Why the warming can't be natural: the nonlinear geophysics of climate closure

S. Lovejoy,
Physics, McGill University,
3600 University st.,
Montreal, Que. H3A 2T8

Abstract:

Claims with extraordinary consequences require extraordinary evidence. It is not sufficient to prove that the industrial epoch warming is "extremely likely" to be anthropogenic: to achieve "climate closure", we must also disprove the converse, the theory that it is a natural fluctuation. This disproof requires an understanding of the strongly nonlinear atmospheric variability: since the 1980's this understanding is one of the important contributions of nonlinear geophysics to atmospheric science. Atmospheric variability spans twenty orders of magnitude in time and ten in space, but it can be tamed with the help of emergent high-level turbulent laws and their anisotropic and multifractal generalizations.

We take a voyage through scales and with the help of some new fluctuation analysis tools, we review the arguments and evidence for wide scale range space-time scaling: high level turbulent laws. We notably find that in addition to the familiar weather and climate regimes, there is an intermediate "macroweather" regime – between 10 days and about 100 years in which the fluctuation exponents are negative so that successive fluctuations tend to cancel each other out, the atmosphere appears stable. We examine the consequences for the extreme fluctuations ("grey swan", "black swan" events) and we apply this to estimating the probabilities and return times of extreme centennial, global scale temperature fluctuations. Even with conservative assumptions, the natural warming hypothesis can be rejected with >99% confidence.

Our understanding of the natural variability allows us to understand and stochastically predict events such as the post 1998 "pause" in the warming that is often claimed by climate skeptics as evidence that the warming has stopped. First, we use the unconditional statistics to show that the pause is simply a natural cooling event that largely masked the ongoing warming. But the pause was preceded by an even larger prepause warming event. By developing a space-time stochastic model – the ScaLIng Macroweather Model (SLIMM) - we were able to accurately forecast the pause, quantitatively showing that it was very likely: without it the warming would have become too strong. The new SLIMM model also allows for regional forecasts from months, year and decades.

1. Introduction:

The atmosphere is a turbulent fluid whose temperature, humidity and wind vary from submillimetric eddies visible in cigarette smoke to huge planetary sized weather systems. It has been changing ever since the earth was formed several

billion years ago and it changes at millisecond scales. Sixty five million years ago, the temperature was five or even ten degrees warmer than today and dinosaurs roamed an ice-free south pole.   As little as fourteen thousand years ago, the earth was still in the throes of an ice age with a global temperatures 2- 4 degrees cooler that today. Historical viniculture records show that in the middle ages, England was significantly warmer than today, yet only centuries later in the "little ice age", Europe had cooled enough so that sixteenth century dutch skaters were immortalized in Breugel's famous paintings.

These facts are supported by several converging lines of evidence and while the quantitative amounts of the warming and cooling are debated, the basic events are undisputed. Indeed, as shown in fig. 1a-e, proxy and instrumental records show that there is strong variability at all observed time scales and fig. 2a quantitatively confirms this with a modern spectrum showing that contrary to conventional wisdom, the "background" spectrum varies by a factor of more than 10^{15} over the range from hours to hundreds of millions of years. In space – and indeed in space-time - Fig. 2b shows that the scaling of satellite infra red radiance extends to planetary scales and up to about 10 days in time. It shows more: it shows that over this range, that the (horizontal – not vertical) spatial and temporal statistics are related by an isotropic space-time scale invariance symmetry.

Without human intervention, over sufficiently long periods, the temperature of the earth can clearly change by several degrees. But what about this: since the end of the 19th century instrumental records show that the earth has warmed by about one degree centigrade. The evidence is all around us: from the melting of polar sea ice – including the summer opening of the Northwest passage – to rising sea levels to deadly heat waves. But what is the cause? Is it simply another natural fluctuation, or is it something different, something artificial, something that only we could have done? More precisely, is a one degree warming of the whole planet *in only a single century* an ordinary - even common - event in the history of the earth, or is it so rare as to be demand a non natural explanation?

The modern answer to this question emerged well before the warming itself was felt or even before human emissions had significantly changed the atmospheric composition. In 1896, in an attempt to understand the causes of the ice ages, Svante Arrhenius estimated that if the concentration of carbon dioxide (CO_2) in the atmosphere was doubled, that global temperatures would rise by 5 – 6° C, quite close to the modern value of 1.5- 4.5^{o}C, (International Panel on Climate Change, fifth assessment report: IPCC AR5). From a scientific point of view, the basic result is straightforward: CO_2 is a "Greenhouse Gas": it lets visible light from the sun through to the surface while absorbing part of the earth's outgoing heat radiation.

Arrhenius's theory signalled the beginning of modern attempts to prove the anthropogenic provenance of a warming that only became strongly apparent in the 1980's. From a purely scientific point of view, the main difficulty is that there are complicated feedbacks between CO_2, water vapour and clouds: these are the main effects responsible for the uncertainty. Arrhenius spent the best part of a year with pencil and paper grappling with these complications; today scientists use the world's most powerful supercomputers.

Is the warming is mostly man-made, through the emission of CO_2 and other

Greenhouse gases, or is it is mostly natural? Today, the theory of anthropogenic warming is entering a mature phase in which continued efforts to prove it more convincingly, are starting to suffer from diminishing returns.   Take for example the IPCC's Fifth Assessment Report   (AR5, 2013): not withstanding massive improvements in computers and algorithms, it cited exactly the same range of temperature increase for a doubling of CO_2 as did the US National Academy of Science report in 1979: 1.5 to 4.5^{o}C. Whereas the fourth report (AR4, 2007) stated that it is "likely that human influence has been the dominant cause of the observed warming since the mid-20th century", six years later, the AR5 only upgraded this to "extremely likely".

In spite of the strong evidence in favour of the anthropogenic theory, it still faces a chorus of denial with entire organizations – such as Canada's "Friends of Science" - dedicated to the proposition that the "The sun is the main driver of climate change. Not CO_2. Not you" a slogan that adorned billboards across Canada in November 2014 and - at least in Quebec - prompted counter-billboards financed by the Association of Science Communicators. Solar, volcanic and internal climate system variability are all invoked by various proponents as plausible – or even proven – alternatives to the anthropogenic theory. So who is right?

In order to break through the impasse, to "close" the debate ([*Lovejoy*, 2015]) it is helpful to recall that science progresses not only from attempting to prove certain theories to be true, but also by rejecting theories that are false. In this, it benefits from a fundamental methodological asymmetry: while no theory can ever be proven true "beyond all doubt", even a single decisive experiment can disprove one that is otherwise highly seductive. In their day–to-day work, scientists constantly reject ideas and theories that are incompatible with observations or with more powerful theories that are known to be true.

In this paper, we review a disproof that is based on a different approach. Rather than exploiting large scale deterministic numerical models, we use past data combined with the new theoretical understanding of the natural variability that has been made possible by advances in nonlinear geophysics, in particular in scaling and multifractals. In section 2 we review elements of turbulence theory; the high level turbulent laws and explain how they can be generalized to take into account both strong nonclassical variability (intermittency, extremes), as well as strong scale dependent anisotropy (especially vertical stratification). In section 3 we discuss fluctuations and use them to give an objective definition of the climate. In section 4 we show how these elements can be combined first to make statistical tests of the giant natural fluctuation hypothesis, and second, by using concrete stochastic models, to accurately hindcast the recent "hiatus" or pause in the warming that is often invoked by skeptics as evidence that the warming is over. In section 5 we conclude.
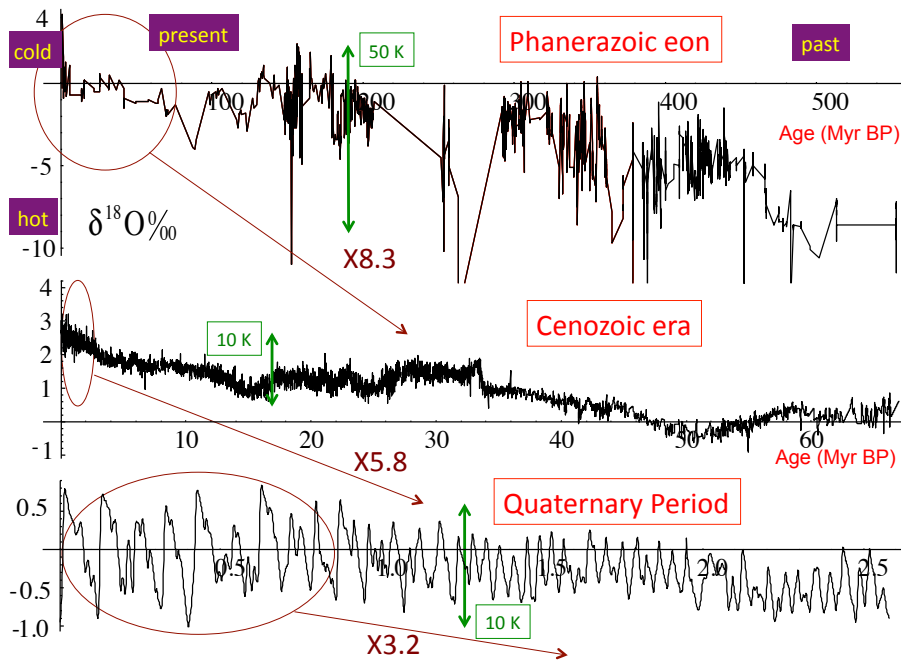
Fig. 1a: $\delta^{18}O$ from assemblies of cores from ocean sediments of benthic organisms, table 1 gives some information on the series.   Large values correspond to small temperatures and visa versa.

      The top series is an update of a global assemblage by [*Veizer et al.*, 1999], covering the Phanerozoic; the current geological eon during which abundant animal life has existed and goes back to the time when diverse hard-shelled animals first appeared: this figure goes as far back as this technique will allow. Although 2980 values were used, but they are far from uniformly distributed; the figure shows a linear interpolation. The corresponding temperature range is indicated based on the "canonical" calibration of -4.5 K/$\delta^{18}O$, and may be as much as a factor 3 too large. Note the negative sign in the calibration: large $\delta^{18}O$ corresponds to small temperatures and visa versa.

      The middle series is from a northern high latitude assemblage by [*Zachos et al.*, 2001] based on global deep-sea isotope records from data compiled from more than 40 Deep Sea Drilling Project and Ocean Drilling Project sites; it has 14828 values covering the period back 67 million years ago, again non uniformly distributed in time and considered to be globally representative. The bottom series is from [*Huybers*, 2007], it uses 2560 data points 12 benthic and 5 planktic $\delta^{18}O$ records over the Quaternary (the recent period during which there were glacials and interglacials; the rough oscillations that are visible, the series is mostly from high northern latitudes).   For both of these series a roughly 50% larger calibration constant -6.5 K/$\delta^{18}O$ was used in order to take into account the larger high latitude variations.

      The ellipses, arrows and numbers indicate the parts of the time axis and zoom factor needed to go from one series to the next. This figure is continued in fig. 1b.
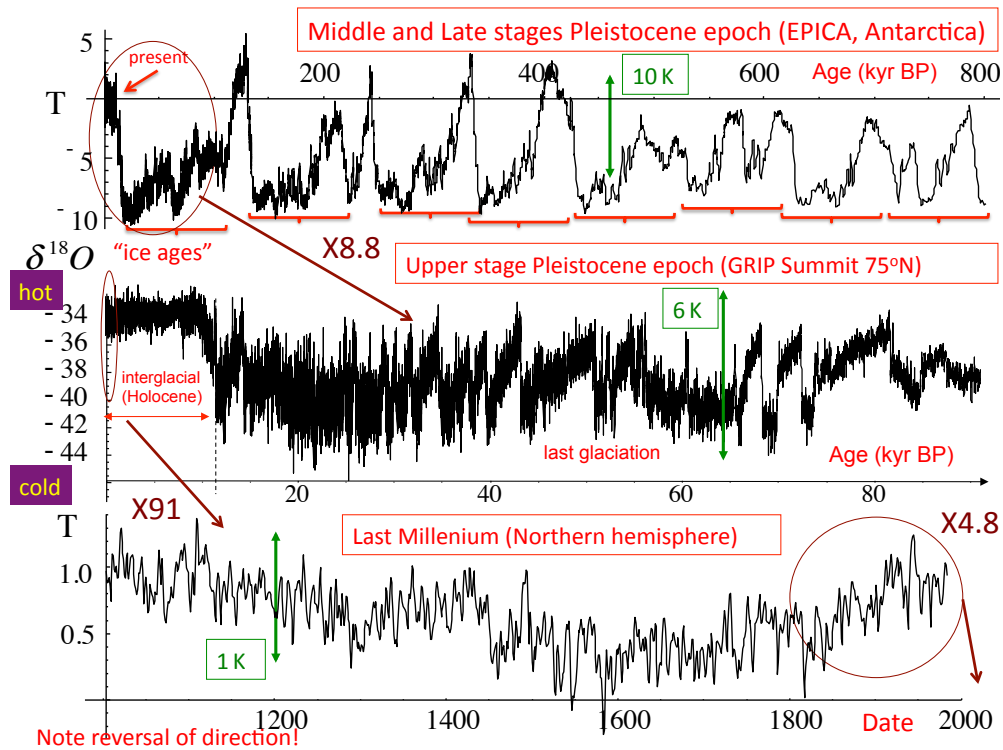
Fig. 1b: The top series is the temperature anomaly from the Epica Antarctic core using a Deuterium based paleotemperature the temperature anomalies are in degrees K (see [*Lovejoy*, 2014b] for information on the data). We may note the loss in resolution (the apparent increase in smoothness) of the curve as we move into the past (to the right); it is an artefact of the compression of the ice column. Clearly, the neat classification of the series into 8 glacial and interglacial (parentheses) epochs is a somewhat subjective simplification of the true variability. Up to at least periods $\approx 10^5$ years, we see that the temperature seems to "wander" i.e. as we consider the change in temperature for increasingly long time periods, the temperature changes more and more. The series in the second row is over the time period indicated by the circle in top row, it is from a high resolution GRIP core (Summit Greenland). The current interglacial - the Holocene – is at the far left and an ellipse indicates the most recent 1000 period. This last millennium is indicated in the bottom series which - conversely to the preceding - shows the present is on the right, the past on the left. This is a multiproxy temperature estimate from [*Moberg et al.*, 2005], the ellipse (right) the industrial epoch of global warming; not all of this variability is natural in origin. The ellipses, arrows and numbers indicate the parts of the time axis and zoom factor needed to go from one series to the next. This figure is continued in fig. 1c.
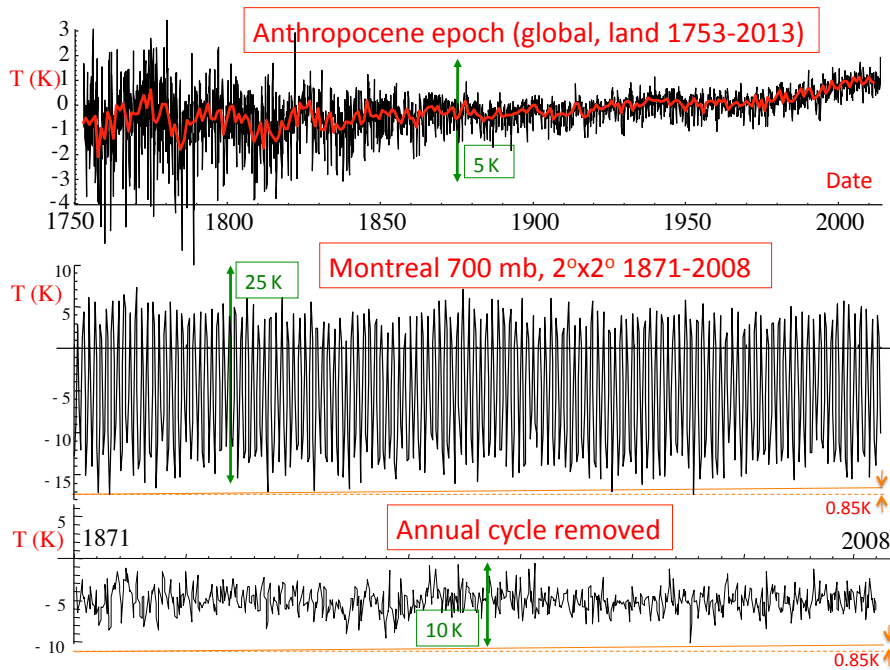
Fig. 1c: The top series the longest available instrumentally based global temperature estimates (monthly, land only, 3129 values, 1753-2013, [*Rohde et al.*, 2013]), the red is the annual average temperature. The data go back to 1753 but due to the very large uncertainties at the early dates (due to limited availability), the thickness of the zigzagging at the far left is large. This covers the epoch of the industrial revolution; the anthropocene, the geological period strongly influenced by humans.

Starting in 1871, reanalysis data at 2°x2°, 6 hour resolution is available from the 20th C reanalysis ([*Compo et al.*, 2011]); data at 700 mb are shown. There were over 2x10^5 values, we averaged so as to only display 720 points (the resolution displayed here is thus about 3 months). The middle series shows the raw data that includes the dominant annual cycle; the bottom series is the same but with this removed.

We also show for reference an estimate the amplitude of the anthropogenic change (from [*Lovejoy*, 2014c] close to the IPCC AR4 estimate); for the global change since 1880, it is ≈0.85 K. For the land only (top series [*Rohde et al.*, 2013]), the estimate is 1.5 K.
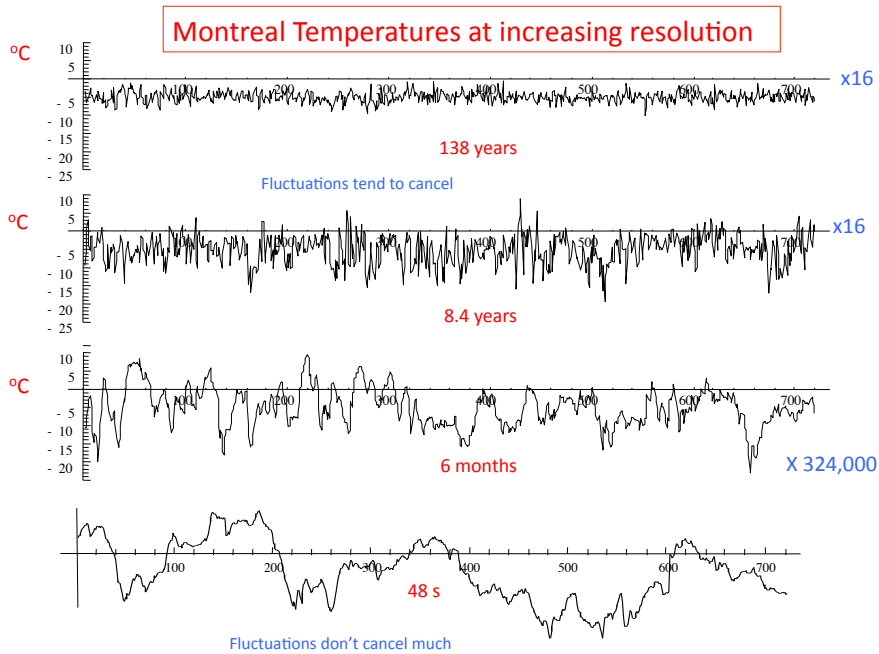
Fig. 1d: The upper left is the same as the lower series in fig. 1c. We successively take the left sixteenth of the series and blow it up by a factor of 16, retaining 720 points at each step until we get to the 6 hour resolution series (third from top), the total length of each series is indicated in red. The bottom series is also from Montreal, but from a millimetre sized thermistor on the roof of the McGill physics building at 0.067 s resolution. The temperature scale is the same for all the series except the bottom one. Higher resolution data would show that the variability continues for at least another 2 orders of magnitude to kHz scales.

Starting at the lower left we see that – as for the Epica series (fig. 1a, top row) – that the temperature appears to wander like a drunkard's walk with temperature differences ΔT = $T(t+\Delta t)$-$T(t)$ tending to grow with time intervals Δt. This character is still apparent at the next (6 hour resolution, lower left) - at least for intervals as long as 10-20% of the series length (i.e. up to 10-20 days long). As we move upwards to longer and longer resolutions to the series indicated 8.5 years (which is at 4 day resolution), notice that the overall variation of the series doesn't change much (i.e. the rough range between the maximum and minimum is nearly independent of the resolution).
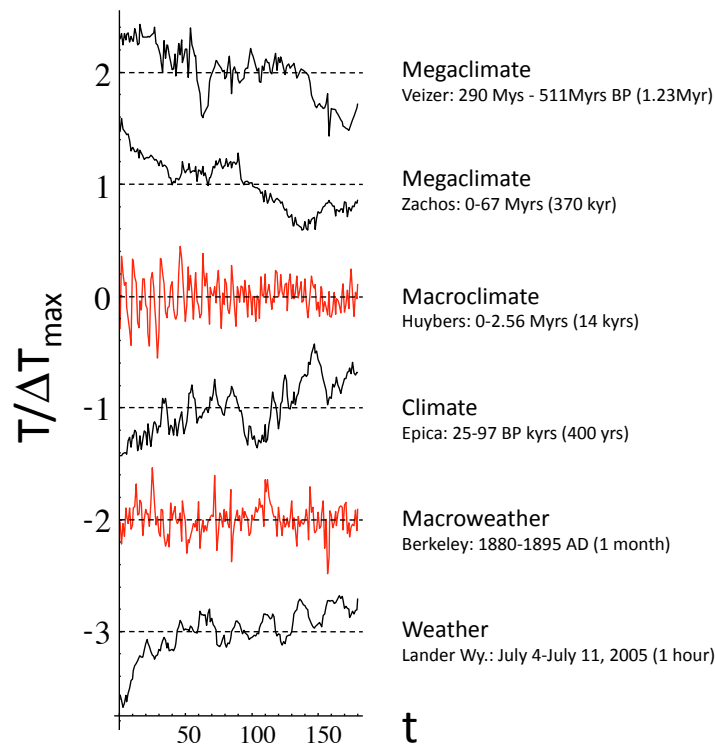
Fig. 1e: Representative series from each of the five scaling regimes taken from fig. 1a-d with the addition of the hourly surface temperatures from Lander Wyoming, (bottom, detrended daily and annually). The Berkeley series was taken from a fairly well estimated period before significant anthropogenic effects and was annually detrended. The Veizer series was taken over a particularly data rich epoch, but there are still traces of the interpolation needed to produce a series at a uniform resolution. In order to fairly contrast their appearances, each series had the same number of points (180) and was normalized by its overall range (the maximum minus the minimum), and each series was offset by 1K in the vertical for clarity. The resolutions were adjusted so that as much as possible, the smallest scale was at the inner scale of the regime indicated. The series resolutions were 1 hour, 1 month, 400 years, 14 kyrs, 370 kyrs and 1.23 Myrs bottom to top respectively. In the macroclimate regime, the inner scale was a bit too small and the series length a bit too long. The resulting megaclimate regime influence on the low frequencies was therefore removed using a linear trend of 0.25 $\delta^{18}O/Myr$. The resolutions and time periods are indicated next to the curves. The black curves have $H>0$, the red, $H<0$, see the parameter estimates in table 2.

From top to bottom the ranges used for normalizing are: 10.1, 4.59, 1.61 (Veizer, Zachos, Huybers respectively, all $\delta^{18}O$), 6.87 K, 2.50 K, 25 K (Epica, Berkeley, Lander).
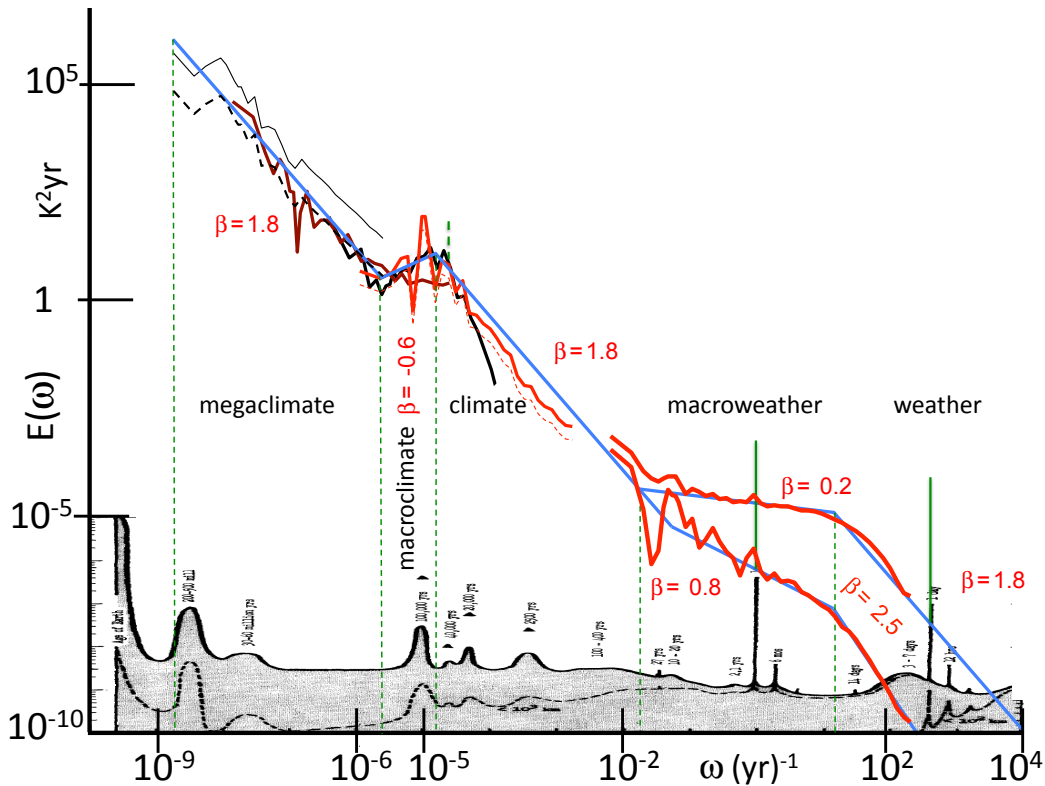
Fig. 2a: A comparison of Mitchell's relative scale, "educated guess" of the spectrum (bottom, [*Mitchell*, 1976]) with modern evidence from spectra of a selection of the series displayed in fig. 1 (the plot is log-log). There are three sets of red lines; on the far right, the spectra from the 1871-2008 20CR (at daily resolution) quantifies the difference between the globally averaged temperature (bottom) and local averages ($2°x2°$, top). This figure has been faithfully reproduced many times.

The upper left red curve is from the calibrated Epica Antarctic core (interpolated to 276 yrs resolution). All the spectra were averaged over logarithmically spaced frequency intervals (10 per order of magnitude), thus "smearing out" the daily and annual spectral "spikes". These spikes have been re-introduced without this averaging, and are indicated by green spikes above the red daily resolution curves. Using the daily resolution data, the annual cycle is a factor ≈ 1000 above the continuum, whereas using hourly resolution data (from the Lander series, fig. 4a), the daily spike is a factor ≈ 3000 above the background. Also shown is the other striking narrow spectral spike at $(41 kyrs)^{-1}$ (obliquity; $\approx$ a factor 10 above the continuum), this is shown in dashed green since it is only apparent in the Huyber series over the period 0.8 - 2.56 Myr BP.

At the upper left, the one brown curve and two black curves are $\delta^{18}O$ spectra from the benthic (i.e. ocean sediment) assemblages, the rightmost black is the Huybers series (at 10 *kyr* resolution), the middle (brown), is the Zachos series (interpolated to 18 *kyrs*), the leftmost (black) is Veizer series (interpolated to 185 *kyrs*). See [*Lovejoy*, 2014b] for more details.

The blue lines have slopes indicating the scaling behaviours ($E(\omega) \approx \omega^{-\beta}$) deduced from the real space Haar analyses (fig. 7). The scaling exponents β are related to the slopes in figure 7 ($\xi(2)/2$; see table 2) by $\beta = 1+\xi(2)$. The thin dashed green lines show the transition frequencies deduced from the spectra; these are at $(20 \text{ days})^{-1}$, $(50 \text{ yrs})^{-1}$, $(80 \text{kyrs})^{-1}$, and $(500 \text{ kyrs})^{-1}$ close to those deduced in real space in fig. 7.
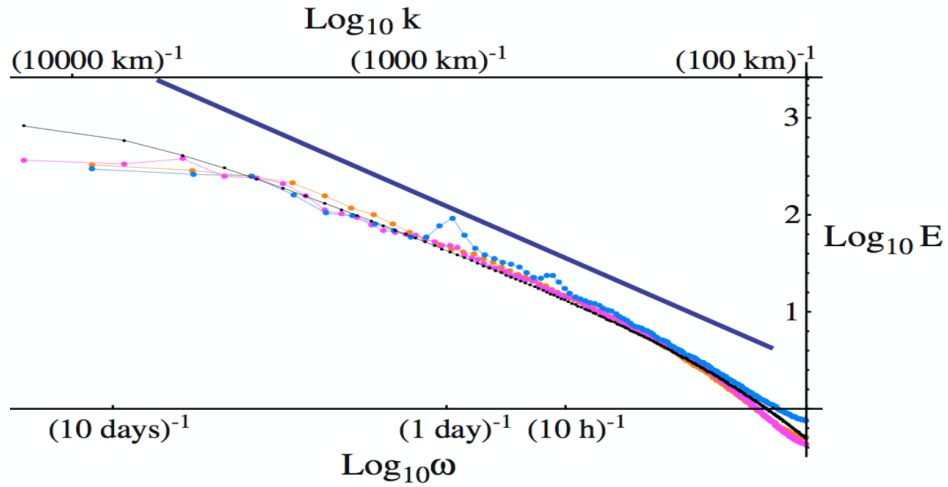

Fig. 2b. 1D spectra of MTSAT thermal IR radiances; 1440 consecutive hourly images at 30 km we used and covered a region over 13,000 km across centered over the tropical Pacific. In black: the theoretical spectrum using parameters estimated in [*Pinel et al.*, 2014] and taking into account the finite space – time sampling volume. The spectra are $E_x(k_x) \approx k_x^{-\beta_x}$, $E_y(k_y) \approx k_y^{-\beta_y}$, $E_t(\omega) \approx \omega^{-\beta_t}$ with $\beta_x \approx \beta_y \approx \beta_t \approx 1.4\pm0.1$; $s \approx 3.4\pm0.1$. The straight line is a reference line with slope −1.5 (blue). Pink, zonal spectrum; orange, the meridional spectrum; blue (with the diurnal spike and harmonic prominent) is the temporal spectrum. Reproduced from [*Pinel et al.*, 2014]. The full (horizontal space - time spectrum thus has the scaling symmetry $P\left(\lambda k_x, \lambda k_y, \lambda \omega\right) = \lambda^{-s} P\left(k_x, k_y, \omega\right)$ where P is the space-time spectral density.

2. Turbulence, scaling, multifractals, emergent laws
2.1 Fluctuations

The undisputed father of GCM's is Lewis F. Richardson, who was the first to write down the modern complete (closed) set of nonlinear, partial differential equations for the evolution of the atmosphere, publishing – at his own expense - the seminal "Weather prediction by numerical process" [*Richardson*, 1922]. In addition,

he spent six weeks with pencil and paper numerically integrating the equations to estimate the pressure change at a single grid point (due to numerical "initialization" issues that were not resolved until the 1970's, he was off by a factor of 100, see the interesting history by [*Lynch*, 2006]). Realizing its importance, Richardson proposed the creation of a "forecasting factory" for numerical weather prediction involving tens of thousands of human computers.

Richardson was more than 30 years ahead of his time - the first numerical weather model was in 1956 – and is therefore rightly revered by meteorologists as a pioneer of numerical weather prediction. However, he is also revered by the (quite different) turbulence community for proposing the first law of turbulence: the "Richardson 4/3 law" of turbulent diffusion [*Richardson*, 1926] which is the precursor of Kolmolgorov's famous law [*Kolmogorov*, 1941]. More recently, he has been recognized as the grandfather of cascade models of turbulence. Indeed, hidden in the middle of his book, Richardson slyly inserted the now iconic poem "Big whorls have little whirls that feed on their velocity and little whorls have smaller whorls and so on to viscosity (in the molecular sense)".

Like the other classical turbulence theorists – Kolmogorov, Obhukhov, Monin, Corrsin, Bolgiano to name a few – Richardson believed that at high enough levels of nonlinearity (quantified by the Reynold's number, the ratio of the typical nonlinear to linear terms in the equations) that new laws would emerge. Although the laws of continuum mechanics are deterministic, the emergent higher level turbulence laws are statistical and govern the behaviour of huge numbers of eddies ("whorls", structures). The situation is analogous to the higher level laws of thermodynamics which "emerge" from the lower level (more basic) laws of statistical mechanics in the thermodynamic limit i.e. for large numbers of degrees of freedom. Note that from a mathematical point of view, if one starts with the simplest system of fluid equations - the incompressible Navier-Stokes equations - then the behaviour in the high Reynold's number limit – "fully developed turbulence" is an open mathematical problem so that – even seventy years after Kolomogorov's law was proposed - there is no mathematically rigorous derivation of any of the proposed laws of turbulence (there are however many, many physical arguments).

The classical turbulence laws can be expressed in the form:

$$\Delta I\left(\underline{\Delta r}\right) = \varphi\left|\underline{\Delta r}\right|^{H} \tag{1}$$

Where $\Delta I\left(\underline{\Delta r}\right)$ is the fluctuation in the quantity I over a vector displacement $\underline{\Delta r}$, φ is a driving, turbulent flux, the vertical lines represent the norm of the vector and H is the fluctuation exponent. The most famous example of eq. 1 is the Kolmogorov law which is recovered by taking $I = v = $ a velocity component, $H = 1/3$ and $\varphi = \varepsilon^{1/3}$ where ε is the flux of energy (per mass) from large to small scales (strictly speaking it is a Fourier space flux). Notice that the form of eq. 1 and the exponent H are scale invariant: they don't change under isotropic scale changes ("zooms") i.e. when $\underline{\Delta r} \rightarrow \lambda\underline{\Delta r}$ where λ is a scale ratio. ΔI changes in a power law way and is said to be "scaling".

Two key aspects of the classical laws make them poor approximations to the atmosphere. The first is that the flux φ was considered to be fairly homogeneous

(uniform), at most having statistics that were no more variable than Gaussian. This is unrealistic since even a cursory consideration of the weather indicates that most of the atmospheric fluxes occur in only small fractions of the available space, especially in storms, even in their centres: the turbulence is highly *intermittent*. When applying equation 1 to the atmosphere, the second limitation is implicit in the use of the vector norm in eq. 1 to quantify the scale of Δr. Since the norm is independent of orientation, it implies that the laws are isotropic, whereas the atmosphere is anisotropic, in particular it is highly stratified: gravity strongly imposes a preferred direction.

2.2 Intermittency

Starting in the 1960's, intermittency was explicitly modelled with the help of Richardson - inspired cascade processes. One first assumes that at large scales, the fluid is stirred in a quasi-steady manner (in the atmosphere by the solar gradient between equator and poles). Since the corresponding energy flux (energy per mass, per time from larges scales to small) is exactly conserved by the nonlinear terms, the latter act to break large eddies up into "daughter eddies", transferring their energy fluxes to smaller and smaller scales until eventually (in the atmosphere at scales of less than a millimetre), viscosity dissipates the energy as heat. Cascade models are phenomenological models of this process; the original versions were "toy" models (believed to capture the basic physics while being relatively simple to analyse) in which the parent eddies are large cubes and the daughters are subcubes, with half the parent diameter (see fig. 3a for a schematic). For each daughter, one flips a coin to decide how the parent energy flux will be multiplicatively modulated over the daughter. In the simplest (fractal) "beta model" ([*Novikov and Stewart*, 1964], [*Mandelbrot*, 1974], [*Frisch et al.*, 1978] the daughters occasionally (with well defined probability) receive zero flux, they are "dead", the others have their fluxes multiplicatively boosted just enough to (on average) conserve the flux. If the process is slightly altered such that rather than receiving a boost or a decrease to zero flux, the alternative is instead a boost or a decrease to a finite positive flux (the "alpha model", [*Schertzer and Lovejoy*, 1985]) then rather than just black or white (active/inactive, dead or alive), there will be intermediate levels of activity, each one concentrated on a different fractal set with a different fractal dimension: the result is a self-similar multifractal (see fig. 2a which is a continuous in scale cascade).

A technical point that is often missed is that there are two variants of the cascade conservation: canonical and microcanonical. The former was described above: at each step in the cascade, the only constraint is that the probabilities of boosts and decreases are such that on average they equal 1 so that on average the fluxes neither increase nor decrease as the cascade proceeds from large to small scales. The alternative, "microcanonical" conservation is much more strict, it requires that at each step of the cascade (i.e. everywhere in space, at every resolution, scale), the multipliers sum exactly to unity (not only on a statistical average). Interestingly, the two state version of this was first discovered by [*de Wijs*, 1951] as model for the distribution of ores in the earth's crust. It was rediscovered by [*Meneveau and Sreenivasan*, 1987] who baptised it the "p-model".

Microcanonical models are still popular because they simplify the cascade mathematics: by construction, if spatially averaged over a microcanonical cascade step, then the small scales are completely averaged out, one is simply left with a low resolution cascade, one constructed with fewer steps. However, this is not true in the case of canonical conservation, indeed, it turns out that if we spatially average a canonical cascade taken to its small scale limit (i.e. after an infinite number of cascade steps), that the small scale activity is not only still present, but for moments exceeding a critical value q_D, that they dominate the statistics and lead to the phenomenon of divergence of statistical moments ([*Mandelbrot*, 1974], [*Schertzer and Lovejoy*, 1987]). This means that the extreme tails of the probability distributions are power laws, a fact we exploit below for statistical testing the natural warming hypothesis.

To make this more precise, consider a multifractal cascade developed over a range of scales from L to l (ratio λ), the statistics may be characterized by considering the various (q^{th} order) statistical moments:

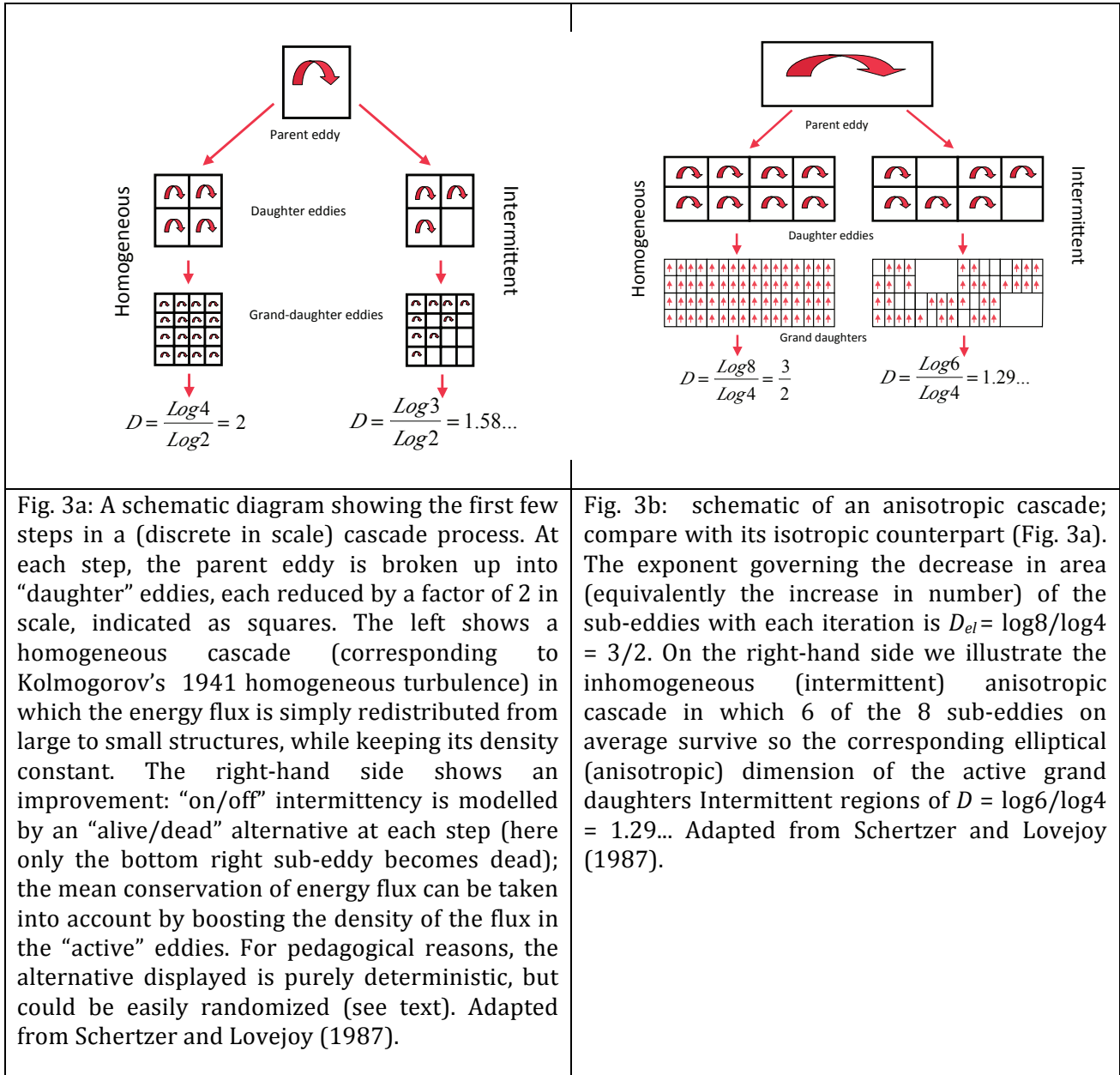$$\left\langle \varphi_\lambda^q \right\rangle = \lambda^{K(q)}; \quad \lambda = L/l \tag{2}$$

where "$< >$" indicates statistical (ensemble) averaging and $K(q)$ is a convex function of the order of moment q. Low order moments (small q) will be dominated by the numerous small fluctuations, high order moments (large q) will be dominated by rare large fluctuations; $K(q)$ generally gives a complete statistical characterization of the cascade at all scales and all intensities. Physically, L is the largest (outer) scale of the cascade, l is the smallest (inner, dissipation) scale. Since the mean flux ($q = 1$) is conserved from scale to scale, $K(1) = 0$ and in addition, for a nonzero cascade, $K(0) = 0$ also. Fig. 4 gives an example of the temperature statistics relevant to both weather and climate. For the general canonical cascades discussed above, the high order statistical moments of the spatially integrated fluxes diverge: $K(q) \to \infty$ for $q > q_D$ where q_D is a critical order typically about 5 - 7 for turbulence (see the review [*Lovejoy and Schertzer*, 2013], ch. 5). A divergence of moments of order $q \geq q_D$ is mathematically equivalent to a power law tail on the probability distribution, hence:
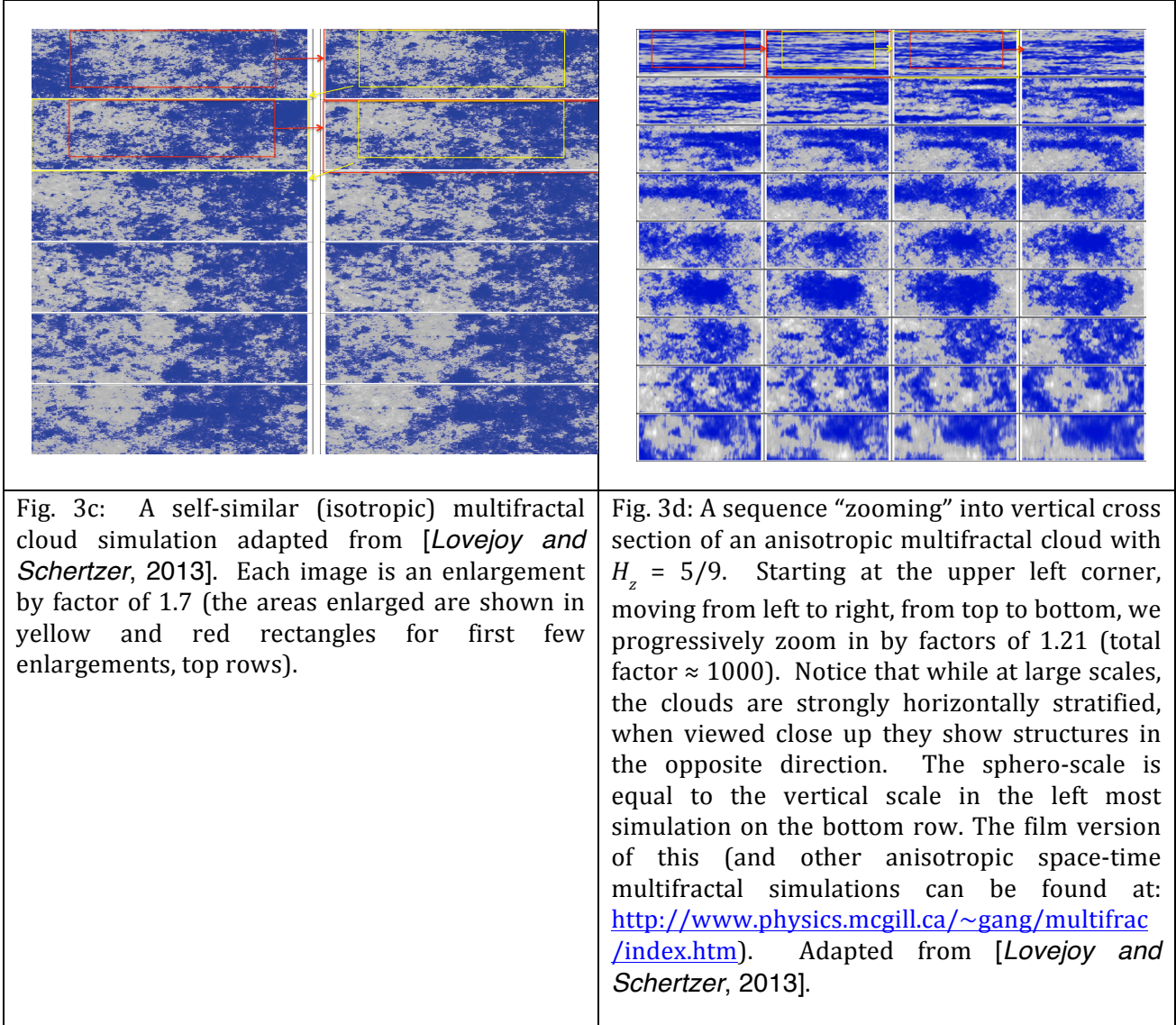
$$\Pr(\Delta I > s) \approx s^{-q_D}; \quad s \gg 1 \tag{3}$$

where "*Pr*" indicates "probability" and s is a threshold. If a random variable ΔI follows eq. 3, then extreme values occur much more often than for classical (exponentially bounded) distributions such as the Gaussian.

In [*Taleb*, 2010], the author popularized the expression "Black Swans", originally to designate events that are not only unexpected because they are rare, but to events that are *epistemologically unexpected* in the sense that they are totally outside the ken of the reigning view, ideology or theory. Inspired by Mandelbrot's use of Levy distributions (which have power law tails but with Levy index $\alpha = q_D$ restricted to values below 2 - the result of additive, not multiplicative processes), Taleb goes on to refer to the corresponding extreme Levy events as "Grey Swans". He justifies this since extreme Levy events can be anticipated but only on the basis of unconventional (non Gaussian) theory. The cascade extension from Levy to more general power law tails (i.e. with $q_D > 2$) should therefore rightly also be referred to

as "Grey Swans", but this term never stuck, hence the term "Black swans" is now often used more generally for any power law extremes (eq. 3).
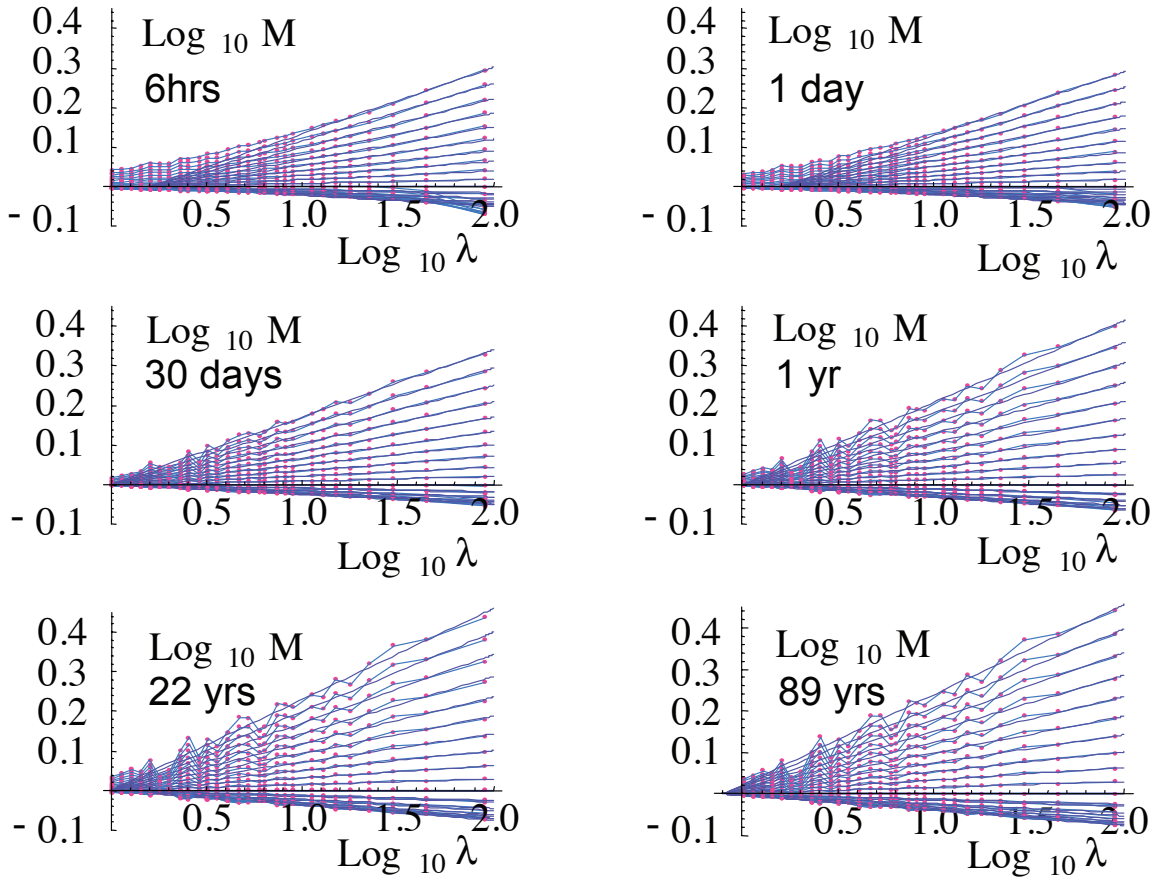

Fig. 3a: A schematic diagram showing the first few steps in a (discrete in scale) cascade process. At each step, the parent eddy is broken up into "daughter" eddies, each reduced by a factor of 2 in scale, indicated as squares. The left shows a homogeneous cascade (corresponding to Kolmogorov's 1941 homogeneous turbulence) in which the energy flux is simply redistributed from large to small structures, while keeping its density constant. The right-hand side shows an improvement: "on/off" intermittency is modelled by an "alive/dead" alternative at each step (here only the bottom right sub-eddy becomes dead); the mean conservation of energy flux can be taken into account by boosting the density of the flux in the "active" eddies. For pedagogical reasons, the alternative displayed is purely deterministic, but could be easily randomized (see text). Adapted from Schertzer and Lovejoy (1987).

Fig. 3b: schematic of an anisotropic cascade; compare with its isotropic counterpart (Fig. 3a). The exponent governing the decrease in area (equivalently the increase in number) of the sub-eddies with each iteration is D_{el} = log8/log4 = 3/2. On the right-hand side we illustrate the inhomogeneous (intermittent) anisotropic cascade in which 6 of the 8 sub-eddies on average survive so the corresponding elliptical (anisotropic) dimension of the active grand daughters Intermittent regions of D = log6/log4 = 1.29... Adapted from Schertzer and Lovejoy (1987).

Fig. 3c: A self-similar (isotropic) multifractal cloud simulation adapted from [*Lovejoy and Schertzer*, 2013]. Each image is an enlargement by factor of 1.7 (the areas enlarged are shown in yellow and red rectangles for first few enlargements, top rows).

Fig. 3d: A sequence "zooming" into vertical cross section of an anisotropic multifractal cloud with H_z = 5/9. Starting at the upper left corner, moving from left to right, from top to bottom, we progressively zoom in by factors of 1.21 (total factor ≈ 1000). Notice that while at large scales, the clouds are strongly horizontally stratified, when viewed close up they show structures in the opposite direction. The sphero-scale is equal to the vertical scale in the left most simulation on the bottom row. The film version of this (and other anisotropic space-time multifractal simulations can be found at: http://www.physics.mcgill.ca/~gang/multifrac/index.htm). Adapted from [*Lovejoy and Schertzer*, 2013].

Fig. 4: The verification of eq. 2 on temperature data from the twentieth century reanalysis (20CR, [*Compo et al.*, 2011]) from zonal (east-west) transects at 45°N using data at 6 hour temporal and 2° spatial resolution, from 1871-2008. $M = \left\langle \varphi_\lambda^q \right\rangle$ (eq. 2) with moments $q = 2$, 1.8, 1.6,…0.2 corresponding to the points and regression lines from top to bottom. Each graph shows the spatial statistics of the fluxes φ_λ estimated at various temporal averaging scales ranging from 6 hours to 89 years by taking the absolute finite difference of the temperature temporally averaged over the indicated duration. Eq. 2 predicts that on such log-log plots, that the lines for different moments q will converge to the "effective" outer scale of the process, i.e. the scale at which the process must start if it is to explain the observed variability at the smaller scales (to the right). In this analysis, the largest scale ($\lambda =1$) corresponds to the largest distance along the 45°N latitude line i.e. 14100km. It is approximately, but not exactly the outer scale. We can see that for all the different time scales, the spatial statistic well obey the predictions of multiplicative cascade theories. The large spatial intermittency – characterized near the mean by the codimension of the mean $C_1 = K'(1)$ slowly increases from 0.095 (6 hours, the weather regime) to 0.13 (89 years, the climate regime) implied by these plots is the statistical expression of the existence of various climate zones. This figure is reproduced from [*Lovejoy and Schertzer*, 2013].

We now illustrate such black swan extremes with an example from the climate that will be relevant later. First consider the probability distribution of daily temperature changes from a single station (fig. 5a). The figure shows the cumulative distributions of the temperature changes accumulating from the largest, not smallest value (it is one minus the usual cumulative distribution function). We see that for both positive and negative temperature changes that the distribution has far more extreme events than would be expected from the classical Gaussian distribution, indeed, the extremes are 7 standard deviation events corresponding to probabilities of less than 10^{-20}. On the other hand, the data closely follow a power law with exponent $q_D \approx 5$. Moving to longer times (fig. 5b), we see the same type of behaviour, even in paleotemperatures. For the latter, modern data (far right) allow the tails to be examined more closely, yielding a more convincing result, again with $q_D \approx 5$. Note that taking differences over longer time scales shifts the tails by a constant factor corresponding to $H = 0.4$ (the left two plots in fig. 5b).

Finally, to evaluate the statistics of natural temperature changes - and to avoid biases due to anthropogenic effects - consider global scale pre-industrial (1500-1900) temperatures. In the pre-industrial period, global scale temperatures can be estimated using "multiproxy" reconstructions. As the name suggests, these statistically combine data from diverse sources ("proxies") typically including tree rings, ice cores, lake sediments in order to estimate the temperature in the absence of instruments, fig. 5c shows the corresponding distributions for differences of 1, 2, 4, ... 64 years. Figure 5c shows that eq. 3 is a reasonable approximation to the tails of the (pre-industrial) distributions of temperature changes. This result will be used below for estimating the probability that the industrial warming is no more than a giant natural fluctuation.
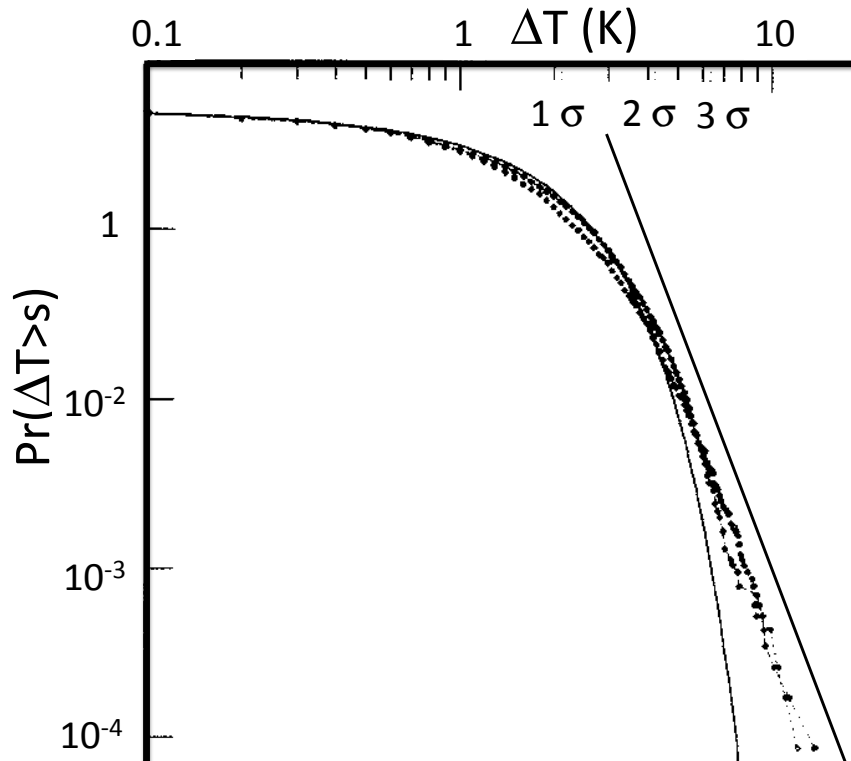
Fig. 5a: The probability distribution of daily temperature differences in daily mean temperatures from Macon France for the period 1949-1979 (10,957 days). Positive and negative differences are shown as separate curves. A best fit Gaussian is shown for reference indicating that the extreme fluctuations correspond to more than 7 standard deviations, for a Gaussian this has a probability of 10^{-20}. The straight reference line (added) has a slope of $-q_D$ with $q_D = 5$. Adapted from [Ladoy et al., 1991].
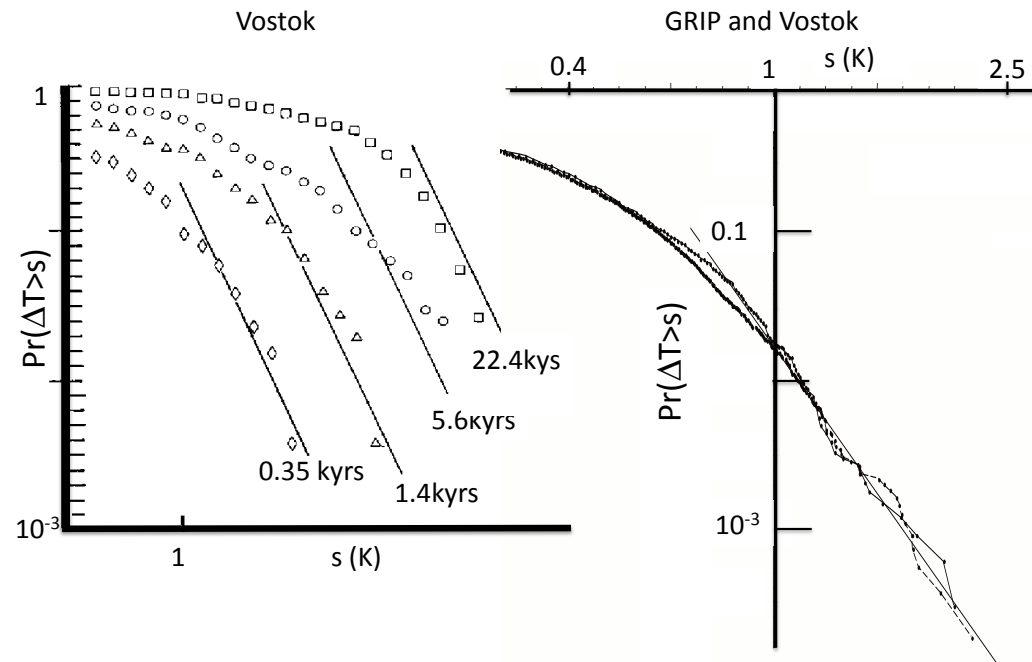
Fig. 5b: Probability distributions of paleotemperature changes for Vostok (left), (Antarctica, paleo temperatures from ^{18}O proxies reproduced from [*Lovejoy and Schertzer*, 1986]) and a modern comparison of GRIP (Greenland) and Vostok right, reproduced from [*Lovejoy and Schertzer*, 2013].   The graphs differ not only due to the much improved sampling density of the more modern data, but also, the rightmost graph is at constant depth intervals (0.55 m for GRIP and 1m for Vostok); this avoids issues of uncertain chronologies. In all cases, the straight reference lines indicate extreme s^{-q_D} behaviour with $q_D = 5$ where s is a temperature change. The reference lines in the left graph are spaced $H\log_{10}4$ apart with scaling exponent $H = 0.4$.
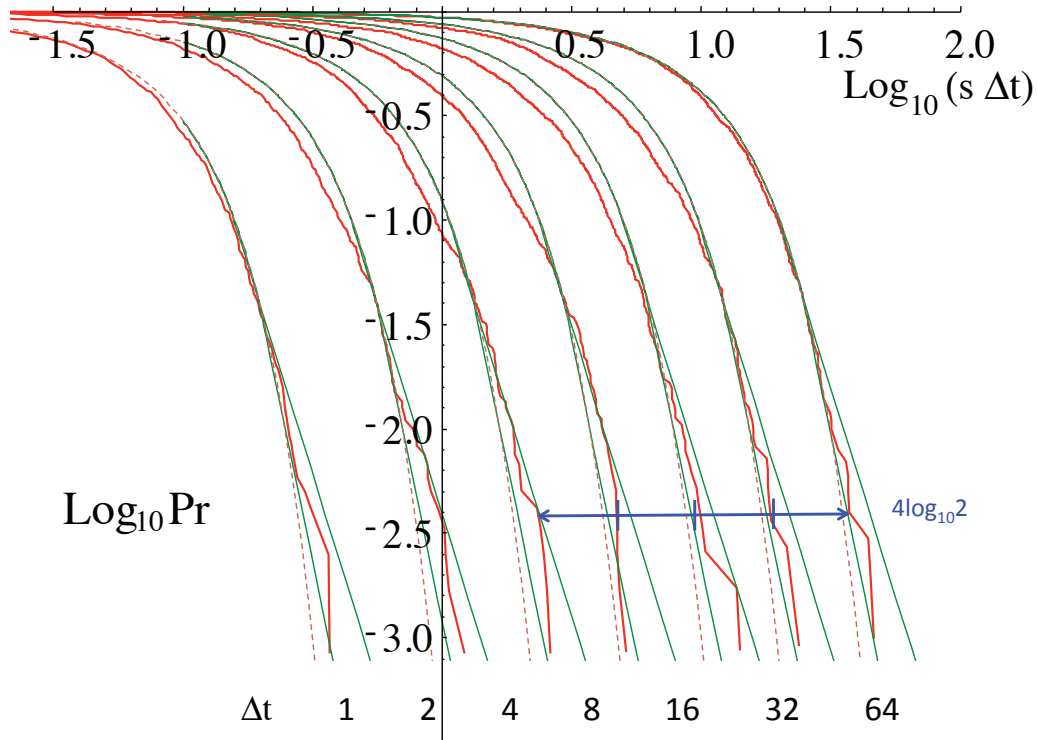
Fig. 5c: The total probability of random absolute pre 1900 temperature differences exceeding a threshold s (in K), using three multiproxies to increase the sample size (the distribution are very similar in form for each of the multiproxies). To avoid excessive overlapping, the latter were compensated by multiplying by the lag Δt (in years, shifting the curves to the right successively by $\log_{10}2 \approx 0.3$), the data are the pooled annual resolution multiproxies from 1500-1900. The blue double headed arrow shows the displacement expected if the difference amplitudes were constant for 4 octaves in time scale (corresponding to $H = 0$ for differences, the standard deviations each octave is indicated by a vertical tick mark on the arrow). The (dashed) reference curves are Gaussians with the corresponding standard deviations and with (thin, straight) tails ($Pr \approx <3\%$) corresponding to bounding s^{-4} and s^{-6} behaviours (reproduced from [*Lovejoy*, 2014c]).

3. The climate
3.1 GCM's: the climate as a boundary value problem

At first, GCM's were weather models designed to model the atmosphere over periods of days: the slowly varying ocean was taken as a fixed lower boundary condition. In order to extend GCM's for modelling the longer time scales associated with the climate, at the very least they had to be coupled to ocean models; modern GCM's are also coupled to cryosphere and carbon cycle models. Changing land use and atmospheric composition (CO_2, methane, aerosol concentrations), are taken into

account as changes in boundary condition as are the external "forcings" consisting of variable solar output and volcanic eruptions.

So how do GCM's work? Due to their sensitive dependence on initial conditions (the "butterfly effect", deterministic chaos), errors grow quickly (due to the scaling, algebraically, not exponentially fast [*Schertzer and Lovejoy*, 2004]) so that for planetary sized structures there are predictability limits of the order of ten days. This means that even with all the correct couplings, that the exact state of the atmosphere cannot be well predicted beyond 10 days or so. Indeed, for climate modelling - and even though all the details are known to be wrong - GCM's are routinely integrated at 10 minute time steps for decades or hundreds of simulated years. However, it is hoped that the resulting fields will have the correct type of statistical variability and that any low frequency trends imposed by slowly varying boundary conditions (such as changing CO_2 concentrations) will change the averages (e.g. of the temperature) in a more or less realistic manner. Mathematically, whereas below ten days, GCM's consider atmospheric forecasting as an *initial* value problem, beyond this, it is considered to be a *boundary* value problem. Notice the irony: that in this "climate" prediction mode, the world's largest supercomputers (some with $\approx 10^6$ coprocessors) are essentially used as random number generators: generating no more than random "weather" so that the climate can be deduced by averaging almost all of it out. We return to this below.

3.2 Fluctuations and the fluctuation exponent

Surely, if scale invariance is a basic symmetry respected by the atmosphere and its models, then shouldn't it be used to categorize the different regimes of atmospheric dynamics? Indeed, since the horizontal velocity field is scaling out to planetary scales (see fig. 2b for the horizontal scaling of satellite radiances, for the wind, see [*Lovejoy et al.*, 2009], [*Pinel et al.*, 2012], or for a review, ch. 2 of [*Lovejoy and Schertzer*, 2013]), this can be used (at least dimensionally) to convert from space to time so that we find that the temporal fluctuations of the wind, temperature and other fields are scaling in time out to scales corresponding to the lifetime of planetary structures: about 10 days (this can be determined from first principles using the energy input from the sun, the size of the planet and the Kolmogorov law, [*Lovejoy and Schertzer*, 2010], [*Lovejoy et al.*, 2014]). Beyond this timescale, one is dealing with the statistics of structures over many lifetimes, new and quite different scaling regimes are established.

To understand the different regimes, let us simplify eq. 1 by considering only the mean fluctuations $\langle \Delta I \rangle$ over time intervals Δt:

$$\langle \Delta I \rangle = \langle \varphi \rangle \Delta t^H \tag{4}$$

Since $\langle \varphi \rangle \approx$ constant ($K(1) = 0$ in eq. 2), the mean behaviour of the fluctuations is determined by the exponent H. For example, when $H>0$, fluctuations tend to grow with the time interval, I will fluctuate like a drunkard's walk - indeed, the usual Brownian motion is the case $H = 1/2$ and φ is a Gaussian (with $K(q) = 0$). When $H<0$, on the contrary, fluctuations will tend to cancel each other out so that averaging over longer and longer intervals tends to converge. For this to be possible, the

notion of fluctuation needs to be appropriately defined with the help of wavelets. When $1>H>0$ using differences $\Delta I = I(t+\Delta t) - I(t)$ is sufficient ("poor man's wavelets"). However any process whose correlations decay with time Δt have average differences that increase with Δt so that when $H<0$ they cannot be used to estimate fluctuations (see eq. 4). In this case one could use "anomaly fluctuations" that are equal to the average over time Δt of the series after its overall mean has been removed. Anomaly fluctuations on the contrary, can only decrease with time scale and therefore cannot be used when $H>0$. A simple definition that combines the advantages of both – and is valid over the range $-1<H<1$ - is simply to use the difference between the average over the first and second halves of the interval: "Haar fluctuations" (the coefficients of Haar wavelets).

Since Haar fluctuations are equal to the difference fluctuation of the anomalies (or equivalently the anomaly fluctuation of the differences), they are easy to apply, and – most importantly - easy to interpret. In contrast, the popular Detrended Fluctuation Analysis [*Peng et al.*, 1994] defines fluctuations as the root mean square residual of a polynomial regression against the running sum of the process. Not only is this mathematically difficult to analyse (DFA fluctuations aren't wavelets) but - more importantly - the interpretation of the fluctuations is so opaque that in DFA papers the authors don't even both to indicate the units of the fluctuations on their scaling plots; they only use them to estimate scaling exponents! A final comment on the exponent H: it is denoted "*H*" in honour of Edwin Hurst but in general – unless the process is Gaussian - it is not the same as the Hurst (i.e. "R/S") exponent (e.g. for a standard random walk -Brownian motion - they both yield $H = \frac{1}{2}$).

In order to understand the *H* exponent, consider the simple (essentially pedagogical) fractal construction shown in fig. 6a, b, that – for want of a better name – we call the "*H* model" (when $1>H>0$ it is close to the "pulse in pulse" model, and has divergence of statistical moments of order $>1/H$, i.e. $\langle \Delta T^q \rangle \to \infty$ for $q>1/H$, see [*Lovejoy and Mandelbrot*, 1985]). To simulate a series with fluctuation exponent *H* over the unit interval, start with the basic fluctuation, the step function labelled "motif" in fig. 6a (top); the dashed line indicates the horizontal axis so that the left half is negative, the right half is (symmetrically) positive. To obtain the 2nd generation of the construction, compress the motif by a factor two in the horizontal and 2^{-H} in the vertical and place the result in the left half of the interval, then multiply it by a random sign. Finally, repeat with another random sign and place the result in the right half interval. The figure shows the result for signs +, - ; this defines the fluctuations at the corresponding reduced scale. Fig. 6b shows the result when this is iterated 8 times; the left column with $H>0$, the right column, $H<0$. The final fractal process is obtained by summing all the contributions. Notice that in the $H>0$ process, the fluctuations decrease with scale so that the process is dominated by the larger scales, conversely for the $H<0$ process. When $H<1$ the process has mean fluctuations $\langle \Delta T(\Delta t) \rangle \propto \Delta t^H$.
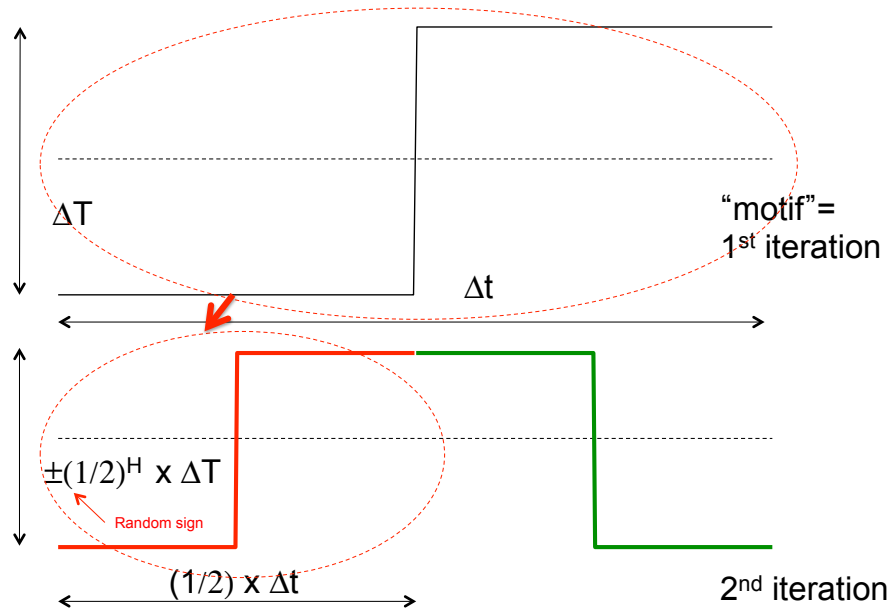
The fractal H model

Fig.6a: The first two steps in the construction of the fractal *H* model. To obtain the second row, the motif (i.e. a basic "fluctuation", top row) is reduced by a factor 2 in the horizontal and by 2^H in the vertical and then multiplied by a random sign, this is placed in the left hand half of the figure; the right hand half has the same shape but with another random sign.
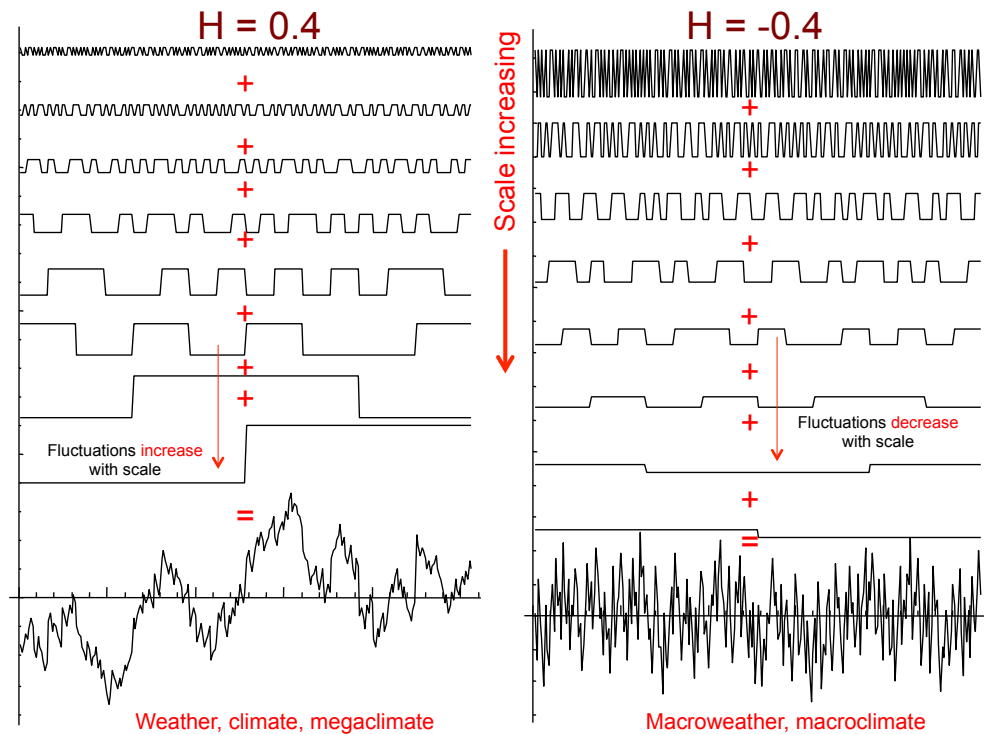
Fig.6b: The first 8 steps in the construction of the fractal *H* model with the sum, bottom series. In the left hand column we show the result for *H*>0, the right, *H*<0. In the *H*>0 case we see that the amplitude of the fluctuations decreases as we go to smaller scales whereas in the *H*<0 case, they increase.

3.3 What is the climate?

When the wind, temperature, humidity, pressure and other atmospheric variables are considered – whether empirically or from GCM's - or from appropriately generalized turbulent cascade processes – it is found that the transition at 5 - 10 days is universally observed to be from high frequency, growing (*H*>0) to low frequency, decreasing (*H*<0) behaviour with successive fluctuations tending to cancel each other out. At first sight, this would appear to validate the dictum "the climate is what you expect, the weather is what you get", ([*Heinlein*, 1973]) i.e. that the climate is a kind of average weather, which is obtained by averaging the weather over longer and longer time intervals. However, analysis of instrumental and paleo (i.e. proxy) data over decades, century and millennial scales shows that the *H*<0 convergence of averages ends at about 30 years (in the industrial epoch) and after about 100 years (pre-industrial), and that at longer scales, the averages begin to change again (i.e. with *H*>0). Fig. 7 is a composite showing the variability over nearly 13 orders of magnitude based on instrumental and paleo data (fig. 2a is the spectral equivalent using largely the same data). For example, the climate "normal" (defined as a thirty year average) itself tends to fluctuate becoming more and more variable right up to tens of thousands of years,

the ice ages.   Empirically, there are therefore three different regimes, not two, the intermediate regime – which is really a kind of "slow" weather is called "macroweather" with the term "climate" being reserved for the longer period (unstable) variations up to ice-age scales of ≈ 100 kyrs [*Lovejoy*, 2013]. If one uses paleo data to extend such analyses to very long scales (up the limit of reliable proxies, 540 Myrs corresponding to the Phanerozoic eon), then in addition one finds a narrow "macroclimate" regime from ≈100 kyrs to ≈1 Myr (with $H \approx$ -0.8) and then a "megaclimate" regime from with (with $H \approx 0.4$) from ≈1 Myr to (at least) 540 Myrs. Interestingly, this latter $H>0$ regime is associated with "unstable" drunkard's walk like behaviour, it is incompatible with Lovelock's Gaia ("living earth") hypothesis [*Lovelock*, 1995] that posits homeostasis (i.e. negative feedbacks that are strong enough to keep the temperature from wandering too far from optimal values for life; [*Lovejoy*, 2014b]).

The weather – macroweather - climate trichotomy is helpful for understanding anthropogenic warming: whereas in pre-industrial times, slowly acting natural processes eventually – at scales of a century or more - begin to dominate the (cancelling, diminishing) macroweather processes, in the industrial era, the anthropogenic effects are stronger than the natural climate processes and they begin to dominate macroweather after only about 30 years. This is important because it means that the (roughly century long) variation since 1880 is mostly due to anthropogenic not natural variability: it allows us to fairly accurately separate out the anthropogenic from the natural variations.
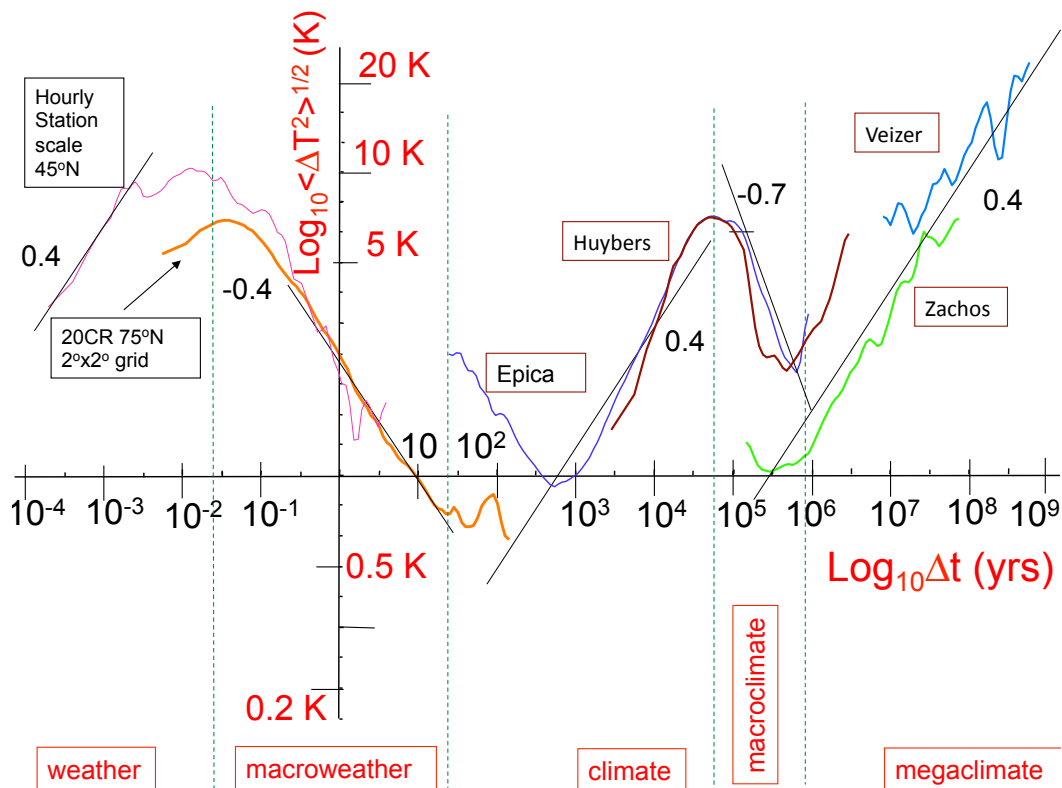
Fig. 7: This is a wide scale range composite series showing atmospheric variability over the range from 1 hour to 553 million years. The 5 regimes: weather, macroweather, climate, macroclimate and megaclimate are also indicated. The curves to tleft are instrumental, the Epica curve is from Antarctica (ice core isotope proxy), the Huybers, Zachos and Veizer curves are from ocean core isotope proxies. Adapted from [*Lovejoy*, 2014b].

3.4 Scaling and anthropogenic warming

The weather, macroweather and then climate picture is based on the corresponding scaling regimes, it characterizes the natural variability. For the industrial period which started roughly 125 years ago, the corresponding duration is still within (although perhaps not far from the limits) of the natural macroweather regime. This means that the probability of natural fluctuations can be safely extrapolated somewhat beyond the limits of the empirical probabilities in fig. 5c. These probabilities can thus be used to estimate how long we may expect to wait for various temperature changes – an estimate of "return times", taking into account the scaling of the probabilities of temperature changes combined with the expected behavior of the extremes (fig. 5c, eq. 3) to estimate the probability of extreme temperature fluctuations. If it is found that the probability of the observed change since 1880 ($\approx 0.9^\circ$ C, see below) is low enough (that the return times are long enough), then we can reject the hypothesis that the fluctuation was caused by natural variability.

The first step is to separate the anthropogenic and natural variability. From fluid mechanical, turbulent and GCM viewpoints, the main anthropogenic forcings (Green House Gases (GHG), aerosols and land use changes) affect the boundary conditions, not the *type* of variability (see above). These arguments support the separation:

$$T_{globe}(t) = T_{anthro}(t) + T_{nat}(t) \tag{5}$$

where T_{globe} is the global temperature anomaly and T_{anthro} and T_{nat} are the contributions of anthropogenic and natural processes. The justification for this is that the anthropogenic forcings since 1880 are of the order of 1 W/m^2 which is less than 1% of the mean solar forcing of 240 W/m^2. The next step is to note that anthropogenic effects are tightly correlated with global economic activity, this justifies using the (relatively well measured) industrial epoch CO_2 forcing, as a linear surrogate for all the anthropogenic forcings (see [*Lovejoy*, 2014c]):

$$T_{anthro}(t) = \lambda_{2xCO2,eff} \log_2 \left(\rho_{CO_2}(t) / \rho_{CO_2,pre} \right) \tag{6}$$

where $\lambda_{2xCO2,eff}$ is the "effective" sensitivity of the climate to a CO_2 doubling, ρ_{CO2} is the global mean CO_2 concentration and $\rho_{CO2,pre}$ is the pre-industrial value (277 ppm; the logarithmic form goes back to [*Arrhenius*, 1896]), it is a consequence of the saturation of the CO_2 absorption bands. Alternatively, we may use the "equivalent CO_2" concentration which results from the conversion of all the GHG and aerosols in CO_2 radiative equivalents. However, CO_2 and CO_{2eq} are very highly correlated (correlation >0.99) so that the residuals - the estimate of the natural variability is nearly the same in both cases. We could note that economic activity is more related to the emission rate rather than the CO_2 concentration (which depends on the

cumulative emissions). However, since economic growth has been roughly exponential - and the integral of an exponential is again an exponential - the two are roughly proportional which is all that we require. More generally, one may assume that the anthropogenic contribution to the temperature is related to the forcing by a non instantaneous but still linear, transfer function. Due to the scaling, this is expected to be a power law – at least above some inner scale, but again, we find that the residuals have nearly the same variability, (work in progress).

Unlike approaches that attempt to separate internal variability from the responses to external natural and anthropogenic forcings, T_{nat} includes any temperature variations that are not anthropogenic in origin, i.e. it includes both "internal" variability and (implicitly) the responses to solar and volcanic forcings which are external but still natural. Similarly, T_{anthro} includes the warming due to the other GHG's as well as the (difficult to estimate) aerosol cooling: $\lambda_{2xCO2,eff}$ is thus the "effective climate sensitivity". It is the sensitivity to the actual (historical) doubling of CO_2, it is thus conceptually distinct from the theoretical/model notions of "equilibrium" and "transient" sensitivity that have been empirically estimated elsewhere (although presumably it is closer to the latter than to the former). It is only the effective climate sensitivity that permits one to estimate the natural variability during the industrial epoch (as a residual).

Fig. 8 shows the results when eqs. 8, 9 are applied using global, annual ρ_{CO2} from 1880 – 2013 using the NASA GISS global temperature series [*Hansen et al.*, 2010]. Without sophisticated statistics, the linear trend is fairly convincing (correlation coefficient 0.94), and the effective sensitivity (the slope) $\lambda_{2xCO2,eff}$ =2.33±0.22 ºC/CO_2 doubling which is close to the equilibrium sensitivity 3±0.75 ºC/CO_2, (conceptually, it is closer to the somewhat smaller transient climate sensitivity). The total anthropogenic warming (the vertical range of the line in fig. 8a) is 0.87±0.11 ºC which is close to the IPCC AR5 estimate 0.85±0.20 ºC (uncertainties are one standard deviations).

While fig. 8a's linearity is impressive, are the residuals (fig. 8b) really reasonable estimates of T_{nat}? One answer is to note that the amplitude of the residues is quite small: ±0.109 ºC. Remarkably, it is virtually the same as the *errors* in GCM hindcasts of global temperatures at one year forecast horizons. For example, over the period 1983- 2004, using different GCM's, and bias correction techniques, [*Smith et al.*, 2007] and [*Laepple et al.*, 2008], obtained RMS hindcast errors of ±0.105 ºC and ±0.106 ºC. (A hindcast is the use of models to make forecasts for historical time periods. For example, we can use the data available at time t to make forecasts for time $t+\Delta t$. In a hindcast, $t+\Delta t$ is in the past, it is forecast using the data up to time t and verified with data from t to $t+\Delta t$). In other words, if all we knew was the global annual CO_2 concentration and the value of $\lambda_{2xCO2,eff}$, then (on average) we could already *predict* the next year's global temperature as well as the GCM's! Clearly, this T_{nat} must be close to the true natural variability. Unsurprisingly, this *unconditional* prediction (i.e. using no information about the actual global temperature series) can be improved even further by exploiting the stochastic climate "memory" to make *conditional* predictions.
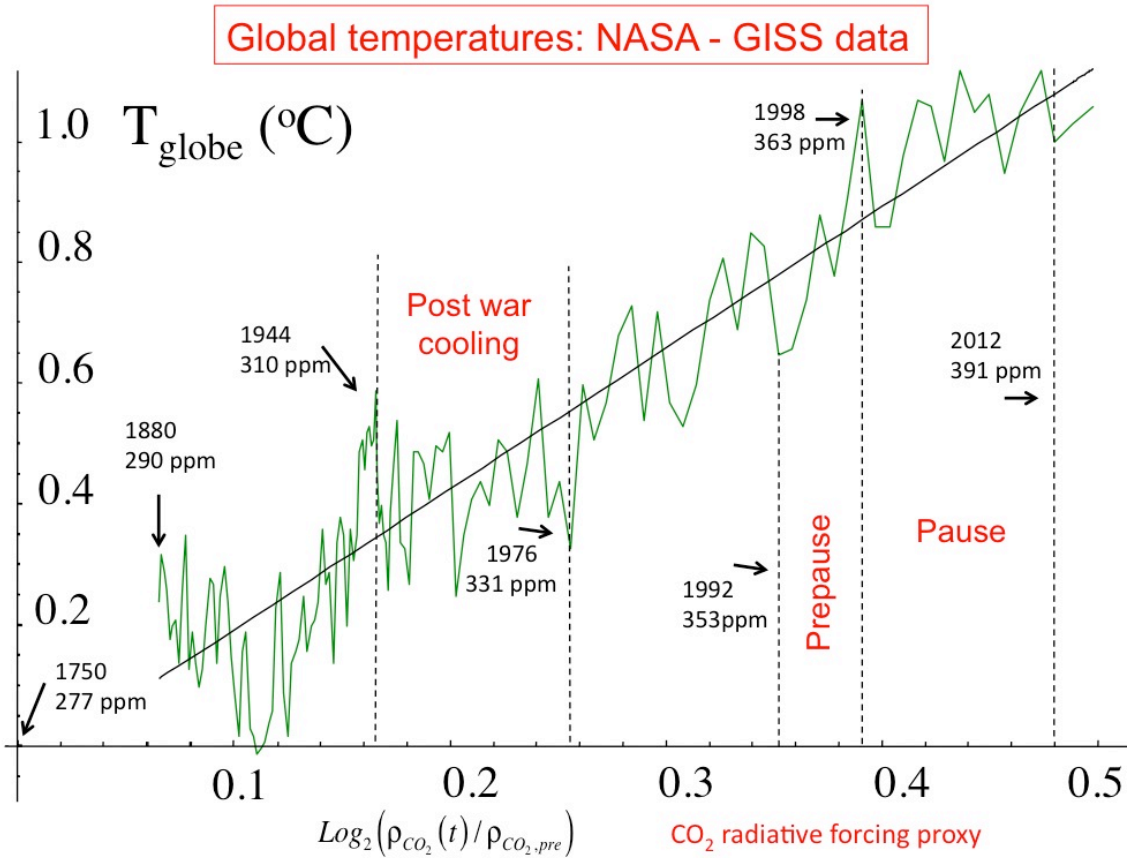
Fig. 8a: Global temperature anomalies (NASA, 1880-2013) as functions of radiative forcing using the CO_2 forcing as a linear surrogate. The line has a slope of 2.33° C per CO_2 doubling. Some of the dates and corresponding annually, globally averaged CO_2 concentrations are indicated for reference; the dashed vertical lines indicate the beginning and end of the events discussed in the text (1944, 1976, 1992, 1998). Adapted from [*Lovejoy*, 2014a].
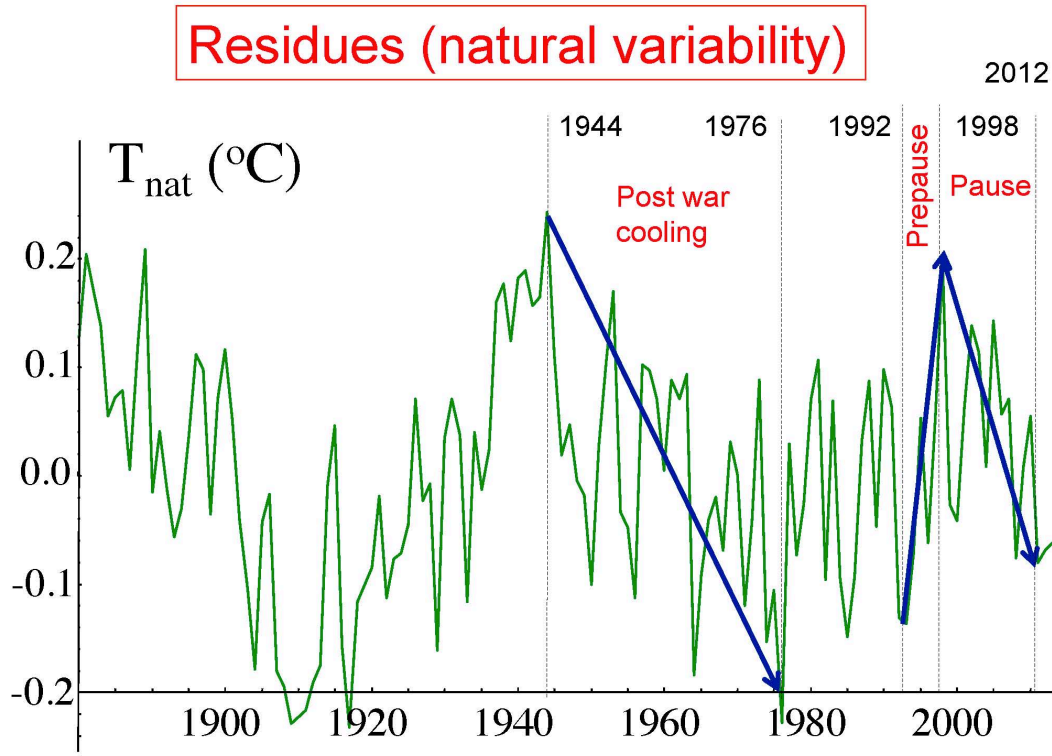
Fig. 8b: The residuals from the straight line in fig. 8a, these are the estimates of the natural variability. The vertical dashed lines are the same as in the previous. The arrows indicate the events discussed in the paper. Adapted from Fig. 1c, [*Lovejoy*, 2014a].

4. The Pause, hiatus, slowdown
4.1 Return periods

We have used (fig. 5c) the multiproxies to estimate the probabilities of natural temperature changes over various time intervals. For the theoretical reasons discussed above, over the empirically reliable range of time intervals (up to 64 years, fig. 5c), these distributions turn out to be nearly scale invariant (independent of the interval), and we can extend the estimates to intervals of 125 years. As discussed earlier, the scaling is associated with power law probabilities. These were used to bound the extreme 3% of the distributions (eq. 3): this takes into account the fact that the extreme changes occur much more frequently than the Gaussian distribution would allow. Finally, we can estimate the expected time interval between temperature changes of various magnitudes - their return periods (we ignore possible clustering of the extremes and estimate the return period as the inverse probabilities taken from fig. 5c). The return periods are shown in fig. 9 where we can see that the warming since 1880 is expected to occur naturally every 1000 to 20,000 years (if we use the traditional Gaussian – the red line - we find periods > 1 Myrs, the warming is nearly a five standard deviation event!). The

possibility that 1880 just happened to be part of such a natural warming can thus be dismissed at the 0.1% level. Similarly, the post war cooling – the largest event in the record since 1880 - should occur every 100 - 150 years. As expected, such an event occurs somewhere in the record (it happens to start in 1944). We can also consider the natural cooling of about 0.3 ᵒC since 1998 - needed to offset the anthropogenic warming over the period and account for the post 1998 flattening in fig. 8a. This "pause" or "hiatus" is much touted by climate skeptics as proof that the warming has stopped and is therefore not anthropogenic. However, from fig. 9, we see that such a cooling is expected to occur every 20 – 50 years, it is not so unusual.   Yet it becomes quite probable when it is noticed that it immediately follows an equally strong warming "prepause" event from 1992-1992 (figs. 8a,b) so that the natural cooling since 1998 is no more than a return to the mean behaviour. Indeed, from fig. 7 we see that it was only in 2012 that the temperature finally went below its long term (anthropogenic) trend.   Recently *Karl et al.* [2015] produced a temperature series with new ocean and other bias corrections. In this warmer series, the amplitude of the corresponding natural cooling is 0.09°C less than that shown in Figure 8b (i.e. about 0.2 instead of 0.3ᵒC). Since the return period for this smaller natural cooling is only about 10 years (fig. 9), decadal trends cannot (and did not) detect any statistically significant pause at all and the authors pronounced the pause nonexistent.
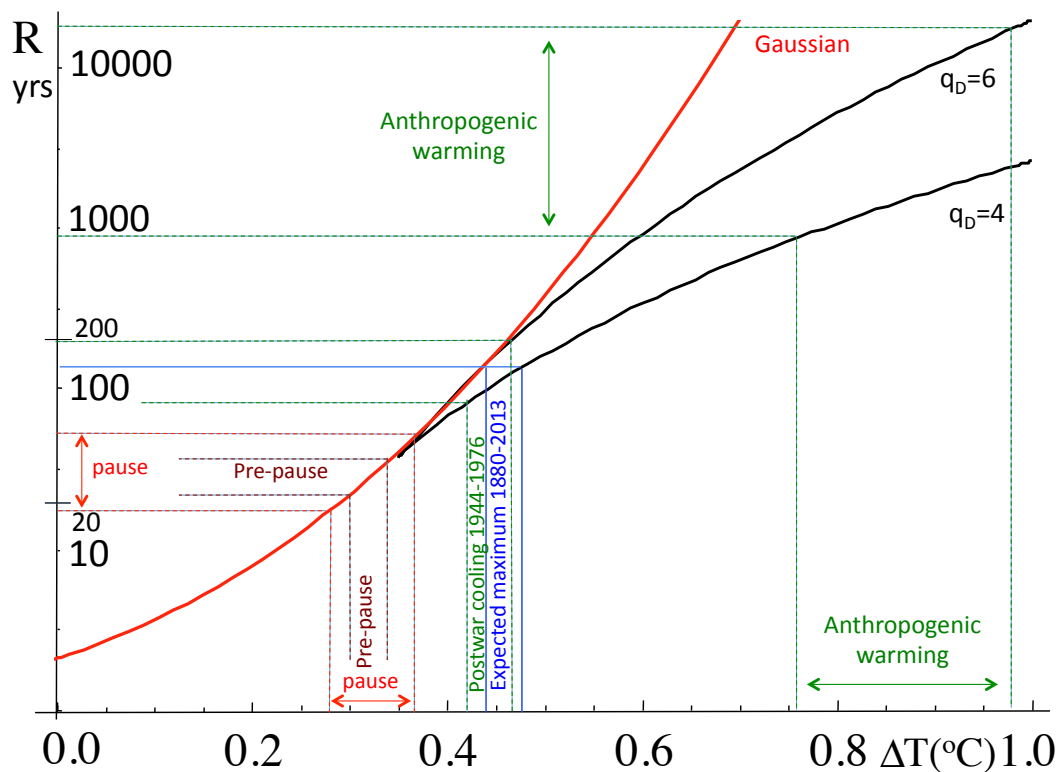

Fig. 9: The typical amount of time one must wait to observe a global scale temperature fluctuation of the amplitude indicated on the horizontal axis, the curves are the scale

invariant regressions to the empirical return times using the classical (Gaussian, red), and bounding hyperbolically tailed distributions (black, exponents q_D for the extreme tails 4, 6 as indicated) distribution. The pairs of vertical lines (one standard deviation bands about the mean) correspond to various events; the pairs of horizontal lines indicate the corresponding return periods from the more extreme ($q_D = 4$, bottom) or less extreme ($q_D = 6$, top) probabilities respectively (see eq. 3). The events, from right to left are: global warming since 1880 (green range 0.76 - 0.98° C), the largest event expected in the 134 years since 1880 (blue, 0.47° C), the postwar cooling (green, 0.42 - 0.47° C), the pre-pause 0.30 - 0.33° C (1992-1998) and "pause" 0.28 - 0.37° C (1998-2012). The horizontal lines indicate the corresponding return periods. Note that these curves should not be used for estimating the return times of temperature changes over periods much longer than a century (the rough duration of the macroweather regime). This figure is simplified with respect to the original [*Lovejoy*, 2014a]).

4.2 Stochastic forecasting: the ScaLINg Macroweather Model (SLIMM):

In the previous subsection, we used the unconditional statistics – the return periods - to argue that the natural cooling that roughly offset the anthropogenic warming since 1998 was not so unusual, noting in particular that it happened to follow an even strong pre-pause warming. While this argument is already fairly convincing, it would be better still to use the conditional statistics i.e. to quantitatively take into account the temperature variations that preceded the pause. In order to do this, we need a stochastic model of the series, in this subsection, we discuss a fairly accurate model, the ScaLIng Macroweather Model (SLIMM).

To understand the physics behind stochastic macroweather models, recall that when GCM's are used for macroweather forecasting, they are pushed far beyond their deterministic predictability limits: the weather they generate is just a "noise" forcing the lower frequencies. In this macroweather regime, control runs – with fixed climate forcings (boundary conditions) - converge slowly (in a power law manner with small negative exponent) to the GCM climates [*Lovejoy et al.*, 2013a]. However, due to model "biases" (in both the means and in the annual cycle), neither the statistics of the driving noise (the weather), nor the model climates are fully realistic.

Following [*Hasselmann*, 1976], alternative stochastic models have been developed, the most sophisticated of which are the Linear Inverse Models (LIM), [*Penland*, 1996; *Penland and Sardeshmuhk*, 1995], [*Newman et al.*, 2003], [*Sardeshmukh et al.*, 2000], [*Sardeshmukh and Sura*, 2009]. In principle stochastic models have the advantage that their statistics can be made realistic and by exploiting empirical data (the system "memory") they can effectively be forced to converge to the real climate, so that for example, a 20 component implementation of the LIM model (with >100 parameters) can already do somewhat better than GCM's for global annual temperature forecasts [*Newman*, 2013].

The key question for stochastic macroweather models is thus how big is the system memory and how best to exploit it? The LIM approach is based on systems of coupled ordinary (integer ordered) differential equations whose solutions are

essentially white noises and their integrals (Ornstein-Uhlenbeck processes). Their low frequency limits are (unpredictable) white noises so that for horizons beyond about 2 years, the errors rapidly increase [*Newman*, 2013]. However, we have seen that over the last decades, a scaling paradigm for atmospheric variability has evolved that implies the existence long-range - potentially huge - memories and these can be exploited for forecasting [*Lovejoy and Schertzer*, 1986], [*Pelletier*, 1998], [*Koscielny-Bunde et al.*, 1998], [*Franzke*, 2010; 2012], [*Rypdal et al.*, 2013], [*Rypdal and Rypdal*, 2014; *Yuan et al.*, 2014], [*Lovejoy*, 2014b], see the reviews: [*Lovejoy and Schertzer*, 2010], [*Lovejoy and Schertzer*, 2012], [*Lovejoy and Schertzer*, 2013].

In a recent paper, we showed how to use the simplest relevant scaling model – fractional Gaussian noise – to exploit the system memory: the ScaLIng Macroweather Model (SLIMM) [*Lovejoy et al.*, 2015]. SLIMM was shown to make skillful hindcasts of natural variability from monthly to decadal scales. The key to overcoming the limitations of an earlier attempt to exploit the scaling ([*Baillie and Chung*, 2002]) was to use the empirical effective climate sensitivity (λ_{2xCO_2}) to remove the anthropogenic effects [*Lovejoy*, 2014c]. For annually, globally averaged temperatures, the resulting two parameter SLIMM model (λ_{2xCO_2} and the scaling exponent *H*, see below) was already generally better than both initialized GCM's and LIM, although LIM was marginally better for horizons up to about 2 years (LDH). The present results can be viewed as the conditional probability extensions of the unconditional (return period) results of the previous section.

In the weather regime, atmospheric dynamics are intermittent; however in the macroweather regime, the intermittency is much weaker so that as a first approximation, nonintermittent (quasi-Gaussian) models may be used (although not for the extremes, see below). The usual starting point (e.g. for LIM) is the ordinary differential equation:

$$\left(\frac{d}{dt} + \tau_w^{-1} \right) T = \sigma \gamma(t)$$

(7)

where τ_w is the weather/macroweather transition scale", σ is the amplitude of the forcing and $\gamma(t)$ is a Gaussian white noise forcing with mean <$\gamma(t)$> =0. To understand this, operate on both sides by $\left(\frac{d}{dt} + \tau_w^{-1} \right)^{-1}$; this exponentially smooths the noise at scale τ_w (denoted by a subscript). At low frequencies, eq. 7 is simply:

$$T(t) = \sigma \gamma_{\tau_w}(t)$$ (8)

so that at low frequencies $T(t)$ is a white noise.

The key to realistic modelling of frequencies lower than τ_w^{-1} (SLIMM) is therefore to use a (low frequency) fractional order generalization:

$$\frac{d^{1/2+H} T}{dt^{1/2+H}} = \sigma \gamma_{\tau_w}(t)$$ (9)

Whose solution is obtained by (Riemann-Liouville) fractional integration of both sides of the equation by order $H+1/2$:

$$T(t) = \sigma G_{H,\tau_w}(t) \qquad (10)$$

Where $G_{H,\tau_w}(t)$ is a fractional Gaussian noise (fGn) process:

$$G_H(t) = K_H \int_{-\infty}^{t} (t-t')^{-(1/2-H)} \gamma(t') dt' \qquad (11)$$

smoothed at scale τ_w (K_H is an appropriate normalization constant). From eq. 11 we see that fGn has long range memory due to the slow fall-off in the weighting (the power law convolution kernel). When $-1/2 < H < 0$, the above process is stationary.

Equivalently, we obtain the same result by simply starting with:

$$\left(\frac{d}{dt} + \tau_w^{-1} \right) T = \sigma G_H(t) \qquad (12)$$

i.e. by replacing the LIM white noise forcing by the SLIMM scaling noise forcing with long range statistical dependency (and hence long range memory). Since $G_{-1/2}(t)$ is a white noise, LIM is recovered with $H = -1/2$. The difference between LIM and SLIMM is (equivalently) either the order of the differential equation (c.f. eqs. 8, 9) or the scaling (long range dependency, memory) of the forcing (c.f. eqs. 7, 12). Note that - unless we are interested in the extremes - this macroweather model is nonintermittent, $K(q) = 0$, an approximation that turns out to be reasonable in the macroweather regime (to see this, in eqs. 1, 2 take $\Delta t = |\Delta r| = L/\lambda$ to yield $\left\langle \Delta I(\Delta t)^q \right\rangle \propto \Delta t^{qH-K(q)}$ and since fGn has $\left\langle \Delta G(\Delta t)_H^q \right\rangle \propto \Delta t^{qH}$ we see that for fGn $K(q) = 0$).

At small scales $G_H(t)$ has singular behaviour so that the temporal resolution τ of $T(t)$ is fundamentally important. Although physically, the weather scales are responsible for the smoothing at τ_w, in practice, we typically have macroweather data averaged at even lower resolutions: for example monthly or annually. The simplest procedure is to introduce the resolution as an averaging procedure yielding $T_\tau(t)$ so that the variance $\left\langle T_\tau^2 \right\rangle = \sigma_T^2 \tau^{2H}$ diverges in the small resolution limit (recall $H < 0$), it follows that the spectrum of T is $E(\omega) \approx \omega^{-\beta}$, ω is the frequency and $\beta = 1 + 2H$. To within an additive and multiplicative constant, the integral of $T(t)$ is a fractional Brownian motion process introduced by [*Kolmogorov*, 1940] and [*Mandelbrot and Van Ness*, 1968]. In [*Lovejoy and de Lima*, 2015] we show how to extend this scalar SLIMM was extended to spatially intermittent space-time SLIMM (accounting for different climatic regions).

To appreciate the huge difference between LIM with exponential correlations, (a continuous version of the discrete autoregressive processes) and SLIMM with power law correlations, we can calculate the fraction of the memory that resides in past event ("innovations, the white noise, $\gamma(t)$ in eqs. 9 or 11). Fig. 10 compares the sizes of the LIM and SLIMM memories when forecasts are made for one (nondimensional) time step into the future. For example for LIM, the data more than 3 time steps in the past contains only $\approx 1\%$ of the memory so that 99% is in the past one to three time steps. In comparison, the SLIMM memory contained in 4 and more past time steps for $H = -0.4, -0.3, -0.2, -0.1$ respectively is $\approx 20\%$, 30%, 50%, 70% of the total memory. When $H = -0.1$ - a typical ocean value – thousands of time

steps in the past are still being "felt" in the sense that they collectively contribute over 20% of the memory. As a technical point, this does *not* mean that we need thousands of past time steps in order to make a good forecast. This is because the temperature values even 10 time steps in the past are also strongly influenced by the distant past innovations. Thanks to the solution of the (mathematical) fGn forecasting problem by [*Gripenberg and Norros*, 1996; *Yaglom*, 1955], and the practical (finite difference) version by [*Hirchoren and Arantes*, 1998] we know that in practice 10-20 past time step temperature data give reasonable forecasts. Whereas in LIM processes, the weights of the past data decrease rapidly from present to past so that the most recent data are the most "influential", in fGn forecasts, the weight of the data furthest in the past is also singularly (strongly) weighted since they are the best available "witnesses" of the distance past.
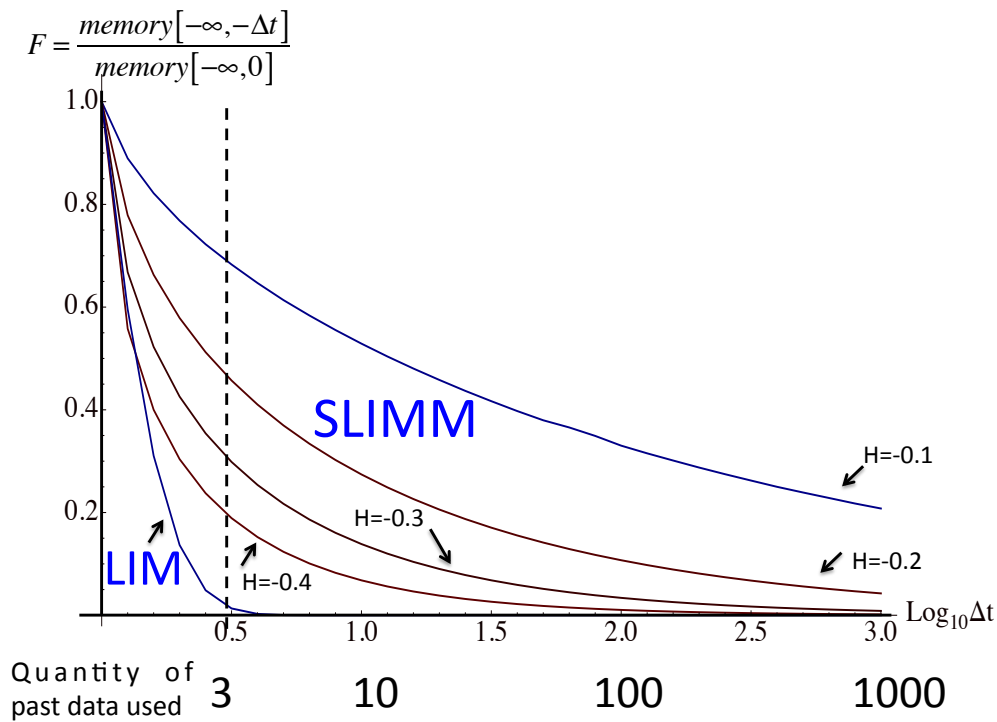

Fig. 10: The large (power law) memory of SLIMM models (based on fractional Gaussian noise, the four rightmost curves) versus the short (exponential) memory of LIM processes (far left). *F* is the fraction of the memory that influences forecasts one time step into the future. The dashed line is roughly the limit of the LIM processes: if one neglects innovations from more than about 3 time steps in the past, only a few percent of the information is lost. The comparable numbers for SLIMM are ≈70%, 50%, 30%, 20% for H =-0.1, -0.2, -0.3, -0.4 respectively (corresponding roughly to the oceans, the globe, land (temperatures) and (typical) precipitation respectively).

4.3 Hindcasting the pause with SLIMM

To illustrate the method, we used the NASA GISS annually globally averaged series from 1880 through 2013 (fig. 8a). The first step was to estimate the natural variability as discussed in section 3.4.   Unlike initialized GCM hindcasts that "optimistically" assume that future solar and volcanic forcings are known, our hindcasts (statistically) take these into account.

In fig. 11 we compare the annual temperature residues (T) and its running sum $S(t) = \sum_{t' \leq t} T(t')$ with their SLIMM hindcasts (red) and theoretical one standard deviation error bars (dashed). As expected, the actual temperatures are seen to lie almost entirely within the limits.   The bottom row (right) shows a blow-up of the temperature hindcast and the actual temperature residues, and fig. 11 shows the difference (error); we see that over the entire period 1998 - 2013, the maximum forecast error is $\approx \pm0.11$ ^{o}C.   However, the error for the hindcast "anomalies" is considerably smaller (i.e. the residues averaged over the hindcast horizon t: $\left(\hat{S}(t) - S(0)\right)/t$): fig. 11 lower left and fig. 12 (blue).   Beyond two years, the anomalies are within the theoretical one standard deviation limits, from 2002-2013, the anomaly errors are $\leq\pm0.02$ ^{o}C i.e. below the estimated temperatures measurement errors (±0.03 ^{o}C, [*Lovejoy et al.*, 2013b]).   (The term "anomaly" is a short hand for the anomaly fluctuation discussed earlier, for the residues this is simply the series at a specific resolution. For example, the RMS accuracy of annual GCM hindcasts are often quoted after averaging them over 4 or 5 years, this is the RMS of the 4 or 5 year anomaly). We could mention that other "mean reverting" processes such as LIM will also show qualitatively the same behaviour: the superiority of SLIMM was quantitatively established using over 100 hindcasts from 1900 to 2003.

Since it can be so accurately hindcast, these results support the (unconditional) statistical analysis of the previous section that the pause must be due to natural variability. This is consistent with [*Steinman et al.*, 2015] who singled out particular high amplitude - but narrow scale range - low frequency natural processes including the Atlantic Multidecadal Oscillation (AMO) and argued that they explain the pause. However, due to the scaling (in space and in time) there is in fact a hierarchy of processes analogous to the AMO ([*Held et al.*, 2010]): our hindcasts statistically account for the whole relevant scale range and show that they are all important.
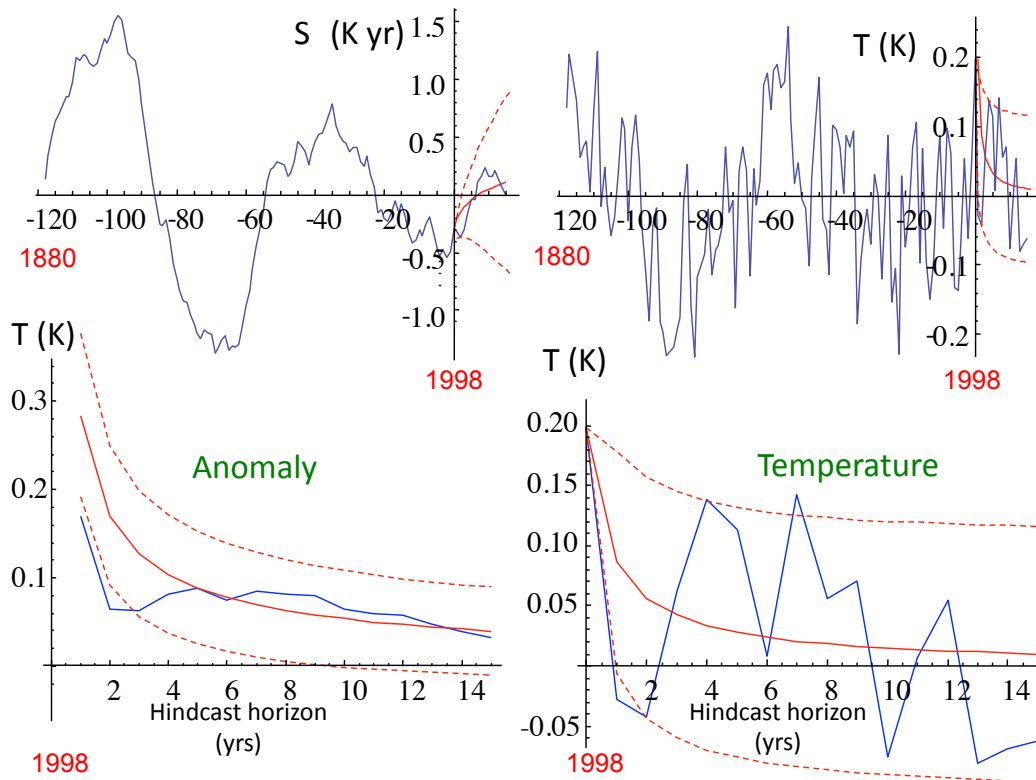
Fig. 11: **Upper left:** The summed (natural) global, annual temperature $S(t)$ (blue) with the hindcast (red) from 1998 shown in red (here and elsewhere, the dashed red lines are one standard deviation error limits).
Upper right: The natural temperature (blue) with the hindcast from 1998 (red).
Lower left: The anomaly defined as the average natural temperature (i.e. residue) over the hindcast horizon (blue), red is the hindcast.
Lower right: The temperature since 1998 (blue) with hindcast (red), a blow-up of the hindcast part of the upper right plot.

A little more effort (using stochastic forecasting) shows that if the anthropogenic warming continues at its present rate, that there is reasonable chance (5%) that the pause will continue until 2019 – 2020. Alternatively, climate skeptics will have to wait another 5 – 6 years before using any continuing pause to reject the anthropogenic warming hypothesis at 95% levels.
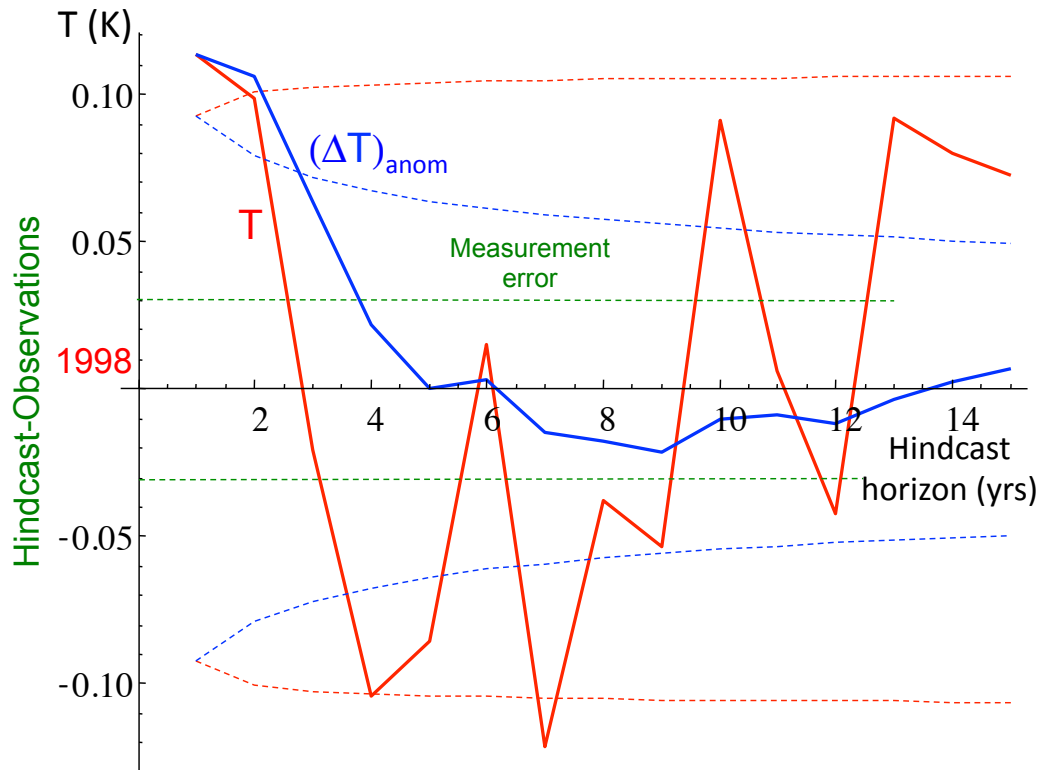
Fig. 12: The hindcast errors: anomalies (blue) and temperatures (red), obtained from fig. 11 lower left and right respectively. The one standard deviation error limits are shown as dashed. The largest absolute errors are ±0.11 K. Also shown is the estimate of the accuracy of the observations (dashed green), about ±0.03K.

5. Conclusions

Due to strong dynamical nonlinearities, the atmosphere is highly variable from planetary down to millimetric dissipation scales, in time from the age of the earth to milliseconds (10, 20 orders of magnitude respectively). The typical ratio of nonlinear to linear terms is about 10^{12} and it is believed that at such high degrees of nonlinearity that new statistical turbulent laws emerge from the deterministic laws of continuum mechanics. These laws are based on scale invariance symmetries, but to be realistic, the scaling must be anisotropic: different in the horizontal and vertical directions. In addition, the intermittency (variability) is so high that they obey nonclassical, multifractal statistics with extreme power law tails. Global atmospheric data and models (GCM's) are now of high enough quality – they span wide enough ranges of scale - that these ideas can be tested, the scaling is indeed accurately obeyed by both models and data and is of the theoretically predicted type including extremes (eq. 3). It was argued that these statistical laws are the high

level (emergent) consequences of the (lower level) deterministic laws of continuum mechanics in the limit of high nonlinearity.

This new understanding of natural variability is based on the intermittent anisotropic generalisations of classical turbulence theory described above, and can be used to statistically test the hypothesis that the industrial epoch warming is simply a large natural fluctuation. By using temperature proxy data over the pre-industrial period 1500- 1900 and the theoretically predicted form of the probabilities, we can estimate how long we must wait for fluctuations of various amplitudes over time scales up to $\approx$ 125 years. The only additional step is to estimate the magnitude of the industrial epoch warming (about 0.9° C for the global average). When this was done, it was found that the probability was so low ($\approx$ 0.1%) that it could be dismissed.

Since 1998, the warming has noticeably slowed down—and due to a lack of a convincing model based explanation—the IPCC AR5 resorted to the vague: "Due to natural variability, trends based on short records are very sensitive to the beginning and end dates and do not in general reflect long-term climate trends" (see [*Hawkins et al.*, 2014]). Our understanding of the natural variability allows us to quantitatively explain the slow down as a cooling fluctuation that masked an ongoing anthropogenic warming trend. A first explanation was based on the *unconditional* statistics the natural cooling (of about 0.3°C) had a fairly short return period (20- 50 years), it is not so unusual. However, we noted that the natural cooling since 1998 immediately followed an even larger pre-pause warming (1992-1998) so that it was plausible that the cooling was simply a return to the long term trend.

In order to confirm this – i.e. to work out the *conditional* statistics, we exploited the scaling to make stochastic forecasts based on the fluctuation exponent *H* and the global temperatures that preceded the "pause". The model we used for the hindcast was fractional Gaussian noise which was a reasonable model for the time series since the latter had low intermittency ($C_1 \approx 0.01$ to 0.02). The resulting conditional hindcast was accurate to within 0.1°C for the entire period following the 1998 peak and the 3 -4 year anomalies were hindcast to with 0.03°C. This was compared to the multimodel mean of CMIP 3 hindcasts that were about 0.2°C too high. The problem was with the models, not the theory of anthropogenic warming itself which accurately predicted the pause as a necessary consequence of the pre-pause warming: without the "slowdown", the warming would have been too strong.

Using fGn as a model for the temporal development of macroweather fields (which have low intermittency) was convenient since the mathematical problem of forecasting fGn has recently been solved. Regional (spatial) extensions of this – with realistic strongly intermittent spatial variability – ScaLINg Macroweather Models (SLIMM) are currently under development since they lead to accurate monthly, seasonal, annual and decadal (i.e. macroweather) forecasts that are currently competitive with conventional methods since they avoid the model "drift" and poor seasonality that plague GCM's while skillfully forecasting the natural variability.

Although scientific opinion has for many years been virtually consensual about the theory of anthropogenic warming, the implications for humanity are so large that professional "skeptics" have continued to denigrate the models and tout the

theory that the warming is no more than a giant natural fluctuation. It is therefore important to close the debate by eliminating this last source of rational criticism. Now that this has been done, any remaining skeptics are no more than deniers. Scientists can move on to understanding (and predicting) space-time climate variability (including regional forecasts) and the rest of the world can move on to dealing with the warming and its potentially catastrophic consequences.

References:

Arrhenius, S. (1896), On the influence of carbonic acid in the air upon the temperature on the ground, *The Philosophical Magazine*, *41*, 237-276.

Baillie, R. T., and S.-K. Chung (2002), Modeling and forecasting from trend-stationary long memory models with applications to climatology, *International Journal of Forecasting*, *18*, 215–226.

Compo, G. P., et al. (2011), The Twentieth Century Reanalysis Project, *Quarterly J. Roy. Meteorol. Soc.*, *137*, 1-28 doi: DOI: 10.1002/qj.776.

de Wijs, H. J. (1951), Statistics of ore distribution, part I, *Geologie en Mijnbouw*, *13*, 365-375.

Franzke, C. (2010), Long-range dependence and climate noise characteristics of Antarctica temperature data, *J. of Climate*, *23*, 6074-6081 doi: 10.1175/2010JCL13654.1.

Franzke, C. (2012), Nonlinear trends, long-range dependence and climate noise properties of temperature, *J. of Climate*, *25*, 4172-4183 doi: 10.1175/JCLI-D-11-00293.1.

Frisch, U., P. L. Sulem, and M. Nelkin (1978), A simple dynamical model of intermittency in fully develop turbulence, *Journal of Fluid Mechanics*, *87*, 719-724.

Gripenberg, G., and I. Norros (1996), On the Prediction of Fractional Brownian Motion *J. Appl. Prob.*, *33*, 400-410.

Hansen, J., R. Ruedy, M. Sato, and K. Lo (2010), Global surface temperature change, *Rev. Geophys.*, *48*, RG4004 doi: doi:10.1029/2010RG000345.

Hasselmann, K. (1976), Stochastic Climate models, part I: Theory, *Tellus*, *28*, 473-485.

Hawkins, E., T. Edwards, and D. McNeall (2014), Pause for thought, *Nature Clim. Change*, *4*, 154-156

Heinlein, R. A. (1973), *Time Enough for Love*, 605 pp., G. P. Putnam's Sons, New York.

Held, I. M., M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis (2010), Probing the Fast and Slow Components of Global Warming by Returning Abruptly to Preindustrial Forcing. , *J. Climate*, *23*, 2418–2427.

Hirchoren, G. A., and D. S. Arantes (1998), Predictors For The Discrete Time Fractional Gaussian Processes, in *Telecommunications Symposium, 1998. ITS '98 Proceedings. SBT/IEEE International*, edited, pp. p49-53, IEEE, Sao Paulo

Huybers, P. (2007), Glacial variability over the last two million years: An extended depth-derived agemodel, continuous obliquity pacing, and the Pleistocene progression., *Quaternary Science Reviews*, *26*((1-2)), 37-55.

Kolmogorov, A. N. (1940), Wienershe spiralen und einige andere interessante kurven in Hilbertschen Raum, *Doklady Academii Nauk S.S.S.R.*, *26*, 115-118.

Kolmogorov, A. N. (1941), Local structure of turbulence in an incompressible liquid for very large Reynolds numbers. (English translation: Proc. Roy. Soc. A434, 9-17, 1991), *Proc. Acad. Sci. URSS., Geochem. Sect.*, *30*, 299-303.

Koscielny-Bunde, E., A. Bunde, S. Havlin, H. E. Roman, Y. Goldreich, and H. J. Schellnhuber (1998), Indication of a universal persistence law governing atmospheric variability, *Phys. Rev. Lett.*, *81*, 729-732.

Ladoy, P., S. Lovejoy, and D. Schertzer (1991), Extreme Variability of climatological data: Scaling and Intermittency, in *Non-linear variability in geophysics: Scaling and Fractals*, edited by D. Schertzer and S. Lovejoy, pp. 241-250, Kluwer.

Laepple, T., S. Jewson, and K. Coughlin (2008), Interannual temperature predictions using the CMIP3 multi-model ensemble mean, *Geophys. Res. Lett.*, *35* doi: L10701, doi:10.1029/2008GL033576, 2008.

Lovejoy, S. (2013), What is climate?, *EOS*, *94, (1), 1 January*, p1-2.

Lovejoy, S. (2014a), Return periods of global climate fluctuations and the pause, *Geophys. Res. Lett.*, *41*, 4704-4710 doi: doi: 10.1002/2014GL060478.

Lovejoy, S. (2014b), A voyage through scales, a missing quadrillion and why the climate is not what ou expect, *Climate Dyn.*, *44*, 3187-3210 doi: doi: 10.1007/s00382-014-2324-0.

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

Lovejoy, S. (2015), Climate Closure, *EOS in press*.

Lovejoy, S., and B. B. Mandelbrot (1985), Fractal properties of rain and a fractal model, *Tellus*, *37 A*, 209.

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 (2010), Towards a new synthesis for atmospheric dynamics: space-time cascades, *Atmos. Res.*, *96*, pp. 1-52 doi: doi: 10.1016/j.atmosres.2010.01.004.

Lovejoy, S., and D. Schertzer (2012), 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, Washington D.C.

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

Lovejoy, S., and M. I. P. de Lima (2015), The joint space-time statistics of macroweather precipitation, space-time statistical factorization and macroweather models *Chaos 25*, 075410 doi: doi: 10.1063/1.4927223.

Lovejoy, S., D. Schertzer, and D. Varon (2013a), Do GCM's predict the climate…. or macroweather?, *Earth Syst. Dynam.*, *4*, 1–16 doi: 10.5194/esd-4-1-2013.

Lovejoy, S., D. Scherter, and D. Varon (2013b), How scaling fluctuation analyses change our view of the climate and its models (Reply to R. Pielke sr.: Interactive comment on "Do GCM's predict the climate... or macroweather?" by S. Lovejoy et al.), *Earth Syst. Dynam. Discuss.*, *3*, C1–C12.

Lovejoy, S., J. P. Muller, and J. P. Boisvert (2014), On Mars too, expect macroweather, *Geophys. Res. Lett.*, *41*, 7694-7700 doi: doi: 10.1002/2014GL061861.

Lovejoy, S., L. del Rio Amador, and R. Hébert (2015), The Scaling LInear Macroweather model (SLIM): using scaling to forecast global scale macroweather from months to decades, *Earth System Dyn. Disc.*, *6*, 489-545 doi: doi:10.5194/esdd-6-489-2015.

Lovejoy, S., A. F. Tuck, D. Schertzer, and S. J. Hovde (2009), Reinterpreting aircraft measurements in anisotropic scaling turbulence, *Atmos. Chem. and Phys.* , *9*, 1-19.

Lovelock, J. E. (1995), *Ages of Gaia*, Oxford University Press.

Lynch, P. (2006), *The emergence of numerical weather prediction: Richardson's Dream*, 279 pp., Cambridge University Press, Cambridge.

Mandelbrot, B. B. (1974), Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier, *Journal of Fluid Mechanics*, *62*, 331-350.

Mandelbrot, B. B., and J. W. Van Ness (1968), Fractional Brownian motions, fractional noises and applications, *SIAM Review*, *10*, 422-450.

Meneveau, C., and K. R. Sreenivasan (1987), Simple multifractal cascade model for fully developped turbulence, *Physical Review Letter*, *59*(13), 1424-1427.

Mitchell, J. M. (1976), An overview of climatic variability and its causal mechanisms, *Quaternary Res.*, *6*, 481-493.

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.

Newman, M. (2013), An Empirical Benchmark for Decadal Forecasts of Global Surface Temperature Anomalies, *J. of Clim.* , *26*, 5260-5269 doi: DOI: 10.1175/JCLI-D-12-00590.1.

Newman, M. P., P. D. Sardeshmukh, and J. S. Whitaker (2003), A study of subseasonal predictability, *Mon. Wea. Rev.*, *131*, 1715-1732.

Novikov, E. A., and R. Stewart (1964), Intermittency of turbulence and spectrum of fluctuations in energy-disspation, *Izv. Akad. Nauk. SSSR. Ser. Geofiz.*, *3*, 408-412.

Pelletier, J., D. (1998), The power spectral density of atmospheric temperature from scales of 10**-2 to 10**6 yr, , *EPSL*, *158*, 157-164.

Peng, C.-K., S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger (1994), Mosaic organisation of DNA nucleotides, *Phys. Rev. E*, *49*, 1685-1689.

Penland, C. (1996), A stochastic model of IndoPacific sea surface temperature anomalies *Physica D*, *98*, 534-558.

Penland, C., and P. D. Sardeshmuhk (1995), The optimal growth of tropical sea surface temperature anomalies, *J. Climate*, *8*, 1999-2024.

Pinel, J., S. Lovejoy, and D. Schertzer (2014), The horizontal space-time scaling and cascade structure of the atmosphere and satellite radiances, *Atmos. Resear.*, *140–141*, 95–114 doi: doi.org/10.1016/j.atmosres.2013.11.022.

Pinel, J., S. Lovejoy, D. Schertzer, and A. F. Tuck (2012), Joint horizontal - vertical anisotropic scaling, isobaric and isoheight wind statistics from aircraft data, *Geophys. Res. Lett.*, *39*, L11803 doi: 10.1029/2012GL051698.

Richardson, L. F. (1922), *Weather prediction by numerical process*, Cambridge University Press republished by Dover, 1965.

Richardson, L. F. (1926), Atmospheric diffusion shown on a distance-neighbour graph, *Proc. Roy. Soc.*, *A110*, 709-737.

Rohde, R., R. A. Muller, R. Jacobsen, E. Muller, S. Perlmutter, A. Rosenfeld, J. Wurtele, D. Groom, and C. Wickham (2013), A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011, *Geoinfor Geostat: An Overview 1:1* doi: doi:http://dx.doi.org/10.4172/2327-4581.1000101.

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

Rypdal, M., and K. Rypdal (2014), Long-memory effects in linear response models of Earth's temperature and implications for future global warming, *J. Climate*, *27 (14)*, 5240 - 5258 doi: doi: 10.1175/JCLI-D-13-00296.1.

Sardeshmukh, P., G. P. Compo, and C. Penland (2000), Changes in probability assoicated with El Nino, *J. Climate*, *13*, 4268-4286.

Sardeshmukh, P. D., and P. Sura (2009), Reconciling non-gaussian climate statistics with linear dynamics, *J. of Climate*, *22*, 1193-1207.

Schertzer, D., and S. Lovejoy (1985), The dimension and intermittency of atmospheric dynamics, in *Turbulent Shear Flow*, edited by L. J. S. B. e. al., pp. 7-33, Springer-Verlag.

Schertzer, D., and S. Lovejoy (1987), Physical modeling and Analysis of Rain and Clouds by Anisotropic Scaling of Multiplicative Processes, *Journal of Geophysical Research*, *92*, 9693-9714.

Schertzer, D., and S. Lovejoy (2004), Uncertainty and Predictability in Geophysics: Chaos and Multifractal Insights, in *State of the Planet, Frontiers and Challenges in Geophysics*, edited by R.S.J. Sparks and C. J. Hawkesworth, pp. 317-334, American Geophysical Union, Washington.

Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy (2007), Improved Surface Temperature Prediction for the Coming Decade from a Global Climate Model *Science*, *317*, 796-799.

Steinman, B. A., M. E. Mann, and S. K. Miller (2015), Atlantic and Pacific multidecadal oscillations and Northern Hemisphere temperatures, *Science*, *347* 988-991 doi: DOI: 10.1126/science.1257856.

Taleb, N. N. (2010), *The Black Swan: The Impact of the Highly Improbable*, 437pp pp., Rrandom House, New York.

Veizer, J., et al. (1999), 87Sr/86Sr, d18O and d13C Evolution Of Phanerozoic Seawater, *Chemical Geology*, *161*, 59-88.

Yaglom, A. M. (1955), Coprrelation theory of processes with random stationary nths increments (Russian). English translation:, Amer. Math. Soc. Trans. Ser., 8, p87-141, *Mat. Sb. N.S.*, *37*, 141-196.

Yuan, N., Z. Fu, and S. Liu (2014), Extracting climate memory using Fractional Integrated Statistical Model: A new perspective on climate prediction, *Nature Scientific Reports*, *4, Article number: 6577* doi: doi:10.1038/srep06577.

Zachos, J., M. Pagani, L. Sloan, E. Thomas, and K. Billups (2001), Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present. , *Science*, *292*(5517), 686-693 doi: Doi: 10.1126/science.1059412.