Constraining two climate field reconstruction methodologies over the North Atlantic realm using pseudo-proxy experiments

Tine Nilsen [a, b, *], Stefanie Talento [c], Johannes P. Werner [d, 1]

[a] Department of Mathematics and Statistics, UiT - The Arctic University of Norway, 9037, Tromsø, Norway
[b] Department of Geography, Justus-Liebig University of Giessen, 35390, Giessen, Germany
[c] Potsdam Institute for Climate Impact Research, 14473, Potsdam, Germany
[d] 63225, Langen, Hesse, Germany

ARTICLE INFO

ABSTRACT

This study presents pseudo-proxy experiments to quantify the reconstruction skill of two climate field reconstruction methodologies for a marine proxy network subject to age uncertainties. The BARCAST methodology (Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time) is tested for sea surface temperature (SST) reconstruction for the first time over the northern North Atlantic region, and compared with a classic analogue reconstruction methodology. The reconstruction experiments are performed at annual and decadal resolution. We implement chronological uncertainties inherent to marine proxies as a novelty, using a simulated age-model ensemble covering the past millennium. Our experiments comprise different scenarios for the input data network, with the noise levels added to the target variable extending from ideal to realistic. Results show that both methodologies are able to reconstruct the Summer mean SST skillfully when the proxy network is considered absolutely dated, but the skill of the analogue method is superior to BARCAST. Only the analogue method provides skillful correlations with the true target variable in the case of a realistic noisy and age-uncertain proxy network. The spatiotemporal properties of the input target data are partly contrasting with the BARCAST model formulations, resulting in an inferior reconstruction ensemble that is similar to a white-noise stochastic process in time. The analogue method is also successful in reconstructing decadal temperatures, while BARCAST fails. The results contribute to constraining uncertainties in CFR for ocean dynamics which are highly important for climate across the globe.

1. Introduction

The well-documented changes in a number of observational variables demonstrate a significant role of anthropogenic activities in modifying the Earth's climate. Yet, the quantification of natural variability in the climate system is crucial to put recent and future climate change into a long-term perspective. Climate Field reconstructions (CFR) offer critical insights into the range and geographic characteristics of historical climate variability prior to the instrumental era, by scaling-up localized and sparse proxy measurements and providing information that can be compared with climate model simulation output.

A number of CFRs have been published over the last decades, quantifying late Holocene climate variations (Masson-Delmotte et al., 2013; Christiansen and Ljungqvist, 2017). CFR provides a distinct advantage over averaged climate reconstructions for instance by studying the response of climate in a region to some external forcing, such as a large tropical volcanic eruption. A well-known class of CFR methodology is multivariate linear regression, where the reconstruction problem can be formulated either using direct or indirect regression between proxies and the climate variable. Regularization of the problem is needed because the system of equations is underdetermined. Examples of CFR methodologies using different regularizations include principal component regression (Luterbacher et al., 2002), canonical correlation analysis (Smerdon et al., 2010) and RegEM (Mann et al., 2008, 2009). A second class of CFR methods includes Bayesian hierarchical models (BHM). The Bayesian methodology differs from the common regression-based method in the sense that model parameters are

* Corresponding author. Department of Mathematics and Statistics, UiT - The Arctic University of Norway, 9037, Tromsø, Norway.
E-mail address: tine.nilsen@uit.no (T. Nilsen).
[1] Independent researcher.

formulated in probability statements representing the current state of knowledge of the climate variable given observation data. This type of method is probabilistic, resulting in reconstruction ensembles (Tingley and Huybers, 2010; Luterbacher et al., 2016; Werner et al., 2018). Yet another class of CFR methodology comprises data assimilation methods (Steiger et al., 2014), typically combined with forward proxy system modeling (Hakim et al., 2016; Tardif et al., 2019). This approach combines a sparse proxy network with climate model simulations through a prior, mapped to the proxy space using proxy system modeling. A Kalman filter is typically used to update the prior. A similar but simpler methodology is the analogue method, or the proxy surrogate reconstruction method (Graham et al., 2007; Franke et al., 2011). Relevant differences between the analogue method and data assimilation techniques are discussed in Gómez-Navarro et al. (2017). Instrumental observations or climate model time slices are used as analogues to a gridded proxy network, but there is no prior distribution or Kalman filtering for the analogue method. The reconstruction at every time step is selected as one single or a function of the closest analogue time slices.

The existing CFRs defined with annual temporal resolution are based on dominantly terrestrial proxy records, representing terrestrial regions (Tingley and Huybers, 2010, 2013; Luterbacher et al., 2016; Werner et al., 2018), the entire globe (Mann et al., 2009; Hakim et al., 2016; Tardif et al., 2019; Neukom et al., 2019), or the global marine realm (Singh et al., 2018). The majority of the published CFRs exhibits annual reconstructed values over the last 1−2 millennia, and so the SST field reconstruction of Tierney et al. (2020) differs in comprising only two time slices. One is an average over the Last Glacial Maximum (LGM) time period 23-19 kyr before present (BP), and the other is averaged for the late Holocene period 4-0 kyr BP. In summary, there are no time-resolving marine CFRs for the past millennium based on marine proxy data alone. This is unfortunate because the oceanic impacts on surface air temperature cannot be independently investigated when using the same, terrestrial proxies to reconstruct both land and ocean temperatures. Given the importance of circulation patterns and heat content of the North Atlantic Ocean in regulating atmospheric variability across the Northern Hemisphere, this missing piece of paleoclimate information has a central role in improving the understanding of natural climate variability.

The spatial patterns of North Atlantic sea surface temperature (SST) and atmospheric sea level pressure gradients drive atmospheric circulation that brings about large-scale changes in precipitation and temperature across the Northern Hemisphere, the strength and trajectory of Atlantic storms as well as ocean biological productivity (Goldenberg et al., 2001; Reid et al., 2003; Beugrand and Reid, 2003; Sutton and Hodson, 2005). Long-term shifts in the distribution of heat in the North Atlantic Ocean could also have significant socioeconomic impacts across the Northern Hemisphere, brought about by changes in the strength of the subpolar gyre and/or associated Atlantic meriodional overturning circulation (AMOC).

Index reconstructions exist for the AMOC and the Atlantic multidecadal variability (AMV) (Gray et al., 2004; Mann et al., 2009; Rahmstorf et al., 2015; Wang et al., 2017). These reconstructions are based entirely or dominantly on terrestrial data: Mann et al. (2009) has a proportion of 91% tree-ring records with marine records constituting just 1.6% of the proxy network. However, we note that the more recent multi-method CFR ensemble of surface temperature is improved in this respect with 19% marine records, all annually resolved (Neukom et al., 2019).

SST reconstruction based on terrestrial proxies relies on documented ocean-atmosphere teleconnections, with the advantage of high-precision dating and data abundance. The SST is measured more directly in marine proxy records (Moffa-Sánchez et al., 2019). The marine records are considered prominent candidates for state-of-the-art marine CFR in this study. The North Atlantic Ocean including the Arctic is particularly suitable for initial testing purposes since the extratropical region is subject to high sampling density, demonstrated by the Ocean2k network (McGregor et al., 2015). From the total of 57 marine proxy records, 31 of these are sampled in the North Atlantic and the Arctic Ocean. A number of these records exhibit subdecadal resolution due to targeted research efforts on high-accumulation sites. Recent proxy development and improved statistical methods in age-model simulation and CFR contribute to our confidence that the field of northern North Atlantic SST can be skillfully reconstructed (Pyrina et al., 2017; Reynolds et al., 2018), despite the shortcomings of the marine proxy data. Proxy drawbacks include age-uncertainties, poor spatial availability, depth and seasonality biases. Additionally and specifically for marine sediment deposits, the proxy forming process is comparable to low-pass filtering, resulting in generic low temporal resolution.

In principle, all types of marine proxy data are subject to chronological uncertainty, either through the age-depth relationship of marine sediments, or in the form of miscounted bands/layers. Bivalve records are exceptions exhibiting annual bands, but cross-dated to achieve absolute dating similar to tree-ring records (Butler et al., 2013; Pyrina et al., 2017; Reynolds et al., 2017, 2018). The temporal resolution for deposits is estimated as a function of depth and accumulation rate. This rate may vary down-core, and is best constrained by radiometric dating techniques and tephra layers as absolute age markers. Age-depth models for deposits can be simulated using Bayesian methods such as BACON (Blaauw and Christen, 2011). The "Banded archive modeling" (BAM) statistical method is designed to simulate age model ensembles for annually banded or layered records such as corals (Comboul et al., 2014).

We propose to generate an entirely marine CFR in this study, using marine data for reconstruction and targeting the SST variable. Two state-of-the-art CFR methods are described and tested in detail, evaluating their skill in reconstructing northern North Atlantic SST over the past millennium with annual and decadal resolution. The reconstruction methods to be tested are the Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time (BARCAST) (Tingley and Huybers, 2010), and the analogue method (Franke et al., 2011). This is the first time that BARCAST applicability is tested for an SST reconstruction. An extended version of the methodology is applied, allowing age-uncertain proxy records following Werner and Tingley (2015). This functionality is critical for handling age uncertainties properly when reconstructing northern North Atlantic SST. By including ensemble chronologies and their errors into the Bayesian reconstruction framework the age-uncertainties are, in theory, reduced compared to other reconstruction methodologies including the analogue method, where age-uncertainties will contribute to increased total uncertainty in the reconstruction (Werner and Tingley, 2015).

Ensembles of annual age-uncertain chronologies are simulated using the BAM technique. The experiment setup is similar to previous BARCAST experiments in terms of age-uncertainties, where the method has proven successful (Werner and Tingley, 2015; Werner et al., 2018). This bridges the gap between terrestrial and marine CFR using BARCAST. It is interesting to test for the pseudo-reality whether the method is able to skillfully reconstruct annual SST, even if this high resolution is difficult to achieve for existing marine proxy networks. We have also implemented one set of pseudo proxy experiments producing decadally-resolved reconstructions, and another set combining BARCAST with a

clustering technique, similar to Talento et al. (2019). The reconstruction skill was demonstrated to be higher for the annually resolved reconstruction than for decadal resolution in that study, although the BARCAST implementation was simple (no age-uncertainty) and the target variable was precipitation instead of SST.

The temporal resolution of North Atlantic proxy deposits varies from subdecadal to multicentennial (McGregor et al., 2015). The BACON technique would produce more realistic age-depth models for marine sediments than BAM, although with decadal resolution at best.

The manuscript is structured as follows. Sect. 2 describes the experiment design, input data and reconstruction methodologies, including their ensemble chronology extensions. The skill metrics are also outlined. In Sect. 3 we present the reconstruction skill results and the implications of age-uncertainties in marine data. A discussion follows in Sect. 4, before finalizing with concluding remarks in Sect. 5.

2. Data and methods

CFRs are generated from a limited number of proxy records, and validating such reconstructions is non-trivial because there is no perfect, independent record of the past climate evolution. The process of pseudo-proxy experiments overcomes this shortcoming by testing CFR methodologies using data from general circulation models (GCM) as a pseudo-reality (Mann et al., 2005, 2007; Lee et al., 2008). Pseudo-proxy experiments are used to test the mean and spatial skill of novel CFR methodologies as well as the sensitivity to the observation data network (Smerdon et al., 2011; Smerdon, 2012; Wang et al., 2014; Gómez-Navarro et al., 2017).

The idea behind our idealized pseudo-proxy experiment is to extract target SST data from a climate model simulation. The target data is sampled in a spatiotemporal pattern that simulates real proxy and instrumental networks. The target data representing proxies are perturbed with noise to simulate real-world proxy data in a systematic manner, while the pseudo-instrumental data is only weakly perturbed with noise of magnitude typical for the real-world instrumental data errors. The pseudo-proxy and pseudo-instrumental data are used as input to the BARCAST and analogue reconstruction techniques, and the resulting reconstructions are compared with the true simulation target. The reconstruction skill is quantified through statistical metrics for a millennium-long validation interval.

2.1. Model data and target

The target variable for our pseudo-proxy experiment is the North Atlantic (57.5°W, 47.5°N to 37.5°E, 82.5°N) summer mean (June, July, August, JJA) SST, originating from one simulation of the Community Earth System Model, Last Millennium Ensemble (CESM-LME, run 001), (Otto-Bliesner et al., 2016). The ensemble simulation covers the time period 850-2005 AD, with transient forcing series based on reconstructed solar insolation (Vieira et al., 2011), volcanic activity (Gao et al., 2008), greenhouse gases (Schmidt et al., 2011), land-use and land cover (Pongratz et al., 2008; Hurtt et al., 2011), orbital parameters (Berger, 1978) and prescribed stratospheric aerosols.

The original SST variable has spatial resolution of ~ 1° in the northern North Atlantic Ocean, and is interpolated to a 5° by 5° regular grid for reconstruction purposes. This is the same resolution as the real-world HadSST.4.0.0.0 observation data (Kennedy et al., 2019).

Confined oceanic regions may be considered less representative

of the larger northern North Atlantic, and are excluded from the reconstruction domain. The excluded regions comprise the Davis Strait, Baffin and Hudson Bay, the Baltic Sea and the Gulf of Bothnia. In the real-world scenario, the temperature signal within these regions is strongly influenced by sea-ice and/or land processes (Moffa-Sánchez et al., 2019; Stramska and Białstrokogrodzka, 2015).

The target data is perturbed with white additive noise to simulate all types of errors and uncertainties in the instrumental and proxy data. This particular type of noise has been implemented and tested earlier for both methodologies, with satisfactory results (Nilsen et al., 2018; Talento et al., 2019).

2.2. Resampling the SST field for the pseudo-instrumental period 1850 - 2005 AD

The pseudo-instrumental network is constructed to mimic the characteristics of the HadSST.4.0.0.0 data set (Kennedy et al., 2019). It covers the period 1850−2005 in an uniform 5° by 5° grid. The re-sampled data is perturbed with additive white noise, resulting in pseudo-instrumental data with signal-to-noise (SNR) ratio of 3 by standard deviation. The noise represents e.g. measurement uncertainties and bias corrections in the real-world data.

2.3. Resampling the SST field for the pseudo-proxy period 850-2005 AD

A realistic proxy network is used as the baseline for selecting pseudo-proxy locations for the time period 850-2005 AD (Fig. 1a, Table A1). The North Atlantic proxy network comprises 22 records, with data originating from marine sediment cores or bivalves. Selected records originate from the North Atlantic Ocean2k
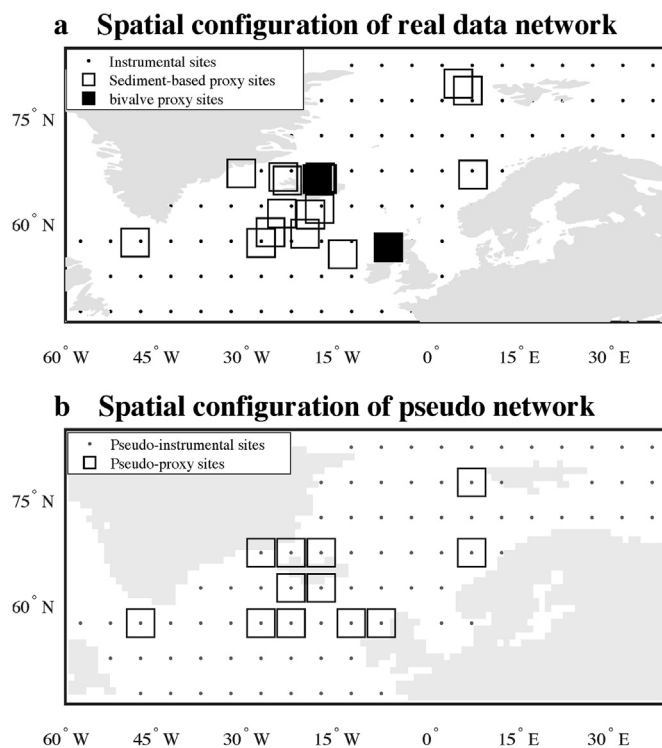
Fig. 1. The reconstruction region extends from 57.5°W, 47.5°N to 37.5°E, 82.5°N. Dots denote locations of (pseudo-)instrumental data, squares denote (pseudo-)proxy locations. (a) Shows the true proxy locations. These are shifted onto the nearest neighboring instrumental grid cell in (b), showing the pseudo-data network. The number of unique proxy sites is reduced from 22 to 12.

contribution to the PAGES2k archives, denoted by "Ocn_" site IDs in Table A1 (McGregor et al., 2015; PAGES2k Consortium, 2017). Ocean2k records were excluded if the locations are outside the grid specified in Sect. 2.1, or if the dating information is missing from the age-models of the publications. Additional proxy records are included in the network even if they are not part of the 2015 version of the Ocean2k database (numbered site IDs in Table A1). Specifically, we sought for reconstructions published later than May 2013, which is the cutoff-period for inclusion in the 2k synthesis (McGregor et al., 2015).

The spatial availability of the pseudo-proxy network is made artificially idealistic with the proxy locations as the baseline. These adjustments are optional, they may affect the reconstruction quality but are not critical for the functionality of the reconstruction methodologies. First, we opt to consider at most one proxy per grid cell for the pseudo-proxy experiments. Secondly, we reduce the number of grid cells by shifting the proxy locations onto the nearest neighboring instrumental location. The reduction of total number of grid nodes is a manner of easing the pressure on the computational resources for BARCAST. The resulting grid has 82 nodes in total, with 12 of these classified as unique pseudo-proxy locations, illustrated in Fig. 1b. The distance between neighboring pseudo-proxy locations ranges from 213 km to 1190 km.

The pseudo-proxy data are perturbed with additive white noise to simulate real-world proxy noise for the full simulation period 850-2005 AD. Three different levels of proxy-noise are tested, as the true SNR of marine proxies is generally unknown and may vary depending on the archive or proxy type. The noise perturbations are such that the resulting pseudo-proxies have a SNR of 10, 1 and 0.5 by standard deviation, respectively. The strongest noise level SNR = 0.5 is considered the most realistic scenario, similar to the estimates obtained by (Laepple and Huybers 2014) for marine proxy records.

2.4. Original BARCAST methodology

BARCAST is a Bayesian hierarchical model with three levels: process, data and prior levels. The evolution of the target variable in space and time is described at the process level. The data level describes the relationship between the target and the instrumental/proxy data, while the statistical information of the priors is incorporated into the last level. The hierarchical structure implies that model parameters are dependent across levels.

The temperature field $\boldsymbol{T}$ is modeled as a multivariate first-order autoregressive model (AR(1)) in time:

$$\boldsymbol{T}_{t+1} - \mu = \alpha(\boldsymbol{T}_t - \mu) + \boldsymbol{\varepsilon}_t \tag{1}$$

Where t denotes time in years, the scalar parameter μ is the mean of the process, and α is the AR(1) coefficient. The innovations (increments) $\boldsymbol{\varepsilon}_t$ are assumed to be IID normal draws $\boldsymbol{\varepsilon}_t \sim N(0, \boldsymbol{\Sigma})$. The spatial covariance matrix depicting the covariance between locations x_i and x_j is:

$$\Sigma_{ij} = \sigma^2 \exp(-\varphi|x_i - x_j|) \tag{2}$$

The spatial e-folding distance is $1/\varphi$. The simplicity of the exponentially decaying covariance structure keeps the computation time low, while the reconstruction skill is satisfactory according to earlier pseudo-proxy studies (Werner et al., 2013; Nilsen et al., 2018; Talento et al., 2019). The BARCAST assumption fits only partly with the behaviour displayed by the SST target (Fig. 2). The correlation is measured for Summer SST between every grid cell in the reconstruction region, but only the median correlations calculated within distance bins separated at 50 km are plotted in the
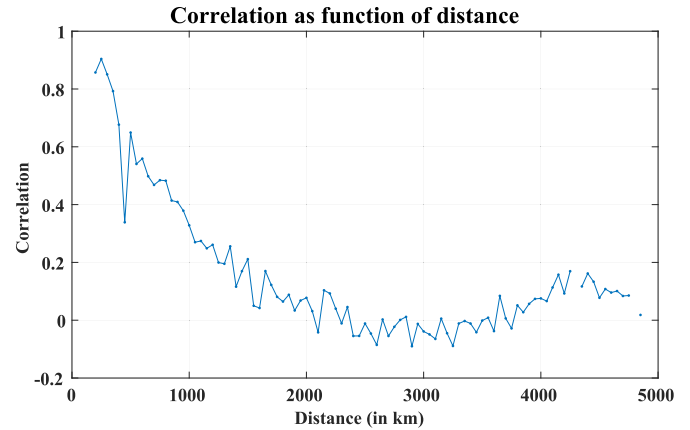

Fig. 2. Correlation of target model SST as function of distance in km. Distances are grouped into 50-km bins, and the median correlation within each bin is plotted.

figure. Initially, the correlations decrease and reach zero at distance ~2000 km, but they increase again as distances exceed 3000 km. We will test how well BARCAST reconstructs the SST field given these teleconnections, that contrasts to some degree the spatial covariance structures assumed within the BARCAST model Eq. (2).

On the data level, the observation equations for the instrumental and proxy data are:

$$\boldsymbol{W}_t = \begin{pmatrix} \mathbf{H}_{I,t} \\ \beta_1 \cdot \mathbf{H}_{P,t} \end{pmatrix} \boldsymbol{T}_t + \begin{pmatrix} \mathbf{e}_{I,t} \\ \mathbf{e}_{P,t} + \beta_0 \end{pmatrix} \tag{3}$$

Where $\mathbf{e}_{I,t}$ and $\mathbf{e}_{P,t}$ are multivariate normal draws $\sim N(0, \tau_I^2 \mathbf{I})$ and $\sim N(0, \tau_P^2 \mathbf{I})$. $\mathbf{H}_{I,t}$ and $\mathbf{H}_{P,t}$ are selection matrices of ones and zeros which at each year select the locations where there are instrumental/proxy data. β_0 and β_1 are parameters representing the bias and scaling factor of the proxy records relative to the temperatures.

The remaining level is the prior. Weakly informative but proper prior distributions are specified for the scalar parameters and the temperature field for the first year in the analysis. The priors for all parameters except φ are conditionally conjugate, meaning the prior and the posterior distribution has the same parametric form. The Markov-Chain Monte Carlo (MCMC) algorithm known as the Gibbs sampler (with one Metropolis step) is used for the posterior simulation (Gelman et al., 2003). Technical details of the Bayesian inference of BARCAST is found in Tingley and Huybers (2010).

2.4.1. BARCAST extension to probabilistic constraining age-uncertain proxy records

Age-uncertain proxies are associated with an ensemble of chronologies: $\{\mathcal{T}_k, k = 1, 2, \ldots M\}$. BARCAST equation (3) is rewritten so that the proxy observations are conditioned on the selected chronology $\mathcal{T}$, and considered a function of location s, rather than the uncertain time point t.

$$\boldsymbol{W}_s | \mathcal{T}, \boldsymbol{T}_s = \beta_0 + \beta_1 \cdot \boldsymbol{\Lambda}_s^{\mathcal{T}} \cdot \boldsymbol{T}_s + \boldsymbol{e}_s \tag{4}$$

where

$$\boldsymbol{e}_s \sim N\left(\boldsymbol{0}\tau^2 \cdot \boldsymbol{I}\right)$$

constitute the independent normal errors at location s. $\boldsymbol{\Lambda}_s^{\mathcal{T}}$ represents the dependence on the chronology, and replaces $\mathbf{H}_{P,t}$ from Eq. (3). Elements are selected from the true temperature field $\boldsymbol{T}_s$ corresponding to the proxy observations $\boldsymbol{W}_s$, depending on the chronology $\mathcal{T}$.

A new conditional posterior is defined to update the probabilities on the members of the ensemble $\mathcal{T}_k$, see Werner and Tingley (2015) for technical details.

Overall, the BARCAST model parameters are unchanged, but the inference on the conditional posteriors now relies also on the currently selected chronology. The climate process and scalar parameters are sampled at each iteration step of the MCMC, and a chronology is then selected according to the conditional posterior.

The so-called "waiting time dilemma" is a deficiency of the described BARCAST extension, implying that the MCMC sampler will quickly iterate towards a limited number of the chronologies that are most compatible with the initial climate estimate. Running a Metropolis-Coupled MCMC sampler (MC^3) compensates this problem without other complications (Werner and Tingley, 2015, and references therein). The Metropolis coupling allows parallel tempering with several "heated" chains, allowing more rapid exploration of the probability space, coupled with one unheated chain to retain a stationary distribution of the MCMC (Wong and Liang, 1997; Li et al., 2009). Coupling is enabled by state-swapping between neighboring chains, i.e. the current values of all parameters, temperatures and the selected chronology may be swapped between parallel runs at certain intervals.

Finally, note that the scalar response parameters β_0, β_1, and τ_P are replaced with vectors in this version. This is in line with the assumption that individual proxies respond differently to surface temperature across the reconstruction domain.

2.5. The analogue method

The analogue method was first introduced for weather fore-casting (Lorenz, 1969), but has since then been modified and applied for reconstruction purposes. The method has been successfully applied to reconstruct SST during the Dansgaard-Oeschger events 5−8 (Jensen et al., 2018), surface air temperatures and precipitation during more recent time periods (Franke et al., 2011; Gomez-Navarro et al., 2015; Gómez-Navarro et al., 2017; Talento et al., 2019). It relies on the concept of combining proxy data and a pool of instrumental or model information identifying, for each time slice in the reconstruction period, one or several time slices in the instrumental or simulated period that could be considered as a potential analogue. In this manuscript, we use proxy and instrumental data, and denote this basic implementation as the "classic" analogue method. Along this we introduce a variation taking age-uncertainties into consideration. In a real-world application this would result in an analogue reconstruction independent from climate model simulations, in contrast to the version using a model-based pool of analogues.

2.5.1. Classic analogue method

To reconstruct the SST field at time t, the set of predictor variables consists of the individual proxy records at time t.

As a first step, a distance metric between time t and each t_i in the instrumental period is defined. We select the distance between t and t_i as the Euclidean distance between the vectors of proxy data at times t and t_i:

$$d(t, t_i) = \sqrt{\sum_{i=1}^{K} \left(\text{Prox}(l_j, t) - \text{Prox}(l_j, t_i)\right)^2} \qquad (5)$$

where $\text{Prox}(l_j, t)$ denotes the proxy information at location l_j and time t. $l_1, ..., l_K$ denote the proxy locations ($K = 12$).

Second, each time slice of the instrumental period is sorted according to its distance to t. Third, the N closest time-steps of the instrumental period are selected as analogues. Finally, the SST

reconstruction at time t is defined as the average of the instrumental SST fields in the N analogues $t_1, ..., t_N$:

$$\text{Reconstruction}(t) = \frac{\text{SST}(t_1) + \cdots + \text{SST}(t_N)}{N} \qquad (6)$$

The number of analogues, N, could range between 1 and the total number of time slices in the instrumental period (156 in the present study). As N grows, there is a tendency for the reconstruction to become more similar to the instrumental-period mean. We select $N = 20$, as it provides the best reconstruction skill results across the different scenarios tested in this paper.

The analogue method can be used to generate an ensemble reconstruction in the case proxy data exhibit age uncertainties. The method is run repeatedly, each time randomly selecting a realistic set of chronologies for the proxies as described in Sect. 2.6.

2.6. Constructing chronology ensembles

Age-uncertainties are included in the pseudo-proxy experiment scheme in order to validate the reconstruction skill for a marine multi-proxy network. We set up one experiment with all pseudo-proxy records subject to age-uncertainties similar to layered-counted proxies. This setup is simplistic rather than realistic: we use the most simple chronology modeling scheme in this initial analysis, which achieves the goal to quantify the reconstruction skill given an age-uncertain data network.

Ensembles of chronologies are generated using the BAM time-scale modeling scheme as in Werner et al. (2018) and Comboul et al. (2014). BAM models the timescale counting procedure as a super-position of two cumulative Poisson processes with error rates {Θ_1, Θ_2}. Between the years i and $i + 1$, the stochastic processes ($P_i^{\Theta_1}$) and ($P_i^{\Theta_2}$) counts the number of missing years or double-counted years respectively (Comboul et al., 2014). As a conservative estimate, we assume for our pseudo-proxy experiments that annual layers are missing or double counted by the same probability ($\Theta^1 = \Theta^2 = 0.02$). These probabilities are identical to Werner and Tingley (2015) for direct comparison. 2% is also the average probability of miscounting in the lake sediments and ice core proxy records used for BARCAST reconstruction in Werner et al. (2018).

The time increment δ_i is given by

$$\forall i \in [1, n-1], \delta i = 1 + P_i^{\Theta_1} - \min\left(P_i^{\Theta_2}, 1\right) \qquad (7)$$

Data and time index i is removed and considered missing from the pseudo-proxy record if $\delta_i > 1$. On the contrary, data is duplicated to the record at time i if $\delta_i < 1$. The time increment δ_i will be exactly one most of the time, and the chronology is unchanged in this case.

One might expect an overall zero-offset due to the symmetry of the error rates, but this is a potential pitfall as demonstrated by Comboul et al. (2014) and in our Fig. 3. The mismatch between an ensemble of 1000 age models and the true age is illustrated for one arbitrary pseudo-proxy in this figure. As layer counting starts from the top or the most recent date, the counting errors accumulate back in time. The possible offset for early dates may have profound effects on the resulting reconstruction product.

2.7. Metrics for reconstruction skill

The deviations between the reconstruction and the simulation target can be measured and quantified using a range of metrics. Using a proper scoring rule is advantageous in the case of a probabilistic forecast, securing that the maximum reward is given when the true probability distribution is reported (Gneiting and Raftery, 2007). The continuous ranked probability score (CRPS) is a proper
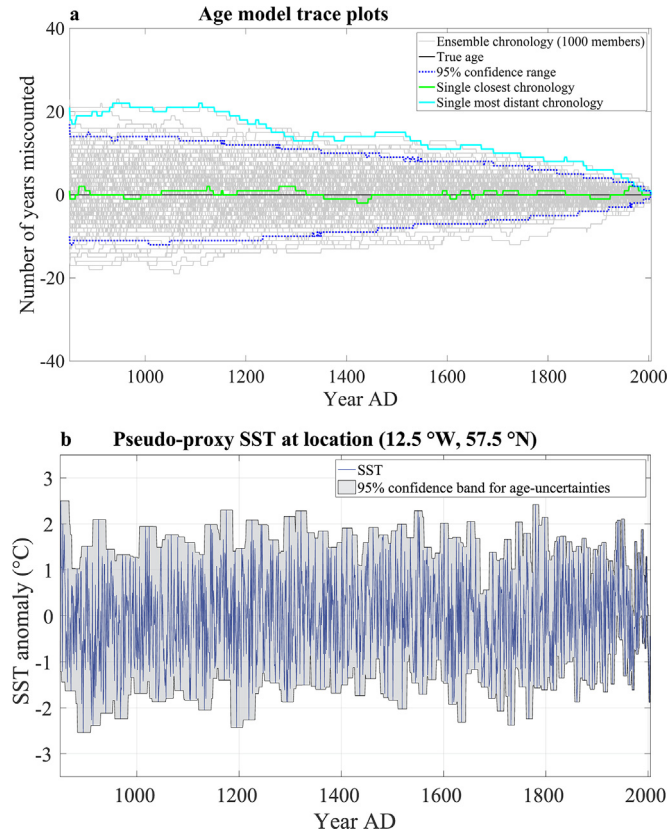
Fig. 3. (a) Trace plots of the age model ensemble generated for one arbitrary pseudo-proxy located at 12.5°W, 57.5°N. The mismatch between true age and counted age is shown on the y-axis, as a function of time. The blue, dotted lines constrains the 95% confidence range of the age models. The trace plot plotted in green (cyan) represents the single age model which is closest (most distant) from the true age, respectively. (b) Time series for the same record plotted in blue. The shaded grey area denotes the 95% confidence range for the age-uncertainties.

scoring rule, and will be used in the following to quantify the reconstruction skill for the BARCAST reconstruction ensemble. The often-used reduction of error (RE) is not a proper scoring rule, and will not be used. RE = 1 implies a deterministic forecast, but the maximum score is obtained when the mean (a point measurement) within the probability distribution P is used instead of the predictive distribution P itself. Additionally and for both reconstruction types, we calculate the Pearson's correlation coefficient and the root-mean-squared error (RMSE) with the true target, in order to have skill measures more suited for comparison.

The concept behind the CRPS is to provide a metric of the distance between the predicted (forecasted) and occurred (observed) cumulative distribution functions of the variable of interest. The lowest possible value for the metric corresponding to a perfect forecast is therefore CRPS = 0. Following Hersbach (2000), the definition of the CRPS can be defined as follows:

$$CRPS(P, x_{target}) = \int_{-\infty}^{\infty} \left[P(x) - u(x - x_{target}) \right]^2 dx \quad (8)$$

Where x is the variable of interest, x_{target} denotes target (validation) data, u is the unit step function and $P(x)$ is the cumulative distribution function of the forecast ensemble with a probability density function (PDF) of $\rho(y)$:

$$P(x) = \int_{-\infty}^{x} \rho(y) dy \quad (9)$$

$\overline{CRPS}$ denotes the temporal average CRPS. $\overline{CRPS}$ is the sum of the average reliability and the average potential CRPS:

$$\overline{CRPS} = \overline{Reli} + \overline{CRPS}_{pot} \quad (10)$$

Where the $\overline{Reli}$ can be interpreted as the validity of the uncertainty bands, or the mean-squared error of the confidence intervals. The $\overline{CRPS}_{pot}$ measures the accuracy of the reconstruction, quantifying the spread of the ensemble and the mismatch between the best estimate and the target variable. See Gneiting and Raftery (2007) for details and elaboration on the CRPS.

3. Results

As mentioned above, pseudo proxy data was generated with three different SNR (10, 1, 0.5), with the last one being the most realistic (albeit still optimistic) experiment. For SNR = 0.5, the experiment was repeated with chronological uncertainties. The reconstructions were generated using annually and decadally resolved input data. For each SNR experiment, the successful BARCAST ensemble of reconstructions consists of ~ 250 members. Similar ensembles are generated for the analogue method when implementing chronological uncertainties. The following analyses are performed on the member-basis, but figures show the ensemble means.

3.1. Skill metrics

The skill results are shown for the Summer mean reconstructions in Figs. 4–6. Note that Figs. 5 and 6, depicting RMSE and CRPS, use the same color bar as in the cross correlation analysis Fig. 4, but best skill is achieved where the RMSE/CRPS is low as opposed to best skill for the highest cross correlations.

Significant correlations between 0 and 1 are plotted in color in Fig. 4. A p-value of 0.05 was used for significance testing, and all significant correlations are positive. The correlations between target and reconstruction are highest directly at the pseudo-proxy locations, and decreases with distance. The correlations also decrease systematically with increasing levels of proxy noise. Regions of minimum significance are detected for the longitude bands ~ 50°N and ~ 82°N for pseudo-proxies without chronological errors (Fig. 4a–f). BARCAST is unable to produce skillful reconstructions when all proxies exhibit chronological errors (Fig. 4g), while the analogue method reconstructs significant, albeit low correlations for 95% of the grid cells (Fig. 4h). The analogue method outperforms BARCAST for all SNR experiments in terms of area average correlation, visualized by the figure box plots. The correlation skill is stable within the BARCAST reconstruction ensembles, with maximum difference in correlation skill of 0.024. The intra-ensemble skill variance is highest away from pseudo-proxy locations. The analogue-chronology ensemble has a maximum intra-ensemble skill difference of 0.31.

Different probabilities for miscounting have been tested for the simulation of chronology ensembles. Changing P1 and P2 from 2% to 5% results in a decreased reconstruction skill for the analogue method, but only minor changes in significance levels. Changing P1 and P2 to 1% is not enough to gain significant skill for BARCAST.

We observe similar skill results for the spatially distributed RMSEs (Fig. 5). Unexpectedly, the skill does not decrease
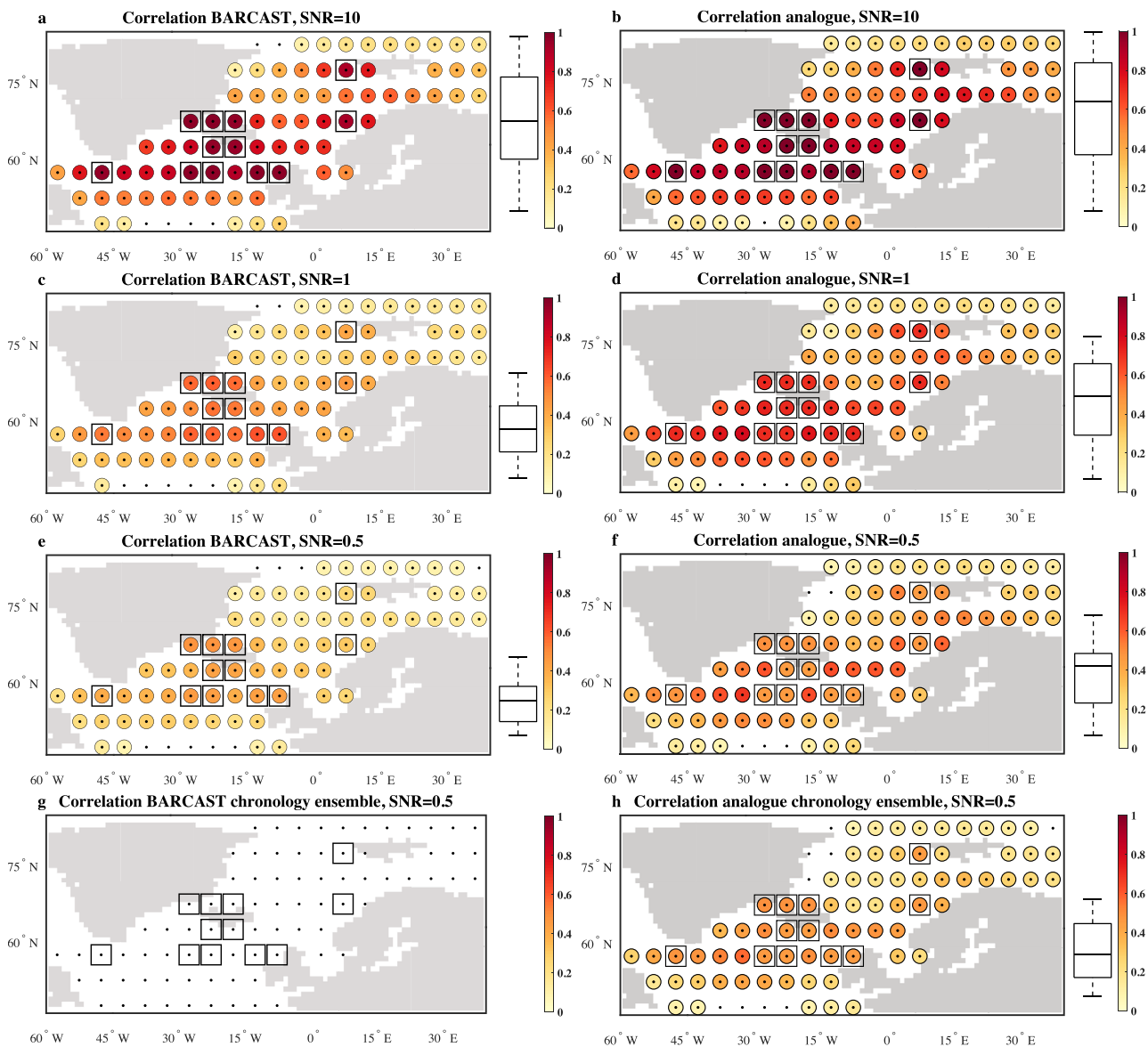
Fig. 4. Colored circles show the correlation between the Summer mean pseudo-reconstruction and the target data, for all signal-to-noise ratios (SNR) by standard deviation. Colors are only shown where the correlation is significantly different from zero. Black dots denote reconstructed grid cells, squares indicate pseudo-proxy locations. Panels (a, c, e, g) are for BARCAST, while panels (b, d, f, h) are for the analogue method. The box plot quantifies the area-average skill.

monotonically for lower SNR, seen for Fig. 5c,e. The analogue method performs better than BARCAST for all SNRs considering the area average skill metric. The intra-ensemble RMSE metrics vary by < 0.035. The BARCAST scores of correlations and RMSEs for SNR = 0.5 are inferior to those obtained in Werner et al. (2013), where the reconstruction target was terrestrial surface air temperature.

Lastly, $\overline{\text{CRPS}}_{\text{pot}}$ and $\overline{\text{Reli}}$ are shown in Fig. 6. All CRPS scores are limited to <1 degree K, with the area average $\overline{\text{CRPS}}_{\text{pot}}$ ranging from 0.45 K to 0.8 K (Fig. 6a, c, e, f). The area average $\overline{\text{Reli}}$ is better than 0.4 K for all SNR experiments (Fig. 6b, d, f, h).

3.1.1. Alternative experiment setups

We dedicated specific experiments to implement (a) decadal resolution of all input data and the reconstruction, and (b) an initial clustering step to the data before applying the reconstruction

methodologies. The motivation for these implementations is the potential of increased reconstruction skill, and/or reduced computation time for BARCAST.

All input data and the chronology ensembles are generated on the annual scale before averaging to decadal resolution. Chronologies were simulated using BAM, as for the annual experiments. The analogue reconstructions are successful using this setup, but BARCAST fails to produce meaningful posterior parameter estimates. Correlation skill results are therefore only shown for the analogue method (Fig. 7), while BARCAST is deemed unsuitable for this particular configuration of decadal SST data. Selected grid cells in the decadal analogue reconstructions are significantly negatively correlated with the target (Fig. 7). In general, the box plots show that the area-average skill is lower than for annual data, and correlations across the region have a larger spread. The number of grid cells without significant correlation is higher.

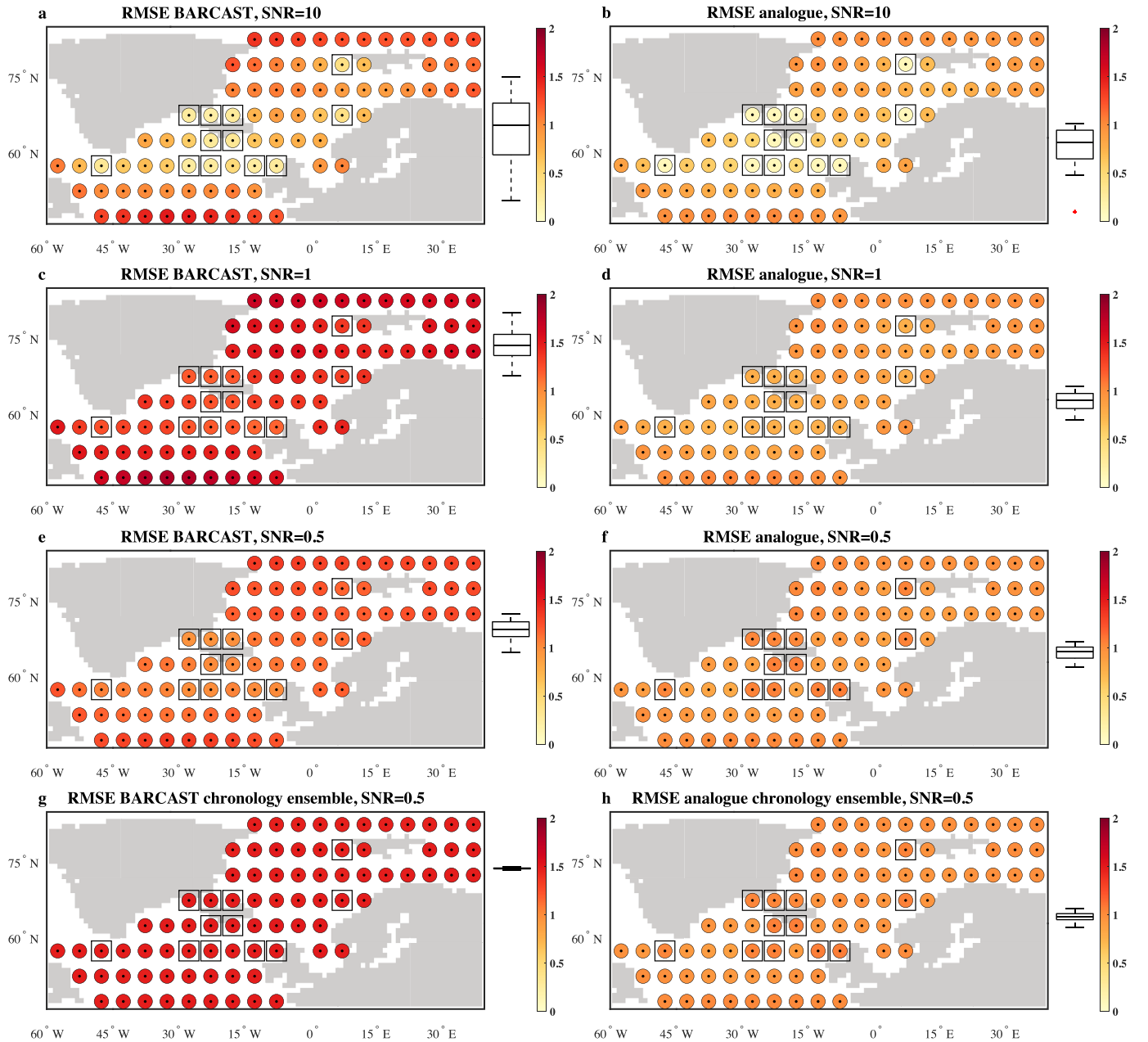The clustering separates the geographical domain of the

Fig. 5. As in Fig. 4, but for the root-mean-square error (RMSE). RMSE values are in temperature units (K).

northern North Atlantic into C clusters, based on similarities in the SST regimes during the instrumental period. The distance between two geographical points is defined as one minus the correlation between the instrumental SST time series at these locations. Then, the analogue method or BARCAST is applied inside each of the C clusters, independently. Finally, the reconstructions obtained within each cluster are merged together, producing a reconstruction covering the whole geographical domain. The results show no improvement in skill when clustering is implemented to the reconstruction schemes. In fact, the BARCAST skill is deteriorated, and we therefore regard the clustering approach as redundant for the SST reconstructions in this study.

In summary, neither the decadal resolution nor the initial clustering step offers advantages for either method in this pseudo proxy setting. The remaining results therefore concern annual Summer mean only.

3.2. BARCAST posterior parameters

The mean values for the posterior distributions of the BARCAST model parameters are listed in Table 1. The SNR of the BARCAST reconstruction is estimated from the posterior parameters:

$$SNR_{rec} = \frac{\beta_1^2 Var(T_t)}{\tau_P^2} \tag{11}$$

The reconstructed SNR is not monotonically decreasing: $SNR_{rec} = 7.2, 1.8, 2.18$ and 0.01. Together with the overestimated $SNR_{rec} = 2.18$ we also find an inflated estimate of the instrumental data error variance $\tau_I^2 = 0.43$ and interpret this as BARCAST incorrectly ascribing measurement errors to the SST predictors for the instrumental period.

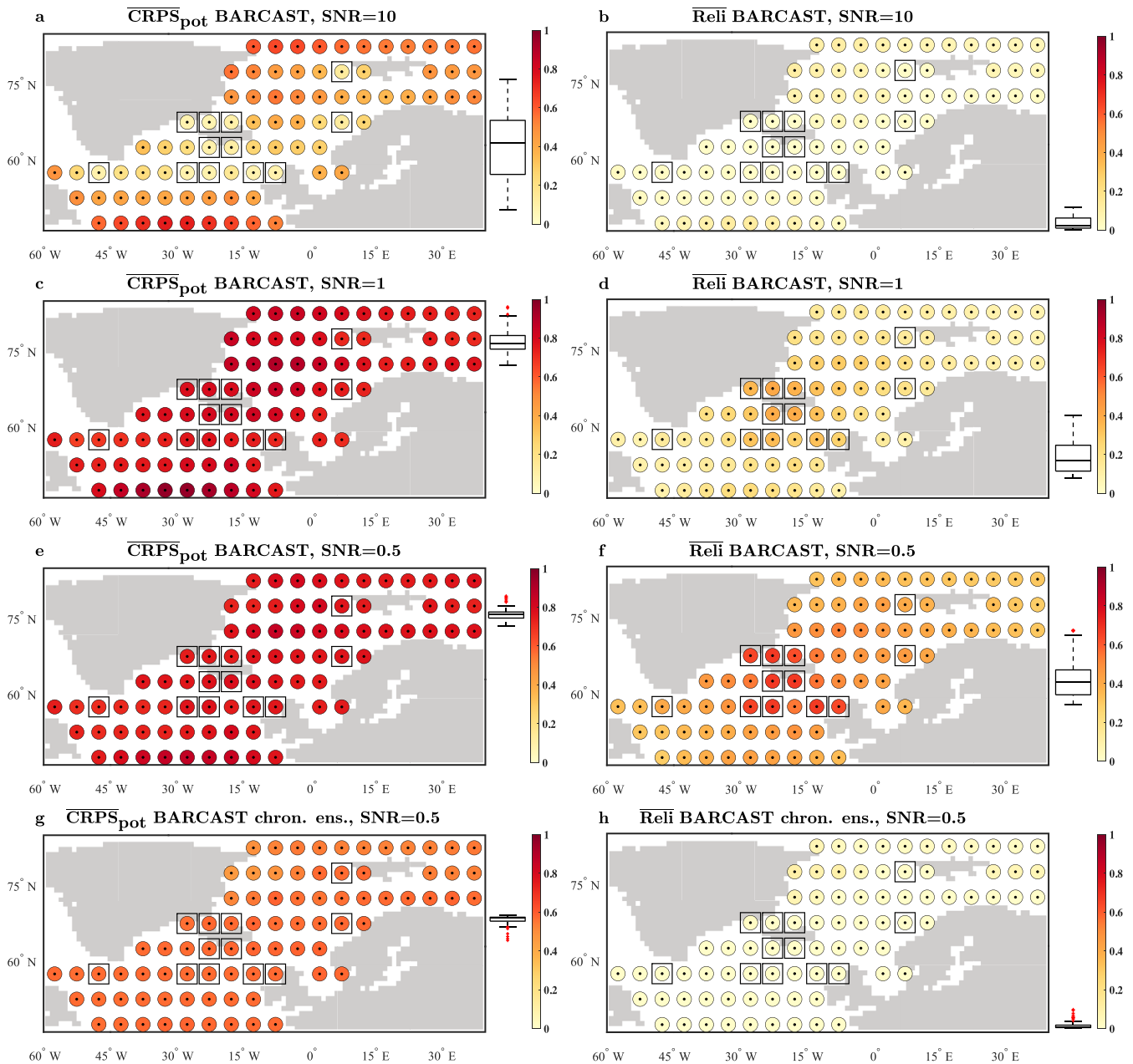The estimated spatial e-folding distance $1/\varphi$ ranges from

Fig. 6. As in Fig. 4, but for the continuous ranked probability (C RPS$_{pot}$ and R eli) only for BARCAST. CRPS values are in temperature units (K).

2115 km for SNR = 10, to 8141 km for SNR = 0.5. Ideally, the spatial correlation structure should not be influenced so strongly from the proxy noise. The teleconnections observed for the input spatial network are likely contributing to this instability.

3.2.1. The AR(1) parameter α

The posterior mean AR(1) coefficient α ranges between 0.08 and 0.12 for the SNR experiments. α can be interpreted as a measure of temporal persistence, where α ~ 0 is similar to a white noise stochastic process. Hence, BARCAST reconstructs the North Atlantic SSTs as very weakly persistent in time. The parameter estimate is unexpected because the ocean has a much larger heat capacity than the atmosphere, and responds more slowly to external forcing. The expected α estimate should be similar, or even higher than that of terrestrial BARCAST reconstructions, where α ~ 0.4 (Tingley and Huybers, 2010; Luterbacher et al., 2016), and α ~ 0.8 (Werner

et al., 2018).

Searching to explain the lack of temporal persistence, we compare the BARCAST posterior α estimates with estimates for the real-world HadSST data, for the simulated model data, and those for the analogue reconstruction. The AR(1) parameter is estimated using the least-squares method and JJA mean SST data. For the instrumental HadSST data set, the subpolar and Arctic regions are missing more data points than the temperate region during the earlier part of the record (figure not shown). The area average α is estimated to 0.32 when data from the full time period 1850−2018 is included (Fig. 8a). The estimate increases to 0.48 when only the past 50 years of data are used for estimation, with better temporal coverage (Fig. 8b).

The AR(1) coefficients for the CESM-LME simulation data are estimated on the grid-cell level using the full simulation period 850-2005 AD. This period is most realistic when comparing our
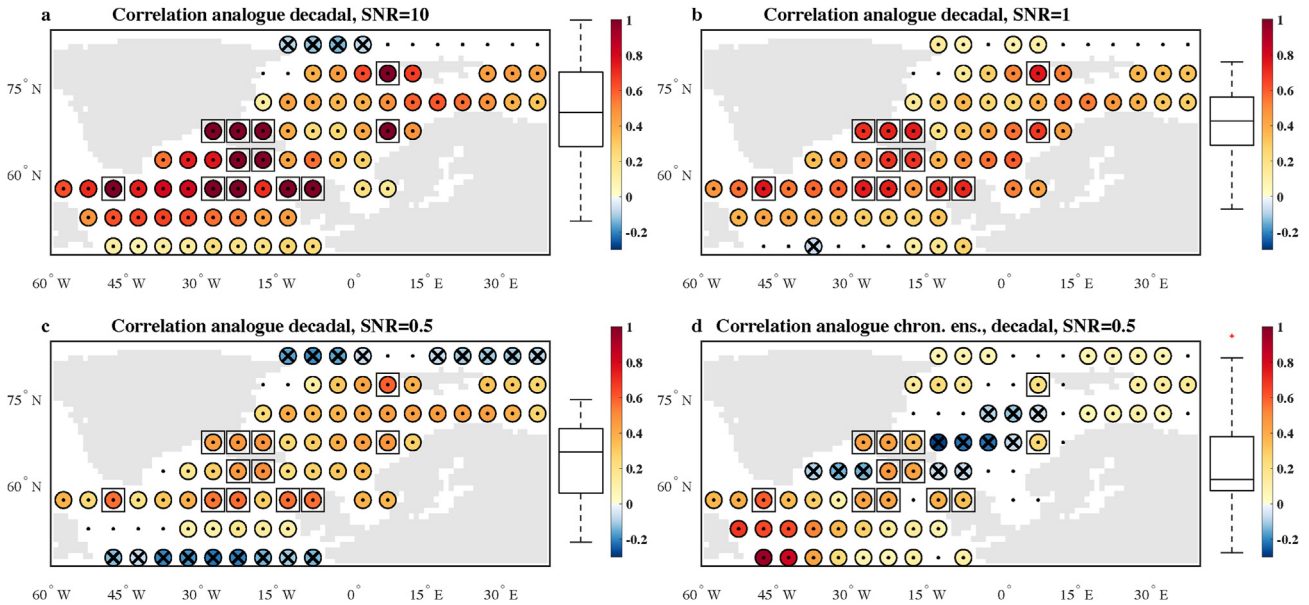
Fig. 7. Colored circles show the correlation between the decadal pseudo-reconstruction and the decadal target data, for all signal-to-noise ratios (SNR) by standard deviation. Colors are only shown where the correlation is significantly different from zero. Black dots denote reconstructed grid cells, squares indicate pseudo-proxy locations. Crosses indicate significant negative correlations. The box plot quantifies the area-average skill.

Table 1
Mean of posterior distribution for the BARCAST parameters.

SNR$_{target}$	SNR$_{rec}$	α	μ	σ^2	$1/\varphi$	τ_I^2	β_0
10	7.2	0.1	0	1.02	2115	0	−0.03
1	1.766	0.09	0	1.9	4222	0	0.06
0.5	2.18	0.12	0	0.82	8141	0.43	0.03
0.5 (chronological errors)	0.01	0.08	0	1.17	2487	0	0.01

BARCAST posterior estimate, although it makes the comparison with HadSST less clear. Spatial resolution of 1° by 1° or 5° by 5° gives nearly identical estimates of the AR(1) coefficients. The area average CESM model run AR(1) coefficient is estimated to 0.20 (Fig. 8c). As a validation exercise, we perform the same analysis for one individual paleoclimate simulation of the MPI-ESM-P (past-1000 r1, Jungclaus et al. (2014)) (Fig. 8d). The area average α is 0.26 for that run, hence the simulated SST for both model runs exhibit less temporal persistence than estimated for the existing land-based BARCAST reconstructions. At the same time, the model SST
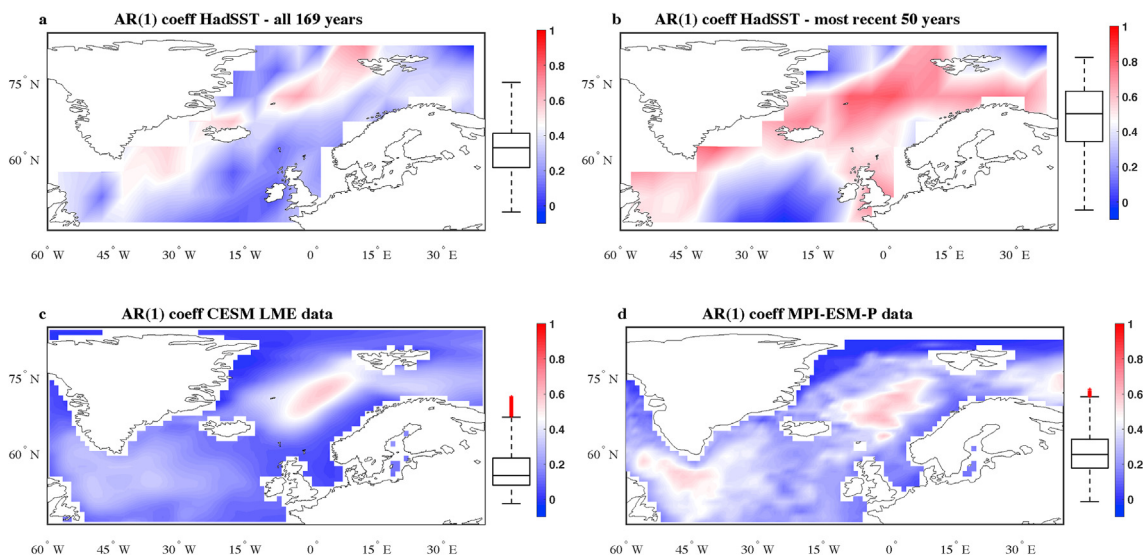
Fig. 8. Maps showing the AR(1) coefficient estimated on the grid-cell level from the instrumental HadSST data (a and b), and two millennium-long climate model simulations (c and d). The coefficients are plotted in color. The area-average coefficient is denoted by the box plot for each panel.

data exhibit a higher value of α than the resulting BARCAST pseudo-reconstruction ensembles.

The area average AR(1) coefficient of the analogue reconstructions vary between 0.11 and 0.21 for the different SNR experiments (figure not shown).

3.2.2. Suitability of the BARCAST AR(1) assumption for CESM SST data

It is possible that the BARCAST AR(1) model assumption itself is improper for the North Atlantic CESM target SST data, and that violation of the BARCAST assumptions causes the underestimation of the posterior AR(1) coefficient α. The AR(1) process is preferred for BARCAST calculations since it facilitates a reconstruction problem without supreme complexity. Previous studies show that a long-range memory (LRM) stochastic model such as the fractional Gaussian noise (fGn) is more consistent with local instrumental SST observations and SSTs from control-run simulations over this region than an AR(1) model (Fredriksen and Rypdal, 2016; Løvsletten and Rypdal, 2016).

An LRM stochastic process exhibits an autocorrelation function (ACF) and a power spectral density (PSD) of a power-law form: $C(t) \sim t^{\beta-1}$, and $S(f) \sim f^{-\beta}$ respectively. The spectral exponent β determines the strength of the persistence. Persistence properties of the target data are analyzed in the spectral domain using the periodogram as the estimator. Hypothesis testing is used to identify the most suitable statistical model similar to Nilsen et al. (2018). See Appendix B for details on how the periodogram is estimated and on the hypothesis testing.

The hypothesis testing provides important information regarding the temporal persistence of the target CESM SST data on the 5° by 5° grid-cell level. The persistence properties vary across the region, and there is no single preferred statistical model for all grid cells. In fact, the results cover all possible outcomes approximately equally: data points consistent with the AR(1) hypothesis only, data points consistent with the fGn hypothesis only, data points consistent with both null hypotheses, and data points inconsistent with both null hypothesis. The AR(1) hypothesis is most appropriate in the region illustrated by red color in Fig. 8c. Weak LRM is identified if the fGn model is preferred, $\beta \leq 0.2$.

Results for one arbitrary grid cell off the Norwegian coast are shown for illustrative purposes in Fig. 9. The AR(1) coefficient α and the LRM parameter β are estimated. These parameters are used to generate Monte Carlo ensembles of AR(1)/fGn processes, and spectra for the 95% confidence ranges of the two processes are plotted together with the CESM target power spectrum (Fig. 9b and c). In this particular location, the CESM power spectrum is consistent with the AR(1) process spectra (Fig. 9b), and not with the fGn process spectra (Fig. 9c). The two processes exhibit spectral differences on all frequencies, visible when plotted together (Fig. 9d).

As a final remark on the results presented in this study, we note that the skill metric results and posterior parameter estimates are not automatically transferable to other ocean regions than the northern North Atlantic. For instance, the temporal covariability of the tropical Pacific SST are strongly influenced by the El Nino Southern Oscillation (ENSO), in addition to the year-to-year variability. Previous results show that the tropical Pacific SSTs are consistent with an autoregressive process of order 1 (AR(1)) null-model (Løvsletten and Rypdal, 2016). In contrast, our results demonstrate only around 25% of the North Atlantic SSTs are consistent with the AR(1) null-model. The tropical Pacific therefore complies better with the BARCAST model assumption. When other conditions are equal, the reconstruction skill may hypothetically be higher for the Pacific region than the North Atlantic.

4. Discussion

This study lays the groundwork for SST field reconstruction covering the northern North Atlantic region, using a marine proxy network subject to age uncertainties. The main objectives have been to (1) test the reconstruction skill of BARCAST and the analogue reconstruction methods using pseudo-proxy experiments, in order to determine whether one or both of the methodologies are suitable for SST field reconstruction. Objective (2) was testing how age-uncertainties in the marine records influence the reconstruction skill.

We find that both CFR methodologies generate skillful annually resolved reconstructions when the input data is perfectly dated (Figs. 4-6a-f). The reconstruction skill decreases away from proxy locations, and decreases with increased levels of proxy noise. The intra-ensemble skill variance is highest for the chronology ensemble version of the analogue method, confirming that repeated runs with random age-model selection actually inflate the reconstruction uncertainties as described in Werner and Tingley (2015). Overall, the analogue method performs better than BARCAST considering both the correlation and the RMSE skill metrics (Figs. 4 and 5). The superiority of the analogue method is likely linked to the flexibility in the modeled proxy-temperature relationship, as the method does not assume a stationary, linear relationship between proxies and the SST signal. Experimenting with decadal resolution and spatial clustering does not improve the reconstruction skill. BARCAST fails to reconstruct decadal temperatures, and deteriorates the skill in the case of clustering. The negative results are in line with Talento et al. (2019), yet we highlight the importance of performing these exercises because the spatiotemporal properties of SSTs differ from air temperatures (Fredriksen and Rypdal, 2016).

As for all CFR products, the reconstructions in our study would benefit from a higher-density proxy network. However, high-
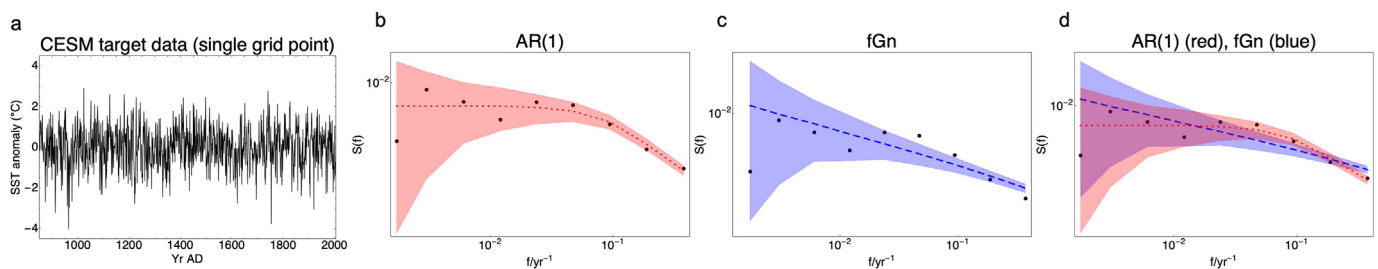
Fig. 9. (a) Arbitrary SST time series of the CESM target data. (b) Log-log power spectra for the time series in (a) shown as black dots, and 95% confidence range based on Monte Carlo ensembles of AR(1) processes shown as red shading. The dashed line marks the ensemble mean of the shaded spectra. (c) as in (b), but for fractional Gaussian noise (fGn) as blue shading and dotted line marks the ensemble mean. (d) Plots (b−c) together.

quality records are preferred for supplementing the network. We obtain significant skill for most of our experiments using this network (Figs. 4–6), so the spatial information in the observation data is considered satisfactory for our purpose.

In addition to the skill results, the posterior estimates of BARCAST parameters also provide indications of the reconstruction quality. The reconstructed temporal covariance is low compared with the target data, the ensemble mean posterior AR(1) parameter $\alpha \leq 0.12$. The reason behind this is unclear, but the spatial teleconnections in the target data could influence the temporal structure due to the coupled spatiotemporal model formulation of BARCAST.

Objective 2 is elaborated because BARCAST is unable to reconstruct skillful annual fields of SST when all proxies are subject to chronological errors in form of miscounted layers (Fig. 4g). This is discouraging in the context of real-world SST reconstruction, since chronological errors occur in the majority of the available marine proxy records. The earlier successful BARCAST reconstructions that implemented chronological uncertainties had the advantages of being applied exclusively to a terrestrial data network and reconstruction domain, in addition to either an evenly sampled domain (Werner and Tingley, 2015), or combined with a dominant proportion of absolutely dated proxies (31 out of 44 proxy records absolutely dated in Werner et al. (2018)). While our marine proxy network has none of these advantages, the three bivalve mollusc records stand out as absolutely dated (site IDs 4, 16 and 17, Table A1). The proxy development for climate reconstruction is relatively new for bivalves, and has been rapidly progressing over the past decade (Butler et al., 2019). Independent of the CFR methodology, prospects are good for a future higher-skill marine reconstruction under the condition that a predominantly absolutely dated proxy network is applied.

Improved annual reconstruction skill will be difficult to achieve using a modified chronology modeling scheme such as BACON instead of BAM, although this modification would imply more realistic age model ensembles. Further work using BACON for age-depth modeling could house potential for marine CFR exhibiting decadal temporal resolution. There are other measures easily available to improve the annual reconstruction skill for the analogue method. The analogue implementation can be modified to construct the pool of analogues from climate model simulations instead of instrumental data. An optimal model pool comprises independent simulations originating from different climate models. This strategy avoids potential influence on the reconstruction skill from model bias. The reconstruction skill is expected to increase with a larger analogue pool (Bothe and Zorita, 2021), with the cost that the model-dependent analogue reconstruction cannot be used for traditional proxy-model comparison studies (PAGES 2k-PMIP3 group, 2015; Moffa-Sánchez et al., 2019). The size of the analogue pool may differ by as much as two orders of magnitudes for the two options. The analogue method was implemented successfully for land temperatures over Europe in Bothe and Zorita (2020), facilitating slightly less than 10 000 model analogues. In comparison, there is only one observed instrumental data set covering the period back to 1850. Using the analogue reconstruction method it is possible to rank individual model simulations by their ability to replicate the climate signal observed in a known proxy network. This is a useful model evaluation tool, applicable when the analogue pool consists of climate model time slices. One strategy for model evaluation is outlined in Gómez-Navarro et al. (2017), using an analogue pool of 16 simulations from seven Earth System Models and the surface temperature proxy archive of PAGES2k Consortium (2013). For every year to reconstruct, the study finds that the closest analogue time slices

were selected from each of the simulations at some point, but two simulations were selected more often than the rest, and two others were selected less frequently. The target simulation is selected as one of several ensemble members from a single model, and the remaining ensemble simulations are not selected more often than simulations from other models.

Finally, extending the analogue method to search for analogues in the space of empirical orthogonal functions (EOF) is another promising option to potentially increase reconstruction skill in the North Atlantic Ocean (Gómez-Navarro et al., 2017).

5. Conclusions

Pseudo-proxy experiments are helpful tools for ranking the reconstruction skills of the BARCAST and analogue reconstruction methodologies in this study. The northern North Atlantic marine proxy network is subject to age-uncertainties and proxy noise, and our analyses demonstrate that the analogue method is superior to BARCAST for SST reconstruction under these premises. A clear deficiency is that observed teleconnections in the climate model simulation SSTs are unaccounted for in the spatiotemporal structure of the BARCAST model assumptions.

Our results suggest that a future annually resolved SST reconstruction based on the BARCAST methodology should rely on a marine proxy network with dominance of absolutely dated records, and preferably uniform sampling in space. A smaller reconstruction region is also recommended without strong teleconnections. The analogue method is computationally cheap to run, allowing users to quickly test different modifications and identifying the optimal implementation for their specific use. The implementation for northern North Atlantic SST reconstruction will benefit from exploring a larger analogue pool based on climate model simulations.

In this study we bridge the gap between terrestrial and marine CFR for the common era. The research topic benefits from exploring a range of advanced statistical tools, the analytics are expected to become wider used in CFR applications after our systematic demonstrations. The experiment results contribute to constraining uncertainties of natural ocean variability influencing climate across the globe. Further CFR comparison with climate model runs also provides an important test bed for understanding multidecadal to centennial climate variability and the climate sensitivity to external forcing, while providing an extended context for anthropogenic warming prior to the instrumental era.

Author statement for manuscript nr JQSR-D-21-00136

All authors have contributed to the completion of the revised manuscript, and approve the final submitted version: Tine Nilsen (TN), Stefanie Talento (ST), and Johannes Werner (JW). TN and JW designed the study. TN has performed the majority of analyses,

produced all figures and tables, and wrote the manuscript draft. ST contributed with source code and analyses for the analogue method. JW has provided BARCAST source code and infrastructure for public access to the code in GitHuB. All authors have worked together to revise the manuscript text.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. List of North Atlantic proxy records

Table A1 lists the proxy records used as basis for the applied pseudo proxy network.

Table A1

List of proxy data series for the reconstruction region, including PAGES Ocean 2k records (McGregor et al., 2015). The proxy data locations are used for the design of the pseudo-proxy network (Fig. 1b).

Site ID	Site name	lat ($°$N), lon ($°$E)	Proxy type for SST rec	Time (yr CE)	References
1	Reykjanes Ridge	58.56, −20.24	Diatom assemblages	18−1582	Berner et al. (2008)
2	Vøring Plateau	66.97, 7.64	Diatom assemblages	4−1995	Berner et al. (2011)
Ocn_055	West Spitzberg, Fram Strait	78.91, 6.77	Dinocyst assemblages	−289−1943	Bonnet et al. (2010)
4	North Icelandic Shelf	66.32, −18.12	Bivalve growth increments	649−2005	Butler et al. (2013)
Ocn_032	MD95-2011	66.97, 7.64	Alkenones (UK37)	10−1440	Calvo et al. (2002)
Ocn_040	ODP984	61.43, −24.08	Foraminifera δ^{18}O and Mg/Ca	152−1396	Came et al. (2007)
7	Gardar Drift	57.26, −27.54	Foraminifera Mg/Ca	70−1390	Farmer et al. (2008)
8	North Icelandic Shelf	66.51, −17.70	Diatom assemblages	15−1875	Jiang et al. (2015)
9	North Icelandic shelf	66.63, −23.85	Diatom assemblages	21−1966	Justwan et al. (2008)
10	Reykjanes Ridge	58.76−25.96	Alkenones (UK37)	100−1260	Marchal et al. (2002)
11	Reykjanes Ridge	57.27, −27.55	Diatom assemblages	2−2004	Miettinen et al. (2012)
12	Southeast Greenland Shelf	67.14, −30.83	Diatom assemblages	6−1910	Miettinen et al. (2015)
13	Iceland Basin	61.29, −19.32	Foraminifera δ^{18}O and Mg/Ca	818−1793	Moffa-Sánchez et al. (2014)
14	Erik Drift	57.30, −48.43	Foraminifera δ^{18}O and Mg/Ca	763−1914	Moffa-Sánchez et al. (2014)
15	Ísafjarðardjúp fjord, Northwest Iceland	66.14,−23.16	Alkenones (UK37)	4−1640	Moossen et al. (2015)
16	North Icelandic Shelf	66.32, −18.12	Bivalve δ^{18}O	953−2000	Reynolds et al. (2016)
17	British Continental Shelf	56.40, −6.20	Bivalve δ^{18}O	1799−2010	Reynolds et al. (2017)
Ocn_020	Feni Drift	55.65, −13.98	Foraminifera δ^{18}O and Mg/Ca	−368−1998	Richter et al. (2009)
Ocn_039	MD992275	66.55, −17.42	Alkenones (UK37)	1−2001	Sicre et al. (2011)
Ocn_058	Western Svalbard	78.91, 6.78	Foraminifera Mg/Ca	−120−2007	Spielhagen et al. (2011)
Ocn_045	RAPiD-12-1K	62.08, −17.82	Foraminifera Mg/Ca	−9918−1950	Thornalley et al. (2009)
22	Fram Strait	79.90, 5.20	Foraminifera assemblages	−12000−1992	Werner et al. (2016)

Appendix B. Spectral analysis of temporal persistence and hypothesis testing

The periodogram is defined here in terms of the discrete Fourier transform H_m as (Malamud and Turcotte, 1999):

$$S(f_m) = \left(\frac{2}{N}\right)|H_m|^2, \quad m = 1, 2, \ldots, N/2$$

For evenly sampled time series $x_1, x_2, \ldots x_N$. The sampling time is an arbitrary time unit, and the frequency is measured in cycles per time unit: $f_m = \frac{m}{N}$. $\Delta f = \frac{1}{N}$ is the frequency resolution and the smallest frequency which can be represented in the spectrum.

Log-binned power spectra are visualized in log−log plots (Fig. 9) since the spectral exponent β can be estimated by a simple linear fit to the spectrum. Log binning of the periodogram is used for analytical purposes, this representation weights all frequencies equally with respect to their contributions to the total variance.

The first null hypothesis tested is that the target data set can be described using an AR(1) process at all frequencies, with the parameter α estimated from the CESM SST data on the grid-cell level. For testing we generate a Monte Carlo ensemble of AR(1) series with a value of the α parameter identical to the target data. The power spectrum of each ensemble member is estimated, and the confidence range for the theoretical spectrum is then calculated using the 2.5 and 97.5 quantiles of the log-binned periodograms of the Monte Carlo ensemble. The second null hypothesis is that the target data set can be described using the fractional Gaussian noise (fGn) stochastic process at all frequencies, with scaling parameter β estimated from the power spectrum of the target data at grid-cell level. The respective null hypothesis is rejected if the log-binned spectrum of the target data is outside of the confidence range for the AR(1) or fGn model at any point.

References

Berger, A.L., 1978. Long-term variations of daily insolation and Quaternary climatic changes. J. Atmos. Sci. 35 (12), 2362−2367.

Berner, K.S., Koç, N., Divine, D., Godtliebsen, F., Moros, M., 2008. A decadal-scale Holocene sea surface temperature record from the subpolar North Atlantic constructed using diatoms and statistics and its relation to other climate parameters. Paleoceanography 23 (2), PA2210.

Berner, K.S., Koç, N., Godtliebsen, F., Divine, D., 2011. Holocene climate variability of the Norwegian Atlantic Current during high and low solar insolation forcing. Paleoceanography 26 (2), PA2220.

Beugrand, G., Reid, P.C., 2003. Long-term changes in phytoplankton, zooplankton and salmon related to climate. Global Change Biol. 9 (6), 801−817.

Blaauw, M., Christen, J.A., 2011. Flexible paleoclimate age-depth models using an autoregressive gamma process. Bayesian Anal 6 (3), 457−474.

Bonnet, S., de Vernal, A., Hillaire-Marcel, C., Radi, T., Husum, K., 2010. Variability of sea-surface temperature and sea-ice cover in the Fram Strait over the last two millennia. Mar. Micropaleontol. 74 (3), 59−74.

Bothe, O., Zorita, E., 2020. Proxy surrogate reconstructions for Europe and the estimation of their uncertainties. Clim. Past 16 (1), 341−369.

Bothe, O., Zorita, E., 2021. Technical note: considerations on using uncertain proxies in the analogue method for spatiotemporal reconstructions of millennial-scale climate. Clim. Past 17 (2), 721−751.

Butler, P.G., Wanamaker, A.D., Scourse, J.D., Richardson, C.A., Reynolds, D.J., 2013. Variability of marine climate on the North Icelandic Shelf in a 1357-year proxy archive based on growth increments in the bivalve Arctica islandica. Palaeogeogr. Palaeoclimatol. Palaeoecol. 373, 141−151.

Butler, Paul G., Freitas, P.S., Burchell, M., Chauvaud, L., 2019. Archaeology and Sclerochronology of Marine Bivalves. In: Smaal, A.C., Ferreira, J.G., Grant, J., Petersen, J.K., Strand, Ø. (Eds.), Goods and Services of Marine Bivalves. Springer International Publishing, Cham, pp. 413−444.

Calvo, E., Grimalt, J., Jansen, E., 2002. High resolution U37K sea surface temperature reconstruction in the Norwegian Sea during the Holocene. Quat. Sci. Rev. 21 (12), 1385−1394.

Came, R.E., Oppo, D.W., McManus, J.F., 2007. Amplitude and timing of temperature and salinity variability in the subpolar North Atlantic over the past 10 k.y. Geology 35 (4), 315−318.

Christiansen, B., Ljungqvist, F.C., 2017. Challenges and perspectives for large-scale temperature reconstructions of the past two millennia. Rev. Geophys. 55 (1), 40−96.

Comboul, M., Emile-Geay, J., Evans, M.N., Mirnateghi, N., Cobb, K.M., Thompson, D.M., 2014. A probabilistic model of chronological errors in layer-counted climate proxies: applications to annually banded coral archives. Clim. Past 10 (2), 825−841.

Farmer, E.J., Chapman, M.R., Andrews, J.E., 2008. Centennial-scale Holocene North Atlantic surface temperatures from Mg/Ca ratios in Globigerina bulloides. Geochem. Geophy. Geosy. 9 (12), Q12029.

Franke, J., González-Rouco, J.F., Frank, D., Graham, N.E., 2011. 200 years of European temperature variability: insights from and tests of the proxy surrogate reconstruction analog method. Clim. Dynam. 37 (1), 133−150.

Fredriksen, H.-B., Rypdal, K., 2016. Spectral characteristics of instrumental and climate model surface temperatures. J. Clim. 29 (4), 1253−1268.

Gao, C., Robock, A., Ammann, C., 2008. Volcanic forcing of climate over the past 1500 years: an improved ice core-based index for climate models. J. Geophys. Res. Atmos. 113 (D23), D23111.

Gelman, A., Carlin, J., Stern, H., Rubin, D., 2003. Bayesian Data Analysis, second ed. Chapman & Hall, New York, ISBN 1-58488-388-X, p. 668.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. 102 (477), 359−378.

Goldenberg, S.B., Landsea, C.W., Mestas-Nuñez, A.M., Gray, W.M., 2001. The recent increase in Atlantic hurricane activity: causes and implications. Science 293 (5529), 474−479.

Gomez-Navarro, J.J., Werner, J., Wagner, S., Luterbacher, J., Zorita, E., 2015. Establishing the skill of climate field reconstruction techniques for precipitation with pseudoproxy experiments. Clim. Dynam. 45 (5−6), 1395−1413.

Gómez-Navarro, J.J., Zorita, E., Raible, C.C., Neukom, R., 2017. Pseudo-proxy tests of the analogue method to reconstruct spatially resolved global temperature during the Common Era. Clim. Past 13 (6), 629−648.

Graham, N.E., Hughes, M.K., Ammann, C.M., Cobb, K.M., Hoerling, M.P., Kennett, D.J., Kennett, J.P., Rein, B., Stott, L., Wigand, P.E., Xu, T., 2007. Tropical Pacific −midlatitude teleconnections in medieval times. Climatic Change 83 (1), 241−285.

Gray, S.T., Graumlich, L.J., Betancourt, J.L., Pederson, G.T., 2004. A tree-ring based reconstruction of the Atlantic Multidecadal Oscillation since 1567 A.D. Geophys. Res. Lett. 31 (12), L12205.

Hakim, G.J., Emile-Geay, J., Steig, E.J., Noone, D., Anderson, D.M., Tardif, R., Steiger, N., Perkins, W.A., 2016. The last millennium climate reanalysis project: framework and first results. J. Geophys. Res. Atmos. 121 (12), 6745−6764.

Hersbach, H., 2000. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. Weather Forecast. 15 (5), 559−570.

Hurtt, G.C., Chini, L.P., Frolking, S., Betts, R.A., Feddema, J., Fischer, G., Fisk, J.P., Hibbard, K., Houghton, R.A., Janetos, A., Jones, C.D., Kindermann, G., Kinoshita, T., Klein Goldewijk, K., Riahi, K., Shevliakova, E., Smith, S., Stehfest, E., Thomson, A., Thornton, P., van Vuuren, D.P., Wang, Y.P., 2011. Harmonization of land-use scenarios for the period 1500−2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. Climatic Change 109 (1), 117−161.

Jensen, M.F., Nummelin, A., Nielsen, S.B., Sadatzki, H., Sessford, E., Risebrobakken, B., Andersson, C., Voelker, A., Roberts, W.H.G., Pedro, J., Born, A., 2018. A spatiotemporal reconstruction of sea-surface temperatures in the North Atlantic during Dansgaard−Oeschger events 5−8. Clim. Past 14 (6), 901−922.

Jiang, H., Muscheler, R., Björck, S., Seidenkrantz, M.-S., Olsen, J., Sha, L., Sjolte, J., Eiríksson, J., Ran, L., Knudsen, K.-L., Knudsen, M.F., 2015. Solar forcing of Holocene summer sea-surface temperatures in the northern North Atlantic. Geology 43 (3), 203−206.

Jungclaus, J.H., Lohmann, K., Zanchettin, D., 2014. Enhanced 20th-century heat transfer to the Arctic simulated in the context of climate variations over the last millennium. Clim. Past 10 (6), 2201−2213.

Justwan, A., Koç, N., Jennings, A.E., 2008. Evolution of the Irminger and East Icelandic Current systems through the Holocene, revealed by diatom-based sea surface temperature reconstructions. Quat. Sci. Rev. 27 (15), 1571−1582.

Kennedy, J.J., Rayner, N.A., Atkinson, C.P., Killick, R.E., 2019. An ensemble data set of sea surface temperature change from 1850: the Met Office Hadley Centre HadSST.4.0.0.0 data set. J. Geophys. Res. Atmos. 124 (14), 7719−7763.

Laepple, T., Huybers, P., 2014. Ocean surface temperature variability: large model−data differences at decadal and longer periods. P. Natl. Acad. Sci. 111 (47), 16682−16687.

Lee, T.C.K., Zwiers, F.W., Tsao, M., 2008. Evaluation of proxy-based millennial reconstruction methods. Clim. Dynam. 31 (2), 263−281.

Li, Y., Protopopescu, V.A., Arnold, N., Zhang, X., Gorin, A., 2009. Hybrid parallel tempering and simulated annealing method. Appl. Math. Comput. 212 (1), 216−228.

Lorenz, E.N., 1969. Atmospheric predictability as revealed by naturally occurring analogues. J. Atmos. Sci. 26 (4), 636−646.

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., Zerefos, C., 2016. European summer temperatures since Roman times. Environ. Res. Lett. 11 (2), 024001.

Luterbacher, J., Xoplaki, E., Dietrich, D., Rickli, R., Jacobeit, J., Beck, C., Gyalistras, D., Schmutz, C., Wanner, H., 2002. Reconstruction of sea level pressure fields over

the Eastern North Atlantic and Europe back to 1500. Clim. Dynam. 18 (7), 545—561.

Løvsletten, O., Rypdal, M., 2016. Statistics of regional surface temperatures after 1900: long-range versus short-range dependence and significance of warming trends. J. Clim. 29 (11), 4057—4068.

Malamud, B.D., Turcotte, D.L., 1999. Self-affine time series: measures of weak and strong persistence. J. Stat. Plann. Inference 80 (1), 173—196.

Mann, M.E., Rutherford, S., Wahl, E., Ammann, C., 2005. Testing the fidelity of methods used in proxy-based reconstructions of past climate. J. Clim. 18 (20), 4097—4107.

Mann, M.E., Rutherford, S., Wahl, E., Ammann, C., 2007. Robustness of proxy-based climate field reconstruction methods. J. Geophys. Res. 112 (D12), D12109.

Mann, M.E., Zhang, Z., Hughes, M.K., Bradley, R.S., Miller, S.K., Rutherford, S., Ni, F., 2008. Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. P. Natl. Acad. Sci. 105 (36), 13252—13257.

Mann, M.E., Zhang, Z., Rutherford, S., Bradley, R.S., Hughes, M.K., Shindell, D., Ammann, C., Faluvegi, G., Ni, F., 2009. Global signatures and dynamical origins of the little ice age and medieval climate anomaly. Science 326, 1256—1260.

Marchal, O., Cacho, I., Stocker, T.F., Grimalt, J.O., Calvo, E., Martrat, B., Shackleton, N., Vautravers, M., Cortijo, E., van Kreveld, S., Andersson, C., Koç, N., Chapman, M., Sbaffi, L., Duplessy, J.-C., Sarnthein, M., Turon, J.-L., Duprat, J., Jansen, E., 2002. Apparent long-term cooling of the sea surface in the northeast Atlantic and Mediterranean during the Holocene. Quat. Sci. Rev. 21 (4), 455—483.

Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., González-Rouco, J., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X., Timmermann, A., 2013. Information from paleoclimate archives. In: Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Book Section 5. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 383—464, 978-1-107-66182-0.

McGregor, H.V., Evans, M.N., Goosse, H., Leduc, G., Martrat, B., Addison, J.A., Mortyn, P.G., Oppo, D.W., Seidenkrantz, M.-S., Sicre, M.-A., Phipps, S.J., Selvaraj, K., Thirumalai, K., Filipsson, H.L., Ersek, V., 2015. Robust global ocean cooling trend for the pre-industrial Common Era. Nat. Geosci. 8 (9), 671—677.

Miettinen, A., Divine, D., Koç, N., Godtliebsen, F., Hall, I.R., 2012. Multicentennial variability of the sea surface temperature gradient across the subpolar North Atlantic over the last 2.8 kyr. J. Clim. 25 (12), 4205—4219.

Miettinen, A., Divine, D.V., Husum, K., Koç, N., Jennings, A., 2015. Exceptional ocean surface conditions on the SE Greenland shelf during the Medieval Climate Anomaly. Paleoceanography 30 (12), 1657—1674.

Moffa-Sánchez, P., Born, A., Hall, I.R., Thornalley, D.J.R., Barker, S., 2014. Solar forcing of North Atlantic surface temperature and salinity over the past millennium. Nat. Geosci. 7 (4), 275—278.

Moffa-Sánchez, P., Moreno-Chamarro, E., Reynolds, D.J., Ortega, P., Cunningham, L., Swingedouw, D., Amrhein, D.E., Halfar, J., Jonkers, L., Jungclaus, J.H., Perner, K., Wanamaker, A., Yeager, S., 2019. Variability in the northern North Atlantic and Arctic oceans across the last two millennia: a review. Paleoceanogr. Paleoclim. 34 (8), 1399—1436.

Moossen, H., Bendle, J., Seki, O., Quillmann, U., Kawamura, K., 2015. North Atlantic Holocene climate evolution recorded by high-resolution terrestrial and marine biomarker records. Quat. Sci. Rev. 129, 111—127.

Neukom, R., Steiger, N., Gómez-Navarro, J.J., Wang, J., Werner, J.P., 2019. No evidence for globally coherent warm and cold periods over the preindustrial Common Era. Nature 571 (7766), 550—554.

Nilsen, T., Werner, J.P., Divine, D.V., Rypdal, M., 2018. Assessing the performance of the BARCAST climate field reconstruction technique for a climate with long-range memory. Clim. Past 14 (6), 947—967.

Otto-Bliesner, B.L., Brady, E.C., Fasullo, J., Jahn, A., Landrum, L., Stevenson, S., Rosenbloom, N., Mai, A., Strand, G., 2016. Climate variability and change since 850 CE: an ensemble approach with the Community Earth System Model. Bull. Am. Meteorol. Soc. 97 (5), 735—754.

PAGES 2k-PMIP3 group, 2015. Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium. Clim. Past 11 (12), 1673—1699.

PAGES2k Consortium, 2013. Continental-scale temperature variability during the past two millennia. Nat. Geosci. 6 (5), 339—346.

PAGES2k Consortium, 2017. A global multiproxy database for temperature reconstructions of the Common Era. Sci. Data 4 (1), 170088.

Pongratz, J., Reick, C., Raddatz, T., Claussen, M., 2008. A reconstruction of global agricultural areas and land cover for the last millennium. Global Biogeochem. Cycles 22 (3), GB3018.

Pyrina, M., Wagner, S., Zorita, E., 2017. Pseudo-proxy evaluation of climate field reconstruction methods of North Atlantic climate based on an annually resolved marine proxy network. Clim. Past 13 (10), 1339—1354.

Rahmstorf, S., Box, J.E., Feulner, G., Mann, M.E., Robinson, A., Rutherford, S., Schaffernicht, E.J., 2015. Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation. Nat. Clim. Change 5 (5), 475—480.

Reid, P.C., Edwards, M., Beaugrand, G., Skogen, M., Stevens, D., 2003. Periodic changes in the zooplankton in the North Sea during the twentieth century linked to oceanic inflow. Fish. Oceanogr. 12 (4—5), 260—269.

Reynolds, D.J., Hall, I.R., Slater, S.M., Mette, M.J., Wanamaker, A.D., Scourse, J.D., Garry, F.K., Halloran, P.R., 2018. Isolating and reconstructing key components of

north Atlantic Ocean variability from a sclerochronological spatial network. Paleoceanogr. Paleoclimatol. 33 (10), 1086—1098.

Reynolds, D.J., Hall, I.R., Slater, S.M., Scourse, J.D., Halloran, P.R., Sayer, M.D.J., 2017. Reconstructing past seasonal to multicentennial-scale variability in the NE Atlantic Ocean using the long-lived marine bivalve mollusk Glycymeris glycymeris. Paleoceanography 32 (11), 1153—1173.

Reynolds, D.J., Scourse, J.D., Halloran, P.R., Nederbragt, A.J., Wanamaker, A.D., Butler, P.G., Richardson, C.A., Heinemeier, J., Eiríksson, J., Knudsen, K.L., Hall, I.R., 2016. Annually resolved North Atlantic marine climate over the last millennium. Nat. Commun. 7 (1), 13502.

Richter, T., Peeters, F., van Weering, T., 2009. Late Holocene (0—2.4kaBP) surface water temperature and salinity variability, Feni Drift, NE Atlantic Ocean. Quat. Sci. Rev. 28 (19), 1941—1955.

Schmidt, G.A., Jungclaus, J.H., Ammann, C.M., Bard, E., Braconnot, P., Crowley, T.J., Delaygue, G., Joos, F., Krivova, N.A., Muscheler, R., Otto-Bliesner, B.L., Pongratz, J., Shindell, D.T., Solanki, S.K., Steinhilber, F., Vieira, L.E.A., 2011. Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0). Geosci. Model Dev. 4 (1), 33—45.

Sicre, M.-A., Hall, I.R., Mignot, J., Khodri, M., Ezat, U., Truong, M.-X., Eiríksson, J., Knudsen, K.-L., 2011. Sea surface temperature variability in the subpolar Atlantic over the last two millennia. Paleoceanography 26 (4), PA4218.

Singh, H.K.A., Hakim, G.J., Tardif, R., Emile-Geay, J., Noone, D.C., 2018. Insights into Atlantic multidecadal variability using the Last Millennium Reanalysis framework. Clim. Past 14 (2), 157—174.

Smerdon, J.E., 2012. Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments. Wiley Interdiscip. Rev. Clim. Change 3 (1), 63—77.

Smerdon, J.E., Kaplan, A., Chang, D., Evans, M.N., 2010. A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millennium. J. Clim. 23 (18), 4856—4880.

Smerdon, J.E., Kaplan, A., Zorita, E., González-Rouco, J.F., Evans, M.N., 2011. Spatial performance of four climate field reconstruction methods targeting the Common Era. Geophys. Res. Lett. 38 (11), L11705.

Spielhagen, R.F., Werner, K., Sørensen, S.A., Zamelczyk, K., Kandiano, E., Budeus, G., Husum, K., Marchitto, T.M., Hald, M., 2011. Enhanced modern heat transfer to the arctic by warm atlantic water. Science 331 (6016), 450—453.

Steiger, N.J., Hakim, G.J., Steig, E.J., Battisti, D.S., Roe, G.H., 2014. Assimilation of time-averaged pseudoproxies for climate reconstruction. J. Clim. 27 (1), 426—441.

Stramska, M., Białstrokogrodzka, J., 2015. Spatial and temporal variability of sea surface temperature in the Baltic Sea based on 32-years (1982—2013) of satellite data. Oceanologia 57 (3), 223—235.

Sutton, R.T., Hodson, D.L.R., 2005. Atlantic Ocean forcing of north American and European summer climate. Science 309 (5731), 115—118.

Talento, S., Schneider, L., Werner, J., Luterbacher, J., 2019. Millennium-length precipitation reconstruction over south-eastern Asia: a pseudo-proxy approach. Earth Syst. Dynam. 10 (2), 347—364.

Tardif, R., Hakim, G.J., Perkins, W.A., Horlick, K.A., Erb, M.P., Emile-Geay, J., Anderson, D.M., Steig, E.J., Noone, D., 2019. Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling. Clim. Past 15 (4), 1251—1273.

Thornalley, D.J.R., Elderfield, H., McCave, I.N., 2009. Holocene oscillations in temperature and salinity of the surface subpolar North Atlantic. Nature 457 (7230), 711—714.

Tierney, J.E., Zhu, J., King, J., Malevich, S.B., Hakim, G.J., Poulsen, C.J., 2020. Glacial cooling and climate sensitivity revisited. Nature 584 (7822), 569—573.

Tingley, M.P., Huybers, P., 2010. A Bayesian algorithm for reconstructing climate Anomalies in space and time. Part I: development and applications to paleoclimate reconstruction problems. J. Clim. 23 (10), 2759—2781.

Tingley, M.P., Huybers, P., 2013. Recent temperature extremes at high northern latitudes unprecedented in the past 600 years. Nature 496 (7444), 201—205.

Vieira, L.E.A., Solanki, S.K., Krivova, N.A., Usoskin, I., 2011. Evolution of the solar irradiance during the Holocene. A&A, p. 531. A6.

Wang, J., Emile-Geay, J., Guillot, D., Smerdon, J.E., Rajaratnam, B., 2014. Evaluating climate field reconstruction techniques using improved emulations of real-world conditions. Clim. Past 10 (1), 1—19.

Wang, J., Yang, B., Ljungqvist, F.C., Luterbacher, J., Osborn, T.J., Briffa, K.R., Zorita, E., 2017. Internal and external forcing of multidecadal Atlantic climate variability over the past 1,200 years. Nat. Geosci. 10 (7), 512—517.

Werner, J.P., Divine, D.V., Ljungqvist, F.C., Nilsen, T., Francus, P., 2018. Spatio-temporal variability of Arctic summer temperatures over the past 2 millennia. Clim. Past 14, 527—557.

Werner, J.P., Luterbacher, J., Smerdon, J.E., 2013. A pseudoproxy evaluation of Bayesian hierarchical modeling and canonical correlation analysis for climate field reconstructions over Europe. J. Clim. 26 (3), 851—867.

Werner, J.P., Tingley, M.P., 2015. Technical Note: probabilistically constraining proxy age-depth models within a Bayesian hierarchical reconstruction model. Clim. Past 11 (3), 533—545.

Werner, K., Müller, J., Husum, K., Spielhagen, R.F., Kandiano, E.S., Polyak, L., 2016. Holocene sea subsurface and surface water masses in the fram strait - comparisons of temperature and sea-ice reconstructions. Quat. Sci. Rev. 147, 194—209. Special Issue: PAST Gateways (Palaeo-Arctic Spatial and Temporal Gateways).

Wong, W.H., Liang, F., 1997. Dynamic weighting in Monte Carlo and optimization. P. Natl. Acad. Sci. USA 94 (26), 14220—14224.