JOURNAL OF GEOPHYSICAL RESEARCH, VOL. ???, XXXX, DOI:10.1029/,

Pairwise Scale-Space Comparison of Time Series with Application to Climate Research

F. Godtliebsen

Department of Mathematics and Statistics, University of Tromsø, Tromsø, Norway

L. Holmström

Department of Mathematical Sciences, University of Oulu, Oulu, Finland.

A. Miettinen

Norwegian Polar Institute, Tromsø, Norway

P. Erästö

Department of Mathematical Sciences, University of Oulu, Oulu, Finland.

D. V. Divine

Department of Mathematics and Statistics, University of Tromsø, Tromsø, Norway

N. Koc

Norwegian Polar Institute, Tromsø, Norway

F. Godtliebsen, Department of Mathematics and Statistics, University of Tromsø, Tromsø, Norway. (e-mail: fred.godtliebsen@uit.no)

Abstract. In this paper, we study how sea surface temperature variations in the North Atlantic and the Norwegian Seas are correlated with the climate in the Northern Hemisphere in late Holocene. The analysis is performed by testing statistical hypotheses through novel scale-space methodologies. In late Holocene, the proposed techniques reveal that the climate development in the subpolar North Atlantic has been incoherent with the development in the Norwegian Sea and the Northern Hemisphere. A prominent discrepancy between the three analyzed series is identified for the periods associated with the Medieval Warm Period and the Little Ice Age. A divergence between the oceanic series and the global Northern Hemisphere temperature estimate detected in the 20^{th} century is in line with the inferred imprint of recent climate change which suggests accentuated warming in par-

L. Holmström, Department of Mathematical Sciences, University of Oulu, Oulu, Finland.

A. Miettinen, Norwegian Polar Institute, Polar Environmental Centre, N-9296 Tromsø, Norway.

P. Erästö, Department of Mathematical Sciences, University of Oulu, Oulu, Finland.

D. V. Divine, Department of Mathematics and Statistics, University of Tromsø, Tromsø, Norway.

N. Koc, Norwegian Polar Institute, Polar Environmental Centre, N-9296 Tromsø, Norway.

ticular over continental regions. Overall, the results obtained by scale-space analysis underscores the significance of the northern North Atlantic in shaping the climate globally, mainly through changes in the strength and structure of the Atlantic meridional overturning circulation.

1. Introduction

A best possible understanding of present and past climate is of utmost importance for producing reliable predictions of future climate scenarios. Today we face changes in the climate all over the world and the observed changes at different locations can show large discrepancies. Here we focus on a particular area of interest by investigating how the trends in sea surface temperature (SST) in the North Atlantic and the Nordic Seas are related with the climate development in the Northern Hemisphere during late Holocene.

To gain insight into this question, we utilize the theory [e.g. *Bjerknes*, 1964] that variability of SST in the North Atlantic and the Nordic Seas has a profound effect on climate in the Northern Hemisphere due to heat release to the atmosphere from the North Atlantic Current (NAC). The NAC plays an important role in the Atlantic Meridional Overturning Circulation (AMOC), which is an essential component of the global climate system [*Wellinga and Wood*, 2002], transporting heat northward via the NAC and ventilating the world ocean through the North Atlantic Deep Water (NADW) formation. The AMOC and regional climate are closely linked [e.g. *Latif et al.*, 2004] and known to vary in a broad range of time-scales [e.g. *Thornalley et al.*, 2009]. The short-term variability is primarily driven by the atmosphere [*Marshall et al.*, 2001], whereas at longer time-scales, the role of the ocean becomes more important [e.g. *Bjerknes*, 1964; *Timmermann et al.*, 1998; *Knight et al.*, 2005].

Historical records and proxy climate data from the Northern Hemisphere have provided evidence for the most recent major climate anomalies, such as the warm Medieval Warm Period (MWP) between 800 and 1400 AD [e.g. *Lamb*, 1965; *Bradley et al.*, 2003; *Mann and Jones*, 2003; *Berner et al.*, 2011], and the following colder era, the Little Ice Age (LIA) between 1400 and 1900 AD [*Grove*, 1988; *Bradley and Jones*, 1993; *Moberg et al.*, 2005; *Mann et al.*, 2008]. Several theories have been proposed to explain the possible cause for these anomalies, such as long-term variability in total solar irradiance [*Shindell et al.*, 2001], sulfate aerosols ejected into the atmosphere by volcanism [*Crowley*, 2000], and changes in large-scale ocean circulation [*Broecker*, 2000; *Crowley*, 2000].

The aim of this paper is to analyze the SST variability in the subpolar North Atlantic and the Nordic Seas in late Holocene and to obtain a better understanding of how the variability in these SSTs correlates with the Northern Hemisphere temperatures. We start out by performing a statistical comparison of two 1200-year-long SST proxy records from the Reykjanes Ridge, in the subpolar North Atlantic, and the Vøring Plateau, in the Norwegian Sea. Specifically, we want to test whether there have been different climatological developments at the Reykjanes ridge and the Vøring plateu for the last 1200 years of the Holocene. If such differences are found we would like to give a characterization of when and at what time scales they have occurred. In addition, we would like to test if there are occasions where changes at both locations have been of the same type, but one has changed more rapidly than the other. Moreover, we would like to give a good characterization of how these two SST series relate to the Northern Hemisphere temperature, presented by *Mann et al.* [2008], for the same time period. Such exploratory data analyses can give new insight into the interpretation of the climatological phenomena observed during this period.

New insights into the phenomena underlying these data sets can be obtained using the methods of time series analysis [e.g. *Box and Jenkins*, 1970; *Brockwell and Davis*, 1991 and *Shumway and Stoffer*, 2000]. A detailed description of the data sets analyzed will be given in Section 4 but an important difference between the three time series should be noted already here, namely that the Northern Hemisphere data set is sampled on a regular grid as five year means while the two SST series are unevenly sampled. We note that compared to the extensive literature on the analysis of evenly sampled signals, fewer papers address unevenly sampled series, e.g. *Lomb* [1976] and *Scargle* [1982]. Many of the methods for unevenly sampled data are based on interpolation [e.g. *Quahabi et al.*, 1998; *Dowski*, 1998]. An alternative approach, frequently used for nonstationary signals, is to explicitly or implicitly use sliding windows, such as short-time FFT and time-varying multitaper methods, see e.g. *Bayram and Baraniuk* [2000] and *Thomson* [2000].

Recently, an important focus in time series analysis has been analysis on several time horizons or scales. A pioneering scale-space analysis of densities and regression curves was given by *Chaudhuri and Marron* [1999]. Their work has in recent years been extended to a large number of situations, see e.g. *Godtliebsen et al.* [2002, 2003, 2004], *Park et al.* [2004, 2007], *Erästö and Holmström* [2005, 2007], *Hannig and Lee* [2005], *Hannig and Marron* [2006] and *Olsen et al.* [2008]. For more references on statistical scale-space methods, see *Holmström* [2010a] for a recent review. We view scale-space methods as particularly useful in climatology since the salient features in a time series may depend heavily on the time horizon it is analyzed on. Scale-space methodologies have in recent years become a useful tool also for geologists, glaciologists and oceanographers; see e.g. *Berner et al.* [2008] and *Miettinen et al.* [2011].

A pairwise scale-space comparison of time series was given by *Park and Kang* [2008]. Here, we develop a technique which is similar to their approach but there are two important differences. First, we compare slopes instead of means since that is more natural in climatology where time-series may exhibit non-stationary behavior with persistent changes in the mean value. Second, we describe methods based both on classical statistical and Bayesian ideas whereas *Park and Kang* [2008] give a procedure motivated from a classical point of view only. The motivation for introducing the Bayesian approach is to see whether two different statistical paradigms give essentially the same results for the data sets analyzed. Such an agreement would be reassuring, bolstering the credibility of the results obtained. Another reason for introducing the Bayesian approach is that, in a scale space context, it can more easily handle complexities such as serial correlation in the time series. The classical scale space methodology used in the comparison is still important, not least because of its much lower threshold for new users.

The paper is organized as follows. In Section 2, we describe our statistical model and give a short outline of the scale-space idea. In Section 3, a description of the methodologies developed for pairwise comparison of time series is given. A description of the climatological data and the results obtained are given in Section 4. A discussion is provided in Section 5. An Appendix contains many of the details of the Bayesian approach. Matlab functions used for our analyses can be downloaded from http://www.unc.edu/~marron/marron_software.html and http://mathstat.helsinki.fi/bsizer/.

2. Model, assumptions and scale-space background

Recall that our aim is a comparison of two time series. We assume that time series k, where k is 1 or 2, follows the simple model

$$y_{k,i} = m_k(x_{k,i}) + \sigma_k(x_{k,i})\varepsilon_{k,i}, \quad k = 1, 2; \quad i = 1, ..., n_k, \tag{1}$$

where m_k and σ_k denote the unknown regression function and noise standard deviation function of time series k, respectively. The $x_{k,i}$ denote the possibly unevenly sampled time points where observations $y_{k,i}$ exist. Note that the sampling in the two time series typically is not the same. The $\varepsilon_{k,i}$ denote independently distributed random errors with mean 0 and variance 1. In the Bayesian model the errors are assumed to be Gaussian with possible correlations within each time series. There are n_k observations in time series k. In the data analyses considered in this paper, $m_k(x_{k,i})$ is the true past temperature at time $x_{k,i}$, $y_{k,i}$ is its proxy-based reconstruction, and $\sigma_k(x_{k,i})\varepsilon_{k,i}$ represents the error in the reconstruction.

For convenience of the reader, we next describe briefly the idea in scale-space methodologies. The notion of "scale" in our scale-space analyses always refers "time-scale". However, the methods developed could conceivably be applied also in other situations, such as in analysis of spatial data where features in different spatial scales would be of interest.

To keep things simple, we assume that we have observed just one time series following the model in equation (1). A traditional analysis will search for the underlying true m through a "smooth" estimate $\hat{m}_h$ where the parameter h controls the degree of smoothness. See e.g. *Fan and Gijbels* [1996] for more details. Then, inference about m is based on $\hat{m}_h$. A major disadvantage with this approach is that $\hat{m}_h$ is a biased estimator of m. The novel idea in a scale-space analysis is that we do not focus on the search for the underlying true m. Instead, we study scale-space versions or smooths of m, denoted by m_h. By this procedure, the estimators $\hat{m}_h$ are unbiased estimators of m_h. Hence, we avoid the bias problems that traditional smoothing methods suffer from. Moreover, we avoid the search for an optimal smoothing level since in a scale-space analysis "all" scales are considered important. Similarly, in a Bayesian scale-space approach, inference is based on the distribution of the smooth m_h, given the observations y_i.

3. Pairwise Scale-Space Comparison

In this section we describe two differently motivated scale-space methodologies for comparing two time series following the model given in Section 2.

3.1. A Classical Approach

Our approach here is a direct application of the original SiZer methodology developed by *Chaudhuri and Marron* [1999]. For time series k, at a particular point x_0 and a given scale h, $\hat{m}_{k,h}(x_0)$ is obtained by fitting the line

$$l(x) = \beta_{k,0} + \beta_{k,1}(x_0 - x)$$

locally to the data $(x_{k,i}, y_{k,i})$. In fact, $\hat{m}_{k,h}(x_0) = \hat{\beta}_{k,0}$, where $(\hat{\beta}_{k,0}, \hat{\beta}_{k,1})$ minimizes

$$\sum_{i=1}^{n_k}[y_{k,i} - (\beta_{k,0} + \beta_{k,1}(x_0 - x_{k,i}))]^2 K_h(x_0 - x_{k,i}),$$

$$K_h(\cdot) = \frac{1}{h}K(\frac{\cdot}{h}),$$

and K is a kernel function, typically a symmetric probability density function. Here, we use a Gaussian kernel. The hypothesis we would like to test, for a given scale h, at the point x_0, is

$$\mathrm{H}_0 : \beta_{1,1}(x_0) = \beta_{2,1}(x_0) \qquad \text{against} \qquad \mathrm{H}_1 : \beta_{1,1}(x_0) \neq \beta_{2,1}(x_0).$$

We do this by rejecting H_0 if

$$\frac{|\hat{\beta}_{1,1}(x_0) - \hat{\beta}_{2,1}(x_0)|}{\widehat{\mathrm{SD}}(\hat{\beta}_{1,1}(x_0) - \hat{\beta}_{2,1}(x_0))} > q \tag{2}$$

where we use the plausible assumption

$$\mathrm{Var}(\hat{\beta}_{1,1}(x_0) - \hat{\beta}_{2,1}(x_0)) = \mathrm{Var}(\hat{\beta}_{1,1}(x_0)) + \mathrm{Var}(\hat{\beta}_{2,1}(x_0))$$

to estimate the denominator in equation (2), and q is a suitable quantile. The value of q is decided in the same way as in *Chaudhuri and Marron* [1999] with

$$\mathrm{ESS}_h(x_0) = \min\{\mathrm{ESS}_{1,h}(x_0), \mathrm{ESS}_{2,h}(x_0)\}$$

where $\mathrm{ESS}_{k,h}(x_0)$ denotes the effective sample size of time series k for scale h and location x_0. The motivation for this approach is that it will be a conservative choice in the sense that we will have more confidence in the features found by our methodology.

In SiZer analyses the results of inferences are visualized with so-called family plots and significance or feature maps, examples of which are shown in Figure 1. In a significance map, a pixel (x, s) corresponding to time x and scale $s = \log_{10}(h)$ is colored blue or red depending on whether the slope of the smooth of the true underlying temperature curve is significantly positive or negative, respectively. Purple indicates non-significance and pixels are colored gray if the data are too sparse to make any conclusions. The SiZer maps for pairwise comparisons (middle panels of Figs. 2 - 4) are interpreted analogously except now inferences are made on the slope of a difference between two time series. The level of significance in all SiZer analyses is 0.05.

3.2. A Bayesian Approach

The Bayesian approach is based on the BSiZer methodology described in *Erästö and Holmström* [2005, 2007] and *Holmström* [2010b]. Denote by $\boldsymbol{y}_k = [y_{k,1}, \ldots, y_{k,n_k}]^T$ the observed time series k, $k = 1, 2$. Let $x_1 < x_2 < \cdots < x_n$ be a grid of time points where one wants to analyze the difference between the slopes of the smooths $m_{1,h}$ and $m_{2,h}$ of the underlying unobserved curves. Let $\boldsymbol{m}'_{k,h} = [m'_{k,h}(x_1), \ldots, m'_{k,h}(x_n)]^T$ be the vector of slopes of $m_{k,h}$ computed on this grid. Our Bayesian scale space analysis uses the posterior distribution of $\boldsymbol{m}'_{1,h} - \boldsymbol{m}'_{2,h}$ given the data $\boldsymbol{y}_1$, $\boldsymbol{y}_2$ to make inferences about the credible features in the difference between the slopes of $m_{1,h}$ and $m_{2,h}$, for a range of time scales h. As in the classical SiZer approach, we also make an independence assumption about the two time series which allows us to obtain a sample from this posterior by sampling separately from the posterior distributions of $\boldsymbol{m}'_{1,h}$ and $\boldsymbol{m}'_{2,h}$ and then simply subtracting the samples. The full details of the Bayesian method are given in the Appendix.

An analog of a SiZer significance map can be obtained by choosing a credibility level $0 < \alpha < 0.5$ and coloring a map pixel (x_j, s) corresponding to time x_j and scale $s = \log_{10}(h)$ blue or red according to whether

$$\mathrm{P}\left\{\boldsymbol{m}'_{1,h}(x_j) - \boldsymbol{m}'_{2,h}(x_j) > 0 \mid \boldsymbol{y}_1, \boldsymbol{y}_2\right\} \geq 1 - \alpha$$

or

$$\mathrm{P}\left\{\boldsymbol{m}'_{1,h}(x_j) - \boldsymbol{m}'_{2,h}(x_j) < 0 \mid \boldsymbol{y}_1, \boldsymbol{y}_2\right\} \geq 1 - \alpha,$$

and purple otherwise, where the probabilities are computed from the generated sample of slope differences. However, instead of using such pointwise inference, the maps are in fact drawn based on the joint posterior probabilities over the grid points x_j's, where a method based on highest pointwise probabilities was used (cf. *Erästö and Holmström* [2005]). We have chosen $\alpha = 0.05$ in all analyses.

Note that we use the same symbol h for the scale space smoothing parameter both in the classical and the Bayesian methods although its technical role in the two approaches is quite different. In the classical SiZer h is the standard deviation (or width in the time domain) of the Gaussian kernel used whereas in the Bayesian BSiZer it controls the roughness penalty in spline smoothing (see the Appendix). Although a spline smoother can be interpreted as an approximate kernel smoother, the relevant smoothing scale ranges of the two methods have very different magnitudes.

4. Results

4.1. Data sets

The two SST series used in this study are diatom based August SST reconstructions with an uneven sampling resolution of $2 - 10$ years from marine sediment cores Rapid 21-COM (hereafter Rapid) from the eastern flank of the Reykjanes Ridge, subpolar North Atlantic [*Miettinen et al.*, 2011; *Miettinen et al.*, 2012], and CR 948/2011 (hereafter CR) from the

Vøring Plateau, the Norwegian Sea [*Andersen et al.*, 2004; *Berner et al.*, 2011]. These two SST series were selected, because a) they represent the highest-resolution SST reconstructions from the northern North Atlantic for the last 1200 years, and b) they are located in critical areas in relation to the NAC, which has an essential role on the North Atlantic climate, i.e., core Rapid 21-COM is influenced by the western branch of the NAC in the south of Iceland, and core CR948/2011 by the eastern branch of the NAC in the Norwegian Sea. The SST reconstructions are based on marine planktonic diatoms and transfer functions. Marine diatoms have proven to be good indicators of surface water conditions in the region [e.g. *Koc-Karpuz and Schrader*, 1990; *Andersen et al.*, 2004; *Berner et al.*, 2008]. A training data set consisting of 139 surface samples with 52 diatom species and modern August SSTs from the Nordic Seas and the North Atlantic [*Andersen et al.*, 2004] was utilized to convert downcore diatom counts to quantitative SST using the weighted averaging partial least squares (WA-PLS) transfer function method [*ter-Braak and Juggins*, 1993]. The WA-PLS diatom transfer function has a RMSE of 0.75 °C, a maximum bias of 0.44 °C and R^2 of 0.96. More details can be found in [*Miettinen et al.*, 2011; *Miettinen et al.*, 2012; *Berner et al.*, 2011].

The Northern Hemisphere surface temperature (hereafter NHem) reconstruction originally named as NH EIV Land+Ocean [*Mann et al.*, 2008]. It is based on a multiple proxy database consisting of tree-rings, marine sediments, speleothems, lacustrine sediments, ice cores, corals, and historical documentary series [*Mann et al.*, 2008]. This proxy database represents a significant extension of the database used in related earlier studies [*Mann et al.*, 1998, 1999; *Juckes et al.*, 2007]. See *Mann et al.* [2008] for further details about this data set.

4.2. Rapid

To get an idea about what significant features can be found in the Rapid time series on different time scales, a SiZer analysis was performed and the result is shown on the first row of Fig. 1. The immediate and overall feature found is an increase in summer SST over the data set, manifested by the color blue for all locations on scales covering the whole period. A closer look at the feature map reveals some features at a centennial time scale. At year 1000 AD, there is a significant increase in the SST reconstruction while there is an abrupt decrease in the SST just after year 1700 AD. Furthermore, there is evidence in the data of a peak in the SST around 1870 AD. This feature seems to be very clear and manifested on scales ranging from 10 to 100 years.

4.3. CR

From the SiZer analysis of the CR time series from the Vøring Plateau displayed in the middle row of Fig. 1, it is clear that, in contrast to the findings for the Reykjanes Ridge, there has been a decrease in temperature on time scales covering the whole period. On scales of length around 100 years, the temperature has decreased more abruptly around years 900, 1100, and 1400 AD. At scales of length 500 years, there is an increasing trend from around year 1500 AD to the present.

4.4. Nhem

In the last row of Fig. 1, the SiZer analysis reveals a millennial-scale decrease in the surface temperature of the Northern Hemisphere. At scales of 10 to 100 years, several features typically associated with major climate transitions of the last Millennium are flagged as significant. In particular, these are the peaks around years 850, 1050, and 1400 AD. A pronounced temperature maximum centered at approximately 1050 AD corresponding to the Medieval Warm Period (MWP, *Lamb* [1965]; *Bradley et al.* [2003]) is detected as significant on scales up to 500 years which is a reflection of a lasting positive surface temperature anomaly from around year 950 to year 1100 AD . Finally, the SiZer map indicates that there is an abrupt decrease in the temperature around year 1420 signifying the onset of the Little Ice Age (LIA, *Moberg et al.* [2005]; *Mann et al.* [2008]). This is a strong feature, visible at scales ranging from 10 to 200 years.

4.5. Rapid vs. CR

By comparing the regional Rapid and CR summer SST reconstructions, we see that the two different methodologies yield very similar results (Fig. 2). Both approaches reveal that the record for Rapid has a significantly larger slope (blue color in Fig. 2) than the record for CR, i.e., in a long term perspective for the last 1200 years, the SST record for Rapid shows a clear warming trend compared with CR, which demonstrates a less pronounced cooling tendency. On a shorter time scale, a blue area over a broad range of scales can be seen around 1400 and 1800 AD showing the most distinct periods when Rapid is increasing (warming) and CR decreasing (cooling).

Analysis of the maps displayed in Figs. 1 and 2 proves that the pairwise scale-space comparison adds important knowledge about the characteristics of the two time series. Note that around year 1800 AD the maps obtained by separate analyses of Rapid and CR (Fig.1) are highlighted blue at a broad range of time scales signifying the period of statistically significant SST increase detected in both series. The pairwise comparison displayed in Fig. 2 also show blue for some scales at this location. The implication of this is that the temperature in Rapid is increasing significantly faster than in CR, a fact that is not clear from the maps of the SiZer analyses of Fig. 1 alone.

Finally, we note that the Bayesian approach reveals some features that are not captured by the classical SiZer approach. The feature flagged around 1200 AD is present only on very small scales. This potential event occurs in the gray area of the classical approach, indicating that inference cannot be performed with this methodology. The same can be stated about the feature around 850 AD. For scales of length around 300 years, the Bayesian approach flags a red area just before 1700 AD. By looking at the observed data in the top panel of Figure 2, there is a vague indication of increase in CR while Rapid is neither increasing nor decreasing. The feature is, however, so vague that it is debatable whether it is actually there or not. It is therefore clear that the agreement between the two approaches is very good.

4.6. Rapid vs. NHem

The comparison of a regional-scale Rapid and global NHem (Fig. 3) series yields results qualitatively similar to the previous analysis between Rapid and CR, indicating an increasing trend for Rapid and a decreasing trend for NHem. Also, the most distinct periods of significantly different temporal evolution of surface temperatures are evident at around 1400 and 1800 AD. However, the trends are reversed for the last century. Whereas Rapid series shows slow cooling, NHem demonstrates a rapid warming trend associated with anthropogenic forcing as indicated by a red area over a broad range of scales. Moreover, a short period of cooling Rapid but warming NHem can be seen around 1750 AD, which is not clear enough to be flagged as significant by classical SiZer. It, however, appears as credible in the Bayesian analysis, as indicated by the red area in the credibility map.

From Figs. 1 and 3 it can be seen that around 1800 AD a similar phenomenon, as described for the comparison of Rapid and CR, is present. This means that also in the comparison of Rapid and NHem, it is clear that the pairwise scale-space comparison complements the information obtained by the two single time series scale-space analyses.

In addition to the discrepancy observed around 1750, there are some other small differences between the classical and the Bayesian analyses but again, many of them occur at scales that are gray regions in the classical approach. Thus, overall, the agreement between the two approaches is remarkably good also in this case.

4.7. CR vs. NHem

The results obtained by comparing CR and NHem (Fig. 4) show distinct differences compared with earlier combinations. First, the regional temperature anomalies are more or less congruent with the global climate development, e.g. the first part of the record until ca. 1400 AD is characterized by the highest SSTs in CR, as well as higher than average surface temperatures in NHem. Secondly, in both reconstructions, the color purple over a broad range of time scales indicates that the derivative is not found to be significantly different from zero. This indicates that the slopes of CR and NHem series in the considered temporal resolution are in phase for most of the investigated period, i.e., they are characterized by a decreasing (cooling) long term trend for the last 1200 years. However, significant differences can be seen in shorter time scales. Red color in a broad range of scales from 800 to 1100 AD indicates a clear cooling trend for CR but a lagged warming trend for NHem suggesting the northern North Atlantic origin of the MWP . Similar periods of the regional surface temperature evolution significantly different from the global climate development can be seen around 1400 AD and in the last century. The opposite situation, namely a stronger warming trend for CR can be seen from around 1500 to 1750 AD.

By comparing Figs. 1 and 4, we infer again that the pairwise comparison contributes additional information around years 900, 1400, and 1900 AD. At approximately 900 and 1400 AD the single time series analyses show a significant decrease of the surface temperature. Note, however, that the pairwise comparison also flags red at this (these) position(s) suggesting that the decrease in the regional CR surface temperature series is significantly steeper than in NHem. After 1900, the separate analyses of CR and NHem flag blue indicating significantly increasing temperatures. But, the increase in CR series appears to be significantly slower than in NHem and the result in the pairwise comparison map is therefore an area flagged as red. From Figure 4 we can see again that for the comparison of these two data sets, the two different statistical approaches show essentially the same feature maps.

5. Discussion and conclusions

We have analyzed the pairwise differences in climate proxy time series using two statistical scale-space paradigms. The original SiZer technique uses classical, "frequentist" statistical reasoning based hypothesis testing while the BSiZer method is based on Bayesian inference that uses posterior probabilities. The regression models employed by the two approaches were also slightly different with SiZer assuming independent errors while BSiZer assumes Gaussian errors with possible temporal correlations. Further, SiZer estimates errors from smoothing residuals while in the Bayesian setting one is able to use prior knowledge e.g. in the form of estimated errors for the reconstructions (cf. the Appendix). The strategies for simultaneous inference or multiple hypothesis testing of features for sets of time points are also different. Despite these contrasts, the two methods produce remarkably similar feature analyses of pairwise differences in the reconstructed temperature time series considered, a reassuring fact that increases our confidence in the robustness of the results. We noted that many of the differences in the feature maps actually occur at least partly in the gray areas of the SiZer maps where this method is unable to produce results due to lack of sufficient data. Here the combination of data and prior information helps the Bayesian method and explains the difference in the results. Posterior analysis of the error covariance structure also suggests that the simpler independent error model of SiZer is probably sufficient here, as the posterior distributions of the off-diagonal elements of the error covariance matrices were highly concentrated near zero.

The results from both statistical methods show statistically significant features from millennial to centennial time scales. The three analyzed series display regional-scale contrasts in climate development in the northern North Atlantic (CR SST vs. Rapid SST) as well as pronounced discrepancies between the regional and global-scale climate variations (North Atlantic records vs. NHem). We note that the difference in seasonal representation between the reconstructions can to some extent bias the inference that follows from our analysis. One can expect, however, that due to the longer time-scales mainly considered here, and because of the negligible relative changes in seasonal orbital forcing over the analyzed 1200-year long period when compared with the entire Holocene [e.g. *Wanner et al.*, 2008], summer and annual mean temperature anomalies are in fact coherent. Besides, the estimate of the Northern Hemisphere surface temperature is largely based on tree ring and latewood density data [e.g. *Mann et al.*, 2008] which are reflective of summer conditions. This suggests that, just like the SST reconstructions from the northern North Atlantic, the NHem series may itself be biased towards the summer season.

A preliminary analysis of the results obtained underscores the significance of the northern North Atlantic in shaping the climate globally, mainly through the changes in the strength and structure of the Atlantic meridional overturning circulation (MOC) [e.g. *Latif et al.*, 2004; *Manabe and Stouffer*, 1998]. A millennial scale progressive synchronous cooling demonstrated by the CR and NHem series until the end of the 1800s signifies a lasting weakening of the eastern branch of the MOC associated with a decreased influx of warm Atlantic waters northward to the Arctic via the North Atlantic Current [*Thornalley et al.*, 2009]. Although the relative roles of various causal factors, both external and internal, behind these changes are still controversial, it had to involve major reorganization in oceanic and atmospheric circulation [e.g. *Trouet et al.*, 2009; *Mann et al.*, 2009].

In shorter, centennial to multicentennial time scales, CR SST series from the Norwegian Sea tends to lead NHem temperatures as can be inferred from the earlier termination of the MWP (flagged red between ca 900-1100 AD in Fig. 4). A delayed response of ca. 50 years to decreasing SST registered in CR in the Norwegian Sea also characterizes the onset of the LIA in the NHem series (Fig. 4) at around 1450 AD. We note that the origin of this lag could be related to a delayed shift in the North Atlantic Oscillation (NAO) phase in response to persistent anomalies in regional sea surface temperatures [e.g. *Trouet et al.*, 2009; *Swingedouw et al.*, 2010; *Miettinen et al.*, 2012]. It is notable that during the LIA, CR series shows generally negative SST anomalies superimposed on a positive trend which is steeper than the one observed in the NHem series (flagged blue during 1500-1800 AD in Fig. 4). This (colder, but warming SST) could suggest that NHem temperatures respond to rising SST only after passing a threshold in the ocean-atmosphere system.

Rapid summer SST series displaying a persistent positive trend throughout the considered time interval seems to stand apart from the variability recorded in CR and NHem records. *Miettinen et al.* [2012] however suggested that the observed statistically significant opposite climate tendencies between the sites in the subpolar North Atlantic and the Norwegian Sea is a surface expression of the lasting changes in the relative strength of the eastern and western branches of the MOC, with a possible amplification through an atmospheric feedback. This apparent SST seesaw in the northern North Atlantic might have an effect on two major anomalies of the European climate of the past Millennium: MWP and LIA. During the MWP, warming of the sea surface in the Norwegian Sea occurred in parallel with cooling in the northern subpolar North Atlantic, whereas the opposite pattern emerged during the LIA.

A divergence between the series detected in the 20^{th} century is in line with the inferred imprint of the recent warming which is generally associated with anthropogenic forcing. Both instrumental data and model based studies agree on accentuated warming in particular over continental regions [e.g. *Karoly and Wu*, 2005; *Knutson et al.*, 2006; *Trenberth et al.*, 2007]. A less pronounced oceanic SST increase is likely to be related to greater evaporation and its heat storage. The recent atmospheric circulation changes, in particular a more positive NAO phase, may also contribute to a moderation of warming trends in subpolar North Atlantic, specifically in the Rapid core region, in the winter half-year. One should also note a distinctive seasonality of the warming pattern with maximum warming in winter and spring [*Knutson et al.*, 2006] which is most likely another forcing factor for a much steeper slope revealed in the NHem record in the twentieth century.

Appendix A: Details of the Bayesian Approach

A1. The model

Write (1) in the form $y_{k,i} = m_k(x_{k,i}) + \varepsilon_{k,i}$, hence absorbing the variances in the variables $\varepsilon_{k,i}$. Denote $\boldsymbol{\varepsilon}_k = [\varepsilon_{k,1}, \ldots, \varepsilon_{k,n_k}]^T$ and, as a slight extension of the model (1), assume that $\boldsymbol{\varepsilon}_k \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_k)$, where $\boldsymbol{\Sigma}_k$ is a general covariance matrix that allows the errors to be correlated. The likelihood of $\boldsymbol{y}_k$ is then the Gaussian

$$p(\boldsymbol{y}_k|\boldsymbol{m}_k, \boldsymbol{\Sigma}_k) \propto |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{y}_k - \boldsymbol{m}_k)^T \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{y}_k - \boldsymbol{m}_k)\right),$$

where $\boldsymbol{m}_k = [m_k(x_{k,1}), \ldots, m_k(x_{k,n_k})]^T$. We assume an inverse Wishart prior for $\boldsymbol{\Sigma}_k$,

$$p(\boldsymbol{\Sigma}_k) \propto |\boldsymbol{\Sigma}_k|^{-(\frac{\nu_k+n+1}{2})} \exp\left(-\mathrm{tr}(\boldsymbol{W}_k \boldsymbol{\Sigma}_k^{-1})\right), \qquad \text{(A1)}$$

where the scale matrix $\boldsymbol{W}_k$ is of the homoscedastic form $\sigma_k^2 \boldsymbol{I}$ and the degrees of freedom ν_k is selected so that the prior is rather uninformative.

For $\boldsymbol{m}_k$ we use a prior that penalizes for roughness in the second derivative of the smooth underlying curve m_k. This idea can be conveniently implemented by assuming that m_k is a natural cubic spline, i.e., a twice continuously differentiable curve that consists of cubic polynomial pieces [e.g. *Green and Silverman*, 1994]. Thus, let the interval $[a, b]$ contain the points $x_{k,i}$, $i = 1, \ldots, n_k$. The spline m_k is then uniquely determined by its values $\boldsymbol{m}_k$ at the knot sequence $x_{k,1}, \ldots, x_{k,n_k}$ because these values determine the interpolating spline uniquely. The prior we use for $\boldsymbol{m}_k$ is the improper Gaussian density

$$p(\boldsymbol{m}_k|\kappa_k) \propto \kappa_k^{\frac{n_k-2}{2}} \exp\left(-\frac{\kappa_k}{2}\boldsymbol{m}_k^T \boldsymbol{L}_k \boldsymbol{m}_k\right), \qquad \text{(A2)}$$

where $\boldsymbol{L}_k$ is a matrix such that

$$\boldsymbol{m}_k^T \boldsymbol{L}_k \boldsymbol{m}_k = \int_a^b [m_k''(x)]^2 \mathrm{d}x$$

and $\kappa_k > 0$ controls the level of roughness penalty. The parameter $\kappa_k > 0$ can be fixed or one can consider it unknown and in that case we specify a Gamma prior for it.

The joint posterior $p(\boldsymbol{m}_k, \boldsymbol{\Sigma}_k, \kappa_k | \boldsymbol{y}_k)$ given the data $\boldsymbol{y}_k$ is now obtained from Bayes' theorem,

$$p(\boldsymbol{m}_k, \boldsymbol{\Sigma}_k, \kappa_k | \boldsymbol{y}_k) \propto p(\boldsymbol{\Sigma}_k) p(\boldsymbol{m}_k | \kappa_k) p(\kappa_k) p(\boldsymbol{y}_k | \boldsymbol{m}_k, \boldsymbol{\Sigma}_k). \tag{A3}$$

We assume that the observations $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are conditionally independent given the underlying curves $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, and other model parameters and that, for the two time series, these parameters also are independent *a priori*. Then the triples $(\boldsymbol{m}_1, \boldsymbol{\Sigma}_1, \kappa_1)$ and $(\boldsymbol{m}_2, \boldsymbol{\Sigma}_2, \kappa_2)$ are independent given the data $\boldsymbol{y}_1$, $\boldsymbol{y}_2$,

$$p(\boldsymbol{m}_1, \boldsymbol{\Sigma}_1, \kappa_1, \boldsymbol{m}_2, \boldsymbol{\Sigma}_2, \kappa_2 | \boldsymbol{y}_1, \boldsymbol{y}_2) = p(\boldsymbol{m}_1, \boldsymbol{\Sigma}_1, \kappa_1 | \boldsymbol{y}_1) p(\boldsymbol{m}_2, \boldsymbol{\Sigma}_2, \kappa_2 | \boldsymbol{y}_2).$$

We can therefore obtain a sample from the posterior $p(\boldsymbol{m}_1, \boldsymbol{m}_2 | \boldsymbol{y}_1, \boldsymbol{y}_2)$ by using Gibbs samplers to generate samples separately from $p(\boldsymbol{m}_1, \boldsymbol{\Sigma}_1, \kappa_1 | \boldsymbol{y}_1)$ and $p(\boldsymbol{m}_2, \boldsymbol{\Sigma}_2, \kappa_2 | \boldsymbol{y}_2)$ and keeping only the parts that correspond to $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$. To get a sample of the slope vectors $\boldsymbol{m}'_{k,h} = [m'_{k,h}(x_1), \ldots, m'_{k,h}(x_n)]^T$ of the smooth $m_{k,h}$ of the curve m_k one first smooths the sample of the $\boldsymbol{m}_k$'s by multiplying the sample vectors by the matrix $(\boldsymbol{I} + h\boldsymbol{L}_k)^{-1}$, effectively a discrete spline smoother. This produces a sample of smooths $\boldsymbol{m}_{k,h} = [m_{k,h}(x_{k,1}), \ldots, m_{k,h}(x_{k,n_k})]^T$ and a second multiplication by an appropriate matrix then produces a sample of the slope vectors $\boldsymbol{m}'_{k,h}$ (cf. *Green and Silverman* [1994]). Finally, a sample from the posterior distribution of the slope difference $\boldsymbol{m}'_{1,h} - \boldsymbol{m}'_{2,h}$ is obtained by forming pairwise differences between samples of $\boldsymbol{m}'_{1,h}$ and $\boldsymbol{m}'_{2,h}$.

A2. Selection of priors

The classical SiZer estimates the errors in (1) from residuals of the smoothed time series. In the Bayesian setting one tries to utilize any prior knowledge one might have about the magnitude of the errors.

The prior distribution (A1) of $\boldsymbol{\Sigma}_k$ has the mean

$$\mathrm{E}(\boldsymbol{\Sigma}_k) = (\nu_k - n_k - 1)^{-1}\boldsymbol{W}_k,$$

where, as noted above, ν_k is the parameter (degrees of freedom) that defines the tightness (informativeness) of the prior and n_k is the length of the time series $\boldsymbol{y}_k$. For the prior parameter $\boldsymbol{W}_k$ we used a diagonal scale matrix $\boldsymbol{W}_k = w_k\boldsymbol{I}_{n_k}$ such that $\mathrm{E}(\boldsymbol{\Sigma}_k) = \sigma_k^2\boldsymbol{I}_{n_k}$, where σ_k is a fixed value. For the time series Rapid and CR described in Sections 4.2 and 4.3 we used the value $\sigma_k = 0.75$, an estimated root mean square error of prediction (RMSEP) reported in *Miettinen at al.*, [2012]. For the time series NHem described Section 4.4 we took $\sigma_k = 0.15$, a value estimated from the error bars in Figure S5a of the Supplement to *Mann et al.* [2008]. Since now $\sigma_k^2\boldsymbol{I}_{n_k} = (\nu_k - n_k - 1)^{-1}w_k\boldsymbol{I}_{n_k}$, we have $w_k/\sigma_k^2 = \nu_k - n_k - 1$. We took $w_k = 5$ and $w_k = 0.5$ for the first two and the third time series, respectively, which corresponds to degrees of freedom ν_k of 149, 219, and 264 for the three time series. With these choices the prior 95% highest density intervals for the diagonal elements of $\boldsymbol{\Sigma}_k$ were approximately [0.45, 1.15], [0.5, 1.15] and [0.11, 0.19] for the three time series and therefore wide enough to allow also the data to have an effect on the posterior errors. It turned out that, with very vague priors, the posterior errors of the first two time series were smaller than the prior values which suggests that the values $\sigma_k = 0.75$ used probably is a bit too large for these temperature reconstructions and that the credibility analysis of their features therefore tends to be somewhat conservative. In contrast, the posterior errors of the third reconstruction were very similar to their prior values.

We used the Gamma(1,1) prior for the parameter κ_k in (A2) in order not to smooth out the finest details in the reconstructions. However, after testing several different priors for κ_k we concluded that both the marginal posterior distribution of κ_k and the credibility maps produced were quite insensitive to a any particular reasonable prior choice.

References

Andersen, C., N. Koc, A. Jennings, and J. T. Andrews (2004), Nonuniform response to the major surface currents in the Nordic Seas to insolation forcing: Implications for the Holocene climate variability, *Paleoceanography*, *19*, PA2003, doi:10.1029/2002PA000873.

Bayram, M., and R.G. Baraniuk (2000), Multiple window time-varying spectrum estimation, *Nonlinear and Nonstationary signal processing*, 292-316, Cambridge University press.

Berner, K. S., N. Koc, D. Divine, F. Godtliebsen, and M. Moros (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*, PA2210, doi:10.1029/2006PA001339.

Berner, K., N. Koc, F. Godtliebsen, and D. V. Divine (2011), Holocene climate variability of the Norwegian Atlantic Current during high and low solar insolation forcing. *Paleoceanography*, *26*, PA2220, doi10.1029/2010PA002002.

Bjerknes, J. (1964), Atlantic air-sea interaction, *Advances in Geophysics*, *10*, 1-82.

Box, G. E. P., and G. M. Jenkins (1970), *Time Series Analysis: Forecasting and Control*, San-Francisco, Holden-Day.

Bradley, R. S., and P. D. Jones (1993), Little Ice Age summer temperature variations: their nature and relevance to recent global warming trends, The Holocene, 3, 367-376.

Bradley, R. S., M. K. Hughes, and H. F. Diaz (2003), Climate in Medieval Time, *Science*, *302*, 404-405.

Broecker, W. S. (2000), Was a change in thermohaline circulation responsible for the Little Ice Age? *Proceedings of the National Academy of Sciences USA*, *97*, 13391342.

Brockwell, P. J., and R. A. Davis (1991), *Time Series: Theory and Methods*, Springer-Verlag, New York.

Chaudhuri, P., and J. S. Marron (1999), SiZer for exploration of structures in curves, *Journal of the American Statistical Association*, *94*, 807-823.

Crowley, T. J. (2000), Causes of Climate Change Over the Past 1000 Years, *Science*, *289*, 270-277.

Dowski, E.R., Whitmore, C.A., and S.K. Avery (1988), Estimation of randomly sampled sinusoids in additive noise, *IEEE Transactions on Acoustics, Speech, and Signal Processing 36*, 1906-1908.

Erästö, P., and L. Holmström (2005), Bayesian multiscale smoothing for making inferences about features in scatter plots, *Journal of Computational and Graphical Statistics*, *14*, 569-589.

Erästö, P., and L. Holmström (2007), Bayesian analysis of features in a scatter plot with dependent observations and errors in predictors, *Journal of Statistical Computation and Simulation*, *77*, 421-431.

Fan, J., and I. Gijbels (1996), *Local polynomical modelling and its applications*, Chapman and Hall, London.

Godtliebsen, F., J. S. Marron, and P. Chaudhuri (2002), Significance in scale-space for bivariate density estimation, *Journal of Computational and Graphical Statistics*, *10*, 1-21.

Godtliebsen, F., J. S. Marron, and P. Chaudhuri (2004), Significance in scale-space, *Image and Vision Computing*, *22*, 1093-1104.

Godtliebsen, F., L. R. Olsen, and J.-G. Winther (2003), Recent developments in time series analysis: Examples of use in climate research, *Geophysical Research Letters*, *30*, 1654-1657.

Green, P. J., and B. W. Silverman (1994), *Nonparametric Regression and Generalized Linear Models. A roughness penalty approach*, Chapman & Hall, London.

Grove, J. M. (1988), The Little Ice Age, *Methuen*, 520 pp.

Hannig, J., and T. Lee (2005), Robust SiZer for exploration of regression structures and outlier detection. *Journal of Computational and Graphical Statistics*, *15*, 101-117.

Hannig, J., and J. S. Marron (2006), Advanced distribution theory for SiZer, *Journal of the American Statistical Association*, *101*, 484-499.

Hocke, K., and N. Kämpfer (2008), Gap filling and noise reduction of unevenly sampled data by means of the Lomb-Scargle periodogram, *Atmospheric Chemistry and Physics Discussions*, *8*, 4603-4623.

Holmström, L. (2010a), Scale space methods, *Wiley Interdisciplinary Reviews: Computational Statistics*, Vol. 2, 2, 150-159.

Holmström, L. (2010b), BSiZer. *Wiley Interdisciplinary Reviews: Computational Statistics*, Vol.2, 5, 526-534.

Juckes, M. N., Allen,M.R., Briffa, K.R., Esper, J., Hegerl, G.C., Moberg, A., Osborn, T.J., and S. L. Weber (2007), Millennial temperature reconstruction intercomparison and evaluation, *Climate of the Past*, *3*, 591-609.

Karoly, D. J., and Q. Wu (2005), Detection of Regional Surface Temperature Trends, *Journal of Climate*, *18*, 4337-4343,doi:10.1175/JCLI3565.1.

Knigth, J. R., R. J. Allan, C. K. Folland, M. Vellinga, and M. E. Mann (2005), A signature of persistent natural thermohaline circulation cycles in observed climate, *Geophysical Research Letters*, *32*, L20808, doi:10.1029/2005GL024233.

Knutson, T. R., T. L. Delworth, K. W. Dixon, I. M. Held, J. Lu, V. Ramaswamy, M. D. Schwarzkopf, G. Stenchikov, and R. J. Stouffer (2006), Assessment of Twentieth-Century Regional Surface Temperature Trends Using the GFDL CM2 Coupled Models, *Journal of Climate*, *19*, 1624-1651, doi:10.1175/JCLI3709.1

Koc-Karpuz, N., and H. Schrader (1990), Surface sediment diatom distribution and Holocene paleo-temperature variations in the Greenland, Iceland and Norwegian Seas through the last 14 ka based on diatoms, *Paleoceanography*, *5*, 557-580.

Lamb, H. H. (1965), The Early Medieval Warm Epoch and its Sequel, *Palaeogeography, Palaeoclimatology, Palaeoecology*, *1*, 13-37.

Latif, M., E. Roeckner, M. Botzet, M. Esch, H. Haak, S. Hagemann, J. Jungclaus, S. Legutke, S. Marsland, U. Mikolajewicz, and J. Mitchell (2004), Reconstructing, monitoring, and predicting multidecadal-scale changes in the North Atlantic Thermohaline Circulation with sea surface temperature, *Journal of Climate*, *17*, 1605-1614.

Lindeberg, T. (1994), *Scale-space theory in Computer Vision*, Kluwer, Dordrecht.

Lomb, N. R. (1976), Least-squares frequency analysis of unequally spaced data, *Astrophys. Space Sci.*, *39*, 447-462.

Manabe, S., and R.J. Stouffer, (1999), The role of thermohaline circulation in climate, *Tellus*, *51*, 91-109.

Mann, M. E., and P. D. Jones (2003), Global surface temperatures over the past two millennia, *Geophysical Research Letters*, *30*, doi:10.1029/2003GL017814.

Mann, M. E, R. S. Bradley, and M. K. Hughes (1998), Global-scale temperature patterns and climate forcing over the past six centuries, *Nature*, *392*, 779-787.

Mann, M. E., R. S. Bradley, and M. K. Hughes (1999), Northern Hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations, *Geophysical Research Letters*, *26*, 759-762.

Mann, M. E., Z. Z. Zhang, M. K. Hughes, R. S. Bradley, S. K. Miller, S. Rutherford, and F. Ni (2008), Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, *PNAS*, *105(23)*, doi:10.1073/pnas.0805721105.

Mann, M. E.,Zhang, Z., Rutherford, S., Bradley, R. S., Hughes, M. K., Shindell, D., Ammann, C., Faluvegi, G., and F. Ni (2009), Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly, *Science*, *326*, 1256-1260, doi: 10.1126/science.1177303.

Marshall, J., Y. Kushnir, D. Battisti, P. Chang, A. Czaja, R. Dickson, J. Hurrell, M. McCartney, R. Saravanan, and M. Visbeck (2001), North Atlantic Climate Variability; phenomena, impacts and mechanisms, *International Journal of Climate*, *21*, 1863-1898.

Miettinen A., D. V. Divine, N. Koc, F. Godtliebsen, and I. R. Hall (2012), Multicentennial variability of the sea surface temperature gradient across the subpolar North Atlantic over the last 2.8 kyr, *Journal of Climate*, doi:10.1175/JCLI-D-11-00581.1, in press.

Miettinen, A., N. Koc, I. R. Hall, F. Godtliebsen, and D. V. Divine, (2011), North Atlantic sea surface temperatures and their relation to the North Atlantic Oscillation during the last 230 years, *Climate Dynamics*, *36*, 533-543, doi:10.1007/s00382-010-0791-5.

Milotti, E. (2005), Sine-fit procedure for unevenly sampled, multiply clocked signals, *Journal of Computational Physics*, *202(1)*, 134-149.

Moberg, A., D. M. Sonechkin, K. Holmgren, N. M. Datsenko, and W. Karlen (2005), Highly variable Northern Hemisphere temperatures reconstructed from low and high resolution proxy data, *Nature*, *433*, 613-617.

Möller-Levet, C. S., F. Klawonn, K.-H. Cho, H. Yin, and O. Wolkenhauer (2005), Clustering of unevenly sampled gene expression time-series data. *Fuzzy Sets and Systems*, *152(1)*, 49-66.

Olsen, L. R., S. H. Sørbye, and F. Godtliebsen (2008), A scale space approach for detecting changes in statistical behaviour of dependent data. *Scandinavian Journal of Statistics*, *35*, 119-138.

Park, C., and K.-H. Kang (2008), SiZer analysis for the comparison of regression curves, *Computational Statistics and Data Analysis*, *52*, 3954-3970.

Park, C., J. S. Marron, and V. Rondonotti (2004), Dependent SiZer: goodness-of-fit tests for time series models, *Journal of Applied Statistics*, *31*, 999-1017.

Park, C., F. Godtliebsen, M. Taqqu, S. Stoev, and J. S. Marron (2007), Visualization and inference based on wavelet coefficients, SiZer and SiNos, *Computational Statistics and Data Analysis*, *51*, 5994-6012.

Ouahabi, A., Depollier, C., Simon,L., and D Kouame (1998), Spectrum estimation from randomly sampled velocity data, *IEEE Transactions on instrumentation and measurement*, *47*, 1005-1012.

Scargle, J. D. (1982), Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data, *Astrophys. J.*, *263*, 835-853.

Shindell, D. T., G. A. Schmidt, M. E. Mann, D. Rind, and A. Waple (2001), Solar forcing of regional climate change during the Maunder Minimum, *Science*, *294*, 2149-2152.

Shumway, R. H., and D. S. Stoffer (2000), *Time Series Analysis and Its Applications*, Springer-Verlag, New York.

Swingedouw, D., L. Terray, C. Cassou, A. Voldoire, D. Salas-Melia, and J. Servonnat (2010), Natural forcing of climate during the last millennium:

fingerprint of solar variability, *Climate Dynamics*, doi:10.1007/s00382-010-0803-05.

ter Braak, C. J. F., and S. Juggins (1993), Weighted averaging partial least squares regression (WA-PLS); an improved method for reconstructing environmental variables from species assemblages, *Hydrobiologia*, *269/270*, 485-502.

Thomson D.J. (2000), Multitaper analysis of nonstationary and nonlinear time series data, *Nonlinear and Nonstationary signal processing*, 317-394. Cambridge University press.

Thornalley, D. J. R., H. Elderfield, and I. N. McCave (2009), Holocene oscillations in temperature and salinity of the surface subpolar North Atlantic, *Nature*, *457*, 711714.

Timmermann, A., M. Latif, R. Voss, and A. Gratzner (1998), Northern Hemispheric Interdecadal Variability: A coupled Air-Sea Mode, *Journal of Climate*, *11*, 1906-1931.

Trenberth, K. E., P. D. Jones, P. Ambenje, R. Bojariu, D. Easterling, A. Klein Tank, D. Parker, F. Rahimzadeh, J. A. Renwick, M. Rusticucci, B. Soden, and P. Zhai (2007), Observations: Surface and Atmospheric Climate Change, In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and

H. L. Miller, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Trouet V., J. Esper, N. E. Graham, A. Baker, J. D. Scourse, and D. C. Frank (2009), Persistent positive North Atlantic Oscillation mode dominated the Medieval Climate Anomaly, *Science*, *324*, 78-80.

Vellinga, M., and R. A. Wood (2002), Global climatic impacts of a collapse of the Atlantic thermohaline circulation, *Climatic Change*, *54*, 251-267.

Wanner, H., Beer, J., Bütikofer, J., Crowley, T. J., Cubasch, U., Flückiger, J., Goosse, H., Grosjean, M., Joos, F., Kaplan, J. O., Küttel, M., Müller, S. A., Prentice, I. C., Solomina, O., Stocker, T. F., Tarasov, P., Wagner, M. and M. Widmann (2008), Mid- to Late Holocene climate change: an overview, *Quaternary Science Reviews*, *27*, 1791-1828, doi: 10.1016/j.quascirev.2008.06.013.

Xian, J., J. Wang, and D.-Q. Dai (2007), Detecting periodically expression in unevenly spaced microarray time series, *Lecture Notes in Computer Science*, doi:10.1007/978-3-540-72584-8, Springer Berlin/Heidelberg.

Øigård, T. A., H. Rue, and F. Godtliebsen (2006), Bayesian multiscale analysis for time series data, *Computational Statistics and Data Analysis*, *51*, 1719-1730.

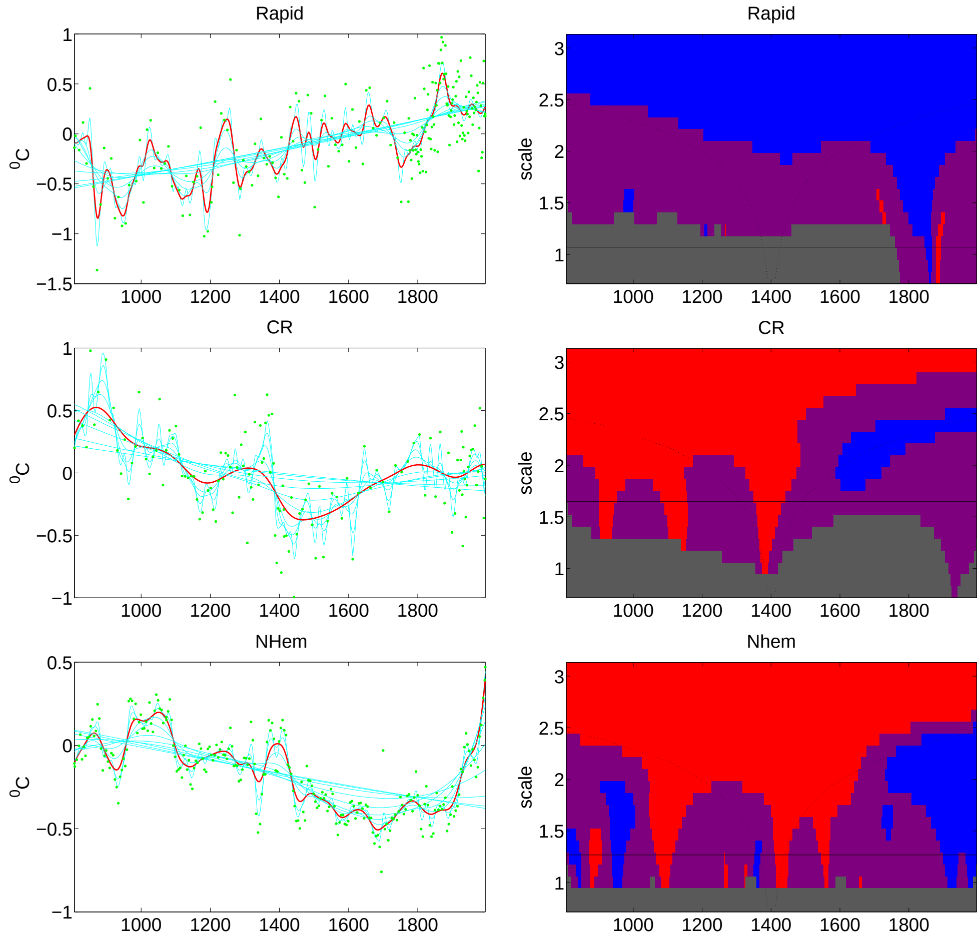

Figure 1. SiZer analyses of the three reconstructed temperature time series considered in the paper: Reykjanes Ridge SST (Rapid), Vøring Plateau SST (CR) and average Northern Hemisphere surface air temperature (NHem). Horizontal time scales indicate calendar years. On each row, the left panel displays a time series of reconstructed past temperatures (dots) together with a family of smooths. The right panel shows a SiZer significance map where, for a given time x and scale $s = \log_{10}(h)$, a pixel (x, s) is colored blue or red depending on whether the slope of the smooth of the true underlying temperature curve is significantly positive or negative, respectively. Purple indicates non-significance and pixels are colored gray if the data are too sparse to make any conclusions. The parallel distance between the dotted lines indicates the effective size of the smoothing kernel used for a particular scale, and hence gives an idea of the corresponding time-scale involved at that level of smoothing. The smoothing level corresponding to the red curve in the left panel is indicated by a horizontal line in the map.

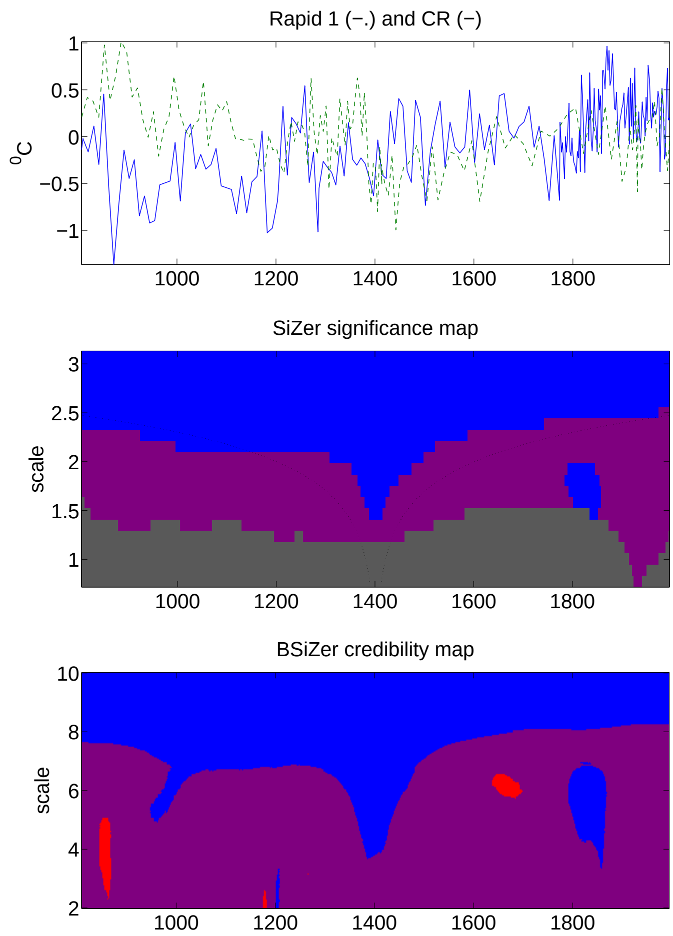

Figure 2. Scale-space comparison of Reykjanes Ridge (Rapid) and Vøring Plateau SST (CR). Horizontal time scales indicate calendar years. Top panel: the two reconstructed temperature time series; Middle panel: SiZer significance analysis of the slope of the difference Rapid - CR. Blue (red) for each time and scale indicates whether the slope of the smooth of the true underlying temperature curve is significantly positive (negative). Purple indicates non-significance and pixels are colored gray if the data are too sparse to make any conclusions. Bottom panel: Bayesian credibility map of the slope of the difference Rapid - CR. The BSiZer credibility map is interpreted analogously with blue (red) color at a pixel indicating a credibly positive (negative) slope, respectively and purple indicating no credible change.

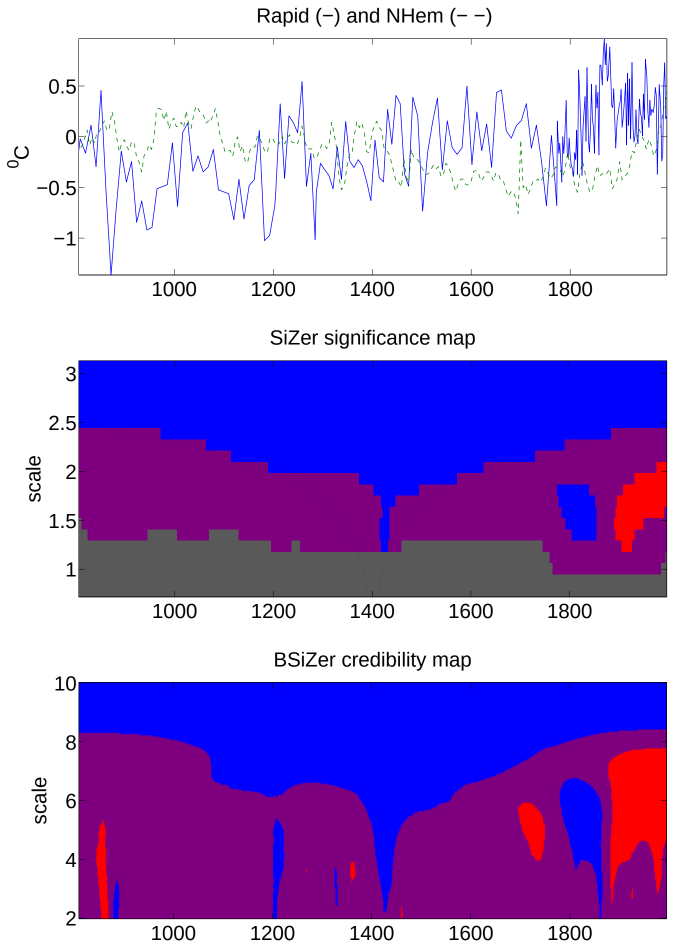

Figure 3. Scale-space comparison of Reykjanes Ridge SST (Rapid) and average Northern Hemisphere surface air temperature (NHem). Horizontal time scales indicate calendar years. Top panel: the two reconstructed temperature time series. Middle panel: SiZer significance analysis of the slope of the difference Rapid - NHem. Blue (red) for each time and scale indicates whether the slope of the smooth of the true underlying temperature curve is significantly positive (negative). Purple indicates non-significance and pixels are colored gray if the data are too sparse to make any conclusions. Bottom panel: Bayesian credibility map of the slope of the difference Rapid - NHem. The BSiZer credibility map is interpreted analogously with blue (red) color at a pixel indicating a credibly positive (negative) slope, respectively and purple indicating no credible change.

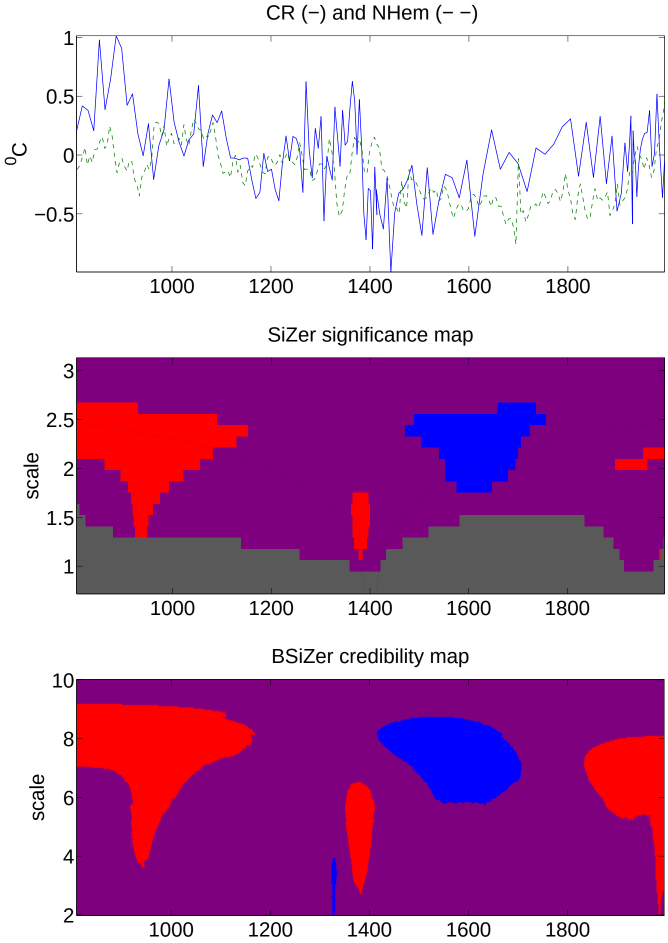

Figure 4. Scale-space comparison of Vøring Plateau SST (CR) and average Northern Hemisphere surface air temperature (NHem). Horizontal time scales indicate calendar years. Top panel: the two reconstructed temperature time series. Middle panel: SiZer significance analysis of the slope of the difference CR - NHem. Blue (red) for each time and scale indicates whether the slope of the smooth of the true underlying temperature curve is significantly positive (negative). Purple indicates non-significance and pixels are colored gray if the data are too sparse to make any conclusions. Bottom panel: Bayesian credibility map of the slope of the difference CR - NHem. The BSiZer credibility map is interpreted analogously with blue (red) color at a pixel indicating a credibly positive (negative) slope, respectively and purple indicating no credible change.