Manuscript prepared for Earth Syst. Dynam. Discuss.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 28 September 2015

Response to reviewer Shaun Lovejoy

T. Nilsen[1], **K. Rypdal**[1], **and H.-B. Fredriksen**[1]

[1]Department of Mathematics and Statistics, UiT The Arctic University of Norway, Norway

Correspondence to: Tine Nilsen (tine.nilsen@uit.no)

Abstract

Shaun Lovejoy's report is not a conventional review of our discussion paper, but rather an essay which is a critique of our paper, but also a review and defence of his own work. A point-to-point response is almost impossible, and we feel confident that this is not the kind of discussion that the reviewer wants. We will therefore address his comments by following the logical structure of the review. Our response contains a number of new analyses and figures. Some of the text and figures may be incorporated in the revised paper and Supplementary Material. Substantial revision of our paper will follow as a result of both reviews, but an outline of this revision will not appear in this response, but at the end of the reponse to reviewer #2.

1 Clarifying the basic issues

1.1 Scaling analysis of evolving systems

On geological time scales the Earth is an evolving system. There are cycles, but the Earth rarely repeats itself. The Eemian was similar to the Holocene, but also very different, the most striking difference being the evolution of human civilizations. Thus, the dynamics of the Earth is non-stationary in a very fundamental sense. This makes scaling analysis, and modelling of Earth processes based on such analysis, a quite problematic issue. It has little meaning to talk about a universal scaling in Earths climate since the scaling characteristic on a given range of scales up to a chosen maximal scale τ_{max} will depend on the eon, era, period, epoch, or age the analysis is done. In other words, the result will depend on the time t around which the time range τ_{max} is centred. A scaling analysis of a given Earth-system variable must therefore be be conditioned by two essential parameters; the range τ_{max} of scales considered, and the positioning t of this range in time. One obvious mathematical and conceptual tool for handling such non-stationary data is the wavelet transform, but more about that later.

The controversy we have with Shaun Lovejoy has its root in different understanding of how this issue of non-stationarity could be handled. Ice cores restrict the information we can obtain to somewhat less than $\tau_{\mathsf{max}} = 1$ Myr BP. This is the range of time scales considered in Lovejoy's work, and the period is the Quaternary (2.5 Myr to present) in which the Earth's climate has been in a bistable state shifting between glacials and interglacials. Lovejoy's methodology and interpretations are based on this choice of the parameters (τ_{max}, t). We don't see anything wrong with that, as long as one is mindful on that this is a choice, and recognises that there are other, equally valid, choices. It is here our views collide, because Lovejoy tells us that his choice is the only one worth to pursue.

1.2 Climate prediction in the Holocene based on scaling analysis

It seems unclear whether scaling analysis per se can be helpful in understanding glacial-interglacial transitions, but if this is the issue, Lovejoy's choice is certainly a reasonable one. If there are strong fluctuations on time scales of millennia they can contain the seed for a flip from an interglacial to a glacial and vice versa. On the other hand, if the issue is understanding of the present and future climate in our present interglacial state, we don't believe this choice is useful, simply because it ignores the knowledge that the Earth at present resides in an interglacial state and probably will continue to do so as long as there is human civilisation and anthropogenic forcing on this planet.

The time series and the wavelet scalogram of the GRIP temperature series for the past $t_{\mathsf{max}} = 90$ kyr illustrates the issue, and shown in Figure 1 of this comment. The central time parameter t is along the horisontal axis and the scale τ along the vertical. We have no data for the future, which means that the transform cannot be computed correctly above the upper white line in the figure. Likewise, the area below the lower white curve is influenced by the interpolation made due to uneven sampling of the time series. It is apparent that the scalogram is different in the first 11.5 kyr (the Holocene) from the remaining 80 kyr (the last glacial). There is generally lower power on all scales in the Holocene, and the increase in power with increasing scale as t is kept constant is lower. This is a signature of the different scaling exponents characterising the interglacial and glacial temperature

fluctuations in Greenland. The difference between glacial and Holocene fluctuations is also illustrated very clearly in Fig. 4a of Lovejoy's report. Unfortunately, the Holocene so far has lasted only about 11.5 kyr, which limits the scales accessible by the wavelet transform (or any other method of fluctuation analysis) to those confined by the upper white line. This means that we cannot say anything certain about Holocene scaling for scales beyond a few kyr.

It is not unlikely that the anthropogenic perturbation has overridden the orbital forcing for several millennia already (early deforestation) and that the Earth will remain in the interglacial state for sufficiently long time to make our remote successors able to establish accurately the scaling properties of Holocene climate up to scales of tens of kyr. It can be argued that the anthropogenic perturbation will change the scaling, and therefore that our successors will find scaling characteristics different from those that has ruled the Holocene until now. But, if we want to use a statistical model based on empirical scaling properties as a tool for prediction, we have to base it on observations in the past. The central issue is then whether we should use the scaling obtained from time series dominated by the glacial state ($\beta \approx 1.7$ for the GRIP record) or a model based on observations from the Holocene only ($\beta \approx 0.5$ for GRIP). Our position is that it is unreasonable to use data from the glacial state to make centennial-to-millennial predictions for the Holocene.

1.3 The issue of uncertainty and hypothesis testing

If we use only Holocene observations, the limited length of the time series confronts us with the issue of uncertainty. This issue is largely ignored in Lovejoy's work, and in section (c) of his review he makes an attempt to justify it. He also argues in section (a) that it is unnecessary to use statistical models or "any assumptions about the scaling or otherwise of the temperatures. Conclusions can be verified using straightforward fluctuation analyses."

In our ears this sounds like a rejection of the scientific method. Without models it is not possible to perform statistical hypothesis testing. Lovejoy's preferred methodology is to perform estimates on some samples and then draw positive conclusions regarding the validity of hypotheses, running a great risk of committing type-I statistical errors (false positives).

The widely accepted approach is to test a hypothesis (e.g., the hypothesis of a scaling break at 100 yr) against a simpler null hypothesis. This null hypothesis *must* take the form of a statistical model, otherwise one cannot assess the probability of obtaining the observed sample under the null hypothesis. The natural choice of a null hypothesis is the simplest possible noise model which seems compatible with the observed data, but still is distinguishable from the alternative hypothesis that we want to test. In section 1.8 we shall demonstrate that a fractional Gaussian noise (fGn) is the proper choice for the Holocene surface temperatures by showing that intermittency is negligible.

One result of our paper is that the fGn (single scaling regime) null hypothesis cannot be rejected for scales up to a millennium by available observation data from the Holocene. A stronger, and more important result, is that considering the scale regime from $10 - 10^4$ yr even a white-noise null model cannot be rejected by most proxies (see Fig. 2c and Table 1 of our paper). Lovejoy quite correctly points out that these findings do not reject the scaling-break hypothesis, but we never claim that. Our claim is that observations are consistent with the fGn null hypothesis, and because of its simplicity, this should be the preferred model for predictions on time scales up to several centuries. Lovejoy et al. have recently published a paper where in effect the fGn-model ($\beta < 1$) is used for prediction on time scales up to a few decades. Following his logic, prediction on longer time scales should employ an fBm-model ($\beta \approx 1.7$), or even some multifractal version.

1.4 The notion of a macroweather-climate scale break

The occurrence of a scale-break around time scales of $\tau_c \sim 100$ yr is used by Lovejoy to justify his notion of “macroweather", as opposed to “climate." GCM experiments with and without full ocean circulation suggest that the main mechanism for the creation of scaling and long-range memory in the climate system is the energy exchange between subsystems with a wide range of response times, and that there are response times both smaller and larger than the “magic” scale $\tau_c \sim 100$ yr of the macroweather-climate transition. Our perception is that when such a transition scale τ_c appears when non-detrending scale estimators are applied to data, it is always explicable as a result of a particular external forcing

or a distinct oscillatory mode. This is why τ_c in Lovejoy's work varies from 10 to 100 yr depending on which part of the Holocene his data cover, and for the Holocene Greenland data $\tau_c \sim 1000$ yr. We find no evidence supporting the notion that there is a universal transition time scale, and that strong large scale variability signifies a new scaling regime characterised by a scaling exponent. When scaling appears to be broken, it brings more useful insight to investigate the particular events that are causing it, rather than taking it as a natural transition to a new scaling regime that is called "climate."

Another point is that scaling is also broken at scales shorter than decadal. One striking example is ENSO, which destroys the scaling properties in the Pacific tropics and subtropics, and even affects the scaling of global temperatures. According to Lovejoy, the timescales of ENSO classifies it as a macroweather fluctuation. We, on the other hand, don't find any compelling reason to classify e.g., Dansgaard-Oesgher events as climate variability and ENSO as macroweather.

For these reasons we are not so enthusiastic about the macroweather-climate notion, but that is not the subject of our paper. The issue is the support one can find in the data, and here the main evidence is presented in fluctuation measures or spectra derived from composites of proxies representing different time intervals and scales. Examples of such composites are presented Figs. 1, 3a,b and 4b in Lovejoy's review report. We will comment on those in the next section.

1.5 The potential fallacies of composite spectra

Some composites shown by Lovejoy are intended to demonstrate the existence of regimes of different scaling, supposed to represent scale regimes at which different physics dominate the fluctuations. As mentioned earlier, fluctuations on long time scales can only be investigated by probing far back in time, possibly into climate states different from the present. In other words, it can be difficult to distinguish a scaling break from an evolutionary change (nonstationarity) of the climate system, i.e., from a change in time of the scaling exponents. An illustrating example is Fig. 1 in Lovejoy's report. The figure shows the square-root of the second-order structure function for 5 different proxies. For the scales up to 400 yr it

shows the fluctuations of the 400 yr long Central England Temperature record (CET), and for a Northern Hemisphere mean temperature for 1850-1969 (*Budyko*, 1969). For a process where fluctuations decrease with incrasing scale ($\beta < 1$), this fluctuation measure gives a flat curve in a log-log plot of fluctuations vs. scale. Hence, for the CET record it is flat because the trend is weak. The NH-mean series, however, covers only the industrial period and is dominated by the anthropogenic warming trend. For a trend-dominated signal the characteristic exponent estimate H for this structure function is positive (actually close to unity). Similar results are obtained on scales 10^2 – 10^5 yr obtained from ice cores, but the nature of the signal causing these similar exponents is profoundly different. It illustrates that scaling analysis based on one single estimator and without careful physical interpretation can be very misleading.

In this plot all observations used for scales < 400 yr are made in the Holocene, while almost all observation for longer scales are based on observations during glacial periods. Hence, what appears as a scaling break around 400 yr when the CET record is used for the short scales, could just as well be due to different scalings in the glacial and interglacial states.

In Fig. 3b in Lovejoy's report the Haar fluctuation of some short multiproxy reconstructions (1500-1979 AD) of Fig. 3a are combined with some very long proxy records. The EPICA record is not so interesting in the present context, since it is necessarily dominated by the glacial state. The *Marcott* (2013) reconstruction, however, covering the entire Holocene, might have the potential to be the answer to our prayers about reconstructions with global distribution that covers the entire Holocene. Lovejoy's analysis, however, suffers from a serious flaw due to uncritical use of these paleoseries. In Fig. 3b he shows the Haar fluctuation for four different series, two for global temperature and two for 30-90°N. Each pair consists of a long series covering the entire Holocene and a short series covering 1500-1900 AD. *Marcott* (2013) write that the long series recover no variability for scales less than 300 yr, 50% of the variability on scales of 1000 yr, and all variability on scales greater than 2000 yr. The short series presumably recover virtually all variability on the (short) scales they cover. This low-pass filtering of the long series by the reconstruction method alters the scaling

and creates a spurious increase of the scaling exponent. This fact is very apparent from Lovejoy's plots, since the fluctuation level derived from the short series is almost one order of magnitude higher than derived from the long series on the scales where the two series overlap. They cannot both be correct, and it is reasonable to assume that the short series give the correct fluctuation on the short time scales around 10^2 yr where the long series is filtered. If we compare this fluctuation level with the level on scales around 10^3 yr derived from the long records, we observe that they are of similar magnitude. In other words, a critical assessment of these data show that the real fluctuation level on centennial and millennial scales are of similar magnitude, consistent with $H \approx 0$, $\beta \approx 1$.

In Figure 2 we demonstrate this by analyzing the reconstruction by *Marcott* (2013) with the periodogram in a particular way to overcome the gradual smoothing as one goes back in time. For the full record, the variability should be trusted only for time scales longer than 2000 years. On shorter time scales, the power is artificially low due to the smoothing. To overcome the smoothing problem, the time series was divided into segments of 2^n * 400 years, with n=0, 1, 2....5 and starting with the most recent period. Segment number: 1=50-450 yr BP, 2=50-850 yr BP, 3= 50-1650 yr BP, 4=50-3250 yr BP, 5=50-6450 yr BP, 6=50-11290 yr BP (longest possible record, shorter than $2^5 * 400$). The periodogram was estimated for each segment, and then a new power spectrum was created using only parts of each segment assumed to be trustworthy with regard to preserved variability. All parts of segment 1 were included, while for segment 2-6 only the low-frequency parts were included (none overlapping). By this composition, the resulting power spectrum represents the variability on all time scales more correctly. The estimated spectrum displays only one scaling regime with $\beta \approx 1.3$, while the spectrum of the full, raw time series exhibits a scale break and $\beta > 2$ in the regime of scales longer than a century, similar to what Lovejoy finds using the Haar fluctuation function.

The point of including Fig. 4b in the review report is hard to understand. The only curve shown here that exhibits a break around 10^2 yr is the mean of a number of ice core records for 10-90 kyr BP. We have never disputed the existence of such a break in the glacial state.

1.6 The issue of scaling – the effect of trends and oscillatory modes

The Haar fluctuation has the advantage with respect to the standard structure function of not going flat for processes with decreasing fluctuations versus scale, and hence works for β both larger and smaller than unity. In Fig. 3a of Lovejoy's report the Haar fluctuation is plotted for the instrumental temperatures and for two multiproxies for the period 1500-1979 AD. All the curves show a break, but at different time scales. The instrumental curve breaks close to 10 yr, while the multiproxies break closer to 50 yr. The slopes after the break are close to those typical for a signal dominated by a trend. For the instrumental record the fluctuation function for the linearly detrended record shows an oscillation after the 10-yr break. This is easy to interpret by noting that the trend in the anthropocene is not linear – a quadratic trend is much more representative of the anthropogenic forcing.

In Figure 3a we show the Haar fluctuation for the full instrumental record, and for the linearly and quadratic detrended record. For the latter, the scale break disappears completely, suggesting that the internal variability follows the same scaling as on the smaller time scales. In Figure 3b we demonstrate the same feature for the Central England Temperature, but for this 350 yr record a linear detrending is suffcient to restore scaling in the Haar structure function for all scales from months to the length of the record.

We agree with Lovejoy that the anthropogenic warming destroys the scaling, but our perception is that it is misleading to think about that as a new scaling regime characterised by a scaling exponent. In the revised manuscript we elaborate on this. We demonstrate that the observations are consistent with both a two-scaling regime model and a one-scaling regime + trend model, but that the latter constitutes a "better" statistical model because it yields much lower errors (uncertainty) on the long time scales. It also has the advantage that part of the model (the trend) makes use of existing knowledge and solid physical understanding. In our opinion statistical modelling should be reserved for those phenomena that cannot be described by simple deterministic models. This is, for example, the idea behind regression analysis; those parts of the variability which can be "explained" by deterministic predictors should be described as such, and in traditional regression the residual is often

assumed to be a Gaussian white noise. In our opinion the main goal of scaling analysis of climatic time series is to establish the true nature of this residual. As will be shown below is that the fractional Gaussian noise is a good model for this residual for Holocene surface temperatures.

In Figure 4 we compute structure functions (empirical moments of order $q = 1, 2, \ldots$) for the global mean surface temperature (GMST) and for its cumulative sum. Structure functions (SF) for the signal itself should be flat for a process where the fluctuations do not increase with increasing scale ($\beta < 1$ if the process is monofractal). For the GMST the high-order SFs are not straight lines in a log-log plot, but curve upwards for scales larger than 10 yr. The reason for this is the trend, which also causes the break in the Haar fluctuation at this scale. However, the collection of SFs for different q contain additional information to the second order statistics expressed by the Haar fluctuation curve, and is more clearly exposed for the detrended signal, as shown in Figure 4b. These SFs are almost flat, but exhibit two peaks corresponding to the annual cycle and another cycle of period of about 20 yr. If there had been an underlying scaling with $\beta > 1$ ($H > 0$) this scaling would have dominated the structure functions and given straight lines with positive slopes $\zeta(q) \equiv \log S_q(\tau)/\log(\tau) = Hq$.

In order to test if the GMST is a monofractal noise process ($H < 0$, $\beta < 1$), we form its cumulative sum (cumsum), which then should be a self-similar process with $H \to H + 1$, $\beta \to \beta + 2$. For scales < 10 yr the structure functions behave as for a monofractal, while for larger scales they bend over. If the scaling function $\zeta(q)$ is computed from the slopes of the SFs up to the 10-yr scale, it is a straight line with a slope (Hurst exponent) $H_u = H + 1 = 0.97$, corresponding to $H = 0.97 - 1 = -0.03$ ($\beta = 0.94$). This is shown in Figure 5a. However, the value of H_u and β close to unity is a typical signature of a trend-dominated process. If we subtract the quadratic trend, and repeat the analysis, the cumsum SFs are still straight and so is the scaling function. But now the Hurst exponent is reduced to $H_u \approx 0.85$ ($\beta \approx 0.70$), as shown in Figure 5b. This is the "true" scaling exponent of the natural variability.

The curving of the SFs for large scales is probably a consequence of the 20 yr oscillation and another oscillation of period around 70 yr. In the second-order Haar fluctuation analysis these oscillations are indistinguishable from true monofractal scaling. Note that it is not the Haar wavelet itself that is the limitation, but the restriction to using only the *second-order* Haar structure function. By using higher-order statistics we reveal the non-scaling nature of some of the fluctuations. By employing the same methodology to the CET and the proxy data, we will observe that the scale-breaks that appear in the Haar fluctuation are due to such non-scaling fluctuations. We will include such evidence in the revised Supplement.

1.7 The issue of scaling – forced versus internal variability

Similar reasoning as presented above pertains also to the multiproxy fluctuations. The periods 1500-1979 AD and 1500-1900 AD are dominated by the warming as the Earth came out of the Little Ice Age (LIA), and the scale break will disappear with detrending. It may be argued that removal of fluctuations by detrending is unjustified, but that depends on whether we prefer to consider scaling as a property of internal climate variability or as a property of the total forced climate signal.

In modern climate science the various natural and anthropogenic drivers of climate variability have been quantified in the form of time series and makes it possible to separate the internal from the forced climate signal based on linear models for the forced response (*Rypdal and Rypdal*, 2014; *Rypdal et al.*, 2015). In General Circulation Models (GCMs) the internal variability can be studied in unforced control simulations, and the scaling of internal and forced temperature fluctuations can be compared (*Østvand et al.*, 2014). The cited studies conclude that the low temperatures during the LIA can be attributed to a combination of volcanic and solar forcing, and hence a detrending of the multiproxy signals is justified if one wants to get closer to the signal representing the internal variability. We conclude that the scale breaks observed in Lovejoy's Fig. 3a arise from the forcing, and is not a property of the internal climate variability. The scale break in the instrumental series is obviously associated with a unique event, the industrial revolution, and this may also be the case with the multiproxy records and the LIA. In order to clarify how unique the fluctuations

like the LIA is, we need to analyse considerably longer time series. This is what we have done in the paper. In some records we have found indications of higher power on low frequencies than consistent with an fGn null hypothesis, but this has disappeared when the forced variability has been separated out.

1.8 The issue of scaling – no intermittency in Holocene temperatures!

Let us first agree with the reviewer that the use of the term "monoscaling" and "multiscaling" towards the end of the paper were misnomers, although it should be clear from the context that what we mean here is "single scaling regime" and "multiple scaling regimes." In his section on "the issue of scaling" Lovejoy presents a lengthy introduction to intermittency and multifractals and end up with the claim that our "restriction to nonintermittent models is unnecessary and unrealistic." And further: "The monofractality - or lack of intermittency - must be quantitatively established not simply assumed a priori."

Since our focus is on Holocene data, we shall establish monofractality of such data here. The evidence presented by Lovejoy for his claim of intermittency is non-Gaussian tails of some PDFs for differences ΔT for time lags $\Delta t = 1, 4, 16, 64$ derived from a multiproxy record. The tails allegedly have the power-law form ΔT^{-5}. This "heavy" tail is in practice indistinguishable from an exponential, and the existence of such tails alone is not a signature of intermittency/multifractality.

For the instrumental global mean temperature the Gaussianity and monofractal scaling was established by *Rypdal and Rypdal* (2010). We refer to this paper for details of the analysis. Since the instrumental data covers a quite limited range of time scales we shall show a similar analysis for the the GRIP ice core δ^{18}O series for the Holocene and the Moberg Northern Hemisphere multiproxy temperature reconstruction. In Figure 6a we show the PDF of the GRIP δ^{18}O anomaly. It is slightly skewed, and a negative tail slightly heavier than a Gaussian. Panel (b) shows a so-called Quantile-Quantile plot, where the quantiles of the data is plotted against those of a normal distribution. The non-Gaussian negative tail shows up as the deviation from the dashed line in the left part of the plot. In panel (c) and (d) we make similar plot as made by Lovejoy, i.e., we plot the probability that $|\Delta(\delta^{18}\mathrm{O})|$ exceeds

a threshold $(\delta^{18}O)_{th}$. Panel (c) is a log-plot and shows that the tail is close to exponential. Panel (d) is a log-log plot and shows that the tail is not a power-law.

A multifractal analysis and test of intermittency can be done by computing structure functions to high order and plot the associated scaling function. Figure 7a shows structure functions and Figure 7b the corresponding scaling function for the GRIP data. The structure functions are straight lines in a log-log plot up to scales of 2000 kyr. The depletion for the highest structure functions beyond that scale is a signature of insufficient statistics on these scales. This is about 1/5 of the total length of the data record, and illustrates our claim that we cannot faithfully estimate scaling properties on scales longer than this. The record is monofractal if the scaling function is close to a straight line. The scaling function shown here indicates that the intermittency is very weak. The main reason for the heavy negative tail is the 8.2 kyr event discussed in the paper. In Figure 8 we show QQ-plots and tail PDFs for the same time record, but truncated at 7.5 kyr BP. Exclusion of the event creates a PDF very close to Gaussian. The remnant of a negative tail still present is probably caused by the forcing from volcanic eruptions. Figures 9 and 10 show similar results for the Moberg multiproxy record, which turns out to be even less intermittent. Instead of $\delta^{18}O$ it is here referred to $|\Delta T|$ exceeding a threshold T_{th}. The conclusion is that the fractional Gaussian noise is a *very* accurate model for Holocene temperatures up to the time scales where the structure function plots start to deviate from straight lines. For the Moberg record this maximal scale is around 400 yr. Beyond this time scale we cannot conclude anything about the scaling from this record with statistical confidence.

As a contrast we show in Figure 11 and 12 a structure function analysis of the GRIP $\delta^{18}O$ record for the last glacial maximum. For scales longer than a few decades (on smaller scales the record is smooth due to interpolation), the second-order structure function suggests mono-scaling with $h \approx 0.3$ ($\beta \approx 1.6$) as shown in Figure 11. However, the higher-order structure functions shown in Figure 12b are not straight in the log-log plot for scales longer than a few decades. If we ignore that and fit straight lines to the SFs in two scale regimes as shown in the figure, we find scaling functions that look multifractal, but this shouldn't lead us to believe that these data can be modelled as a simple multifractal process. The skewness

of the PDF shown in Figure 12a, and the curved SF suggest that this record requires more complex modelling.

2 Uncritical treatment of paleoseries?

The selection of multiproxy series in our paper is far from uncritical, but was a careful assessment based on Lovejoy's previous works. The Marcott reconstructions are interesting and should be discussed in the paper, but as discussed in section 1.5, we believe that the reviewer's analysis of these series are flawed.

Lovejoy is probably right in that the multicentennial variability of the multiproxy records is not a scientifically settled issue, but that only adds to the statistical uncertainty of low-frequency scaling which is the main proposition of our paper.

2.1 Greenland Holocene ice core records and the Berner SST reconstruction

For the Holocene the Greenland record behaves similar to instrumental records for continental interiors. These are characterised by low persistence ($\beta \approx 0$). The Berner reconstruction is similar to instrumental SST-records, which are strongly persistent and sometimes with $\beta > 1$. This does not mean that the Greenland record is more "exceptional" than Berner SST. Both are local temperature proxies, but since 70% of the Earth surface is ocean, the SST is more representative for the global temperature. The Haar fluctuation analysis of the Berner record for the Holocene shown in Lovejoy's Fig. 4b does not show a clear scale break, but a rather flat Haar fluctuation spectrum, corresponding to $\beta \approx 1$. Hence this record seems to confirm our conclusions, rather than refuting them.

It seems rather odd that Lovejoy now argues so strongly that the Greenland records are useless for scaling assessments, considering that he has used them extensively in the past for this purpose, but then using records that span mostly the glacial period. We conclude from this that he now admits that the scale-break hypothesis has been based on non-representative data, but that he believes more recent data are more representative and support the hypothesis.

3 Additional technical points

3.1 Statistical testing and significance

The reviewer's comments on this issue is consistent with the fact that error bars are absent in the figures he presents in his review report, and generally throughout his papers. Error analysis is very important because, if done properly, it forces us to be precise on (i) which estimator we use for hypothesis testing, (ii) which null hypothesis we choose, and (iii) which (alternative) hypothesis we want to test. Without these elements, the concept of statistical significance has no precise meaning.

(i) *Which estimator do we use, and why?*

In this paper we have chosen to estimate the power spectral density (PSD) by means of the periodogram. We have chosen to estimate the PSD because it is widely used and does not eliminate trends. It also works equally well for processes with fluctuations growing ($\beta > 1$) and decreasing ($\beta < 1$) as scale increases. The Haar wavelet has many of the same properties, but as far as we know, neither the periodogram nor the Haar wavelet has been tested for biases and errors as we did with the DFA and the Mexican hat wavelet in *Rypdal et al.* (2013). This should be done, but there is little reason to believe that biases and errors for either are very different from those we have tested. DFA(n) with $n \geq 2$ is insensitive to a linear trend, which means that for an ensemble of fGns with a linear trend superposed it will give an unbiased estimate of the scaling exponent for the underlying fGn. In this respect it performs similar to the Mexican hat wavelet, which eliminates linear trends because it is a symmetric wavelet. Estimators that eliminate a linear trend effectively uses the shorter time scales for the estimation of the scaling exponent. Therefore they are not suitable for detecting a scale break or a trend. This is why we have not used such an estimator in our analysis, and the reviewer's comments on DFA and trend detection is therefore completely beside the point. Whether we use the periodogram or the Haar wavelet in the analysis is probably unimportant. The important thing is to estimate the error bars for the estimator

under a null hypothesis, as we do for the periodogram by the blue-shaded areas in all figures where we show power spectra.

A principal point is that there is no such thing as a "correct" or "incorrect" estimator, but some are more sensitive than others when it comes to detctection of particular features. If an estimator fails to reject the null hypothesis, it just means that the alternative hypothesis is insignificant with respect to this particular estimator. If we are able to find another estimator that allows us to reject the null hypothesis from the data, while the alternative hypothesis is not rejected, then the alternative is significant under the chosen null.

From these considerations one might be lead to search for estimators that are particularly sensitive to the large scales, such as the Haar wavelet. The problem, however, is that such estimators are also particularly sensitive to the large scales in the random noise that constitutes the null hypothesis, giving rise to large error bars at these scales. This is exactly what is observed is our spectra, where the error bars are much wider at the low frequencies.

(ii) *The choice of null hypothesis, and why?*

The null hypothesis cannot be chosen subjectively to the same extent as the estimator, since the null should represent a plausible explanation of the observed data provided the alternative hypothesis is false. In section 1.8 we demonstrated that a Gaussian monoscaling process with $\beta < 1$ is the appropriate null model for Holocene temperatures.

(ii) *The choice of alternative hypothesis*

One of the most serious problems in Lovejoy's reasoning is that there is no clear hypothesis, i.e., no quantitative model for the large-scale fluctuations. The notion of a "break in scaling" is used, and figures are presented that give the reader the idea that there is a regime for large scales that can be characterised by a scaling exponent ($\beta > 1$) which is larger than for the short scales. This interpretation is underscored by straight lines drawn in the log-log plots of the Haar fluctuation for the large scales. In the text of his review report, however, he

downplays the significance of self-similar scaling on the large scales and reduces the issue to a question of the existence of large fluctuations on these scales, regardless of the cause and nature of these fluctuations. They could just as well be trends caused by specific, well understood forcings, or well established internal oscillations – all fluctuations that do not exhibit scaling.

In our paper we draw a distinction between these two alternative hypotheses. In the figures where we compute periodograms with the blue-shaded error fields (red-shaded for the Moberg record in Fig. 2c) we just compute the 95% error bars corresponding to the fGn null hypothesis, and compare it with the estimated spectrum from the actual observed record. For the Holocene records, the fGn hypothesis for scales up to a millennium cannot clearly be rejected by the observations we have considered. But this is not a very strong result. The difference between the Medieval Warm Anomaly (MWA) and the Little Ice Age (LIA) is so large in the Moberg record that it is not a very likely outcome of an fGn-fluctuation. But it is very explicable as a result of a combination of volcanic and solar forcing (volcanic more important than solar). This was shown in Figures 2b and 3c of *Østvand et al.* (2014), where the residual after subtracting the forced response was clearly within the confidence range of the fGn. Thus, the fGn null for *internal* variability is not rejected by the Moberg record.

Our main focus, however, is on the hypothesis of a low-frequency scaling regime with $\beta > 1$ in the Holocene. We then have to estimate β from the few independent measurements we have for those scales. The result for the Moberg record is shown in Figure 2b, and for the other multiproxies in Figure 3 and Table 1. Supposing that the scale break is at 100 yr, we have effectively $N/100$ independent observations for estimation of the scaling exponent of this regime, if the resolution is annual and N is the length of the record measured in in years. For the CET record we then have 4 independent measurements. For the Moberg record we have 20, and for the Marcott reconstruction 100. If we could trust that the Marcott reconstruction has effective resolution < 100 yr, it would have been the best candidate for estimation of the scaling exponent of the low-frequency regime. But we know that this reconstruction suppresses variability on scales less than 2000 yr.

3.2 The "rule of the thumb"

The "ill-starred rule of the thumb" of not drawing conclusions about scaling properties for scales corresponding to more than one fourth of the length of the sample at hand, is not particular to the DFA estimator. For the large scales (compared to the length of the sample record) the uncertainty is large for any estimator, and if we use these scales when estimating scaling exponents from single realisations in the MC ensemble, we end up with large uncertainties in the estimated exponents. In other words, the rule of the thumb is chosen such that the estimate of the scaling exponent from a realisation of a known monoscaling process is reasonably well confined. If we use all available scales for this estimate, the estimate is so uncertain that it gives no useful information about the scaling exponent of the underlying process, even if it is perfectly monoscaling. The reviewer writes about the significance of "events that are four or five standard deviations from the mean." But that is not what we observe in Holocene climate. In the spectra we hardly find events that are two standard deviations from the mean.

3.3 Mexican Hat and Morlet wavelet vs. Haar wavelet

The reviewer's comments about our use of wavelets are also beside the point. We don't use the wavelets for scaling analysis, only for demonstrating the influence of the 8.2 kyr event on the power spectra. We have decided to use the term "scalogram" instead of "wavelet power spectrum" from now on and in the revised paper. The Mexican hat and Morlet wavelets are similar to local Fourier transforms and hence their scalograms are suitable as local supplements to the periodograms. A known weakness of the Haar wavelet is stronger spectral leakage (see e.g., textbook of Percival and Walden, 2008), which is not so important for scaling analysis, but a drawback if we want to study the effect of local events on spectra. The Mexican hat and Morlet are also more sensitive to oscillations, and can be used to detect those in the noise. In this respect the Haar wavelet (there are different versions) performs more similar to the DFA.

Lovejoy has developed an algorithm for using the Haar wavelet on unevenly spaced data. This algorithm is explained in appendix A of his 2014b paper. We agree that linear interpolation is not optimal, but in our case it is not a problem. Our wavelet scalograms show the lower white curve where interpolation has an effect. We are interested in power increases on centennial time scales, and interpolation does not affect these time scales.

4 Concluding remarks

Lovejoy's review contains no concrete suggestions for revision of our paper. We cannot agree with his assertion in his section (a) that the paper is particularly technical, and that these technicalities obscure the real issue. On the contrary, some of his comments on technical points in his section (c) are largely beside the point and only serve to obscure the issue. Other comments in this section are at odds with sound approaches to statistical hypothesis testing.

Our "uncritical use of paleoseries" claimed in his section (b) is the same use as made by Lovejoy himself in quite recent papers, but we shall happily include a critical analysis of the Holocene multiproxies of *Marcott* (2013) in a revised paper.

In spite of our disagreements, the review has been helpful in bringing the nature of these to the forefront, and we will try to incorporate the essence of this response document into the paper. Some text and a few figures will be included in the main paper, and further text and figures in the Supplementary Material.

Acknowledgements. This work was funded by project no. 229754 under the the Norwegian Research Council KLIMAFORSK programme. We acknowledge in-depth discussions with Martin Rypdal during the work with this response, and also have made the use of routines for generating Monte Carlo ensembles of fractional Gaussian noises and for structure function and scaling functions produced by Martin Rypdal and Ola Løvsletten.

References

Berner, K.S., Koc, N., Divine, D., Godtliebsen, F., and Moros, M.: 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, 10.1029/2006PA001339, 2008.

Budyko, M. I..: The effect of solar radiation variations on the climate of the Earth, Tellus, 21, 611-619, 1969. Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, J. Geophys. Res, 111, D12 106, 2006.

Marcott, S. A., Shakun, J. D., Clark, P. U., and Mix, A. C.: A reconstruction of the regional and global temperature for the past 11.300 years, Science, 339, 1198, doi: 10.1126/science.1228026 (2013). Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, J. Geophys. Res, 111, D12 106, 2006.

Rypdal, M., and Rypdal, K.: Testing Hypotheses about Sun-Climate Complexity Linking, Phys. Rev. Lett. 104, 128501, doi: 10.1103/PhysRevLett.104.12850, 2010.

Rypdal, K., Østvand, L., and Rypdal, M.: Long-range memory in Earth's surface temperature on time scales from months to centuries, J. Geophys. Res, 118, 7046-7062, doi:10.1002/jgrd.50399, 2013.

Rypdal, M., and Rypdal, K.: Long-memory effects in linear-response models of Earth's temperature and implications for future global warming, J. Climate, 27, 5240-5258, doi:10.1175/JCLI-D-13-00296.1, 2014.

Rypdal, K., Rypdal, M., and H.-B. Fredriksen: Spatiotemporal Long-Range Persistence in Earth's Temperature Field: Analysis of Stochastic-Diffusive Energy Balance Models, (in press) J. Climate, 2015.

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

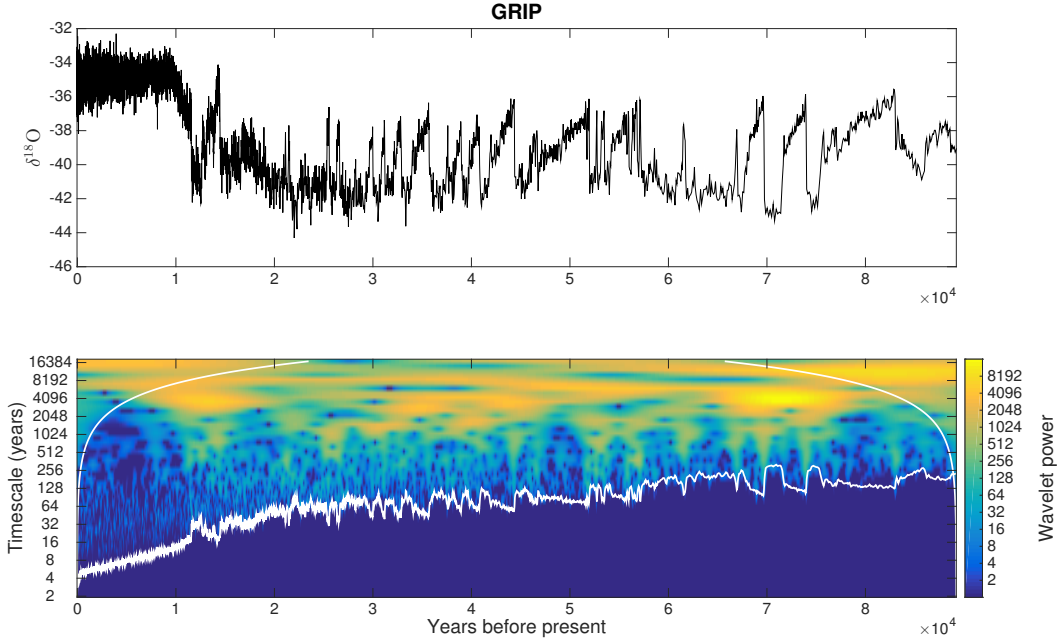

Figure 1. Upper panel: The δ^{18}O proxy rime series for Greenland temperature from the GRIP ice core for the period 0 – 90 kyr BP. Lower panel: The Morlet wavelet scalogram for the signal in the upper panel.

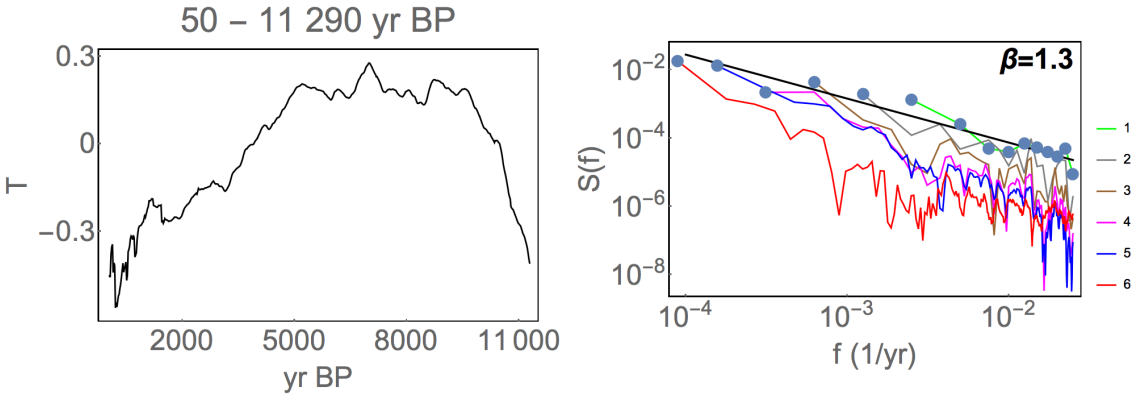

Figure 2. (a): The Marcott global multi proxy reconstruction. (b): Periodograms of time series in (a) by dividing into segments of 2^n * 400 years, with n=0, 1, 2....5 and starting with the most recent period. Segment number: 1=50-450 yr BP, 2=50-850 yr BP, 3= 50-1650 yr BP, 4=50-3250 yr BP, 5=50-6450 yr BP, 6=50-1 290 yr BP (longest possible record, shorter than $2^5 * 400$). The periodogram is estimated for each segment, and then a new power spectrum created using only parts of each segment assumed to be trustworthy with regard to preserved variability (the blue dots). All parts of section 1 are included, while for section 2-6 only the low-frequency parts are included (none overlapping).

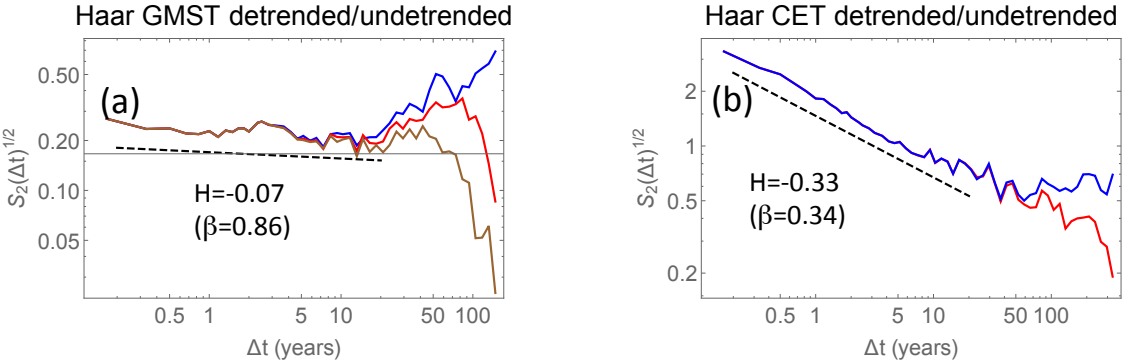

Figure 3. (a): Haar fluctuation for GMST. Undetrended (blue), linearly detrended (red), and quadratically detrended (brown). (b): Haar fluctuation for CET. Undetrended (blue), linearly detrended (red).

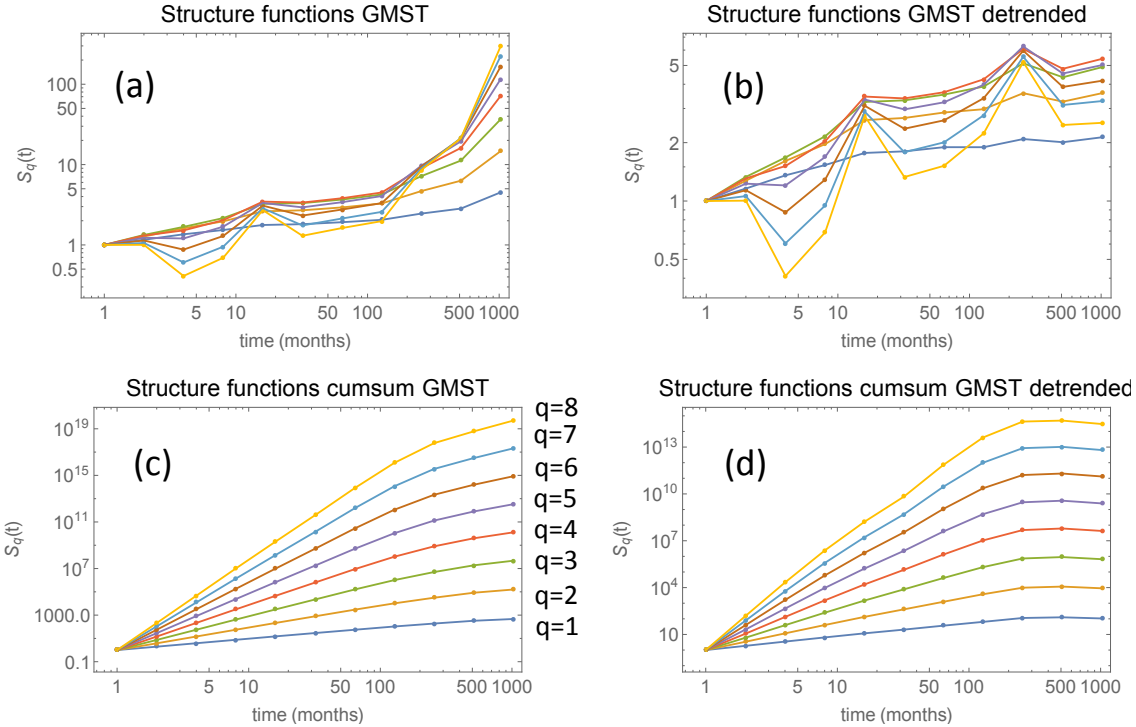

Figure 4. (a): Structure function estimates (empirical moments) $S_q(\tau) = (N-\tau)^{-1} \sum_{i=1}^{N-\tau} |T(t_i + \tau) - T(t_i)|^q$ for the GMST (HadCrut3) monthly record 1880-2010; $T(t_i)$; $i = 1, \ldots, N$. (b): Structure function for the quadratically detrended GMST (residual after subtraction of a second-order polynomial fit). (c): Structure function for the cumulative sum $y_{t_i} = \sum_{j=1}^{i} T(t_i)$. (d): Structure function for the cumulative sum of the quadratically detrended GMST.

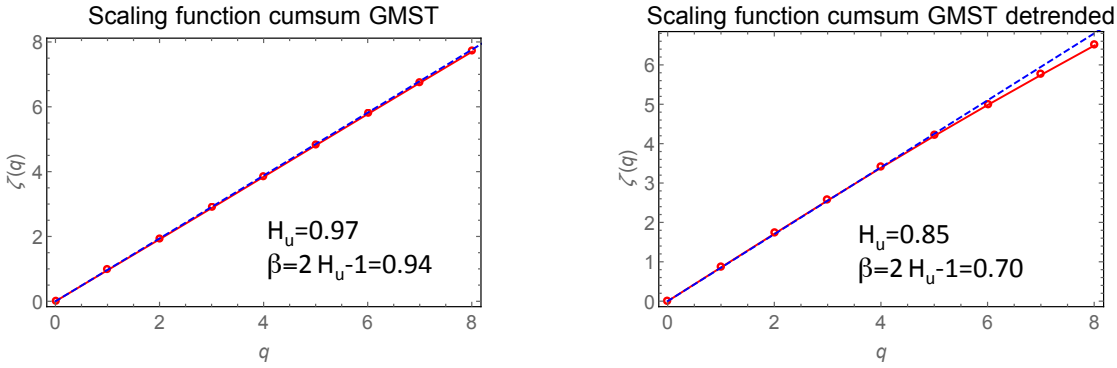

Figure 5. (a): The scaling function $\zeta(q)$ defined through the relation $S_q(\tau) = \tau^{\zeta(q)}$ for the cumulative sum (i.e., as the slope of the $\log S_q(\tau)$ vs. $\log \tau$). The slopes have been computed from the structure functions in Fig. 4c. (b): The same as in (a), but for the cumulative sum of the quadratically detrended GMST.

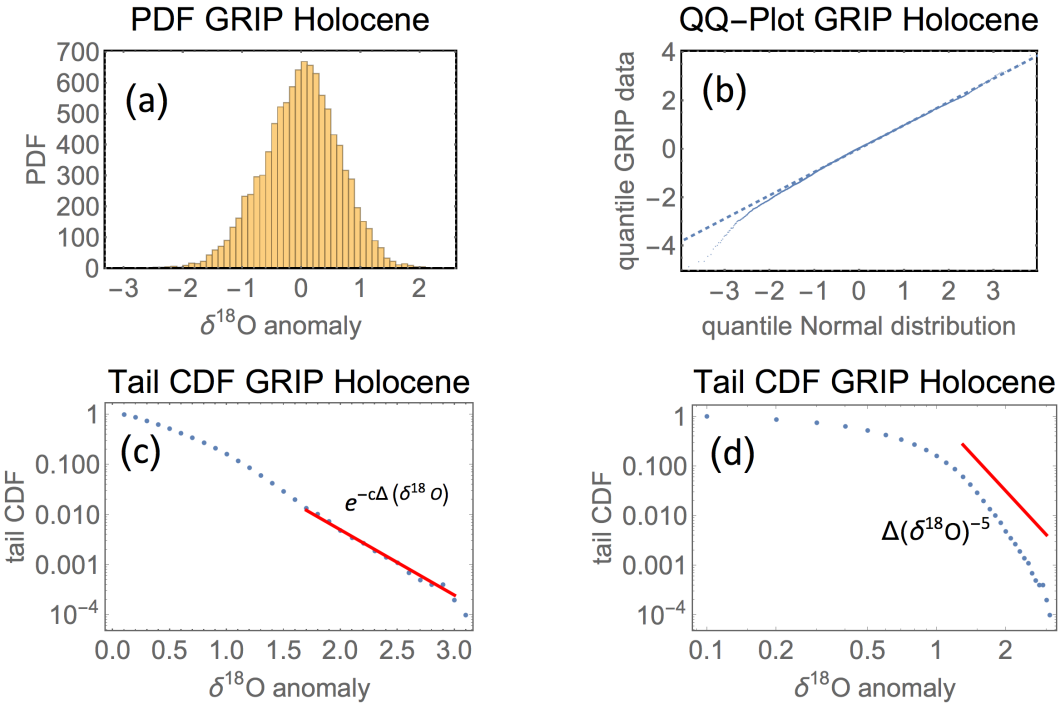

Figure 6. (a): Probability Density Function (PDF) of GRIP temperature anomaly for Holocene data 0 – 10500 yr BP. QQ-plot for these data. (c): Tail Cumulative Probability Function (probability for $\Delta(\delta^{18}O)$> threshold) for the data in log-plot. (d): Same as (c) in a log-log plot.

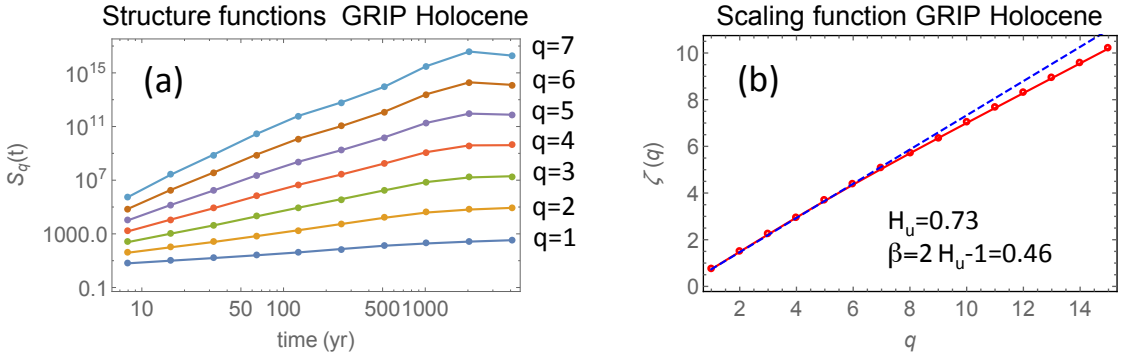

Figure 7. (a): Empirical moments (structure function estimates) for the GRIP Holocene data for moment orders $q = 1, \ldots, 7$. (b): The corresponding scaling function estimate $\zeta(q)$ for $q = 1, \ldots, 15$ estimated as the slope of the empirical moment curves in the range 8-1024 yr. The dashed line has slope $H_u = 0.73$ (Hurst exponent), which corresponds to the spectral index $\beta = 2H_u - 1 \approx 0.46$.

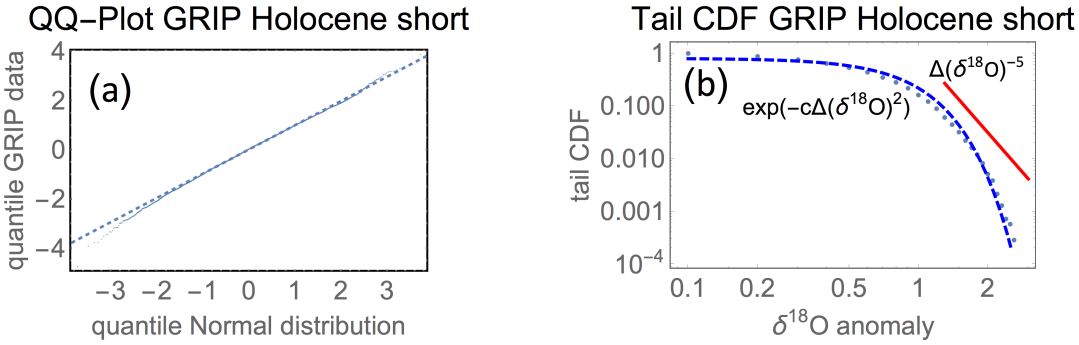

Figure 8. Analysis of the GRIP Holocene data in the range 0 – 7500 yr, i.e., without the 8.2 kyr event. (a): The QQ-plot. (b): The tail CDF (bullets) and the fitted Gaussian for these data (dashed).

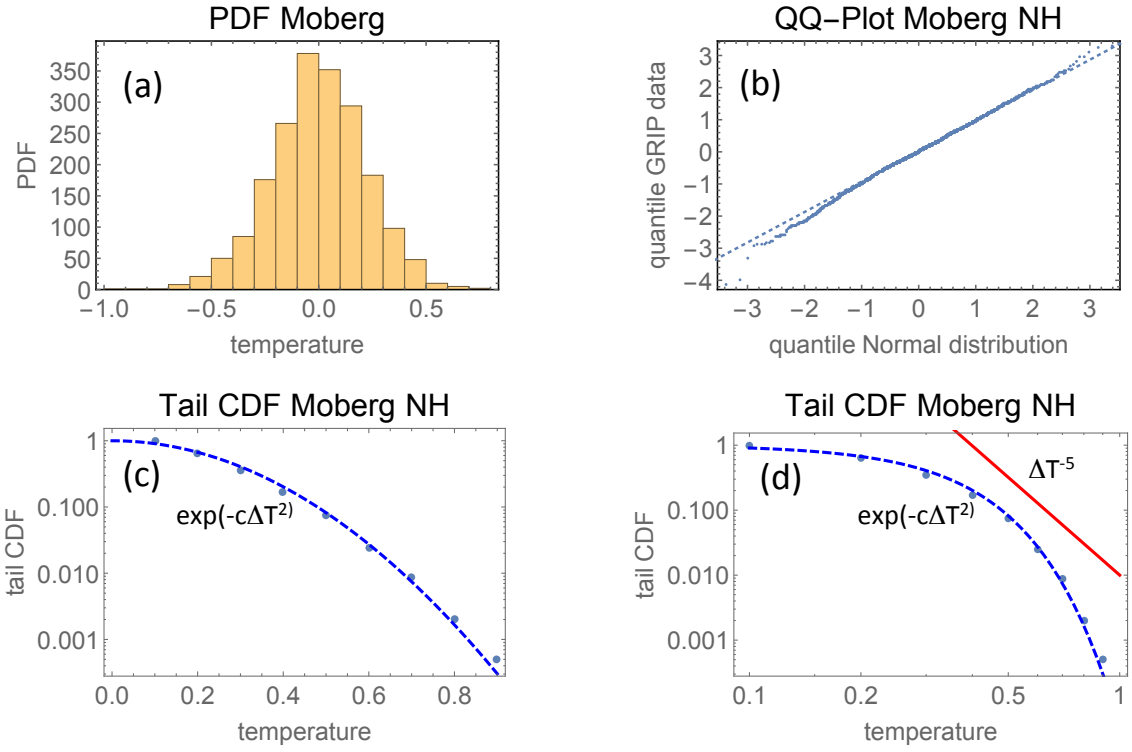

Figure 9. (a): PDF for the Moberg Northern Hemisphere multiproxy data. (b) QQ-plot for these data. (c): The tail CDF (bullets) and the fitted Gaussian (dashed) in a log-plot. (d): The same as in (c) in a log-log plot. The red line is a plot of the power-law function ΔT^{-5}.

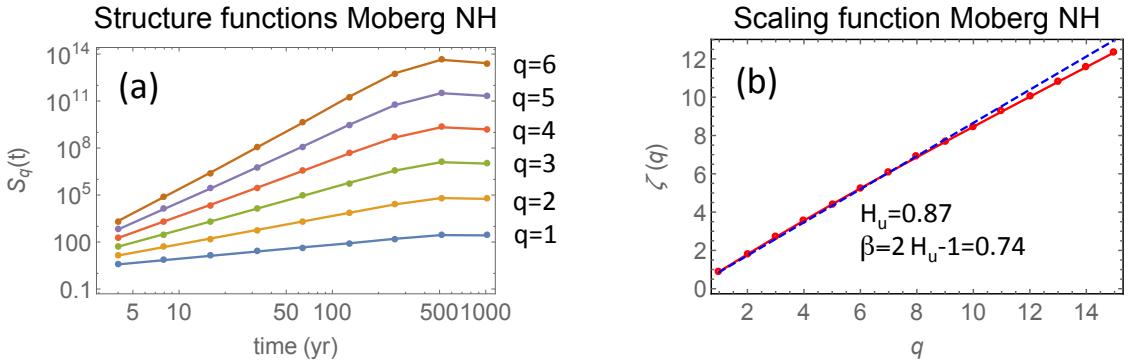

Figure 10. (a): Empirical moments for the Moberg Northern Henisphere data for moment orders $q = 1, \ldots, 6$. (b): The corresponding $\zeta(q)$ for $q = 1, \ldots, 15$ estimated as the slope of the empirical moment curves in the range 4-256 yr. The dashed line has slope $H_u = 0.87$, which corresponds to $\beta = 2H_u - 1 \approx 0.74$.

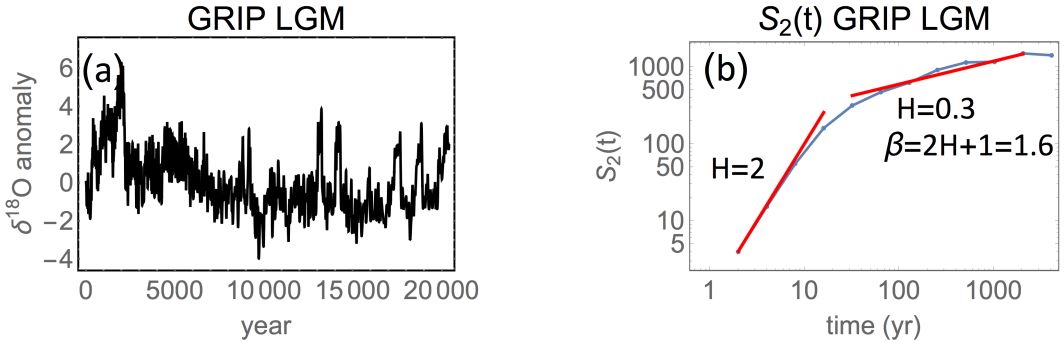

Figure 11. (a): δ^{18}O anomaly time series for 20 kyr of the Last Glacial Maximum (LGM) (b): The second order structure function $S_2(t)$ for this time series (not of its cumulative sum). The steep slope on short time scales up to ≈ 30 yr is due to the smoothness from interpolation, i.e., the actual time resolution is not better than a few decades. The curve is not completely straight in the range above 30 yr, but a slope computed on the scales $32 < t < 256$ yr is $H \approx 0.3$. The corresponding Hurst exponent is $H_u = H + 1 = 1.3$, and the spectral exponent is $\beta = 2H_u - 1 = 1.6$.

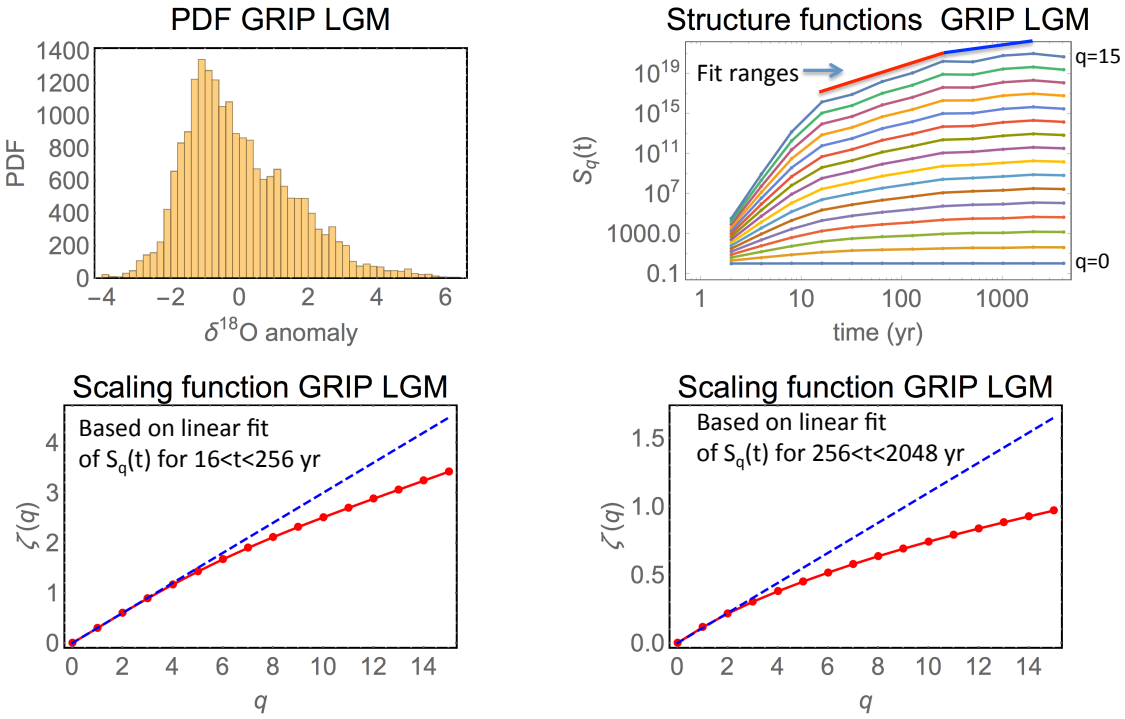

Figure 12. (a): PDF for the LGM data. (b) Empirical moments for the LGM data (not for the cumulative sum) for moment orders $q = 1, \ldots, 15$. (c): The corresponding scaling function $\zeta(q)$ for $q = 1, \ldots, 15$ estimated as the slope of the empirical moment curves in the range 16-256 yr. The dashed line has slope $H = 0.3$, which corresponds to $H_u = H + 1 = 1.3$ and $\beta = (1 + H_u)/2 \approx 1.15$. (d): The $\zeta(q)$ for $q = 1, \ldots, 15$ estimated as the slope of the empirical moment curves in the range 256-2048 yr. The dashed line has slope $H = 0.11$, which corresponds to $H_u = H + 1 = 1.11$ and $\beta = 2H_u - 1 \approx 1.22$.

Manuscript prepared for Earth Syst. Dynam. Discuss.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 28 September 2015

Response to reviewer 2

T. Nilsen[1], **K. Rypdal**[1], **and H.-B. Fredriksen**[1]

[1]Department of Mathematics and Statistics, UiT The Arctic University of Norway, Norway

Correspondence to: Tine Nilsen (tine.nilsen@uit.no)

1 General comments

Reviewer: The manuscript is well written, though with short passages that are technical.

Response: The most technical parts, about the Lomb-Scargle periodogram (LSP), will be moved to the Supplementary Material (SM). This is not only because they were technical, but because the revised manuscript does not rely extensively on the use of this spectral estimator. On the other hand, in the revision we will define some other estimators in the Methods section. This involves a few more equations, but they are simple and are included as an aid for readers who are not familiar with these techniques.

Reviewer: It is not overly exciting, as the relevance of the scaling models for future prediction is not made very clear.

Response: We encourage to read our comprehensive response to reviewer Lovejoy. Our paper deals with the most fundamental issues in statistical modelling of climatic records. Has scaling analysis any meaning without invoking statistical models, and how shall we select the model to use? The relevance for prediction is discussed there. In the revision we will discuss this issue more explicitly and in greater depth.

Reviewer: As a manuscript with paleoclimate focus this manuscript may be more suitable for a journal like "Climate of the Past".

Response: In the revision it will be even more clear that our purpose is to establish the correct statistical scaling model for Holocene climate as a tool for prediction of future climate on centennial time scales. This paper should be suitable for Earth System Dynamics.

Reviewer: At present several data (independence of the temperature reconstructions, calibration of the ice core data, effects of noise) and method aspects (bias in the estimation

procedure around $\beta = 1$, use of Lomb-Scargle periodograms) are not sufficiently discussed and justified in the manuscript to allow for firm conclusions.

Response: All these issues are discussed in our responses to specific comments below. Revisions will be made where they are appropriate. That said, we are a bit skeptical to the use of the phrase "firm conclusions." Many of the results we have obtained are of the negative type, i.e., we show that one cannot reject a null hypothesis, and hence renders the alternative hypothesis (e.g., the hypothesis of two scaling regimes) unsupported by the data. This means that we point out that previously published conclusions are subject to Type-I statistical errors (false positives = detecting an effect that is not present). The scientific literature is littered with Type-I errors, presumably because detecting new effects are perceived as much more exciting than proving that exciting findings may have less exciting explanations.

Reviewer: I suggest that the manuscript be subjected to minor revisions before it can be published.

Response: The paper will be subject to a major revision. The revision will be outlined at the end of this response.

2 Specific comments

Reviewer: Page 1202, line 15: Is this an important model for temperature prediction on centennial timescales?

Response: A recent paper proposes that the macroweather-climate regime model (SLIM) can be used as a basis for temperature prediction up to a few decades, (*Lovejoy et al.*, 2015). According to the multiple scale regimes hypothesis, prediction on centennial time scales the alternative SLIM-model should be based on a model with $\beta \sim 1.5$ or larger. The

properties of such a prediction model would be considerably different from one with e.g., $\beta \sim 0.7$. This issue will be discussed in greater depth in the revision.

Reviewer: Page 1205, line 10: It is not obvious why a Lomb-Scargle periodogram should deal better with irregularly sampled data than conventional techniques or other techniques such as spectral estimation via "correlation slotting" . There is evidence that it does not (Rehfeld Kurths, 2011). Uneven sampling causes bias when overcome by standard interpolation (leading to a red bias), but the LS periodogram also shows considerable high-frequency biases (blue bias) which may be worse. The argument that "the data is not standard, therefore the methods have to be special" is not necessarily true. A much more critical point, with a potentiallly larger impact on the scaling question, is the question of signal-to-noise ratios, and/ or the global representativeness of the used proxy records (see e.g. Laepple and Huybers, 2014).

Response: We agree with the reviewer that it is not obvious a priori that the Lomb-Scargle periodogram (LSP) performs better than other spectral techniques, so this will be discussed more in the revised version of our paper. We have looked into the papers by *Rehfeld et al.* (2011) and *Broersen et al.* (2000), and find that irregularly sampled data causes various problems for all spectral techniques. Slotting can be problematic because the covariance estimators may not be positive semi-definite, and could hence give negative values in the spectrum. The skill of the LS periodogram is as demonstrated by *Rehfeld et al.* (2011) dependent on the skewness of the distribution of sampling intervals. The bias will therefore differ from dataset to dataset, so we cannot determine just by reading these other papers which method is the most suitable for our purpose. In the originally submitted manuscript, the skill of the LSP for our data is illustrated in figure 1 of the main article. Surrogate data was generated with sampling times equal to the low resolution GRIP Holocene/last glacial period. The figure does indeed show a slight underestimation for all values of β, due to the blue bias.

We have recently become aware that one important aspect of the ice core data has been overlooked when we designed the surrogate data: namely the gradual, natural smoothing of the ice core proxy data as one goes back in time. Our surrogate data are sampled at irregular time steps, but as point measurements. In reality, the given values of stable isotopes from ice cores are generally averages over the time period covered by the sliced equal-length samples. The real proxy values will therefore exhibit less variability than if they were instantaneous measurements of the true temperature. This smoothing effect will increase as one goes back in time due to compression of the core. The smoothing effect constitutes a red bias that works in the opposite direction of the blue bias introduced by the LSP. Taking this into account we have created new surrogate data that resembles proxy data to a larger degree than before. The new surrogates are created by first generating regularly sampled time series, which are then averaged over time intervals equal to the GRIP time intervals to obtain irregularly sampled surrogate data. The results are shown in Figure 1 of this response, and demonstrates that the blue bias from the LSP is approximately balanced by the red bias from the smoothing.

Despite our findings of the generally good performance of the LSP for our purpose, a second look at the interpolation + standard periodogram approach shows that it is actually adequate for most of the data studied in our paper. Hence, we have decided to use it in the revised version of the paper, and put results from LSP, including figure 1, in the revised supplementary material. Hopefully this will also meet the complaints of both reviewers that the paper is "too technical."

Reviewer: Page 1206, line 22: This length limit assumes independence of the data points - which is not justified for proxy records.

Response: This "rule of the thumb" is not based on an assumption of independence of the data points, but on experience with application of various estimators on ensembles of surrogate fGns with relevant scaling exponents ($0 < \beta < 1$). For scales longer than $N/4$ the statistical spread of the fluctuation function or the power spectral density gets so large that

it contributes to unacceptable uncertainty in the estimate of the scaling exponent. The rule is illustrated by the increasing uncertainty spread towards lower frequencies in our plots of power spectra. Here it is also clear that the spread increases for lower β (weaker dependence between points).

Reviewer: Page 1210, line 12: Which figures in the supplement are referred to?

Response: It is referred to section S1 in the supplement, and fig S1 and S2. We will make this clear in the revised version.

Reviewer: Why is the case of $\beta = 1$ so awfully difficult? This is a significant drawback of the methodology!

Response: In figure 1 of our main article, a negative bias is seen for synthetic surrogate data with β slightly above unity. This bias is explained in more detail in section S1 (p1 l 5-10) but we realize that we do not explicit refer to this explanation in section 3.2 of the main article. It is not completely clear to us if the reviewer finds the explanation in the supplementary material unsatisfactoy, or if the main error is that we do not explicitly refer to the supplement for details about the bias. In the revised paper we will refer to the right section in the supplementary material for a detailed explanation of this bias. To make sure the bias for synthetic data with β=1 is properly explained, we comment on this below.

Synthetically generated discrete fGns and fBms exhibit scaling only asymptotically, i.e., for time scales much larger than the sampling time. At the shortest time scales the variability is higher than what we expect from a continuous process. This effect is larger for the case β just above 1 than for the other values of β in our study. It leads to a flattening of the spectrum as the frequency approaches the Nyquist frequency and hence reduces the estimated β if this frequency range is used in the fitting of a straight line to the spectrum in a log-og plot. For very long time series it is no problem to leave out this range in the fitting, and then there will be no bias. For short time series, including those with approximately 2000

data points used here, it is not possible to completely avoid this effect without increasing the uncertainty of the estimate (a smaller fitting range enhances uncertainty).

Reviewer: And: Is the LS-Periodogram actually doing a better job here than interpolation+(multitaper) spectral estimation?

Response: The main motivation for using the LSP instead of interpolation and a standard spectral estimator has been to avoid the red bias on high frequencies associated with interpolation. The LSP may be affected by the blue bias on the same frequencies. A red or blue bias in the power spectrum may incorrectly be interpreted as a scale break, and it may influence the frequency range we choose for estimation of a spectral exponent. The periodogram is our preferred estimator for the power spectral density, and we have therefore repeated our scaling analysis using interpolation plus periodogram for the high-resolution GRIP data set and the EPICA data set with lower temporal resolution. The other ice core data sets analysed in the supplementary material have similar temporal resolution as EPICA. The interpolation + periodogram approach works very well for both data sets in the frequency range up to the frequency corresponding to inverse of the mean sampling interval, i.e., the interpolation has very little effect for this frequency range. The spectra obtained are very similar to the LSP's, and the estimates of the scaling exponent β are similar. Since the two methods both perform well for our purpose, we choose to employ the more standard technique in the main article.

Reviewer: P1211, Line 15: Why are the wavelet analyses necessary? What are they supposed to contribute to the main message of the manuscript?

Response: This is discussed at length in our response to reviewer Lovejoy.

Reviewer: P1211, line 24: Are these reconstructions actually independent, or do they use the same background proxy series for reconstruction?

Response: First, let us acknowledge a small error that was discovered after the mansucript was published. On page 1212 line 10 it is referred to the Mann et al. (2009) reconstruction, but this is not correct. The reconstruction that was used is *Mann et al.* (2008), the error will be corrected both in the main text and in the bibliography.

Regarding the independence of the proxy records, all papers except one provide the detailed information necessary to track the origin of the records. *Briffa et al.* (2001) use tree-ring records but does not state explicitly the source. For the remaining 6 proxy/multiproxy reconstructions, there are data sets that are used in more than one reconstruction, including tree-ring records and low-resolution proxy records. Independence is therefore not guaranteed, but the reconstruction methods are different. The records are therefore processed in different ways, which may influence the scaling of the resulting reconstructions.

Reviewer: Page1212, line 23: If these are temperature reconstructions, what were the calibration models? Linear regression based on O18? Or nonlinear models? Were they adjusted for Holocene and Glacial differences? Same for l. 4 on p1213 (deuterium)

Response: This is an error from our side. The records that are used are not converted to temperature, they are δ^{18}O records (GRIP) and the EPICA record is δD. What was meant in the present text was that the δ^{18}O/ δD proxies are related to temperature. This will be corrected in the revised paper.

Reviewer: Page 1216, line 10: Why should DO-events flatten the spectrum on centennial and shorter timescales? This is not obvious (to me).

Response: This was an inaccurate statement from our side, and we will remove this claim from the revised paper. The DO events affect the entire spectrum spectrum. Preliminary

spectral analysis of the stadial and interstadial stages separately suggest scaling with $\beta \sim 1$ in both stages, so the discontinuities associated with the transitions between the two stages and their distribution in time, is what alters the spectrum an gives rise to what appears as a scale break. It should be mentioned that we have checked that this is not an artifact due to some blue bias of the LSP, since we observe the same flattening for interpolation + periodogram. Further detailed analysis of the scaling structure of the glacial state (which is very complex and very interesting) is beyond the scope of the present paper.

Reviewer: Page 217, lines 24/25: Basically, what you write here is that there is no power in this test of beta differences. So why use it?

Response: This simple test on synthetic LRM processes demonstrates that it is relatively easy to detect a scale-break in a time series. On the other hand, without testing the statistical significance of the break, the data analysis is not enough to claim that the scale-break is a real phenomenon. The fGn is defined from only one scaling exponent, but spurious breaks may occur at random without being significant.

Reviewer: Page 1218. These paragraphs are nice to read.

Response: Thank you! They will be expanded.

Reviewer: Page 1219, line 28: This sentence, and the following one, is hard to digest, in particular to the nice and easy to read discussion before.

Response: In the revision, we shall expand the discussion section and in particular the part dealing with intermittency and multifractality.

3 Outline of revision

In the abstract we add the modified sentences:

- A model for internal variability with only one regime is simpler, and allows more certain predictions on time scales of centuries when combined with existing knowledge of radiative forcing.
- Spectra from a number of late Holocene multiproxy temperature reconstructions, and one from the entire Holocene, have also been analysed, without identifying a significant scale break.

In section 1 (Introduction) we make the definitions of "scaling regime" and "scale break" more precise, and we give a very brief introduction to the concept of fluctuation analysis and the Haar fluctuation function.

In section 2 (The concept of multiple scaling regimes) we only make a few minor updates.

In section 3 (Methods) we move the description of the Lomb-Scargle periodogram to the Supplementary Material, and we add one subsection on structure functions and the scaling function and one on the Haar fluctuation function. Both are accompanied with a figure where the methods are applied to the instrumental global temperature record and used to contrast the differences between modelling the record as a process consisting of two scaling regimes or modelling it as a fractional noise plus a trend.

In Section 4 (Data) we add a subsection on the Marcott multiproxy series.

In section 5 (Results) we expand the discussion of the Moberg multiproxy record and add two figures. One employs structure function and scaling function analysis to show the non-intermittent nature of this record, and also to demonstrate the effect of introducing a model that involves an oscillatory trend. In another figure, the contrast between such a model and

a two-scaling regime model is shown by using the Haar fluctuation analysis on ensembles of realisations of these two models.

A new subsection is added where the Marcott reconstruction is analysed, correcting for the gradual smoothing (reduction of the high-frequency variability) of the proxy signal as one goes backward in time towards the early Holocene (see the discussion in our response to Lovejoy).

In section 6 (Discussion and conclusion) we will include of some of the material in the response to Lovejoy. In particular we will discuss in more depth the ambiguities arising when performing scaling analysis on data from evolving or otherwise non-stationary systems. Here we will add a wavelet scalogram of GRIP data illustrating the difference between Holocene and glacial climate, and also the change in variability between stadials and interstadials during the ice age. This scalogram also show that the longest interstadials exhibits a variability very similar to the Holocene.

In the Supplementary Material some of the material and figures shown in our response to Lovejoy will be included.

Acknowledgements. This work was funded by project no. 229754 under the the Norwegian Research Council KLIMAFORSK programme.

References

K. R. Briffa and T. J. Osborn and F. H. Schweingruber and I. C. Harris and P. D. Jones and S. G. Shiyatov and E. A. Vaganov: Low-frequency temperature variations from a northern tree ring density network, J. Geophys. Res.-Atmos, 106, D3, doi:10.1029/2000JD900617, 2001.

Broersen, P.M.T. and S. de Waele and R. Bos: The accuracy of time series analysis for laser-doppler velocimetry, 10th International Symposium on Application of Laser Techniques to Fluid Mechanics, Lisbon, July 10-13, 2000

Lovejoy, S. and del Rio Amador, L. and Hébert, R.: The Scaling LInear Macroweather model (SLIM): using scaling to forecast global scale macroweather from months to decades, Earth System Dynamics Discussions, 6, 1, doi: 10.5194/esdd-6-489-2015, 2015

Mann, Michael E. and Z. Zhang and M. K. Hughes and R. S. Bradley and S. K. Miller and S. Rutherford and F. Ni: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, PNAS, 106, 36, doi: 10.1073/pnas.0805721105, 2008.

Marcott, S. A., Shakun, J. D., Clark, P. U., and Mix, A. C.: A reconstruction of the regional and global temperature for the past 11.300 years, Science, 339, 1198, doi: 10.1126/science.1228026 (2013).

Rehfeld, K. and N. Marwan and J. Heitzig and J. Kurths: Comparison of correlation analysis techniques for irregularly sampled time series, Nonlinear Processes in Geophysics, 18, 389-404, doi:10.5194/npg-18-389-2011, 2011

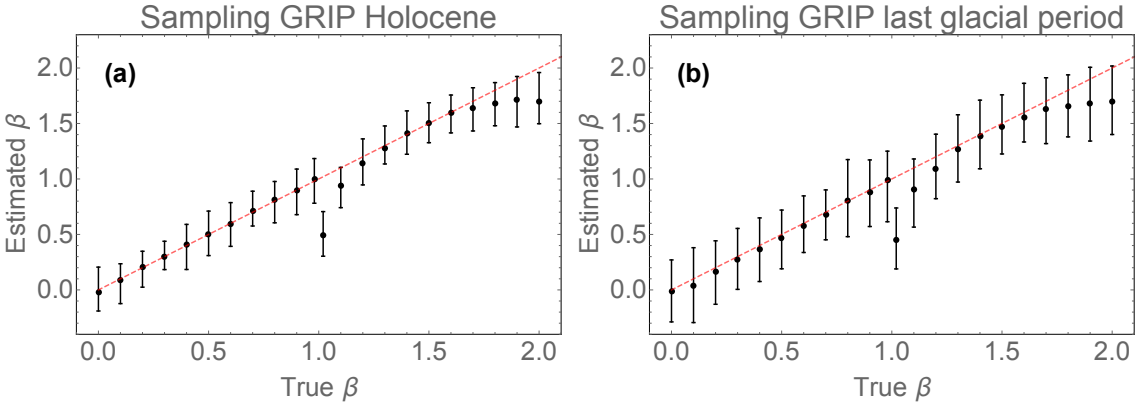

Figure 1. This figure is similar to figure 1 of our paper, but with other surrogate data. These new surrogate data are created by averaging regularly sampled data over time intervals equal to the GRIP sampling intervals, rather than interpolating to compute point measurements at the GRIP sampling times. The red bias introduced by the averaging is approximately balancing the blue bias shown in figure 1 of our paper.

List of relevant changes in the revised manuscript

Due to the untraditional form of review from Shaun Lovejoy, we did not write a point-by point response, and the manuscript has undergone a major revision. We have compiled a list below of important changes based on both reviews. All sections have been changed in some way, so we recommend reading the markup file for the revised paper to get a better overview of the changes. Changes are also marked in red in the revised supplementary material.

1) The terms "monoscaling" and "multiscaling" has been exchanged with "single scaling regime " and multiple scaling regimes" throughout the revised manuscript.
2) Since both reviewers questioned the relevance of the wavelet power spectra, we have included a section explaining the relevance of each method applied. See section 3 in the revised manuscript.
3) The term "wavelet power spectra" has been replaced by "wavelet scalogram" throughout the paper, since this term is more commonly used for the method.
4) The wavelet scalograms that were included in the supplementary material have been removed. A new wavelet scalogram has been included in the main paper, for the last 85 kyr of the GRIP $\delta^{18}O$-series. See section 6 and figure 15.
5) We have included structure/scaling function analysis to highlight the general importance of higher-order statistics and to justify a monoscaling model (fractional Gaussian noise) for the internal temperature variability in the Holocene. We refer to sections 1, 3.3, 5.1 and 6 in the revised manuscript.
6) The Haar fluctuation function is discussed because it is strongly advocated as the ultimate scaling analysis technique by e.g., Lovejoy and Schertzer. (2012). See sections 3.4 and 5.1 in the revised manuscript.
7) For the unevenly sampled ice core data, we have decided to use linear interpolation and standard spectral analysis (periodogram) instead of the Lomb-Scargle periodogram in the main paper. We have included a section where different spectral methods are discussed. Testing of the Lomb-Scargle periodogram and scaling analysis of ice core data using this method has been moved to the supplementary material. See sections 3.1 and 5.3 in the revised manuscript, and S1 (in revised supplementary material).
8) The temperature reconstruction by Marcott et al. (2013) is discussed, and scaling analysis is included. See section 4.2 and 5.2 in the revised manuscript.
9) The issue of scaling in nonstationary time series is discussed in more detail in section 6 of the revised manuscript.

- S. Lovejoy and D. Schertzer. (2012): Haar wavelets, fluctuations and structure functions: convenient choices for geophysics, *Nonlinear Proc. Geoph*.,**19**, 5, DOI =10.5194/npg-19-513-2012.
- S. A. Marcott, J. D. Shakun, P. U. Clark and A. C. Mix (2013): A Reconstruction of Regional and Global Temperature for the Past 11,300 Years, *Science*, **339**, 6124, doi = 10.1126/science.1228026.

Manuscript prepared for Earth Syst. Dynam.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 27 October 2015

Are there multiple scaling regimes in Holocene temperature records?

Tine Nilsen[1], Kristoffer Rypdal[1], and Hege-Beate Fredriksen[1]

[1]Department of Mathematics and Statistics, University of Tromsø The Arctic University of Norway, Tromsø, Norway

Correspondence to: Tine Nilsen (tine.nilsen@uit.no)

Abstract. The concept of multiple scaling regimes in temperature time series is examined, with emphasis on the question whether or not a mono-scaling model can be rejected from observation data from the Holocene. A model for internal variability with only one regime is simpler, and allows more certain predictions on time scales of centuries when combined with existing knowledge of radiative forcing. Our analysis of spectra from stable isotope ratios from Greenland and Antarctica ice cores shows that a scale break around centennial time scales is evident for the last glacial period, but not for the Holocene. Spectra from a number of late Holocene multiproxy temperature reconstructions, and one from the entire Holocene, have also been analysed, without identifying a significant scale break. Our results indicate that a mono-scaling model cannot be rejected as a null model for the Holocene climate up to at least millennial time scales, although it can be rejected for the glacial climate state. The scale break observed from the glacial time ice core records is likely caused by the influence of Dansgaard-Oeschger events and teleconnections to the Southern hemisphere on centennial time scales. From our analysis we conclude that the two-regime model is not sufficiently justified for the Holocene to be used for temperature prediction on centennial time scales.

1 Introduction

The main focus of this paper is the scaling properties in paleotemperature records at centennial and millenial time scales. In particular we study the differences in variability between glacial/interglacial time periods, and we discuss the justification of separating temperature variability on different time scales into distinct scaling regimes. The notion of "scaling" in climatic time series is based on the observation that the natural variability of the Earth's surface temperature can be modelled as a persistent stochastic process, with superposed trends and quasi-periodic modes representing variability which is not included in the noise background. There is a considerable body of literature suggesting that long-range memory (LRM) stochastic processes are good statistical models for de-seasonalised local and global temperature records on time scales from months up to a century or more (Koscielny-Bunde et al., 1996; Rybski et al., 2006; Efstathiou et al., 2011; Rypdal et al., 2013; Østvand et al., 2014). The standard continuous-time stochastic LRM processes are the fractional Gaussian noise

(fGn) and fractional Brownian motion (fBm). The latter is the cumulative integral of the former, and both are said to be scale-invariant (or scaling). The strength of persistence, or memory, in an LRM stochastic process is described by the spectral exponent β; the power spectral density (PSD) takes a power-law form $S(f) \sim f^{-\beta}$. The fGn has $-1 < \beta < 1$ and stationary variance, while the fBm has $1 < \beta < 3$ and a non-stationary variance that grows in time like $\sigma(t) \sim t^{\beta-1}$. The fGn is persistent (exhibits long-range memory) if $\beta > 0$, and is anti-persistent if $\beta < 0$.

Because the terms "scaling regime" and "scale-break/deviation from scaling" might be ambigous, the terms are briefly explained in the following. According to the glossary of Kantelhardt (2011) a scaling regime can be identified only if a power law is valid for scales spanning at least one order of magnitude, be it frequency or time scale. "Deviation from scaling" is synonymous with violation of Kantelhardt's definition. The term "break in scaling" is used to separate scaling regimes that exists in a single time series, where each regime complies with Kantelhardt's definition and is valid for at least one order of magnitude.

Ditlevsen et al. (1996) analysed the scaling in high-resolution ice core data from Greenland. Two different overlapping time series were used to create a composite power spectrum, and from this a break in scaling was identified around centennial time scales. On time scales shorter than centennial the spectrum was flat ($|\beta| \ll 1$), while on longer time scales a non-stationary regime with $\beta \approx 1.6$ was found. One of the time series covers 0-91 kyr BP, and the other 0-3 kyr BP. This procedure of combining different time series into one power spectrum is problematic since the two time series reflect different climate states with different variability. The longer time series is dominated by the glacial state, while the short one contains only Holocene data. The different variability of the two states is seen clearly by direct inspection of the data, e.g., from comparing the Holocene part of the GRIP ice core (Fig. 9a) and the last glacial period from the same ice core (Fig. 11a). The standard deviation in the Holocene time series is less than half of the glacial one, and the latter looks more bursty. The records in Fig. 9a and Fig. 11a can be associated with different stochastic processes. The Holocene record is similar to an fGn with low persistence, while the records from the last glacial period exhibits strong intermittency and is associated with a high spectral exponent, $\beta > 1$.

A simple measure of scaling is the fluctuation analysis (FA) (for a brief review see Rypdal et al. (2013)). It defines the fluctuation function $F(\Delta t)$ as the standard deviation of the data record after it has been filtered by a simple moving average with window width Δt, and hence measures the fluctuation magnitude as a function of scale Δt. The fluctuations are scaling if $F(\Delta t) \sim \Delta t^h$, where h is the scaling exponent (Lovejoy and co-workers denote this exponent by H, and call it the *Hurst exponent*. This convention deviates from the mainstream literature on fractal processes, where the Hurst exponent is $H = h + 1$). The PSD is then of the power-law form with $\beta = 2h + 1$. An issue with FA is which mean value to relate the fluctuation deviation to; the local mean in the window, or the mean of the entire data record. The latter is problematic if the fluctuations are monotonically

growing with increasing scale. This problem can be circumvented by convolving the data record with the simple antisymmetric Haar-wavelet (Lovejoy and Schertzer., 2012) rather than performing the moving average. For fluctuations growing with scale it measures fluctuation differences versus scale, whereas for fluctuations decreasing with scale it measures fluctuation relative to the local mean. One feature that the PSD, FA and Haar fluctuation share with many other measures of scaling is that it is sensitive to trends and large-scale oscillations, i.e., it is often not able to discriminate between such variability and true scaling behaviour.

Pelletier (1998) estimated the power spectra and scaling exponents from a Deuterium record from the Vostok ice core as well as from instrumental local data, and also created composite spectra from the records. Huybers and Curry (2006) and Lovejoy and Schertzer (2012) have studied the scaling in multiple proxy data sets covering time scales from years to millions of years. Both report a break in scaling from fluctuations decreasing with scale ($h < 0,\ \beta < 1$) to fluctuations increasing with scale ($h > 0,\ \beta > 1$) on a transition time scale $\tau_c \sim 10^2$ yr. The break in scaling is seen from composite spectra of paleotemperature records based on different proxies and reconstruction techniques, where many of the records span hundreds of kyr. Since glaciation is the dominating climate state in the Quaternary, the spectra obtained in those papers are typical for glacial climate. Huybers and Curry (2006) suggest that the power-law continuum in the spectrum of surface temperature on time scales between one year and a century is a result of an inverse cascade in frequency space driven by the seasonal cycle forcing. The non-stationary scaling regime from century time scale and longer is proposed to be the result of a nonlinear response to the Milankovitch cycle forcing. From the composite spectra, they infer scaling exponents in the range $\beta = 0.37 - 0.56$ for time scales $\tau < 10^2$ yr, and $\beta = 1.29 - 1.64$ at longer time scales. Lovejoy and Schertzer (2012) introduce three different scaling regimes: the "weather" regime ($\beta \approx 2$ for time scales up to 10 days), the "macroweather" regime ($\beta \approx 0.2$ for time scales from 10 days to 10^2 yr), and the "climate" regime ($\beta \approx 1.4$ for time scales from 10^2 yr and longer). Common for the studies mentioned is that they don't make a distinction between glacial and Holocene spectra.

Other results reported in the literature support our hypothesis of different scaling in glacial and interglacial climate, with the scale break at centennial time scales absent for the Holocene. Blender et al. (2006) analysed the scaling properties of a 10 kyr long general circulation model (GCM) climate simulation, and no scale break can be detected at centennial time scales. Lovejoy et al. (2013) make similar observations and conclude that GCM's do not predict climate, only macroweather. Roe and Steig (2004) found by using a short-range memory autoregressive (AR) model that the characteristic time scales for paleotemperature ice core records were significantly shorter during the Holocene than during the last glacial period. This study is important for our reasoning, but the idea needs to be adapted to a long-memory model. We will separate the ice core records into glacial and interglacial time series and demonstrate the fundamentally different scaling properties of these climate states, and we will analyse other temperature reconstructions for the Holocene in search for a detectable

scale break.

The paper is organized as follows: in Sect. 2 we address the issues of uncertainties and limitations of proxy-based reconstructions, and the implications for the existence of separate scaling regimes are discussed. Sect. 3 describes the scaling analysis methods employed, and information about the model and the data used can be found in Sect. 4. The results from the analysis are presented in Sect. 5, and discussion and conclusion follow in Sect. 6.

2 The concept of multiple scaling regimes in the Holocene

Lovejoy and Schertzer (2012) identify two scaling regimes in a number of Holocene temperature records. In instrumental data the transition time τ_c is found to be 10-30 yr, in proxy/multiproxy reconstructions it is 40-100 yr, while for one of the ice core paleotemperature records it is approximately 2000 yr. Hence, it seems difficult to identify a universal τ_c from the data examined in that paper. For the proxy/multiproxy reconstructions that were analysed in Lovejoy and Schertzer (2012), the time period 1500-1979 was selected because it was common to all reconstructions, and the medieval warm period was avoided. However, by starting in the Little Ice Age the series are strongly influenced by steadily increasing solar, as well as anthropogenic, forcing. A pronounced linear trend has strong effect on the estimate of the scaling exponent from power spectra unless the time series is linearly detrended. This is also the case for the Haar fluctuation analysis, which was also applied in Lovejoy and Schertzer (2012).

Many of the papers cited in Sect. 1 present composite spectra based on instrumental and/or proxy data for scaling analysis. There are many problems related to this, in addition to the already mentioned aspect of combining time series from the Holocene/glacial climate state. The various data sets are representative for different degrees of spatial averaging. This will affect the shape of the spectra and the estimates of the scaling exponent, because the high-frequency variability is reduced with increasing degree of spatial averaging. For the instrumental data we can obtain global averages, while proxy/multiproxy time series represent local, regional or at best hemispheric temperature. There is an important difference between composite spectra and spectra from multiproxy reconstructions. Multiproxy reconstruction methods generally take geographical weighting into account, and the aim is to obtain realistic high/low frequency variability throughout the time period covered by the reconstruction. Most composite spectra, on the other hand, do not handle these aspects in a satisfactory manner. An example of a well designed composite spectrum can be found in Laepple and Huybers (2014), where the proxy records have been corrected for noise, and the instrumental data used were extracted from the same location where the proxy records were sampled. No attempt was made to estimate a scaling exponent from this spectrum, but it is clear that no scale-break can be observed around centennial time scales. Correcting for proxy noise is outside the scope of the present study,

and we will therefore avoid studying composite spectra based on different reconstructions spanning different time scales and climatic states, and on different proxies and reconstruction techniques.

The techniques used to estimate the scaling exponent have inherently higher uncertainties on the longest time scales, due to sparse data on these time scales. A rule of thumb is that the scaling properties for a time series of length N should be estimated only up to time scales $N/4$, since the uncertainty on time scales longer (frequencies lower) than this is too large to make meaningful estimates. Suppose, for instance, we want to establish that we have scaling in annual data on scales up to 100 yr. Then we need a series which is 400 yr long. If we want to establish the existence of a different scaling regime on time scales longer than 100 yr for a time series, we need to know with reasonable certainty the spectral estimates up to one millenium. As we will demonstrate in Sect. 5, this implies that we need record lengths spanning several millennia to bring the uncertainty of β below the limit needed to reject the mono-scaling hypothesis. More generally, we need a record length of the order of 10^2 times the smallest time scale of a scaling regime.

Instrumental temperature data are not included in our analysis of multiple scaling regimes because previous studies do not show pronounced breaks in the scaling after detrending to account for influences from anthropogenic warming (Rypdal et al., 2013). The series are too short to detect scale breaks at centennial time scales. However, we use instrumental data in Sect. 3 for illustration of interesting features of various techniques for scaling characterisation. Detection of scaling properties in regional or hemispheric proxy/multiproxy temperature time series is possible but not optimal, since these records generally cover the past 2000 years or less, and, even though they are usually given with annual resolution, some are effectively filtered to vary smoothly on annual time scale, and have effective resolution from 5 to 10 yr. For some of the available proxy/multiproxy reconstructions an enhanced power can be inferred relative to a mono-scaling spectrum on time scales longer than a century, but data are too sparse to show that this enhanced power represents a new scaling regime.

The deep ice cores sampled at Greenland and in Antartica provide the most suitable data sets for studying possible scale breaks in both the Holocene and the last glacial period, due to the high temporal resolution and long duration. It was argued by Lovejoy and Schertzer (2012) that the Holocene part of the Greenland ice core GRIP $\delta^{18}O$ is exceptionally stable compared to other proxy-based reconstructions of North-Atlantic temperature, but we observe similar stability also for the stable isotope records from Greenland ice cores GISP2 and NGRIP, and the Antarctic ice cores EPICA, Vostok and Taylor dome. When comparing the Holocene stable isotope records from deep ice core records with temperature reconstructions from other regions and based on other proxies, one needs to keep in mind that local variability and the sensitivity of the various proxies to changes in their environment will affect the resulting proxy record. A proxy record is based on an imperfect recorder of a climatic variable of interest, and some sources of noise are not related to climatic processes. Unfortunately, there is no such thing as a temperature reconstruction based on a single proxy that is representative for global temperature variability. In this manner, all temperature reconstructions

are exceptional, because the proxies record some local variability and are subject to noise from a number of sources.

3 Methods

In our data analysis we have used a number of tools, which are described in detail in the following subsections. In general, the periodogram estimator for the power spectral density (PSD) is applied for the scaling analysis of paleotemperature time series in our study, and are the basis for significance testing when studying potential scale breaks in the time series. We have chosen the PSD because most readers are familiar with it, and because it is fully adequate for the purpose. Wavelet scalograms are not used to estimate scaling properties, but to visualize particular features in some of the time series. Structure functions are discussed to point out the general importance of higher-order statistics and to justify a monoscaling model (fractional Gaussian noise) for the internal temperature variability in the Holocene. The Haar fluctuation function is discussed because it is strongly advocated as the ultimate scaling analysis technique by e.g., Lovejoy and Schertzer. (2012).

3.1 Estimation of Power Spectral Density

The periodogram is applied as an estimator for power spectral density (PSD) for evenly sampled time series of length N. The stable istope records from Greenland and Antarctic ice cores have been linearly interpolated to obtain even sampling in time. All other records were already provided with even sampling. The periodogram is defined here in terms of the discrete Fourier transform H_m as $S(f_m) = (2/N)|H_m|^2$, $m = 1, 2, \ldots, N/2$. The sampling time is the time unit, and the frequency is measured in cycles per time unit: $f_m = m/N$. $\Delta f = 1/N$ is the frequency resolution and the smallest frequency which can be represented in the spectrum, while $f_{N/2} = 1/2$ is the Nyquist frequency (the highest frequency that can be resolved). The periodogram has a poor signal to noise ratio, but since we are interested in studying the overall shape (scaling) of the spectrum, and not the power at specific spectral peaks this is not a problem here. By presenting the periodogram in a log-log plot, the scaling exponent β can be estimated by a linear fit to the power spectrum; $\log S(f) = -\beta \log f + c$. In the present study the periodogram is log-binned before fitting to ensure that all time scales are weighted equally (Østvand et al., 2014).

We have also considered other spectral estimators for the unevenly sampled stable isotope data from ice cores, such as the Lomb-Scargle periodogram (LSP), (Lomb, 1976; Scargle, 1982), or correlation slotting, Rehfeld et al. (2011). The main motivation for looking into different spectral methods is to compare biases in the spectra that could be wrongfully interpreted as breaks in the scaling. The papers by Rehfeld et al. (2011) and Broersen et al. (2000) demonstrate that irregularly sampled data cause various problems for all spectral techniques. Slotting can be problematic because the covariance estimators may not be positive semi-definite, and could hence give negative values in

the spectrum. Interpolation leads to underestimation of the spectral power at high frequencies, while the Lomb-Scargle periodogram suffers from the opposite bias: overestimation of the spectral power at high frequencies. The skill of the LS periodogram is, as demonstrated by Rehfeld et al. (2011), dependent on the skewness of the distribution of sampling intervals. The bias will therefore differ from dataset to dataset. We have tested the performance of the method on surrogate data mimicking the ice core proxy data under study. The detailed results for this test are shown in supplementary material. In general the method performs well, but not much better than interpolation plus standard periodogram. Results presented in the main paper are obtained using only interpolation and the standard periodogram. Scaling analysis of the ice core data based on the LSP is included in the supplementary material section 1.

3.2 Wavelet scalogram

The continuous wavelet transform is the convolution between a time series $x(t)$ and the rescaled mother wavelet $\Psi(t)$;

$$W(t,\tau;x(t),\Psi(t)) = \int_{-\infty}^{\infty} x(t')\frac{1}{\sqrt{\tau}}\Psi^*\left(\frac{t'-t}{\tau}\right)dt', \tag{1}$$

where the asterisk indicates complex conjungate. The wavelet scalogram (WS) is defined as $|W(t),\tau)|^2$, and is plotted versus time and time-scale. The WS is used here as a supplementary tool to the Fourier spectra. Time segments before and after the time interval where we have data were padded with zeros, as described in Torrence and Compo (1998). The region in (t,τ)-space affected by edge effects is the region above the white line in the upper part of the WS-plot shown in, e.g., Fig. 10. Due to the uneven sampling of the data in this study, linear interpolation has been performed prior to computing the WS. At each time t there is a characteristic sampling period in the orginal time series, and hence a Nyquist period. This Nyquist period is marked as the lower white curve in the WS plots. The WS below that curve does not reflect observed variability.

We have chosen two wavelet functions as the basis for our study: the Morlet wavelet which is complex valued, and the Mexican hat wavelet (second derivative of a Gaussian) which is real valued. The wavelet scalograms from these two wavelet functions provide different information. The Mexican Hat wavelet function resolves the timing of spectral peaks precisely, while the scale resolution is poor. For the Morlet wavelet function the opposite is true.

3.3 Structure functions and scaling function

A plethora of estimators have been developed for computing characteristic exponents for monoscaling long-range memory (LRM) processes. They all have strengths and weaknesses, but have in common that they give nonsense if the signal is not a monofractal process. Common for many papers by "LRM-skeptics" is the uncritical use of cookbook recipes for such estimators to data that are not

fractal, for instance climatic time series dominated by a specific trend (e.g., Mann (2011)). Hence, more important than estimating a characteristic exponent is to examine the general scaling characteristics of the data. A classical and useful method is to examine the probability density functions (PDFs) of the time series as it varies on different time scales. Rather than computing and plotting the PDFs, it is more common to compute the statistical moments of order q of the distribution, and then plot these moments as a function of time scale Δt. Given a stochastic process $x(t)$, the moments

$$S_q(\Delta t) \equiv E[|x(t+\Delta t) - x(t)|^q], \tag{2}$$

are called the structure functions (SFs) of the process. If the process is sampled at discrete times $t = 1, \ldots, N$ the empirical moments $\hat{S}_q(\Delta t) = (N-\Delta t)^{-1} \sum_{i=1}^{N-\Delta t} |x(i+\Delta t) - x(i)|^q$ constitute estimates of the structure functions. For large Δt (when Δt is no longer a small fraction of N) the number of independent terms becomes small, and the statistical uncertainty of the estimate becomes large. A useful rule is that this limits the scales we can investigate to $\Delta t < N/4$.

Let us assume that $x(t)$ is a self-similar, Gaussian, non-stationary stochastic process, i.e., a fractional Brownian motion (fBm). Then the fluctuation function $F(\Delta t)$ is monotonically increasing ($h > 0$, $\beta = 2h + 1 > 1$), and the structure functions defined on $x(t)$ take the form,

$$S_q(\Delta t) \equiv E[|x(t+\Delta t) - x(t)|^q] = \Delta t^{\zeta(q)}, \tag{3}$$

where $\zeta(q) = hq$ is the *scaling function* of the self-similar (monofractal) process. By taking the logarithm of Eq. (3) we find the linear relationship between $\log S_q(\Delta t)$ and $\log \Delta t$, where $\zeta(q)$ is the constant of proportionality, hence the SFs appear as straight lines in log-log plots with slope $\zeta(q)$. The scaling $\zeta(q)$ is a linear function of q only if the process is monofractal. If the SFs have the form Eq. (3) (i.e., if the SFs are straight lines in log-log plot) so that $\zeta(q)$ is defined, but $\zeta(q)$ is not a linear function, then the process is multifractal. If the SFs are not straight, the process is neither multi- nor monofractal, but it may still have a bursty or *intermittent* appearance.

If the process is stationary, with decreasing fluctuation function ($h < 0$, $\beta < 1$), it is denoted a fractional Gaussian noise (fGn), and the SFs are constant (flat) and contain no other information than the stationarity. However, as we shall see below, the SFs may still contain some useful information about the deviation from scaling if there are trends or oscillations in the data. In order to expose the scaling properties (if any) of the process, the trick is to form the cumulative sum $y(i) = \sum_{j=1}^{i} x(j)$ and then compute the SFs from this sum. The resulting SFs take the form

$$S_q(\Delta t) \equiv E[|y(t+\Delta t) - y(t)|^q] = \Delta t^{\zeta(q)} \tag{4}$$

where $\zeta(q) = H_u q$ is the scaling function, and H_u the scaling exponent, of the cumulative sum $y(t)$. H_u is also called the Hurst exponent of the stationary proxess $x(t)$, and is related to the spectral exponent through $H_u = (\beta+1)/2$. Since we have that $h = (\beta-1)/2$ for a nonstationary process we often see the relation $H_u = h+1$, although for SFs the two scaling exponents are defined for different

classes of processes. However, there are other estimators, like the periodogram and the wavelet variance (including the Haar wavelet fluctuation employed by Lovejoy and Schertzer (2012)), which work on both stationary and nonstationary processes. For these estimators this relation is meaningful.

The usefulness of the structure-function approach is illustrated in Figure 1. Panel (a) shows the eight structure functions corresponding to $q = 1, 2, \ldots, 8$ for the monthly global mean surface temperature (GMST) for the period 1880 – 2010 derived from the HadCRUT3 data set.

The underlying scaling of the noise is exposed by computing the SFs for the cumulative sum, as shown in Figure 1b. However, the corresponding scaling function, shown by the upper line in Figure 1d, has the slope $H_u \approx 1$, which is always the case for a signal dominated by a strong trend. The true scaling of the noise appears after a second-order polynomial fit to the record has been subtracted. The SFs for the cumulative sum of the detrended record is shown in Figure 1c, and the corresponding scaling function by the lower line in Figure 1d. This line has Hurst exponent $H_u \approx 0.85$. The straight appearance of the scaling function tells us that the GMST is monofractal, and simple tests on the PDFs at different scales show that it is Gaussian (Rypdal and Rypdal, 2010).

The scaling functions in Figure 1d have been computed from the slopes of the SFs in the regime of scales $\Delta t < 10$ yr where the SF-curves are straight. The bending of these curves for large scales are due to oscillatory modes on periods around 20 yr and 70 yr. Estimators employing only second-order statistics, like the periodogram or the Haar fluctuation employed by Lovejoy and Schertzer (2012) are not able to distinguish between scale-invariant fluctuations and variability due to trends or oscillations. Careful application of higher-order statistics like higher-order SFs has this ability. This can give us the possibility to separate distinct non-scaling dynamical features from the scaling, persistent noise background. In section 5.1 we shall demonstrate this usefulness on a multiproxy temperature reconstruction spanning two millennia.

3.4 The Haar fluctuation function

The Haar fluctuation was briefly mentioned in Sect. 1. The simple definition given in Lovejoy et al. (2013) starts by defining the fluctuation on scale Δt as,

$$\Delta x_t(\Delta t) = \left| \frac{2}{\Delta t} \sum_{i=t}^{t+\Delta t/2} x_i - \frac{2}{\Delta t} \sum_{i=t+\Delta t/2}^{t+\Delta t} x_i \right|, \tag{5}$$

and then the Haar structure function is given by

$$S_q^{\text{Haar}}(\Delta t) = \frac{1}{N - \Delta t} \sum_{t=1}^{N-\Delta t} \Delta x_t(\Delta t)^q. \tag{6}$$

The Haar fluctuation function used extensively by Lovejoy and Schertzer in arguing for the existence of transitions between scaling regimes is defined as

$$F^{\text{Haar}}(\Delta t) = \sqrt{S_2^{\text{Haar}}(\Delta t)}. \tag{7}$$

In Figure 2 we illustrate some features of the Haar fluctuation applied to the instrumental GMST. Figure 2a shows the GMST and a trend computed by fitting a second order polynomial. The black thick curve in Figure 2b is the Haar fluctuation function computed from the record in a log-log plot (the upper curve is shifted by a factor 10).

There are (at least) two different ways to model this record as a simple stochastic process. One is to assume that it is a linear combination of an fGn ($\beta < 1$) and an fBm ($\beta > 1$). The former will dominate the fluctuation function on the small scales, and the latter on the long scales. Hence this is a model that exhibits a scale break and two scaling regimes. We have estimated the slopes of the fluctuation function in these two regimes, and found $H \approx -0.1$ ($\beta = 2h + 1 \approx 0.8$) on the short scales, and $h = 0.3$ ($\beta \approx 1.6$) on the long scales. We then computed the weights of each process from using the estimated variance of the GMST on the shortest and longest scales, respectively, and computed an ensemble of realisations of their linear combination. The red curves in Figure 2b constitute 20 realisations in such an ensemble (multiplied by a factor 10). These curves demonstrate that the observed Haar-fluctuation is consistent with this model, but also that the uncertainty in the model prediction on scales longer than a decade is so large that that the observed fluctuation function here could also be consistent with a model where $h < 0$ ($\beta < 1$). Another obvious way to model the record is as a linear combination of the fGn and the quadratic trend. The resulting ensemble is shown as the blue curves in Figure 2b. The observed record is consistent with this model too, but the big difference is that in this case the model prediction on large scales is much more certain, and hence constitutes a "better" statistical model. An extra bonus is that the quadratic trend is physically well understood, since it corresponds closely to present knowledge about greenhouse forcing.

Thus, we have the choice between explaining the observation with a poor statistical model (many parameters, large prediction uncertainty, and no physics explaining the scale break) and a much better model (fewer parameters, lower uncertainty, and a clear physical explanation).

A common estimator for scaling exponents is the wavelet variance, i.e., to plot the variance of the wavelet coefficients versus scale in a log-log plot, and it is common to normalise the wavelet such that the slope for an fGn will be the spectral exponent β. In practice this can be obtained by squaring the Haar fluctuation $S_2^{1/2}(\Delta t)$ and multiplying by the scale Δt, i.e., we compute $\Delta t S_2(\Delta t)$. In Figure 3 we generate an ensemble of 10 fGns of 1000 data points with $H = -0.1$ ($\beta = 0.8$) and plot $S_2^{1/2}(\Delta t)$ for all realisations in the ensemble in Figure 3a , and $\Delta t S_2(\Delta t)$ for the same ensemble in Figure 3c. If we use the entire ensemble to estimate the slope we will get a quite accurate result (getting better the larger the ensemble), but if we estimate the slope from one realisation we make greater errors if we include the longer time scales in the fit. This is why some authors recommend not to include longer scales than $1/4$ of the record length and others recommend no more than $1/10$. If we fit a straight line we have no reason to believe that the Haar fluctuation gives less accurate estimates than the Haar wavelet variance. However, by inspecting the two for one particular realisation in the ensemble, as is done in Fig. 3b and d, we observe that the Haar

fluctuation can present a break in the curve that visually is much more pronounced than in the corresponding Haar wavelet variance. This is nothing but a visual illusion (the two plots contain the same information); a curve that changes its slope from negative to positive is more easily perceived to represent different qualities than the curve that only changes its positive slope somewhat. Since all the curves are produced from realisations of the fGn, all scale breaks are spurious, and caused by the diverging statistical uncertainty in the high-scale end. Thus, if the underlying scaling is close to $1/f$ noise, then weak trends or oscillations tend to appear as breaks in the Haar fluctuation curve, but are much less visible in the wavelet variance curve. A sound approach to graphical tools like this should avoid visualisations like the Haar fluctuation, which exaggerates such spurious breaks.

4 Data

The scaling is analysed in seven proxy/multiproxy temperature reconstructions representing late Holocene temperature, one temperature reconstruction representing the entire Holocene time period, in addition to six reconstructions of stable isotope ratios from the deep ice cores GRIP, GISP2 and NGRIP from Greenland, and EPICA, Taylor dome and Vostok from Antarctica. Information and analysis results from GISP2, NGRIP, Taylor and Vostok is provided in supplementary material. From the available ice core time series we extract sub-series covering only the Holocene and only the last glacial period, respectively. For the GRIP ice core we also extract a time series covering 0 – 85 kyr BP. Since the exact timing of the transition between the Holocene and the last glacial period is slightly different for Greenland and Antarctica, we have chosen the start and end of the time series carefully for each series, such that the transition is not contained in any of the "Holocene only" or the "glacial only" time series.

4.1 Proxy/multiproxy late Holocene temperature reconstructions

We have chosen seven proxy- or multiproxy based temperature reconstructions for our study, and in order to avoid the trend effect from anthropogenic warming we have discarded data after 1850 AD (see Table 1). All time series are given with annual resolution. A few of the reconstructions are based partly on the same raw proxy records, but we include all since the reconstruction methods are different. The Jones et al. (1998) multiproxy reconstruction represents northern hemisphere temperature. The Briffa et al. (2001) reconstruction represents the continental region 20° N – 90° N and is constructed from tree rings. The Esper et al. (2002) reconstruction is also based on tree rings and represent the continental region 30° N – 80° N. The Huang (2004) reconstruction is based on borehule temperatures, integrated with instrumental temperatures and the multiproxy reconstruction by Mann et al. (1999). The Moberg et al. (2005) multiproxy reconstruction represent northern hemisphere temperature, and is smoothed on the shortest time scales, so estimates of the scaling exponents are restricted to time scales from 4 years and longer. The Mann et al. (2009) multiproxy reconstruction

represents global temperature, and is smoothed up to decadal time scales. The Neukom et al. (2014) multiproxy reconstruction represents southern hemisphere temperature.

4.2 Multiproxy full Holocene temperature reconstruction

The temperature reconstruction described in Marcott et al. (2013) is included in our study because it covers the entire Holocene time period. The reconstruction is based on 73 proxy records with temporal resolution varying from 20 to 500 years. The spatial distribution of proxy data is near global, and there is a high percentage of data sets from marine sites. The proxy records were interpolated to 20-year resolution before constructing the temperature reconstruction, and the final record is presented with a 20-year resolution. This data set is unique in our study because the reconstructed temperature gets gradually smoother as one goes back in time. This is observed from the time series itself in Figure 8a. From the supplementary material of the Marcott et al. (2013) paper it is clear that the proxy records covering the most recent time also in general exhibit the best temporal resolution. The reconstructed temperature data for the past 1500 years therefore represent high-frequency variability in a more realistic way than the remaining part of the reconstruction.

4.3 The GRIP ice core

The European multinational research project "Greenland Ice Core Project" (GRIP) completed drilling a 3028 m deep ice core from central Greenland in 1992 (Dansgaard et al., 1993). Two GRIP data sets are used in this study, one with high temporal resolution covering 0 – 91 kyr BP (Ditlevsen et al., 1996), and one with lower temporal resolution covering 0 – 250 kyr BP (Greenland Ice-Core Project (GRIP) Members, 1993; Johnsen et al., 1997). The high-resolution data set was provided by Peter Ditlevsen at the Centre for Ice and Climate, Niels Bohr Institute, University of Copenhagen, personal communication. Both data sets are used to estimate the scaling exponents, but the results shown in Sect. 5 are for the high resolution time series. Both temperature reconstructions are based on δ^{18}O.

4.4 The EPICA ice core

The European Project for Ice Coring in Antarctica (EPICA) drilled two deep ice cores in Antarctica between 1996 and 2006. Here we focus on the core from dome C at the East Antarctic Plateau, covering the past 740 000 years (EPICA community members, 2004; Jouzel et al., 2007). The temperature reconstruction is based on δD.

5 Results

5.1 Results for late Holocene multiproxy reconstructions

Three approaches are used to detect a scale break from the spectra of the seven multiproxy temperature reconstructions. The first is to assume a scale break at exactly 100 years, and then estimate β for long and short time scales, and determine the uncertainties for each estimate. By this approach we demonstrate that scale breaks may occur by chance from a mono-scaling model, without being statistically significant. The second approach is to use a procedure for automatic detection of a scale break from a two-scaling regimes hypothesis, and show that a wide range of time scales τ_c for the break, and a wide range of scaling exponents β_1, β_2, arise by applying the procedure to a Monte Carlo ensemble of monoscaling time series. We also employ the structure function approach, and show that the enhanced power at large scales is associated with an oscillation with characteristic scale around 500 yr.

Fig. 4 illustrates the procedure and results for the Moberg temperature reconstruction, using the first approach. The scaling exponent β is estimated from the standard periodogram of the reconstructed data, for time scales shorter than 10^2 yr ($\beta_{1,\mathrm{data}}$) and for time scales longer than 10^2 yr ($\beta_{2,\mathrm{data}}$), as shown in Fig. 4b. A Monte Carlo (MC) ensemble of synthetic fGn's with 2000 members is then constructed with $\beta_{1,\mathrm{data}}$, and from the spectra (Fig. 4c), the same estimation technique is used to estimate $\beta_{1,\mathrm{MC}}$ and $\beta_{2,\mathrm{MC}}$ for each realization. From the distribution of the estimated $\beta_{1,\mathrm{MC}}$ and $\beta_{2,\mathrm{MC}}$, the 95% confidence ranges are computed. Fig. 4d shows the mean and 95% confidence range for $\beta_{2,\mathrm{MC}}$. Since the blue line ($\beta_{2,\mathrm{data}}$) is within the confidence range for a MC ensemble of fGn's with β=0.8, the single-scaling regime hypothesis cannot be rejected. Results for all seven reconstructions are shown in Table 1.

For the Esper at al. (2002) reconstruction the estimate of $\beta_{1,\mathrm{data}}$ is slightly outside the confidence range, but this is due a bias of the synthetic fBm for β slightly higher than unity (see supplementary material section 1.1 for further details). This deviation should therefore be ignored.

From the second approach we obtain for each reconstruction two values of β and a time for the scale break. The procedure is to fit two line segments with slopes β_1 and β_2 to the log-log spectrum, such that they join at $f = f_c = 1/\tau_c$. The two slopes and the transition frequency f_c are the parameters to be fitted by an ordinary least-square procedure. Results for the seven temperature reconstructions are provided in Table 2, where also the differences in β-values are included. The scale-break hypothesis of Lovejoy and Schertzer (2012) states that the the difference $\beta_2 - \beta_1$ should be around unity. This procedure has also been tested on a Monte Carlo ensemble of mono-scaling fGn's. Fig. 5 shows a histogram of the differences in estimated β_2 and β_1. The histogram shows that the scale breaks detected by this procedure in the multiproxy records are not unlikely to be detected in records with a single scaling regime, i.e., their detection does not reject the single-scaling regime hypothesis. A histogram of τ_c also shows a broad distribution, (figure not shown).

In Figure 6a we plot SFs for the cumulative sum of the Moberg multiproxy reconstruction. The SFs are straight in the log-log plot up to around 500 yr, but then there is a broad bump. By examining the record it becomes apparent that this bump is associated with an oscillation with period of order of a millennium that involves the Medieval Warm Anomaly (MWA) high and the Little Ice Age (LIA) low. The fact that this oscillation shows up in the high-order SFs indicates that its amplitude is larger than consistent with the underlying persistent noise. If we fit the SFs by straight lines up to 500 yr we obtain the scaling function in Figure 6b with slope $H_u \approx 0.87$, in very good agreement with what was found from the instrumental data. If this oscillation were a manifestation of a new scaling regime with $\beta > 1.4$ ($h > 0.2,\ H_u > 1.2$), we should expect the SFs in Figure 6a to be straight with slope q on scales $> \tau_c$, and the scaling function obtained by fitting lines to the SFs on these scales to have slope $H_u = 1$ (this is easily demonstrated by Monte Carlo simulations). What we observe, however, is a downward bend caused by the oscillation discussed above. It can be correctly argued that this bend is not statistically significant, since we only have one sample of it, but it demonstrates very clearly that there are cases where fluctuation measures like the periodogram or the Haar fluctuation function will suggest a new scaling regime with higher scaling exponents (see Figure 7b), while the SF method will suggest oscillations.

The issue of modelling the fluctuations on multi-century time scale as a second scaling regime or an oscillation is illustrated in Figure 7. The idea is the same as in Figure 2, but now the "trend" is modelled as a growing oscillation on the form $At\sin[\omega(t-\varphi)]$. The Moberg record and the fitted trend is shown in Figure 7a, and Figure 7b shows the Haar fluctuation of the record along with realisations of a model comprised of a linear combination of an fGn and an fBm (red curves), and realizations of a model comprised of a linear combination of an fGn and the oscillatory trend. Again, the observation is consistent with both models, but the latter exhibits smaller uncertainties at longer scales, and hence is a better statistical model. Rypdal and Rypdal (2014) attribute the oscillatory trend to a combination of volcanic and solar forcing, and show that the residual after subtracting the response to this forcing is well modelled as an fGn with h compatible with what was used for the short scales in Figure 7b. Thus, compared to a two-scaling regime model, the simpler and more accurate statistical model of the Moberg record is to model the internal variability as a persistent fGn for all scales up to the length of the record, superposed on a forced oscillatory trend.

An important feature shown by the scaling functions in Figures 1 and 6 is that the background noise in GMST is monofractal. It is also Gaussian. This means that a fractional Gaussian noise is a good model for these fluctuations, and hence that all essential information is contained in the scaling exponent and the variance. This is also true for ice core data in the Holocene, while during the last glaciation, ice core data are neither monofractal nor multifractal. In an accompanying paper from our research group (Rypdal and Rypdal, 2015) it is demonstrated that if the transitions between stadial and interstadials associated with DO events and glacial/interglacial transitions are removed from the ice-core records, the remaining fluctuations scales roughly as a $1/f$ noise ($\beta \approx 1$) on time scale

longer than a century. In other words, over the length of the Antarctic ice core record (800 kyr) the temperature variability can be described as a series of glacial/interglacial transitions, and within the glacial periods; a series of stadial/interstadial transitions, superposed on a background $1/f$ noise. This suggests that Holocene variability should also exhibit this scaling of the climate noise, and the analysis of the Moberg record we have made here does not reject that hypothesis.

5.2 Results for the full Holocene multiproxy reconstruction

The reconstruction by Marcott et al. (2013) has been analysed with the periodogram in a particular way to take into account the increasing smoothness of the record as one goes backward in time. If we compute the standard periodogram for the full time series, the resulting spectral exponent is $\beta = 2.9$. The power is artificially low at high frequencies, and this is corrected by dividing the time series into segments S_n of lengths $2^n \times 400$ yr, with $n = 0, 1, 2 \ldots, 5$, and starting with the most recent period. Hence, $S_1 = 50 - 450$ yr BP, $S_2 = 50 - 850$ yr BP, $S_3 = 50 - 1650$ yr BP, $S_4 = 50 - 3250$ yr BP, $S_5 = 50 - 6450$ yr BP, and $S_6 = 50 - 11290$ yr BP (longest possible record, shorter than $2^5 \times 400$). The periodogram was estimated for each segment, and then a new power spectrum was created using only parts of each segment assumed to be trustworthy with regard to preserved variability. All of S_1 was included, while for $S_2, \ldots, S_6$ only the low-frequency parts were included (none overlapping). By this composition, the resulting power spectrum represents the variability on all time scales more correctly. Figure 8c shows the spectra of all six segments, in addition to the corrected spectrum (blue dots, black line). The value of β=1.3 is estimated from this line. The corrected spectrum still does not represent the true scaling of the global temperature, but it is a better representation than the periodogram of the original record.

5.3 Results for ice core time series

For the time series plots, time on the horizontal axis is given in years BP (before present), where "present" is defined as 1950 AD. The spectral analysis is presented in a double-logarithmic plot. The raw periodogram is plotted in gray, while the log-binned version is marked by black points. The spectral index β is estimated from the log-binned periodogram in the region shown by the blue line. Finally, the blue, shaded area indicate the 95% confidence range estimated from an ensemble of synthetic fGn's/fBm's with β and variance estimated from the log-binned periodogram. The plot of the wavelet scalogram is included in this section only for the GRIP Holocene/past85 kyr record.

For the last glacial period, we present time series and periodograms for a time interval of ≈ 80 kyr. Spectral analysis results are also included for a combined Holocene/last glacial period time series from the GRIP ice core to illustrate that analysis of such records will be dominated by the glacial climate and suppress the characteristics of Holocene climate.

5.3.1 Results from the GRIP ice core

Fig. 9a shows the δ^{18}O time series of the Holocene part of the high-resolution GRIP ice core, and Fig. 9b the periodogram from the same time series. Fig. 9c displays the same time series as shown in (a), but with the earliest 2500 yr removed. Fig 9d shows the periodogram for the time series in (c). The rationale for removing the earliest part of the Holocene record can be seen from Fig. 9a, where one observes a decrease in δ^{18}O around 8 kyr BP. This particular decrease is often observed in paleotemperature records from the northern hemisphere, and especially in records from the North-Atlantic region. The feature is known as the 8.2 kyr event, and the temperature change was probably caused by a large pulse of freshwater into the North-Atlantic Ocean associated with the collapse of the Laurentide ice sheet (Alley and Agustsdottir, 2005). In Fig. 9b, β is estimated to be ≈ 0.3 for time scales up to 10^3 yr. No scale break is detected on centennial time scales. The low value of β is typical for local temperature data from continental sites (Blender and Fraedrich, 2003; Fraedrich and Blender, 2003). On time scales longer than a millennium we can infer a higher β, but still $\beta < 1$. Since the 8.2 kyr event might affect the scaling we also analysed the shorter record (Fig. 9c). The periodogram for this time series is essentially flat. Fig. 10 shows the Mexican hat and Morlet wavelet scalograms for the full Holocene section of the GRIP ice core. The 8.2 kyr event clearly increases the power at millennial time scales, and this event is the source of the increased power observed at that time scale in Fig. 9b. From the periodogram of the Holocene part of the low-resolution GRIP time series we estimate $\beta \approx 0.1$ (not shown in figure).

Fig. 11a displays the δ^{18}O time series for the GRIP ice core from the last glacial period, and Fig. 11b the periodogram for the same time period. In Fig. 11a the Dansgaard-Oeschger (DO) events are observable as rapid warming over decadal time scales, followed by more gradual cooling (Bond and Lotti, 1995). In Fig. 11b we find $\beta \approx 1.8$ for time scales longer than 10^2 yr and shorter than 10^4 yr. On centennial time scales and shorter, the spectrum is flatter. This means that a hypothesis of a scale break at centennial time scales is plausible under glacial climate conditions, even though such a scale break could not be identified from the Holocene time series. From the low-resolution GRIP data set we estimate $\beta \approx 1.3$ for time scales longer than centennial, and a scale break is seen at this scale (figure not shown).

Fig. 12a shows the past 85 kyr time series of the high-resolution GRIP ice core, and Fig. 12b the periodogram for the same time series. In Fig. 12b, $\beta \approx 1.6$ for time scales longer than centennial, and a scale break is visible at this scale. The periodogram in Fig. 12 is very similar to that in Fig. 11, indicating that the information from the Holocene is suppressed to a high degree in the periodogram of this time series.

5.3.2 Results from the EPICA ice core

Fig. 13a shows the Holocene time series of the EPICA ice core, and Fig. 13b the periodogram for the same time series. In Fig. 13a the Antarctic equivalent to the Northern Hemisphere Holocene climate optimum (HCO) occurred between 11500 and 9000 yr BP (Masson et al., 2000). In Fig. 13b, $\beta \approx 0$ for time scales shorter than 10^3 yr.

Fig. 14a shows the time series of the EPICA ice core from the last glacial period, and Fig. 14b the periodogram for the same time period. We observe from the time series that the fluctuations do not coincide with the DO events in the GRIP ice core with respect to timing and amplitude. Like the glacial part of the GRIP ice core the EPICA glacial time series Fig. 14a has higher fluctuation levels than the Holocene counterpart. In Fig. 14b we estimate $\beta \approx 1.6$ for 10^3 yr $< \tau < 10^4$ yr. The scale break in this figures appears at 10^3 yr.

6 Discussion and conclusions

In this paper we have examined a number of paleoclimatic temperature records to assess the feasibility of detection with confidence multiple scaling regimes in Holocene climate, and in particular a break in scaling around centennial time scales. Seven proxy/multiproxy reconstructions from the late Holocene, and one for the entire Holocene, have been selected for analysis due to high temporal resolution and coverage in time, and six reconstructions from deep ice cores sampled at Greenland and Antarctica also meet our requirements for temporal coverage and resolution.

For the seven proxy-based temperature reconstructions, our first approach was to assume a break at exactly 100 years. Obviously there are few data points available for estimation on the longer scales using this procedure and the estimated values of β_2 are within the uncertainties of a mono-scaling model for all seven reconstructions. The scale break is therefore not statistically significant. For the second approach, our systematic procedure detects a break in scaling for all reconstructions. The time scale for the break varies significantly between reconstructions and is in most cases not even located near centennial time scales. The differences $\beta_2 - \beta_1$ varies over a great range and takes on both positive and negative values. This procedure has also also been tested on a Monte Carlo ensemble of fGn's and demonstrates that we will find such apparent breaks even in data that should not have breaks.

The discussion and correction of the temperature reconstruction by Marcott et al. (2013) illustrates the potential pitfall of uncritically selecting paleoclimatic time series for scaling analysis. The time series may be an excellent temperature reconstruction for many purposes, but for scaling analysis one needs to correct for the the fact that the data are increasingly low pass filtered as one goes backward in time.

In ice-core data, a scale break at centennial time scales can only be seen in records from the last glacial period. The time series for the Holocene from the GRIP and EPICA ice cores both exhibit

weak persistent scaling (Figs. 9 and 13). The scaling exponent is estimated to $\beta \approx 0.3$ and $\beta \approx 0$ for the two ice cores respectively up to millenium time scale. No break in scaling can be observed at centennial time scale. The low value of β obtained is consistent with the scaling exponents observed over land from the paleoclimate model run presented in Blender et al. (2006). On time scales longer than millenial we do not have enough data points to make confident estimates of β. From the wavelet scalograms we argue that the increase in power seen at the longest time scales in the GRIP periodogram can be attributed to the 8.2 kyr event.

The scaling properties of the GRIP and EPICA last glacial period are significantly different from the Holocene. A scale break at centennial time scales is identified with confidence from Figs. 11 and 14. We interpret this scale break as being associated with the variability of Dansgaard-Oeschger events and teleconnections to the Southern Hemisphere, (WAIS Divide Project Members, 2015). A number of theories and models exist for the mechanism of these events, see e.g., Dokken et al. (2013). This and other studies indicate that this variability is internal and not a direct response to external forcing.

From the GRIP time series including both the Holocene period and the last glacial period we obtain a power spectrum very similar to that of the glacial climate, Fig. 12. Since the glacial climate state is the dominating state in the Quaternary, the Holocene temperature variability is strongly suppressed when time series covering 100 kyr or longer are used to estimate scaling exponents. Our analyses of Holocene records, on the other hand, show that a scale break on centennial time scales is not a universal feature, and in those cases it appears to be present, it cannot be detected with sufficient certainty.

Faced with the results we have presented here one may ask what the practical implications are. Is scaling in climatic time series a useful concept? Our perception is that a scaling law may be useful as a statistical (stochastic) model when a causal description turns out to be very complex, i.e., when the viable alternative is something like a general circulation model. Such a statistical model does not have to exhibit long-memory scaling (a more standard model is a short-memory autoregressive process), but there is strong evidence, for internal variability of surface temperature data, that an fGn is a much better model than an AR(1) process for time scales at least up to centuries (Rypdal and Rypdal, 2014). Thus, for prediction on time scales up to decades, a mono-scaling model with $\beta < 1$ is what should be used, (Lovejoy et al., 2015). More interesting, however, is whether long-memory scaling is important for prediction on century time scales and beyond, and here the issue of non-stationary scaling ($\beta > 1$) for such time scales becomes crucial. What is the proper value of β to use in such prediction efforts in a warming Holocene climate? The conclusion we draw from our results is that, unless we ignore the knowledge that the present climate state of the Earth is an interglacial, we should still use $\beta < 1$. One argument that can be raised against that conclusion is that, even though we cannot reject the hypothesis of single-regime scaling based on available Holocene data, we cannot exclude that two scaling regimes is true either. Moreover, the latter is supported

from records spanning hundreds of kyr which encompass both glacial and interglacial climate. This stalemate reflects that we are faced with a model selection problem where the outcome depends on which available knowledge we prefer to emphasise.

The multiple regime model stresses the information we have on scaling in the second-order statistics such as power spectra ($\beta > 1$) on time scales up to hundreds of kyr (Quaternary scaling), and infers that this scaling should be a guideline for prediction independent of whether the initial state is glacial or interglacial. It essentially ignores the fact that Quaternary climate is characterised by several intermittencies, (Lovejoy and Schertzer, 2012). The dominating Quaternary climate state is the glacial, and temperature proxies from the last glacials take the form of a non-Gaussian intermittent stochastic process, displayed in its full glory by the Dansgaard-Oeschger events. One type of intermittency implies that the probability density functions (PDFs) are heavy-tailed on short time scales and approach Gaussian on longer time scales. This typically happens if the signal is bursty, but without long-range correlation between bursts. In this case the high-order structure functions are not straight lines in a log-log plot. Another, and more restricted, class of intermittent processes are those that are multifractal. Here we have correlated bursts, straight structure functions, but curved scaling function. Hence structure functions of higher order than two are needed to characterise the process. For a Gaussian process the power spectral density can be inferred from the second-order structure function and hence does not convey information beyond second-order statistics. Moreover, Quaternary climate is characterised by the glacial-interglacial transitions, which adds more intermittency, and all this intermittency makes prediction based only on Quaternary scaling very difficult.

The single-regime model, on the other hand, ignores the information available on time scales beyond the Holocene, but makes use of the fact that our present climate state is an interglacial, and that second-order statistics is sufficient to describe the scaling on the time-scales that is available to us in the Holocene. As discussed above, and by Rypdal and Rypdal (2015), single-regime scaling can be rejected by data that goes way beyond the Holocene if this scaling is supposed to account for DO-events and glacial/interglacial transistions, but there is no statistically significant empirical evidence that the scaling inferred by glacial-state data is present in the interglacial climate state (Rypdal and Rypdal, 2015).

The issue discussed in this paper is an example of a more general problem concerning scaling analysis that needs to be addressed in a systematic manner. On geological time scales the Earth is an evolving system. There are cycles, but the Earth rarely repeats itself. The Eemian was similar to the Holocene, but also very different, the most striking difference being the evolution of human civilisations. Thus, the dynamics of the Earth is non-stationary in a very fundamental sense. This makes scaling analysis, and modelling of Earth processes based on such analysis, a quite problematic issue. It has little meaning to talk about a universal scaling in Earths climate since the scaling characteristic on a given range of scales up to a chosen maximal scale $\tau_{\max}$ will depend on the eon, era, period, epoch, or age the analysis is done. In other words, the result will depend on the time t around which

the time range $\tau_{\max}$ is centred. A scaling analysis of a given Earth-system variable must therefore be conditioned by two essential parameters; the range $\tau_{\max}$ of scales considered, and the positioning t of this range in time. The time series and the wavelet scalogram of the GRIP temperature series for the past $t_{\max} = 90$ kyr illustrates the issue, as shown in Figure 15. The central time parameter t is along the horisontal axis and the scale τ along the vertical. We have no data for the future, which means that the transform cannot be computed correctly above the upper white line in the figure. Likewise, the area below the lower white curve is influenced by the interpolation made due to uneven sampling of the time series. It is apparent that the scalogram is different in the first 11.5 kyr (the Holocene) from the remaining 80 kyr (the last glacial). There is generally lower power on all scales in the Holocene, and the increase in power with increasing scale as t is kept constant is lower. We can also observe from this scalogram that the longest interstadials (warm stages) associated with DO-events exhibit variability very similar to the Holocene.

The literature reveals that there is no consensus of how this issue of non-stationarity could be handled. Ice cores restrict the information we can obtain to somewhat less than $\tau_{\max} = 1$ Myr BP. This is the range of time scales considered in the work of Lovejoy and Schertzer (2012) and Huybers and Curry (2006), and the period is the Quaternary in which the Earth's climate has been in a bistable state shifting between glacials and interglacials. The methodology and interpretations are based on this choice of the parameters $(\tau_{\max}, t)$. We don't see anything wrong with that, as long as one is mindful on that this is a choice, and recognises that there are other, equally valid, choices. If the issue is understanding of the present and future climate in our present interglacial state, we don't believe this choice is useful, simply because it ignores the knowledge that the Earth at present resides in an interglacial state and probably will continue to do so as long as there is human civilisation and anthropogenic forcing on this planet. Nevertheless, the recent results by (Rypdal and Rypdal, 2015) suggests that something useful can be learnt about Holocene climate from data spanning the last glacials and interglacials, namely that periods between the abrupt climate transitions are characterised by monoscaling of $1/f$-type, very similar to the scaling of the temperature variability in our present climate.

Acknowledgements. This paper was supported by the the Norwegian Research Council (KLIMAFORSK programme) under grant no. 229754. We are grateful for discussions with Martin Rypdal and Dmitry Divine, technical assistance from Ola Løvsletten and Peter Ditlevsen for providing the high-resolution GRIP data set.

References

Alley, R. B. and Agustsdottir, A. M.: The 8k event: cause and consequences of a major Holocene abrupt climate change, Quaternary Science Reviews, 24, 1123–1149, doi:10.1016/j.quascirev.2004.12.004, 2005.

Blender, R. and Fraedrich, K.: Long time memory in global warming simulations, Geophys. Res. Lett., 30, doi:10.1029/2003GL017666, 2003.

Blender, R., Fraedrich, K., and Hunt, B.: Millennial climate variability: GCM-simulation and Greenland ice cores, Geophys. Res. Lett., 33, doi:10.1029/2005GL024919, 2006.

Bond, G. C. and Lotti, R.: Iceberg Discharges into the North Atlantic on Millennial Time Scales During the Last Glaciation, Science, 267, 1005–1010, doi:10.1126/science.267.5200.1005, 1995.

Briffa, K. R., Osborn, T. J., Schweingruber, F. H., Harris, I. C., Jones, P. D., Shiyatov, S. G., and Vaganov, E. A.: Low-frequency temperature variations from a northern tree ring density network, J. Geophys. Res.-Atmos, 106, 2929–2941, doi:10.1029/2000JD900617, 2001.

Broersen, P. M. T., de Waele, S., and Bos, R.: The accuracy of time series analysis for laser-doppler velocimetry, Selected Papers from the 10th International Symposium Lisbon, Portugal July 10?13, 2000, in: Laser Techniques for Fluid Mechanics, 2000.

Dansgaard, W., Johnsen, S. J., Clausen, H. B., Dahl-Jensen, D., Gundestrup, N. S., Hammer, C. U., Hvidberg, C. S., Steffensen, J. P., Sveinbjörnsdottir, A. E., Jouzel, J., and Bond, G.: Evidence for general instability of past climate from a 250-kyr ice-core record, Nature, 364, 218–220, doi:10.1038/364218a0, 1993.

Ditlevsen, P. D., Svensmark, H., and Johnsen, S.: Contrasting atmospheric and climate dynamics of the last-glacial and Holocene periods, Nature, 379, 810–812, doi:10.1038/379810a0, 1996.

Dokken, T. M., Nisancioglu, K. H., Li, C., Battisti, D. S., and Kissel, C.: Dansgaard-Oeschger cycles: Interactions between ocean and sea ice intrinsic to the Nordic seas, Paleoceanography, 28, 491–502, doi:10.1002/palo.20042, 2013.

Efstathiou, M. N., Tzanis, C., Cracknell, A. P., and Varotsos, C. A.: New features of land and sea surface temperature anomalies, Int. J. Remote Sensing, 32, 3231–3238, doi:10.1080/01431161.2010.541504, 2011.

EPICA community members: Eight glacial cycles from an Antarctic ice core, Nature, 429, 623–628, doi:10.1038/nature02599, 2004.

Esper, J., Cook, E. R., and Schweingruber, F. H.: Low-Frequency Signals in Long Tree-Ring Chronologies for Reconstructing Past Temperature Variability, Science, 295, 2250–2253, doi:10.1126/science.1066208, 2002.

Fraedrich, K. and Blender, R.: Scaling of Atmosphere and Ocean Temperature Correlations in Observations and Climate Models, Phys. Rev. Lett., 90, 108 501, doi:10.1103/PhysRevLett.90.108501, 2003.

Greenland Ice-Core Project (GRIP) Members: Climate instability during the last interglacial period recorded in the GRIP ice core, Nature, 364, 203–207, doi:10.1038/364203a0, 1993.

Huang, S.: Merging information from different resources for new insights into climate change in the past and future, Geophys. Res. Lett., 31, doi:10.1029/2004GL019781, 2004.

Huybers, P. and Curry, W.: Links between annual, Milankovitch and continuum temperature variability, Nature, 441, doi:10.1038/nature04745, 2006.

Johnsen, S. J., Clausen, H. B., Dansgaard, W., Gundestrup, N. S., Hammer, C. U., Andersden, U., Andersen, K. K., Hvidberg, C. S., Dahl-Jensen, D., Steffensen, J. P., Shoji, H., Sveinbjörnsdottir, A. E., White, J., Jouzel, J., and Fisher, D.: The 18-O record along the Greenland Ice Core Project deep ice core

and the problem of possible Eemian climatic instability, J. Geophys. Res.-Oceans, 102, 26, 397–26, 410, doi:10.1029/97JC00167, 1997.

Jones, P. D., Briffa, K. R., Barnett, T. P., and Tett, S. F. B.: High-resolution palaeoclimatic records for the last millennium: interpretation, integration and comparison with General Circulation Model control-run temperatures, The Holocene, 8, 455–471, doi:10.1191/095968398667194956, 1998.

Jouzel, J., Masson-Delmotte, V., Cattani, O., Dreyfus, G., Falourd, S., Hoffmann, G., Minster, B., Nouet, J., Barnola, J. M., Chappellaz, J., Fischer, H., Gallet, J. C., Johnsen, S., Leuenberger, M., Loulergue, L., Luethi, D., Oerter, H., Parrenin, F., Raisbeck, G., Raynaud, D., Schilt, A., Schwander, J., Selmo, E., Souchez, R., Spahni, R., Stauffer, B., Steffensen, J. P., Stenni, B., Stocker, T. F., Tison, J. L., Werner, M., and Wolff, E. W.: Orbital and Millennial Antarctic Climate Variability over the Past 800,000 Years, Science, 317, 793–796, doi:10.1126/science.1141038, 2007.

Kantelhardt, J. W.: Fractal and Multifractal Time Series, in: Mathematics of Complexity and Dynamical Systems, edited by Meyers, R. A., pp. 463–487, Springer New York, doi:10.1007/978-1-4614-1806-1_30, 2011.

Koscielny-Bunde, A. B., Havlin, S., and Goldreich, Y.: Analysis of daily temperature fluctuations, Physica A, 231, 393–396, doi:10.1016/0378-4371(96)00187-2, 1996.

Laepple, T. and Huybers, P.: Ocean surface temperature variability: Large model-data differences at decadal and longer periods, P. Natl. A. Sci., 111, 16 682–16 687, doi:10.1073/pnas.1412077111, 2014.

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

Lovejoy, S. and Schertzer., D.: Haar wavelets, fluctuations and structure functions: convenient choices for geophysics, Nonlinear Processes in Geophysics, 19, 513–527, doi:10.5194/npg-19-513-2012, 2012.

Lovejoy, S. and Schertzer, D.: Low Frequency Weather and the Emergence of the Climate, pp. 231–254, 196, American Geophysical Union, doi:10.1029/2011GM001087, 2012.

Lovejoy, S., Schertzer, D., and Varon, D.: Do GCMs predict the climate ... or macroweather?, Earth Syst. Dynam., 4, 439–454, 2013.

Lovejoy, S., del Rio Amador, L., and Hébert, R.: The Scaling LInear Macroweather model (SLIM): using scaling to forecast global scale macroweather from months to decades, Earth System Dynamics Discussions, 6, 489–545, 2015.

Mann, M.: On long range dependence in global surface temperature series, Climatic Change, 107, 267–276, doi:10.1007/s10584-010-9998-z, 2011.

Mann, M. E., Bradley, R. S., and Hughes, M. K.: Northern Hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations, Geophys. Res. Lett., 26, 759–762, doi:10.1029/1999GL900070, 1999.

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

Marcott, S. A. a. D. S., Clark, P. U., and Mix, A. C.: A Reconstruction of Regional and Global Temperature for the Past 11,300 Years, Science, 339, 1198–1201, doi:10.1126/science.1228026, 2013.

Masson, V., Vimeux, F., Jouzel, J., Morgan, V., Delmotte, M., Ciais, P., Hammer, C., Johnsen, S., Lipenkov, V. Y., Mosley-Thompson, E., Petit, J.-R., Steig, E. J., Stievenard, M., and Vaikmae, R.: Holocene Cli-

765 mate Variability in Antarctica Based on 11 Ice-Core Isotopic Records, Quaternary Res., 54, 348–358, doi:10.1006/qres.2000.2172, 2000.

Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M., and Karlén, W.: Highly variable Northern Hemisphere temperatures reconstructed from low-and high-resolution proxy data, Nature, 433, 613–617, doi:10.1038/nature03265, 2005.

770 Neukom, R., Gergis, J., Karoly, D. J., Wanner, H., Curran, M., Elbert, J., Gonzalez-Rouco, F., Linsley, B. K., Moy, A. D., Mundo, I., Raible, C. C., Steig, E. J., van Ommen, T., Vance, T., Villalba, R., Zinke, J., and Frank, D.: Inter-hemispheric temperature variability over the past millennium, Nature Climate Change, 4, 362–367, doi:10.1038/nclimate2174, 2014.

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

Pelletier, J. D.: The power spectral density of atmosphereic temperature from time scales of 10^{-2} to 10^{6} yr., Earth and Planetary Science Letters, 158, 157–164, doi:10.1016/S0012-821X(98)00051-X, 1998.

Rehfeld, K., Marwan, N., Heitzig, J., and Kurths, J.: Comparison of correlation analysis techniques for irregu-
780 larly sampled time series, Nonlinear Processes in Geophysics, 18, 389–404, doi:10.5194/npg-18-389-2011, 2011.

Roe, G. H. and Steig, E. J.: Characterization of Millennial-Scale Climate Variability, J. Climate, 17, 1929–1944, doi:10.1175/1520-0442(2004)017<1929:COMCV>2.0.CO;2, 2004.

Rybski, D., Bunde, A., Havlin, S., and von Storch, H.: Long-term persistence in climate and the detection
785 problem, Geophys. Res. Lett., 33, doi:10.1029/2005GL025591, 2006.

Rypdal, K., Østvand, L., and Rypdal, M.: Long-range memory in Earth's surface temperature on time scales from months to centuries, J. Geophys. Res., 118, 7046–7062, doi:10.1002/jgrd.50399, 2013.

Rypdal, M. and Rypdal, K.: Testing Hypotheses about Sun-Climate Complexity Linking, Phys. Rev. Lett., 104, 128 501–4, doi:10.1103/PhysRevLett.104.128501, 2010.

790 Rypdal, M. and Rypdal, K.: Long-memory effects in linear-response models of Earth's temperature and implications for future global warming, J. Climate, 27, doi:10.1175/JCLI-D-13-00296.1, 2014.

Rypdal, M. and Rypdal, K.: Late Quaternary temperature variability described as abrupt transitions on a $1/f$ noise background, submitted to Earth Syst.Dyn. Discuss., 2015.

Scargle, J. D.: Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly
795 sampled data, Astrophys. J., 263, 835–853, doi:10.1086/160554, 1982.

Torrence, C. and Compo, G. P.: A Practical Guide to Wavelet Analysis, Bull. Amer. Met. Soc., 79, 61–78, doi:10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2, 1998.

WAIS Divide Project Members: Precise interpolar phasing of abrupt climate change during the last ice age, Nature, 520, 661–665, doi:10.1038/nature14401, 2015.

Table 1. Results using approach 1 for multiproxy temperature reconstructions

Reconstruction	Time period	$\beta_{1,\mathrm{data}}$	conf. range for $\beta_{1,\mathrm{MC}}$	$\beta_{2,\mathrm{data}}$	conf. range for $\beta_{2,\mathrm{MC}}$
Jones et al. 1998	1000 - 1850	0.5	(0.4, 0.7)	1.2	(-0.8, 1.7)
Briffa et al. 2001	1402 - 1850	0.6	(0.4, 0.8)	2.9	(-2.0, 3.0)
Esper et al. 2002	831 - 1850	1.3	(0.8, 1.2)	1.2	(0.2, 3.3)
Huang 2004	1500 - 1850	0.7	(0.6, 1.0)	2.3	(-4.4, 6.0)
Moberg et al. 2005	0 - 1850	0.8	(0.6, 1.0)	1.2	(0.0, 1.5)
Mann et al. 2008	500 - 1850	2.5	(1.9, 2.6)	1.6	(1.5, 3.1)
Neukom et al. 2014	1000 - 1850	0.6	(0.4, 0.8)	1.3	(-0.8, 1.9)

Table 2. Results using approach 2 for multiproxy temperature reconstructions

Data set	$\beta_{1,\mathrm{data}}$	$\beta_{2,\mathrm{data}}$	τ_c (yr)	$\beta_2 - \beta_1$
Jones et al. 1998	0.5	0.9	38	0.4
Briffa et al. 2001	0.9	0.2	22	-0.7
Esper et al. 2002	1.4	1.0	38	-0.4
Huang 2004	0.8	2.2	94	1.4
Moberg et al. 2005	0.7	2.6	353	1.9
Mann et al. 2008	3.1	0.9	47	-2.2
Neukom et al. 2014	0.5	0.8	9	0.3

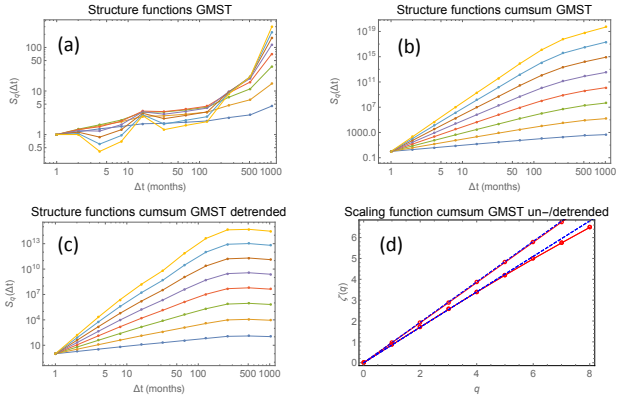

Figure 1. (a): Structure function estimates (empirical moments) $S_q(\Delta t) = (N-\Delta t)^{-1} \sum_{i=1}^{N-\Delta t} |T(t_i+\Delta t) - T(t_i)|^q$ for the GMST (HadCRUT3) monthly record 1880-2010; $T(t_i);\ i = 1, \ldots, N$. (b): Structure function for the cumulative sum $y_{t_i} = \sum_{j=1}^{i} T(t_j)$. (c): Structure function for the cumulative sum of the quadratically detrended GMST. (d): Scaling functions for the undetrended cumulative sum (upper line) and the detrended cumulative sum (lower line).

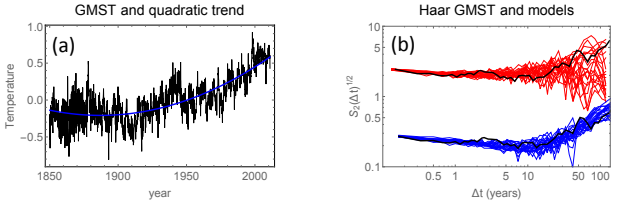

Figure 2. (a): The instrumental global mean surface temperature (GMST) 1850 – 2010 (black). A second order polynomial least-square fit to the GMST record (blue). (b): Black curves are the Haar fluctuation function of the GMST, the upper is multiplied by 10. The red curves are Haar fluctuation functions of 20 realisations of a model comprised of a linear combination of an fGn with $h = -0.1$ and an fBm with $h = 0.3$. The blue curves are the same of a model comprised of a linear combination of an fGn with $h = -0.1$ and the second-order polynomial trend.

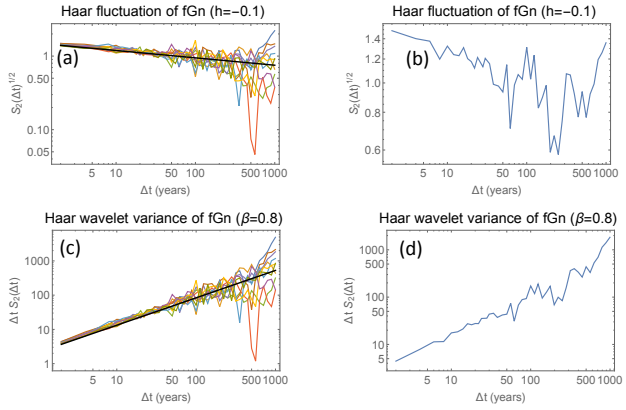

Figure 3. (a) The Haar fluctuation of 20 realisations of an fGn with h=-0.1, corresponding to $\beta = 0.8$. (b) The Haar-fluctuation of one realisation in the ensemble. (c) The Haar wavelet variance of the same 20 realisations as in panel (a). (d) The Haar wavelet variance of the same realisation as in panel (b).

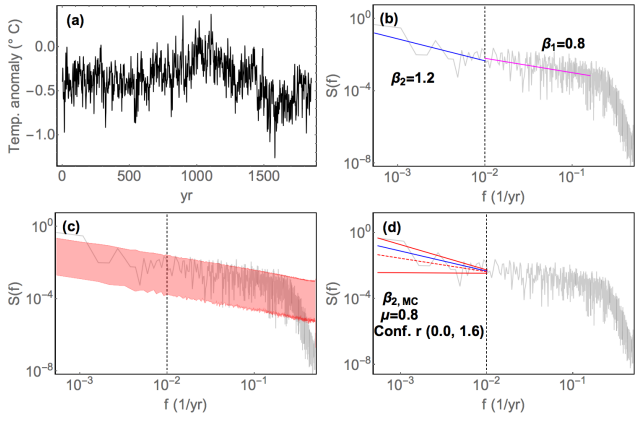

Figure 4. a) The Moberg et al. (2005) reconstructed temperature for the Northern hemisphere. (b) Estimated values of β_1 and β_2. (c) 95% confidence range for periodograms in Monte Carlo study. (d) 95% confidence range for estimates of β_2.

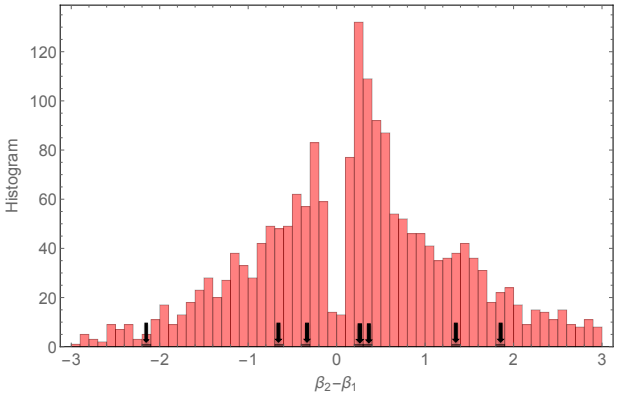

Figure 5. Differences in β_2 and β_1 for a Monte Carlo ensemble with 2000 members of synthetic LRM processes with β=0.7. The black arrows indicate the differences from the multiproxy reconstructions.

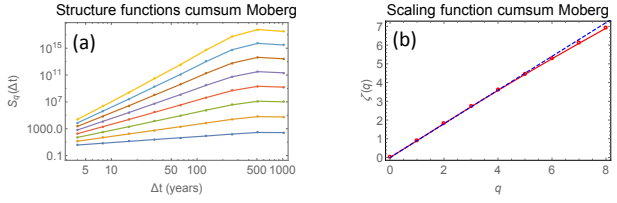

Figure 6. (a): Structure functions for the cumulative sum of the Moberg NH reconstruction year 0–1979. (b): Scaling functions for the cumulative sum computed from straight line fits to the SFs in the scale range 1 – 500 yr.

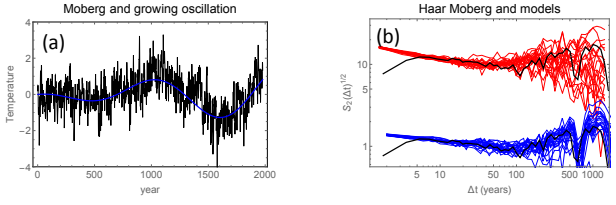

Figure 7. (a): The Moberg reconstruction (black). A fit of the function $At\sin[\omega(t-\varphi)]$ (blue). (b): Black curves are the Haar fluctuation function of the Moberg record, the upper is multiplied by 10. The red curves are Haar fluctuation functions of 20 realisations of a model comprised of a linear combination of an fGn with $h=-0.2$ and an fBm with $h=0.3$. The blue curves are the same of a model comprised of a linear combination of an fGn with $h=-0.2$ and the trend $At\sin[\omega(t-\varphi)]$.

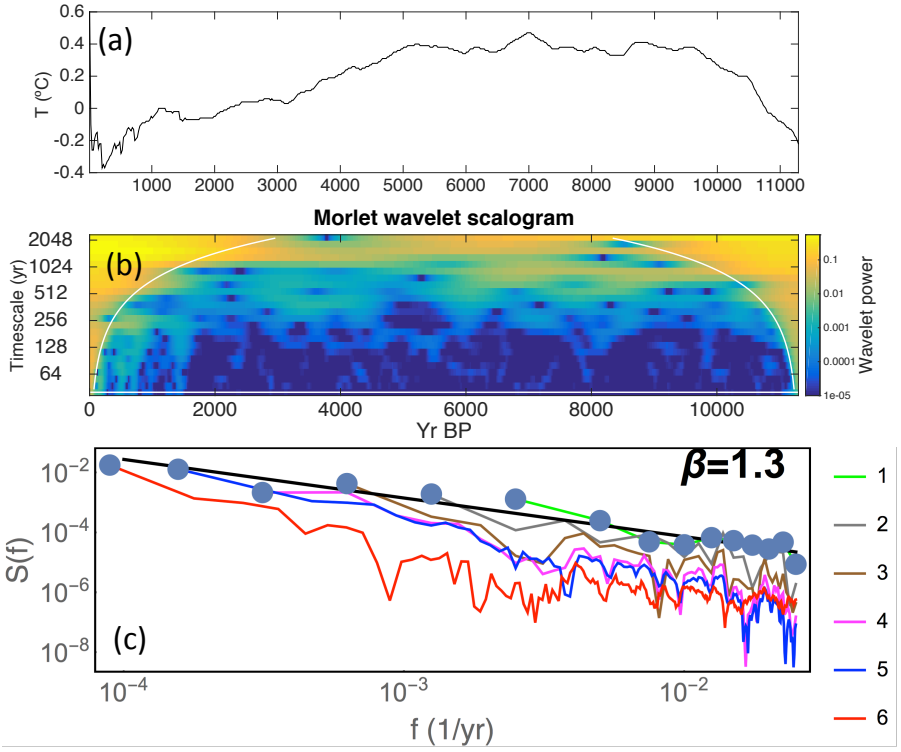

Figure 8. (a) The reconstructed time series covering the entire Holocene time period. (b) Wavelet scalogram for the same time period. (c) Section 1-6 of the reconstruction described in section 5.2, and composite spectrum (black line, blue dots). The estimated β is estimated for this line.

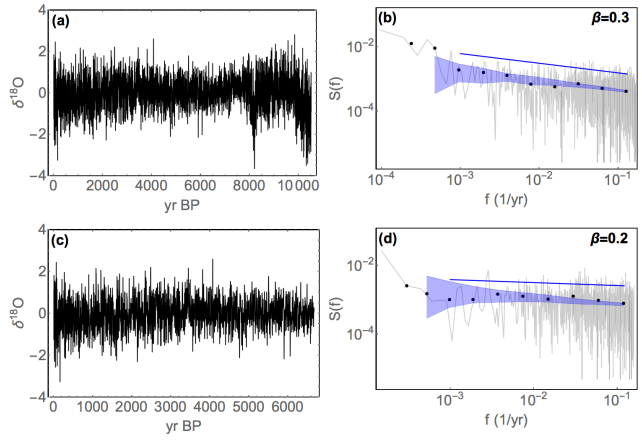

Figure 9. (a): δ^{18}O anomalies from the Holocene part of the high-resolution GRIP ice core. (b): Periodogram. The raw periodogram is shown in gray, the log-binned version by black dots. β is estimated from the log-binned periodogram in the region marked by the blue line. The confidence range is shown by the blue, shaded area, estimated from a Monte Carlo ensemble of synthetic fGns with the estimated value of β and variance from the log-binned periodogram. (c) Same figure as in (a) except the oldest section has been removed. (d) Periodogram for the time series in (c).

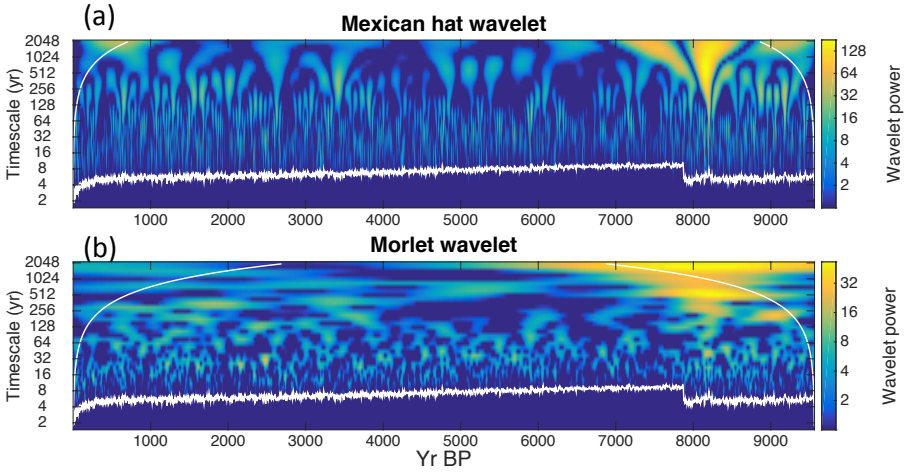

Figure 10. Top: the Mexican hat wavelet scalogram for the Holocene part of the GRIP ice core, and bottom: the Morlet wavelet scalogram for the same time series. The lower white curve in each plot denotes the varying Nyquist frequency, and the upper white curve the area affected by edge effects. Studies are restricted to the area between the two curves. The color bar to the right of the figure is used to indicate the the magnitude of the wavelet power.

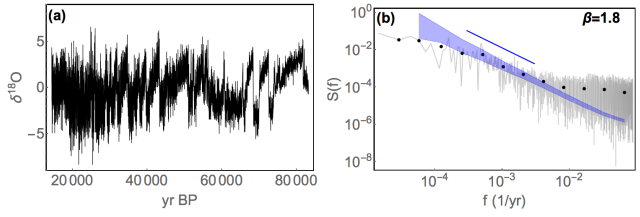

Figure 11. (a): δ^{18}O anomaly time series from the last glacial period of the high-resolution GRIP ice core. (b): Periodogram for the time series in (a).

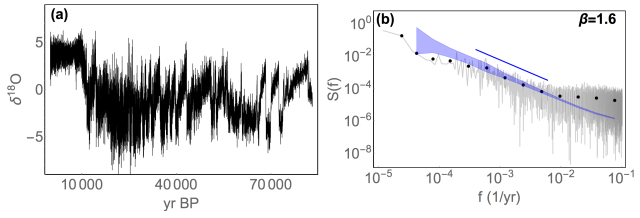

Figure 12. (a): δ^{18}O anomalies from the past 85 kyr of the high-resolution GRIP ice core. (b): Periodogram for the same time series.

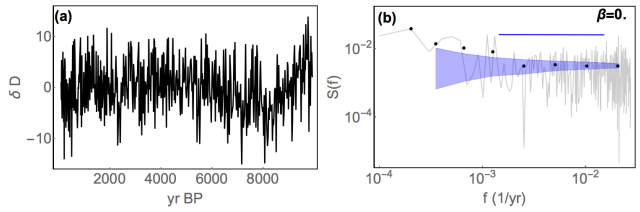

Figure 13. (a): δD anomalies from the Holocene part of the EPICA ice core. (b): Periodogram.

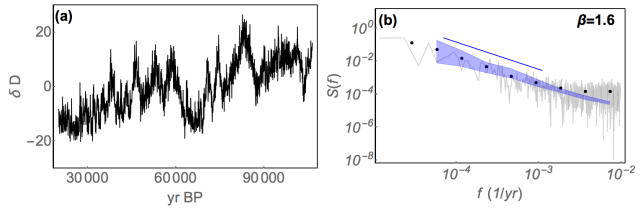

Figure 14. (a): δD anomalies from the last glacial period part of the EPICA ice core. (b): Periodogram for the time series in (a).

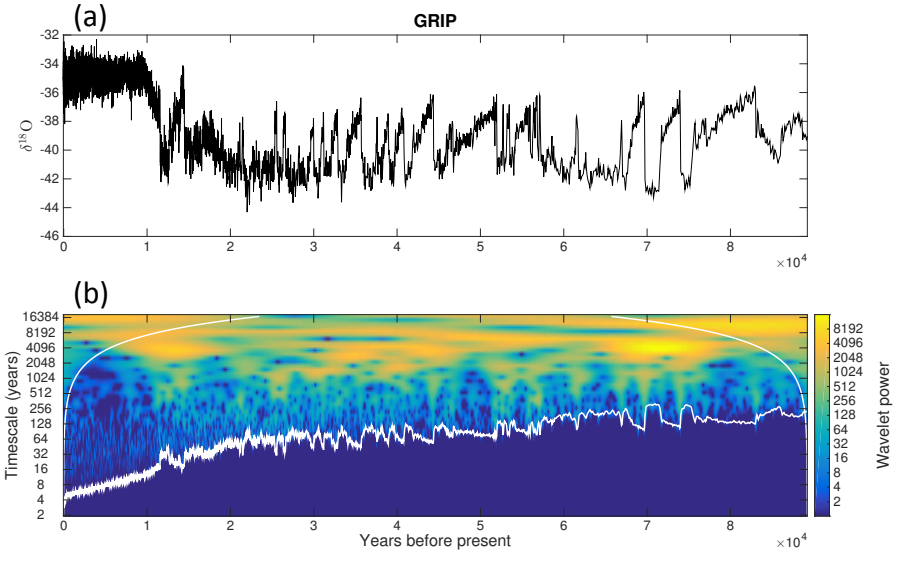

Figure 15. Upper panel: The δ^{18}O proxy time series for Greenland temperature from the GRIP ice core for the period 0 – 90 kyr BP. Lower panel: The Morlet wavelet scalogram for the signal in the upper panel.

Manuscript prepared for Earth Syst. Dynam.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 27 October 2015

Supplementary material to: Are there multiple scaling regimes in Holocene temperature records?

Tine Nilsen[1], Kristoffer Rypdal[1], and Hege-Beate Fredriksen[1]

[1]Department of Mathematics and Statistics, University of Tromsø The Arctic University of Norway, Tromsø, Norway

1 The Lomb-Scargle periodogram as an alternative spectral estimator for the unevenly sampled ice core data

While the standard periodogram is based on Fourier analysis (Schuster, 1898), the LSP is based on a least-squares fit of sinusoids to the data. We use the R package *lomb*, and the power is normalized by multiplying by σ^2/N, where σ^2 is the variance. For time series of N data points $Y_j = Y(t_j)$ collected at times t_j where $j = 1, 2, ...N$, with mean value $\bar{Y}$, the Lomb-Scargle periodogram is defined as,

$$S_N(\omega) = \frac{1}{2\sigma^2}\left\{\frac{\left[\sum_j (Y_j - \bar{Y})\cos\omega(t_j - \tau)\right]^2}{\sum_j \cos^2\omega(t_j - \tau)} + \frac{\left[\sum_j (Y_j - \bar{Y})\sin\omega(t_j - \tau)\right]^2}{\sum_j \sin^2\omega(t_j - \tau)}\right\}, \quad (1)$$

with

$$\tau = \left(\frac{1}{2\omega}\right)\tan^{-1}\left[\frac{\sum_j \sin 2\omega t_j}{\sum_j \cos 2\omega t_j}\right] \quad (2)$$

1.1 Test of the Lomb-Scargle periodogram - surrogate data mimicking ice core proxy data

The Lomb-Scargleperiodogram (LSP) was orginially designed to detect a periodic signal hidden in noise, and has been applied to periodic data with random missing values such as astronomical observations (Lomb, 1976; Scargle, 1982), biological rhythms (Ruf, 1999; Van Dongen et al., 1999), and heart-rate signals (Laguna et al., 1998). In the proxy-based climatic time series we don't expect perfect periodicities, and data points are not randomly missing. The sampling interval rather increases systematically as one goes backward in time. It is therefore not a priori obvious that this method can give confident estimates of the scaling exponent for the ice core time series. Pelletier (1998) used the Lomb periodogram to estimate the scaling exponent for temperature inferred from the Vostok ice core, but did not discuss the sensitivity of the method.

We perform such a test on ensembles of synthetic fGn's and fBm's, respectively. For each realization in the ensemble, the scaling exponent is estimated from the LSP. The mean exponent value as well as the error based on the 2.5% and 97.5% quantiles are estimated. The LSP gives the same results as the ordinary periodogram when the sampling is even. The method is tested by removing data points from the synthetic data sets and repeat the estimation for each realization. Data is removed in such a manner that the sampling intervals become identical to those we find in ice core data. We have chosen a section of the low-resolution GRIP δ^{18}O time series covering the Holocene, and another one from the last glacial period as the basis for this test (see the main article, Sect. 4.3 for data description). The test is carried out by first generating surrogate data sets where the time step is chosen to be the least time step in the observed record. The surrogate time series is then interpolated, and we sample this interpolation function at the times known from the ice core time series. The "resampled" time series has the same number of data points as the GRIP Holocene/last glacial period time series, and β is estimated in the range $\frac{1}{20}$ - $\frac{1}{333}$ yr^{-1} for the Holocene, and $\frac{1}{200}$ - $\frac{1}{3333}$ yr^{-1} for the last glacial period.

Fig. 1 shows the result when surrogate data are sampled like the low-resolution GRIP Holocene (Fig. 1a,c) and low-resolution GRIP last glacial period (Fig. 1b,d). In Fig. 1a,b, observe a slight negative bias for β, caused by the LSP. The difference between the upper and lower panels is due to an important detail when generating our surrogate data. In the top panel, our surrogate data are sampled at irregular time steps, but as point measurements. In reality, the given values of stable isotopes from ice cores are generally averages over the time period covered by the sliced equal-length samples. The real proxy values will therefore exhibit less variability than if they were instantaneous measurements of the true temperature. This smoothing effect will increase as one goes back in time due to compression of the core. The smoothing effect constitutes a positive bias that works in the opposite direction of the negative bias introduced by the LSP. Taking this into account we have created surrogate data (Fig. 1c,d) that resembles proxy data to a larger degree than in the topmost panel.

Because the least time step between values is different in the GRIP Holocene and last glacial period, the lengths of the synthetic data sets are also different. This leads to different error bars for the estimates. From Fig. 1c,d we observe that the estimated β is close to the true β, except for true β slightly greater than unity. The negative bias observed for β slightly higher than unity is expected, and is present also in the standard periodogram. It is due to a spectral feature arising when the (continuous-time) fBm is sampled at discrete times: the spectrum flattens when the Nyquist frequency is approached for β slightly above unity, and leads to an underestimation of β if these frequencies are used in the fitting of a straight line in the log-log plot. Bias and errors are similar to the standard periodogram and shows that the LSP is a very useful substitution for the periodogram for these data records.

Another test of the LSP is also presented for data with constant sampling intervals where data points are missing randomly. This test is more general, since such data are not dealt with in the paper.

1.2 Test of the Lomb-Scargle periodogram - random removal of data

Systematically we remove 25%, 50% and 75% of the data points from Monte Carlo ensembles of synthetic fGn's and fBm's. Fig. 2 shows the estimated scaling exponent vs. the true exponent for the even sampling case, and when 25%, 50% and 75% of the data points are removed.

In Fig. 2b,c,d we observe that when more data is removed, β is gradually more underestimated. This can be explained from Fig. 3 which shows increasing power at the high frequencies when data is removed. The LSP method fits sinusoids to the data on different frequencies, and the low frequency variability is mostly unaffected by the removal of data. For the high frequencies on the other hand, removal of data allows good fit of sinusoids with larger amplitude.

Fig. 2 demonstrates that serious underestimation of β can result from applying the LSP to time series with randomly missing data, if the true β is larger than unity.

1.3 Spectral analysis of ice core data using the Lomb-Scargle periodogram

As a supplement to the spectral analysis in the main article, the same data are here analysed using the LSP. The resulting spectra and scale-break features are similar as the results when interpolation and the standard periodogram are used.

For the GRIP high resolution ice core, fig 4 shows the δ^{18}O time series and corresponding LSP for the Holocene. The topmost figures are for the full Holocene time period, where $\beta \approx 0.3$ on time scales up to 10^3 yr. The lower panel is for the most recent 7.5 kyr of the Holocene. The spectrum is essentially flat for the same time period.

Fig 5 shows the δ^{18}O time series for the last glacial period of the high-resolution GRIP ice core, and the LSP for the same time period. $\beta \approx$1.4 on time scales longer than 10^2 yr and shorter than 10^4 yr. Fig 6 shows the δ^{18}O time series for the past 85 kyr of the high-resolution GRIP ice core, and the LSP for the same time period. The estimated value of β is the same as in fig 5 on the same time scales.

For the EPICA ice core, fig 7 shows the δD time series and corresponding LSP for the Holocene. $\beta \approx$0.2 on time scales up to millenia. For the last glacial period of EPICA, fig 8 shows the δD time series and corresponding LSP. $\beta \approx$1.8 on time scales longer than 10^3 yr and shorter than 10^4 yr.

2 Analysis of additional four deep ice core time series

To support our results in the main article, we present here the analyses of additional four deep ice cores from Greenland and Antarctica. The Greenland ice cores are GISP2, (Grootes et al., 1993;

Grootes and Stuiver, 1997) and NGRIP, (North Greenland Ice-Core project (NGRIP) members, 2004; Andersen et al., 2006; Vinther et al., 2006; Rasmussen et al., 2006; Svensson et al., 2008; Wolff et al., 2010). From Antarctica we include the Taylor dome ice core, (Steig et al., 2000), and Vostok, (Lorius et al., 1985; Jouzel et al., 1987, 1993). We include only results using interpolation and the standard periodogram, and not the LSP.

2.1 Results from the GISP2 and NGRIP ice cores

Figure 9 shows (a) the δ^{18}O time series of the Holocene part of the GISP2 ice core, and (b) the periodogram for the same time series. No scale break is detected on centennial time scales, and the estimated $\beta \approx 0.2$. (c) shows the δ^{18}O time series of the Holocene part of the NGRIP ice core, and (d) the periodogram for the same time series, where we estimate $\beta \approx 0.35$ and no scale break at centennial time scales.

Figure 10 displays (a) The δ^{18}O time series for the GISP2 ice core from the last glacial period, and (b) the periodogram for the same time series, where we estimate $\beta \approx 1.1$. (c) shows the δ^{18}O time series for the NGRIP ice core from the last glacial period, and (d) the periodogram for the same time series with estimated $\beta \approx 1.4$.

2.2 Results from the Taylor and Vostok ice core

Figure 11 shows (a) the Holocene time series of the Taylor ice core, and (b) the periodogram for the same time series where we estimate $\beta \approx 0.2$ for time scales shorter than 10^3 yr. (c) shows the Holocene time series from the Vostok ice core, and (d) the periodogram for the same time series. Due to the poor temporal resolution of the Vostok paleotemperature time series, the periodogram has few data points and the uncertanties are therefore larger for this time series than others studied. Further, the periodogram is affected by the smoothing of the δD time series. This is observed as an abrupt decrease in power at the highest frequencies, and a linear fit should be avoided in this area. We obtain $\beta \approx 0.1$ from the periodogram, but note that this time series can not be used to infer the scaling at centennial time scales.

Figure 12 shows (a) The δ^{18}O time series for the Taylor ice core from the last glacial period, and (b) the periodogram for the same time period where we estimate $\beta \approx 1.9$. (c) shows the δD time series for the Vostok ice core from the last glacial period, and (d) the periodogram for the same time period with estimated $\beta \approx 1.7$.

References

Andersen, K. K., Svensson, A., Johnsen, S. J., Rasmussen, S. O., Bigler, M., Röthlisberger, R., Ruth, U., Siggaard-Andersen, M.-L., Steffensen, J. P., Dahl-Jensen, D., Vinther, B. M., and Clausen, H. B.: The Greenland Ice Core Chronology 2005, 15-42 ka. Part 1: constructing the time scale, Quaternary Sci. Rev., 25, 3246–3257, Shackleton special issue 24, 2006.

Grootes, P. M. and Stuiver, M.: Oxygen 18/16 variability in Greenland snow and ice with 10^3 to 10^5-year time resolution, Journal of Geophysical Research: Oceans, 102, 26 455–26 470, doi:10.1029/97JC00880, 1997.

Grootes, P. M., Stuiver, M., White, J. W. C., Johnsen, S., and Jouzel, J.: Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores, Nature, 366, 552–554, doi:10.1038/366552a0, 1993.

Jouzel, J., Lorius, C., Petit, J. R., Genthon, C., Barkov, N. I., Kotlyakov, V. M., and Petrov, V. M.: Vostok ice core: a continuous isotope temperature record over the last climatic cycle (160,000 years), Nature, 329, 403–408, doi:10.1038/329403a0, 1987.

Jouzel, J., Barkov, N. I., Barnola, J. M., Bender, M., Chappellaz, J., Genthon, C., Kotlyakov, V. M., Lipenkov, V., Lorius, C., Petit, J. R., Raynaud, D., Raisbeck, G., Ritz, C., Sowers, T., Stievenard, M., Yiou, F., and Yiou, P.: Extending the Vostok ice-core record of palaeoclimate to the penultimate glacial period, Nature, 364, 407–412, doi:10.1038/364407a0, 1993.

Laguna, P., Moody, G. B., and Mark, R. G.: Power Spectral Density of Unevenly Sampled Data by Least-Square Analysis: Performance and Application to Heart Rate Signals, IEEE T. BIO-MED. ENG., 45, 698–715, doi:10.1109/10.678605, 1998.

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

Lorius, C., Jouzel, J., Ritz, C., Merlivat, L., Barkov, N. I., Korotkevich, Y. S., and Kotlyakov, V. M.: A 150,000-year climatic record from Antarctic ice, Nature, 316, 591–596, doi:10.1038/316591a0, 1985.

North Greenland Ice-Core project (NGRIP) members: High-resolution record of Northern Hemisphere climate extending into the last interglacial period, Nature, 431, 147–151, doi:10.1038/nature02805, 2004.

Pelletier, J. D.: The power spectral density of atmosphereic temperature from time scales of 10^{-2} to 10^6 yr., Earth and Planetary Science Letters, 158, 157–164, doi:10.1016/S0012-821X(98)00051-X, 1998.

Rasmussen, S. O., Andersen, K. K., Svensson, A. M., Steffensen, J. P., Vinther, B. M., Clausen, H. B., Siggaard-Andersen, M.-L., Johnsen, S. J., Larsen, L. B., Dahl-Jensen, D., Bigler, M., Röthlisberger., R., Fischer, H., Goto-Azuma, K., Hansson, M. E., and Ruth, U.: A new Greenland ice core chronology for the last glacial termination, J. Geophys. Res-Atmos., 111, doi:10.1029/2005JD006079, 2006.

Ruf, T.: The Lomb-Scargle Periodogram in Biological Rhythm Research: Analysis of Incomplete and Unequally Spaced Time-Series, Biol. Rhythm Res., 30, 178–201, doi:10.1076/brhm.30.2.178.1422, 1999.

Scargle, J. D.: Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly sampled data, Astrophys. J., 263, 835–853, doi:10.1086/160554, 1982.

Schuster, A.: On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena., Terrestrial Magnetism, 3, 13–41, doi:10.1029/TM003i001p00013, 1898.

Steig, E. J., Morse, D. L., Waddington, E. D., Stuiver, M., Grootes, P. M., Mayewski, P. A., Twickler, M. S., and Whitlow, S. I.: Wisconsinan and Holocene Climate History from an Ice Core at Taylor Dome, Western Ross Embayment, Antarctica, Geogr. Ann. A., 82, 213–235, 2000.

Svensson, A., Andersen, K., Bigler, M., Clausen, H., Dahl-Jensen, D., Davies, S., Johnsen, S., Muscheler, R., Rasmussen, S., Röthlisberger, R., Seierstad, I., Steffensen, J., and Vinther, B.: A 60 000 year Greenland stratigraphic ice core chronology, Clim. Past, 4, 47–57, doi:10.5194/cp-4-47-2008, 2008.

Van Dongen, H. P. A., Olofson, E., VanHartevelt, J. H., and Kruyt, E. W.: Searching for Biological Rhythms: Peak Detection in the Periodogram of Unequally Spaced Data, J. Biol. Rhythm, 14, 617–620, doi:10.1177/074873099129000984, 1999.

Vinther, B. M., Clausen, H. B., Johnsen, S. J., Rasmussen, S. O., Andersen, K. K., Buchardt, S. L., Dahl-Jensen, D., Seierstad, I. K., Siggaard-Andersen, M.-L., Steffensen., J. P., Svensson, A., Olsen, J., and Heinemeier, J.: A synchronized dating of three Greenland ice cores throughout the Holocene, J. Geophys. Res-Atmos., 111, D13 102, doi:10.1029/2005JD006921, 2006.

Wolff, E. W., Chappellaz, J., Blunier, T., Rasmussen, S., and Svensson, A.: Millennial-scale variability during the last glacial: The ice core record, Quaternary Sci. Rev., 29, 2828–2838, doi:10.1016/j.quascirev.2009.10.013, 2010.

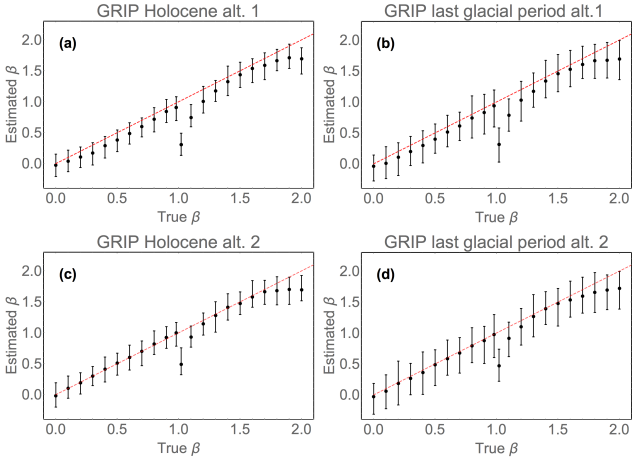

Figure 1. Value of estimated β vs. true β from 100 realizations of synthetic LRM processes. (a) Temporal sampling equal to a point measurement from the low-resolution GRIP Holocene time series. (b) Temporal sampling equal to a point measurement of the low-resolution GRIP last glacial time series. (c,d) As in a,b, respectively, but with data points corresponding to temporal averages.

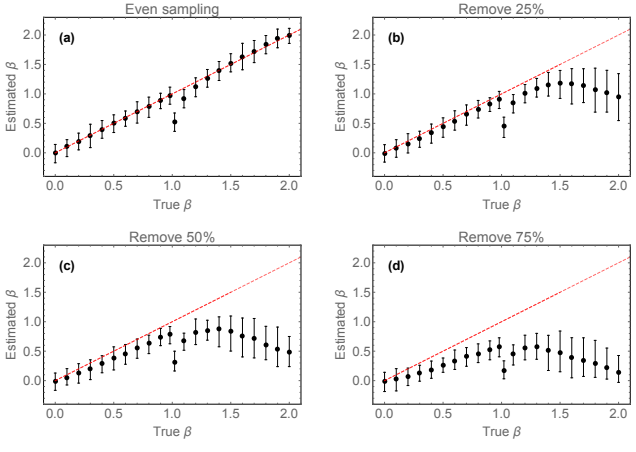

Figure 2. Value of estimated β vs. true β from 100 realizations of synthetic data. (a) Evenly sampled time series, 12576 data points. (b) 25% removed. (c) 50% removed. (d) 75% removed.

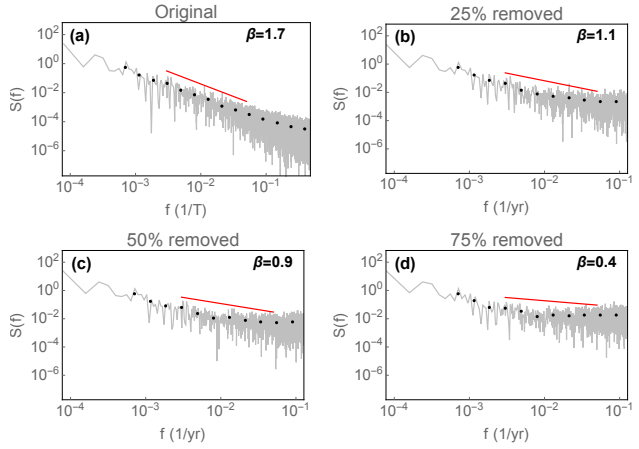

Figure 3. LSP for an fBm with β=1.6, with 12 576 data points. (a) Original time series. (b) 25% removed. (c) 50% removed. (d) 75% removed.

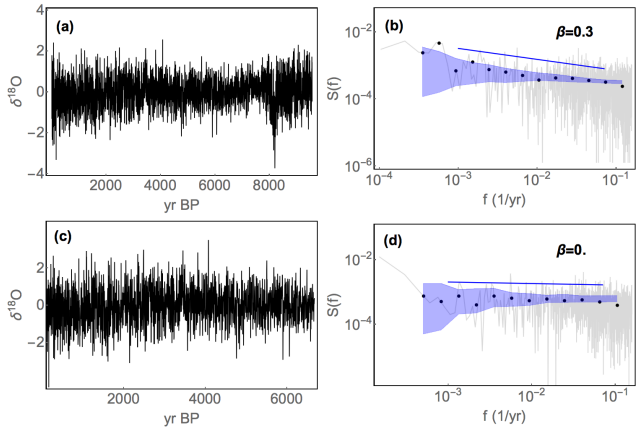

Figure 4. (a): δ^{18}O anomalies from the Holocene part of the high-resolution GRIP ice core. (b): Lomb-Scargle periodogram. The raw LSP is shown in gray, the log-binned version by black dots. β is estimated from the log-binned LSP in the region marked by the blue line. The confidence range is shown by the blue, shaded area, estimated from a Monte Carlo ensemble of synthetic fGns with the estimated value of β and variance from the log-binned LSP. (c) Same figure as in (a) except the oldest section has been removed. (d) Lomb-Scargle periodogram for the time series in (c).

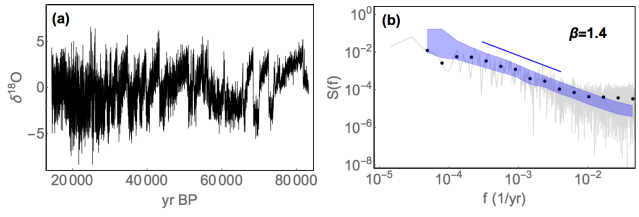

Figure 5. (a): δ^{18}O anomalies from the last glacial period of the high-resolution GRIP ice core. (b): Lomb-Scargle periodogram.

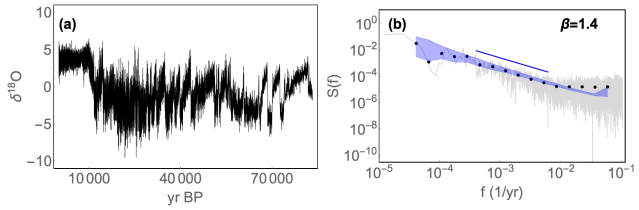

Figure 6. a): δ^{18}O anomalies from the past 85 kyr of the high-resolution GRIP ice core. (b): Lomb-Scargle periodogram

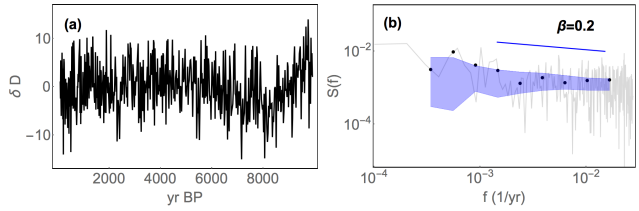

Figure 7. a): δD anomalies from the Holocene part of the EPICA ice core. (b): Lomb-Scargle periodogram

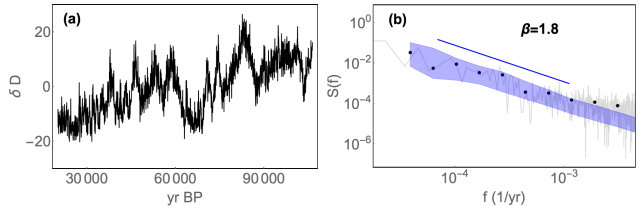

Figure 8. a): δD anomalies from the last glacial period of the EPICA ice core. (b): Lomb-Scargle periodogram

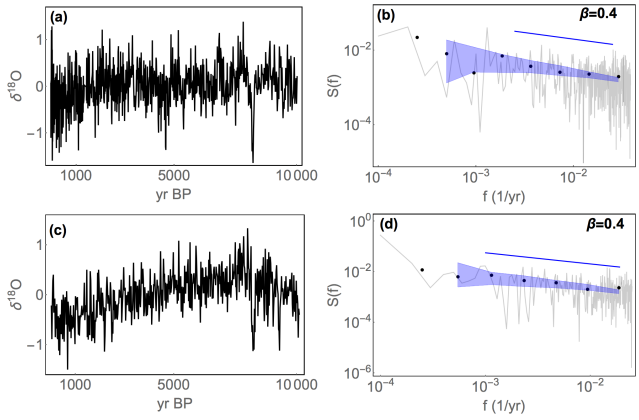

Figure 9. (a): δ^{18}O from the Holocene part of the GISP2 ice core. (b): periodogram for the time series in (a). (c) δ^{18}O from the Holocene part of the NGRIP ice core. (d): periodogram for the time series in (c). In (b) and (d) the raw periodogram is shown in gray and the log-binned version by black dots. β is estimated from the log-binned periodogram in the region marked by the blue line. The confidence range is shown by the shaded area, estimated from a Monte Carlo ensemble of synthetic fGn with the value of β found from the log-binned periodogram.

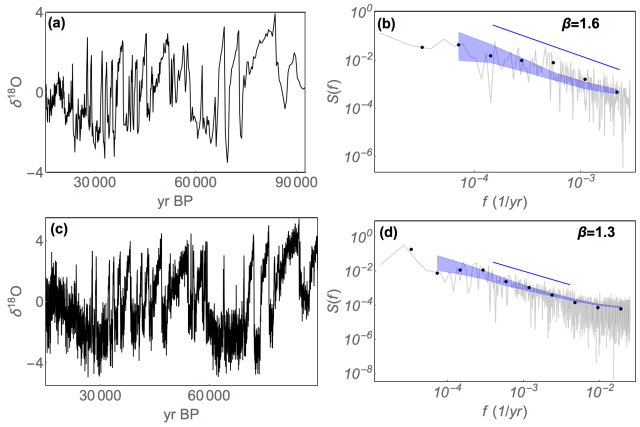

Figure 10. (a):^{18}O from the last glacial period part of the GISP2 ice core. (b): periodogram for the time series in (a). (c): δ^{18}O time series from the last glacial period of the NGRIP ice core. (d): periodogram for the time series in (c).

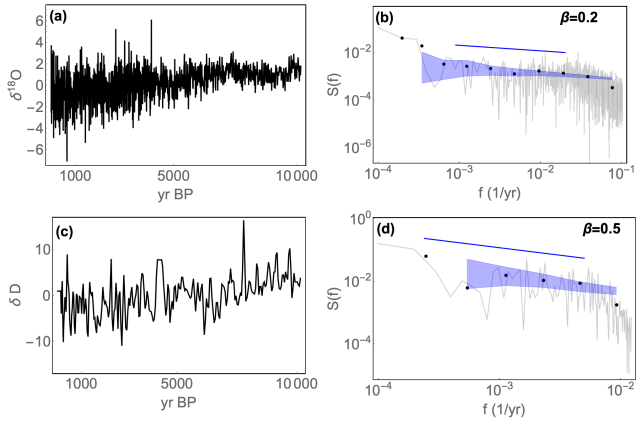

Figure 11. (a): δ^{18}O from the Holocene part of the Taylor ice core. (b): periodogram for the time series in (a). (c) δD from the Holocene part of the Vostok ice core. (d): periodogram for the time series in (c).

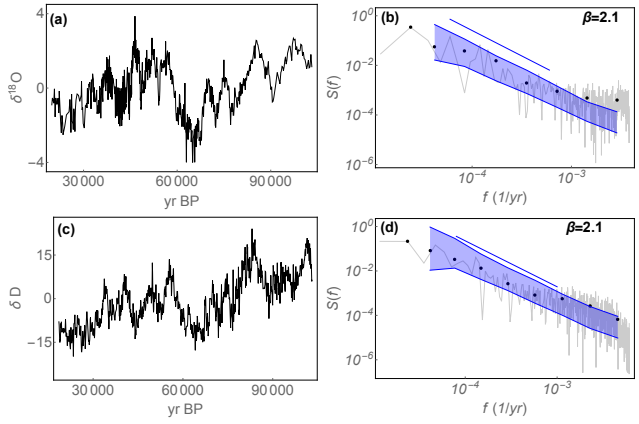

Figure 12. a): δ^{18}O from the last glacial period part of the Taylor ice core. (b): periodogram for the time series in (a). (c):δD time series from the last glacial period of the Vostok ice core. (d): periodogram for the time series in (c).