1242 − 1967	1	20 yr averages
Arc_018	Camp Century	6-7-877,17,-61.13	ic	151 − 1995	1	not annual as per original
Arc_075	Austfonna	D'Arrigo et al. (2005) 79.83,24.02	ic	1400 − 1998	1	interpolated onto
Arc_059	Devon Ice Cap	D'Arrigo et al. (2006) 75.33,-82.5	ic	−421 − 1971	5	not annual
Arc_045	Penny Ice Cap P96	Schneider et al. (2015) 67.25,-66.75	ic	65 − 1980	25	not annual
Arc_001	Blue Lake	6-7-868.09,-150.47	ls	730 − 2000	1	very nonlinear response
Arc_014	Lake Lehmilampi	Helama et al. (2014) 63.62,29.1	ls	251 − 1800	1	exact interpretation, annual and centennial
Arc_022	Hvítárvatn	64.6,-19.8	ls	−1 − 2000	summer+1	not annual
Arc_069	Kongressvatnet	McCarroll et al. (2013) 78.02,17,13.93	ls	232 − 2008	10	not annual
Arc_067	Hallet Lake	6-761.5,-146.2	ls	116 − 2005	11	not annual
Arc_050	Lake Hampträsk	Lindholm et al. (2011) 60.28,25.42	ls	1359 − 1994	14	not annual
Arc_070	Lake E	6-7-867,-50.7	ls	−3642 − 1876	19	not annual
Arc_043	Braya Sø	MacDonald et al. (1998) 67,-50.7	ls	−998 − 1999	29	not annual
Arc_040	Moose Lake	661.35,-143.6	ls	−718 − 1963	36	not annual
Arc_051	Lake Pieni-Kauro	Melvin et al. (2013) 64.28,30.12	ls	−1462 − 1979	44	not annual
Arc_041	Hudson Lake	6-7-861.9,-145.66	ls	−837 − 1997	47	not annual
Arc_054	Lake 4	Briffa et al. (2013) 65.1,-83.79	ls	−634 − 1997	50	not annual
Arc_042	Screaming Lynx Lake	6-766.07,-145.4	ls	−1067 − 1988	51	not annual

42

Table A4. Summary of some of major features of the new Arctic2K reconstruction ensemble. Anomalies are given reative to the period of 1850–2000

Feature name	~~Yukon~~ Year/Period ~~-140.7~~considered, years CE	~~D'Arrigo et al. (2006)~~Anomaly value ~~300~~(STD)°C	~~67.9~~Note ~~1177~~
Warmest century	~~2000~~901–1000	~~summer~~ 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)°C/century	before the onset CWP
Largest warming trend magnitude	centered at 1905	0.01 ± 0.001	per year, centennial scale
Largest warming trend magnitude	centered at 1477	0.04 ± 0.01	per year, ca. 30 years scale
Largest cooling trend magnitude	centered at 1040	-0.006 ± 0.002	per year, centennial scale
Largest cooling trend magnitude	centered at 1450	-0.05 ± 0.01	per year, ca. 30 years scale

880   The input proxy data is available ~~for download at NOAA URL GOES HERE~~ either through the individual publications (see tables), the majority is 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 homepage `http://browse.ceda.ac.uk/browse/badc/cru/data/cru_ts/cru_ts_3.24.01`
885 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).

References

890 Altekar, G., Dwarkadas, S., Huelsenbeck, J. P., and Ronquist, F.: Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylo-
genetic inference, Bioinformatics, 20, 407–415, doi:10.1093/bioinformatics/btg427, http://bioinformatics.oxfordjournals.org/content/20/
3/407.abstract, 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, Journal of Climate,
895 26, 3001–3012, doi:10.1175/jcli-d-11-00139.1, 2013.

Babst, F., Poulter, B., Bodesheim, P., Mahecha, M. D., and Frank, D. C.: Improved tree-ring archives will support earth-system science,
Nature Ecology & Evolution, 1, 1–2, 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,
The Holocene, 9, 79–84, doi:10.1191/095968399672825976, 1999.

900 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
Science Reviews, 29, 3489 – 3504, doi:10.1016/j.quascirev.2010.06.041, aPEX: Arctic Palaeoclimate and its Extremes, 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, Climate of the Past, 10, 181–198, doi:10.5194/cp-10-181-2014, http://www.clim-past.net/10/181/
905 2014/, 2014.

Björklund, J. A., Gunnarson, B. E., Krusic, P. J., Grudd, H., Josefsson, T., Östlund, L., and Linderholm, H. W.: Advances towards improved
low-frequency tree-ring reconstructions, using an updated *Pinus sylvestris L.* MXD network from the Scandinavian Mountains, Theoretical
and Applied Climatology, 113, 697–710, doi:10.1007/s00704-012-0787-7, 2012.

Björklund, J. A., Gunnarson, B. E., Seftigen, K., Esper, J., and Linderholm, H. W.: Blue intensity and density from northern Fennoscandian
910 tree rings, exploring the potential to improve summer temperature reconstructions with earlywood information, Climate of the Past, 10,
877–885, doi:10.5194/cp-10-877-2014, 2014.

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 Research, 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
915 recent temperature and radial tree growth spanning 2000 years across northwest Eurasia, Phil. Trans. R. Soc. B, 363, 2271–2284, 2008.

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 Science
Reviews, 72, 83–107, doi:10.1016/j.quascirev.2013.04.008, 2013.

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

Chaudhuri, P. and Marron, J. S.: SiZer for exploration of structures in curves, Journal of the American Statistical Association, 94, 807–823,
1997.

Christiansen, B.: Reconstructing the NH mean temperature: Can underestimation of trends and variability be avoided?, Journal of Climate,
925 24, 674–692, 2011.

44

Christiansen, B. and Ljungqvist, F. C.: Challenges and perspectives for large-scale temperature reconstructions of the past two millennia, Reviews of Geophysics, 55, 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 Research, 66, 342–355, doi:10.1016/j.yqres.2006.06.005, 2006.

930 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, Climate of the Past, 10, 825–841, doi:10.5194/cp-10-825-2014, http://www.clim-past.net/10/825/2014/, 2014.

Comiso, J. C., Parkinson, C. L., Gersten, R., and Stock, L.: Accelerated decline in the Arctic sea ice cover, Geophysical Research Letters, 35, 2008.

935 Cook, E. R., Briffa, K. R., and Jones, P. D.: Spatial regression methods in dendroclimatology – A review and comparison of 2 techniques, International Journal of Climatology, 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, Journal of Paleolimnology, 41, 77–94, doi:10.1007/s10933-008-9252-0, 2008.

940 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, Journal of Paleolimnology, 41, 77–94, doi:10.1007/s10933-008-9252-0, 2009.

Cropper, J. P. and Fritts, H. C.: Tree-Ring Width Chronologies from the North American Arctic, Arctic and Alpine Research, 13, 245, doi:10.2307/1551032, 1981.

945 Curry, J. A., Schramm, J. L., and Ebert, E. E.: Sea ice-albedo climate feedback mechanism, Journal of 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, doi:10.1029/2004GL022055, 2005.

Davi, N.: Boreal temperature variability inferred from maximum latewood density and tree-ring width data, Wrangell Mountain region, Alaska, Quaternary Research, doi:10.1016/s0033-5894(03)00115-7, 2003.

950 Divine, D. and Dick, C.: Historical variability of sea ice edge position in the Nordic Seas, Journal of Geophysical Research, 111, C01 001, doi 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 Research, 30, doi:10.3402/polar.v30i0.7379, 2011.

955 Divine, D. V., Godtliebsen, F., and Rue, H.: A modelling approach to assessing the timescale uncertainties in proxy series with chronological errors, Climate of the Past Discussions, 8, 31–61, doi:10.5194/cpd-8-31-2012, http://www.clim-past-discuss.net/8/31/2012/, 2012.

Dobrovolný, P., Moberg, A., Wilson, R., Brázdil, R., Pfister, C., Glaser, R., van Engelen, A., Limanówka, D., Kiss, A., Riemann, D., Halíčková, M., Macková, J., Luterbacher, J., and Böhm, R.: Temperature reconstruction of Central Europe derived from documentary evidence since AD 1500, Climatic Change, 101, 69–107, 2010.

960 D'Arrigo, R., Wilson, R., and Jacoby, G.: On the long-term context for late twentieth century warming, Journal of Geophysical Research, 111, doi:10.1029/2005jd006352, 2006.

Earl, D. J. and Deem, M. W.: Parallel tempering: Theory, applications, and new perspectives, Physical Chemistry Chemical Physics, 7, 3910–3916, doi:10.1039/B509983H, http://dx.doi.org/10.1039/B509983H, 2005.

Emile-Geay, J. and Tingley, M.: Inferring climate variability from nonlinear proxies: application to palaeo-ENSO studies, Climate of the Past, 12, 31–50, 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, Nature Climate Change, doi:10.1038/NCLIMATE1589, 2012.

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, Geophysical Research Letters, 25, 1749–1752, doi:10.1029/98GL01177, 1998.

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

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

Gneiting, T. and Raftery, A. E.: Strictly proper scoring rules, prediction, and estimation, Journal of the American Statistical Association, 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, Climate Dynamics, 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, Journal of Geophysical Research, 111(D10), D07 110, doi:10.1029/2005JD006494, 2006.

Grinsted, A., Moore, J. C., and Jevrejeva, S.: Reconstructing sea level from paleo and projected temperatures 200 to 2100 AD, Climate Dynamics, 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, Journal of Geophysical Research, 102(C12), 26 455–26 470, doi: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 2,000 years, Climate Dynamics, 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, International Journal of Climatology, 34, 623–642, doi: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, doi:10.2478/s13386-013-0163-0, 2014.

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

Hind, A., Zhang, Q., and Brattström, G.: Problems encountered when defining Arctic amplification as a ratio, Scientific Reports, 6, 2016.

Hofer, D., Raible, C. C., and Stocker, T. F.: Variations of the Atlantic meridional overturning circulation in control and transient simulations of the last millennium, Climate of the Past, 7, 133–150, doi:10.5194/cp-7-133-2011, http://www.clim-past.net/7/133/2011/, 2011.

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, Journal of Geophysical Research, 106, 14 265–14 274, doi: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, The Holocene, 10, 9–19, doi:10.1191/095968300676746202, http://hol.sagepub.com/content/10/1/9.abstract, 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.

Hunke, E. C., Lipscomb, W. H., and Turner, A. K.: Sea-ice models for climate study: retrospective and new directions, Journal of Glaciology, 56, 1162–1172, doi: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, doi:10.1017/CBO9781107415324, www.climatechange2013.org, 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, Climatic Change, 14, 39–59, doi:10.1007/bf00140174, 1989.

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, Journal of Geophysical Research, 117, D05 127, doi: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, doi: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, Journal of Paleolimnology, 16, 239–255, doi:10.1007/BF00176939, 1996.

Leopardi, P.: A partition of the unit sphere into regions of equal area and small diameter, Electronic Transactions on Numerical Analysis, 25, 309–327, 2006.

Li, Z., Protopopescu, V. A., Arnold, N., Zhang, X., and Gorin, A.: Hybrid parallel tempering and simulated annealing method, Applied Mathematics and Computation, 212, 216–228, doi:10.1016/j.amc.2009.02.023, 2009.

Lindholm, M., Jalkanen, R., Salminen, H., Aalto, T., and Ogurtsov, M.: The height-increment record of summer temperature extended over the last millennium in Fennoscandia, The Holocene, 21, 319–326, doi:10.1177/0959683610378875, 2011.

Ljungqvist, F., Krusic, P. J., Brattström, G., and Sundqvist, H. S.: Northern Hemisphere temperature patterns in the last 12 centuries, Climate of the Past, 8, 227–249, 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, 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, Environmental Research Letters, 11, 024 001, http://stacks.iop.org/1748-9326/11/i=2/a=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 and Alpine Research, 30, 334, doi:10.2307/1552005, 1998.

47

1040   Mann, M. E., Zhang, Z., Rutherford, S., Bradley, R. S., Hughes, M. K., Shindell, D., Ammann, C., Faluvegi, G., and Ni, F.: Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly, Science, 326, 1256–1260, 2009.

Maslanik, J., Stroeve, J., Fowler, C., and Emery, W.: Distribution and trends in Arctic sea ice age through spring 2011, Geophysical Research Letters, 38, doi:10.1029/2011GL047735, L13502, 2011.

Masson-Delmotte, V., Landais, A., Stievenard, M., Cattani, O., Falourd, S., Jouzel, J., Johnsen, S. J., Dahl-Jensen, D., Sveinsbjornsdottir, A.,
1045   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, Journal of Geophysical Research, 110(D9), 14 102, doi: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, The Holocene, 23, 471–484,
1050   doi:10.1177/0959683612467483, 2013.

McKay, N. P. and Kaufman, D. S.: An extended Arctic proxy temperature database for the past 2,000 years, Scientific Data, 1, 140 026, doi: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
1055   human activity, Reviews of Geophysics, 52, 185–217, doi:10.1002/2013RG000431, 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, The Holocene, 23, 364–373, doi:10.1177/0959683612460791, 2013.

Miettinen, A., Divine, D., Koç, N., Godtliebsen, F., and Hall, I. R.: Multicentennial Variability of the Sea Surface Temperature Gradient across the Subpolar North Atlantic over the Last 2.8 kyr*,+, Journal of Climate, 25, 4205–4219, doi:10.1175/JCLI-D-11-00581.1, 2012.

1060   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, doi:10.1002/2015PA002849, 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 Science Reviews, 29, 1779–1790, 2010.

Mitchell, J., Dzerdzeevskii, B., and Flohn, H.: Climatic change: report of a working group of the Commission for Climatology, Technical
1065   Note 79, World Meteorological Organisation, 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, Journal of Paleolimnology, 25, 503–517, doi: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 Science Reviews, 47, 1–14, 2012.

1070   Nicolle, M., Debret, M., Massei, N., Hormes, A., Colin, C., DeVernal, A., Divine, D., Werner, J., Korhola, A., and the Arctic 2K working Group: Climate variability in subarctic area for the last two millennia, submitted to Climate of the Past, 2017.

Nilsen, T., Werner, J. P., and Divine, D. V.: How wrong are climate field reconstruction techniques in reconstructing a climate with long range memory?, in prep.

Nilsen, T., Rypdal, K., and Fredriksen, H.-B.: Are there multiple scaling regimes in Holocene temperature records?, Earth System Dynamics,
1075   7, 419–439, doi:10.5194/esd-7-419-2016, http://www.earth-syst-dynam.net/7/419/2016/, 2016.

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

O'Connor, F. M., Boucher, O., Gedney, N., Jones, C., Folberth, G., Coppell, R., Friedlingstein, P., Collins, W., Chappellaz, J., Ridley, J., et al.: Possible role of wetlands, permafrost, and methane hydrates in the methane cycle under future climate change: a review, Reviews of Geophysics, 48, 2010.

Ojala, A. E. and Alenius, T.: 10 000 years of interannual sedimentation recorded in the Lake Nautajärvi (Finland) clastic–organic varves, Palaeogeography, Palaeoclimatology, Palaeoecology, 219, 285 – 302, doi:http://dx.doi.org/10.1016/j.palaeo.2005.01.002, http://www.sciencedirect.com/science/article/pii/S0031018205000180, 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), Climate of the Past, 9, 2379–2389, doi:10.5194/cp-9-2379-2013, http://www.clim-past.net/9/2379/2013/, 2013.

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

Otterå, O. H., Bentsen, M., Drange, H., and Suo, L.: External forcing as a metronome for Atlantic multidecadal variability, Nature Geoscience, 3, 688–694, doi:10.1038/ngeo955, 2010.

Overpeck, J., Hughen, K., Hardy, D., Bradley, R., Case, R., Douglas, M., Finney, B., Gajewski, K., Jacoby, G., Jennings, A., et al.: 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, Nature Geoscience, 6, 339–346, 2013.

PAGES 2k Consortium: A global multiproxy database for temperature reconstructions of the Common Era, Scientific Data, 4, 170 088, doi:10.1038/sdata.2017.88, 2017.

Perovich, D., Grenfell, T., Light, B., and Hobbs, P.: Seasonal evolution of the albedo of multiyear Arctic sea ice, Journal of Geophysical Research: Oceans, 107, 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, Annals of Glaciology, 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, Geophysical Research Letters, 35, 2008.

Porter, T. J., Pisaric, M. F., Kokelj, S. V., and deMontigny, P.: A ring-width-based reconstruction of June–July minimum temperatures since AD 1245 from white spruce stands in the Mackenzie Delta region, northwestern Canada, Quaternary Research, 80, 167–179, doi:10.1016/j.yqres.2013.05.004, 2013.

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.

Schleussner, C.-F., Divine, D. V., Donges, J. F., Miettinen, A., and Donner, R. V.: Indications for a North Atlantic ocean circulation regime shift at the onset of the Little Ice Age, Climate Dynamics, 45, 3623–3633, doi:10.1007/s00382-015-2561-x, 2015.

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, Geophysical Research Letters, 42, 4556–4562, doi:10.1002/2015gl063956, 2015.

Schulz, M., Prange, M., and Klocker, A.: Low-frequency oscillations of the Atlantic Ocean meridional overturning circulation in a coupled climate model, Climate of the Past, 3, 97–107, doi:10.5194/cp-3-97-2007, http://www.clim-past.net/3/97/2007/, 2007.

Schweingruber, F. H. and Briffa, K. R.: Tree-Ring Density Networks for Climate Reconstruction, in: Climatic Variations and Forcing Mechanisms of the Last 2000 Years, edited by Jones, P. D., Bradley, R. S., and Jouzel, J., chap. 2, p. 43–66, Springer Nature, doi:10.1007/978-3-642-61113-1:etal3, 1996.

Seidenkrantz, M.-S., Aagaard-Sørensen, S., Sulsbr"uck, H., Kuijpers, A., Jensen, K., and Kunzendorf, H.: Hydrography and climate of the last 4400 years in a SW Greenland fjord: implications for Labrador Sea palaeoceanography, The Holocene, 17, 387–401, doi:10.1177/0959683607075840, 2007.

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 Research, 54, 113–128, 2012.

Sigl, M., Winstrup, M., McConnell, J., Welten, K., Plunkett, G., Ludlow, F., Büntgen, U., Caffee, M., Chellman, N., Dahl-Jensen, D., et al.: 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, Geophysical Research Letters, 35, 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, Journal of Paleolimnology, 41, 209–224, doi:10.1007/s10933-008-9258-7, 2009.

Thornalley, D. J., Elderfield, H., and McCave, I. N.: Holocene oscillations in temperature and salinity of the surface subpolar North Atlantic, Nature, 457, 711–714, doi:10.1038/nature07717, 2009.

Tingley, M. P.: A Bayesian ANOVA scheme for calculating climate anomalies, with applications to the instrumental temperature record, Journal of Climate, 25, 777–791, doi: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, Journal of Climate, 23, 2759–2781, doi: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, Journal of Climate, 23, 2782–2800, doi: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, doi: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, doi:10.1002/env.2368, http://dx.doi.org/10.1002/env.2368, env.2368, 2016.

Vermeer, M. and Rahmstorf, S.: Global sea level linked to global temperature, Proceedings of the National Academy of Sciences, 106, 21 527–21 532, 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 Science Reviews, 29, 522 – 538, doi:http://dx.doi.org/10.1016/j.quascirev.2009.11.002, http://www.sciencedirect.com/science/article/pii/S0277379109003655, 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.

Wahl, E. R. and Smerdon, J. E.: Comparative performance of paleoclimate field and index reconstructions derived from climate proxies and noise-only predictors, Geophysical Research Letters, 39, L06 703, doi:10.1029/2012GL051086, 2012.

Walsh, J. E., Fetterer, F., Scott Stewart, J., and Chapman, W. L.: A database for depicting Arctic sea ice variations back to 1850, Geographical Review, 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, Geophysical Research Letters, 42, 7162–7170, doi: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, Climate of the Past, 12, 171–188, doi:10.5194/cp-12-171-2016, http://www.clim-past.net/12/171/2016/, 2016.

Werner, J. and Tingley, M.: Technical Note: Probabilistically constraining proxy age–depth models within a Bayesian hierarchical reconstruction model, Climate of the Past, 11, 533–545, doi:10.5194/cpd-10-4499-2014, 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, Journal of Climate, doi: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, The Holocene, 24, 198–208, doi:10.1177/0959683613516815, 2014.

Young, N. E., Schweinsberg, A. D., Briner, J. P., and Schaefer, J. M.: Glacier maxima in Baffin Bay during the Medieval Warm Period coeval with Norse settlement, Science Advances, 1, e1500 806, doi:10.1126/sciadv.1500806, 2015.

Zhang, H., Werner, J., Luterbacher, J., Garcia-Bustamante, E., González-Rouco, F., Wagner, S., Zorita, E., Fraedrich, K., Jungclaus, J., and Zhu, X.: Gridded decadal warm season temperature over East Asia in the past 1200 years, in review.

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, Environmental Research Letters, 10, 084 020, http://stacks.iop.org/1748-9326/10/i=8/a=084020, 2015.

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, Climate of the Past, 12, 1297–1312, doi:10.5194/cp-12-1297-2016, 2016.