The macroweather to climate transition in the Holocene: regional and epoch to epoch variability

(comments on "Are there multiple scaling regimes in
Holocene temperature records?"by T. Nilsen, K. Rypdal, and H.-B. Fredriksen)

General comments:

This paper addresses a basic issue in climate science: the nature of the Holocene temperature variability. Overall, I welcome this contribution, it will help to continue and deepen a debate on the nature of the variability, and on the fundamental atmospheric scaling regimes that goes back to [*Lovejoy and Schertzer*, 1986], [*Shackleton and Imbrie*, 1990].

That being said, I have several criticisms, the main ones being:

a) the authors have tended to obscure the fundamental issues by making them appear unnecessary technical.

b) they have been uncritical about the significance of several of the paleoseries they have employed.

c) The data analysis techniques that have issues.

Rather than give a traditional review of this paper, let me debate the issues and comment when appropriate on the authors' claims. Such a debate could be useful for the broader community.

-Shaun Lovejoy

a) Obscuring the basic issue:

i) Increasing and decreasing fluctuations: what is τ_c?

There is no debate that starting at scales of 5 - 10 days, that as time scales increase, that atmospheric fluctuations in general - and temperature fluctuations in particular - tend to *decrease* with scale out to at least 100 years. In the pre-industrial epoch, at century scales and for global (or hemispheric) averages they reach ≈ ±0.1K. This is not just a technical point, it has a basic and simple physical meaning: successive fluctuations tend to cancel out so that averaging over longer and longer periods tends to converge to the atmosphere's "climate state", this is the macroweather regime.

At some point –at time scale τ_c - this trend reverses itself, flucutations start to increase, the climate regime. We can be quite certain of this because of the existence glacial-interglacial transitions. From a wide range of evidence, we know that at 30- 50 kyr scales (i.e. for to time periods separated by 60 – 100 kyrs) that temperature fluctuations are typically of the order ±2 to ±3K (corresponding to typical differences of 4 - 6K between glacials and interglacials). This result goes back to [*Lovejoy and Schertzer*, 1986] (the "glacial-interglacial window", fig. 1: the authors should cite this earlier work) and has been amply confirmed since. Modern data enables us to even nuance this by considering the typical difference between glacial and interglacial temperatures as functions of latitude, estimating "high latitude amplification factors". In the industrial period, the reversal from decreasing to increasing fluctuations τ_c occurs at about 10- 20 years (global averages), this is the scale beyond which the anthropogenic fluctuations dominate the natural ones.

It is important to underline that none of these conclusions requires *any assumptions about the scaling or otherwise of the temperatures,* and they can be verified using straightforward fluctuation analyses (simpler and easier to interpret) than those used in this paper. Also, the authors should state these facts: without them the motivation and significance of their work is obscure.

To put it another way: up to some critical time scale τ_c, we *must* have decreasing fluctuations, convergence to climate, whereas for time scales longer than τ_c, fluctuations *must* start to increase - the climate state itself change from glacial to interglacial. Finding τ_c – including its geographical variation and epoch to epoch variation - is thus the basic question in paleoclimate. Notice that a priori, the critical scale τ_c is a property that does not depend on the existence of scaling, that is more fundamental than a break in the scaling (more on this later).

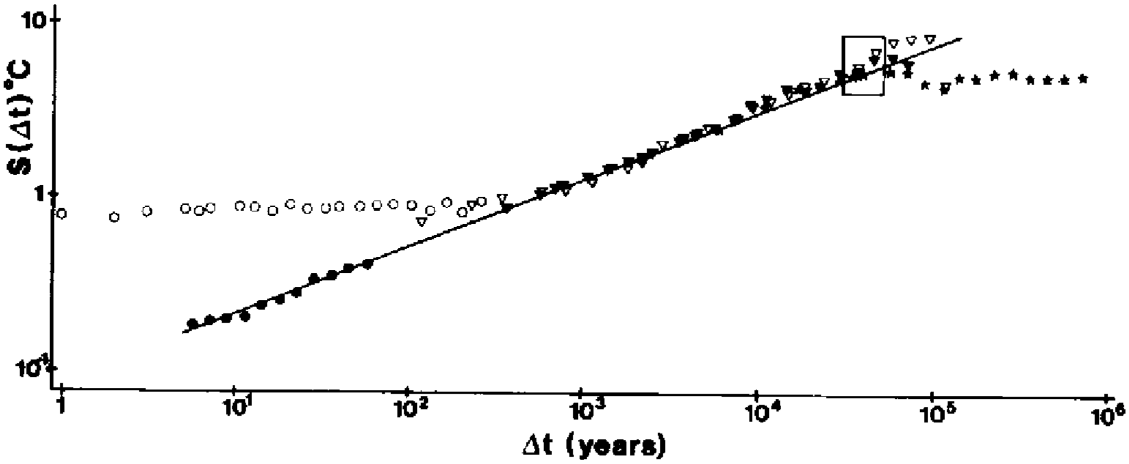

Fig. 1: The composite RMS ($S(\Delta t)=\left\langle \Delta T^2\right\rangle^{1/2}$) of temperature fluctuations estimated as differences (this is the usual "structure function"). The Manley central England (local) series is shown as open circles, the Budyko northern hemisphere averaged fluctuations are closed circles, Greenland (Camp Century) ice cores are the open triangles, the closed triangles are the Vostok Antarctic core and the asterixes are an ocean core. The rectangle is the "interglacial window" and the straight reference line is $S(\Delta t) \approx 0.077 \Delta t^{0.4}$ so that ignoring intermittency, H =0.4. Reproduced from [*Lovejoy and Schertzer*, 1986].

ii) The issue of scaling

In their paper, the authors do not mention this basic high versus low frequency question. They reduce the issues to the more narrow and technical one of: "are there multiple scaling regimes in the Holocene?". They then use a rather specialized (and not so realistic) model: fractional Gaussian noise (fGn) and fractional Brownian motion (fBm) which are both Gaussian and nonintermittent. Fluctuations $\Delta T(\Delta t)$ in the temperature $T(t)$ vary as $\Delta T(\Delta t) \approx \varphi \Delta T^H$ where φ is a driving flux and H is the fluctuation exponent. For both fGn and fBm, φ itself is a Gaussian with trivial scaling properties, all the scale dependence is determined by H which is >0 for fBm, and <0 for fGn (corresponding to fluctuations increasing and decreasing with time scale Δt respectively). Since the scaling depends on the single exponent H (equivalent here to the spectral exponent: $\beta = 2H+1$), the latter property is usually termed "monoscaling" in order to distinguish it from the general case of "multiscaling".

In this more general multiscaling case, φ itself has nontrivial scaling properties: $\left\langle \varphi_{\Delta t}^{q} \right\rangle \approx \Delta t^{-K(q)}$ where the subscript indicates the flux estimated at resolution Δt and $K(q)$ is convex exponent which a function of the order of moments q ("<.>" means ensemble averaging), it characterizes the intermittency by quantifying how far the process is from being Gaussian (which has $K(q) = 0$). Since the mean (q =1 moment) of the flux is by definition independent of scale, $K(1)$=0 and $\left\langle \Delta T(\Delta t) \right\rangle \propto \Delta t^{H}$ so in general that H characterizes the behavior of the mean fluctuations. Monoscaling is thus characterized by a single exponent whereas multiscaling involves an infinite hierarchy (fortunately, the latter can often be reduced to a two parameter functional form due to a multiplicative version of the central limit theorem,- but that's another story).

The usual use of the terms "monoscaling" and "multiscaling" therefore have noting to do with breaks in the scaling, they refer to the type of scaling (nonintermittent versus intermittent). The authors' cause confusion by using the expression "multiscaling" to instead designate monofractal scaling that is broken at least one scale. Rather than make a qualitative distinction between regimes where the mean fluctuations are decreasing or increasing (based on the sign of the exponent H), the authors use the distinction "persistent' and "antipersistent" which has no relevance except in their restrictive Gaussian framework. The same comment applies to their use of the term "stationary" or "nonstationary". On the one hand, these terms refer to either an infinite ensemble or to a theoretical model: empirical time series are thus neither stationary nor nonstationary. In the authors' case, the β>1 (fBm, H>0) model is only nonstationary if it is assumed to hold out to infinite times, which is a totally academic situation of no interest here. What it does however imply is that the series appears to nonstationary in the sense that fBm tends to "wander" like a drunkard's walk (which itself has H =1/2), but this wandering is simply a consequence of the fact that for fBm, H>0 so that mean fluctuations increase with scale. If over a range, H>0, no new information is added by saying that the process is "*apparently* nonstationary over the range".

This restriction to nonintermittent models is unrealistic and is unnecessary. On the one hand - even as far back as to [*Lovejoy and Schertzer*, 1986] - it was known that the actual distribution of temperature changes has “fat” power law tails (exponent $q_D \approx 5$). This means that the extremes are far “heavier” than Gaussians allow a fact that has also been exploited for statistical testing of the natural warming hypothesis for the industrial scale warming (see [*Lovejoy*, 2014a]). It has also been proposed as an explanation for the Dansgaard-Oeschger (DO) events that the authors invoke as a mechanism for *breaking* the scaling. On the contrary, [*Lovejoy and Schertzer*, 2013] (see fig. 2a,b below) argue that where rather than providing a mechanism that breaks the scaling, on contrary the DO are extreme events that occurs as a *consequence* of the scaling! On the other hand, there exists a simple parameter C_1 (“the codimension of the mean”, the rate of change of $K(q)$ near the mean: $K'(1)$) that conveniently quantifies the deviations from monofractality - when $C_1 = 0$, the process is quasi-Gaussian, nonintermittent. For climate series this was investigated in some detail in ch. 10, 11 of ([*Lovejoy and Schertzer*, 2013]. It was found that in the decreasing fluctuation (macroweather regime) that indeed the intermittency was low ($C_1 \approx 0.01$-0.03), but that in the increasing climate regime, it was significantly larger ($C_1 \approx 0.1$), being nearly the same as for atmospheric turbulence, hence displaying strong intermittency.

Since it turns out that over the macroweather regime the intermittency is low, for some purposes one can indeed use fGn as an approximation – for example in forecasting macroweather temperature (e.g. [*Lovejoy*, 2015]). However, the monofractality - or lack of intermittency - must be quantitatively established, not simply assumed *a priori*. This is especially true since it turns out to be a poor approximation for the tails and hence – precisely where the authors use it - for statistical testing.

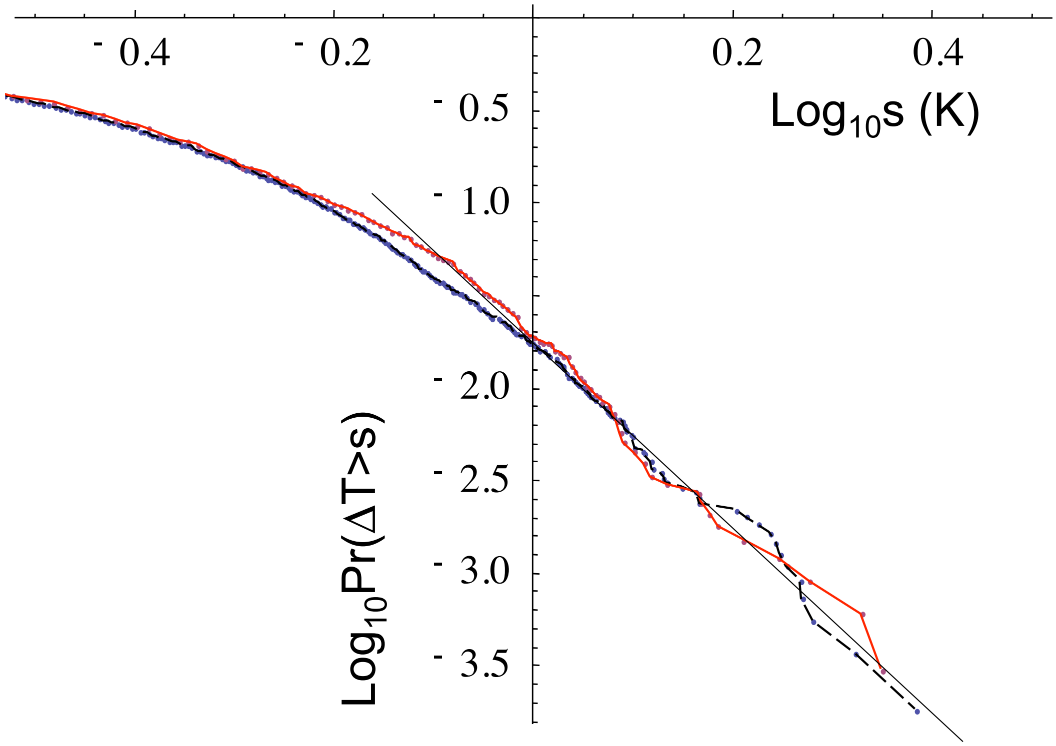

Fig. 2a: The probability distributions of changes in Greenland paleotemperature (GRIP blue, 5425 pts, from 248 kyrs BP) and Vostok (red, 3300 points, from 420 kyrs BP). The reference line has absolute slope = q_D =5 corresponding to a probability tail of for the probability of a random temperature fluctuation ΔT exceeding a fixed thereshold s. Reproduced from fig. 5.21 of [*Lovejoy and Schertzer*, 2013].

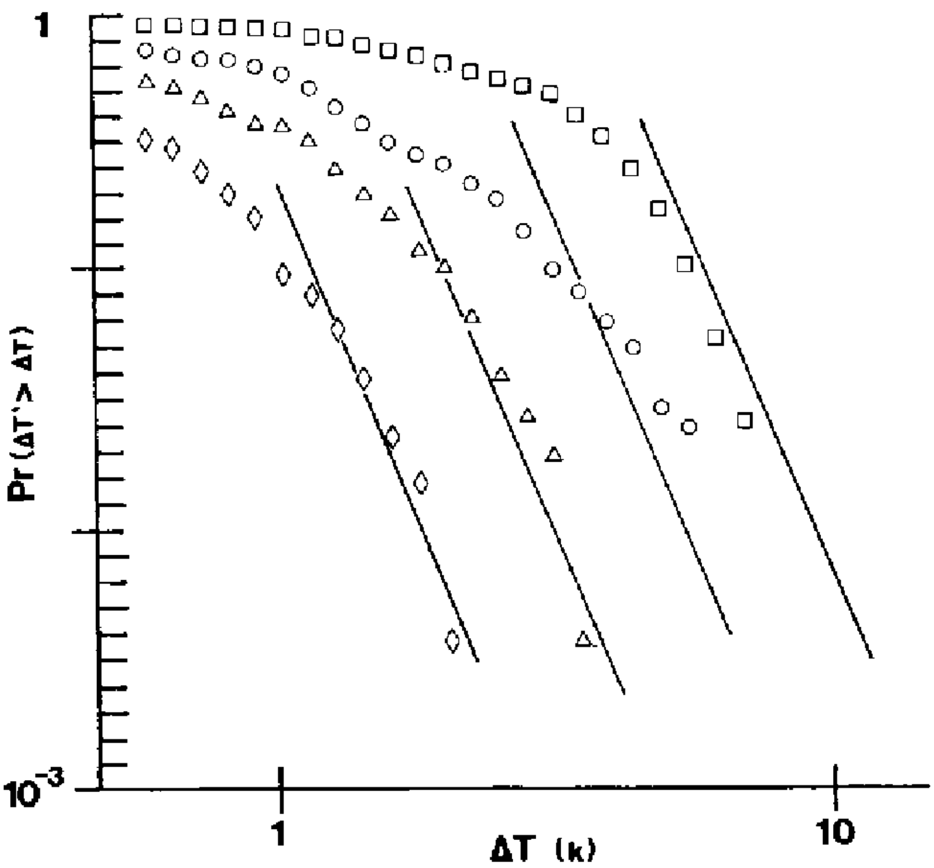

Fig. 2b: A log-log plot of $Pr(\Delta T' > \Delta T)$ which is the probability of temperature fluctuation ΔT' exceeding a fixed threshold ΔT. The data used are from the Antarctic (Vostok) series used in fig. 1 (with the linear temperature calibration discussed in [*Lovejoy and Schertzer*, 1986]). The curve to the far left is for fluctuations over intervals (Δt) of 350 years. The remaining curves, from left to right are obtained by increasing Δt by factors of 4 (i.e. 1400, 5600, 22400 years respectively). The straight lines indicate the functions $Pr(\Delta T' > \Delta T) \approx \Delta T^{-q_D}$ with $q_D = 5$ and the amplitude to the fluctuations varying as $\Delta T^* \approx \Delta t^H$ with $H = 0.4$. Reproduced from [*Lovejoy and Schertzer*, 1986].

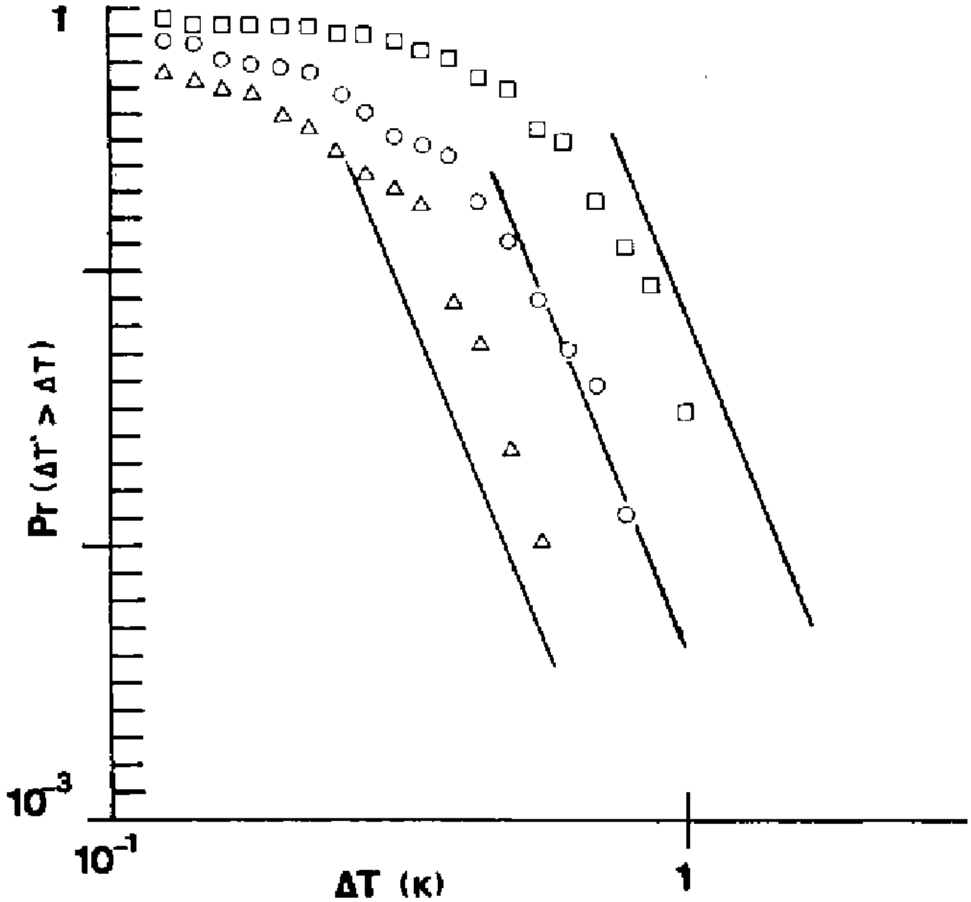

Fig. 2c: The same as fig.2a,b except for the northern hemisphere annual average (Jones) data. The curves from left to right are for Δt = 1, 4, 16, 64 years. The straight lines are for the same values of H, q_D and in the previous. This has been updated in [*Lovejoy*, 2014a] which confirms that the same exponent q_D is valid. Reproduced from [*Lovejoy and Schertzer*, 1986].

b) Uncritical treatment of paleoseries

i) Multiproxies:

The authors use multiproxy reconstructions and high latitude ice core isotope paleo temperatures. First, let us discuss the multiproxies. As with any new technique, the early multiproxies had imperfections, notably at low frequencies. As the methodologies and data improved, these were given more care. Specifically, as pointed out in [*Lovejoy and Schertzer*, 2012a], the low frequency early (pre 2003) multiproxies, have fairly systemic differences with respect to the post 2003 multiproxies, see fig. 3a that used the following: [*Moberg et al.*, 2005] wavelet based, [*Huang*, 2004] Huang borehole based and [*Ljungqvist*, 2010]). Fig. 3b shows an even more recent multiproxy covering the entire Holocene [*Marcott et al.*, 2013] that shows systematically increasing (not decreasing fluctuations) with τ_c presumably ≈ 100 years.

When the authors make statistical Monte Carlo, fGn based tests about single or multiple scaling ranges, in addition to the questionable usefulness of the fGn model for this, more fundamentally, they miss the key point that the multiproxies are really quite far from perfect and that their tests mostly shed light on the low frequency limitations of the multiproxies, not (as they suppose) on the statistical variations of well measured quantities that differ only in standard measurement errors. In plain English: the multicentennial behaviour of the temperature in the last millennium is not a scientifically settled issue!

ii) ice cores:

Ice cores are more reliably converted to temperatures, but when using them, the authors have ignored evidence (fig. 3a, b), that the Greenland Holocene ice cores are very exceptional (to a lesser extent, so are the Vostok Holocene cores). This can be seen in fig. 4a, b where we see that whereas the Greenland Holocene is less variable by 3- 4 standard deviations when compared to earlier 10 ky periods, (bottom dashed red), but that the Sea Surface Temperatures (SST) 1500 km distant (from ocean cores from [*Berner et al.*, 2008], dashed red top) has almost the same variability as the previous Greenland 10 kyr periods! Also the Vostok Holocene is only very slightly outside of its previous one standard deviation limits so that it is really the Greenland ice cores that are exceptional (this is confirmed by the analysis of the Marcotte Holocene multiproxy, fig. 4b).

The finding that the near-Greenland SST shows a highly "unstable" climate ([*Berner et al.*, 2008])) combined with the new [*Marcott et al.*, 2013] Holocene reanalysis (fig. 4b) brings into serious doubt the significance of the Holocene Greenland cores for informing us on the global Holocene climate. Greenland seems to be simply exceptional.

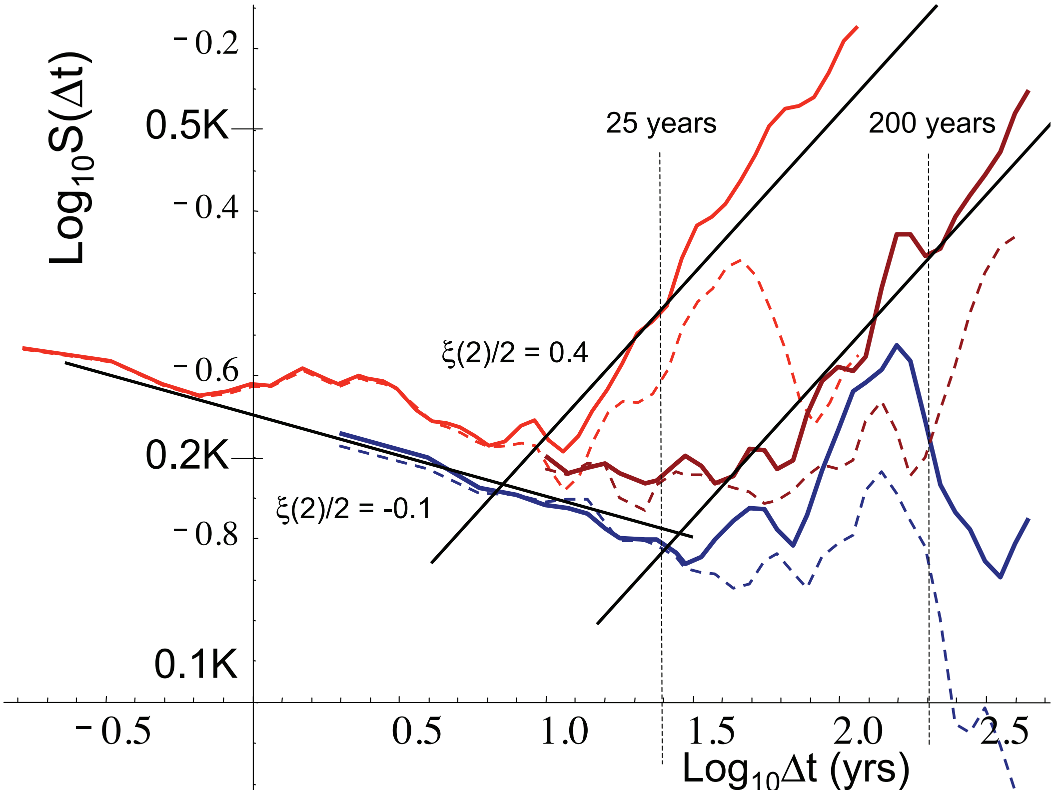

Fig. 3a: The RMS Haar fluctuation ($S(\Delta t)$) for the mean of the pre-and post-2003 series from 1500 to 1979 (bottom and middle solid lines respectively and excluding the Crowley series because of its poor resolution), along with the mean of the globally averaged monthly resolution surface series (NOAA CDC, NASA GISS, HadCRUT3) (solid, top). In order to assess the effect of the twentieth-century warming, the structure functions for the multiproxy data were recalculated from1500–1900 only (the dashed lines that join the solid lines at small lags) and for the instrumental surface series with their linear trends from 1880–2008 removed (the data from 1880–1899 are too short to yield a meaningful S(Δt) estimate for the lower frequencies of interest). Although in all cases the large Δt variability is reduced, the basic power-law trend seems to remain, although the transition scale τc increases (especially for the post-2003 reconstructions). Note that the decrease in S(Δt) for the linearly detrended surface series over the last factor of 2 or so in lagΔt is a pure artefact of the detrending. Reference lines corresponding to β=0.8 and 1.8 have been added. Reproduced from Lovejoy and Schertzer (2012a).

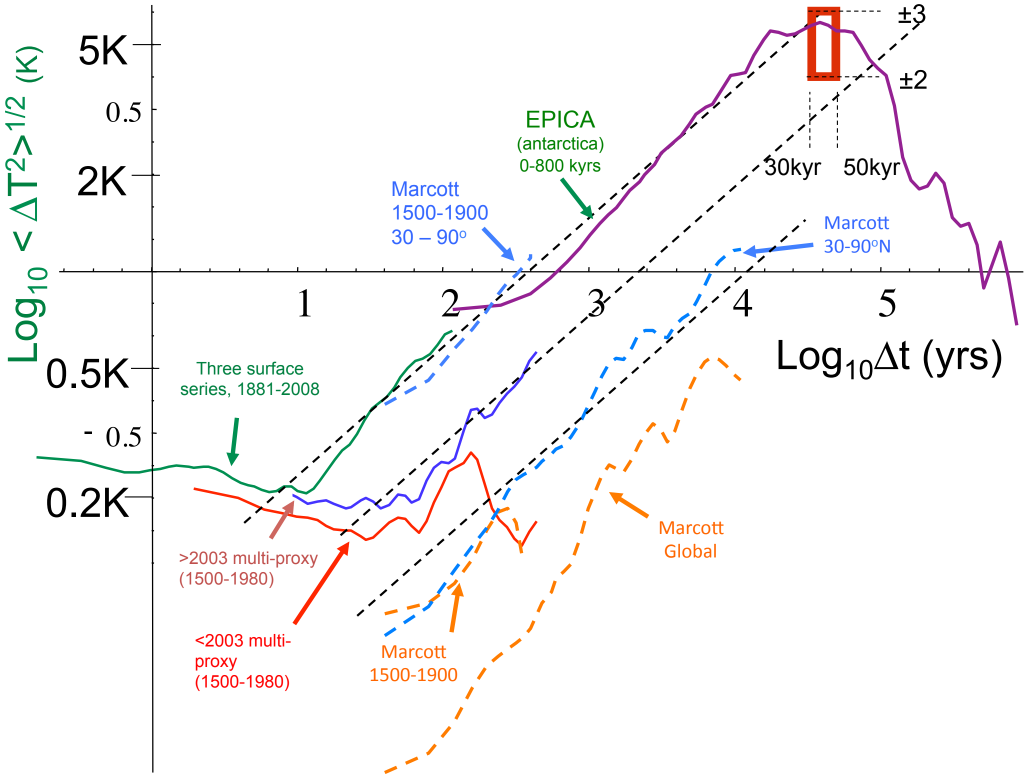

Fig. 3b: Comparison of the RMS Haar fluctuations $S(\Delta t)=\left\langle \Delta T^{2}\right\rangle^{1/2}$ from: surface series analyses analysed in fig. 2a (green), the pre and post 2003 multiproxies (from fig. a) as well as various analyses of the recent [*Marcott et al.*, 2013] reconstruction based on 73 very long proxies at 20 year resolution. Although the absolute amplitude of the fluctuations depends on the time period and latitudinal band, the RMS fluctuations are increasing – not decreasing - with time scale. It appears that in the period prior to 1500 A.D, the variability is however a bit too weak.

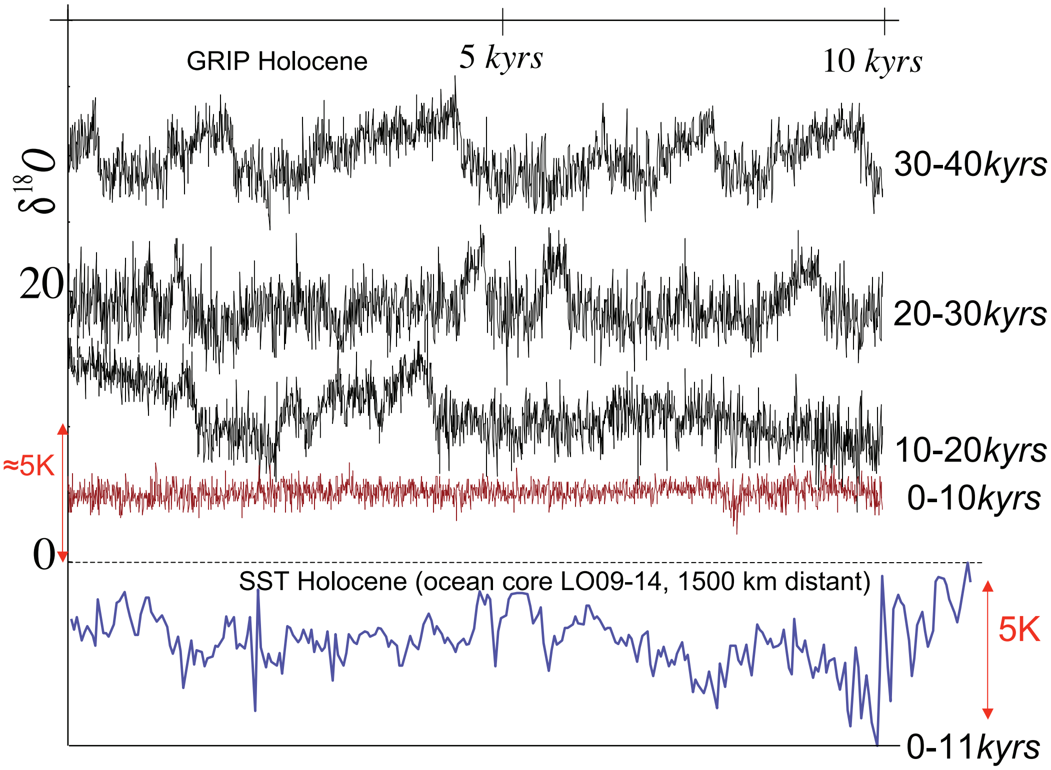

Fig. 4a: The top part shows four successive 10 kyr sections of the 5.2-year resolution GRIP data, the most recent to the oldest from bottom to top. Each series is separated by 10mils in the vertical for clarity (vertical units: mils–i.e.parts per thousand of isotope excess). For reference, a 5K corresponding temperature spread is also shown using a calibration constant of 0.5K/mil. We see that the bottom Holocene GRIP series is indeed relatively devoid of low-frequency variability compared to the previous 10kyr sections, a fact confirmed by statistical analysis discussed in the text and shown in fig. 4b. In contrast, the bottom curve shows the (much lower resolution but on the same scale) paleo-SST curve from ocean core LO09–14 (Berner et al., 2008), taken from a location only 1500km distant and displaying far larger variability: see fig.4b. Adapted from[*Lovejoy and Schertzer*, 2012a], reproduced from [*Lovejoy and Schertzer*, 2013].

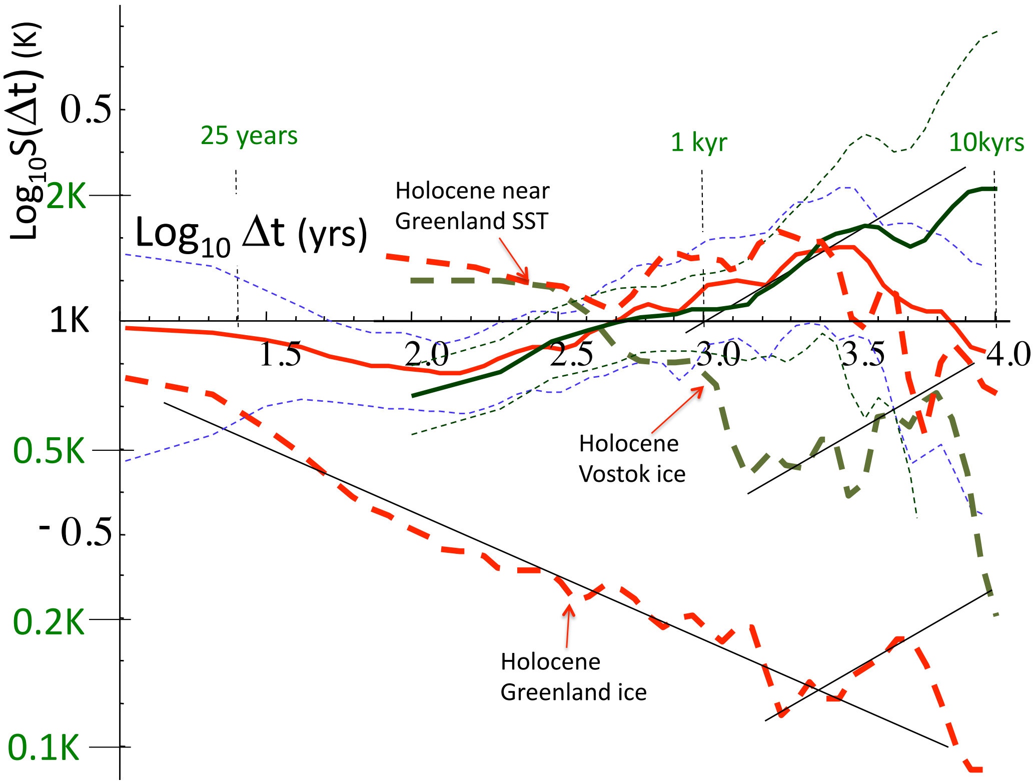

Fig. 4b: A comparison of the RMS Haar structure function ($S(\Delta t)$) for both Vostok and GRIP high resolution cores (resolutions 5.2 and 50 years respectively over the last 90 *kyrs*). For Vostok we used the Petit et al calibration, for GRIP, 0.5 *K/mil*. The series were broken into 10 *kyr* sections. The dashed curves show the most recent of these (roughly the Holocene), with Greenland GRIP and Bernier et al data red and Vostok green). The solid red, and green are the mean of the eight 10 – 90 *kyr* GRIP and Vostok RMS fluctuations ($S(\Delta t)$). The one standard deviation variations about the mean are indicated by very thin dashed lines. Also shown are reference lines with slopes $\xi(2)/2$ = -0.3, 0.4 corresponding to β = 0.4, 1.8 respectively. Although the Holocene is exceptional for both series, for GRIP it is exceptional by many standard deviations. For the Holocene we can se that $\tau_c \approx 1$ *kyr* for Vostok, and ≈ 2 *kyr* for GRIP, although for the previous 80 *kyrs*, we see that $\tau_c \approx 100$ *yrs* for both. Adapted from [*Lovejoy and Schertzer*, 2012a], and [*Lovejoy and Schertzer*, 2013].

c) Technical points:

i) The authors reiterate an unfortunate "rule of thumb" (p1206) to effect that the longest factor of 4 at the low frequencies are too unreliable "for meaningful estimates". This ill- starred rule of thumb originated in the Detrended Fluctuation Analysis (DFA) literature and is based on the fact that as we move to low frequencies there are fewer and fewer statistically independent samples hence that the error in their estimation increases. The situation is analogous to the significance of extreme values. While it may be true that extremes are difficult to quantify, the largest (or smallest) value in series or in a ensemble of random variables is on the contrary highly statistically significant. If in a sample of 1000 values, one finds events that are at four or five standard deviations from the mean, then the event is extremely statistically significant even though assigning it a precise probability value it difficult.

In the present case, eliminating the factor 4 lowest frequencies in the instrumental temperatures (e.g. since 1880) virtually eliminates the possibility of detecting the anthropogenic signal above the noise: the latter is indeed only significant in the extreme three or four lowest frequencies. Quite concretely, its employment in DFA analyses of station temperature records has meant that a series of papers were blind to this very strong effect. Applying the rule in the case of last millennium multiproxies would imply that the multiproxies had nothing to say about the medieval warming or other low frequency events. To summarize, the lowest frequencies are often extremely statistically significant and that is precisely why considerable effort is made to extend series over as long periods as possible. To properly quantify their significance, one needs to make an assumption about the stochastic (or deterministic) process generating the series - here for example is it mono or multiscaling (multifractal) - and with which exponents? Without an explicit stochastic or dynamical model, if one wants to estimate scaling exponents, then it is hard to go much beyond the obvious i.e. that one may expect larger random deviations at the long times than at the short times.

A final comment on the authors use of statistical testing: using the fGn model, the authors fail to reject the hypothesis that there is a single scaling regime. This fact in no way obliges us to accept that there is no break!

ii) The authors use the Mexican Hat and Morley wavelets even though these are difficult to implement on non uniformly spaced data and their interpretations are nontrivial. Why not use the much simpler Haar wavelets (see [*Lovejoy and Schertzer*, 2012b] (which for an interval Δt are simply given by the difference between the means of the first and second halves of the intervals) and which can easily be estimated from irregularly space data (see appendix A of [*Lovejoy*, 2014b]). In regions where the fluctuations increase with scale (Δt), the Haar can be interpreted as differences, and in regions where it decreases with scale, they can be interpreted as anomalies. Unlike the DFA or other standard wavelets, the fluctuations defined by Haar wavelets are easy to use, easy to understand, and have the added bonus that the actual values of the fluctuations (or mean or RMS, or other statistic of the fluctuations) can be interpreted directly. The difficulty in understanding the significance of the fluctuations defined by Mexican Hat or Morley wavelets (or most other wavelets!) presumably explains why the authors' wavelet results (fig. 5) are given no units (or even dimensions) and indeed, this it is traditional (almost universal) in DFA analyses to give the fluctuation functions without even indicating their units or dimensions!

References:

Berner, K. S., K. N., D. Divine, F. Godtliebsen, and M. Moros (2008), A decadal-scale Holocene sea surface temperature record from the subpolar North Atlantic constructed using diatoms and statistics and its relation to other climate parameters *Paleoceanography*, *23* doi: Doi:10.1029/2006pa001339.

Huang, S. (2004), Merging Information from Different Resources for New Insights into Climate Change in the Past and Future, *Geophys.Res, Lett.* , *31*, L13205 doi: doi : 10.1029/2004 GL019781.

Ljungqvist, F. C. (2010), A new reconstruction of temperature variability in the extra - tropical Northern Hemisphere during the last two millennia, *Geografiska Annaler: Physical Geography*, *92 A*(3), 339 - 351 doi: DOI : 10.1111/j .1468 - 0459.2010 .00399.x.

Lovejoy, S. (2014a), Scaling fluctuation analysis and statistical hypothesis testing of anthropogenic warming, *Climate Dynamics*, *42*, 2339-2351 doi: 10.1007/s00382-014-2128-2.

Lovejoy, S. (2014b), A voyage through scales, a missing quadrillion and why the climate is not what ou expect, *Climate Dyn.*, *doi: 10.1007/s00382-014-2324-0.*

Lovejoy, S. (2015), Using scaling for macroweather forecasting including the pause, *Geophys. Res. Lett.*, *in press* doi: DOI: 10.1002/2015GL065665.

Lovejoy, S., and D. Schertzer (1986), Scale invariance in climatological temperatures and the local spectral plateau, *Annales Geophysicae*, *4B*, 401-410.

Lovejoy, S., and D. Schertzer (2012a), Low frequency weather and the emergence of the Climate, in *Extreme Events and Natural Hazards: The Complexity Perspective*, edited by A. S. Sharma, A. Bunde, D. N. Baker and V. P. Dimri, pp. 231-254, AGU monographs.

Lovejoy, S., and D. Schertzer (2012b), Haar wavelets, fluctuations and structure functions: convenient choices for geophysics, *Nonlinear Proc. Geophys.* , *19*, 1-14 doi: 10.5194/npg-19-1-2012.

Lovejoy, S., and D. Schertzer (2013), *The Weather and Climate: Emergent Laws and Multifractal Cascades*, 496 pp., Cambridge University Press, Cambridge.

Marcott, S. A., 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*, 1198 doi: doi: 10.1126/science.1228026.

Moberg, A., D. M. Sonnechkin, K. Holmgren, N. M. Datsenko, and W. Karlén (2005), Highly variable Northern Hemisphere temperatures reconstructed from low- and high - resolution proxy data, *Nature*, *433*(7026), 613-617.

Shackleton, N. J., and J. Imbrie (1990), The δ18O spectrum of oceanic deep water over a five-decade band, *Climatic Change*, *16*, 217-230.