



Seasonal reconstructions coupling ice core data and an isotope-enabled climate model – methodological implications of seasonality, climate modes and selection of proxy data

Jesper Sjolte¹, Florian Adolphi^{1,3}, Bo M. Vinther⁴, Raimund Muscheler¹, Christophe Sturm², Martin Werner⁵, and Gerrit Lohmann⁵

¹Department of Geology – Quaternary Science, Lund University, Sölvegatan 12, 22362 Lund, Sweden

²Swedish Meteorological and Hydrological Institute, 60176 Norrköping, Sweden

³Climate and Environmental Physics & Oeschger Centre for Climate Change Research, Physics Institute, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland

⁴Physics of Ice, Climate and Earth, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

⁵Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Sciences, Bussestr. 24, 27515 Bremerhaven, Germany

Correspondence: Jesper Sjolte (jesper.sjolte@geol.lu.se)

Received: 4 November 2019 – Discussion started: 19 November 2019

Revised: 2 July 2020 – Accepted: 5 August 2020 – Published: 11 September 2020

Abstract. The research area of climate field reconstructions has developed strongly during the past 20 years, motivated by the need to understand the complex dynamics of the earth system in a changing climate. Climate field reconstructions aim to build a consistent gridded climate reconstruction of different variables, often from a range of climate proxies, using either statistical tools or a climate model to fill the gaps between the locations of the proxy data. Commonly, large-scale climate field reconstructions covering more than 500 years are of annual resolution. In this method study, we investigate the potential of seasonally resolved climate field reconstructions based on oxygen isotope records from Greenland ice cores and an isotope-enabled climate model. Our analogue-type method matches modeled isotope patterns in Greenland precipitation to the patterns of ice core data from up to 14 ice core sites. In a second step, the climate variables of the best-matching model years are extracted, with the mean of the best-matching years comprising the reconstruction. We test a range of climate reconstructions, varying the definition of the seasons and the number of ice cores used. Our findings show that the optimal definition of the seasons depends on the variability in the target season. For winter, the vigorous variability is best captured when defining the season as December–February due to the dominance

of large-scale patterns. For summer, which has weaker variability, albeit more persistent in time, the variability is better captured using a longer season of May–October. Motivated by the scarcity of seasonal data, we also test the use of annual data where the year is divided during summer, that is, not following the calendar year. This means that the winter variability is not split and that the annual data then can be used to reconstruct the winter variability. In particular when reconstructing the sea level pressure and the corresponding main modes of variability, it is important to take seasonality into account, because of changes in the spatial patterns of the modes throughout the year. Targeting the annual mean sea level pressure for the reconstruction lowers the skill simply due to the seasonal geographical shift of the circulation modes. Our reconstructions based on ice core data also show skill for the North Atlantic sea surface temperatures, in particular during winter for latitudes higher than 50° N. In addition, the main modes of the sea surface temperature variability are qualitatively captured by the reconstructions. When testing the skill of the reconstructions using 19 ice cores compared to the ones using eight ice cores, we do not find a clear advantage of using a larger data set. This could be due to a more even spatial distribution of the eight ice cores. However, including European tree-ring data

to further constrain the summer temperature reconstruction clearly improves the skill for this season, which otherwise is more difficult to capture than the winter season.

1 Introduction

Knowledge of past climate is essential to understand the range and processes of natural climate variability and impact of internal and external forcing, as well as it serving as baseline to assess anthropogenic influences. The widespread implementation of weather observations dates back to about 1850, with sparse coverage in the early years. In order to investigate changes in weather and climate, as well as to evaluate climate models, so-called reanalysis data sets have been developed. Reanalysis data are gridded data products based on assimilation of weather observations using climate models. The use of reanalysis data sets has seen a wide range of applications due to the gridded data format and global coverage. However, due to being limited to the instrumental period, there is a strong incentive to develop similar products reaching further back in time.

In extratropical regions water-stable isotopes from archives of paleo-precipitation are widely used as climate proxies for temperature; however the variability in water isotopes in precipitation is also related to atmospheric circulation. As vapor condensates from an air parcel during advection or ascent, the water molecules incorporating heavy isotopes (^{18}O , D) condensate more readily than lighter molecules. This means that the isotopic composition depends on the initial vapor content of the air parcel as well as its condensation history. The covariability in vapor content and temperature results in the correlation between local temperature and the isotope composition of precipitation, while the dependency of the isotope content on the pathway results in the connection to atmospheric circulation. The δ notation is commonly used for the isotope ratio of a sample, giving the relative deviation from an isotopic standard (Craig, 1961).

Ice cores are some of the most important archives of the isotope composition of past precipitation. Some Greenland ice cores offer seasonal resolution and in some cases even higher resolution (Furukawa et al., 2017); however the average annual ice accumulation must be larger than 0.2 m yr^{-1} in order for the annual cycle of the isotope composition not to be completely smoothed out by diffusion in the firn (Johnsen, 1977). If the annual cycle is partly preserved, it can be reconstructed by mathematical back-diffusion of the data. It has been shown that seasonal ice core data have high correlation with local temperature and circulation patterns and that the summer and winter data reflect distinctly different spatial and temporal climate variability (Vinther et al., 2010). In particular, Vinther et al. (2010) showed that the seasonal winter $\delta^{18}\text{O}$ has better coherency with annual mean temperature than annual mean $\delta^{18}\text{O}$. This is due to weaker connection of the summer $\delta^{18}\text{O}$ with summer temperature and larger

variability in both winter $\delta^{18}\text{O}$ and temperature, which then dominates the annual signal Vinther et al. (2010).

When studying climate further back than the earliest widespread weather observations, we rely on climate proxy data, such as ice cores. Inherent uncertainties in proxy data include age model uncertainties, seasonality and if there is a stationary relationship between proxy and climate. This means that proxy data sets must be carefully chosen and evaluated, and the data must be well studied to understand the relationship to climate before being incorporated in climate field reconstructions. Pioneering examples of climate field reconstructions include Mann et al. (1998), which regressed climate patterns based on observations on a collection of climate proxy data to obtain a global gridded data set of temperature, and Luterbacher et al. (2001, 2004), which reconstructed European sea level pressure and temperature with a similar regression technique but also using early weather observations as well as historical documentation of weather variability.

Inspired by the techniques used for weather forecasts and reanalysis data, recent climate field reconstructions employ assimilation of climate proxy data using a climate model. The Last Millennium Reanalysis Project (LMR) (Hakim et al., 2016; Tardif et al., 2019) aims to make a global reanalysis using a wide range of proxy data. Their method includes proxy system modeling to link the proxies to the variables of the climate model. The regional studies of Sjolte et al. (2018) and Klein et al. (2019) are climate field reconstructions using Greenland and Antarctic ice core records, respectively. In the case of these two studies, an isotope-enabled climate model was used for the assimilation of isotope records from ice cores, which eliminates the step of calibrating the proxy records to a given environmental variable, such as temperature. These studies all use different statistical approaches when performing the assimilation procedure, where LMR employs a Kalman filter, Klein et al. (2019) a particle trajectory approach and Sjolte et al. (2018) a variation in the analogue method, where the matching of model output to proxy data is done based on empirical orthogonal functions (EOFs). For a brief review of uses of the analogue method see Bothe and Zorita (2019). Common to the studies named above is the use of a static model ensemble. The latter means that there are no constraints on which model year can be chosen as analogy for a given year of the proxy data. This is mainly done for practical reasons since one avoids having to run ensemble simulations step-by-step as it is done for meteorological reanalysis data. One point that sets the study by Sjolte et al. (2018) apart from the other studies mentioned in this section is the use of seasonal proxy data in order to focus on reconstructing the winter season only, as opposed to targeting the variability in the annual mean. As mentioned above in connection with the study by Vinther et al. (2010), the Greenland ice core data show distinctly different variability between summer and winter. Such differences in variability may originate in the relation between climate proxy records

Table 1. Tree-ring sites used to constrain summer reconstructions of 1241–1970, with correlations with observed mean temperature from Wilson et al. (2016) for the indicated months. JA: July–August; JAS: July–September; AMJJAS: April–September; MJJA: May–August; JJA: June–August; MJ: May–June.

Location	Site name	Long.	Lat.	Time period	Corr. with CRUTS3.2 1901–present
Scotland	SCOT	57.08	−3.44	1200–2010	JA: 0.75
E Alps – Tyrol	TYR	47.30	12.30	1053–2003	JAS: 0.72
Jämtland	JAEM	63.30	13.25	783–2011	AMJJAS: 0.75
Tjeggelvas, Arjeplog, Ammarnäs composite	TAA	65.54–66.36	16.06–18.12	1200–2010	MJJA: 0.81
North Fenno	EFmean	66–69	19–32	750–2010	JJA: 0.76
Forfjorddalen	FORF	68.47	15.43	978–2005	JA: 0.71
Tatra	TAT	48–49	19–20	1040–2010	MJ: 0.45
Southern Finland	SFIN	62.19.30	28.19.30	760–2000	MJJA: 0.71

and climate variability, for example due to different climate sensitivity through the seasons, or due to climate variability itself, for example the change of circulation regimes during the year (Hurrell et al., 2003). Due to these questions of seasonality climate field reconstructions targeting the annual mean could therefore, by the nature of both the climate proxies and climate variability, have limited skill. This could to a large extent depend on the definition of the year and may bias reconstructions towards specific seasons despite the use of annual data. The issue with seasonality could in particular play a role when it comes to atmospheric circulation regimes, which shall return to later in Sect. 4.1.

In this study we will investigate the methodological implication of extracting seasonal and annual climate information from Greenland ice cores using a coupled model–data approach. We will use the method by Sjolte et al. (2018) with an extended data set including summer and annual isotope data from ice cores, as well as tree-ring chronologies from Europe. In combining model output with these data sets, we reconstruct sea level pressure (SLP), surface air temperature (T2m) and sea surface temperature (SST). We will test the following:

- the influence the number of ice cores assimilated for the reconstruction
- if the definition of the seasons impact the skill and recorded climate variability in the reconstructions
- if annual data can be used to reconstruct winter variability
- to which extent the governing atmospheric circulation modes can be reconstructed using summer, winter and annual data
- if including tree-ring data can improve the reconstruction for the summer season

- if the reconstructions capture variations in the North Atlantic SSTs, hereunder the main modes of the SST variability.

2 Data

In this study we use the seasonal $\delta^{18}\text{O}$ ice core data of Vinther et al. (2010). We use the data for summer (May–October), winter (November–April) and winter-centered annual mean (August–July) as defined by Vinther et al. (2010). To achieve the longest possible data set with the best regional coverage we chose eight cores covering 1241–1970, and for the largest data set possible we chose all 19 cores covering 1777–1970 (Fig. S1 in the Supplement).

In addition to using the ice core data, we produce reconstructions for summer where tree-ring data are used to further constrain temperature. Tree-ring chronologies using primarily maximum latewood density as climate proxy can have a strong sensitivity to summer temperature. Such records are compiled in Wilson et al. (2016). From this compilation we select tree-ring records that cover the entire study period (1241–1970) and correlate well with local temperature. This leaves us with eight tree-ring records from Europe (Table 1).

We use the isotope-enabled version of ECHAM5/MPI-OM (Werner et al., 2016) in T31L19 configuration, which corresponds to $3.75^\circ \times 3.75^\circ$ horizontal resolution using 19 vertical hybrid levels. The model includes isotope tracers in a fully coupled hydrological cycle, with fractionation taken into account for all phase transitions. The simulation covers the years 800–2005 with natural and anthropogenic forcings, including greenhouse gases, volcanic aerosols, total solar irradiance, land use and orbital forcing. See Sjolte et al. (2018) for full details on the model run.

To evaluate the skill of the reconstructions we use the 20th Century Reanalysis Version 2c (20CR) (Compo et al., 2011) for the period 1851–1970, as well as the accompanying COBE SST data (Ishii et al., 2005). We mainly use 20CR to assess the skill in spatial correlation patterns and

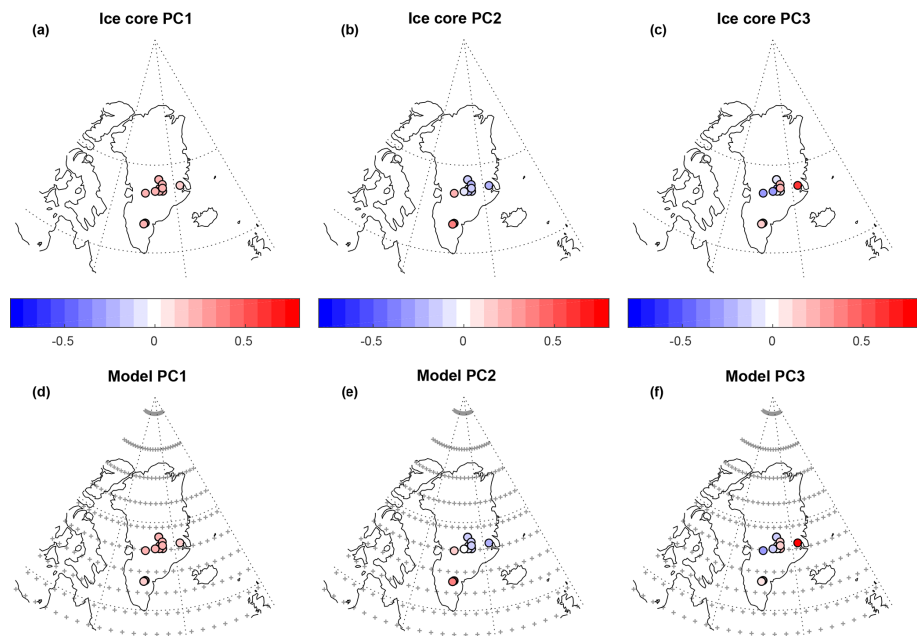


Figure 1. (a–c) Loadings of the first three PCs of ice core $\delta^{18}\text{O}$ for winter using 19 cores. Panels (d)–(f) are the same as (a)–(c) but for modeled precipitation weighted $\delta^{18}\text{O}$ for November–April at the sites of the 19 ice cores. Results for summer and annual data are very similar (not shown). In (d)–(f) the crosses mark the model grid showing the horizontal model resolution of $3.75^\circ \times 3.75^\circ$.

Table 2. Reconstructions featured in this study. A total of 12 reconstructions are done using six data sets – e.g., both the reconstructions for JJA (summer) and sum50 use the same ice core data representing the summer season May–October but targeting the differently defined summer seasons by extracting either JJA or May–October from the model output. The number of ensemble members (no. ens.) are given in parenthesis for each set of seasons. The winter reconstruction for DJF (winter) using eight ice cores covering 1241–1970 is published in Sjolte et al. (2018).

Data set and time span	19 cores, 1777–1970	Eight cores, 1241–1970
Seasons (no. ens.)	JJA/sum50 (31)	JJA/sum50 (39)
Seasons (no. ens.)	DJF/win50 (34)	DJF/win50 (39)
Seasons (no. ens.)	DJF/win100 (33)	DJF/win100 (39)

assessing modes of variability. 20CR has well-known biases (Reeves Eyre and Zeng, 2017), and care should be taken when performing detailed analysis using this data set. In addition to the evaluation using 20CR, we compare the reconstructions to the southwest Greenland temperature data compiled by Vinther et al. (2006), which is continuous for 1874–1970, as well as data from Stykkishólmur, Iceland, which covers 1830–1970 (Jónsson, 1989). These data are the longest-running instrumental temperature data available relatively close to the ice core sites used here. Finally, the station-based record of the North Atlantic Oscillation (NAO) by Jones et al. (1997) is used for evaluating the reconstructed NAO for the period 1824–1970.

We follow the convention of using the term principal components (PCs) for the time series of the main modes of variability, while using the term EOFs for the spatial patterns of the modes. The method of Ebisuzaki (1997) is used to calcu-

late the significance when correlating filtered time series in order to take autocorrelation into account.

3 Methods

3.1 Selection of model analogues based on ice core data

We use the reconstruction method of Sjolte et al. (2018) to produce a number of reconstructions of different length, different definitions of the seasons as well as varying the number of proxy records in the data set. The reconstruction method can be classified as assimilation of proxy data using the analogue method with a fixed model ensemble. This method identifies analogues, i.e., years, in a climate model simulation most closely matching the annual or season spatial pattern in a set of proxy data. In order to capture the

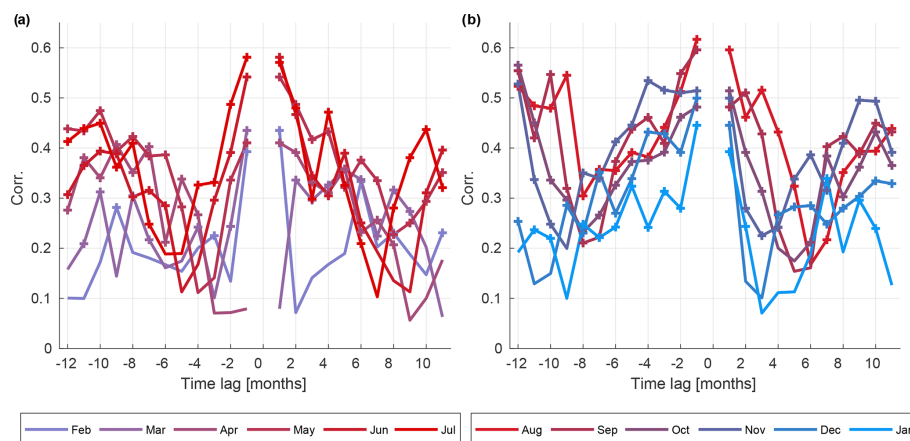


Figure 2. Autocorrelation analysis for PC1 of monthly 20CR SLP (1851–2010). Panel (a) shows results for February–July, and (b) shows August–January. The symbol “+” indicates significant correlations ($p < 0.01$).

characteristic regional variability in Greenland $\delta^{18}\text{O}$ – and to smooth out the noisy signal of individual ice cores – the matching of the model output is done using EOFs. Conventionally, proxy data need to be calibrated to a given climate variable, e.g., temperature, in order to be compared to a climate model. The use of an isotope-enabled climate model makes it possible match the proxy data with modeled patterns without calibration, since the proxy itself is included in the model output. This important feature of the method means that we include the governing processes of the variability in the proxy data, capturing the integrative nature of isotope proxies and the information that lies therein (see Introduction). The work flow of the reconstruction is to (i) calculate the PCs from the respective covariance matrix of the ice core $\delta^{18}\text{O}$ ($\text{PC}_{\text{icecore}}$) and modeled $\delta^{18}\text{O}$ (PC_{model}), retaining the first three PCs, and evaluate the modeled patterns for a given model year (t') against the ice core patterns (Fig. 1) for each proxy year (t) using Eq. (1); (ii) sort the model simulation by comparing the isotope patterns each year of the model simulation to the isotope patterns each year of the ice core data, using the normalized PCs to achieve equal weighting for the regional variability; and (iii) define the best-matching model years as ensemble member one, the second-best-matching years as ensemble member two, and so on, and test how many ensemble members to retain ($p < 0.01$) by calculating the chi-squared statistic between the modeled and the ice core PCs; and (iv) extract the climate field variables from the selected model ensemble and calculate the ensemble mean, which comprises the climate reconstruction.

$$\chi^2_{\text{Match-IC}}(t) = \frac{1}{3} \sum_{k=1}^3 (\text{PC}(k, t')_{\text{model}} - \text{PC}(k, t)_{\text{icecore}})^2 \quad (1)$$

The number of ensemble members (see Table 2) depends on the degrees of freedom, i.e., the length of the reconstruction, and how many closely matched model analogues are found. In order to assess the quality of the matching exercise, we

extract the ensemble mean reconstructed $\delta^{18}\text{O}$ at the ice core sites and correlate it against the ice core $\delta^{18}\text{O}$. This tests if matching the modeled PCs to the ice core PCs captures the variability in the original ice core data. The performance is similar for summer, winter and annual data, and the signal of the ice core data is well captured, with correlations ranging from 0.4 to 0.8 (Fig. S2). The highest correlations are seen for sites with multiple ice cores and high accumulation rate, both of which reduce noise. In summary, the reconstructed $\delta^{18}\text{O}$ captures the regional variability in the ice core data well based on matching the normalized PC_{model} and $\text{PC}_{\text{icecore}}$.

As outlined in the Introduction, the definition of the seasons or year is an important parameter for the reconstruction. This applies both in terms of the seasonality of the proxy data and the target season of the reconstruction. Following the study of Vinther et al. (2010), we will use the definitions of summer as May–October (sum50), winter as November–April (win50) and winter-centered annual mean August–July (win100) for the ice core data. These definitions will also be applied to the target seasons of the reconstructions, as well as the widely used definitions of summer (JJA) and winter (DJF). We investigate the seasonal and annual variability using these different definitions with two data sets for *short* (1777–1970; 19 ice cores) and *long* (1241–1970; eight ice cores) reconstructions, resulting in a total of 12 reconstructions, where one for DJF covering 1241–1970 was published by Sjolte et al. (2018) (see Table 2).

3.2 Constraining summer reconstructions using tree-ring data

For the summer season we test incorporating tree-ring data to further constrain the reconstruction. We choose a simple approach of incorporating the data, which can serve as a pilot study for further tests of adding more data to the reconstruction. For the test we sort the preselected 39 ensemble members ($t'_{\text{IC-ENS}}$) based on the ice core data (Table 2) us-

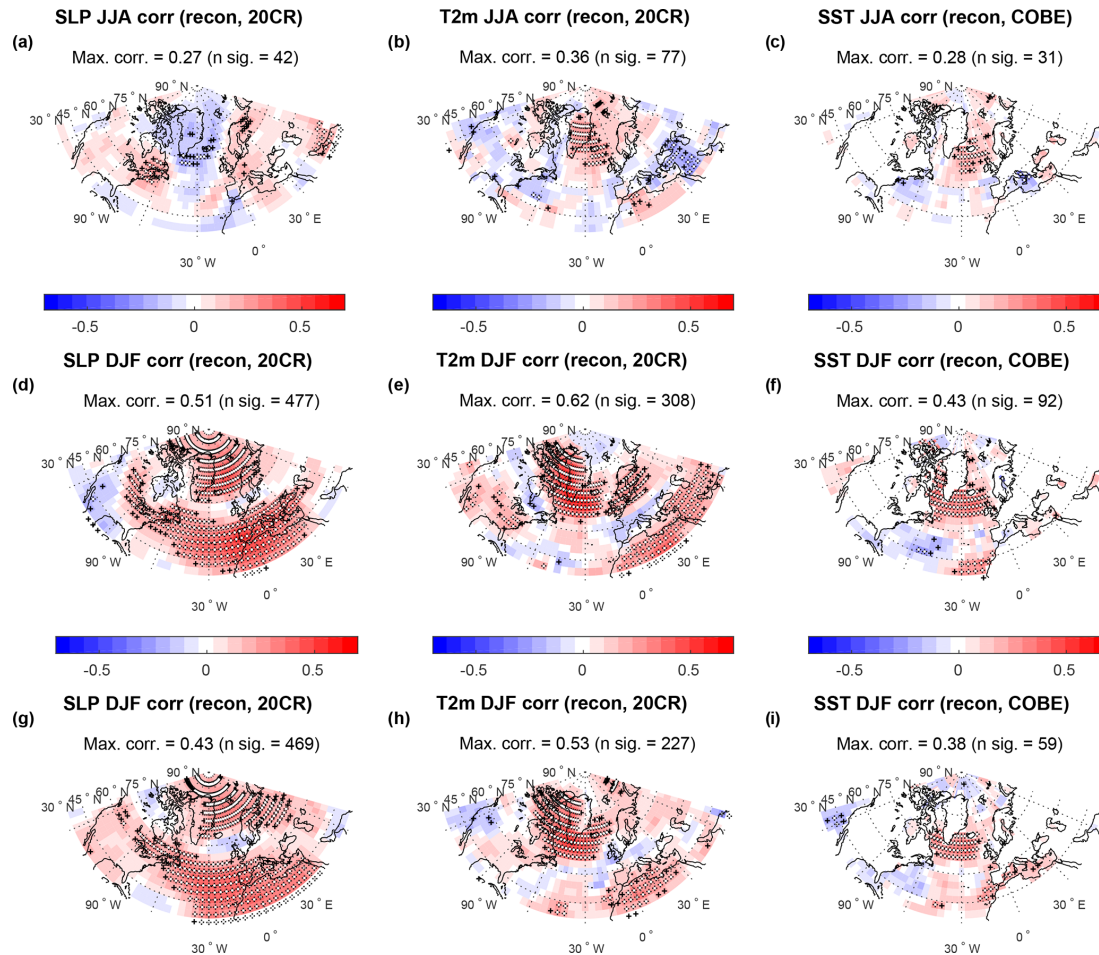


Figure 3. (a–c) Correlation between reconstructed (eight ice cores) and reanalysis SLP, T2m and COBE SST for JJA. The reanalysis data have been interpolated to the model grid ($3.75^\circ \times 3.75^\circ$). Black markers indicate $p < 0.05$, and white markers indicate $p < 0.025$. Also indicated is the maximum correlation (Max. corr.) and the number of significantly correlated grid points ($n \text{ sig.}$) ($p < 0.05$). Panels (d)–(f) are the same as (a)–(c) but for DJF. Panels (g)–(i) are the same as (a)–(c) but for DJF reconstructed from the winter-centered annual mean ice core data. Figure S7 shows corresponding figures for the reconstructions using 19 ice cores.

ing a chi-squared fit of normalized modeled temperature at the eight tree-ring sites (T_{model}) against the normalized tree-ring data (T_{trees}) (see Eq. 2).

$$\chi_{\text{Match-TR}}^2(t) = \frac{1}{8} \sum_{k=1}^8 (T(k, t'_{\text{IC-ENS}})_{\text{model}} - T(k, t)_{\text{trees}})^2 \quad (2)$$

The fit is done using the JJA temperature from the model, which are the best months to use with respect to seasonal sensitivity for these eight tree-ring records (Wilson et al., 2016). In a next step we test the ensemble mean temperature reconstruction against the time series of the tree-ring data at each site, by calculating the correlation with the tree-ring data while increasing the number of ensemble members from 1 to 39 (Fig. S3). Although a chi-squared test of the fit of the reconstructed temperature shows that including 24 ensemble members provides a good fit ($p < 0.01$), the correlation decreases quite rapidly when including more ensemble mem-

bers, and we choose to include only 20. With this ensemble we capture the variability in the tree-ring data relatively well for the whole period of the reconstruction (Fig. S4). The correlation goes to zero when including all of the 39 ensemble members, indicating that without the tree-ring data the reconstruction using only the eight ice cores and the model has no predictive skill of the summer temperature in Europe.

4 Results

4.1 The seasonal variability in observations and when combining proxy data and model output

In the Introduction we mentioned seasonality, definition of seasons and shifts in circulation patterns as potential limiting factors for the skill of climate field reconstructions. In general, seasonal dependency on climate variables, temporal resolution and the precision of the chronology of proxy

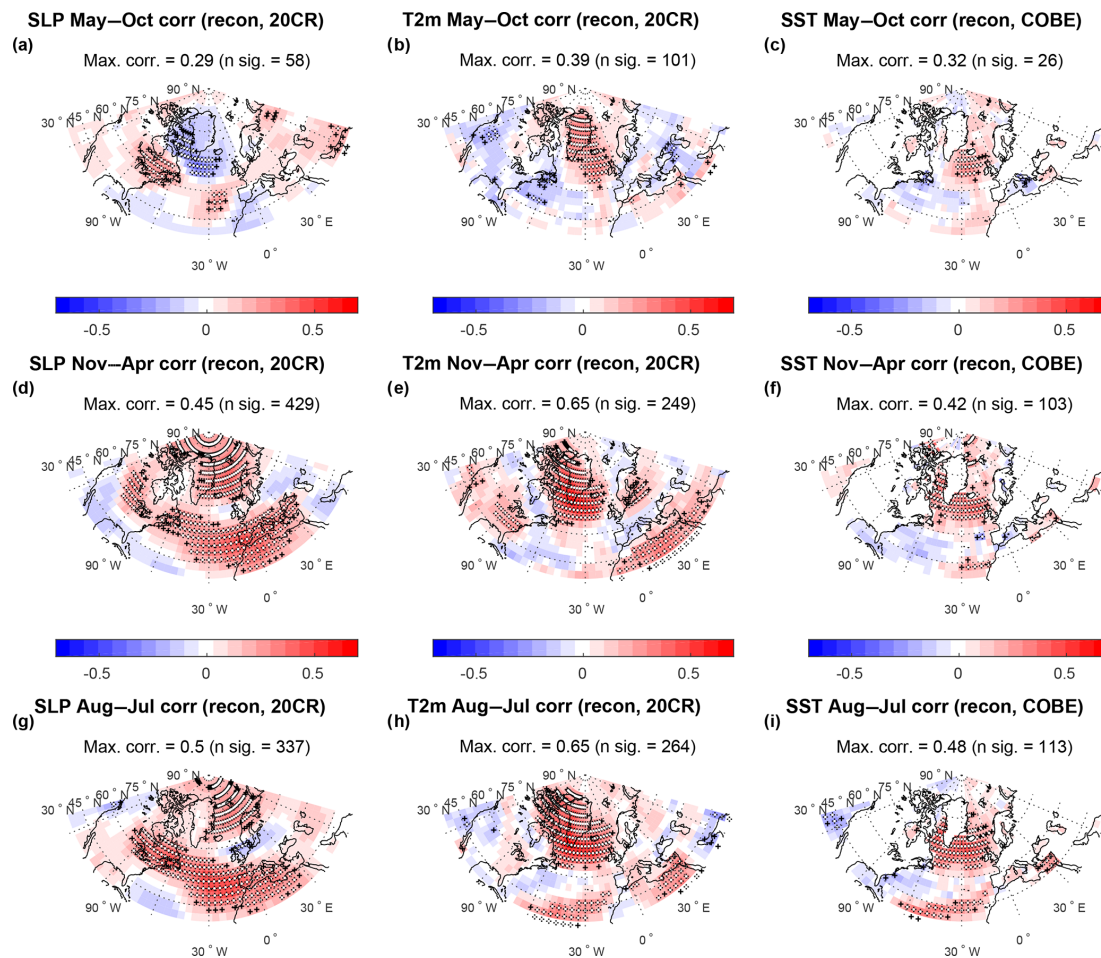


Figure 4. (a–c) Correlation between reconstructed (eight ice cores) and reanalysis SLP, T2m and COBE SST for sum50 (May–October). The reanalysis data have been interpolated to the model grid ($3.75^\circ \times 3.75^\circ$). Black markers indicate $p < 0.05$, and white markers indicate $p < 0.025$. Also indicated is the maximum correlation (Max. corr.) and the number of significantly correlated grid points ($n \text{ sig.}$) ($p < 0.05$). Panels (d)–(f) are the same as (a)–(c) but for win50 (November–April). Panels (g)–(i) are the same as (a)–(c) but for the winter-centered annual mean (win100, August–July). Figure S8 shows corresponding figures for the reconstructions using 19 ice cores.

records sets a limit on the temporal resolution of climate field reconstructions. Seasonal resolution is likely the highest possible resolution which can be attained due to these different factors. The subseasonal autocorrelation structure of atmospheric variability is a key factor in how well seasonal proxy data can represent climate variability. This can be illustrated by investigating the monthly autocorrelation during the year of the first leading mode of sea level pressure in the North Atlantic region, the NAO. We found that, for example, the second and third leading modes are too dissimilar between summer, autumn, winter and spring to allow a meaningful study of the monthly autocorrelation of these modes, as they simply represent different teleconnection patterns during each season. Figure 2 shows the monthly autocorrelation of each month of the PC-based NAO calculated from the 20CR. These figures show that during the cold season the NAO has the weakest autocorrelation with

other months, as well as weaker year-to-year autocorrelation compared to summer. While the lower autocorrelation during winter shows stochastic nature of the variability, it is also during winter that the NAO variability is the most vigorous. Thus, the portion of a given climate signal that can be reconstructed is a balance of what is recorded in the proxy at a certain resolution, as well as the strength and autocorrelation of the signal sampled at this resolution. It is noteworthy that Fig. 2 also illustrates that targeting the calendar year in a reconstruction (or any sort of analysis) splits up the variability midwinter and mixes the variability in two consecutive winters that have little variability in common. This is the motivation for using the definition of winter-centered annual mean for the annual data in this study.

Vinther et al. (2010) tested the ice core data used in this study using correlation with observed temperature, leading to the division of the in seasons using the definition of

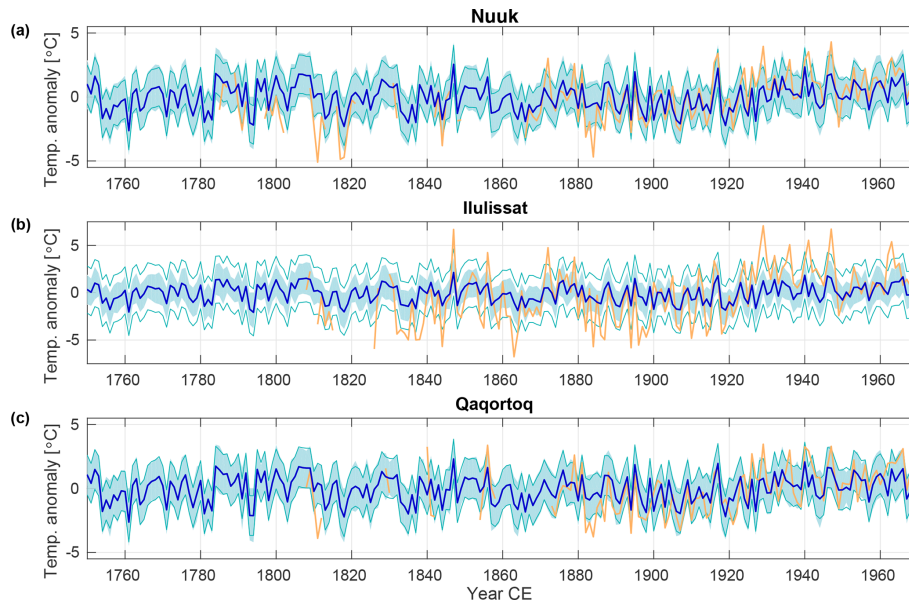


Figure 5. Time series of November–April (sum50) temperature for observed (yellow) and reconstructed ensemble mean (sum50, eight ice cores) (dark blue) from (a) Nuuk, (b) Ilulissat and (c) Qaqortoq. Light-blue shading is the 1σ spread of the reconstructed temperature, and the green lines indicate the RMSE between the observed and reconstructed temperature.

sum50, win50 and win100 as outlined in Sect. 3. Due to the changes in the patterns and variability in the circulation modes from summer to winter we furthermore test the seasonality in terms of circulation modes. We do this by performing monthly reconstructions for pressure and correlating the time series of the corresponding main modes of circulation against that of the modes of the 20CR. This is done using the same method as for the seasonal reconstructions but only picking individual months from the matching year of the model simulation. We do not suggest that it is feasible to reconstruct climate on monthly timescales using seasonal ice core data. This exercise is purely for testing purposes. The monthly reconstructions are done for each data set (sum50, win50, win100; for eight ice cores and 19 ice cores) for the months that each data set is assumed to cover, e.g., May–October for sum50. The overall results show that the different reconstructed surface pressure modes, as represented by the first three PCs, do not peak in skill during the same months (Figs. S5 and S6). For example, for win50 PC1 has highest skill for February–April, while the skill for PC2 peaks January–February. This type of behavior is repeated for the sum50 and win100 data sets. The differentiated seasonality in the skill of the reconstructed modes can originate from (i) the sensitivity of the Greenland $\delta^{18}\text{O}$ to different modes; (ii) the changes in circulation modes during the season; (iii) the autocorrelation structure of circulation, as discussed above; and (iv) model biases in circulation modes – and combinations of these influences. The difference in the reconstructions using eight ice cores and 19 ice cores, respectively, is mainly seen for win100, where more monthly

reconstructions show significant skill across the year when using more ice cores in the reconstruction. Furthermore, the monthly skill for the win100 data set indicates that it is feasible to reconstruct the winter circulation (e.g., DJF). This test suggests that in order to get the highest average skill possible for all modes during winter the reconstruction should target DJF, while for summer the full span of the season (May–October) is likely better, also taking into account the higher monthly autocorrelation during the warm season. The EOF patterns of surface pressure will be discussed further in Sect. 4.2.2.

4.2 Evaluation of reconstructions

In the following sections we evaluate and compare the reconstructions using different methods. We start with point-by-point correlation maps for the North Atlantic sector of the reconstructions to 20CR SLP and T2m, as well as the COBE SSTs. This is a general evaluation in terms of spatial coverage and skill of the reconstructions. We also include a comparison to the longest instrumental records of temperature from Greenland and Iceland. Next we evaluate the skill of the reconstructions in terms of atmospheric circulation modes. In the final part of the evaluation we investigate to which extent the main patterns of North Atlantic SSTs and their variability can be reconstructed using the method of this study. We would like to emphasize that none of these reconstructions have been calibrated to observations. Instead, the model provides us directly with the physical variables of SLP, T2m and SST from the model years where modeled and measured

Table 3. Summary of maximum correlations (Max. corr.) and number of grid points with significant correlation (n sig.; $p < 0.05$) from Figs. 3 and 4, as well as Figs. S7, S8 and S9. Results in columns sum50* and JJA* are for the reconstructions using tree-ring data.

	JJA		JJA*		Sum50		Sum50*		DJF		DJF _{win100}		DJF _{win100}		Win50		Win50		Win100		Win100	
	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long
CorrSLP	0.32	0.27	0.33	0.33	0.31	0.29	0.26	0.26	0.48	0.51	0.39	0.43	0.43	0.48	0.45	0.43	0.43	0.43	0.43	0.50	0.43	0.50
n_{SLP}	86	42	120	120	69	58	61	61	433	477	511	469	469	451	429	310	310	310	310	337	310	337
CorrT _{2m}	0.40	0.36	0.63	0.63	0.38	0.39	0.59	0.59	0.63	0.62	0.56	0.53	0.53	0.62	0.65	0.62	0.62	0.62	0.62	0.65	0.62	0.65
n_{T2m}	116	77	247	247	135	101	266	266	283	308	218	227	227	311	249	225	225	225	225	264	225	264
CorrSST	0.31	0.28	0.50	0.50	0.36	0.32	0.49	0.49	0.41	0.43	0.42	0.38	0.38	0.43	0.42	0.48	0.48	0.48	0.48	0.48	0.48	0.48
n_{SST}	71	31	116	116	56	26	138	138	85	92	95	59	59	95	103	124	124	124	124	113	124	113

$\delta^{18}\text{O}$ patterns match. The evaluation of these reconstructions are thus done using completely independent data sets.

4.2.1 Reconstructed temperature and sea level pressure

Investigating the results for correlations and the spatial patterns of skill for SLP, T2m and SSTs reveals a complex interplay of factors influencing the reconstructions for different seasons, as well as how the different definition of seasons influence the skill. Reconstructions for the summer season show the least skill but perform better using the extended definition of the target season (May–October) (Fig. 4) rather than JJA (Fig. 3). The summer reconstruction also appears to benefit the most from including 19 ice cores rather than eight. This can be seen also be seen from Table 3, which summarizes the maximum correlation and number of significantly correlated grid points compared to 20CR for all reconstructions in this study. Including more cores and using the extended season likely reduces noise in the reconstruction. Using the extended season also smooths out the variability in the 20CR data, which can partly account for the higher skill of the short sum50 reconstruction for summer. The summer reconstructions using eight ice cores show no significant skill for Europe, which is in line with the correlation analysis with European tree-ring data (see Sect. 3.2). However, the evaluation of the summer reconstructions using 19 ice cores shows patches of significant correlation in Europe. The reconstructions for winter show the highest skill of the reconstructions, in line with the findings of Vinther et al. (2010) that $\delta^{18}\text{O}$ is found to be a more efficient climate proxy during winter (Sjolte et al., 2011, 2014). This can partly be due to the climate variability in extratropical North Atlantic region being most vigorous during winter causing a large signal-to-noise ratio in $\delta^{18}\text{O}$ records with respect to their ability to record circulation changes. All of the these factors contribute to better reconstructions for winter compared to summer, both in terms of spatial skill and strength of correlation with 20CR. This includes significant temperature skill in northern Europe, which is probably due to the reconstruction capturing the main modes of SLP. We will return to this topic in Sect. 4.2.2. As opposed to summer, the winter reconstructions for DJF performs better, rather than the extended season November–April. This is probably due to the migration of circulation patterns and low autocorrelation of atmospheric circulation during winter as discussed in Sect. 4.1.

One of the questions of this study is about the use of annual data for reconstructions of climate and atmospheric circulation. For the reconstructions targeting the winter-centered annual mean (win100) the skill and patterns of correlation are reminiscent of that of the winter reconstructions, although clearly with less areal coverage of significant correlation for SLP. We interpret this as being due the migration of the circulation patterns with the seasons, as discussed above. How-

Table 4. Correlation between reconstructed and observed temperature for Greenland coastal stations (1874–1970) and the Icelandic station, Stykkishólmur (1831–1970). Bold marks $p < 0.05$; the symbol * marks $p < 0.10$. The low-pass filter is a decadal FFT (fast Fourier transform) filter.

19 ice cores	Sum50	Low pass	Win50	Low pass	Win100	Low pass
Stykkishólmur	0.32	0.48	0.33	0.44	0.25	0.17
Nuuk	0.19	0.33	0.58	0.35	0.52	0.47
Ilulissat	0.18*	0.35	0.53	0.48	0.45	0.40
Qaqortoq	0.24	0.34	0.56	0.40*	0.53	0.44*
SWG index	0.22	0.38*	0.59	0.42*	0.52	0.44
Eight ice cores	Sum50	Low pass	Win50	Low pass	Win100	Low pass
Stykkishólmur	0.33	0.38*	0.33	0.53	0.28	0.37*
Nuuk	0.24	0.36*	0.60	0.53	0.58	0.56
Ilulissat	0.19	0.39*	0.56	0.60	0.50	0.52
Qaqortoq	0.27	0.45	0.60	0.65	0.59	0.59
SWG index	0.26	0.44	0.63	0.63	0.58	0.56
19 ice cores	JJA	Low pass	DJF	Low pass	DJF _{win100}	Low pass
Stykkishólmur	0.27	0.58	0.35	0.41	0.27	0.20
Nuuk	0.10	−0.03	0.56	0.49	0.45	0.45
Ilulissat	0.09	0.27	0.56	0.67	0.41	0.42
Qaqortoq	0.22	0.43*	0.52	0.49	0.48	0.41*
SWG index	0.15	0.29	0.58	0.57	0.47	0.45
Eight ice cores	JJA	Low pass	DJF	Low pass	DJF _{win100}	Low pass
Stykkishólmur	0.29	0.50	0.27	0.41	0.29	0.34
Nuuk	0.18*	0.14	0.57	0.53	0.49	0.47
Ilulissat	0.07	0.29	0.58	0.70	0.47	0.55
Qaqortoq	0.25	0.43*	0.57	0.62	0.52	0.50
SWG index	0.21	0.35	0.61	0.67	0.52	0.52

ever, for SSTs the win100 reconstruction shows the highest spatial skill of all the reconstructions, including better capturing low-latitude variability, with the correlation pattern being reminiscent of the spatial pattern of Atlantic Multidecadal Oscillation (AMO)-type variability. As with the extended summer season, part of the increase in skill for the win100 SST reconstruction could also originate from a smoother signal for annual data – in both observations and reconstruction, where some of the noise is reduced compared to seasonal data, but some of the signal is also lost. Targeting the winter season (DJF) using the winter-centered annual data results in a clear gain in skill for SLP, while the skill for SST is somewhat reduced, although retaining the overall correlation pattern of the winter-centered annual mean reconstruction. This indicates that it is feasible to reconstruct winter variability from annual data, if the definition of the winter-centered annual mean is used for the proxy data. Seasonal $\delta^{18}\text{O}$ data are increasingly sparse going back in time, and using winter-centered annual mean data could be an alternative for reconstructing winter variability beyond the reach of seasonal $\delta^{18}\text{O}$ data when seasonality in the ice can still be defined from e.g., aerosol records.

To further assess the skill of the reconstructed temperature we compare to data from three stations on the Greenland coast and one Icelandic station. Vinther et al. (2010) showed that the first principal component (PC1) of Greenland isotope data (20 cores) has strong correlation ($r = 0.71$) with the stacked Greenland coastal data (southwest Greenland temperature, SWG, index) during winter (November–April), while PC1 of the isotope data for summer is most strongly correlated with data from Iceland ($r = 0.55$) (May–October). Here we compare the reconstructed site temperature both to data from each of the stations and to the SWG index. The highest correlations are found for the eight-core win50 reconstruction at Nuuk and Qaqortoq with a correlation of 0.6 at both sites (Fig. 5 and Table 4). It is also for this reconstruction we find the highest correlation of 0.63 with the SWG index. While the correlation for Ilulissat is similar to the correlation for Nuuk and Qaqortoq, the observed higher amplitude is not captured by the reconstruction, which is probably due to subgrid variability resolved neither by reconstruction nor the model. The 19-core reconstructions have slightly lower correlations with the Greenland temperature data. This could be due to a weighting of the variability more to the east, as most of the additional cores in the shorter reconstructions are

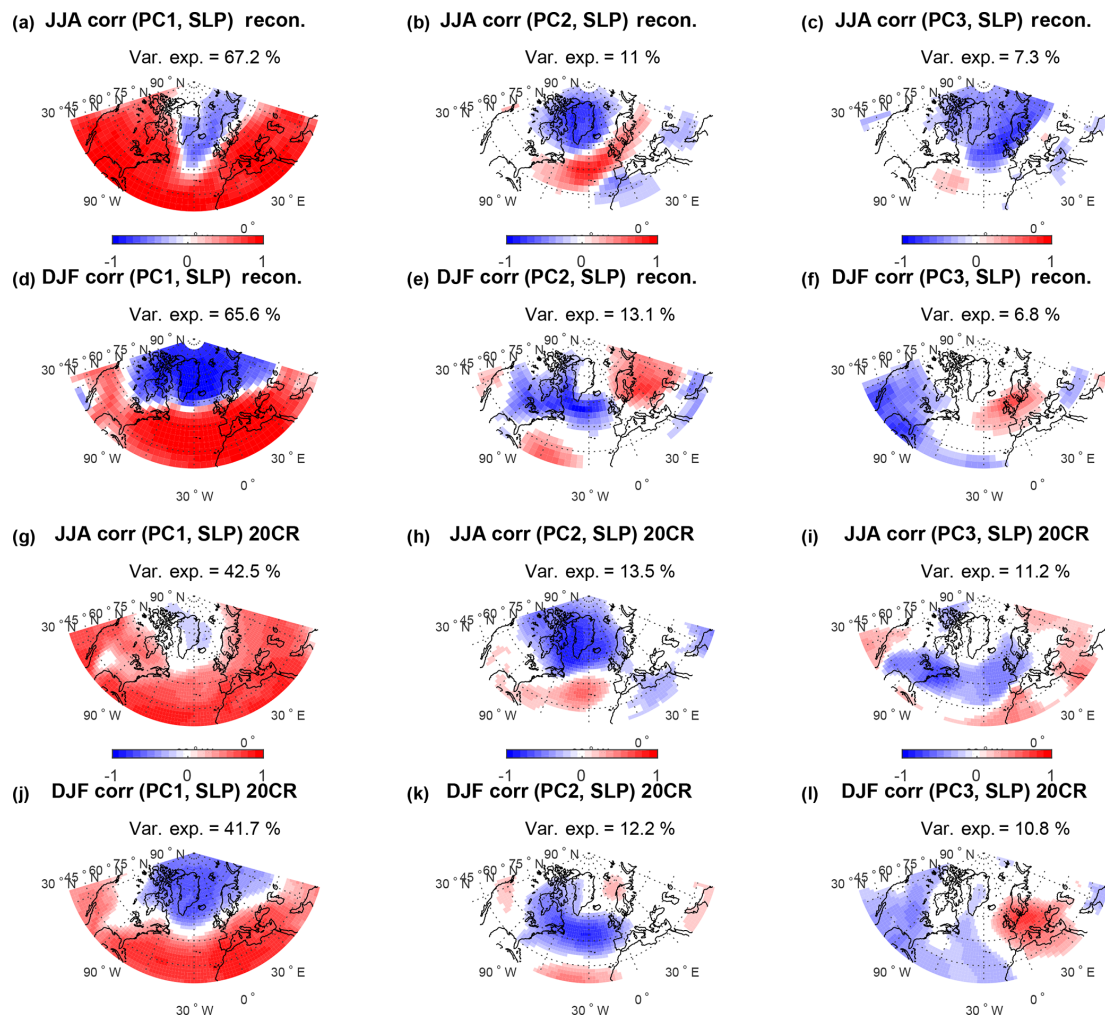


Figure 6. (a–c) Regression of the first three reconstructed PCs of SLP on reconstructed (eight ice cores) JJA SLP, which corresponds to the reconstructed EOF patterns. Panels (d)–(f) are the same as (a)–(c) but for DJF. These plots, with the addition of the plots for DJF reconstructed (eight ice cores) from the winter-centered annual mean ice core data, are shown in Fig. S10, as well as corresponding plots for sum50, win50 and win100 shown in Fig. S11. (g)–(i) Regression of the first three 20CR PCs of SLP on 20CR JJA SLP, which corresponds to the EOF patterns. Panels (j)–(l) are the same as (g)–(i) but for DJF. These plots for 20CR data are also shown in Fig. S12, as well as corresponding plots for sum50, win50 and win100 shown in Fig. S13. The time period for all plots is 1851–1970. Data only shown for $p < 0.05$.

to the east of the ice divide. For the summer reconstructions, the correlations with the Greenland station data are below 0.3. However, the eight-core sum50 reconstruction captures a substantial part of the longer-term variability with a correlation of 0.44 to the decadal filtered SWG index. With respect to the definition of the winter season, the DJF reconstructions appear to better capture the long-term variability, with slightly higher correlation for the filtered data compared to the win50 reconstructions. The win100 and the win100 DJF reconstructions both show only slightly lower correlations than the win50 and DJF reconstructions, indicating that for temperature alone the seasonal data are less crucial than for reconstruction SLP, at least when comparing locally to the Greenland coastal data.

The correlations with the Icelandic temperature data show correlations around 0.3 for all reconstructions, with most of the summer reconstructions showing higher correlations for long-term variability compared to the winter reconstructions. This indicates a similar behavior as for the ice core PC1 correlation with respect to the winter data responding more to the western Greenland temperature and the summer data having better coherency with Icelandic data. The predominance of the summer signal east of Greenland also results in the reconstructions based on the winter-centered annual mean not having very high skill for Icelandic temperatures, at least for the long-term variability.

Comparing the summer reconstructions including tree-ring data with the 20CR, we find that the skill for SLP, T2m

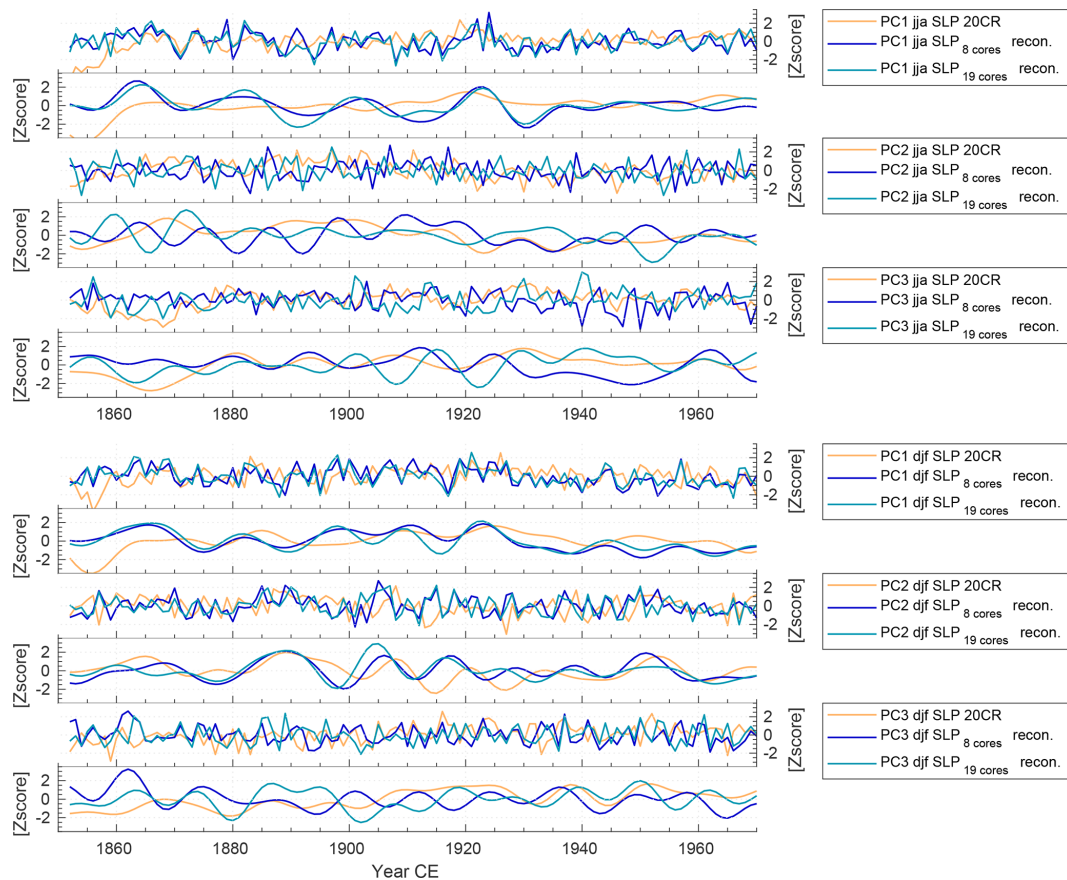


Figure 7. Time series of reconstructed PC1, PC2 and PC3 of SLP using eight ice cores (dark blue) and 19 ice cores (light blue) compared to PC1, PC2 and PC3 of 20CR SLP (yellow). Smoothed curves are using a decadal FFT filter. Top six plots are for JJA, and bottom six plots are for DJF.

and SST has increased considerably compared to the summer reconstructions only using eight ice cores (Table 3 and Fig. S9). The skill is particularly improved for temperature in the eastern sector of the domain, while the skill for SLP is still low near Greenland, although the skill has clearly increased over northern Europe for JJA.

4.2.2 Main modes of atmospheric variability

Sjolte et al. (2018) showed that the winter variability in the first two PCs of the SLP in the North Atlantic region could be reconstructed with good skill using the analogue method based on eight ice cores. Here we evaluate all the different reconstructions of this study for the first three PCs, including the spatial patterns of the loading of the PCs (EOFs). For the DJF and win50 reconstructions, EOF1, 2 and 3 all qualitatively match that of the 20CR (Fig. 6). The reconstructed EOF patterns for SLP are very similar for the reconstructions using eight and 19 ice cores, respectively, and we only show the patterns for the reconstructions using eight ice cores. There are some indications that EOF2 of the reconstructions summarizes some of the variability assigned to

EOF3 of the 20CR as also discussed by Sjolte et al. (2018). For summer the reconstructed EOFs capture many of the same features of the 20CR but less clearly than for the winter reconstructions. For example, the reconstructed JJA pattern for EOF1 shows differences in 20CR south of Greenland (Fig. 6), which probably partly explains the low skill for summer SLP in this region shown in Sect. 4.2.1. The origin of this problem is probably a bias for large-scale summer variability in the ECHAM5/MPIOM model (Jungclaes et al., 2006). This means that the main modes of the original model simulation (not shown) do not correspond to the main observed modes, except for winter NAO, which the model captures. It is only after matching the model output to the proxy data that the main modes align with the observed patterns.

The maps of the EOF patterns illustrate the point made earlier about the differences in the modes of SLP variability from season to season. The patterns change not only from summer to winter but also depending of the definition of the season, e.g., JJA versus May–October (Fig. S11). Furthermore, the EOFs of the winter-centered annual mean appear as mixtures of summer and winter variability, carrying most

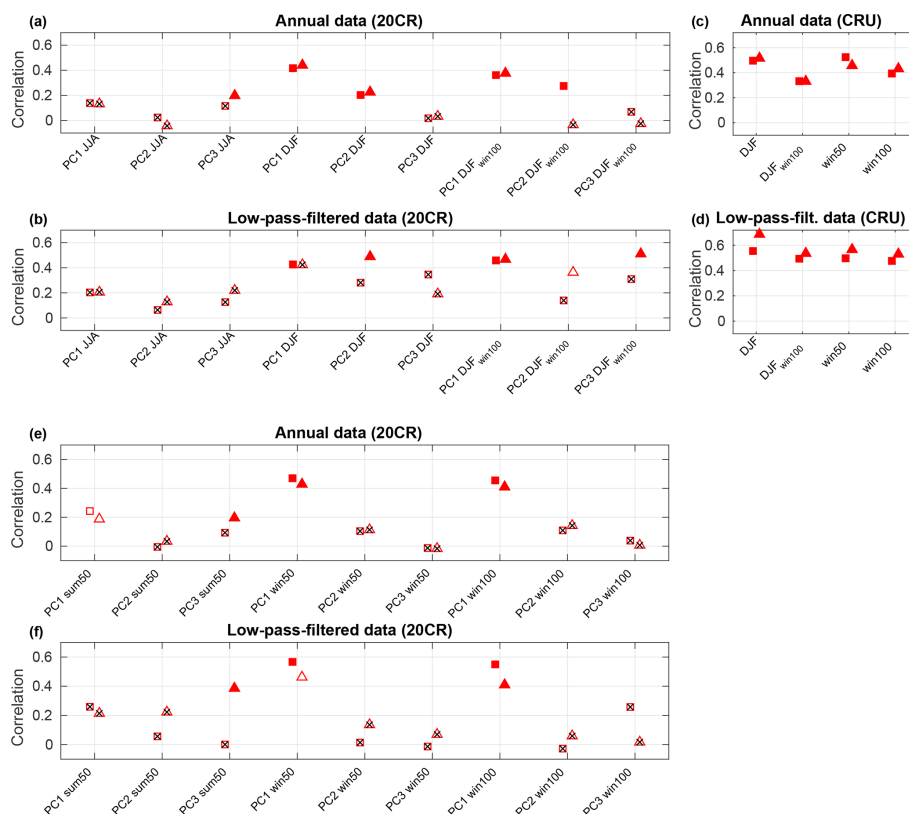


Figure 8. (a, b, e, f) Correlation analysis for reconstructed PC1, PC2 and PC3 of SLP using eight ice cores (triangles) and 19 ice cores (squares) correlated with PC1, PC2 and PC3 of 20CR SLP covering 1851–1970; (c, d) correlation analysis for reconstructed PC1 SLP using eight ice cores (triangles) and 19 ice cores (squares) correlated with station-based NAO covering 1824–1970. The station-based NAO is only valid for winter and annual data due to the seasonal shift in the centers of action. Open markers indicate significance of $p < 0.1$, and full markers indicate $p < 0.05$, while crossed-out markers indicate $p > 0.1$.

likeness to the winter patterns, again showing the problem of using the annual mean SLP as target for reconstructions.

Common for all the different reconstructions is that they all assign more variability to EOF1 and less to EOF3 compared to 20CR, while EOF2 is fairly similar to 20CR in terms of the explained variance. This could be due to sole use of Greenland ice core data, which could skew variability to be dominated more by NAO-type variability. For DJF the model simulation itself does not have a high bias in the explained variance in NAO-type variability.

From the time series of EOFs (PCs) it is evident that the reconstructions have realistic amplitudes of the year-to-year variability (Fig. 7). In other words, the spectrum of the reconstructions are similar to actual weather variability as also found for the DJF reconstruction by Sjolte et al. (2018). Correlating the reconstructed PCs to that of the 20CR (see Fig. 8) shows that (i) the variability in PC1 is well captured by the winter and annual data; (ii) only the win50 DJF reconstruction has skill for PC2; (iii) the summer reconstructions have some skill for PC3; and (iv) in some instances the decadal filtered data capture a significant part of the 20CR variability, even with no correlation for annual data – e.g., PC2 and

PC3 of DJF win100 (eight cores). The very low values 1851–1860 in the 20CR PC1 is possibly a bias in the reanalysis and is not seen in the Jones et al. (1997) NAO time series (not shown). Comparing the reconstructions for winter and annual data to the Jones et al. (1997) NAO results in higher correlations than for 20CR, also for the filtered data. For summer it is not meaningful to use the station-based NAO due to the shifted centers of action during summer compared to winter. As discussed in Sect. 4.2.1 the skill for SLP improves locally when including tree-ring data to constrain the summer reconstructions. However, the skill for the circulation patterns is not improved by including the tree-ring data.

4.2.3 North Atlantic sea surface temperature

The correlation maps with the COBE SSTs (Figs. 3 and 4) indicate that the reconstructions are particularly well suited to investigate the SST variations in the region 50–70° N, 70–0° W. For this purpose we define a North Atlantic SST index as the mean SST for the aforementioned area. Although the year-to-year variations in the reconstructions are somewhat noisy compared to the variations in the COBE

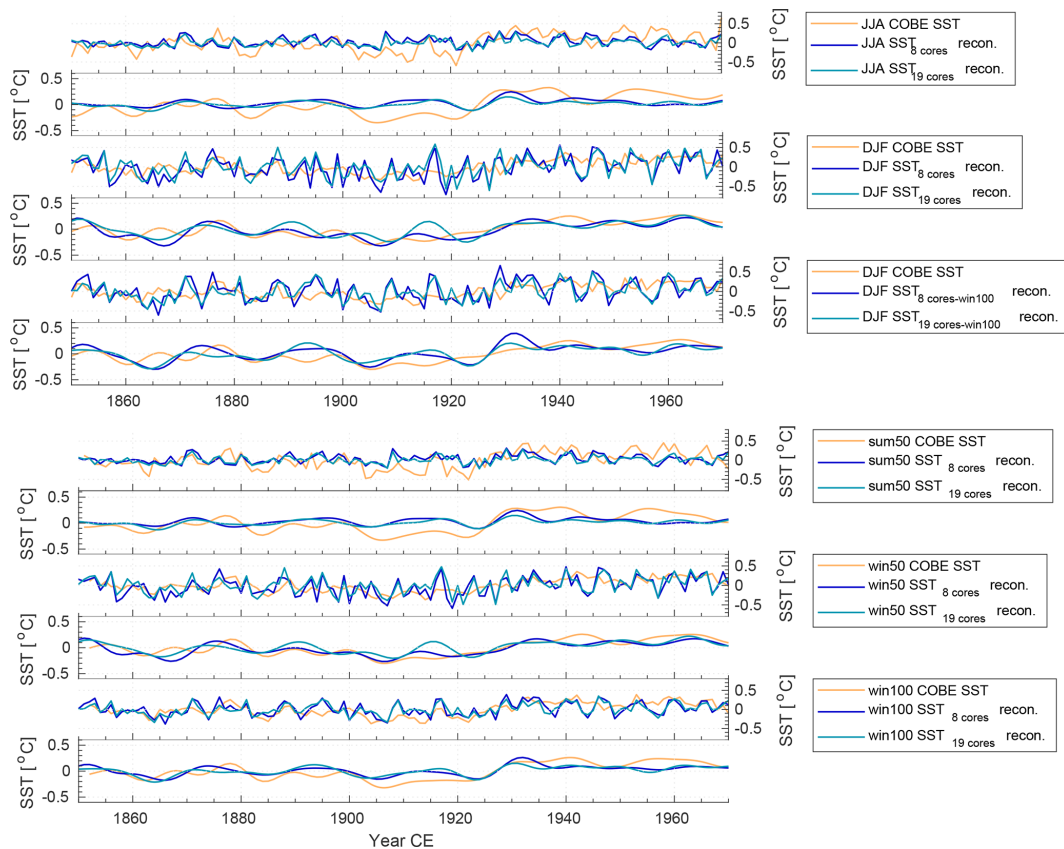


Figure 9. Time series of the North Atlantic SST index (50–70° N, 70–0° W) for reconstructions using eight ice cores (dark blue) and 19 ice cores (light blue) compared to COBE SSTs (yellow). Smoothed curves are using a decadal FFT filter. The top six plots are for JJA, DJF and DJF reconstructed using the winter-centered annual mean, while the bottom six plots are for sum50 (May–October), win50 (November–April) and the winter-centered annual mean win100 (August–July).

SSTs, the reconstructions have significant skill for all investigated seasons, most notably for winter and annual data (Fig. 9). For decadal filtered data, the win50 DJF and win50 reconstructions (eight cores) explain more than 50 % of the COBE North Atlantic SST variability ($r = 0.72$ and $r = 0.74$, respectively) (Fig. 10). While the long-term SST changes for summer are underestimated, the reconstructions of winter SST match the COBE amplitudes of the decadal–multidecadal SST variability very well. As mentioned in Sect. 4.2.1 the skill for temperature and SST is markedly improved when including the tree-ring data in the summer reconstructions. This is also seen in the higher correlations and stronger significance for the North Atlantic SST index for these reconstructions (Fig. 10).

To further investigate how much information of the North Atlantic SST variability is obtainable using this type of reconstruction, we also compared the patterns and variability in the main modes of reconstructed SSTs to that of the COBE SSTs (Fig. 11). As the skill of the reconstructions decreases with the distance from the proxy sites, we calculated the modes using data from 30–70° N for the reconstructions, while we used 0–70° N for the COBE data. Generally

the reconstructions qualitatively capture the spatial characteristics of the EOF1, 2 and 3 patterns of the COBE data, as well as the variability in the PCs (Fig. 12). Again, the match appears to be better for the winter season. The PCs of the reconstructed SSTs are correlated with the reconstructed PCs of SLP, indicating that the SST variability captured by the reconstruction is related to atmosphere–ocean interaction of the main circulation modes (not shown). EOF1 of the SSTs is also correlated with the North Atlantic SST index discussed above, and the pattern is akin to AMO-type variability associated with long-term variation in the NAO (McCarthy et al., 2015). EOF2 of the SSTs can be related to subpolar gyre-type variability connected with the frequency of the weather patterns Atlantic Ridge and Blocking (Moffa-Sanchez et al., 2014; Moreno-Chamarro et al., 2017). Only the reconstructed PC1 for winter and annual SSTs shows consistent skill compared to the COBE SSTs, although the win50 PC3 (19 cores) also has significant correlation for both annual and decadal filtered data (Fig. 13).

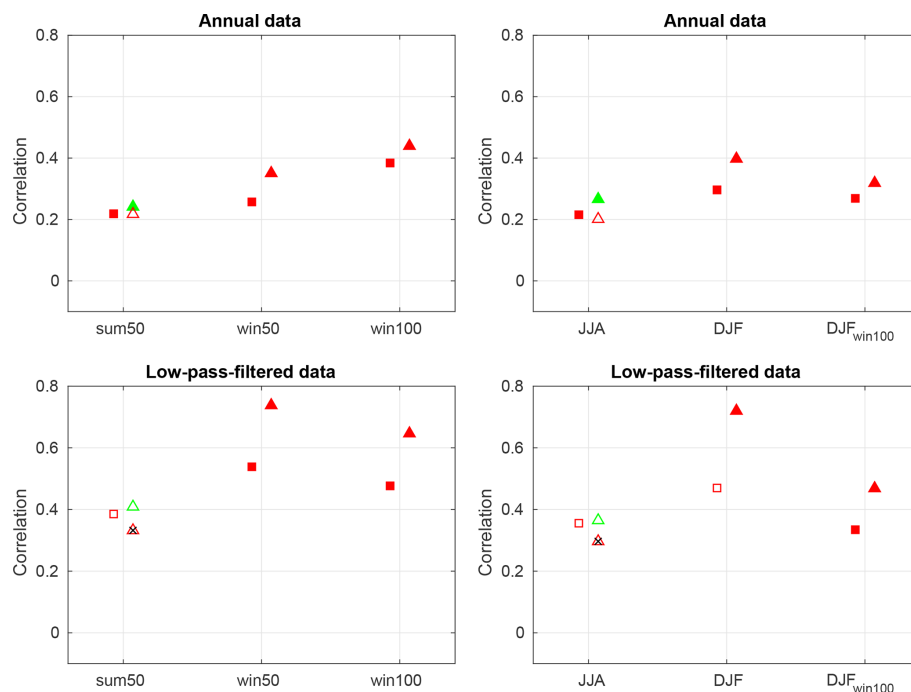


Figure 10. Correlation analysis of the North Atlantic SST index (50–70° N, 70–0° W) for reconstructions using eight ice cores (triangles) and 19 ice cores (squares) correlated with COBE SSTs covering 1851–1970. The green markers are for the reconstructions including tree-ring data. Open markers indicate significance of $p < 0.1$, and full markers indicate $p < 0.05$, while crossed-out markers indicate $p > 0.1$.

4.3 Comparison to other millennium-length reconstructions

While an exhaustive comparison to other reconstructions is beyond the scope of this study, we briefly compare our reconstructions to two other data sets. We limit ourselves to reconstructions that are based on data entirely independent from this study and also cover the span of our longest reconstructions (1241–1970). We first compare to the temperature index for central Europe by Glaser and Riemann (2009), which is based entirely on historical documentation and early instrumental data. Due to less available data in the early part of the millennium, the reconstruction by Glaser and Riemann (2009) is only in seasonal resolution prior to 1500 CE, while monthly data are available after this. Due to this change in resolution and variability, we only show the comparison to Glaser and Riemann (2009) for the period after 1500 CE. In the comparison we use our reconstructions including tree-ring data for summer (JJA, sum50), as the reconstructions relying solely on ice core data (eight ice cores) do not have skill in Europe for summer. Judging from the moving correlation, there is fairly good correspondence between our reconstructions and the temperature index of Glaser and Riemann (2009) for the period after 1600 CE, apart from a distinct spell of out-of-phase variability around 1650 CE for the summer season (Fig. 14). The correlation is most consistent for DJF, although the decadal to multidecadal variability also appears coherent for the summer season. For the period prior

to 1500 CE (no shown) little coherency is seen between our reconstructions and the temperature index of Glaser and Riemann (2009). As the temperature index of Glaser and Riemann (2009) relies on a relatively few data for the early part of their reconstruction, it is tempting to conclude that the loss of correlation is due to this, as our reconstructions are produced with the same number of records and same method throughout the reconstructions. Despite this, the comparison provides support for the validity of our seasonal temperature reconstructions extending further back than the comparison to reanalysis data.

In a second comparison we include the recent DJF NAO reconstruction by Cook et al. (2019), which is based on drought data from tree rings. For reference we also include the comparison to the model-constrained NAO reconstruction by Ortega et al. (2015) also shown in Sjolte et al. (2018), although this reconstruction is also partly based on Greenland ice core data. From the moving correlation, there is little correspondence between our NAO reconstruction and that of Cook et al. (2019) prior to the instrumental record (Fig. 15). Unlike our method, the method of Cook et al. (2019) involves calibration to observed the NAO. Also, for the decadal to multidecadal timescales, the variability in the reconstructions diverge prior to the instrumental record, including the reconstruction by Ortega et al. (2015). This indicates that the long standing problem of incoherence between different NAO reconstructions prior to the instrumental record is still valid (Pinto and Raible, 2012). The reconstructions shown

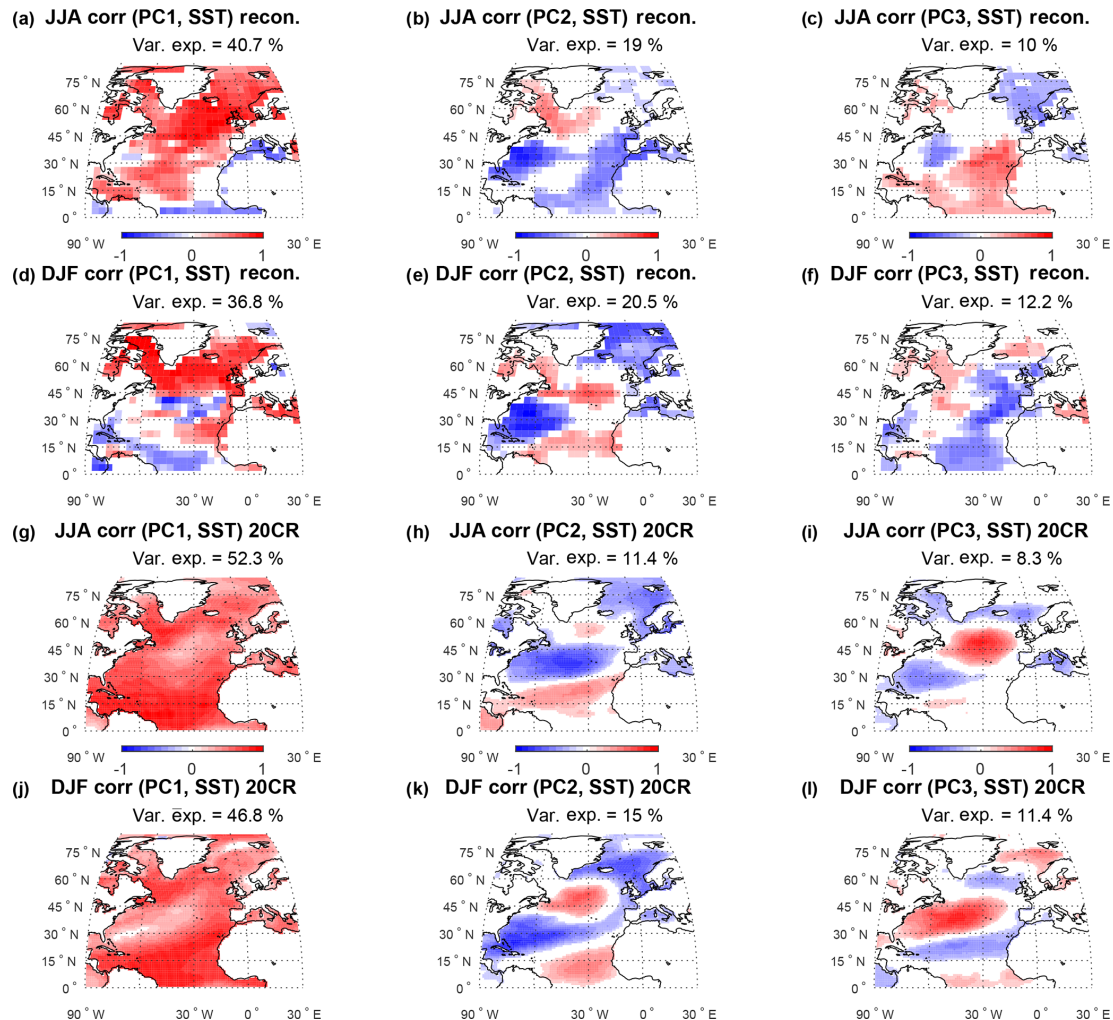


Figure 11. (a–c) Regression of the first three reconstructed PCs of SSTs on reconstructed sum50 JJA SSTs, which corresponds to the reconstructed EOF patterns. Panels (d)–(f) are the same as (a)–(c) but for DJF. Corresponding plots for reconstructions of sum50 (May–October), win50 (November–April) and the winter-centered annual mean (win100, August–July) are shown in Fig. S14. (g–i) Regression of the first three COBE SST PCs on COBE JJA SSTs, which corresponds to the reconstructed EOF patterns. Panels (j)–(l) are the same as (g)–(i) but for DJF. A corresponding figure for 20CR sum50 (May–October), win50 (November–April) and the winter-centered annual mean win100 (August–July) can be found in the Fig. S15. The time period for all plots is 1851–1970. Data only shown for $p < 0.05$.

in Fig. 15b–c are scaled to the decadal variability in the observed NAO to facilitate comparing the interannual variability. It is clear that the interannual amplitude of the reconstruction by Ortega et al. (2015) is underestimated, while our reconstruction appears to be only slightly underestimated in amplitude, and the reconstruction by Cook et al. (2019) could have a somewhat overestimated interannual variability. Factors which could contribute to the lack of correlation between our and the reconstruction by Cook et al. (2019) are that the relationship between drought and winter NAO is not stationary in time (López-Moreno and Vicente-Serrano, 2008) and that the number of records in the reconstruction by Cook et al. (2019) decrease strongly back in time prior to 1700 CE.

5 Discussion and conclusions

In this study we tested climate reconstructions of summer, winter and annual climate variability, based on a data set of eight ice cores covering 1241–1970 and an extended data set of 19 cores covering 1777–1970. While the increased number of ice cores can reduce noise in the reconstructions, the more geographically uneven distribution of the additional cores appears to have some negative effects on the skill of the reconstructions. This means that the overall added value of more ice core data seems less than the drawbacks of the much shorter time span being covered. Unfortunately it is not possible to test the reconstructions of eight versus 19 cores on truly equal terms, as the EOFs of the eight ice cores for shorter time periods are dependent on the exact choice of the

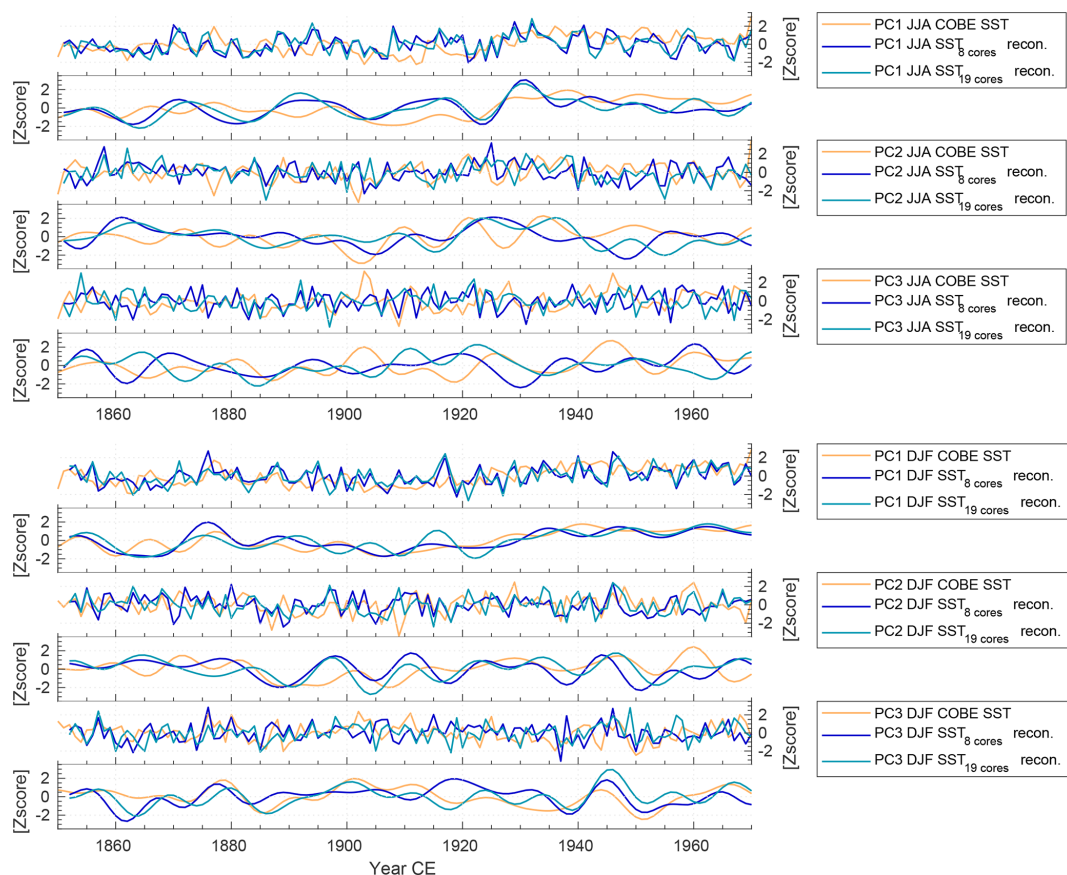


Figure 12. Time series of reconstructed PC1, PC2 and PC3 of SSTs using eight ice cores (dark blue) and 19 ice cores (light blue) compared to PC1, PC2 and PC3 of COBE SSTs (yellow). Smoothed curves are using a decadal FFT filter. Top six plots are for JJA, and bottom six plots are for DJF.

investigated time period. This is due to poor statistics in determining the EOFs when the number of ice cores is low and the data sample is short.

The inherent properties of climate variability with respect to autocorrelation and changes in governing weather patterns as illustrated in Sect. 4.1 are probably the reasons for the differences in skill seen for the reconstructions using different definitions of the target season. One consequence is that the skill for secondary circulation modes is better for the reconstructions targeting DJF rather than November–April; secondly, using the wider definition of summer (May–October) may reduce some noise in the temperature reconstruction, an effect which likely also can be seen for the temperature reconstructions of the winter-centered annual mean. Additionally, reconstruction of the DJF atmospheric circulation using winter-centered annual mean ice core data is attainable, which opens up the possibility of extending the winter reconstructions further back than with seasonal data. This could be done by using high-resolution chemistry data (e.g., Rasmussen et al., 2006) to define the seasons in the ice core data, even though the annual cycle in the ice core isotope data cannot be recovered.

The evaluation of correlation with the North Atlantic SSTs shows a particularly strong sensitivity to SST variability north of 50° N. This is in principle true for all seasons but in particular in winter, where the amplitude of the decadal changes in SSTs are captured by the reconstruction. This is achieved without tuning the reconstruction to observations. This indicates a clear potential for reconstructing AMO-like variability. Furthermore, the reconstructions yield qualitatively similar main patterns of variability as those based on observations (EOF1, 2 and 3). These SST patterns are connected to the main atmospheric modes of variability.

The reconstructions in this study only based on ice core data are using what one might call a minimal proxy data set. The thought behind it is to select few – but high-quality, well-dated and well-studied – proxy data rather than a large collection of data where the links between climate parameters and all proxy data have not been tested in detail. Furthermore, the use of isotope records has the property discussed in the Introduction of recording not only local information, while the assimilation using an isotope-enabled climate model allows coupling the model and proxy data without calibration. However, it is clear that the skill of the summer reconstructions

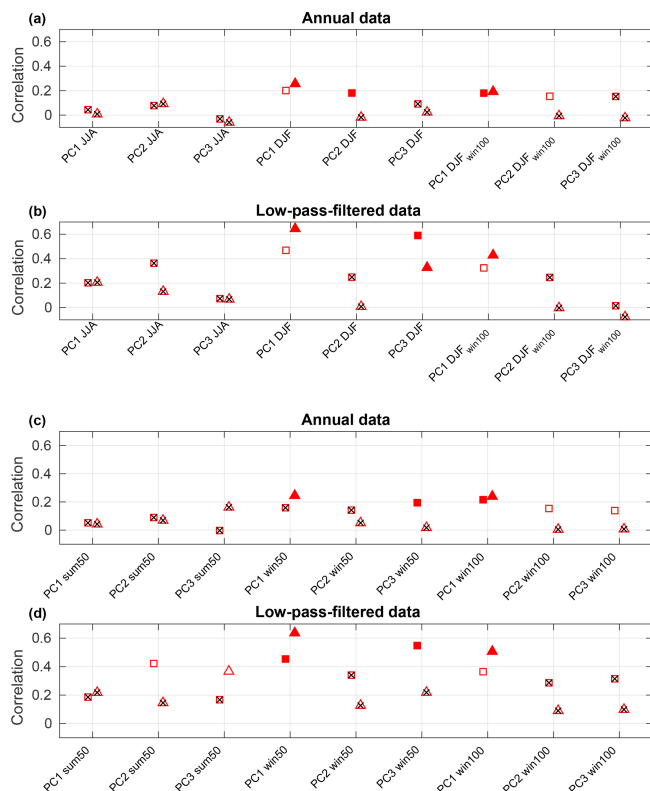


Figure 13. Correlation analysis for reconstructed PC1, PC2 and PC3 of SSTs using eight ice cores (triangles) and 19 ice cores (squares) correlated with PC1, PC2 and PC3 of COBE SSTs covering 1851–1970. Open markers indicate significance of $p < 0.1$ and full markers indicate $p < 0.05$, while crossed-out markers indicate $p > 0.1$.

is generally lower than the winter reconstructions. For this reason we also perform a test including European tree-ring data for two additional reconstructions for summer (JJA and May–October) covering 1241–1970. For these reconstructions the skill for temperature is clearly improved, although for SLP the skill only improves locally with no improvement of the skill for the main modes of circulation.

For model assimilation-type climate reconstructions the performance of the climate model is an important parameter. All climate models have biases that can influence the patterns of the reconstructed climate variability. Here we have mainly discussed the model bias in SLP during summer as this is the most prominent model-related problem found for our reconstructions. Given the relatively coarse model resolution ($3.75^\circ \times 3.75^\circ$), using a model with finer resolution and better representation of orography, atmospheric circulation and physics would probably yield a better climate reconstruction. However, the model used in this study fundamentally performs well when it comes to mimicking the variability in the isotopic composition of Greenland precipitation, which is what allows us to use the method of matching the ice core EOF patterns.

Different strategies can be chosen for attaining an uncertainty estimate of the reconstructions based on the analogue method. Bothe and Zorita (2019) presents different options: (i) uncertainty based on the fit of the analogues to the proxy data, (ii) a fixed distance allowed for the fit of the analogues but variable number of analogues and (iii) uncertainty estimated from the ensemble spread of model analogues. Our method employs a fixed number of model analogues (e.g., 39 for DJF 1241–1970), and the ensemble spread is therefore the most natural choice of uncertainty estimate. When comparing to other data sets, the RMSE can also be used along with the correlation coefficient as a measure of how well reconstruction matches the variability. This can for example reveal cases where the correlation is good but the amplitude of the variability does not match (see Fig. S16). In Fig. 5, where we plot time series of Greenland coastal temperature, we both show the ensemble spread and the RMSE with respect to the observations. Except for Ilulissat, which has very high observed variability, the ensemble spread and RMSE are very similar. This indicates that the ensemble spread is a good measure of the uncertainty at a grid point scale. In the comparison to other NAO reconstructions, we also show the ensemble spread and the RMSE with respect to the observations (Fig. 15). In this case the RMSE is well within the envelope of the ensemble spread of our reconstructed NAO, indicating that the spread is a relatively conservative measure of uncertainty. In addition we have investigated the quality of the fit over time (chi-squared distance for each time step) to see if there are trends or periods of very poorly fitting model analogues. Although there are years where we have trouble finding a good model analogue, the fit is on average throughout the records as good as for 1851–1970 where the reconstructions are evaluated. For example, there are no large decadal trends in the fit. From a statistical point of view, the reconstructions are therefore equally valid any time during the reconstruction as there is no calibration involved in the method.

The approach of using an ensemble of analogues not only improves the reconstruction in terms of correlation with observations but also reduces the variability when producing the ensemble mean due to averaging out some of the variability (Gómez-Navarro et al., 2017). Using the example of the Greenland coastal temperature again (Fig. 5), the amplitude of the year-to-year variability is somewhat underestimated in the reconstruction, while the decadal-scale variability is well captured. This smoothing of the high-frequency variability in the reconstruction can to a certain extent be attributed to the ensemble approach but also to the relatively coarse resolution of the model, which also smooths out variability. On the other hand, the SST reconstruction (Fig. 9) shows an overestimated variability in winter, which could be due to using an atmospheric signal to reconstruct ocean variability, while the amplitude is underestimated for summer. This contrast can probably be explained by the lower skill for summer, which causes loss of variability due to lack of coherency in the

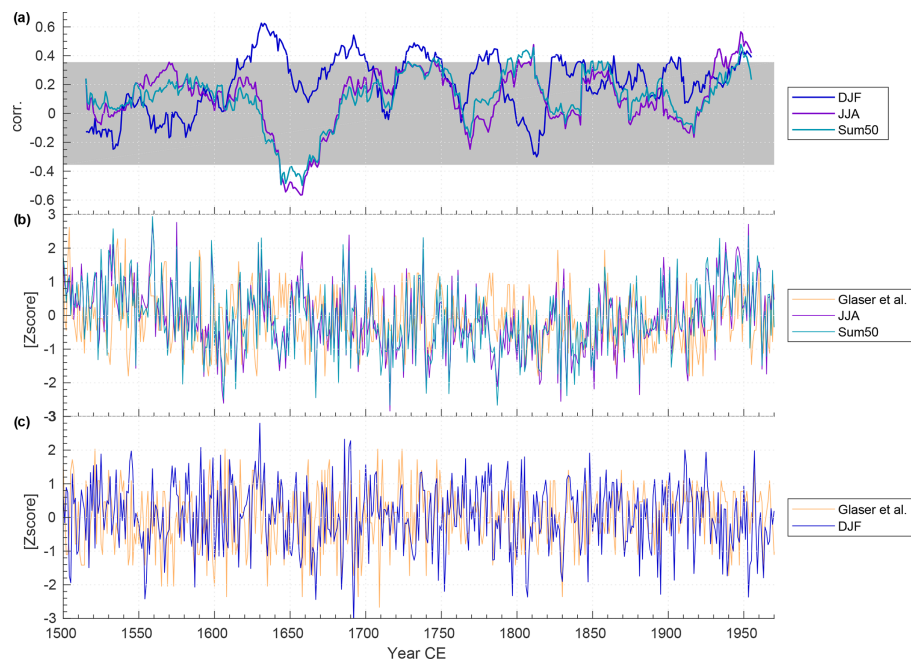


Figure 14. (a) Moving 31-year correlation between the Glaser and Riemann (2009) central Europe temperature index (JJA, DJF) and reconstructed temperature from this study (JJA, sum50, DJF). Correlations beyond the gray shaded area are significant ($p < 0.05$). (b) Time series of the Glaser and Riemann (2009) central Europe temperature index (JJA) and reconstructed temperature from this study for summer (JJA, sum50). (c) Time series of the Glaser and Riemann (2009) central Europe temperature index (DJF) and reconstructed temperature from this study (DJF). For the reconstructed temperature from this study we extract the area mean temperature (T2m) for the box 50–60° N and 0–20° E using only values for land.

ensemble. For the reconstruction of atmospheric circulation (SLP), the amplitude of year-to-year variability is well preserved, and the ensemble averaging appears to have a minor effect on the high-frequency variability (Figs. 8 and 15). One factor in preserving the year-to-year atmospheric variability is that we are sampling from a model simulation where, for example, the NAO has a nearly white power spectrum (not shown), and given that the ensemble spread is relatively large (Fig. 15), this spectrum will be preserved in the reconstruction.

To attain the best possible reconstruction of climate variability, taking into account the nature of the target for the reconstruction is important. This is illustrated by the dependency of the skill of the climate reconstructions on the definition of seasonality, due to the seasonal changes of the patterns or variability. For winter a narrow definition of the season (DJF) yields better performance for circulation patterns. Furthermore, in some cases a wider definition of the season, e.g., for summer and annual data, can provide better performance for temperature due to better capturing the signal during months of higher autocorrelation and less variability.

Further development of seasonal climate field reconstructions requires a larger data set of well-study proxy records. Isotope records of tree-ring cellulose from regions with sustained winter snow are potential sources for expanding the spatial coverage for winter (Seftigen et al., 2011; Ed-

wards et al., 2017). In more temperate climates such records could be used for reconstructing summer variability (Labuhn et al., 2016). Speleothem data could potentially also be used; however it is a challenge to find high-resolution continuous data sets due to growth hiatuses (e.g., de Jong et al., 2013). Newly updated isotope-enabled climate models (e.g., Cauquoin et al., 2019) show the continual development of this field. This makes running new millennium-length model simulations attractive for the purpose of providing better sampling pools for finding model analogues to match the proxy data. Although not shown in this study, reconstruction of precipitation is also possible using the analogue method. However, in particular for precipitation, better model resolution is important to capture storm tracks and orographic effects. Finally, the indication found in this study that it is possible to capture the main SST patterns of the North Atlantic makes this approach a good supplement to marine records due to better precision of the dating of terrestrial records.

Code and data availability. The code for the reconstruction method as well as the reconstructions shown in this paper are available upon request to the corresponding author. The time series for PC1 and PC2 of reconstructed DJF SLP in the North Atlantic region previously published by Sjolte et al.

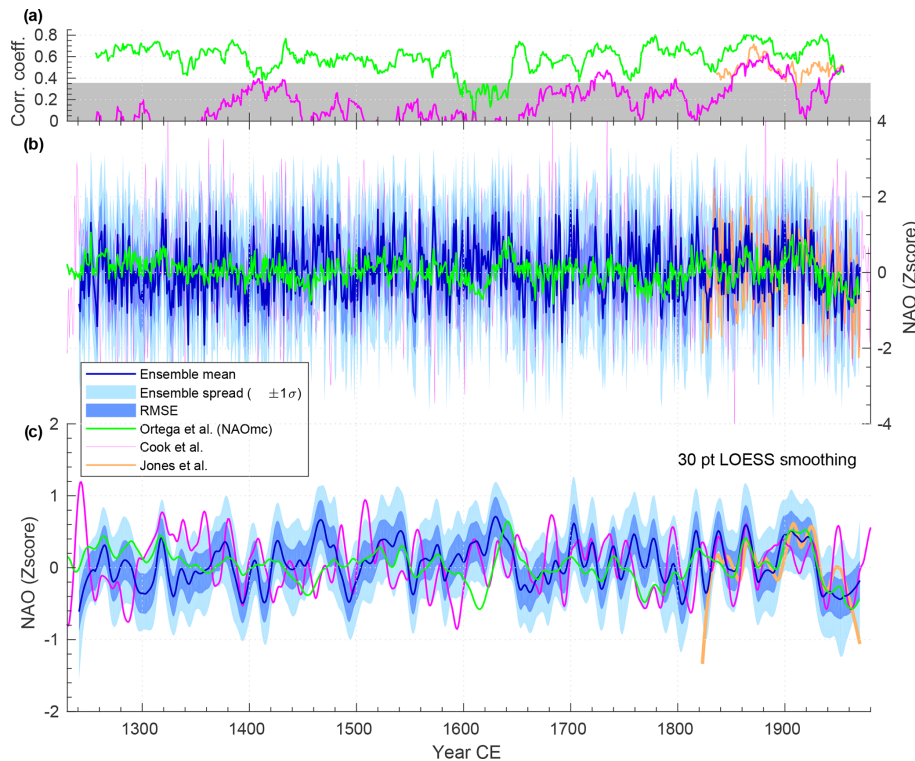


Figure 15. (a) Moving 31-point correlation between reconstructed DJF NAO from this study and Cook et al. (2019) (magenta), Ortega et al. (2015) (green) and observed NAO (yellow) (Jones et al., 1997). Correlations beyond the gray shaded area are significant ($p < 0.05$). (b) Ensemble mean reconstructed NAO (PC1 of reconstructed SLP, Hurrell et al., 2003) with error estimated by ensemble spread and RMSE, compared to observed NAO (Jones et al., 1997) and NAO reconstructions by Cook et al. (2019) and Ortega et al. (2015). The amplitude of all time series are scaled to fit the decadal variability in the observed NAO. (c) Same as (b), except filtered with a 30 point “loess” filter.

(2018) are available from the PANGAEA open-access data library (<https://doi.org/10.1594/PANGAEA.892841>).

Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/cp-16-1737-2020-supplement>.

Author contributions. JS developed the method, did the reconstructions, performed the analysis and wrote the first version of the manuscript. FA contributed to the writing, method development and analysis. BMV provided seasonal and annual ice core data. RM contributed to method development. CS contributed to the initial study and method development. MV and GL provided technical support and insights on climate modeling. All authors discussed and edited the paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Support for the Twentieth Century Reanalysis Project version 2c data set is provided by the U.S. Department of Energy, Office of Science Biological and Environmental

Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office. COBE SST data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at <https://www.esrl.noaa.gov/psd/> (last access: 30 August 2016). The authors thank two anonymous referees and the editor for insightful comments and helpful suggestions.

Financial support. This work was supported by the Swedish Research Council (grants nos. DNR2011-5418 & DNR2013-8421 to Raimund Muscheler), the Crafoord Foundation and the strategic research program of Modelling the Regional and Global Earth system (MERGE) hosted by the Faculty of Science at Lund University. Florian Adolphi was supported by the Swedish Research Council (DNR grant no. 2016-00218 to Florian Adolphi).

Review statement. This paper was edited by Nerilie Abram and reviewed by two anonymous referees.

References

- Bothe, O. and Zorita, E.: Proxy surrogate reconstructions for Europe and the estimation of their uncertainties, *Clim. Past*, 16, 341–369, <https://doi.org/10.5194/cp-16-341-2020>, 2020.
- Cauquoin, A., Werner, M., and Lohmann, G.: Water isotopes – climate relationships for the mid-Holocene and preindustrial period simulated with an isotope-enabled version of MPI-ESM, *Clim. Past*, 15, 1913–1937, <https://doi.org/10.5194/cp-15-1913-2019>, 2019.
- Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason Jr., B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Broennimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, O., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, *Q. J. Roy. Meteorol. Soc.*, 137, 1–28, <https://doi.org/10.1002/qj.776>, 2011.
- Cook, E. R., Kushnir, Y., Smerdon, J. E., Williams, A. P., Anchukaitis, K. J., and Wahl, E. R.: A Euro-Mediterranean tree-ring reconstruction of the winter NAO index since 910CE, *Clim. Dynam.*, 53, 1567–1580, <https://doi.org/10.1007/s00382-019-04696-2>, 2019.
- Craig, H.: Isotopic Variations in Meteoric Waters, *Science*, 133, 1702–1703, <https://doi.org/10.1126/science.133.3465.1702>, 1961.
- de Jong, R., Kamenik, C., and Grosjean, M.: Cold-season temperatures in the European Alps during the past millennium: variability, seasonality and recent trends, *Quaternary Sci. Rev.*, 82, 1–12, <https://doi.org/10.1016/j.quascirev.2013.10.007>, 2013.
- Ebisuzaki, W.: A method to estimate the statistical significance of a correlation when the data are serially correlated, *J. Climate*, 10, 2147–2153, [https://doi.org/10.1175/1520-0442\(1997\)010<2147:AMTETS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<2147:AMTETS>2.0.CO;2), 1997.
- Edwards, T. W., Hammarlund, D., Newton, B. W., Sjolte, J., Linderson, H., Sturm, C., Amour, N. A. S., Bailey, J. N.-L., and Nilsson, A. L.: Seasonal variability in Northern Hemisphere atmospheric circulation during the Medieval Climate Anomaly and the Little Ice Age, *Quaternary Sci. Rev.*, 165, 102–110, <https://doi.org/10.1016/j.quascirev.2017.04.018>, 2017.
- Furukawa, R., Uemura, R., Fujita, K., Sjolte, J., Yoshimura, K., Matoba, S., and Iizuka, Y.: Seasonal-Scale Dating of a Shallow Ice Core From Greenland Using Oxygen Isotope Matching Between Data and Simulation, *J. Geophys. Res.-Atmos.*, 122, 10873–10887, <https://doi.org/10.1002/2017JD026716>, 2017.
- Glaser, R. and Riemann, D.: A thousand-year record of temperature variations for Germany and Central Europe based on documentary data, *J. Quaternary Sci.*, 24, 437–449, <https://doi.org/10.1002/jqs.1302>, 2009.
- Gómez-Navarro, J. J., Zorita, E., Raible, C. C., and Neukom, R.: Pseudo-proxy tests of the analogue method to reconstruct spatially resolved global temperature during the Common Era, *Clim. Past*, 13, 629–648, <https://doi.org/10.5194/cp-13-629-2017>, 2017.
- Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., and Perkins, W. A.: The lastmillennium climate reanalysis project: Framework and first results, *J. Geophys. Res.-Atmos.*, 121, 6745–6764, <https://doi.org/10.1002/2016JD024751>, 2016.
- Hurrell, J. W., Kushnir, Y., Ottersen, G., and Visbeck, M.: An Overview of the North Atlantic Oscillation, 1–35, American Geophysical Union, <https://doi.org/10.1029/134GM01>, 2003.
- Ishii, M., Shouji, A., Sugimoto, S., and Matsumoto, T.: Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the Kobe Collection, *Int. J. Climatol.*, 25, 865–879, <https://doi.org/10.1002/joc.1169>, 2005.
- Johnsen, S. J.: Stable isotope homogenization of polar firn and ice, in: *Proceedings of Symposium on Isotopes and Impurities in Snow and Ice*, 118, 210–219, 1977.
- Jones, P., Jonsson, T., and Wheeler, D.: Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland, *Int. J. Climatol.*, 17, 1433–1450, [https://doi.org/10.1002/\(SICI\)1097-0088\(19971115\)17:13<1433::AID-JOC203>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-0088(19971115)17:13<1433::AID-JOC203>3.0.CO;2-P), 1997.
- Jónsson, T.: The observations of Jon Thorsteinsson in Nes and Reykjavik 1820–1854, Tech. rep., Icel. Met. Office Report, 1989.
- Jungclauss, J. H., Keenlyside, N., Botzet, M., Haak, H., Luo, J.-J., Latif, M., Marotzke, J., Mikolajewicz, U., and Roeckner, E.: Ocean Circulation and Tropical Variability in the Coupled Model ECHAM5/MPI-OM, *J. Climate*, 19, 3952–3972, <https://doi.org/10.1175/JCLI3827.1>, 2006.
- Klein, F., Abram, N. J., Curran, M. A. J., Goosse, H., Goursaud, S., Masson-Delmotte, V., Moy, A., Neukom, R., Orsi, A., Sjolte, J., Steiger, N., Stenni, B., and Werner, M.: Assessing the robustness of Antarctic temperature reconstructions over the past 2 millennia using pseudoproxy and data assimilation experiments, *Clim. Past*, 15, 661–684, <https://doi.org/10.5194/cp-15-661-2019>, 2019.
- Labuhn, I., Daux, V., Girardclos, O., Stievenard, M., Pierre, M., and Masson-Delmotte, V.: French summer droughts since 1326 CE: a reconstruction based on tree ring cellulose $\delta^{18}\text{O}$, *Clim. Past*, 12, 1101–1117, <https://doi.org/10.5194/cp-12-1101-2016>, 2016.
- López-Moreno, J. I. and Vicente-Serrano, S. M.: Positive and Negative Phases of the Wintertime North Atlantic Oscillation and Drought Occurrence over Europe: A Multitemporal-Scale Approach, *J. Climate*, 21, 1220–1243, <https://doi.org/10.1175/2007JCLI1739.1>, 2008.
- Luterbacher, J., Xoplaki, E., Dietrich, D., Jones, P. D., Davies, T. D., Portis, D., Gonzalez-Rouco, J. F., von Storch, H., Gyalistras, D., Casty, C., and Wanner, H.: Extending North Atlantic oscillation reconstructions back to 1500, *Atmos. Sci. Lett.*, 2, 114–124, <https://doi.org/10.1006/asle.2002.0047>, 2001.
- Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M., and Wanner, H.: European seasonal and annual temperature variability, trends, and extremes since 1500, *Science*, 303, 1499–1503, <https://doi.org/10.1126/science.1093877>, 2004.
- Mann, M. E., Bradley, R. S., and Hughes, M. K.: Global-scale temperature patterns and climate forcing over the past six centuries, *Nature*, 392, 779–787, <https://doi.org/10.1038/33859>, 1998.
- McCarthy, G. D., Haigh, I. D., Hirschi, J. J. M., Grist, J. P., and Smeed, D. A.: Ocean impact on decadal Atlantic climate variability revealed by sea-level observations, *Nature*, 521, 508–U172, <https://doi.org/10.1038/nature14491>, 2015.
- Moffa-Sanchez, P., Born, A., Hall, I. R., Thornalley, D. J. R., and Barker, S.: Solar forcing of North Atlantic surface temperature and salinity over the past millennium, *Nat. Geosci.*, 7, 275–278, <https://doi.org/10.1038/ngeo2094>, 2014.

- Moreno-Chamarro, E., Zanchettin, D., Lohmann, K., and Jungclauss, J. H.: An abrupt weakening of the subpolar gyre as trigger of Little Ice Age-type episodes, *Clim. Dynam.*, 48, 727–744, <https://doi.org/10.1007/s00382-016-3106-7>, 2017.
- Ortega, P., Lehner, F., Swingedouw, D., Masson-Delmotte, V., Raible, C. C., Casado, M., and Yiou, P.: A model-tested North Atlantic Oscillation reconstruction for the past millennium, *Nature*, 523, 71+, <https://doi.org/10.1038/nature14518>, 2015.
- Pinto, J. G. and Raible, C. C.: Past and recent changes in the North Atlantic oscillation, *Wiley Interdisciplinary Reviews-Climate Change*, 3, 79–90, <https://doi.org/10.1002/wcc.150>, 2012.
- Rasmussen, S. O., Andersen, K. K., Svensson, A. M., Steffensen, J. P., Vinther, B. M., Clausen, H. B., Siggaard-Andersen, M.-L., Johnsen, S. J., Larsen, L. B., Dahl-Jensen, D., Bigler, M., Röthlisberger, R., Fischer, H., Goto-Azuma, K., Hansson, M. E., and Ruth, U.: A new Greenland ice core chronology for the last glacial termination, *J. Geophys. Res.-Atmos.*, 111, D06102, <https://doi.org/10.1029/2005JD006079>, 2006.
- Reeves Eyre, J. E. J. and Zeng, X.: Evaluation of Greenland near surface air temperature datasets, *The Cryosphere*, 11, 1591–1605, <https://doi.org/10.5194/tc-11-1591-2017>, 2017.
- Seftigen, K., Linderholm, H. W., Loader, N. J., Liu, Y., and Young, G. H.: The influence of climate on $\delta^{13}C/\delta^{12}C$ and $\delta^{18}O/\delta^{16}O$ ratios in tree ring cellulose of *Pinus sylvestris* L. growing in the central Scandinavian Mountains, *Chem. Geol.*, 286, 84–93, <https://doi.org/10.1016/j.chemgeo.2011.04.006>, 2011.
- Sjolte, J., Hoffmann, G., Johnsen, S. J., Vinther, B. M., Masson-Delmotte, V., and Sturm, C.: Modeling the water isotopes in Greenland precipitation 1959–2001 with the meso-scale model REMO-iso, *J. Geophys. Res.-Atmos.*, 116, D18105, <https://doi.org/10.1029/2010JD015287>, 2011.
- Sjolte, J., Hoffmann, G., and Johnsen, S. J.: Modelling the response of stable water isotopes in Greenland precipitation to orbital configurations of the previous interglacial, *Tellus B*, 66, 22872, <https://doi.org/10.3402/tellusb.v66.22872>, 2014.
- Sjolte, J., Sturm, C., Adolphi, F., Vinther, B. M., Werner, M., Lohmann, G., and Muscheler, R.: Solar and volcanic forcing of North Atlantic climate inferred from a process-based reconstruction, *Clim. Past*, 14, 1179–1194, <https://doi.org/10.5194/cp-14-1179-2018>, 2018.
- Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., Anderson, D. M., Steig, E. J., and Noone, D.: Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling, *Clim. Past*, 15, 1251–1273, <https://doi.org/10.5194/cp-15-1251-2019>, 2019.
- Vinther, B., Jones, P., Briffa, K., Clausen, H., Andersen, K., Dahl-Jensen, D., and Johnsen, S.: Climatic signals in multiple highly resolved stable isotope records from Greenland, *Quaternary Sci. Rev.*, 29, 522–538, <https://doi.org/10.1016/j.quascirev.2009.11.002>, 2010.
- Vinther, B. M., Andersen, K. K., Jones, P. D., Briffa, K. R., and Cappelen, J.: Extending Greenland temperature records into the late eighteenth century, *J. Geophys. Res.-Atmos.*, 111, D11105, <https://doi.org/10.1029/2005JD006810>, 2006.
- Werner, M., Haese, B., Xu, X., Zhang, X., Butzin, M., and Lohmann, G.: Glacial–interglacial changes in $H_2^{18}O$, HDO and deuterium excess – results from the fully coupled ECHAM5/MPI-OM Earth system model, *Geosci. Model Dev.*, 9, 647–670, <https://doi.org/10.5194/gmd-9-647-2016>, 2016.
- Wilson, R., Anchukaitis, K., Briffa, K. R., Büntgen, U., Cook, E., D’Arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B., Hegerl, G., Helama, S., Klesse, S., Krusic, P. J., Linderholm, H. W., Myglan, V., Osborn, T. J., Rydval, M., Schneider, L., Schurer, A., Wiles, G., Zhang, P., and Zorita, E.: Last millennium northern hemisphere summer temperatures from tree rings: Part I: The long term context, *Quaternary Sci. Rev.*, 134, 1–18, <https://doi.org/10.1016/j.quascirev.2015.12.005>, 2016.