

# Late Quaternary temperature variability described as abrupt transitions on a $1/f$ noise background

Martin Rypdal<sup>1</sup> and Kristoffer Rypdal<sup>1</sup>

<sup>1</sup>Department of Mathematics and Statistics, University of Tromsø The Arctic University of Norway, Tromsø, Norway

Correspondence to: Martin Rypdal (martin.rypdal@uit.no)

**Abstract.** In order to have a scaling description of the climate system that is not inherently non-stationary, the rapid shifts between stadials and interstadials during the last glaciation (the Dansgaard-Oeschger events) cannot be included in the scaling law. The same is true for the shifts between the glacial and interglacial states in the quaternary climate. When these events are omitted from a scaling analysis the climate noise is consistent with a  $1/f$  law on time scales from months to  $10^5$  years. If the records analysed include the shift events, the effect is to create a break the scaling from a  $1/f$  law to a  $1/f^\beta$  law, with  $1 < \beta < 2$ . No evidence of multifractal intermittency has been found in any of the temperature records investigated, and the events are not a natural consequence of multifractal scaling.

## 1 Introduction

The temporal variations in Earth's surface temperature are well described as *scaling* on an extended range of time scales. In this parsimonious characterisation, a parameter  $\beta$  describes how the fluctuation levels on the different time scales are related to each other. The  $\beta$ -parameter can be defined via the scaling of the spectral density function of the signal by the relation

$$S(f) = \langle |\tilde{T}(f)|^2 \rangle \sim f^{-\beta}, \quad (1)$$

where  $\tilde{T}(f)$  is the Fourier transform of the time record  $T(t)$  and  $\langle \dots \rangle$  denotes an ensemble average. An alternative is to measure the range of the variability on the longest time scales within a time window of length  $\Delta t$  by

$$T_{\Delta t}(t) = \left| \frac{2}{\Delta t} \sum_{i=t}^{t+\Delta t/2} T(t) - \frac{2}{\Delta t} \sum_{i=t+\Delta t/2}^{t+\Delta t} T(t) \right|, \quad (2)$$

and to define  $\beta$  via the following relation (Lovejoy and Schertzer, 2013):

$$\langle |T_{\Delta t}(t)|^2 \rangle \sim \Delta t^{\beta-1}. \quad (3)$$

In this description, the temperature fluctuations would decrease with scale if  $\beta < 1$ , implying that the climate fluctuations become less prominent as we consider longer time scales, a picture which

is somewhat different from the rich long-range variability indicated by proxy reconstructions of past climate. On the other hand, a value  $\beta > 0$  would imply that variability increases with scale, a property that (if it were valid on a large range of time scales) would lead to levels of temperature variability inconsistent with reality. It is therefore a natural *a priori* working hypothesis, that Earth’s typical temperature fluctuations, the climate noise, is characterised by  $\beta \simeq 1$ . Such a process is called a  $1/f$  noise.

The  $1/f$  description of Earth’s temperature is of course an idealised model. The reality is that the climate system consists of many components that respond to perturbations on different characteristic time scales, and the temperature signal can be seen as an aggregation of signals with different time-scale characteristics. Since it is difficult to recognise pronounced time scales in the temperature records, a scaling description is both convenient and accurate. However, we are aware that the scaling is not perfect, and that there are structures in the climate system that deviate from the scaling law. One example is the El Niño Southern Oscillation (ENSO), which places larger fluctuations on the times scales of a few years than what can be expected from a scaling model. Other examples are the Dansgaard-Oeschger (DO) cycles in the Greenland climate during the last glacial period, encompassing repeated and rapid shifts between a cold stadial state and a much warmer interstadial state. The result of this phenomenon is that the glacial climate in Greenland has much larger millennial-scale fluctuations than what can be expected from a  $1/f$  description. However, as we demonstrate in this paper, the temperature variations of both the stadial and interstadial climate states fit well with the  $1/f$ -scaling, telling us that the deviation from  $1/f$  scaling in the glacial climate arise from these regime shifting events. As we go to even longer time scales, we also observe anomalous fluctuation levels on time scales from  $10^4$  to  $10^5$  years that can be identified with the shifting between glacial and interglacial conditions.

One could argue that the DO cycles and the glaciation cycles are intrinsic to the climate system and should not be treated as special events, and their variations should be reflected in a scaling description of the climate. This idea was forwarded by Lovejoy and Schertzer (1986), elaborated in many later papers, and expanded to time scales up to almost a Gyr in Lovejoy (2014). Here several scaling regimes are proposed, including a “break” in the scaling law with an exponent  $\beta \approx 1.8$  on time scales longer than a century. A scaling model invoking two scaling regimes can account for the millennial-scale temperature fluctuations that are produced by the DO cycles, which are anomalous with respect to a  $1/f$  model. However, the estimated scaling exponent will depend on the average “density” of DO events in the ice-core record used for the estimate, and since the events are not uniformly distributed over time, there is no uniquely defined scaling exponent for the last glacial period. Moreover, the scaling law would not be useful as a climate-noise model to use as a null hypothesis for determining the significance of particular trends and events, such as the anthropogenic warming over the last centuries.

The main message of this paper is that the  $1/f$  noise characterisation of the temporal fluctuations in global mean surface temperature is very robust. It is an accurate description for the Holocene climate, but it is also valid under both stadial and interstadial conditions during glaciations, and during both glacial and interglacial conditions in the quaternary climate. The  $1/f$  character of the climate noise provides us with robust estimates of future natural climate variability, even in present state of global warming. Such an estimate would of course be invalidated by a future regime shift (a tipping point) to a warmer climate state provoked by anthropogenic forcing. A future observed change in the  $1/f$  character of the noise could therefore be taken as an early warning signal for such a shift.

## 2 Data, methods and results

The analysis in this work is based on four data sets for temperature fluctuations: the HadCRUT4 monthly global mean surface temperature (Morice et al., 2012) in the period 1880-2011 CE (Common Era), the Moberg Northern Hemisphere reconstruction for annual mean temperatures in the years 1-1978 CE (Moberg et al., 2005), as well as temperature reconstructions from the North Greenland Ice Core Project (NGRIP) (Andersen et al., 2004) and the European Project for Ice Coring in Antarctica (EPICA) (Augustin et al., 2004). For the NGRIP ice core we have used 20-yr means of  $\delta^{18}\text{O}$  going back 60 kyr. For the EPICA ice core we have temperature reconstructions going back over 300 kyr, but the data is sampled at uneven time intervals and the time between subsequent data points becomes very large as we go back more than 200 kyr. In addition we have used annual data for radiative forcing in the time period 1880-2011 CE (Hansen, 2005) to remove the anthropogenic component in HadCRUT4 data. Plots of all four data records are shown in Fig. 1.

### 2.1 Global versus local scaling

On the face of it, it is difficult to discern scaling laws for the climate noise on time scales longer than millennia, since we do not have high-resolution global (or hemispheric) temperature reconstructions for time periods longer than two kyr. The ice core data available only allow us to reconstruct temperatures locally in Greenland and Antarctica, and we know from the instrumental record that local and regional continental temperatures scale differently from the global mean surface temperatures on time scales shorter than millennial. The differences we find are that local temperature scaling exponents  $\beta_l$  are smaller than global temperature exponents  $\beta_g$ , and that the ocean temperatures scale with higher exponents than land temperatures. Since there are strong spatial correlations in the climate system, it is possible that all local temperatures are scaling with a lower exponent than the global. In (Rypdal et al., 2015) this phenomenon is illustrated in an explicit stochastic spatio-temporal model. In this model, which is fitted to observational instrumental data, we find the relationship  $\beta_g = 2\beta_l$ . This relationship is derived under the highly inaccurate assumption that all local temperatures scale

with the same exponent, but it is still a useful approximation in the following, where we will argue  
 95 that we can use local and regional temperature records to discern the scaling of the global mean  
 surface temperature on time scales of 10 kyr and longer. We do this by showing that the assumption  
 that  $\beta_g/\beta_l > 1$  is valid on very long times scales leads to the impossible result that the variance of  
 global averages becomes larger than the mean variance of local averages. Thus we conclude that  $\beta_l$   
 converges to  $\beta_g$  on sufficiently long time scale, and we estimate an upper limit for that time scale.

Let us denote by  $\sigma_g$  and  $\sigma_l$  the standard deviations of the global surface temperature and a local  
 temperature respectively, on a monthly time scale. From Eq. (3) it follows that the ratio between the  
 variances for the global and local temperatures at time scale  $\Delta t$  is

$$\rho = \left(\frac{\sigma_g}{\sigma_l}\right)^2 \left(\frac{\Delta t}{\tau}\right)^{\beta_g - \beta_l},$$

where  $\tau = 1$  month. Unless we expect global temperatures to have larger variations than the local  
 temperature at time scale  $\Delta t$  (the global temperature can not have a larger standard deviation than  
 the average standard deviation of the local temperatures) we must have  $\rho > 1$ , or equivalently,

$$\Delta t < \tau \left(\frac{\sigma_l}{\sigma_g}\right)^{2/(\beta_g - \beta_l)}.$$

100 On the time scale of months, the fluctuation levels of local continental temperatures is about two  
 orders of magnitude larger than the fluctuation level for the global mean temperature. If we also use  
 $\beta_g = 1$  and  $\beta_l = 1/2$  we obtain the condition  $\Delta t < 10^5$  months  $\sim 10$  kyr, i.e., on time scales longer  
 than 10 kyr the ratio  $\beta_g/\beta_l$  can no longer be larger than unity. A similar estimate can be obtained  
 from the NGRIP ice core data. In the Holocene the 20-yr resolution temperature reconstructions from  
 105 Greenland has a standard deviation which is about five times greater than the 20-yr moving average  
 of the Moberg reconstruction for the Northern hemisphere. Applying the same argument restricts  
 the time scale for which Greenland scaling exponent is smaller than the global scaling exponent to  
 approximately 10 kyr.

Based on the reasoning above, we expect scaling of the ice core data to be similar to the global  
 110 scaling on sufficiently long time scales. In the remainder of this paper we demonstrate that the scaling  
 in the ice core data on time scales up to hundreds of kyr is similar to the  $1/f$  scaling we observe in  
 global temperature up to a few millennia. This suggests that the  $1/f$  scaling on very long time scales  
 in ice core data is a reflection of the scaling in global temperatures on these scales.

## 2.2 Methods for estimation of scaling

115 We use two methods to analyse the scaling of temperature records. The first is a simple periodogram  
 estimation of the spectral power density  $S(f)$ . This estimator can also be applied to data with un-  
 even time sampling using the Lomb-Scargle method (Lomb, 1976). The other method is to take the  
 wavelet transform of the temperature data:

$$W_{\Delta t}(t) = \frac{1}{\sqrt{\Delta t}} \int T(t') \psi\left(\frac{t-t'}{\Delta t}\right) dt' \quad (4)$$

120 and construct the mean square of the wavelet coefficients; the wavelet variance. This is a standard technique for estimating the scaling exponent  $\beta$  (Malamud and Turcotte, 1999), and it is known that

$$\langle |W_{\Delta t}(t)|^2 \rangle \sim \Delta t^\beta. \quad (5)$$

We choose to use the so-called Haar wavelet

$$125 \quad \psi(t) = \begin{cases} 1 & t \in [0, 1/2) \\ -1 & t \in [1/2, 1) \\ 0 & \text{otherwise} \end{cases},$$

and the integral in Eq. (4) is computed as a sum. With this wavelet we have the relation

$$W_{\Delta t}(t) = T_{\Delta t}(t) \sqrt{\Delta t}, \quad (6)$$

between the wavelet transform and the Haar fluctuation of Eq. (2). The power spectral density and the wavelet variance are equivalent representations of the second order statistics of the time record, one in frequency domain and the other in time domain, and Eqs. (1) and (5) show that they are characterised by the same exponent  $\beta$  if there is scaling of the second moment. By the Wiener-Khinchin theorem, these second-order moments are also equivalent to the autocorrelation function. Hence, scaling in the second-order statistics plays a special rôle, irrespective of the scaling or non-scaling of other moments.

135 The wavelet variance method can be adapted to the case of unevenly sampled data using the method described in Lovejoy (2014). In the present work, we obtain very similar results using the periodogram and the wavelet variance estimators. Claims have been made that higher-order statistics in the form of a multifractal characterisation is an essential part of the statistical description of these data (e.g., Lovejoy and Schertzer (2013), Chapter 11). For this reason we include a brief analysis of higher moments of the data in Sect. 2.4, and discuss their significance in Sect. 2.5.

### 2.3 Results of second-order analysis

In Fig. 2 we show the wavelet fluctuation  $\langle |W(t, \Delta t)|^2 \rangle$  estimated for two different segments of the NGRIP data. Both time series have the same number of data points and both represent time intervals of 8500 years. The differences between the two time series is that one contains DO cycles, whereas the other does not. The estimated wavelet fluctuations and the spectral density scale very differently for the two time series, and this motivates us to separate stadial and interstadial conditions when we analyse the scaling in NGRIP data. This separation is shown in Fig. 1(a), where the red curve represents the  $\delta^{18}\text{O}$  concentration in interstadial periods and the blue curve represents the  $\delta^{18}\text{O}$  concentration in stadial periods. We have followed Svensson et al. (2008) in defining the dates for the onsets of the interstadials and we have defined the start dates for the stadial periods to be just after

the rapid temperature decrease that typically follows the slow cooling in the interstadial periods. In Fig. 3 we show the spectral density function and the wavelet scaling function for the stadial data (red diamonds) and the interstadial periods (purple triangles), which both display an approximate  $1/f$  scaling, but where the fluctuation variance in the stadial data is larger than in the interstadial data.

155 These results are different from what is obtained when considering the NGRIP data (during the last glaciation) as a single time series (shown as blue diamonds). If we were to define a single scaling exponent for the whole time series, then we would obtain an estimate  $\beta \approx 1.4$ .

Fig. 3 shows that the scaling of the stadial and interstadial NGRIP data are similar to the scaling of global temperatures on shorter time scales during the Holocene. We have included an analysis of  
160 the instrumental temperature record both with (green triangles) and without the anthropogenic component (green disks). The anthropogenic component can be removed by subtracting the response to the anthropogenic forcing in a simple linear response model of the type considered in (Rypdal and Rypdal, 2014). We have also included an analysis of the Moberg Northern Hemisphere reconstruction (black squares), and we observe that the composite scaling wavelet variance function and the  
165 composite spectral density function obtained by combining the instrumental data with the Moberg reconstruction, is consistent with a  $1/f$  model on time scales from months to centuries. Since the NGRIP data also shows  $1/f$  scaling, and since we believe that the scaling of the NGRIP data is a reflection of global scaling on time scales longer than a millennium, it is illustrative to adjust the fluctuation levels of the NGRIP data so that its Holocene part has a standard deviation close to that of the  
170 standard deviation of the 20-year means of the Moberg reconstruction in the same time period. This means that we use the adjusted NGRIP data as a proxy for global temperature on millennial scales. The effect of this adjustment is only a vertical shift of the wavelet scaling function and the spectral density functions in the double-logarithmic plots, so that it becomes easier to compare the scaling of the NGRIP data with the Moberg reconstruction and the instrumental data. We do not apply any  
175 adjustments of the fluctuation levels of the stadial and interstadial periods relative to each other. The same adjustment is applied to the EPICA data, and here we also consider the scaling of the glacial and interglacials separately as shown in Fig. 1(b). The scaling estimated from the EPICA data for glacial periods (black crosses in Fig. 3) follows closely the scaling of the NGRIP data analysed as a single time series (blue diamonds). This shows that the glacial climates have similar characteristics  
180 in Greenland and in Antarctica. Careful examination of the figure shows that the fluctuations grow slightly faster with the scale  $\Delta t$  in the NGRIP time series than for the glacial periods of the EPICA time series. This is expected since the regime shifting events in Antarctica associated with the DO cycles are much less pronounced than in Greenland (WAIS Divide Project Members, 2015). In the EPICA data we cannot estimate a scaling exponent for the dynamics in periods without regime shifts,  
185 but our results for the EPICA data are consistent with a description of the climate as a  $1/f$  climate noise plus regime shifts. If we analyse the EPICA data without omitting the interglacials, then the

fluctuations increase even faster with the scale  $\Delta t$  (orange stars in Fig. 3). This effect is completely analogous to the effect of shifting between the stadial and interstadial conditions during glaciations.

## 2.4 A note on multifractal processes

190 The exponent  $\beta$  is well-defined as long as the power spectral density function  $S(f)$  is a power law in  $f$ , or equivalently if the wavelet variance  $\langle |W(t, \Delta t)|^2 \rangle$  is a power law in  $\Delta t$ . As mentioned in Sect. 2.3; if well-defined, the  $\beta$  exponent is related to the temporal correlations in the signal via simple formulas. In fact, for a (zero-mean) stationary process  $T(t)$  with  $-1 < \beta < 1$  we have  $\langle T(t)T(t + \Delta t) \rangle \sim (\beta + 1)\beta\Delta t^{\beta-1}$  and for (a zero-mean) process with stationary increments and  
 195  $1 < \beta < 3$  we have  $\langle \Delta T(t)\Delta T(t + \Delta t) \rangle \sim (\beta - 1)(\beta - 2)\Delta t^{\beta-3}$ , where  $\Delta T(t)$  is the increment process of  $T(t)$  (Rypdal and Rypdal, 2012). Thus, the results presented so far in this paper do not rely on any assumptions of self-similar or multifractal scaling. It is only assumed that the second-order moments  $\langle |W(t, \Delta t)|^2 \rangle$  are well approximated by power-laws over an extended range of time scales.

200 A more complete scaling analysis can be performed if one imposes the more restrictive assumption that the wavelet-based structure functions  $\langle |W(t, \Delta t)|^q \rangle$  are power-laws in  $\Delta t$ , not only for  $q = 2$ , but for an interval of  $q$ -values. The time record can be classified as multifractal only if this is true. It is then possible to define a scaling function  $\eta(q)$  via the relation

$$\langle |W_{\Delta t}(t)|^q \rangle \sim \Delta t^{\eta(q)}. \quad (7)$$

205 Lovejoy and Schertzer (2013) defines a scaling function  $\xi(\Delta t)$  from the moment  $\langle |T_{\Delta t}(t)|^q \rangle$ . From Eq. (5) we observe that their scaling function is related to ours by  $\eta(q) = \xi(q) + q/2$ .

By Eqs. (5) and (7) we observe that  $\eta(2) = \beta$ . If  $T(t)$  is self-similar (or if  $T(t)$  is the increment process of a self-similar process in the case  $\beta < 1$ ) we have  $\eta(q) = \beta q/2$ , but in general, the  $\eta(q)$  may be concave (it bends down). Processes that exhibit power-law structure functions and strictly  
 210 concave scaling functions can be characterised as multifractal intermittent. A monofractal is a special case of the multifractal class. For a monofractal (monoscaling) process, the scaling function is linear in  $q$ .

If Fig. 4 we present a crude multifractal analysis of the data sets considered in this paper using  $q$ -values in the range from 0.1 to 4. For the Holocene we find linear scaling functions for both the  
 215 instrumental record and the Moberg Northern Hemisphere reconstruction, and in the NGRIP data we find linear scaling functions for the stadial periods and the interstadial periods when these are analysed separately, although, as we have already seen, there is a deviation from the  $1/f$  scaling in the stadial periods for time scales shorter than about 200 yrs. If the NGRIP record is analysed with both stadial and interstadial stages included, then it is not clear how to define the scaling function  
 220 since the shifts between the two types of stages causes a “break” in the power-law scaling of the wavelet-based structure functions. If we define  $\eta(q)$  using the time scales shorter than 2 kyr we

obtain a linear scaling function corresponding to  $\beta = 1.14$ , and if we use the time scales longer than 4 kyr we obtain a linear scaling function corresponding to  $\beta = 1.78$ . In neither case do we obtain a strictly concave scaling function. A linear scaling function is also obtained if we disregard the “break” in the scaling and fit power laws using all the available time scales. In this case the scaling function corresponds to  $\beta = 1.26$ . For the periods of the EPICA record that corresponds to ice ages, we find wavelet-based structure functions that are closer to power-laws than what is observed in the NGRIP record. This is expected since the abrupt transitions between cold and warm periods is much less pronounced in Antarctica than in Greenland. The scaling function for the ice-age periods in the EPICA data is linear and corresponds to  $\beta = 1.18$ .

The results discussed above show that from this analysis there is no evidence of multifractal intermittency in the temperature records analysed in this paper. This is not very surprising and could be suspected by direct inspection of the data record. The trained observer would use the fact that if  $\eta(q)$  is strictly concave, then the kurtosis of  $W_{\Delta t}(t)$ ,

$$\frac{\langle |W_{\Delta t}(t)|^4 \rangle}{\langle |W_{\Delta t}(t)|^2 \rangle^2} \sim \Delta t^{\eta(4)-2\eta(2)},$$

is decreasing as a power-law function of  $\Delta t$ , and is therefore leptokurtic<sup>1</sup> on the shorter time scales  $\Delta t$ . Multifractal intermittency in addition implies that the amplitudes of the random fluctuations are clustered in time, on all time scales, as observed in intermittent turbulence or financial time series (see e.g., Bouchaud and Muzy (2003)). These are *not* prominent features in the time series analysed in this paper. For the NGRIP data, the  $\delta^{18}\text{O}$  ratio slightly deviates from a normal distribution as a result of the DO events, but this is not well described by a multifractal model since that would require the wavelet-based structure functions to be power-laws in  $\Delta t$ . In fact, what we show in this paper is that the effect of DO events is to break the scaling, rather than to produce multifractal scaling.

Admittedly this multifractal analysis is a crude first-order characterisation. Our crude analysis suggests that the records analysed are most reasonably modeled as monofractal. However, to establish this with confidence we need to perform statistical hypothesis testing. The strategy for such testing must consist of two elements. First, we have to test whether we can reject the hypothesis that the observed records are realisations of a multifractal (with monofractal as a special case) stochastic process. If this hypothesis can be rejected, there is no point in discussing whether the process is multifractal or monofractal. If we cannot reject the multifractal hypothesis, we must test if we can reject that this multifractal is a monofractal. The outcome of these tests depends on the lengths of the observed records, since rejection of the various null hypotheses depends on the statistical uncertainty associated with realisations of the null models. Monte Carlo simulations of these null models is the simplest tool to establish these uncertainties. In a forthcoming paper we will perform this rigorous testing of the multifractal hypothesis for the data analysed in the present paper, in addition to a wide

<sup>1</sup>A distribution is leptokurtic if it has high kurtosis compared with a normal distribution. This means that the probability density function has a high central peak and fatter tails.



selection of forcing data and climate model data. The results presented here should therefore be taken as preliminary.

## 2.5 A note on non-fractal processes that scale in the second moment

In this paper we have focused on scaling in second-order statistics, or more precisely, on modeling the temperature records as stochastic processes that exhibit scaling of the second moment, but not necessarily of other moments. Reviewer Shaun Lovejoy strongly opposes this approach. He considers it as a return to old quasi-Gaussian ideas that disregards the developments of multifractal formalism and multiplicative cascades, and in his last referee comment he raises doubts about the existence of processes that exhibit scaling in the second moment, but not in other moments. Here we will not only demonstrate the existence of such processes, but explain that the serious fallacy of Lovejoy's approach is that he fails to distinguish between multifractal noises and non-Gaussian noises that cannot be modeled within the multiplicative cascade paradigm. Examples of the latter is the large class of Lévy noises. In less technical terms, the issue is that a multifractal noise may consist of uncorrelated random variables (e.g., their signs may be uncorrelated), but they will never be independent (e.g., their squares will be correlated). A Lévy noise, on the other hand, consists of independent random variables, which implies that all powers of the variables will be uncorrelated. Empirical multifractal analysis methods typically fail to distinguish between these different classes of processes because they implicitly assume a multifractal model. Often, the distinction is not easy to make, because a non-Gaussian Lévy noise may have a bursty (intermittent) appearance, and analysis must be designed to separate multifractal clustering (correlation in higher powers) from intermittency of non-Gaussian independent variables. Long-range memory in the process does not make the distinction less relevant. Such processes may easily be produced from those discussed above by convolving the zero-memory processes with a memory response kernel.

From a physical viewpoint, it is very important to distinguish between these two classes of stochastic processes. The multifractal processes are based on a turbulent cascade paradigm and the dynamical description is fundamentally nonlinear. The Lévy noises, and their long-memory cousins, may arise from non-Gaussian, independent fluctuations on the short time scales, e.g., jumps with randomly distributed waiting times.

We distinguish between a Lévy *noise*  $T(t)$  and a Lévy *process*  $X(t)$ . The latter is a continuous-time stochastic process with stationary, identical and independently distributed (i.i.d.) increments, i.e., for any  $\tau$  the increments  $X(t+\tau) - X(t)$  have a well defined distribution which is independent of  $t$ . The discrete-time process  $T(t) = X(t+1) - X(t)$ , where  $t$  is the set of nonnegative integers, is a Lévy noise. Theoretical results on the fluctuation statistics on varying time scales of Lévy noises are most conveniently obtained by means of the standard structure functions of the underlying Lévy process (rather than the wavelet structure function defined in Sect. 2.4), i.e., we define

$$S_q(\Delta t) \equiv \langle |X(t + \Delta t) - X(t)|^q \rangle. \quad (8)$$

For a process which belongs to the multifractal class we have

$$S_q(\Delta t) \sim \Delta t^{\zeta(q)}, \quad (9)$$

where the scaling function  $\zeta(q)$  is related to  $\eta(q)$  for the wavelet moments and  $\xi(q)$  for the Haar fluctuation by  $\zeta(q) = \eta(q) + q/2 = \xi(q) + q$ . In Appendix A we show that for a Lévy process the following relations hold for the second and fourth moments;

$$S_2(\Delta t) = \langle Y^2 \rangle \Delta t, \quad (10)$$

$$S_4(\Delta t) = 3\langle Y^2 \rangle^2 \left[ \Delta t^2 + \left( \frac{1}{3} \text{kurt}[Y] - 1 \right) \Delta t \right], \quad (11)$$

where  $Y \equiv X(1)$  and  $\text{kurt}[Y] \equiv \langle Y^4 \rangle / \langle Y^2 \rangle^2$  is the kurtosis (flatness) of  $Y$ . For a Gaussian process  $\text{kurt}[Y] = 3$ , and hence  $S_4(\Delta t) \propto \Delta t^2$ . In this case  $S_q(\Delta t) \propto \Delta t^{q/2}$ ,  $X(t)$  is a Wiener process, and  $T(t)$  is a Gaussian white noise. For a non-Gaussian Lévy noise, Eq. (11) provides the key to distinguish it from multifractal noise. For  $\Delta t \sim \text{kurt}[Y]/3 - 1$  there is a break in the scaling of  $S_4(\Delta t)$ . In fact, for  $\Delta t \ll \text{kurt}[Y]/3 - 1$  moments higher than  $q = 2$  will scale more or less like  $\Delta t^1$  (i.e.,  $\zeta(q) \rightarrow 1$  for large  $q$ ), while for  $\Delta t \gg \text{kurt}[Y]/3 - 1$  they will scale like  $\Delta t^{q/2}$ . The latter corresponds to the scaling of a Gaussian white noise, which is quite obvious, since the random variables are independent and the central limit theorem implies that the fluctuations are Gaussian on the long time scales. On the other hand, on the short time scales when the fluctuations are still non-Gaussian, the scaling function  $\zeta(q)$  bends over to become flat for large  $q$ , which is just the behavior we find for multifractals. Hence, by leaving out the scales  $\Delta t > \text{kurt}[Y]/3 - 1$  from the analysis we will be led to the conclusion that the non-Gaussian Lévy process is multifractal. The trace-moment analysis employed by Shaun Lovejoy (Schertzer and Lovejoy, 1987; Lovejoy and Schertzer, 2013) is designed to conceal the scaling behavior on these scales and is not suitable as a model selection test to distinguish multifractals from non-Gaussian Lévy noises or their long-memory derivatives.

In Fig. 5 we present analysis of a synthetic jump-diffusion process, which belongs to the class of Lévy noises. The details of this process is explained in Appendix B. The second-order structure function is a power law (a straight-line in the log-log plot), but the other structure functions are not. If a scaling function is produced by fitting a straight line to the structure functions on the long time scales, and computing the slopes, we find the scaling function of a white Gaussian noise (the red line in Fig. 5d). If the same is done on the short time scales, the scaling function has the concave appearance of a multifractal (the blue curve). In Fig. 6, we show the same for a jump-diffusion process with memory, produced by convolving the Lévy noise with a memory kernel. Hence, the distinction between multifractals and other non-Gaussian processes is not limited to processes of independent random variables.

### 3 Discussion and concluding remarks

Accurate characterization of the climate noise is essential for the detection and evaluation of anthropogenic climate change. For instance, when we apply standard statistical methods for estimating the significance of a temperature trend, the result depends crucially on the so-called error model, i.e., the model for the climate noise that is used as a null hypothesis. There is strong evidence that the temperature fluctuations are better described by scaling models than by so-called red-noise models (or AR(1)-type models). However, simply characterising the climate noise as scaling does not specify an error model. The exponent in the scaling law (the  $\beta$  parameter) must also be determined, and it is usually determined from the same signal as we are testing for trends. If we do that without detrending we risk to estimate a too high  $\beta$  for the error model, which yields a trend test with weak statistical power, i.e., we may fail to detect a trend even if it is present. It is possible to improve the statistical power in a logically consistent way by detrending prior to estimating  $\beta$ , but the approach is often (incorrectly) criticized for being circular, since  $\beta$  should be estimated under the assumption that the error model (null hypothesis) is true. However, since de-trending only has a small effect if the null hypothesis is true, de-trending is valid under both the null hypothesis and the alternative hypothesis.

Another approach, which is the motivation for this paper, is to characterise the scaling of the climate noise from pre-industrial temperature records. If we are to use the scaling exponent estimated from pre-industrial records to demonstrate the anomalous climate event associated anthropogenic influence, we must be confident that the temperature scaling does not change significantly over time. We must also be confident that the scaling is robust, in the sense that it is not too sensitive to moderate changes in the climate state. The results presented in this paper suggest that, unless the climate system experiences dramatic regime shifting events, we can be confident that the natural fluctuations in global surface temperature is approximated by  $1/f$ -type scaling on a large range of time scales. This result makes it easy to determine, on any time scale, if the observed increase in global mean surface temperature is inconsistent with the natural variability, and by how much.

The  $1/f$  scaling described here is the same scaling observed in numerous publications by Shaun Lovejoy and Daniel Schertzer, e.g., Lovejoy et al. (2013); Lovejoy and Schertzer (2013); Lovejoy (2014). The important difference in our interpretations is that they believe this scaling is limited to the scale range from a few months to a century (the “macroweather regime”). Our analysis suggests that this scaling is an expression of Nature’s internal “humming” in Quaternary surface temperature variability on all scales up to hundreds of kyr. A natural conclusion drawn from this interpretation is that description of the observed deviations from this scaling caused by DO events and glacial/interglacial transitions should be sought in dynamical-stochastic models rather than in general scaling laws (Braun et al., 2011; Rypdal, 2016).

## Appendix A: Lévy processes

360 Let  $X(t)$  be a random walk process with stationary and independent increments. This means that the random variables  $X(t+1) - X(t)$  are independent and all have the same distribution. If we normalise so that  $X(0) = 0$  we can write  $X(t+1) - X(t) \stackrel{d}{=} X(1)$ , where  $\stackrel{d}{=}$  denotes equality in distribution for random variables. A random walk of this type is called a Lévy process (see Appelbaum (2004) for a review), and for it to be well-defined in continuous time one must assume that  $X(1)$  belongs  
365 to the class of infinitely divisible random variables. This is a technical requirement that ensures that an increment  $X(t + \Delta t) - X(t)$  has a well-defined distribution for arbitrary small  $\Delta t$ . Since a Lévy process has stationary and independent increments, it is uniquely defined by the (infinitely divisible) distribution of  $X(1)$ . In the following we denote  $Y = X(1)$ .

The characteristic function of the random variable  $Y$  is defined as the  $\phi_Y(u) = \langle e^{iuY} \rangle$ . If  $Y$  has a  
370 probability density function  $p_Y$ , then

$$\phi_Y(u) = \int e^{iuy} p_Y(y) dy$$

is the Fourier transform of  $p_Y$ . When working with Lévy processes it is common to define the function  $\psi(u)$  via the relation  $\phi_Y(u) = e^{\psi(u)}$ , and  $\psi(u)$  is usually called the Lévy exponent. Note that since  $\phi_Y(0) = 1$  we have  $\psi(0) = 0$ . Since  $\psi(u)$  defines the random variable  $Y$  it also determines  
375 the Lévy process uniquely. If  $t$  is an integer the value of  $X(t)$  is a random variable that can be written as a sum of lag-1 increments:

$$X(t) = (X(1) - X(0)) + (X(2) - X(1)) + \dots + (X(t) - X(t-1)).$$

This is a sum of  $t$  independent copies of the random variable  $Y$ , and therefore the characteristic function of  $X(t)$  is

380 
$$\phi_{X(t)}(u) = \langle e^{iuX(t)} \rangle = e^{t\psi(u)}.$$

In general there is a simple relation between the  $n$ 'th moment of a random variable and the  $n$ 'th derivative of its characteristic function evaluated in  $u = 0$ . Using this relation we can express the moments of  $X(t)$  via the formula (Gardiner, 2009),

$$\langle X(t)^n \rangle = i^{-n} \frac{d^n}{du^n} \Big|_{u=0} e^{t\psi(u)}.$$

385 This implies that  $\langle X(t) \rangle = -i\psi'(0)t$ , so if we assume that the process does not have a linear drift, then we must have  $\psi'(0) = 0$ . The second moment is computed the same way:

$$\langle X(t)^2 \rangle = -\frac{d^2}{du^2} \Big|_{u=0} e^{t\psi(u)} = t^2 e^{t\psi(0)} \psi'(0)^2 + t e^{t\psi(0)} \psi''(0) = -\psi''(0)t = \langle Y^2 \rangle t.$$

Since increments are stationary we have the following result for the second order structure function:

$$S_2(\Delta t) = \left\langle |X(t + \Delta t) - X(t)|^2 \right\rangle = \langle X(\Delta t)^2 \rangle = \langle Y^2 \rangle \Delta t.$$

390 This is an important result, because it shows that the second structure function of *any* Lévy processes scales linearly with the time scale  $\Delta t$ . It is also interesting to compute the fourth structure function  $S_4(\Delta t)$ , because for a self-similar process with  $S_2(\Delta t) \propto \Delta t$  one must have  $S_4(\Delta t) \propto \Delta t^2$ . For a Lévy process  $X(t)$  with zero drift we have

$$\begin{aligned}
S_4(\Delta t) &= \langle X(\Delta t)^4 \rangle = \frac{d^4}{du^4} \Big|_{u=0} e^{\Delta t \psi(u)} \\
395 \quad &= \Delta t^4 e^{\Delta t \psi(0)} \psi'(0)^4 + 6\Delta t^3 e^{\Delta t \psi(0)} \psi'(0)^2 \psi''(0) + 3\Delta t^2 e^{\Delta t \psi(0)} \psi''(0)^2 \\
&+ 4\Delta t^2 \psi^{(3)}(0) e^{\Delta t \psi(0)} \psi'(0) + \Delta t \psi^{(4)}(0) e^{\Delta t \psi(0)} \\
&= 3\psi''(0)^2 \Delta t^2 + \psi^{(4)}(0) \Delta t,
\end{aligned}$$

and since  $\psi''(0) = -\langle Y^2 \rangle$  and  $\psi^{(4)}(0) = \langle Y^4 \rangle - 3\langle Y^2 \rangle^2$ , we can write

$$S_4(\Delta t) = 3\langle Y^2 \rangle^2 \Delta t^2 + (\langle Y^4 \rangle - 3\langle Y^2 \rangle^2) \Delta t. \quad (\text{A1})$$

400 The ratio between the second and first terms in the above equation is

$$\frac{\langle Y^4 \rangle - 3\langle Y^2 \rangle^2}{3\langle Y^2 \rangle^2 \Delta t} = \frac{\text{kurt}[Y] - 3}{3\Delta t},$$

and hence we can conclude that  $S_4(\Delta t) \sim \Delta t^2$  for  $\Delta t \gg \text{kurt}[Y] - 3$  and  $S_4(\Delta t) \sim \Delta t$  for  $\Delta t \ll \text{kurt}[Y] - 3$ .

## Appendix B: A Poisson jump process

405 A Poisson jump process is defined via the Lévy exponent

$$\psi(u) = \lambda \int (e^{iux} - 1) dP_J(x),$$

where  $P_J(x)$  is the distribution function for the jumps and  $\lambda$  is the rate for the occurrence of jumps.

For simplicity we can imagine a process where we have positive jumps of fixed size  $x_+$ :

$$dP_J(x) = \delta(x - x_+)$$

410 and

$$\psi(u) = \lambda(e^{iux_+} - 1).$$

Hence the probability density function for  $X(t)$  becomes

$$\begin{aligned}
p_{X(t)}(x) &= \frac{1}{2\pi} \int e^{-iux} e^{t\psi(u)} du = \frac{1}{2\pi} \int e^{-iux} e^{t\lambda(e^{iux_+} - 1)} du \\
&= \frac{e^{-\lambda t}}{2\pi} \int e^{-iux} e^{t\lambda e^{iux_+}} du = \frac{e^{-\lambda t}}{2\pi} \int e^{-iux} \sum_{n=0}^{\infty} \frac{1}{n!} (t\lambda e^{iux_+})^n du \\
415 \quad &= \frac{e^{-\lambda t}}{2\pi} \sum_{n=0}^{\infty} \frac{1}{n!} (\lambda t)^n \int e^{iunx_+} e^{-iux} du \\
&= e^{-\lambda t} \sum_{n=0}^{\infty} \frac{1}{n!} (\lambda t)^n \delta(x - nx_+)
\end{aligned}$$

There is a drift

$$\langle X(t) \rangle = \int x P_{X(t)}(x) dx = x_+ e^{-\lambda t} \sum_{n=0}^{\infty} \frac{1}{n!} (\lambda t)^n n = \lambda t,$$

and

$$\langle |X(t) - \lambda t|^q \rangle = \int |x - \lambda t|^q P_{X(t)}(x) dx = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{1}{n!} (\lambda t)^n |nx_+ - \lambda t|^q.$$

### Appendix C: Lévy processes convolved with memory kernels

Let  $X(t)$  be a Lévy process and for  $\beta \in (1, 3)$ , define a fractional Lévy process by

$$\begin{aligned} Z(t) &= \int_{-\infty}^0 \left( (t-s)^{\frac{\beta}{2}-1} - (-s)^{\frac{\beta}{2}-1} \right) dX(s) + \int_0^t (t-s)^{\frac{\beta}{2}-1} dX(s) \\ 420 \quad &= \int_{-\infty}^{\infty} \left( (t-s)_+^{\frac{\beta}{2}-1} - (-s)_+^{\frac{\beta}{2}-1} \right) dX(s) = \int_{-\infty}^{\infty} K_{\beta}(t, s) dX(s), \end{aligned}$$

where

$$K_{\beta}(t, s) = (t-s)_+^{\frac{\beta}{2}-1} - (-s)_+^{\frac{\beta}{2}-1} = (t-s)^{\frac{\beta}{2}-1} \Theta(t-s) - (-s)^{\frac{\beta}{2}-1} \Theta(-s).$$

Note that

$$K_{\beta}(at, as) = (at-as)^{\frac{\beta}{2}-1} \Theta(at-as) - (-as)^{\frac{\beta}{2}-1} \Theta(-as) = a^{\frac{\beta}{2}-1} K(t, s),$$

425 and since  $\langle dX(t) dX(s) \rangle \propto \delta(t-s) dt$  we have

$$\langle Z(t)^2 \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K_{\beta}(t, s) K_{\beta}(t, s') E[dX(s) dX(s')] = \int_{-\infty}^{\infty} K_{\beta}(t, s)^2 ds.$$

This implies that

$$\langle Z(at)^2 \rangle = \int_{-\infty}^{\infty} K_{\beta}(at, s)^2 ds = \int_{-\infty}^{\infty} K_{\beta}(at, as)^2 a ds = a^{\beta-1} E[Z(t)^2],$$

$$\text{i.e. } \langle Z(\Delta t)^2 \rangle = \langle Z(1)^2 \rangle \Delta t^{\beta-1}.$$

430 *Acknowledgements.* This paper was supported by the the Norwegian Research Council (KLIMAFORSK programme) under grant no. 229754.

## References

- Andersen, K. K., Azuma, N., Barnola, J. M., Bigler, M., Biscaye, P., Caillon, N., Chappellaz, J., Clausen, H. B., Dahl-Jensen, D., Fischer, H., Flückiger, J., Fritzsche, D., Fujii, Y., Goto-Azuma, K., Grønbold, K.,  
 435 Gundestrup, N. S., Hansson, M., Huber, C., Hvidberg, C. S., Johnsen, S. J., Jonsell, U., Jouzel, J., Kipfstuhl, S., Landais, A., Leuenberger, M., Lorrain, R., Masson-Delmotte, V., Miller, H., Motoyama, H., Narita, H., Popp, T., Rasmussen, S. O., Raynaud, D., Röthlisberger, R., Ruth, U., Samyn, D., Schwander, J., Shoji, H., Siggard-Andersen, M. L., Steffensen, J. P., Stocker, T., Sveinbjörnsdóttir, A. E., Svensson, A., Takata, M., Tison, J. L., Thorsteinsson, T., Watanabe, O., Wilhelms, F., and White, J. W. C.: High-resolution record of  
 440 Northern Hemisphere climate extending into the last interglacial period, *Nature*, 431, 147–151, 2004.
- Appelbaum, D.: Lévy processes - from probability to finance and quantum groups, *Notices of the American Mathematical Society*, 51(11), 1336–1347, 2004.
- Augustin, L., Barbante, C., Barnes, P. R. F., Marc Barnola, J., Bigler, M., Castellano, E., Cattani, O., Chappellaz, J., Dahl-Jensen, D., Delmonte, B., Dreyfus, G., Durand, G., Falourd, S., Fischer, H., Flückiger, J., Hansson, M. E., Huybrechts, P., Jugie, G., Johnsen, S. J., Jouzel, J., Kaufmann, P., Kipfstuhl, J., Lambert, F., Lipenkov, V. Y., Littot, G. C., Longinelli, A., Lorrain, R., Maggi, V., Masson-Delmotte, V., Miller, H., Mulvaney, R., Oerlemans, J., Oerter, H., Orombelli, G., Parrenin, F., Peel, D. A., Petit, J.-R., Raynaud, D., Ritz, C., Ruth, U., Schwander, J., Siegenthaler, U., Souchez, R., Stauffer, B., Peder Steffensen, J., Stenni, B., Stocker, T. F., Tabacco, I. E., Udisti, R., van de Wal, R. S. W., van den Broeke, M., Weiss, J., Wilhelms, F., Winther, J.-G.,  
 445 Wolff, E. W., and Zucchelli, M.: Eight glacial cycles from an Antarctic ice core, *Nature*, 429, 623–628, 2004.
- Bouchaud, J.-P. and Muzy, J.-F.: Financial Time Series: From Batchelier’s Random Walks to Multifractal Cascades, in *The Kolmogorov Legacy in Physics, Lecture Notes in Physics*, 636, 229–246, Springer-Verlag, Berlin Heidelberg, 2003.
- Braun, H., Ditlevsen, P., Kurths, J., and Mudelsee, M.: A two-parameter stochastic process for Dansgaard?  
 455 Oeschger events, *Paleoceanography*, 26, PA3214, 2005, doi:10.1029/2011PA002140.
- Gardiner, C. W.: *Stochastic Methods. A Handbook for the Natural and Social Sciences*. Fourth Edition. Chapter 2, Springer, 2009.
- Hansen, J.: Earth’s Energy Imbalance: Confirmation and Implications, *Science*, 308, 1431–1435, 2005.
- Lomb, N. R.: Least-squares frequency analysis of unequally spaced data, *Astrophys Space Sci*, 39, 447–462,  
 460 1976.
- Lovejoy, S. and Schertzer, D.: Scale invariance in climatological temperatures and the local spectral plateau, *Annales Geophysicae*, 4,B, 401-410 , 1986.
- Lovejoy, S., D Schertzer, and D. Varon: Do GCMs predict the climate ... or macroweather?, *Eart Syst. Dynam.*, 4,439-454, 2013.
- 465 Lovejoy, S. and Schertzer, D.: *The Weather and Climate: Emergent Laws and Multifractal Cascades*, Cambridge University Press, New York, NY, 2013.
- Lovejoy, S.: A voyage through scales, a missing quadrillion and why the climate is not what you expect, *Clim Dyn*, 44, 3187–3210, 2014.
- Malamud, B. D. and Turcotte, D. L.: Self-affine time series: I. Generation and analyses, *Adv. Geophys*, 40, 470 1–90, 1999.

WAIS Divide Project Members: Precise interpolating phasing of abrupt climate change during the last ice age, *Nature*, 520, 661–665, 2015.

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, 2005.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, *J. Geophys. Res.*, 117, D08 101, 2012.

Rypdal, K., Rypdal, M., and Fredriksen, H.-B.: Spatiotemporal Long-Range Persistence in Earth’s Temperature Field: Analysis of Stochastic-Diffusive Energy Balance Models, *J. Climate*, 28, 8379–8395, 2015.

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, 2014.

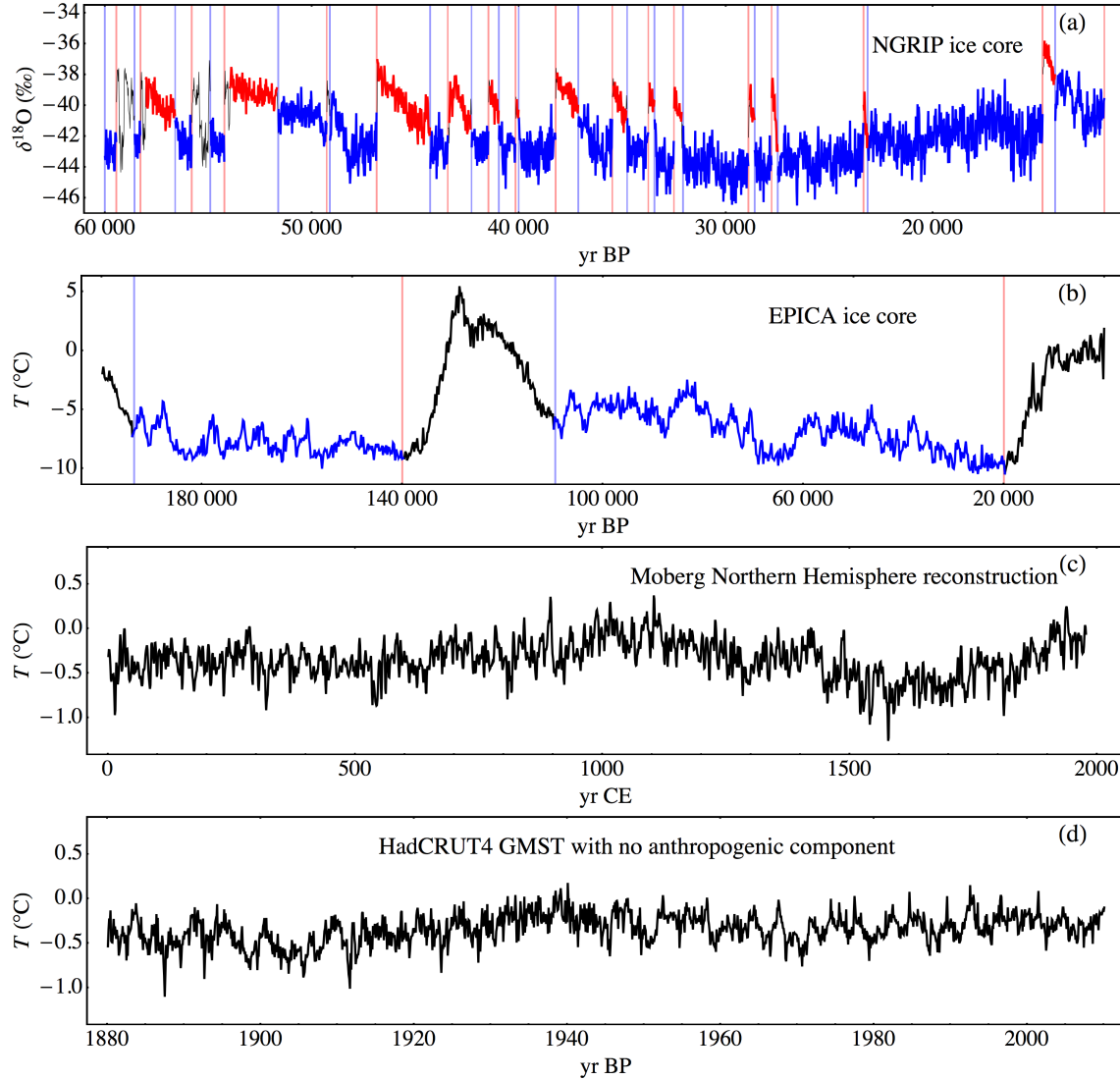
Rypdal, M. and Rypdal, K.: Is there long-range memory in solar activity on timescales shorter than the sunspot period?, *J. Geophys. Res.*, 117, A04103, 2012.

Rypdal, M.: Early-Warning Signals for the onsets of Greenland Interstadials and the Younger Dryas-Preboreal transition, *J. Climate*, in press , 2014, doi:10.1175/JCLI-D-15-0828.1.

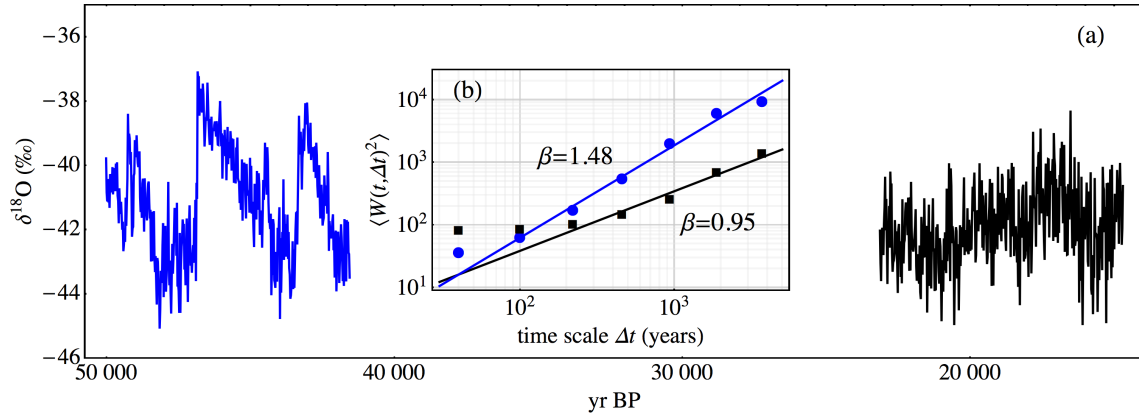
Schertzer, D., and Lovejoy, S.: Physical Modeling and Analysis of Rain and Clouds by Anisotropic Scaling Multiplicative Processes, *J. Geophys. Res.*, 92, 9693–9714, 1987.

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

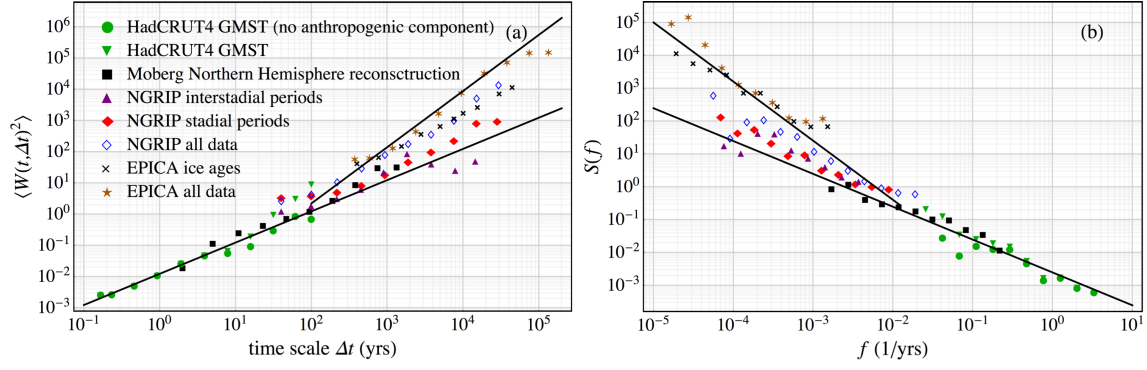




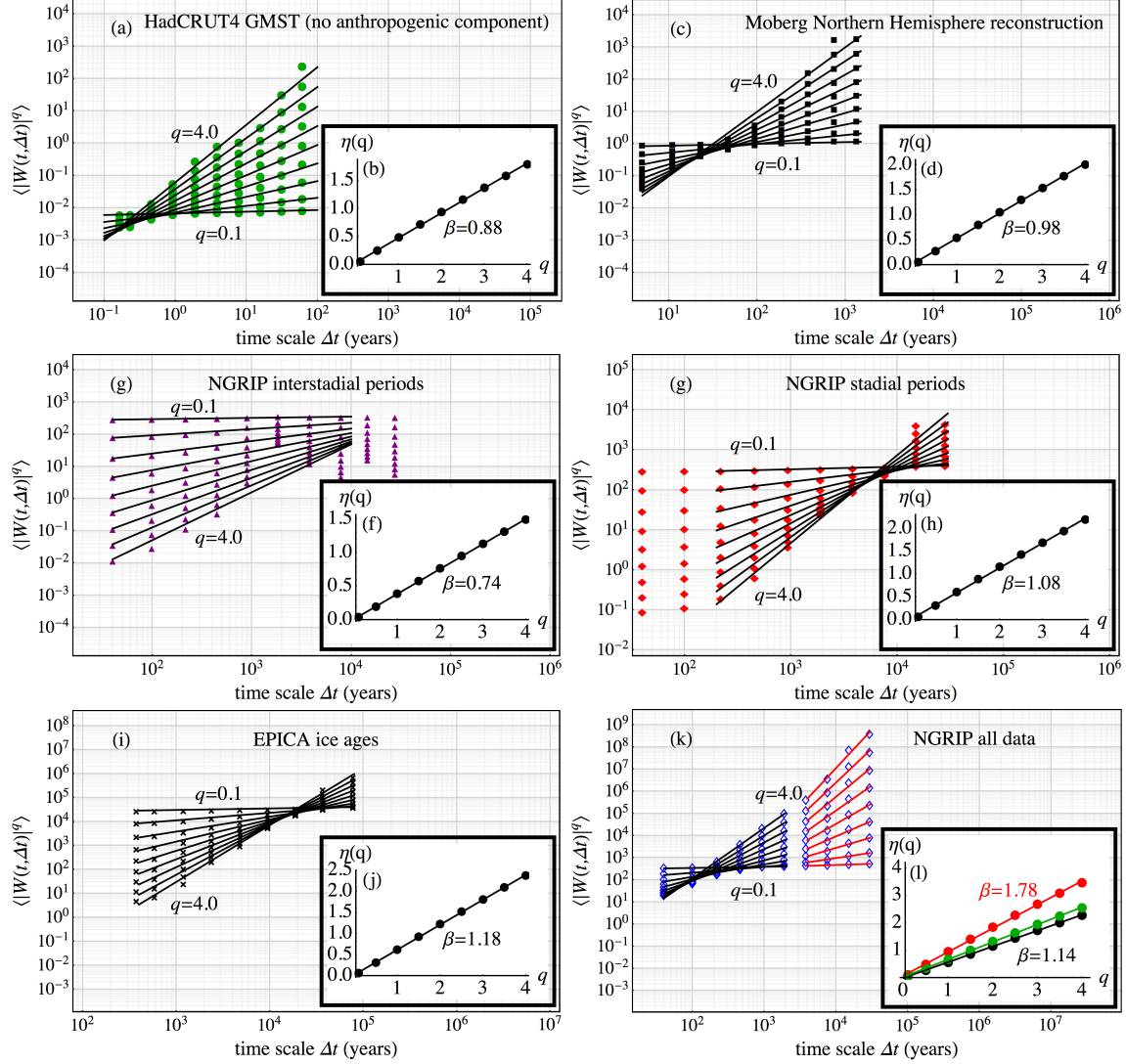
**Figure 1.** (a): The  $\delta^{18}\text{O}$  concentration in the NGRIP ice core dating back to 60 kyr before present (BP). Here present means AD 2000 (= 2000 CE). The data is given as 20-year mean values. The time series is split into stadal (blue) and interstadial (red) periods. (b): The temperature reconstruction from the EPICA ice core. The shown time series is sampled with a time resolution of roughly 200 years. The temperature curve in the glacial periods is given in a blue color. (c): The Moberg reconstruction for the mean surface temperature in the Northern Hemisphere. The data is given with annual resolution. (d): The HadCRUT4 monthly global mean surface temperature where the anthropogenic component has been removed using a linear-response model.



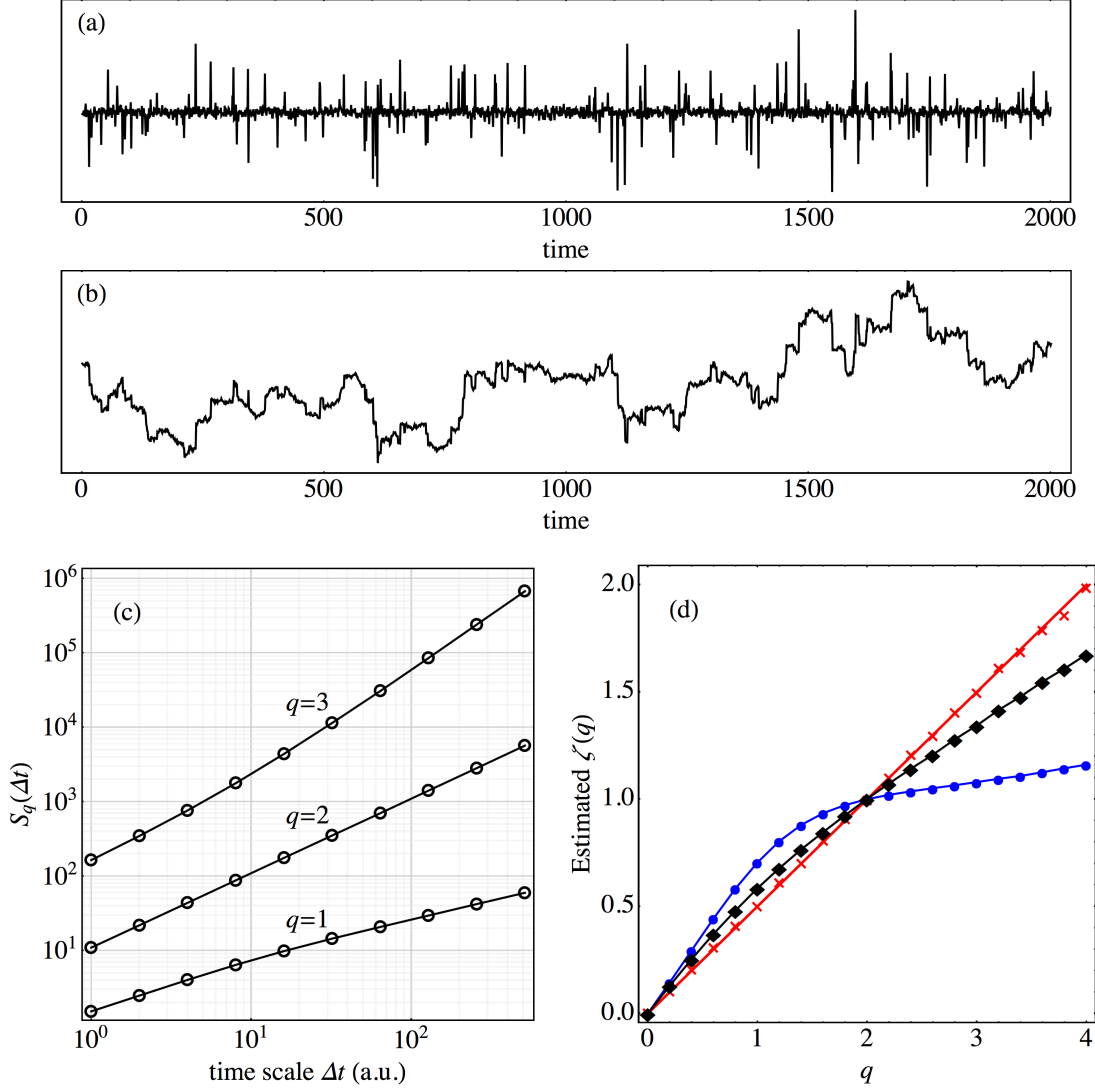
**Figure 2.** (a): The  $\delta^{18}\text{O}$  concentration in the NGRIP ice core. The data is given as 20-yr mean values. Two different parts of the the times series is shown. The blue curve represents the  $\delta^{18}\text{O}$  concentration in a time period starting approximately 50 kyr before present (BP) and has a duration of approximately 8500 years. As in Fig. 1, present means AD 2000 (= 2000 CE). The black curve represents the  $\delta^{18}\text{O}$  concentration in a long stadial period that started about 22 kyrs BP and has a duration of approximately 8500 years. (b): The wavelet scaling functions estimated from the two parts of the NGRIP data set. The blue points are the estimates from the part of the NGRIP ice core that is shown as a blue curve in (a), and which contains DO cycles. The black points are the estimates from the the part of the NGRIP ice core that is shown as a black curve in (a), and which does not contain any DO cycles.



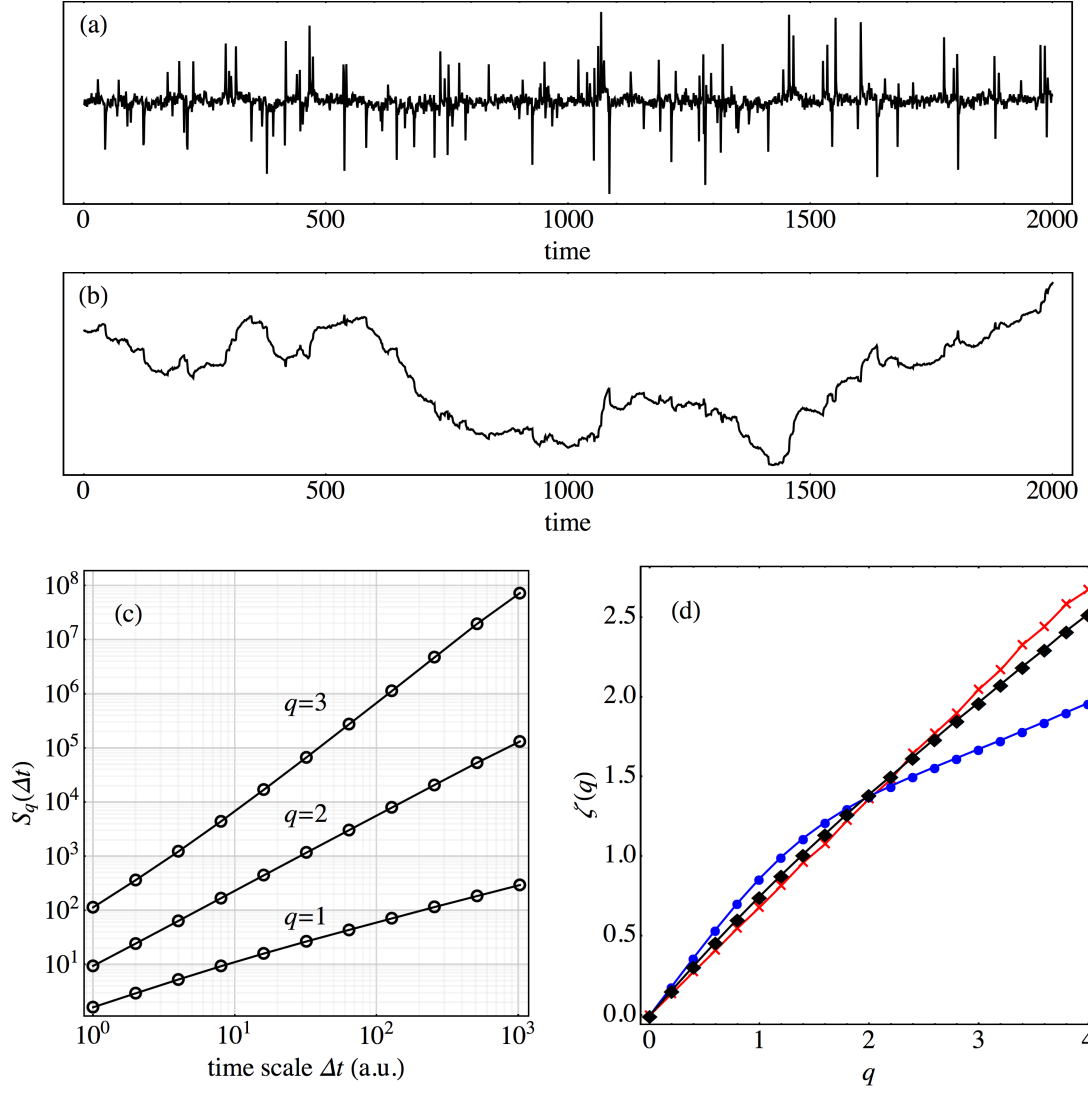
**Figure 3.** (a): For each time series considered in this paper we show double-logarithmic plots of the wavelet fluctuation  $\langle |W(t, \Delta t)|^2 \rangle$  as a function of the time scale  $\Delta t$ . The green triangles and the green circles represent the the HadCRUT4 monthly global mean surface temperatures with and without the anthropogenic component respectively. The black circles is the analysis of the Moberg Northern Hemisphere reconstruction. The analysis of the 20-yr mean NGRIP data is shown as the blue diamonds, the purple triangles and the red diamonds. The blue diamonds show the results of the analysis of the entire dataset dating back to 60 kyrs BP. The red diamonds are the results of the analysis performed on the stadial periods only, and the purple triangles are the results of the analysis of the interstadial periods only. The results for the EPICA ice core data are shown as the orange stars and the black crosses. The orange stars are obtained by analysis of the entire data set dating back 200 kyrs, and the black crosses are obtained by only analysing the two most recent glaciations. The two solid lines have slopes  $\beta = 1$  and  $\beta = 1.8$ . (b): As in (a), but instead of the wavelet fluctuation function we show the spectral density function  $S(f)$ . The two solid lines have slopes  $-\beta$  with  $\beta = 1$  and  $\beta = 1.8$ .



**Figure 4.** (a): The estimated wavelet-based structure functions  $\langle |W(t, \Delta t)|^q \rangle$  for the HadCRUT4 monthly global mean surface temperature where the anthropogenic component has been removed using a linear-response model. The lines show the fitted power-law functions  $c_q \Delta t^{\tau(q)}$ . The  $q$ -values are  $q = 0.1, 1.0, 1.5, \dots, 4.0$ . (b) The scaling function  $\tau(q)$  obtained from the fitted power-laws in (a). The line is a linear fit to the estimated scaling function, and the slope of this line is  $\beta/2$  with  $\beta = 0.88$ . (c-d): As (a) and (b) but in this case for the Moberg Northern Hemisphere reconstruction. (g-f):



**Figure 5.** (a): The increments of a jump-diffusion process shown in (b). This is a non-Gaussian independent noise process. (b): A realisation of a jump-diffusion process, and the cumulative sum of the signal in (a). This process is the sum of a Brownian motion and a Poisson jump process as described in Appendix B. The jump distribution is Gaussian with a standard deviation that is ten times greater than the standard deviation of the increments of the Brownian motion. (c):  $S_q(\Delta t)$  for  $q = 1, 2, 3$  for the jump-diffusion process as computed from a large ensemble of realisations of the process. (d): Scaling function  $\zeta(q)$  estimated from structure functions like those in (c). The red line is estimated by computing the slope of the structure-function curves on the longest time scale ( $\Delta t = 500$ ). The blue curve is estimated from the slopes at the shortest time scale ( $\Delta t = 1$ ). The black curve by estimating the slope of the straight line drawn between the end points of the structure-function curves.



**Figure 6.** As Fig. 5, but for a jump-diffusion process with memory as described in Appendix C. The parameter value  $\beta = 0.4$  is used.