# Harnessing butterflies: theory and practice of the Stochastic Seasonal to Interannual Prediction System (StocSIPS)

S. Lovejoy, L. Del Rio Amador, R. Hébert
Physics, McGill University,
3600 University st.
Montreal, Que. H3A 2T8
Canada

## Abstract:

The atmosphere is governed by continuum mechanics and thermodynamics yet simultaneously obeys statistical turbulence laws.    Up until its deterministic predictability limit ($\tau_w \approx 10$ days), only General Circulation Models (GCM's) have been used for prediction; the turbulent laws being still too difficult to exploit. However, beyond $\tau_w$ - in macroweather - the GCM's effectively become stochastic with internal variability fluctuating about the model– not the real world – climate and their predictions are poor.  In contrast, the turbulent macroweather laws become advantageous notably due to a) low macroweather intermittency that allows for a Gaussian approximation, and b) thanks to a statistical space-time factorization symmetry that (for predictions) allows much decoupling of the strongly correlated spatial degrees of freedom.   The laws imply new stochastic predictability limits.  We show that pure macroweather  - such as in GCM's  without external forcings (control runs), can be forecast nearly to these limits by the ScaLIng Macroweather Model (SLIMM) that exploits huge system memory that forces the forecasts to converge to the real world climate.

To apply SLIMM to the real world requires preprocessing to take into account anthropogenic and other low frequency external forcings.  We compare the overall Stochastic Seasonal to Interannual Prediction System (StocSIPS, operational since April 2016) with a classical GCM (CanSIPS) showing that StocSIPS is superior for forecasts two months and further in the future, particularly over land.  In addition, the relative advantage of StocSIPS increases with forecast lead time.

In this chapter we review the science behind StocSIPS and give some details of its implementation and we evaluate its skill both absolute and relative to CanSIPS.

## 1. Introduction

### 1.1 Deterministic, stochastic, low level, high level laws

L. F. Richardson's "Weather forecasting by numerical process" (1922) opened the era numerical weather prediction. Richardson not only wrote down the modern equations of atmospheric dynamics, but he pioneered numerical techniques for

their solution and even laboriously attempted a manual integration. Yet this work also contained the seed of an alternative: buried in the middle of a paragraph, he slyly inserted the now iconic poem: "Big whirls have little whirls that feed on their velocity, little whirls have smaller whirls and so on to viscosity (in the molecular sense)". Soon afterwards, this was followed by the Richardson 4/3 law of turbulent diffusion [*Richardson*, 1926], which today is celebrated as the starting point for modern theories of turbulence including the key idea of cascades and scale invariance. Unencumbered by later notions of meso-scale, with remarkable prescience, he even proposed that his scaling law could hold from dissipation up to planetary scales, a hypothesis that has been increasingly confirmed in recent years. Today, he is simultaneously honoured by the Royal Meteorological Society's Richardson prize as the father of numerical weather prediction, and by the Nonlinear Processes division the European Geosciences Union's Richardson medal as the grandfather of turbulence approaches.

Richardson was not alone in believing that in the limit of strong nonlinearity (high Reynolds number, *Re*), that fluids would obey new high level turbulent laws. Since then, Kolmogorov, Corrsin, Obhukhov, Bolgiano, and others proposed analogous laws, the most famous of which is the Kolmogorov law for velocity fluctuations (it is nearly equivalent to Richardson's law). While the laws of continuum mechanics and thermodynamics are deterministic, the classical turbulent laws characterize the statistics of fluctuations as a functions of space-time scale; they are stochastic. Just as the laws of statistical mechanics are presumed to be compatible with those of continuum mechanics - and even though no proof (yet) exists – the latter are also presumed to be compatible with the higher level turbulence laws, see the comprehensive review [Lovejoy and Schertzer, 2013].

If both continuum mechanics and turbulent laws are valid, then both are potentially exploitable for making forecasts. Yet for reasons that we describe below, for forecasting, only the brute force integration of the equations of continuum mechanics -  General Circulation Models (GCM's) - have been developed to any degree. In this paper we review an early attempt to directly exploit the turbulent laws for macroweather forecasting i.e. for forecasts beyond the deterministic predictability limit ($\approx$ 10 days).

## 1.2  The Status of the Turbulent laws

The classical turbulent laws are of the scaling form: fluctuation $\approx$ (turbulent flux)x(scale)$^H$ where $H$ is the fluctuation exponent (for the Kolmogorov law, $H=1/3$, see below). The scaling form is a consequence of the scale invariance of the governing laws; symbolically, (*laws*)-> (*scale change by factor $\lambda$*) -> $\lambda^H$ (*laws*), (note that the scale change must be anisotropic, see [Schertzer et al., 2012]). The atmosphere has structures spanning the range of scales from planetary to submillimetric with $Re \approx 10^{12}$: making it in principle an ideal place to test such high $Re$ theories. However, the classical laws were based on very restrictive assumptions, they used unrealistic notions of turbulent flux and scale. In particular, the fluxes (which are actually in Fourier space and typically go from small to large wavenumbers) were assumed to be homogeneous or at least quasi Gaussian. However a basic feature of atmospheric dynamics is that almost all of the energy

and other fluxes are sparsely distributed in storms - and in their centres - and this enormous turbulent intermittency, was not taken into account.  In addition, the classical notion of scale was naïve: it was taken to be the usual Euclidean distance between two points i.e. it was isotropic, the same in all directions.

To be realistic, [*Schertzer and Lovejoy*, 1985] argued that the classical laws needed to be generalized precisely to take into account intermittency and anisotropy (especially stratification) and introduced the main tools: multifractal cascade processes and Generalized Scale Invariance.  Profiting from the golden age of geophysical data (remotely sensed, in situ and airborne), models and reanalyses (model – data hybrids), a growing body of work has largely vindicated this view, and has resulted in a quantitative characterization of the relevant multifractal hierarchy of exponents over wide ranges of space and time scales.  While the laws are indeed of the (generalized) scaling form indicated above, with only a few exceptions the values of the exponents still have not been derived theoretically.  They are nevertheless robust with quite similar values being found in diverse empirical data sets as well as in GCM outputs.

While large scale boundary conditions clearly affect the largest scales of flows, at small enough scales, the latter become unimportant so that for example in the atmosphere for scales below about 5000km, the predictions of turbulent cascade theories are accurate to within typically ±0.5% (see e.g. ch. 4 of [*Lovejoy and Schertzer*, 2013]), although at larger scales, deviations are important.  If the turbulent laws are insensitive to driving mechanisms and boundary conditions, then they should be "universal", operating for example in other planetary atmospheres.  This prediction was largely confirmed in a quantitative comparison of turbulent laws on Earth and on Mars.  It turns out with the exception of the largest factor of five or so in scale, that statistically, we are twins with our sister planet [*Chen et al.*, 2016], see fig. 1a, b!

## 1.3 Status of forecasts based on the Classical laws and their prospects with turbulence laws

Over the last decades, conventional numerical approaches have developed to the point where they are now skilful up until nearly their theoretical (deterministic) predictability limits - itself close to the lifetimes of planetary structures (about 10 days, see below).  Actually - due to stochastic parametrisations - state of the art ensemble GCM forecasts are stochastic – deterministic hybrids, but this limit is still fundamental.  At the same time, the strong intermittency (multifractality) over this range has meant that stochastic forecasts based on the turbulent laws must be mathematically treated as (state) *vector* anisotropic multifractal cascade processes, the mathematical understanding of which is still in its infancy (see e.g. [*Schertzer and Lovejoy*, 1995]), GCMs are the only alternative.  However, if we consider scales of many lifetimes of planetary structures - the macroweather regime - then the situation is quite different.  On the one hand because of the butterfly effect (sensitive dependence on initial conditions), in macroweather the GCMs become fully stochastic – and in fact, even in the weather regime.  On the other hand, as pointed out in [*Lovejoy and Schertzer*, 2013], [*Lovejoy and de Lima*, 2015], [*Lovejoy et al.*,

2015] in their macroweather limit, the turbulence laws become much simpler and – as we review below – can already be used to yield monthly, seasonal annual and decadal forecasts that are comparable or better than the GCM alternatives. The stochastic forecasts that we describe here effectively harness the butterfly effect. Significantly, their forecasts already appear to be close to new - stochastic - predictability limits.

As we review below, there are two principle reasons that macroweather turbulent laws are tractable for forecasts. The first is that macroweather intermittency is generally low enough that a Gaussian model is a workable approximation (although not for the extremes) - and the corresponding prediction problem has been mathematically solved. This is the basis of the ScaLIng Macroweather Model (SLIMM, [*Lovejoy et al.*, 2015]) that is the core of the Stochastic Seasonal and Annual Prediction System (StocSIPS) that we describe in this review paper. The second is that in macroweather, the usual size-lifetime relations breakdown, being replaced by new ones and an important new property called "statistical space-time factorization" (SSTF) holds (at least approximately). It turns out that the SSTF effectively transforms the forecast problem from a familiar deterministic nonlinear PDE *initial value* problem into a stochastic, fractional order linear ODE *past value* problem. In contrast at macroweather time scales, a fundamental GCM limitation comes to the fore: each GCM converges to its own model climate, not to the real world climate. While this was not important at shorter weather scales, now it becomes a fundamental obstacle. We conclude that for macroweather forecasting, the turbulent approach becomes attractive while the GCM approach becomes unattractive. Below, we compare the skills of the two different approaches and underline the advantages of exploiting the turbulent laws.

Before describing the forecast model and its skill (section 3), and comparing stochastic hindcasts with GCMs (section 4), we first discuss macroweather statistics (section 2).
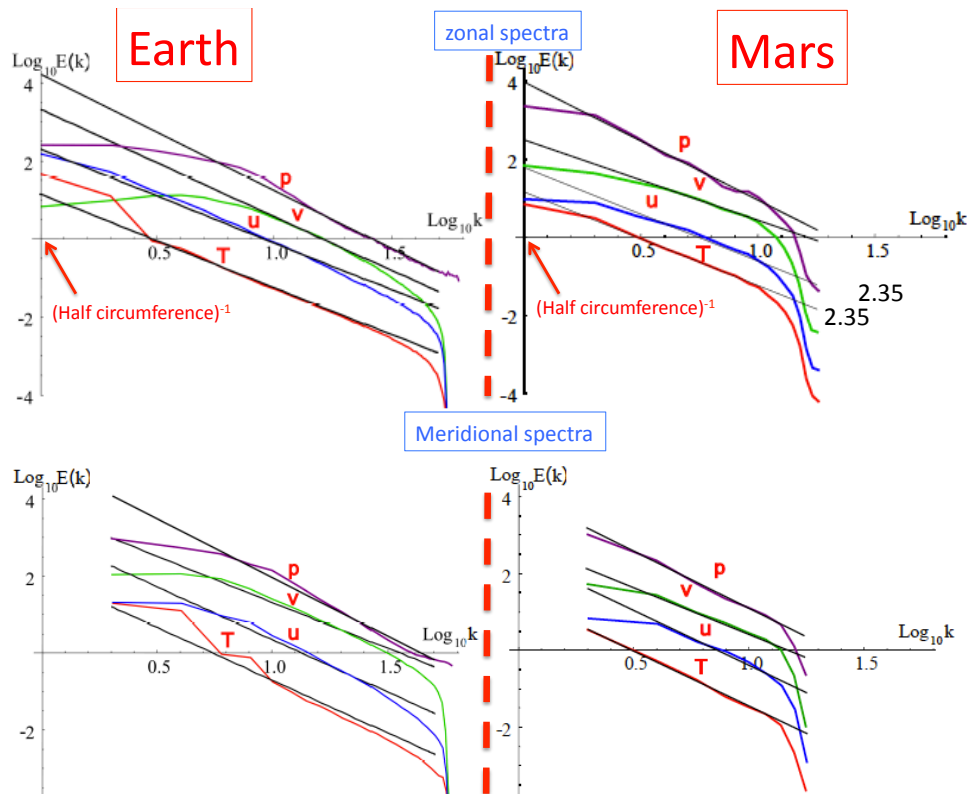
Fig. 1a: **Top row**: The zonal spectra of Earth (top left) and Mars (top right) as functions of the nondimensional wave numbers for the pressure (p, purple), meridional wind (v, green), zonal wind (u, blue), and temperature (T, red) lines. The data for Earth were taken at 69% atmospheric pressure for 2006 between latitudes ±45°. The data for Mars were taken at 83% atmospheric pressure for Martian Year 24 to 26 between latitudes ±45°. The reference lines (top left, Earth) have absolute slopes, from top to bottom: 3.00, 2.40, 2.40, and 2.75 (for *p, v, u,* and *T*, respectively). Top right (Mars) have reference lines with absolute slopes, from top to bottom: 3.00, 2.05, 2.35, and 2.35 (for *p, v, u,* and *T*, respectively). The spectra have been rescaled to add a vertical offset for clarity and wavenumber *k* = 1 corresponds to the half circumference of the respective planets.

**Bottom row:** The same as top row except for the meridional spectra of Earth (left) and Mars (right). The reference lines (left, Earth) have absolute slopes, from top to bottom: 3.00, 2.75, 2.75, and 2.40 (for *p, v, u,* and *T*, respectively). The reference lines (right, Mars) have absolute slopes, from top to bottom: 3.00, 2.40, 2.80, and 2.80 (for *p, v, u*, and *T*, respectively). The spectra have been rescaled to add a vertical offset for clarity. Adapted from [*Chen et al.*, 2016].
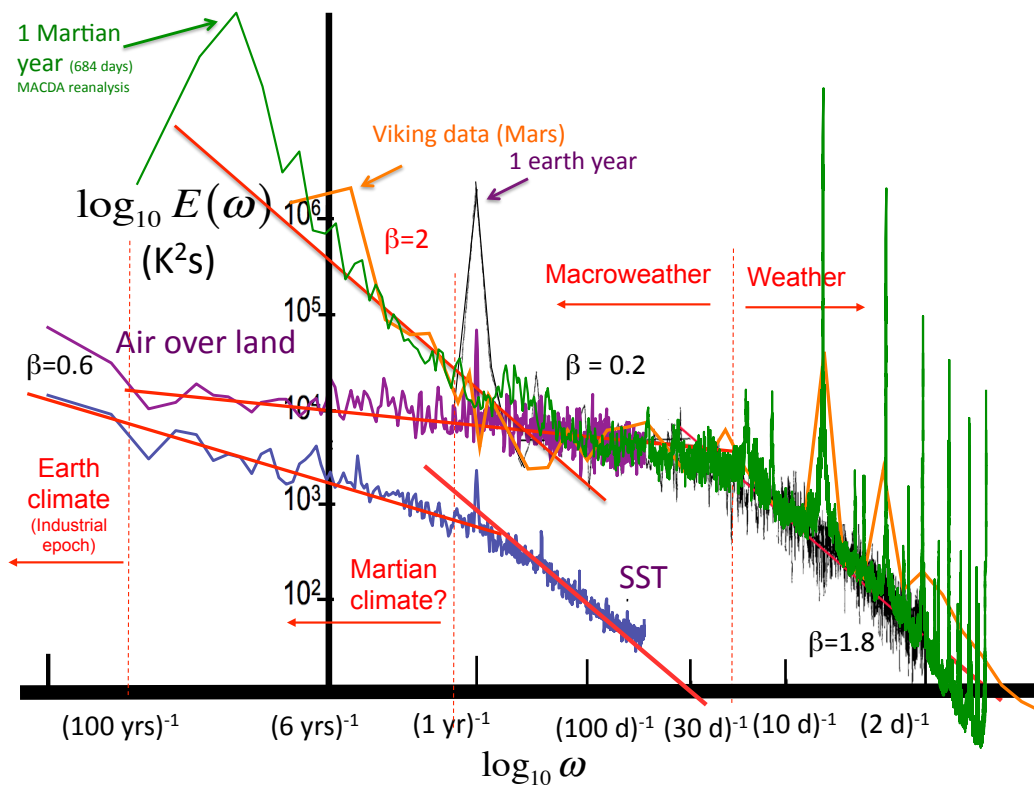
Fig. 1b: The three known weather - macroweather transitions: air over the Earth (black and upper purple), the Sea Surface Temperature (SST, ocean) at 5º resolution (lower blue) and air over Mars (Green and orange). The air over earth curve is from 30 years of daily data from a French station (Macon, black) and from air temps for last 100 years (5ºx5º resolution NOAA NCDC), the spectrum of monthly averaged SST is from the same data base (blue, bottom). The Mars spectra are from Viking lander data (orange) as well as MACDA Mars reanalysis data (Green) based on thermal infrared retrievals from the Thermal Emission Spectrometer (TES) for the Mars Global Surveyor satellite. The strong green and orange "spikes" at the right are the Martian diurnal cycle and its harmonics. Adapted from [*Lovejoy et al.*, 2014].
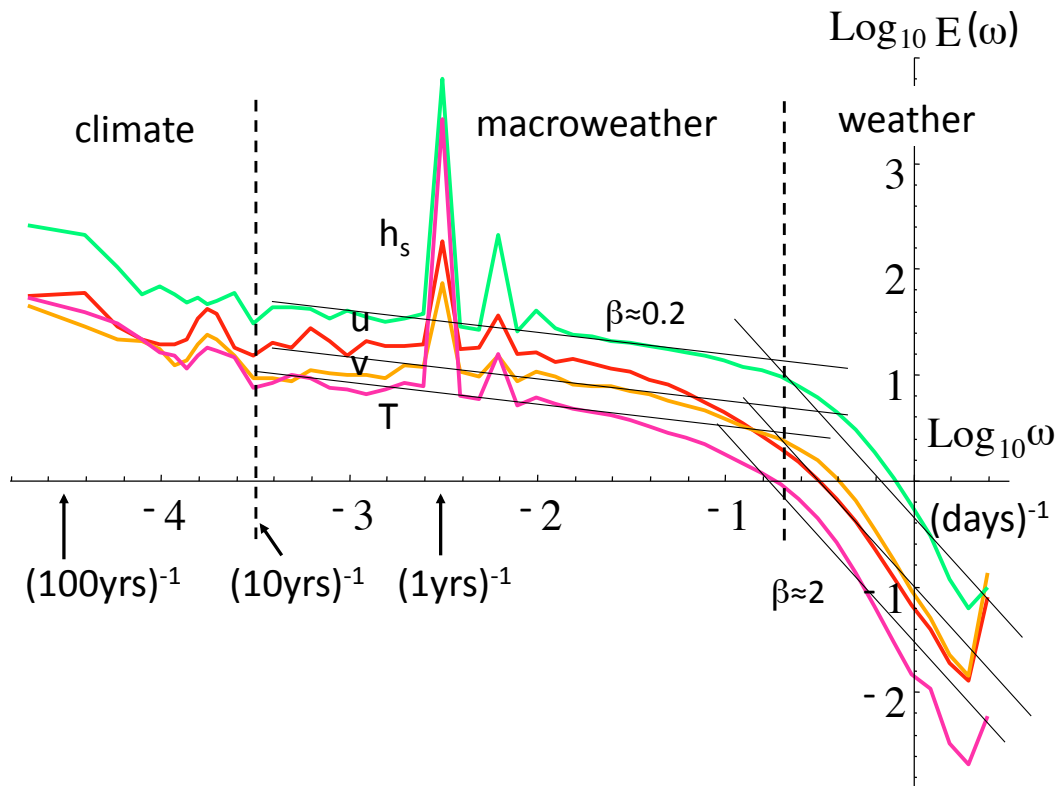
Fig. 1c: Spectra from the 20CR reanalysis (1871-2008) at 45oN for temperature (*T*), zonal and meridional wind (*u, v*) and specific humidity ($h_s$). The reference lines have correspond to, $\beta_{mw}$ = 0.2, $\beta_w$ = 2 left to right respectively. Adapted from [*Lovejoy and Schertzer*, 2013].
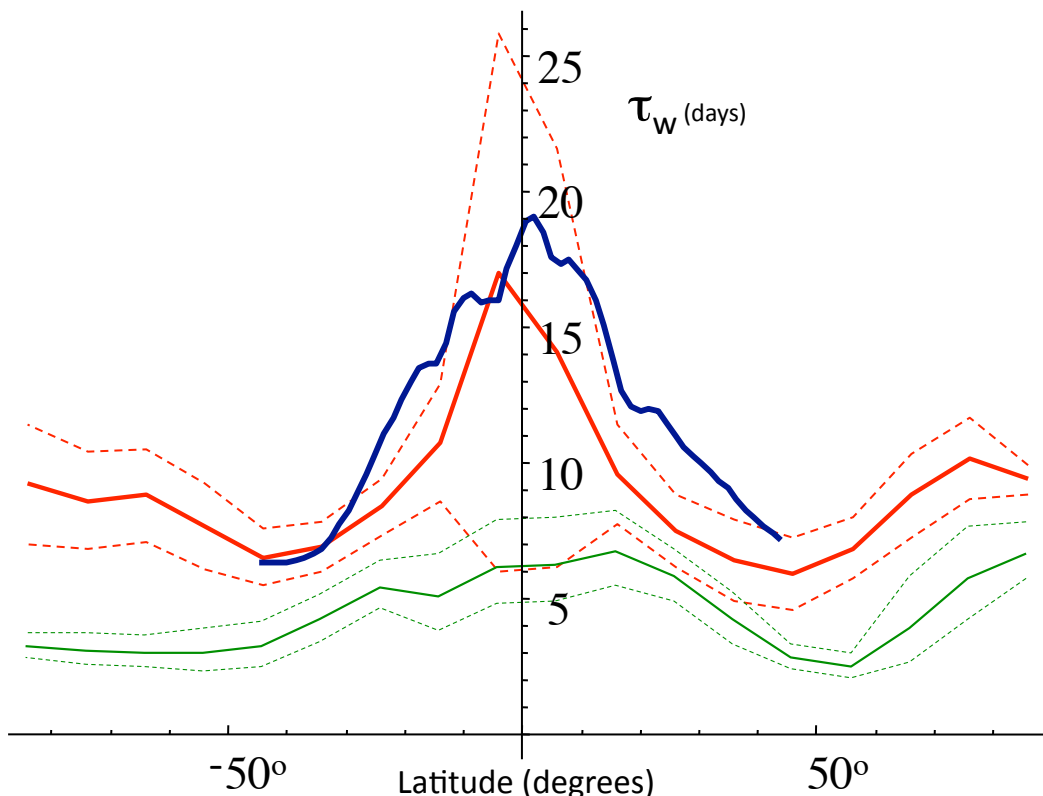
Fig. 2: The weather-macroweather transition scale $\tau_w$ estimated directly from break points in the spectra for the temperature (red) and precipitation (green) as a function of latitude with the longitudinal variations determining the dashed one standard deviation limits. The data are from the 138 year long Twentieth Century reanalyses (20CR, [*Compo et al.*, 2011]), the $\tau_w$ estimates were made by performing bilinear log-log regressions on spectra from 180 day long segments averaged over 280 segments per grid point. The blue curve is the theoretical $\tau_w$ obtained by estimating the distribution of $\varepsilon$ from the ECMWF reanalyses for the year 2006 (using $\tau_w = \varepsilon^{-1/3}L^{2/3}$ where $L$ = half earth circumference), it agrees very well with the temperature $\tau_w$. $\tau_w$ is particularly high near the equator since the winds tend to be lower, hence lower $\varepsilon$. Similarly, $\tau_w$ is particularly low for precipitation since it is usually associated with high turbulence (high $\varepsilon$). Reproduced from [*Lovejoy and Schertzer*, 2013].

## 2. Macroweather statistics

### 2.1 The transition from weather to macroweather

Ever since the first atmospheric spectra [*Panofsky and Van der Hoven*, 1955; *Van der Hoven*, 1957], it has been known that there is a drastic change in atmospheric statistics at time scales of several days. At first ascribed to "migratory pressure systems", termed a "synoptic maximum" [*Kolesnikov and Monin*, 1965], it was eventually theorized as baroclinic instability ([*Vallis*, 2010]). However, its

presence in all the atmospheric fields (fig. 1c), its true origin and its fundamental implications could not be appreciated until the turbulent laws were extended to planetary scales.

The key point is that the horizontal dynamics are controlled by the energy flux ε to smaller scales (units W/Kg, also known as the "energy rate density"). Although this is the same dimensional quantity upon which the Kolmogorov law is based ($\Delta v = \varepsilon^{1/3}L^{1/3}$ for the velocity difference $\Delta v$ across a distance $L$), it had not been suggested that it hold up to planetary scales; Kolmogorov himself believed that it would not hold to more than several hundred meters. Indeed as pointed out in [*Lovejoy et al.*, 2007] on the basis of state of the art dropsonde data, the original Kolmogorov law is isotropic and doesn't appear to hold anywhere in the atmosphere (at least at scales above ≈ 5 m)! However, the recognition that an anisotropic generalization of the Kolmogorov law could account for the horizontal statistics (with the vertical being controlled by buoyancy force variance fluxes and Bolgiano-Obhukhov statistics), explains how it is possible for the horizontal Kolmogorov law to hold up to planetary scales (see fig. 1a, for the space-time scaling up to planetary scales, see also fig. 3 for IR radiances). The classical lifetime – size ($L$) relation is then obtained by using dimensional analysis on ε: $\tau \approx \varepsilon^{-1/3}L^{2/3}$ where $L$ is the horizontal extent of a structure, (no longer an isotropic 3D estimate of its size). This law has been validated in both Lagrangian and Eulerian frames, see [*Radkevitch et al.*, 2008], [*Pinel et al.*, 2014], fig. 3).
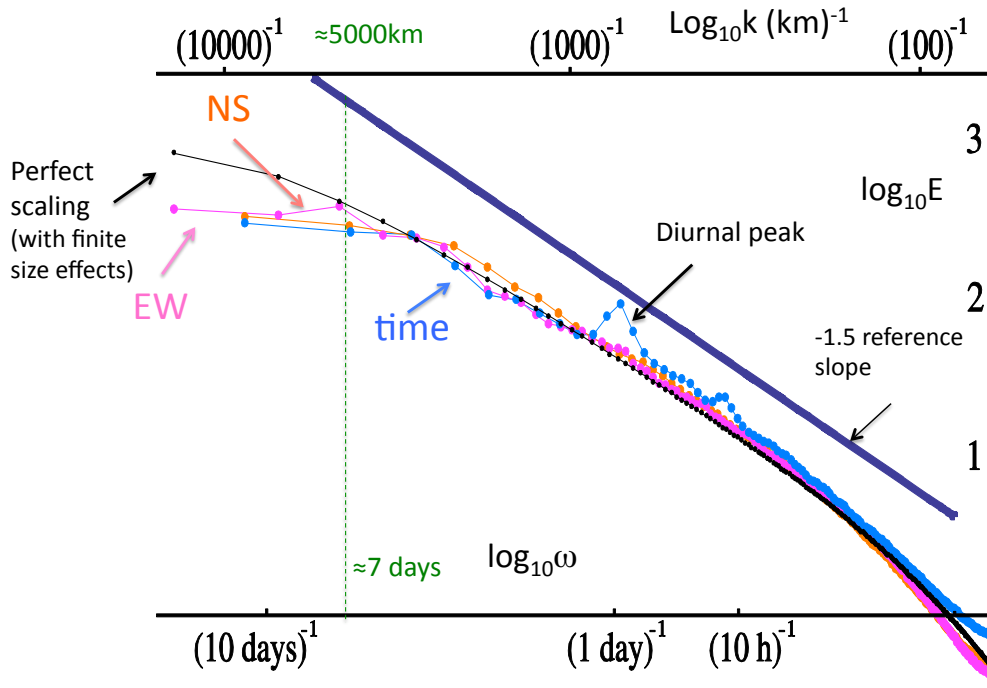
Fig. 3:  The zonal, meridional and temporal spectra of 1386 images (~ two months of data, September and October 2007) of radiances fields measured by a thermal infrared channel (10.3-11.3 µm) on the geostationary satellite MTSAT over south-west Pacific at resolutions 30 km and 1 hr over latitudes 40°S – 30°N and longitudes 80°E – 200°E.  With the exception of the (small) diurnal peak (and harmonics), the rescaled spectra are nearly identical and are also nearly perfectly scaling (the black line shows exact power law scaling after taking into account the finite image geometry.  Reproduced from [*Pinel et al.*, 2014].

If one estimates $\varepsilon$ by dividing the total tropospheric mass by the total solar power that is transformed into mechanical energy (about 4% of the total), then one finds $\varepsilon \approx 1$ mW/Kg which is close to the directly estimated empirical value (it even explains regional variations, see fig. 2).   Using $\varepsilon \approx 1$ mW/Kg, $L$ = 20,000km (the largest great circle distance) this value implies that the lifetime of planetary structures and hence the weather-macroweather transition is $\tau_w \approx 5$- 10 days.  When the theory is applied to the ocean (which is similarly turbulent with $\varepsilon \approx 10^{-8}$ W/Kg), one obtains a transition at about 1- 2 years (also observed, [*Lovejoy and Schertzer*, 2010], fig. 1b).  Finally, it can be used to accurately estimate $\varepsilon \approx 40$ mW/Kg on Mars and hence the corresponding Martian transition scale at about 1.8 sols (fig. 1b, [*Lovejoy et al.*, 2014]).

From the point of view of turbulent laws, the transition from weather to macroweather is a "dimensional transition" since at time scales longer than $\tau_w$, the spatial degrees of freedom are essentially "quenched" so that the system's dimension is effectively reduced from 1+3 to 1 ([*Lovejoy and Schertzer*, 2010]). Using spectral analysis fig. 4 shows that simple multifractal turbulence models reproduce the transition.  GCM control runs i.e. with constant external forcings (see section 2.2 and fig. 5c)  - also reproduce realistic macroweather variability, justifying the term "macroweather".   However in forced GCM's - as with instrumental and multiproxy data beyond a critical time scale $\tau_c$, the variability starts to increase again (as in the weather regime) and the true climate regime begins;  $\tau_c \approx 10$ years in the anthropocene, and $\tau_c > \approx 100$ years in the pre-industrial epoch, (see section 2.2, fig. 5).
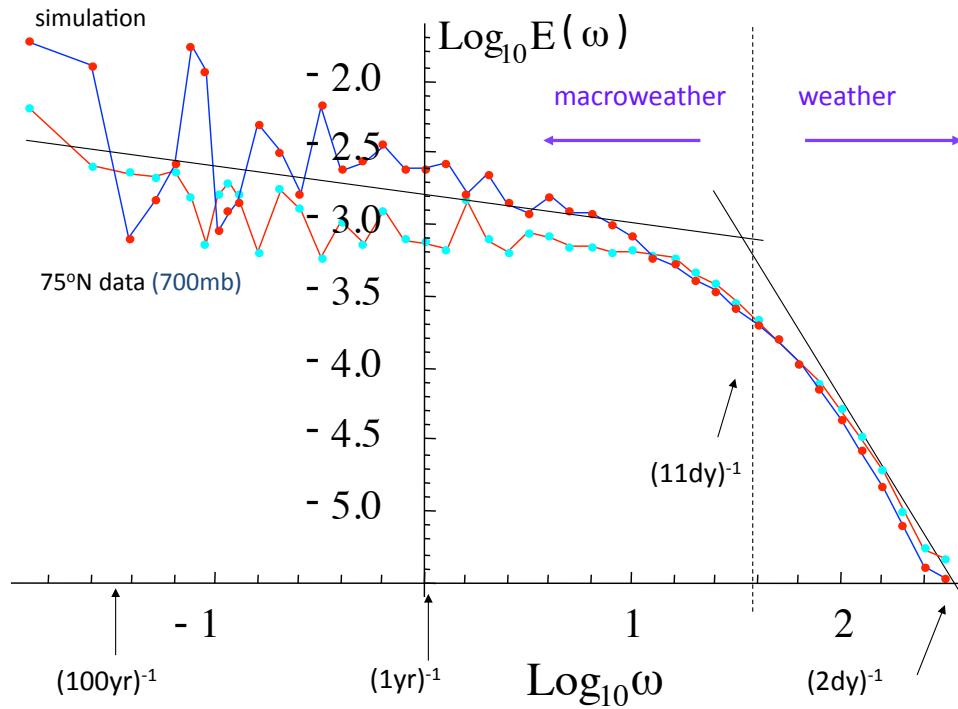
Fig. 4: A comparison of temperature spectra from a grid point of the 20CR data (bottom, orange line) and from a turbulence cascade model (top, blue line) showing that it well reproduces the weather-macroweather transition. Reproduced from [*Lovejoy and Schertzer*, 2013].

In order to understand the key difference between weather, macroweather and the climate, rather than spectra, it is useful to consider typical fluctuations. Classically - for example in the Kolmogorov law - fluctuations were taken to be differences i.e. $\Delta T(\Delta t)$:

$$\left(\Delta T\left(\Delta t\right)\right)_{diff} = T\left(t\right) - T\left(t - \Delta t\right)$$

(1)

While this is fine for weather fluctuations - these typically increase with scale $\Delta t$ - it is not adequate for those that typically decrease with $\Delta t$, and as we shall see this includes macroweather fluctuations. For these, we often consider "anomalies"; for example for the temperature anomaly $T(t)$ is the temperature with both the annual cycle and the overall mean of the series removed so that $<T>=0$ where "$<.>$" indicates averaging. For such zero mean anomaly series $T(t)$, define the $\Delta t$ resolution anomaly fluctuation by:

$$\left(\Delta T\left(\Delta t\right)\right)_{anom} \approx \frac{1}{\Delta t} \int_{t-\Delta t}^{t} T\left(t'\right) dt'$$

(2)

(as for differences, in $\Delta T(\Delta t)$ we suppressed the $t$ dependence since we assume that the fluctuations are statistically stationary). Since $T(t)$ fluctuates around zero, averaging it at larger and larger $\Delta t$ tends to decrease the fluctuations so that the decreasing classical anomaly fluctuations and the increasing difference fluctuations will each have restricted and incompatible ranges of validity.

In general, average fluctuations may either increase or decrease depending on the range of $\Delta t$ considered so that we must define fluctuations in a more general way; wavelets provide a fairly general framework for this. A simple expedient combines averaging and differencing while overcoming many of the limitations of each: the Haar fluctuation (from the Haar wavelet). It is simply the difference of the mean over the first and second halves of an interval:

$$\left(\Delta T(\Delta t)\right)_{Haar} = \frac{2}{\Delta t}\int_{t-\Delta t/2}^{t} T(t')dt' - \frac{2}{\Delta t}\int_{t-\Delta t}^{t-\Delta t/2} T(t')dt'$$

(3)

 (see [*Lovejoy and Schertzer*, 2012] for these fluctuations in a wavelet formalism). In words, the Haar fluctuation is the difference fluctuation of the anomaly fluctuation, it is also equal to the anomaly fluctuation of the difference fluctuation. In regions where the fluctuations decrease with scale we have:

$$\left(\Delta T(\Delta t)\right)_{Haar} \approx \left(\Delta T(\Delta t)\right)_{anom} \qquad \text{(decreasing with } \Delta t) \qquad (4)$$

$$\left(\Delta T(\Delta t)\right)_{Haar} \approx \left(\Delta T(\Delta t)\right)_{diff} \qquad \text{(increasing with } \Delta t)$$

In order for eq. 4 to be reasonably accurate, the Haar fluctuations in eq. 3 need to be multiplied by a calibration factor; here we use the canonical value 2 although a more optimal value could be tailored to individual series.

Over ranges where the dynamics have no characteristic time scale, the statistics of the fluctuations are power laws so that:

$$\left\langle \left| \Delta T(\Delta t) \right|^q \right\rangle \propto \Delta t^{\xi(q)}$$

(5)

the left hand side is the $q^{th}$ order structure function and $\xi(q)$ is the structure function exponent. "<>" indicates ensemble averaging; for individual series this is estimated by temporal averaging (over the disjoint fluctuations in the series). The first order ($q = 1$) case defines the "fluctuation exponent" $\xi(1) = H$:

$$\left\langle \left| \Delta T(\Delta t) \right| \right\rangle \propto \Delta t^{H}$$

(6)

In the special case where the fluctuations are quasi-Gaussian, $\xi(q) = qH$ and the Gaussian white noise case corresponds to $H = -1/2$. More generally, there will be "intermittency corrections" so that:

$$K(q) = qH - \xi(q) \qquad\qquad\qquad (7)$$

where $K(q)$ is a convex function with $K(1) = 0$. $K(q)$ characterizes the multifractality associated with the intermittency.

Eq. 6 shows that the distinction between increasing and decreasing mean fluctuations corresponds to the sign of $H$. It turns out that the anomaly fluctuations are adequate when $-1 < H < 0$ whereas the difference fluctuations are adequate when

0<*H*<1.  In contrast, the Haar fluctuations are useful over the range -1<*H*<1 which encompasses virtually all geoprocesses, hence its more general utility.  When *H* is outside the indicated ranges, then the corresponding statistical behaviour depends spuriously on either the extreme low or extreme high frequency limits of the data.

## 2.2 The low frequency macroweather limit and the transition to the climate

We have argued that there is a drastic statistical transition in all the atmospheric fields at time scales of 5 – 10 days, and that the basic equations have no characteristic time scale.  However, it was noted since [*Lovejoy and Schertzer*, 1986] (fig. 5a) that global temperature differences tend to increase in a scaling manner right up to the ice age scales: the glacial-interglacial "window" at about 50kyrs (a half cycle) over which fluctuations are typically of the order ±2 to ±4 K.
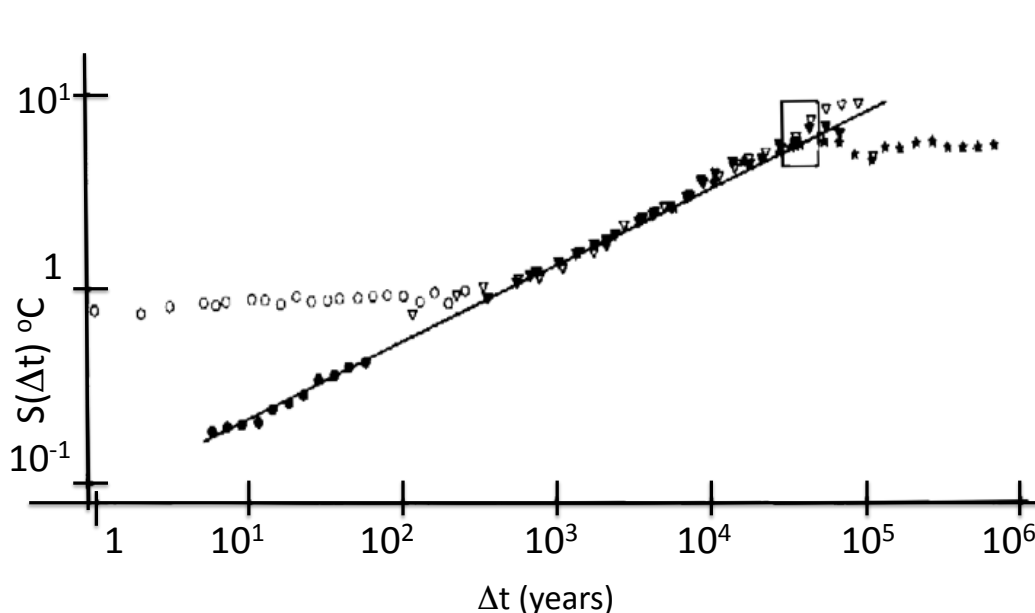


Fig. 5a:  The RMS difference structure function estimated from local (Central England) temperatures since 1659 (open circles, upper left), northern hemisphere temperature (black circles), and from paleo temperatures from Vostok (Antarctic, solid triangles), Camp Century (Greenland, open triangles) and from an ocean core (asterixes).   For the northern hemisphere temperatures, the (power law, linear on this plot) climate regime starts at about 10 years.  The rectangle (upper right) is the "glacial-interglacial window" through which the structure function must pass in

order to account for typical variations of ±2 to ±3K for cycles with half periods ≈ 50 kyrs. Reproduced from [*Lovejoy and Schertzer*, 1986].
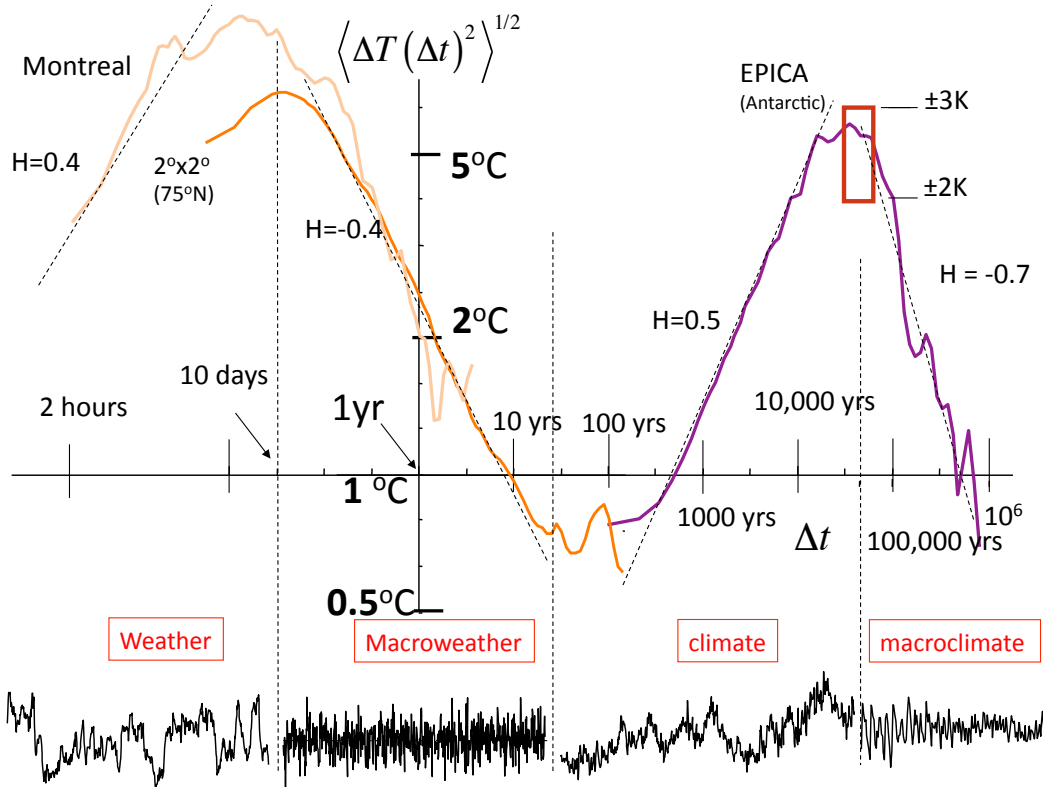


Fig. 5b: A composite RMS Haar structure function from (daily and annually detrended) hourly station temperatures (left), 20CR temperatures (1871-2008 averaged over 2° pixels at 75°N) and paleo-temperatures from EPICA ice cores (right) over the last 800kyrs. The glacial-interglacial window is shown upper right rectangle. Adapted from [*Lovejoy*, 2015a].
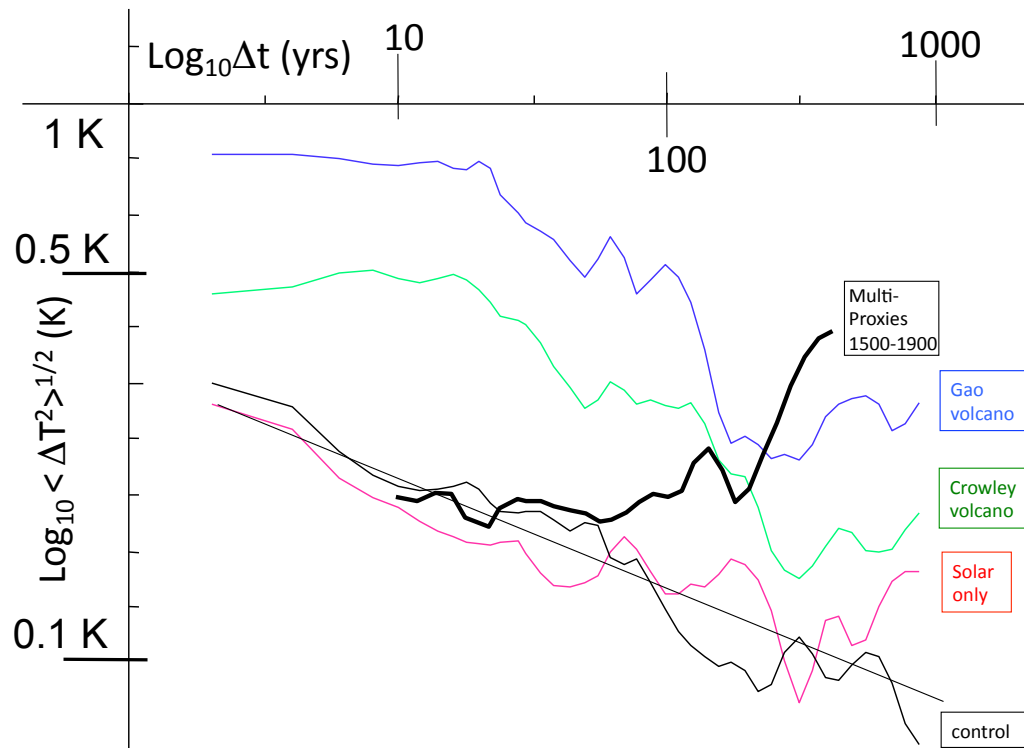
Fig. 5c: Haar fluctuation analysis of globally, annually averaged outputs of past Millenium simulations over the pre-industrial period (1500-1900) using the NASA GISS E2R model with various forcing reconstructions. Also shown (thick, black) are the fluctuations of the pre-industrial multiproxies showing that they have stronger multi centennial variability. Finally, (bottom, thin black), are the results of the control run (no forcings), showing that macroweather (slope<0) continues to millennial scales. Reproduced from [*Lovejoy et al.*, 2013].
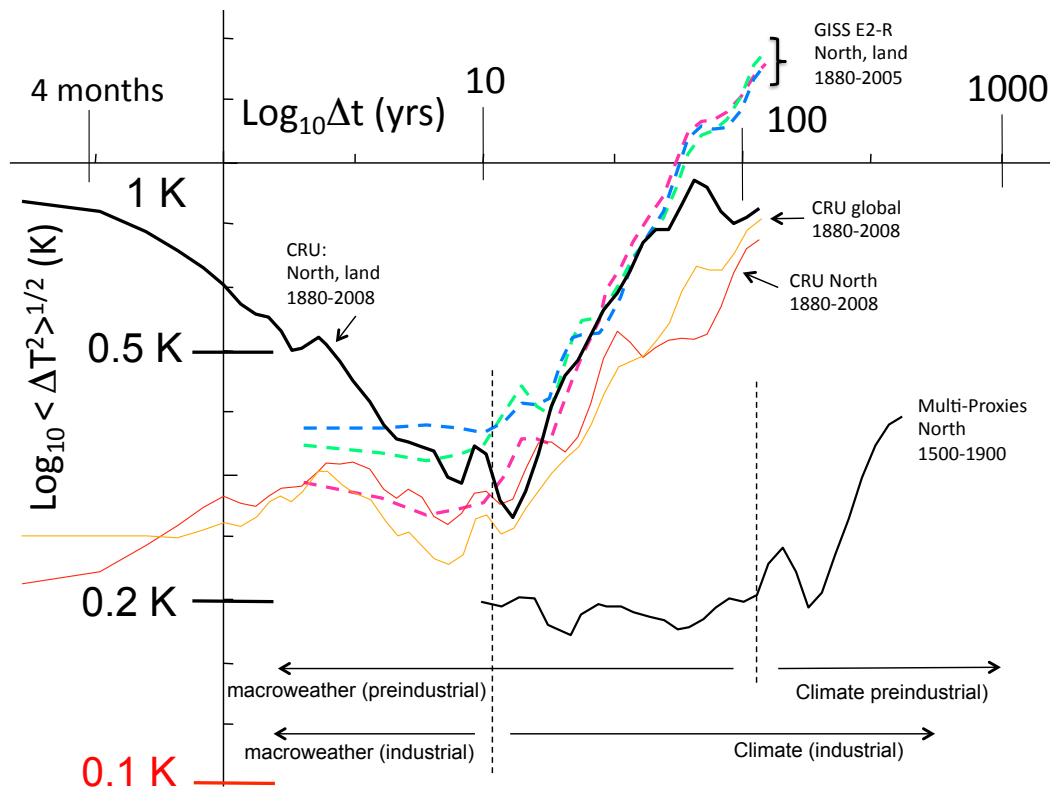
Fig. 5d:  Haar fluctuation analysis of  Climate Research Unit (CRU, HadCRUtemp3 temperature fluctuations), and globally, annually averaged outputs of past Millenium simulations over the same period (1880-2008) using the NASA GISS E2R model with various forcing reconstructions (dashed).  Also shown are the fluctuations of the pre-industrial multiproxies showing the much smaller centennial and millennial scale variability that holds in the pre-industrial epoch.  Reproduced from [*Lovejoy et al.*, 2013].

Fig. 5a shows the root mean square second order structure function defined by difference fluctuations $\left\langle \Delta T \left( \Delta t \right)_{diff}^2 \right\rangle^{1/2}$ for both local and hemispherically averaged temperatures.  From the above discussion, we anticipate that it will give spurious results in the regions where the true fluctuations decrease with scale; indeed, the local (central England) series (upper left in fig. 5a and ocean cores beyond ≈ 100 kyrs, upper right) are spuriously flat (i.e. the differences do not reflect the underlying scaling of the fluctuations that are in fact decreasing over these ranges). This is confirmed using both more modern data as well as Haar rather than difference fluctuations, in fig. 5b that shows a composite of temperature variability over the range of scales of hours to nearly a million years.   From fig. 5b, it can be seen that the drastic weather-macroweather spectral transition corresponds to a change in the sign of *H* for *H*>0 to *H*<0, i.e. from fluctuations increasing to fluctuations decreasing with scale.  The bottom of the figure shows extracts of

typical data at the corresponding resolutions, when $H>0$, the signal "wanders" like a drunkard's walk, when $H<0$, successive fluctuations tend to cancel out.

Moving to the longer time scales, one may also note that beyond a decade or two, the fluctuations again increase with scale. In reality, as one averages from weeks to months to years, the temperature fluctuations are indeed averaged out, appearing to converge to a fixed climate.  However, starting at decades, this apparent fixed climate actually started to fluctuate, varying up to ice age scales in much the same way as the weather varies (with nearly the same exponent $H \approx 0.4$, see fig. 5b).  While the adage says "The climate is what you expect, the weather is what you get", the actual data indicate that "Macroweather is what you expect, the climate is what you get".
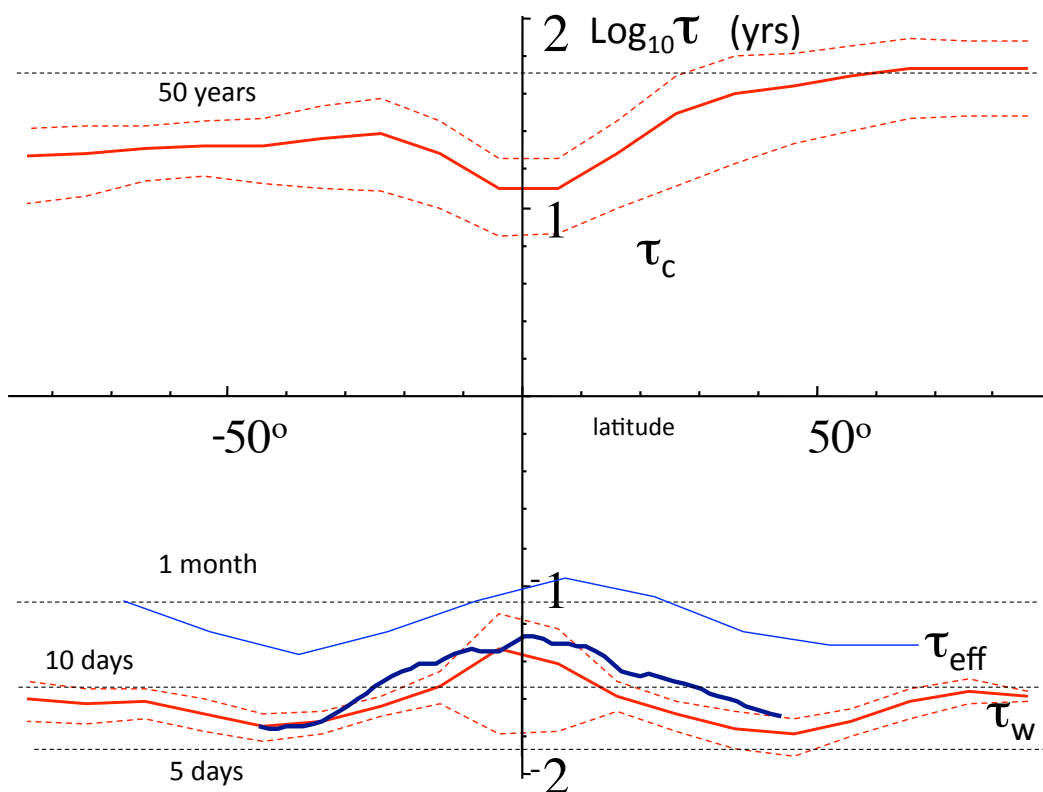


Fig. 6: Variation of $\tau_w$ (bottom) and $\tau_c$ (top) as a function of latitude as estimated from the 138 year long 20CR reanalyses, 700mb temperature field (the $\tau_c$ estimates are only valid in the anthropocene).  The bottom red and thick blue curves for $\tau_w$ are from fig. 2; also shown at the bottom is the effective external scale ($\tau_{eff}$) of the temperature cascade estimated from the European Centre for Medium-Range Weather Forecasts interim reanalysis for 2006 (thin blue).  The top $\tau_c$ curves were estimated by bilinear log-log fits on the Haar structure functions applied to the same 20CR temperature data.  The macroweather regime is the regime between the top and bottom curves.

The annual and decadal scales in figs. 5a and Fig 5b are from the anthropocene, it is important to compare this with the preindustrial variability. This comparison is shown in detail in figs. 5c and 5d that include comparisons with GCM outputs. From the figures we see that in the anthropocene, macroweather ends (scale $\tau_c$) at around a decade or so; fig. 6 give estimates of $\tau_c$ averaged over fixed latitudes showing that it is a little shorter in the low latitudes. We have seen (fig. 4) that without external forcing, turbulence models when taken to their low frequency limit reproduce macroweather statistics; the same is true of GCMs in their "control run" mode (fig. 5c). These results are important for macroweather forecasting since they represent a potential calculable climate perturbation to the otherwise (pure internal variability) macroweather behaviour.

In order to reproduce the low frequency climate regime characterized by increasing fluctuations, we therefore need something new: either a new source of internal variability, or external forcings. Figure 5d shows that whereas in the anthropocene, the GCM's with Green House Gas (GHG) forcings do a good job of reproducing the variability, in the preindustrial period (fig. 5c), their centennial and millennial scale variability seems to be too weak (at least when using current estimates of "reconstructed" solar and volcanic forcings [*Lovejoy et al.*, 2013]).

The usual way to understand the low frequencies is to consider them as responses to small perturbations, indeed, even the strong anthropogenic forcing is less than 1% of the mean solar flux and may be considered this way. This smallness is the usual justification for making the approximation that the external forcings (whether of natural or anthropogenic origin) yield a roughly linear response, indeed, this is the basis of linearized energy balance models and it can also be supported from a dynamical systems point of view [*Ragone et al.*, 2015].

In order to avoid confusion, it is worth making these notions more precise. For simplicity, consider the atmosphere with fixed external radiative forcing $F(\underline{r})$ at location $\underline{r}$, (e.g. corresponding to GCM control runs). For this fixed forcing, the (stochastic) temperature field is:

$$T_F(\underline{r},t) = \langle T_F(\underline{r}) \rangle + T_F'(\underline{r},t) \tag{8}$$

where the ensemble average is independent of time (since the past forcing is fixed) and $T'$ (with $<T'> = 0$) is the random deviation. If we identify $\langle T_F(\underline{r}) \rangle$ with the climate and $T_F'(\underline{r},t)$ with the internal variability, then:

$$T_{F,\mathrm{int}ernal}(\underline{r},t) = T_F(\underline{r},t) - T_{F,c\lim}(\underline{r}); \quad T_{F,c\lim}(\underline{r}) = \langle T_F(\underline{r},t) \rangle; \quad T_{F,\mathrm{int}ernal}(\underline{r},t) = T_F'(\underline{r},t)$$

$$\tag{9a}$$

For simplicity, we have ignored the annual cycle, the internal variability is somewhat different than the notion of temperature anomalies discussed in section 4).

Now increase the forcing from $F(\underline{r}) \rightarrow F(\underline{r}) + \Delta F(\underline{r},t)$ so that the climate part increases from $\langle T_F(\underline{r}) \rangle \rightarrow \langle T_{F+\Delta F}(\underline{r},t) \rangle$ i.e. $T_{F,c\lim}(\underline{r}) \rightarrow T_{F+\Delta F,c\lim}(\underline{r},t)$ and:

$$\Delta T_{\Delta F,c\lim}(\underline{r},t) = T_{F+\Delta F,c\lim}(\underline{r},t) - T_{F,c\lim}(\underline{r}) \tag{9b}$$

is the change in the climate response to the changed forcing.    The generalized climate sensitivity λ can then defined as:

$$\lambda\left(\underline{r},t\right)=\frac{\partial T_{F,c\lim}\left(\underline{r},t\right)}{\partial F\left(\underline{r},t\right)}\approx\frac{\Delta T_{F,c\lim}\left(\underline{r},t\right)}{\Delta F\left(\underline{r},t\right)} \tag{10}$$

GCM's make many realizations (sometimes from many models- "multimodel ensembles") and this equation may be used to determine the climate response and generalized sensitivity (the more common equilibrium and transient climate sensitivities are discussed momentarily).    If $t$ is a future time, then $T_{F+\Delta F}\left(\underline{r},t\right)$ is a prediction of the future state of the atmosphere including the internal variability and the changed forcing, whereas $T_{F+\Delta F,c\lim}\left(\underline{r},t\right)$ is called a climate "projection". Sometimes climate projections and sensitivities are estimated from single GCM model runs by estimating the ensemble averages by temporal averages over decadal time scales.

We can now state the linear response assumption:

$$\Delta T_{c\lim}\left(\underline{r},t\right)=G\left(\underline{r},t\right)*\Delta F\left(\underline{r},t\right) \tag{11}$$

where $G(\underline{r},t)$ is the system Green's function, in this context, it is also known as the Climate Response Function (CRF), "∗" means convolution.  Eq. 11 is the most general statement of linearity for systems whose physics is the same at all times and locations (it assumes that only the differences in times and locations between the forcing and the responses are important).   To date, applications of CRF's have been limited to globally averaged temperatures and forcings so that the spatial ($\underline{r}$) dependence is averaged out; for simplicity, below we drop the spatial dependence.

The CRF is only meaningful if the system is linear, in which case it is the response of the system to a Dirac function forcing.   The simplest CRF is itself a Dirac function possibly with a lag $\Delta t \geq 0$, i.e. $G\left(t\right)=\lambda\delta\left(t-\Delta t\right)$, (sensitivity $\lambda$).   Such CRF's have been used with some success by [*Lean and Rind*, 2008] and  [*Lovejoy*, 2014a] to account for both anthropogenic and natural forcings.  Rather than characterize the system by a response to Dirac forcing, it is more usual to characterize it by its responses to a step function $F(t)$ (the Equilibrium Climate Sensitivities, ECS) or to a linearly increasing $F(t)$ ("ramps"; Transient Climate Responses, TCR).   Since step functions and ramps are simply the first and second integrals of the Dirac function, if the response is linear (eq. 11), then knowledge of these responses as functions of time is equivalent to the CRF (note that usually the ECS is defined as the response after an infinite time, and TCR after a finite conventional period of 70yrs).

Traditionally, Green's functions are deduced from linear differential operators arising from linear differential equations.  For example, by treating the ocean as a homogeneous slab, the linearized energy balance equation may be used to determine the CRF, but the latter is an integer ordered ordinary differential equation for the mean global temperature which leads to exponential CRF's (e.g. [*Schwartz*, 2012], [*Zeng and Geil*, 2017]).   Such CRF's are unphysical since they break the scaling symmetry of the dynamics; the dynamical ocean is better modelled as a hierarchy of slabs each with its own time constant (rather than a unique slab

with a unique constant).    To model this in the linear energy balance framework requires introducing differential terms of fractional order; these generally lead to the required scaling CRF's (SCRF) and will be investigated elsewhere.

Rather than determine the CRF from differential operators, they can be determined directly from the symmetries of the problem.  In this case (considering only the temporal CRF, $G(t)$), the three relevant symmetries are:  a) that the physics is stationary in time, b) that the system is causal, c) that there is no characteristic time scale.  From these three symmetries we obtain $G(t) \propto t^{H_R - 1}\Theta(t)$ where $H_R$ is the SCRF response exponent and $\Theta(t)$ is the Heaviside function (=0 for $t<0$, = 1 for $t \geq 0$), necessary to ensure causality of the response.

Before continuing, we must note that such pure power law SCRF's are unusable due to either high or low frequency divergences; in this context the divergences are aptly called "runaway Green's function effect" ([*Hébert and Lovejoy*, 2015]) so that truncations are needed.  For forcings that have infinite "impulses" (such as step functions or ramps whose temporal integrals diverge), when $H_R>0$ a low frequency temperature divergences will occur, unless $G(t)$ has a low frequency cutoff whereas whenever $H_R<0$, the cutoff must be at high frequencies. For example, [*Rypdal*, 2015; *Rypdal and Rypdal*, 2014] use an SCRF with exponent $H_R>0$ (without cutoff) so that low frequency temperature divergences occur unless all the forcings return to zero quickly enough.   This is why [*Hebert et al.*, 2017] use $H_R <0$ but introduce a high frequency cut-off $\tau$ in order to avoid the divergences: $G(t) = \lambda_H \left( t/\tau + 1 \right)^{H_R - 1}\Theta(t)$; $\lambda_H$ is a generalized sensitivity.  In this case, the cut-off should correspond to the smallest time scale over which the linear approximation is valid.  While the most general (space-time) linear approximation (i.e. with $G(\underline{r},t)$) may be valid at shorter time scales, if we reduce the problem to a "zero dimensional" (globally averaged) series $T(t)$, then clearly a linear response is only possible at scales over which the ocean and atmosphere are strongly coupled.   The breakthrough in understanding and quantifying this was to use Haar fluctuations to show that the coupling of air temperature fluctuations over land and SST fluctuations abruptly change from very low to very high at the ocean weather-ocean macroweather transition scale of $\tau$ = 1- 2 years (see fig. 7).  A truncated SCRF with this $\tau$ and with $H_R \approx$ -0.5 allows [*Hebert et al.*, 2017] to make future projections based on historical forcings as well as to accurately project the forced response of GCM models.
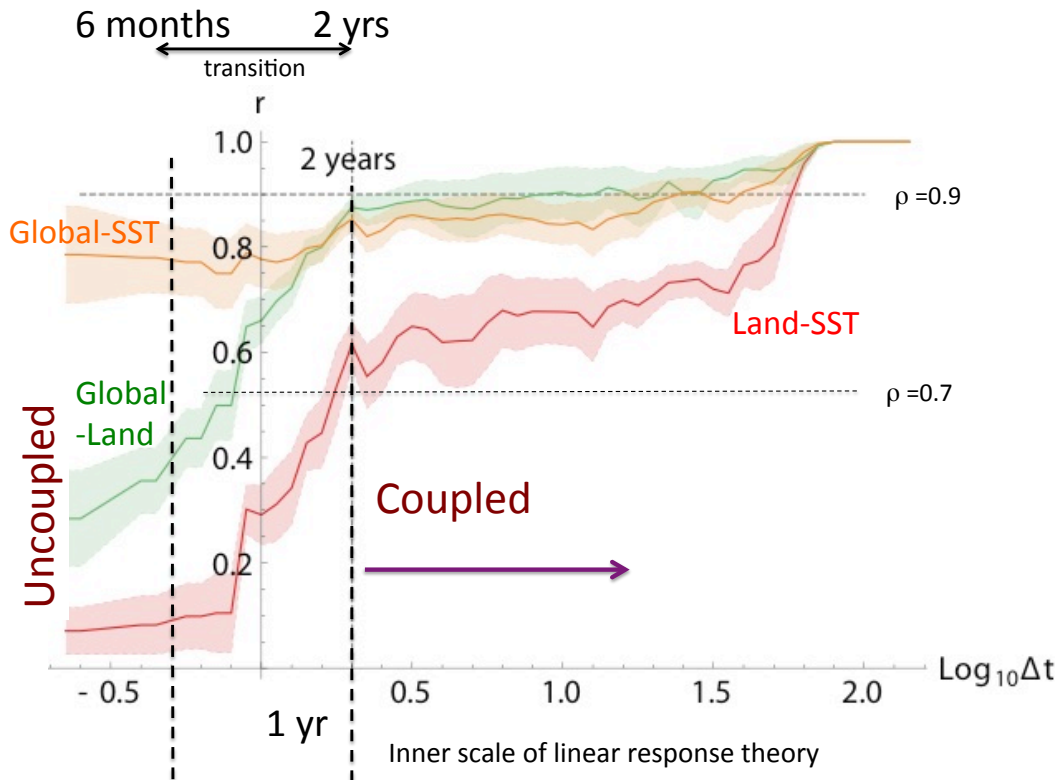
Fig. 7:  The correlations quantifying the coupling of global, land and ocean temperature fluctuations.  At each scale $\Delta t$, the correlation coefficient $r$ of the corresponding Haar fluctuations was calculated for each pair of the monthly resolution series.  The key curve is the correlation coefficient of globally averaged air over land with globally averaged sea surface temperature (SST, bottom, red). One can see that there is a sharp transition at $\tau \approx 1$ -2 years from very low correlations, to very high correlations corresponding to uncoupled and coupled fluctuations. Reproduced from [*Hebert et al.*, 2017].

## 2.3 Climate zones and Intermittency: in space and time

We have argued that macroweather is the dynamical regime of fluctuations with time scales between the lifetimes of planetary structures ($\tau_w$) and the climate regime where either new (slow) internal processes or external forcings begin to dominate ($\tau_c$).  We have seen that a key characteristic is that mean fluctuations tend to decrease with time scale so that the macroweather fluctuation exponent $H<0$. However in general, fluctuations require an infinite hierarchy of exponents for their characterization (the entire function $K(q)$ in eq. 7).  In particular, when $K(q)$ is large, the process is typically "spikey" with the spikes distributed in a hierarchical manner over various fractal sets.

To see this, consider the data shown in fig. 8a (macroweather time series and spatial transects, top and bottom respectively).  Fig. 8b compares the root mean square (RMS, exponent $\xi(2)/2$) and mean fluctuation (exponent $H = \xi(1)$) of

macroweather temperature temporal data (bottom) and for the transect (top). When the system is Gaussian, $\xi(q) = qH$ so that $K(q) = 0$  and we obtain $\xi(2)/2 = \xi(1)$ so that the lines in the figure will be parallel.  We see that to a good approximation this is indeed true of the nonspikey temporal series (fig. 8a, top). However, the spatial transect is highly spikey (fig. 8a, bottom) and the corresponding statistics (the top lines in fig. 8b) tend to converge at large $\Delta t$.  To a first approximation, it turns out that $\xi(2)/2 - \xi(1) \approx K'(1) = C_1$ which characterizes the intermittency near the mean.    However, there is a slightly better characterisation of $C_1$ (described in [*Lovejoy and Schertzer*, 2013], ch. 11), using the intermittency function (see fig. 8c and caption) whose theoretical slope (for ensemble averaged statistics) is exactly $K'(1) = C_1$.  As a point of comparison, recall that fully developed turbulence in the weather regime typically has $C_1 \approx 0.09$, (see [*Lovejoy and Schertzer*, 2013], table 4.5).    The temporal macroweather intermittency ($C_1 \approx 0.01$) is indeed small whereas the spatial intermittency is large ($C_1 \approx 0.12$).

The strong spatial intermittency is the statistical expression of the existence of climate zones ([*Lovejoy and Schertzer*, 2013]).  However we shall see that due to space-time statistical factorization (next subsection), each region may be forecast separately.  In addition, a low intermittency (Gaussian) approximation can be made for the temporal statistics.  Note that in spite of this Gaussian approximation for forecasts, there is evidence that the 5th and higher moments of the temperature fluctuations diverge (i.e. power probability distributions) so that the Gaussian approximation fails badly for the extreme 3% or so of the fluctuations (see [*Lovejoy and Schertzer*, 1986], [*Lovejoy*, 2014a]).
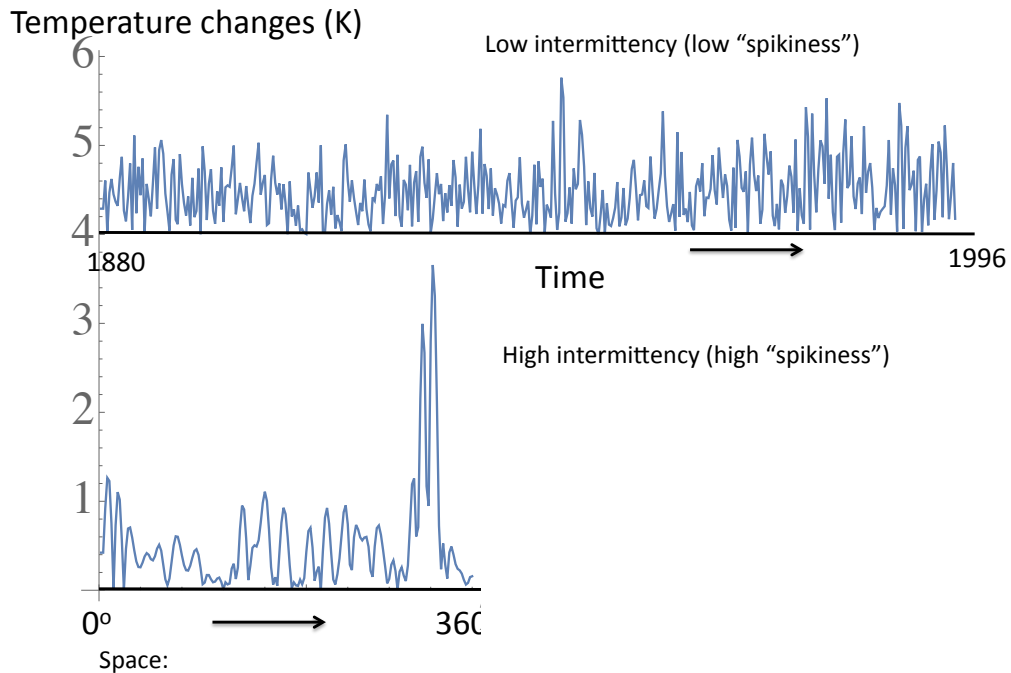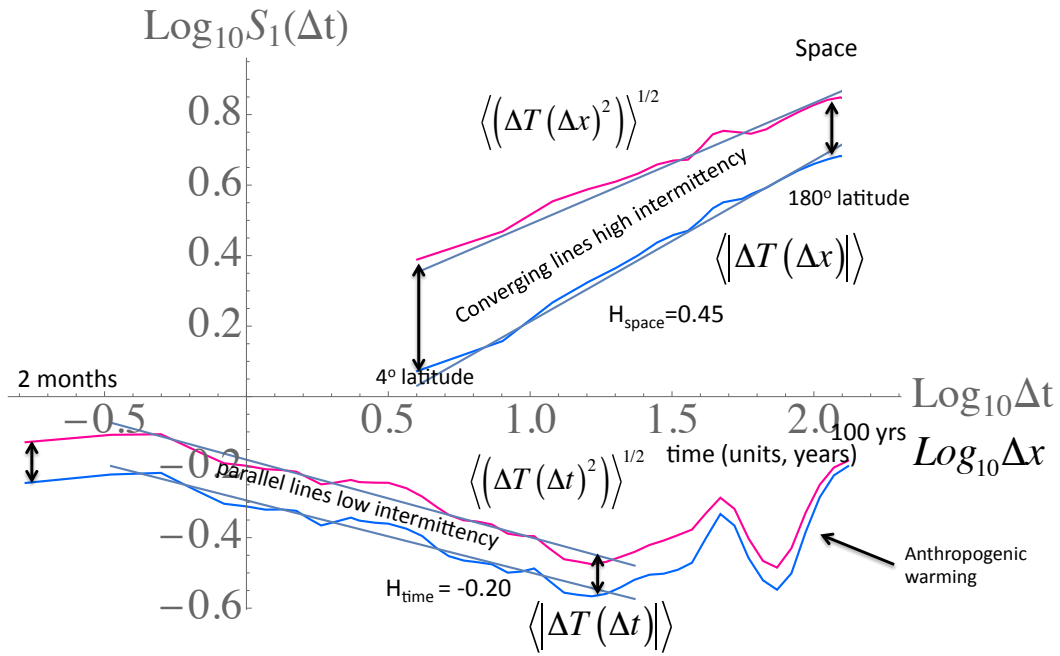
Fig. 8a:  A comparison of temporal and spatial macroweather series at $2^o$ resolution. The top are the absolute first differences of a temperature time series at monthly resolution (from $80^o$ E, $10^o$ N, 1880 -1996, displaced by 4K for clarity), and the bottom is the series of absolute first differences of a spatial latitudinal transect (annually averaged, 1990 from $60^o$ N), as a function of longitude.  Both use data from the 20CR.  One can see that while the top is noisy, it is not very "spikey".

$Log_{10}S_1(\Delta t)$

Space

$\left\langle\left(\Delta T(\Delta x)^2\right)\right\rangle^{1/2}$

0.8

0.6

0.4

0.2

180° latitude

$\left\langle\left|\Delta T(\Delta x)\right|\right\rangle$

Converging lines high intermittency

$H_{space}=0.45$

2 months

4° latitude

$-0.5$            0.5            1.0            1.5            2.0            $Log_{10}\Delta t$

100 yrs            $Log_{10}\Delta x$

time (units, years)

$\left\langle\left(\Delta T(\Delta t)^2\right)\right\rangle^{1/2}$

parallel lines low intermittency

$-0.4$

$-0.6$

$H_{time}$ = -0.20

$\left\langle\left|\Delta T(\Delta t)\right|\right\rangle$

Anthropogenic warming

top is spatial fluctuations, bottom temporal at equator

Fig. 8b:  The first order and RMS Haar fluctuations of the series and transect from fig. 8a.  One can see that in the spikey transect, the fluctuation statistics converge at large lags (time scale $\Delta t$), the rate of the converge is quantified by the intermittency parameter $C_1$.  The series (bottom) is less spikey, converges very little and has low $C_1$ (see fig. 8c).
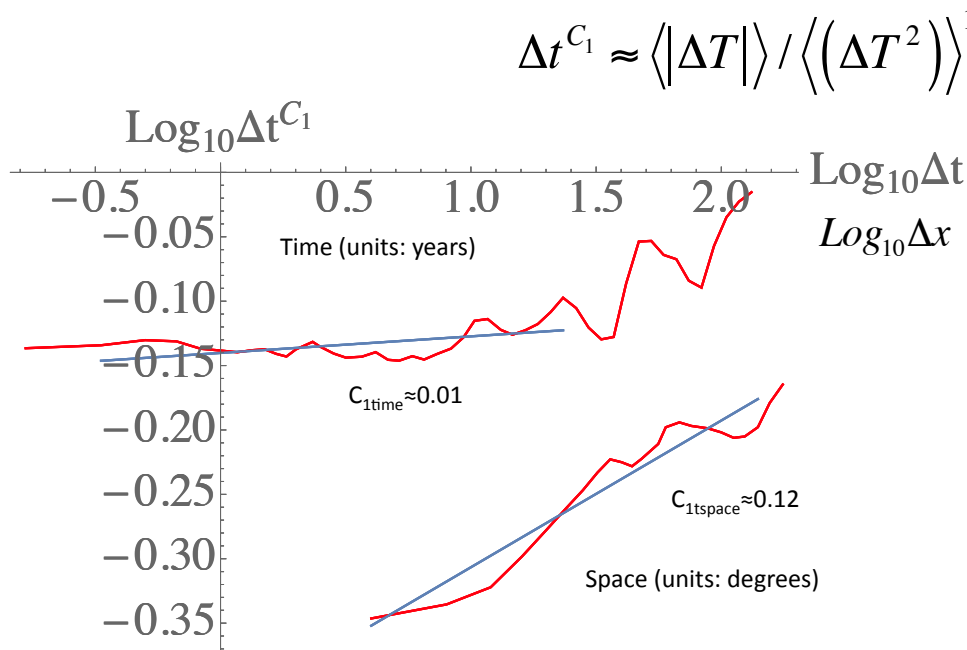
$$\Delta t^{C_1} \approx \left\langle \left| \Delta T \right| \right\rangle / \left\langle \left( \Delta T^2 \right) \right\rangle^{1/2}$$



Fig. 8c: A comparison of the intermittency function $F = \left\langle \left| \Delta T \right| \right\rangle \left( \left\langle \left| \Delta T \right|^{1+\Delta q} \right\rangle \right) / \left( \left\langle \left| \Delta T \right|^{1-\Delta q} \right\rangle \right)^{1/\Delta q}$ for the series and transect in the figs 8a, b, quantifying the difference in intermittencies: in time $C_1 \approx 0.01$, in space, $C_1 \approx 0.12$. Since $K'(1) = C_1$, when $\Delta q$ is small enough (here, $\Delta q = 0.1$ was used), we have $F(\Delta t) = \Delta t^{C_1}$. The break in the temporal scaling at about 20 – 30 years is due to anthropogenic forcings.

## 2.4 Scaling, space-time statistical factorization and size-lifetime relations

In the previous section we saw that there was evidence for scaling separately both in space and in time with the former being highly intermittent (multifractal) and the latter being nearly Gaussian (fig. 8). However, in order to make stochastic macroweather forecasts, we need to understand the *joint* space-time macroweather statistics and these turn out to be quite different from those in the weather regime. For the latter, recall that there exist well-defined statistical relations between weather structures ("meso-scale complexes", "storms", "turbulence" etc.) of a given size $L$ and their lifetimes $\tau$. Indeed, the textbook space-time "Stommel" diagrams that adorn introductory meteorology textbooks show log spatial scale versus log temporal scale plots with boxes or circles corresponding to different morphologies and phenomenologies and these typically occupy the diagonals. These diagrams are

usually interpreted as implying that each factor of two or so in spatial scale corresponds to fundamentally different dynamical processes, each with its own typical spatial extent and corresponding lifetime. However, as pointed out in [*Schertzer et al.*, 1997], the part of the diagram occupied by realistic structures and processes are typically not only on diagonals (implying a scaling space-time relation), but are on the precise diagonal whose slope has the value 2/3, theoretically predicted by the (Lagrangian, co-moving) size-lifetime relation discussed above: $\tau = \varepsilon^{-1/3} L^{2/3}$. The usual interpretation is an example of the "phenomenological fallacy" [*Lovejoy and Schertzer*, 2007]: rather than refute the scaling hypothesis, the Stommel diagrams support it.

As usual, the Eulerian (fixed frame) space-time relations are much easier to determine empirically, although theoretically their relation to Lagrangian statistics is not trivial. In a series of papers based on high resolution lidar data ([*Lilley et al.*, 2008; *Lovejoy et al.*, 2008; *Radkevitch et al.*, 2008] and then geostationary IR data (fig. 3, [*Pinel et al.*, 2014]), an argument by [*Tennekes*, 1975] about the small structures being "swept" by larger ones was extended to the (atmospheric) case assuming that there was no scale separation between small and large horizontal scales. It was concluded that the corresponding Eulerian (i.e. fixed frame) space-time relation generally had space-time spectra of the form:

$$P_{xyt}\left(k_x, k_y, \omega\right) = \left[\!\!\left[\left(k_x, k_y, \omega\right)\right]\!\!\right]^{-s} \tag{12}$$

where $P_{xyt}$ is the space-time spectra density:

$$P_{xyt}\left(k_x, k_y, \omega\right) \propto \left\langle \left|\tilde{T}\left(k_x, k_y, \omega\right)\right|^2 \right\rangle \tag{13}$$

and $\left[\!\!\left[\left(k_x, k_y, \omega\right)\right]\!\!\right]$ is the wavenumber $(k_x, k_y)$ - frequency $(\omega)$ scale function nondimensionalized by the large scale turbulent velocities (i.e. using $\varepsilon$ and the size of the earth). The analogous (real space) second order joint space-time structure function statistics:

$$S_{xyt}\left(\Delta x, \Delta y, \Delta t\right) = \left\langle \Delta T\left(\Delta x, \Delta y, \Delta t\right)^2 \right\rangle \tag{14}$$

were of the form:

$$S_{xyt}\left(\Delta x, \Delta y, \Delta t\right) = \left[\!\!\left[\left(\Delta x, \Delta y, \Delta t\right)\right]\!\!\right]^{\xi(2)} \tag{15}$$

where $\left[\!\!\left[\left(\Delta x, \Delta y, \Delta t\right)\right]\!\!\right]$ is the real space (nondimensional) scale function for horizontal lag $(\Delta x, \Delta y)$ and temporal lag $\Delta t$. The scale functions relevant here satisfy the isotropic scaling: $\left[\!\!\left[\lambda^{-1}\left(\Delta x, \Delta y, \Delta t\right)\right]\!\!\right] = \lambda^{-1}\left[\!\!\left[\left(\Delta x, \Delta y, \Delta t\right)\right]\!\!\right]$ and $\left[\!\!\left[\lambda\left(k_x, k_y, \omega\right)\right]\!\!\right] = \lambda\left[\!\!\left[\left(k_x, k_y, \omega\right)\right]\!\!\right]$ where $\lambda$ is a scale reduction factor. This is directly confirmed in fig. 3 for IR radiances.

In the simplest cases (with no mean advection and ignoring weak scaling singularities associated with waves, [*Pinel and Lovejoy*, 2014]), and retaining only a single spatial lag $\Delta x$, and wavenumber $k_x$, the nondimensional scale functions reduce to the usual vector norms i.e. they are of the form:

$$\llbracket (\Delta x, \Delta t) \rrbracket = \left( \Delta x^2 + \Delta t^2 \right)^{1/2} \tag{16}$$

$$\llbracket (k_x, \omega) \rrbracket = \left( k_x^2 + \omega^2 \right)^{1/2} \tag{17}$$

With $s = d + \xi(2)$ with $d$ = the dimension of space-time, in this example $d = 2$.

In order to define a relationship between a structure of extent $L$ with the lifetime $\tau$, we can use $S_{xt}$. For example, if we wait at a fixed location ($\Delta x = 0$) for a time $\tau$, we may ask what distance $L$ must we go at a given instant ($\Delta t = 0$) in order to expect the same typical fluctuation? This gives us an implicit relation between $L$ and $\tau$: $S_{xt}(0, \tau) = S_{xt}(L, 0)$; in this simple case (eqs. 15, 16) this implies $\tau = L$ for the nondimensional variables so that the dimensional relationship would correspond to a constant speed relating space and time. A similar relation would be obtained by using the same argument in Fourier space on the spectral density $P$.

What is the space-time relation in macroweather where we consider temporal averages over periods $> \tau_w$, typically months or longer? In this case, we average over many lifetimes of structures of all sizes, so clearly size-lifetime relations valid in the weather regime must break down. [*Lovejoy and Schertzer*, 2013] and [*Lovejoy and de Lima*, 2015] argued on theoretical, numerical and empirical grounds that – at least to a good approximation – that the result is statistical space-time factorization (SSTF). The application of the SSTF to the second order statistics means:

$$P_{xt}(k_x, \omega) = P_x(k_x) P_t(\omega)$$

$$R_{xt}(\Delta x, \Delta t) = R_x(\Delta x) R_t(\Delta t); \tag{18}$$

Note that in real space we have used correlation functions $R_{xt}(\Delta x, \Delta t) = \langle T(t, x) T(t - \Delta t, x - \Delta x) \rangle$ rather than Haar structure functions $S$; in macroweather ($H < 0$) they are essentially equivalent. However for small lags in time, one effectively goes outside the macroweather regime and $\Delta t = 0$ is problematic. When both $H_t < 0$ and $H_x < 0$ we can avoid issues that arise at small $\Delta t$, $\Delta x$ by using correlation functions (fig. 9a) (for the case $H_t < 0$, $H_x > 0$, see section 10.3 of [*Lovejoy and Schertzer*, 2013]).

Using the autocorrelations to obtain space-time macroweather relations, we obtain $R_{xt}(0, \tau) = R_{xt}(L, 0)$ so that using factorization and the identity $R_t(0) = R_x(0)$ the implicit $\tau$-$L$ relation is:

$$R_t(\tau) = R_x(L) \tag{19}$$

This is valid if both space and time have $H < 0$; if there is scaling, we have $R_t(\tau) \propto \tau^{H_t}$ and $R_x(L) \propto L^{H_x}$ with exponents $H_t < 0$, $H_x < 0$. The lifetime of a macroweather structure of size $L$ is thus:

$$\tau \propto L^{H_x/H_t} \tag{20}$$

Which - unless $H_x = H_t$ - is quite different from the lifetime size relationship in the weather regime; fig. 9a shows that $\tau \propto L^{0.65}$, for macroweather temperature and

precipitation.  Fig. 9a, shows that empirically the factorization works well for both temperature and precipitation data, and fig. 9b shows that it is also (even better) obeyed by the GISS E2R GCM; Del Rio Amador 2017 shows that it holds very accurately for 36 CMIP5 control runs.

It turns out that the SSTF is important for macroweather forecasting.  This is because, using means square estimators, since it implies that no matter how strong the correlations (teleconnections), if one has long time series at each point, pixel or region, that no further improvement can be made in the forecast by adding co-predictors such as the temperature data at other locations (Del Rio Amador et al 2017).  This effectively means that the original nonlinear initial value PDE problem has been effectively transformed into a linear but fractional ordered ODE "past value" problem, we pursue this in the next sections.
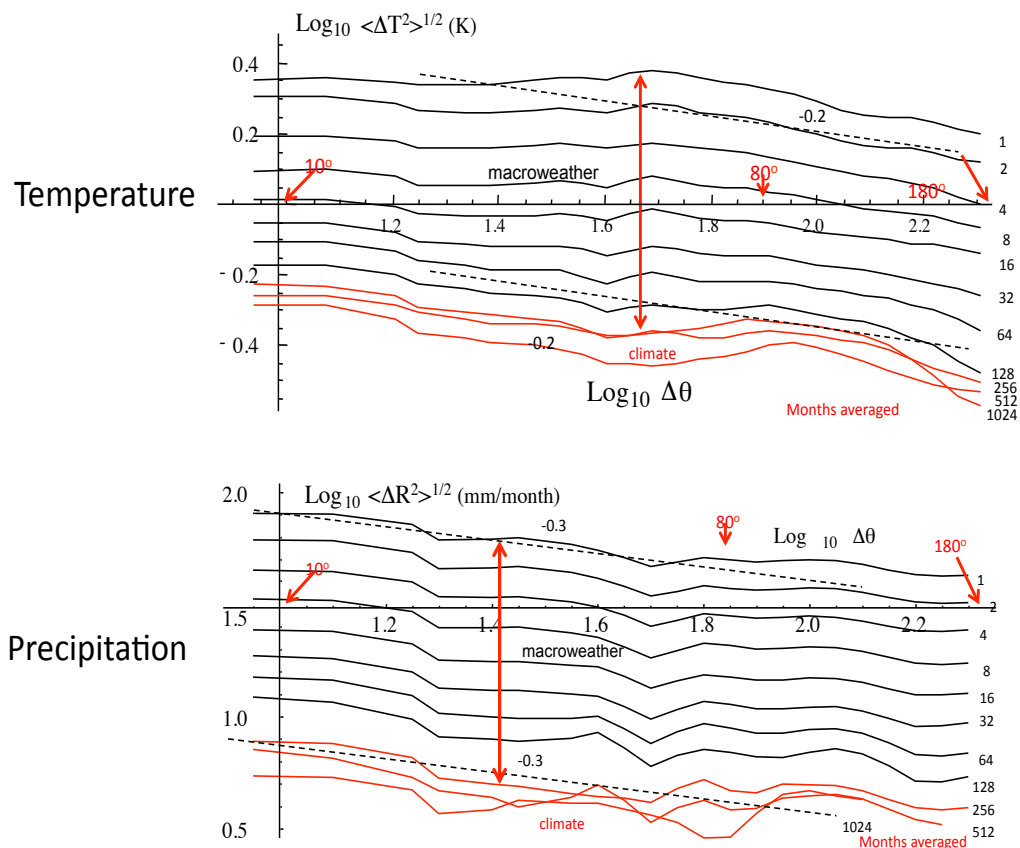


Fig. 9a: The joint space ($\Delta\theta$ i.e. angle subtended) time ($\Delta t$) RMS fluctuations of temperature (top, adapted from [*Lovejoy*, 2017]) and precipitation (bottom, adapted from [*Lovejoy and de Lima*, 2015]).  In both cases, zonal spatial anomaly fluctuations are given for data averaged over 1, 2, 4, .. 1024 months (since the temporal $H<0$ this is an anomaly fluctuation).  The temperature data are from the HadCRUtemp3 data base and the precipitation data from the Global Historical Climate Network, both at 5°, monthly resolutions and spanning the 20th century.  On

this log-log plot, SSTF implies $S_{\theta t}(\Delta\theta,\Delta t)=S_{\theta}(\Delta\theta)S_{t}(\Delta t)$ so that the curves will be parallel. If in addition, they respect spatial scaling then they will be linear and if they respect the temporal scaling then as we double the temporal resolution (top to bottom), they will be equally spaced (separated by log $2^H$).  Eventually (red), the temporal scaling breaks down (at $\tau_c \approx 256$ months).  Over the regimes where both SSTF and scaling hold we have for temperature, $S_{\theta,t}(\Delta\theta,\Delta t)\approx\Delta\theta^{-0.2}\Delta t^{-0.3}$ and for precipitation $S_{\theta,t}(\Delta\theta,\Delta t)\approx\Delta\theta^{-0.3}\Delta t^{-0.4}$.  The double headed red arrows show the corresponding total predicted range over macroweather time scales.
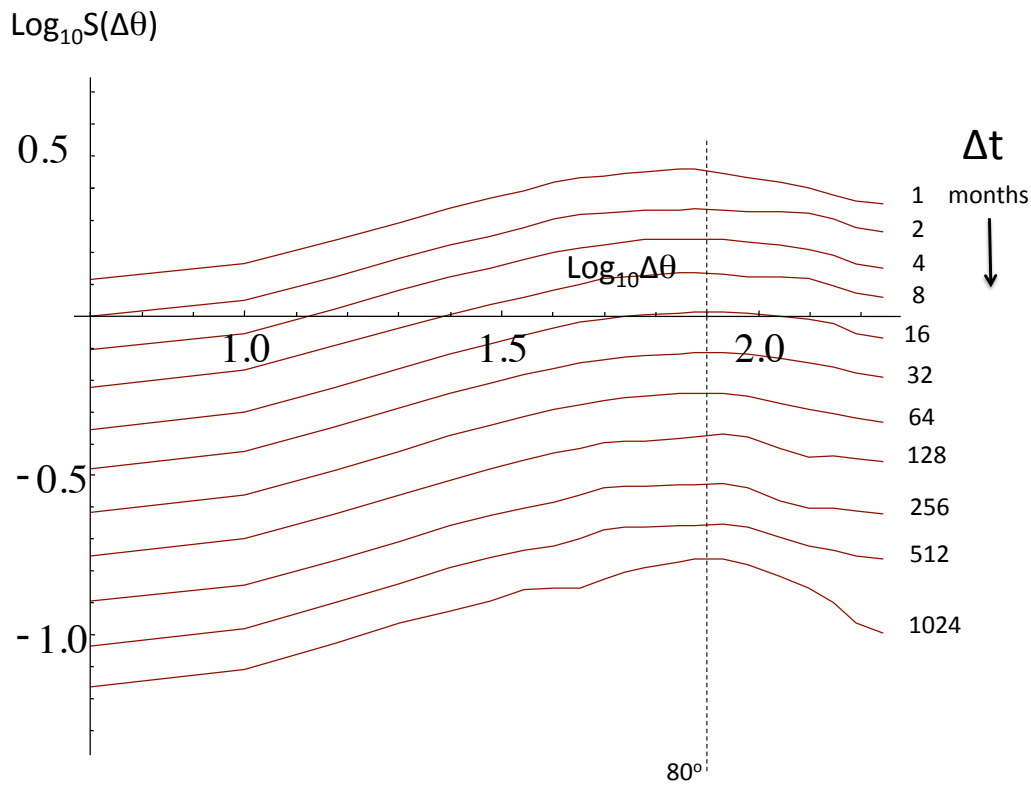


Fig. 9b: The same as fig. 9a, but for temperature fluctuations from GISS-E2R historical simulations from 1850.  In this case, rather than using anomalies (which were the only data available for fig. 9a), we used the difference between two realizations of the same historical simulation (i.e. with identical external boundary conditions) obtained by slightly varying the initial conditions.  The temporal behaviour of this plot shows rapidly the model climate is approached under temporal averaging, and how it varies as a function of angular scale.  Again we see that the joint fluctuations have nearly exactly the same shapes (confirming SSTF); over the ranges where the scaling holds, the joint structure function is: $S_{\theta,t}(\Delta\theta,\Delta t)\approx\Delta\theta^{0.3}\Delta t^{-0.4}$.  This plot shows that GCM's obey the SSTF very accurately, a fact confirmed in section 4 by the success by which they can be predicted by SLIMM.

## 3. Macroweather forecasting

### 3.1 The fractional Gaussian noise model and some of its characteristics

We have argued that macroweather is scaling but with low intermittency, so that a Gaussian forecasting model may be an acceptable approximation. The simplest such model is fractional Gaussian noise (fGn). We now give a brief summary of some useful properties of fGn; for a longer review see [*Lovejoy et al.*, 2015] and for an extensive mathematical treatment see [*Biagini et al.*, 2008].

Over the parameter range of interest -1/2<H<0, fGn is essentially a smoothed Gaussian white noise and it's mathematical definition raises similar issues. For our purposes, it most straightforward to use the framework of generalized functions and start with the unit Gaussian white noise $\gamma(t)$ which has $\langle\gamma\rangle = 0$ and is "$\delta$ correlated":

$$\langle\gamma(t)\gamma(t')\rangle = \delta(t-t') \tag{21}$$

where "$\delta$" is the Dirac function. The $H$ parameter fGn $G_H(t)$ is thus:

$$G_H(t) = \frac{c_H}{\Gamma(1/2+H)}\int_{-\infty}^{t}(t-t')^{-(1/2-H)}\gamma(t')dt'; \quad -1 < H < 0 \tag{22}$$

The constant $c_H$ is a constant chosen so as to make the expression for the statistics particularly simple, see below. Mathematically $\gamma(t)$ is thus the density of the Wiener process $W(t)$, often written $\gamma(t)dt = dW$: just as the Dirac function is only meaningful when integrated, the same is true of $\gamma(t)$. For fGn, we shall see below that $G_H(t)dt = dB_{H'}$ where $B_{H'}$ is a generalization of the Wiener process, fractional Brownian motion (fBm, parameter $H' = 1+ H$) and $B_{H'}$ reduces to a Wiener process when $H' = 1/2$. $G_H(t)$ is thus the (singular) density of an fBm measure. In practice, we will always consider $G_H(t)$ smoothed over finite resolutions so that whether we define $G_H(t)$ indirectly via fBm or directly as a smoothing of eq. 21 the result is equivalent.

We can see by inspection of eq. 21, that $G_H(t)$ is statistically stationary and by taking ensemble averages of both sides of eq. 21 we see that the mean vanishes: $\langle G_H(t)\rangle = 0$. When $H = -1/2$, the process $G_{-1/2}(t)$ itself is simply a Gaussian white noise. Although we justified the use of fGn as the simplest scaling process, it could also be introduced as the solution of a stochastic fractional ordered differential equation:

$$\frac{d^{H+1/2}T}{dt^{H+1/2}} = \gamma(t) \tag{23}$$

the solution of which is $T(t) \propto G_H(t)$.

Now, take the average of $G_H$ over $\tau$, the "$\tau$ resolution anomaly fluctuation":

$$G_{H,\tau}(t) = \frac{1}{\tau}\int_{t-\tau}^{t}G_H(t')dt' \tag{24}$$

If $c_H$ is now chosen such that:

$$c_H = \left( \frac{\pi}{2\cos(\pi H)\Gamma(-2H-2)} \right)^{1/2} \tag{25}$$

then we have:

$$\left\langle G_{H,\tau}(t)^2 \right\rangle = \tau^{2H}; \quad -1 < H < 0 \tag{26}$$

This shows that a fundamental property of fGn is that in the small scale limit ($\tau$->0), the variance diverges and $H$ is scaling exponent of the root mean square (RMS) value. This singular small scale behaviour is responsible for the strong power law resolution effects in fGn. Since $\left\langle G_H(t) \right\rangle = 0$, sample functions $G_{H,\tau}(t)$ fluctuate about zero with successive fluctuations tending to cancel each other out; this is the hallmark of macroweather.

*Anomalies:*

An anomaly is the average deviation from the long term average and since $\left\langle G_H(t) \right\rangle = 0$, the anomaly fluctuation over interval $\Delta t$ is simply $G_H$ at resolution $\Delta t$ rather than $\tau$:

$$\left(\Delta G_{H,\tau}(\Delta t)\right)_{anom} = \frac{1}{\Delta t}\int_{t-\Delta t}^{t} G_{H,\tau}(t')dt' = \frac{1}{\Delta t}\int_{t-\Delta t}^{t} G_H(t')dt' = G_{H,\Delta t}(t); \quad \Delta t > \tau \tag{27}$$

Hence using eq. 26:

$$\left\langle \left(\Delta G_{H,\tau}(\Delta t)\right)_{anom}^2 \right\rangle = \Delta t^{2H}; \quad -1 < H < 0 \tag{28}$$

*Differences:*

In the large $\Delta t$ limit we have:

$$\left\langle \left(\Delta G_{H,\tau}(\Delta t)\right)_{diff}^2 \right\rangle \approx 2\tau^{2H}\left(1 - (H+1)(2H+1)\left(\frac{\Delta t}{\tau}\right)^{2H}\right) \tag{29}$$

Since $H<0$, the differences asymptote to the value $2\tau^{2H}$ (double the variance). Notice that since $H<0$, the differences are not scaling with $\Delta t$.

*Haar Fluctuations*

For the Haar fluctuation we obtain:

$$\left\langle \left(\Delta G_{H,\tau}(\Delta t)\right)_{Haar}^2 \right\rangle = 4\Delta t^{2H}\left(2^{-2H} - 1\right); \quad \Delta t \geq 2\tau \tag{30}$$

this scales as $\Delta t^{2H}$ and does not depend on the resolution $\tau$. This relation can be used to estimate the spatial variation of $H$, fig. 10 gives the sptial distribution using 20CR data. It can be seen that $H$ is near zero over the oceans and is lower over land,

typical values being -0.1 and -0.3 respectively.   Below, we see that this corresponds to large memory (and hence forecast skill) over oceans and lower memory and skill over land.
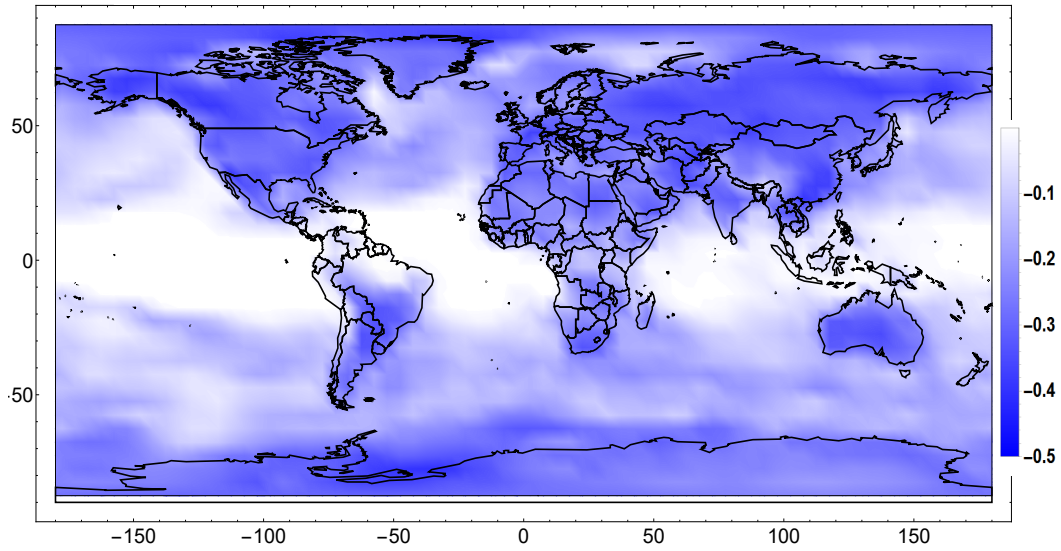


Fig.10:  The spatial distribution of the exponent $H$ estimated at $5^{o}X5^{o}$ resolution using monthly resolution data from the NCEP reanalyses (1948- 2010) and estimated by a maximum liklihood method.  The mean was -0.11±0.09.

*Autocorrelations:*

$$\langle G_{\tau,H}(t)G_{\tau,H}(t-\Delta t)\rangle = \quad R(\widehat{\Delta t})=\frac{\tau^{2H}}{2}\left[\left(\widehat{\Delta t}+1\right)^{2H+2}+\left(\widehat{\Delta t}-1\right)^{2H+2}-2\widehat{\Delta t}^{2H+2}\right]; \quad \widehat{\Delta t}=\frac{\Delta t}{\tau} \quad (31)$$

*Spectra:*

Since fGn is stationary, its spectrum is given by the Fourier transform of the autocorrelation function.  Note that in the above, $\Delta t>0$; since the autocorrelation is symmetric for the Fourier transform with respect to $\Delta t$, we use the absolute value of $\Delta t$.  We obtain:

$$E(\omega)=\frac{\Gamma(3+2H)\sin\pi H}{\sqrt{2\pi}}|\omega|^{-\beta}; \quad \beta=1+2H$$

$$(32)$$

*Relation to fBm:*

It is more common to treat fBm whose differential $dB_{H'}(t)$ is given by:

$$dB_{H'} = G_H(t)dt; \quad H' = H+1; \quad 0 < H' < 1 \tag{33}$$

so that:

$$\Delta B_{H'}(\tau) = B_{H'}(t) - B_{H'}(t-\tau) = \int_{t-\tau}^{t} G_{H'}(t')dt' = \tau G_{H',\tau}(t) \tag{34}$$

with the property:

$$\left\langle \Delta B_{H'}(\Delta t)^2 \right\rangle = \Delta t^{2H'} \tag{35}$$

(eq. 29). While this defines the increments of $B_{H'}(t)$ and shows that they are stationary, it does not completely define the process. For this, one conventionally imposes $B_{H'}(0)=0$, and this leads to the usual definition:

$$B_{H'}(t) = \frac{c_{H'}}{\Gamma(H'+1/2)} \int_{-\infty}^{0} \left( (t-s)^{H'-1/2} - (-s)^{H'-1/2} \right) \gamma(s)ds + \frac{c_{H'}}{\Gamma(H'+1/2)} \int_{0}^{t} (t-s)^{H'-1/2} \gamma(s)ds \tag{36}$$

[*Mandelbrot and Van Ness*, 1968]. Whereas fGn has a small scale divergence that can be eliminated by averaging over a finite resolution $\tau$, the fGn integral $\int_{-\infty}^{t} G_H(t')dt'$ on the contrary has a low frequency divergence. This is the reason for the introduction of the second term in the first integral in eq. 36: it eliminates this divergence at the price of imposing $B_{H'}(0) = 0$ so that fBm is nonstationary (although its increments are stationary, eq. 34).

A comment on the parameter $H$ is now in order. In treatments of fBm, it is usual to use the parameter $H$ confined to the unit interval i.e. to characterize the scaling of the increments of fBm. However, fBm (and fGn) are very special scaling processes, and even in low intermittency regimes such as macroweather – they are at best approximate models of reality. Therefore, it is better to define $H$ more generally as the fluctuation exponent (eq. 6); with this definition $H$ is also useful for more general (multifractal) scaling processes although the interpretation of $H$ as the "Hurst exponent" is only valid for fBm). When $-1<H<0$, the mean at resolution $\tau$ (eq. 24) defines the anomaly fluctuation, so that $H$ is equal to the fluctuation exponent for fGn, in contrast, for processes with $0<H<1$, the fluctuations scale as the mean differences and eq. 35 shows that $H'$ is the fluctuation exponent for fBm. In other words, as long as an appropriate definition of fluctuation is used, $H$ and $H' = 1+H$ are fluctuation exponents of fGn, fBm respectively. The relation $H' = H+1$ follows because fBm is an integral order 1 of fGn. Therefore, since the macroweather fields of interest have fluctuations with mean scaling exponent $-1/2<H<0$, we use $H$ for the fGn exponent and $\frac{1}{2}<H'<1$ for the corresponding integrated fBm process.

We can therefore define the resolution $\tau$ temperature as:

$$T_{\tau}(t) = \sigma_T G_{H,\tau}(t) = \sigma_T \frac{B_{H'}(t) - B_{H'}(t-\tau)}{\tau} \tag{37}$$

Using eq. 26, the $\tau$ resolution temperature variance is thus:

$$\left\langle T_\tau^2 \right\rangle = \sigma_T^2 \tau^{2H} \tag{38}$$

From this and the relation $T_\tau(t) = \sigma_T G_{H,\tau}(t)$, we can trivially obtain the statistics of $T_\tau(t)$ from those of $G_{H,\tau}(t)$.

## 3.2 Mean Square (MS) estimators for fGn and the ScaLIng Macroweather Model (SLIMM)

The Mean Square (MS) estimator framework is a general framework for predicting stochastic processes, it determines predictors that minimize the prediction error variance, see e.g. [*Papoulis*, 1965]. Since Gaussian processes are completely determined by their second order statistics, the MS framework therefore gives optimum forecasts for fGn.

Our problem is to use data $T_\tau(s)$ at times $s<0$ (or equivalently, the innovations $\gamma(s)$) to predict the future temperature $T_\tau(t)$ at times $t>0$. Denoting this predictor by $\hat{T}_\tau(t)$ MS theory then shows that the latter is given by a linear combination of data, i.e. either the $T_\tau(s)$ or equivalently by a linear combination of past white noise "innovations" $\gamma(s)$:

$$\hat{T}_\tau(t) = \int_{-\tau_0 < s \leq 0} M_T(t,s) T_\tau(s) ds$$

$$\hat{T}_\tau(t) = \int_{-\tau_0 < s \leq 0} M_\gamma(t,s) \gamma(s) ds \tag{39}$$

where $M_T$, $M_\gamma$ are the predictor kernels based on past temperatures and past innovations respectively, and the range of integration is over all available data, the range $-\tau_0 < s \leq 0$. The simplest problems are those where the range extends to the infinite past ($\tau_0 \to \infty$), but practical predictions require the solution for finite $\tau_0$.

The prediction error is thus:

$$E_T(t) = T_\tau(t) - \hat{T}_\tau(t) \tag{40}$$

and from MS theory, the basic condition imposed by minimizing the error variance $\left\langle E_T^2(t) \right\rangle$ is:

$$\left\langle E_T(t)\hat{T}_\tau(t) \right\rangle = \left\langle E_T(t)T_\tau(s) \right\rangle = \left\langle E_T(t)\gamma(s) \right\rangle = 0; \quad t > 0; \quad s \leq 0 \tag{41}$$

This equation states that the (future) prediction error $E_T(t)$ is statistically independent of the predictor $\hat{T}_\tau(t)$ or, equivalently, it is independent of the past data $T_\tau(s)$, $\gamma(s)$ upon which the predictor is based. This makes intuitive sense: if there was a nonzero correlation between the available data and the prediction error, then there would still information in the data that could be used to improve the predictor and reduce the error. Since GCM forecasts are not MS, they do not satisfy this orthogonality condition. On the one hand, this explains how they can have negative skill (see below), on the other, it justifies complex GCM post-processing

that exploit past data to reduce the errors.  Indeed, a condition used to optimize post-processing corrections is actually close to the orthogonality condition.

In [*Lovejoy et al.*, 2015], the mathematically simplest predictor was given in the case of infinite past data but using the innovations $\gamma(s)$:

$$\hat{T}_\tau(t) = \int_{-\infty}^{0} M_\gamma(t,s)\gamma(s)ds$$

$$M_\gamma(t,s) = \frac{c_H \sigma_T}{\tau \Gamma(H+3/2)}\left[(t-s)^{H+1/2} - (t-\tau-s)^{H+1/2}\right]$$

(42)

The error is:

$$E_T = T_\tau(t) - \hat{T}_\tau(t) = \frac{c_H \sigma_T}{\tau \Gamma(H+3/2)}\left[\int_0^t (t-s)^{H+1/2}\gamma(s)ds - \int_0^{t-\tau}(t-\tau-s)^{H+1/2}\gamma(s)ds\right]$$

(43)

Since $\hat{T}(t)$ depends only on $\gamma(s)$ for $s<0$ and $E_T$ on $\gamma(s)$ for $s>0$, it can be seen by inspection that the orthogonality condition (eq. 41) holds.  Using this MS predictor, we can define the Mean Square Skill Score (MSSS) or "skill" for short:

$$MSSS = S_k(t) = 1 - \frac{\left\langle E_T(t)^2 \right\rangle}{\left\langle T_\tau(t)^2 \right\rangle}$$

(44)

For MS forecasts, we can use the orthogonality condition to obtain equivalently;

$$S_k(t) = \frac{\left\langle \hat{T}_\tau^2(t) \right\rangle}{\left\langle T_\tau^2(t) \right\rangle}$$

(45)

which shows that for MS forecasts, the skill is the same as the fraction of the variance explained by the predictor.

Using the predictor (eq. 42) we can easily obtain the skill for fGn forecasts:

$$S_k(\lambda) = \left[\frac{F_H(\infty) - F_H(\lambda)}{F_H(\infty) + \dfrac{1}{2H+2}}\right]; \qquad \lambda = t/\tau; \quad \lambda \geq 1$$

(46)

where the auxiliary function $F_H$ is given by:

$$F_H(\lambda) = \int_0^{\lambda-1}\left((1+u)^{H+1/2} - u^{H+1/2}\right)^2 du; \quad \lambda \geq 1$$

(47)

with:

$$F_H(\infty) = \pi^{-1/2}2^{-(2H+2)}\Gamma(-1-H)\Gamma(3/2+H)$$

(48)

and the asymptotic expression:

$$F_H(\lambda) = F_H(\infty) - \frac{(H+1/2)^2}{-2H}\lambda^{2H} + \dots$$

(49)

[*Lovejoy et al.*, 2015].  For any system that has quasi-Gaussian statistics and scaling fluctuations with -1/2<*H*<0 the theoretical skill, eq. 46 represents a stochastic

predictability limit, of similar fundamental significance to the usual deterministic predictability limits arising from sensitive dependence on initial conditions. In section 4.2, we show that CMIP5 GCMs can indeed be predicted to nearly this limit using the MS approach outlined here.

Although the MSSS is commonly used for evaluating forecasts, the correlation coefficient between the hindcast and the temperature is occasionally used:

$$\rho_{\hat{T},T}(t,\tau) = \frac{\left\langle \hat{T}_\tau(t)T_\tau(t)\right\rangle - \left\langle \hat{T}_\tau(t)\right\rangle\left\langle T_\tau(t)\right\rangle}{\left\langle \hat{T}_\tau(t)^2\right\rangle^{1/2}\left\langle T_\tau(t)^2\right\rangle^{1/2}}$$

(50)

Since $<T> = 0$, the upper right cross term vanishes and using orthogonality $\left\langle T_\tau(t)\hat{T}_\tau(t)\right\rangle = \left\langle \hat{T}_\tau(t)^2\right\rangle$ we obtain:

$$\rho_{\hat{T},T}(t,\tau) = S_k(t,\tau)^{1/2}$$

(51)

Therefore, MS forecast skill can equivalently be quantified using either correlations or MSSS.
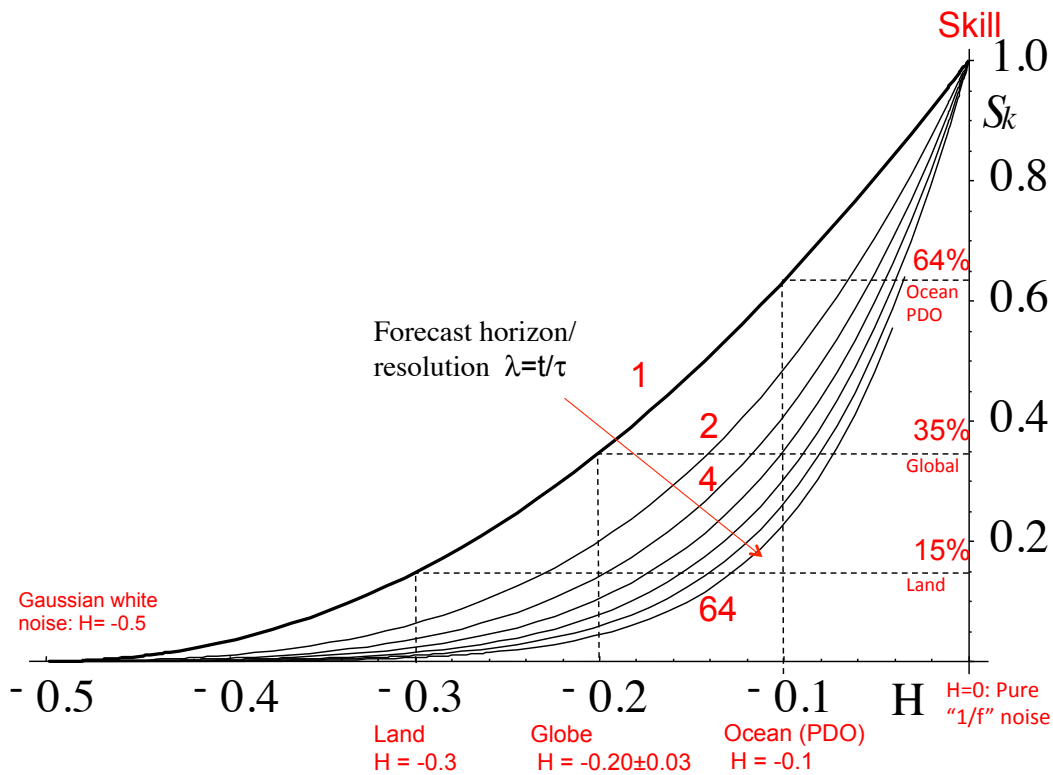


Fig. 11a: Forecast skill for nondimensional forecast horizons λ = (horizon/resolution) = 1,2 4, 8,…64 (left to right) as functions of *H*. For reference, the rough empirical values for land, ocean and the entire globe (the value used here, see below) are indicated by dashed vertical lines. The horizontal lines show the fraction of the variance explained (the skill, $S_k$, eq. 46) in the case of a forecast of

resolution $\tau$ data at a forecast horizon $t = \tau$ ($\lambda = 1$; corresponding to forecasting the anomaly fluctuation one time unit ahead).

Fig. 11a shows the theoretical skill as a function of $H$ for different forecast horizons.  To underscore the huge memory implied by the power law kernel $M_\gamma$, we can compare the fGn kernel with that of the exponential kernels that arise in autoregressive (AR) type processes.  This is relevant here since the main existing stochastic macroweather forecasts techniques ("Linear Inverse Modelling, LIM, see the next subsection) are vector AR processes that reduce to scalar AR processes in an appropriately (diagonalized) frame.  If for simplicity, we consider only forecasts one time step into the future (i.e. horizon $\tau$, for a process resolution $\tau$), then the fraction f($\lambda$) of the predictor variance that is due to innovations at times $\lambda\tau$  or further in the past,  can be written in the same form as for fGn:

$$f\left(\lambda\right) = \frac{I\left(\lambda\right)}{I\left(0\right)}; \quad I\left(\lambda\right) = \int_{-\infty}^{-\lambda}\left(g\left(s\right) - g\left(s-1\right)\right)^2 ds \tag{52}$$

where  $g\left(s\right) = \left(-s\right)^{1/2+H}$  for fGn  (for  SLIMM  predictions)  and  $g\left(s\right) = e^s$  for  AR processes.  The comparison is shown in fig. 11b, it can be seen that almost all the information needed to forecast an AR process is in the most recent 3 steps, whereas for SLIMM, with $H$ =-0.1 (appropriate for forecasting the globally averaged temperature), roughly 20% comes from innovations more than 1000 steps in the past.  Significantly, we will see that this does not mean that we need such long series to make good forecast; this is because even relatively short series with $H$ = -0.1 have information from the distant past; this is discussed below.
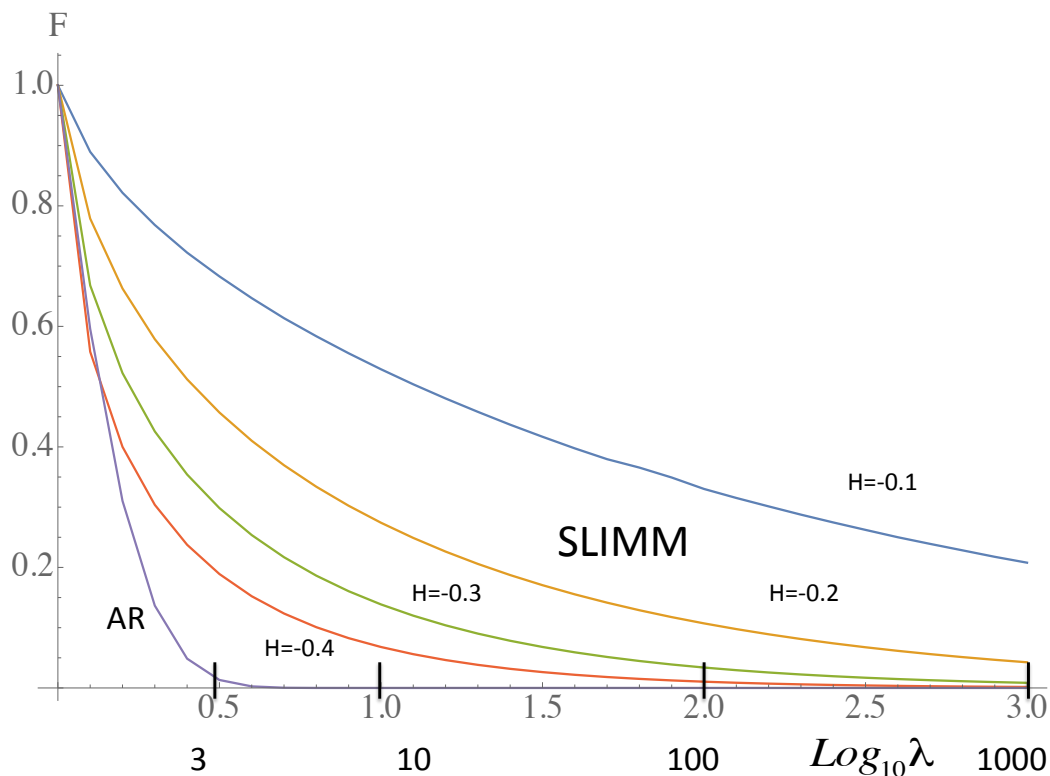
Fig.11b:  The fraction of the prediction variance of a forecast one time step ahead that is due to innovations further in the past than $\lambda$ time units (one unit= resolution $\tau$).   The right four curves are for SLIMM ($H$ = -0.1, -0.2, -0.3, -0.4), and the far left curve is for an autoregressive process $F = f(\lambda)$= Fraction of total memory used in forecasts one step into the future.

### 3.3 SLIMM prediction skill and alternative stochastic macroweather prediction systems

Following [*Hasselmann*, 1976] who proposed the use of stochastic differential equations to understand low frequency weather (i.e. macroweather), attempts have been made to use this for monthly, Seasonal to Interannual forecasts.  The basic idea is to model the atmosphere as an Ornstein-Uhlenbeck process, i.e. the solution of $\dfrac{dT}{dt}+T/\tau=\gamma(t)$ where $\tau$ is the basic time scale and $\gamma$ is a white noise forcing.  The idea is that the weather acts essentially as a random white noise perturbation to the temperature $T$.  Fourier analysis shows that the spectrum is $E(\omega)\propto 1/(\omega^2+\tau^{-2})$ so that at high frequencies, $E(\omega)\propto \omega^{-2}$ whereas at low frequencies, $E(\omega)\approx$ constant. The process is thus an (unpredictable) white noise; this can be seen directly by taking the low frequency limit $dT/dt\approx 0$  in the equation.   From an empirical point of view, there are two scaling regimes (exponents $\beta$ = 0, 2), corresponding to $H$ = ($\beta$-

1)/2 = -1/2 and H= ½, respectively but neither are realistic: for example, the true values for the temperature are closer to ≈ -0.1, ≈ 0.4 for macroweather, weather respectively with the former showing significant spatial variations, see fig. 10. The key point is that models based on integer order differential equations implicitly assume that the low frequencies are unpredictable whereas on the contrary, the temporal scaling implies long range dependencies, a large memory. From the point of view of differential equations, we thus require terms of fractional order (see eq. 22).

Over the decades, the Hasselman inspired approach has been significantly developed, in the framework of "Linear Inverse Modelling" (LIM), sometimes also called the "Stochastic Linear Framework" (SLF), although the latter is somewhat a misnomer since it restrictively excludes fractional ordered (but still linear) terms (for LIM, SLF see e.g. [*Penland*, 1996; *Penland and Sardeshmuhk*, 1995; *Sardeshmukh et al.*, 2000], [*Newman*, 2013]). The essential development is the extension of scalar Ornstein-Uhlenbeck processes to vector processes with each component being a significant macroweather variable (e.g. an El Nino index, an ocean temperature at a particular grid point etc.). Typical implementations such as described in [*Newman*, 2013] involve 20 components (implying hundreds of empirical parameters). When diagonalized, the system reduces to decoupled Ornstein-Uhlenbeck processes whose longest characteristic times are about 1 year, and beyond this, the system has little skill, see fig. 12a.

Because its theoretical basis is weak and it involves a large number of empirical parameters, LIM is an example of what is commonly termed an "empirically based" approach. Other such approaches have been proposed, notably by [*Suckling et al.*, 2016] and they have had some success by using carefully chosen climate indices that are linearly related to macroweather variables of interest and using empirically determined time delays. In contrast, SLIMM is based on fundamental space-time scale symmetries that we argue are respected by the dynamical equations.

In order to use SLIMM for forecasts, it is important to first remove the low frequency responses to anthropogenic forcings, failure to do so ([*Baillie and Chung*, 2002]) leads to poor results. For annually, globally averaged temperatures, it turns out that reasonable results can be obtained using the $CO_2$ radiative forcing (proportional to $logCO_2$ concentration) as a linear surrogate for all anthropogenic forcings (fig. 12b). SLIMM then forecasts the internal variability: the residuals. The reason that this works so well is presumably that all anthropogenic effects are linked through the economy and the economy is well characterized by energy use and hence by $CO_2$ emissions.

When SLIMM hindcasts are made for hemispheric and global scales, [*Lovejoy et al.*, 2015], they are generally better than LIM and GCM forecasts (fig. 12a). In addition, [*Lovejoy*, 2015b] made global scale SLIMM forecasts and showed that they could accurately (to within about ±0.05°C for three year anomalies) forecast the so-called "pause" in the warming (1998-present). In comparison, CMIP3 GCM predictions were about 0.2°C too high. While the cause of the GCM over-prediction is currently debated (e.g. [*Schmidt et al.*, 2014], [*Guemas et al.*, 2013], [*Steinman*

*et al.*, 2015]), the SLIMM prediction was successful largely because as fig. 12b shows, the pause was simply a natural cooling event that followed the enormous "pre-pause" 1992-1998 warming, with all of this superposed on a rising anthropogenic warming trend.

The forecast solution adopted in these publications showed that the fGn model was worth pursuing. However, the technique they used was based on $M_\gamma$, i.e. finding the optimum predictor using the innovations $\gamma(s)$ directly (obtained by numerically inverting eq. 22) and assuming that the available data extended into the infinite past. It is much more convenient to use the past data $T(s)$ and to take into account the fact that the past data are only finite in extent. Since an fGn process at resolution $\tau$ is the average of the increments of an fBm, process, it suffices to forecast fBm so that in the operational version of SLIMM described below, we therefore availed ourselves of the mathematical solution of the prediction problem of finding the kernel $M_T(t,s)$ in eq. 39) for both finite and infinite past data. [*Gripenberg and Norros*, 1996] mathematically solved the fBm solution with $\frac{1}{2}<H'<1$ and this was numerically investigated by [*Hirchoren and D'attellis*, 1998].
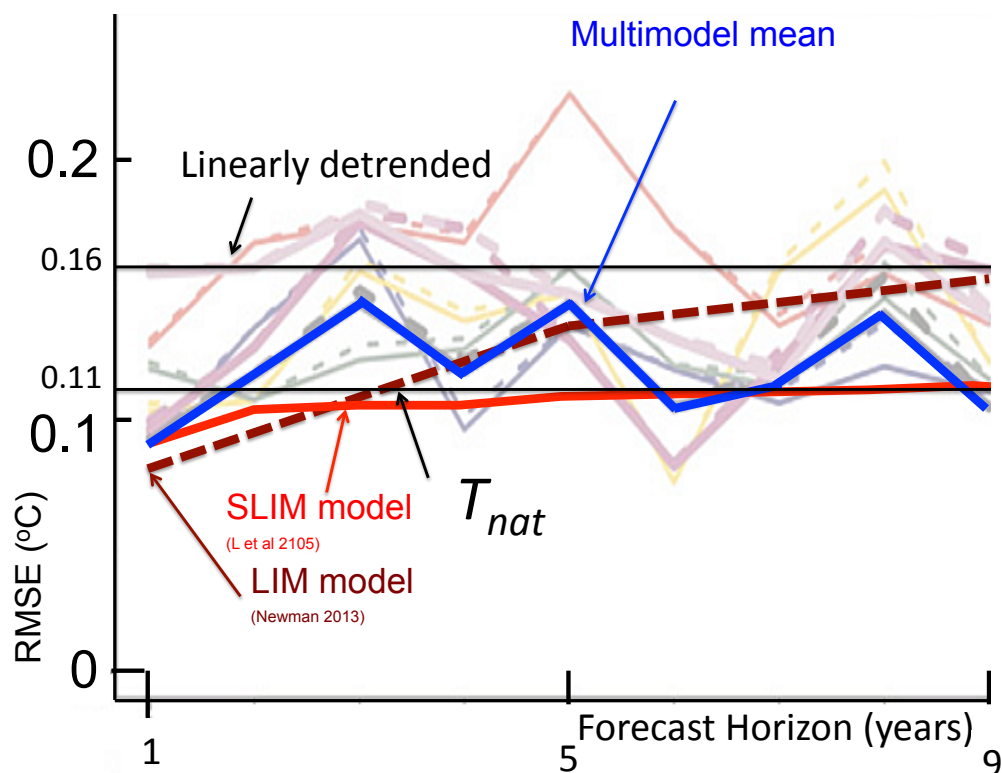


Fig. 12a: A comparison of Root Mean Square Error of hindcasts of various global annual temperatures for horizons of 1 to 9 years: the (GCMbased) ENSEMBLES experiment (from [*García-Serrano and Doblas-Reyes*, 2012], LIM [*Newman*, 2013] and SLIMM [*Lovejoy et al.*, 2015]. The light lines are from individual members of the ENSEMBLE experiment; the heavy line is the multimodel ensemble.

This shows the RMSE comparisons for the global mean surface temperatures compared to NCEP/NCAR (2m air temperatures). Horizontal reference lines indicate the standard deviations of $T_{nat}$ (bottom horizontal line, the RMS of the residuals after removing the anthropogenic forcing using the CO2 as a linear surrogate, itself nearly equivalent to the pre-industrial variability, [*Lovejoy*, 2014a]) and of the RMS of the residuals of the linearly detrended temperatures (top horizontal line). Also shown are the RMSE for the LIM model and the SLIMM. Adapted from [*Lovejoy et al.*, 2015].
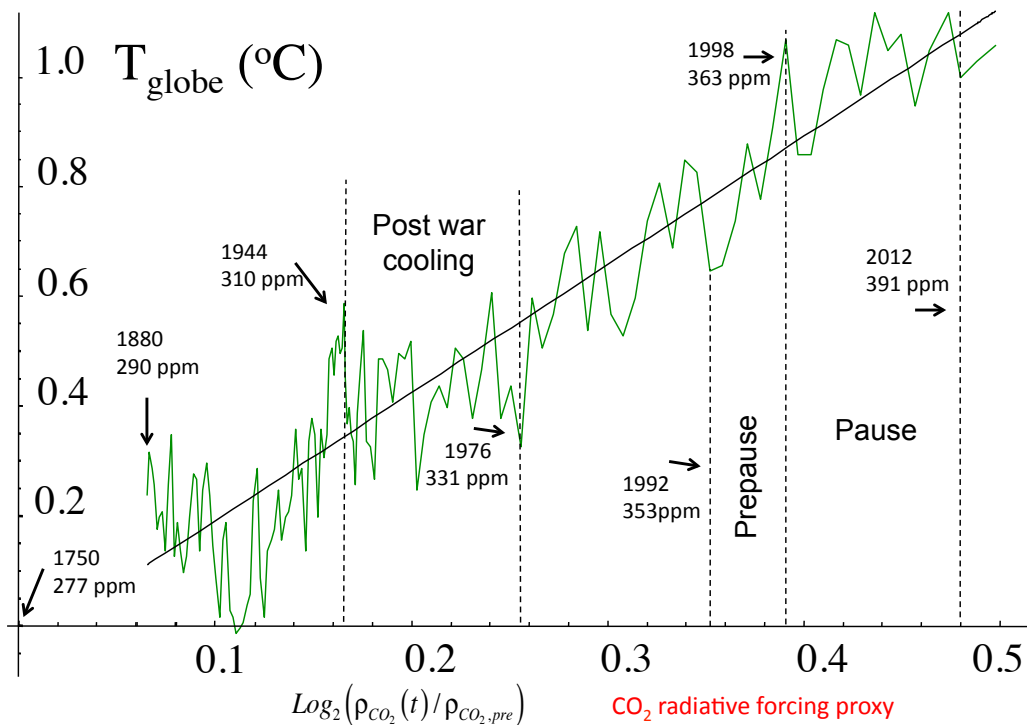


Fig. 12b:  The NASA GISS globally, annually averaged temperature series from 1880-2013 plotted as a function of $CO_2$ radiative forcing.  The regression slope indicated corresponds to 2.33±0.22 K/$CO_2$ doubling.  The internal variability forecast by SLIMM are the residuals (see fig. 12c).  Adapted from [*Lovejoy*, 2014b].
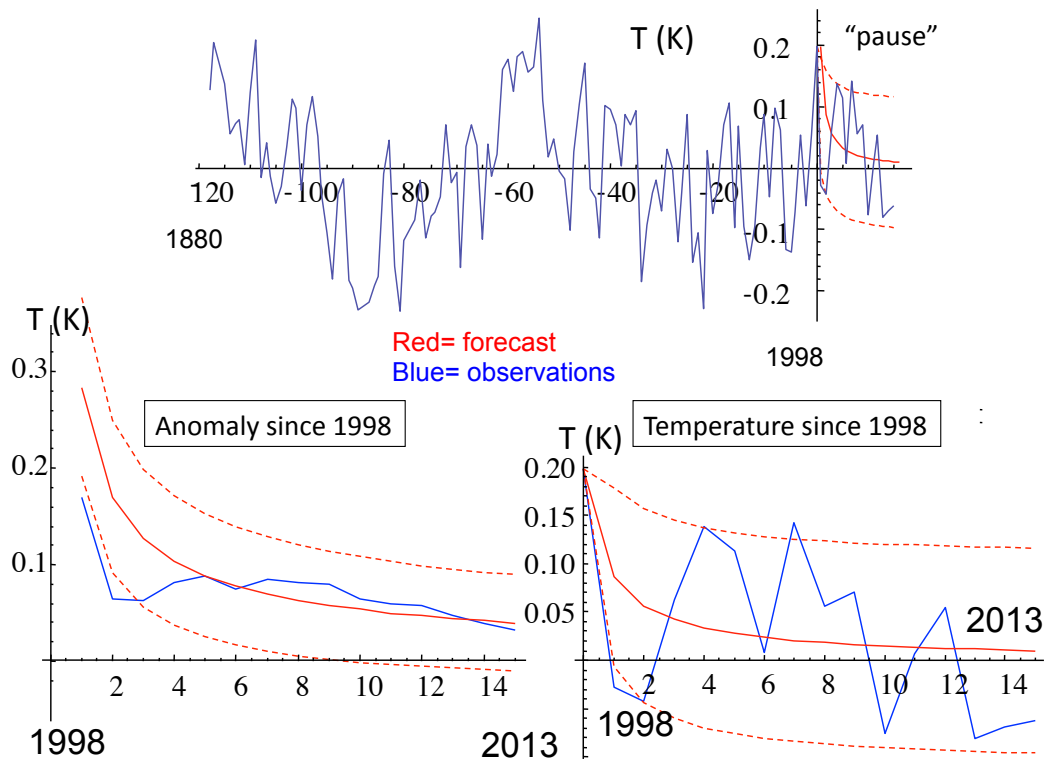
Fig. 12c: **Top**: The residuals temperature of fig. 12b after the low frequency anthropogenic rise has been removed (blue) with the hindcast from 1998 (red). **Bottom left**: The anomaly defined as the average natural temperature (i.e., residual) over the hindcast horizon (blue), red is the hindcast. **Bottom right**: The temperature since 1998 (blue) with hindcast (red), a blowup of the hindcast part of the top right. Adapted from [*Lovejoy*, 2015b].

We saw that the (infinite past) innovation kernel $M_\gamma$ (eq. 42) gave a strong (even singular) weight to the recent past, forecasting AR processes has an analogous strong weighting of the recent data. However, [Gripenberg and Norros, 1996] found something radically new in the case of finite data: the most ancient available data also had a singular weighting! In their words, this was because "the closest witnesses to the unobserved past have special weight", see fig. 12d for a graphical example.
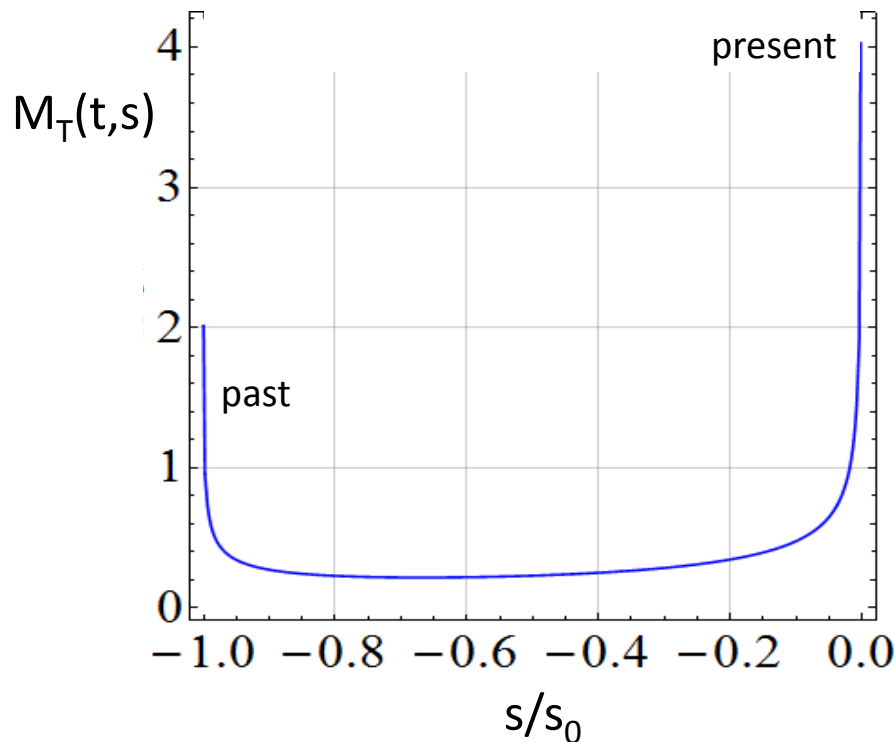
Fig. 12d: This shows the kernel $M_T(t,s)$ (eq. 39, the discrete case) when the data extends to $s_0 = \tau_0$ in the past with parameter $H = -0.1$. Note the strong weighting on both the most recent (right) and the most ancient available data (left). Reproduced from [*Del Rio Amador*, 2017].

## 4. Stochastic predictability limits and forecast skill

### 4.1 Stochastic predictability limits: StocSIPS hindcasting skill demonstrated on CMIP5 control runs

We are used to the deterministic predictability limits that arise from the "butterfly effect" - sensitive dependence on initial conditions - we argued that this limit (the inverse Lyapunov exponent of the largest structures) was roughly given by the lifetime of planetary structures: $\tau_w = \varepsilon^{-1/3}L^{2/3}$ [*Schertzer and Lovejoy*, 2004]. However, we also argued that when taken way beyond this limit, that both the GCMs and the atmosphere should be considered stochastic. More precisely, we argued that fGn provides a good approximation for the temporal variability, and that due to SSTF, that attempting to use spatial correlations for co-predictors may not lead to an improvement when compared to direct predictions that exploit the huge memory of the system. However, SSTF does not necessarily extend from temperatures to other

series such as climate indices. It is possible that use of the latter as co-predictors may yield larger skills.

Since fGn has stochastic predictability limits that determine its skill, eq. 46, these should therefore be relevant in both GCMs and in real macroweather. However, the latter and in externally forced GCM's, as discussed in section 4.2 there are low frequency responses to climate forcings, and these must be forecast separately (using linearity eq. 11) from the internal macroweather variability modelled by fGn processes. This means that the best place to test our predictors is on unforced GCM's i.e. on control runs. For this purpose we used 36 globally and monthly averaged CMIP5 models control. For each, we estimated the relevant exponent $H$ by determining the value that made the predictor best satisfy the orthogonality condition (eq. 41); this was slightly more accurate than using either spectra or Haar fluctuation analysis ([*Del Rio Amador*, 2017]). While each model had somewhat different exponents, we found a mean $H = -0.11 \pm 0.09$ theoretically implying a huge memory (see e.g. fig. 11a,b). We used the discrete $M_T$ kernel (following [*Hirchoren and Arantes*, 1998]) and produced 12 month hindcasts comparing both the theoretical skill and the actual hindcast skill, see figs. 13a. Fig. 13b shows that the control runs were hindcast very nearly to their theoretical limits. It is thus quite plausible that the theoretical stochastic predictability limit eq. 46 really is an upper bound on the skill of macroweather forecasts.
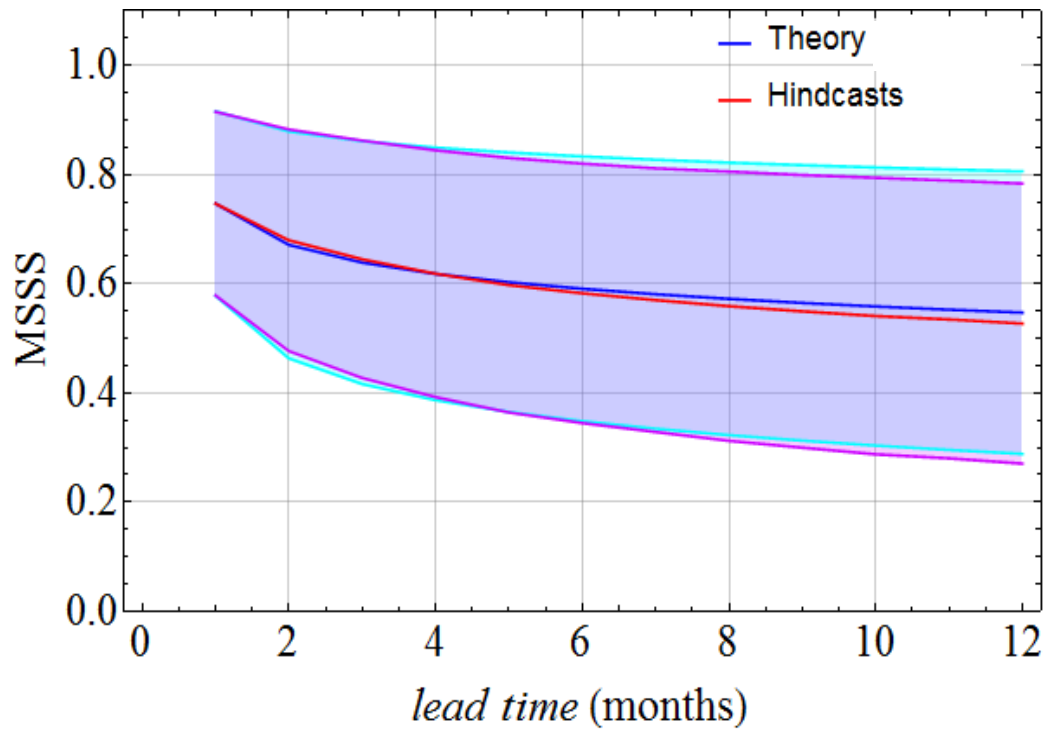
Fig. 13a: The MSSS for hindcasting 36 CMIP5 GCM control runs, each at least 2400 months long. Each GCM had a slightly different H and hence different theoretical predictability. The graph shows that both the means and the spreads of theory and practice (SLIMM hindcasts) agree very well. Reproduced from [*Del Rio Amador*, 2017].
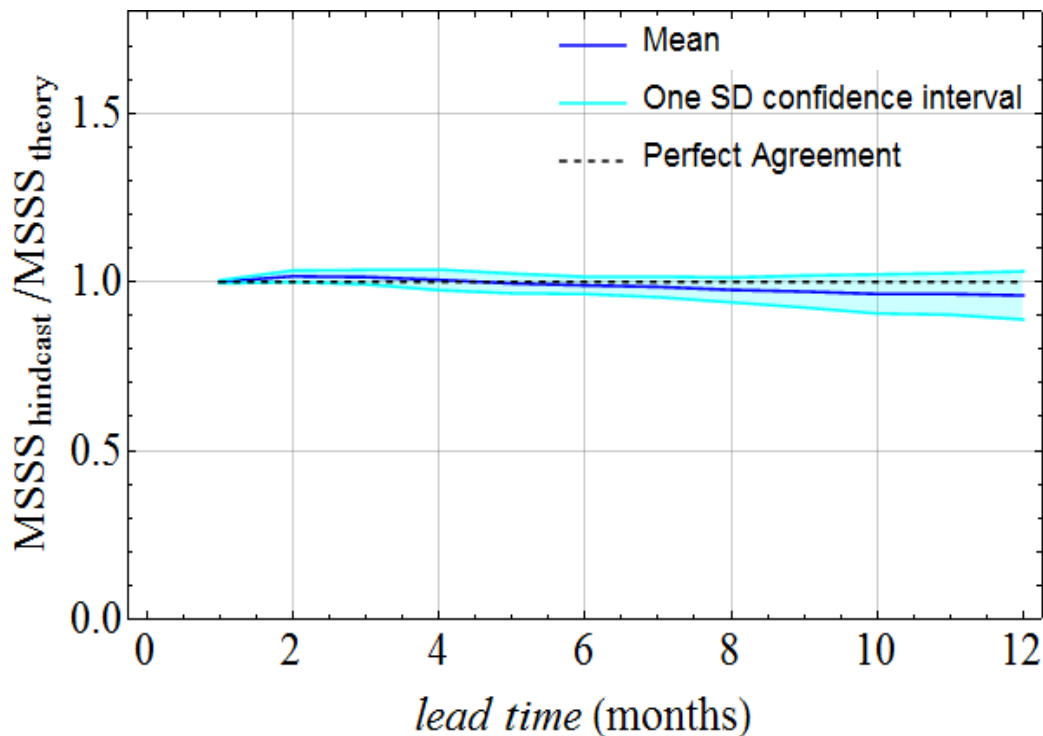
Fig. 13b: The ratio of the actual MSSS hindcast skill to theortical MSSS skill evaluated for the CMIP5 control runs used in fig. 13a.  Reproduced from [*Del Rio Amador*, 2017].

## 4.2 Regional forecasting

In the previous section, we saw that without external forcings, we can make global scale macroweather forecasts that nearly attain their theoretical limits, and in section 3.3 (the pause), we already indicated that by appropriately removing the low frequencies (in that case, the anthropogenic forcings), we could also make accurate global scale real world forecasts.   Due to SSTF, we argued that if at a given location long series were available, they could be forecast directly, that using information at other locations as co-predictors would not increase the overall skill. In this section, we therefore discuss regional forecasts at $5^o$ resolution.   This resolution was chosen because it the smallest that is available from both historical data and reanalysis data sets that we used.

The various steps in the forecast are illustrated in fig. 14 using the pixel over Montreal as an example.  The first step is to remove the low frequencies that are not due to internal macroweather variability; failure to remove them will lead to serious biases since the SLIMM forecast assumes a long term mean equal to 0 and the ensemble forecast is always towards this mean. The low frequencies have both a mean component (mostly anthropogenic in origin but also one due to internal variability), but also a strong annual cycle that slowly evolves from one year to the

next. Using the knowledge (fig. 5d) that the scaling is broken at decadal scales, we can use a high pass filter to separate out these from the internal variability. Similarly, the annual cycle can be forecast by using the past thirty years of data in order to make running estimates of the relevant Fourier coefficients (only keeping those for the annual cycle and 6, 4, and 3 month harmonics). The various steps are shown in fig. 14. Finally the anomalies (lower right) were forecast using SLIMM.
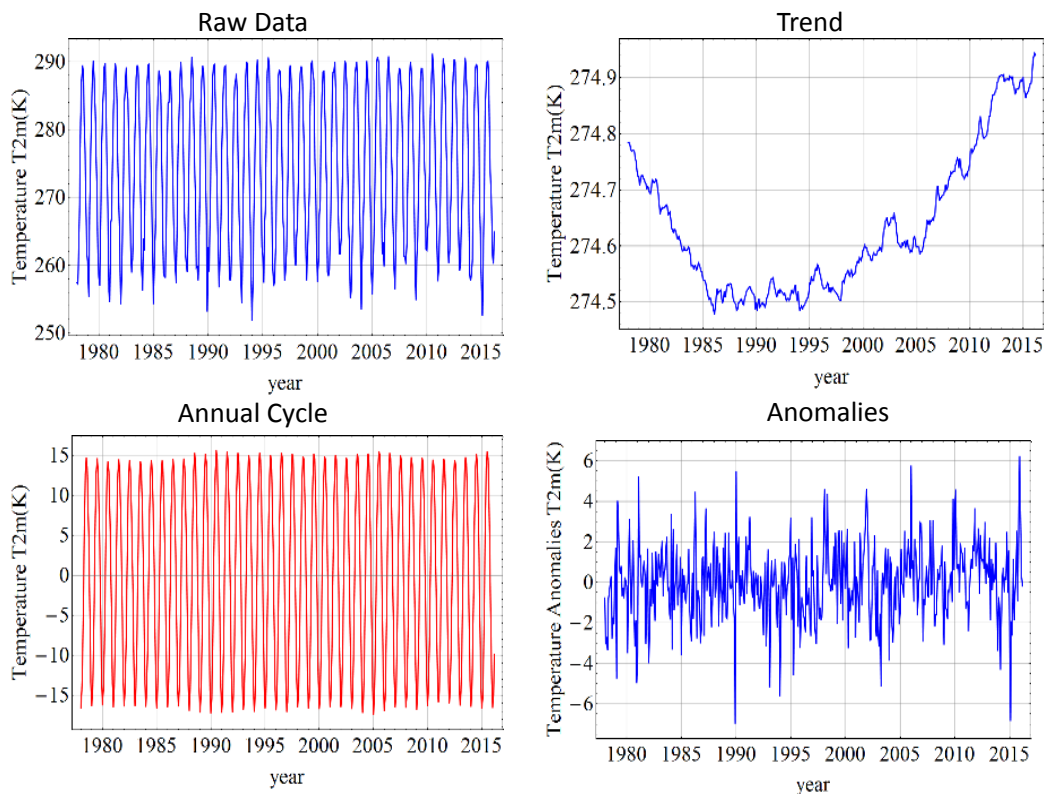


Fig. 14: An example of forecasting the temperature at Montreal using the National Centers for Environmental Prediction (NCEP) reanalysis (at 5º×5º resolution?). The top left shows the raw monthly data, the bottom left shows the mean annual cycle as deduced using a (causal) thirty year running estimate, the upper right shows the low frequency (a causal thirty year running average) trend and the bottom right shows the resulting anomalies that were forecast by SLIMM. Reproduced from [*Del Rio Amador*, 2017].

The regional variation of the skill of the resulting StocSIPS hindcasts is shown in fig. 15, we can see that it is close to the theoretical maximum.
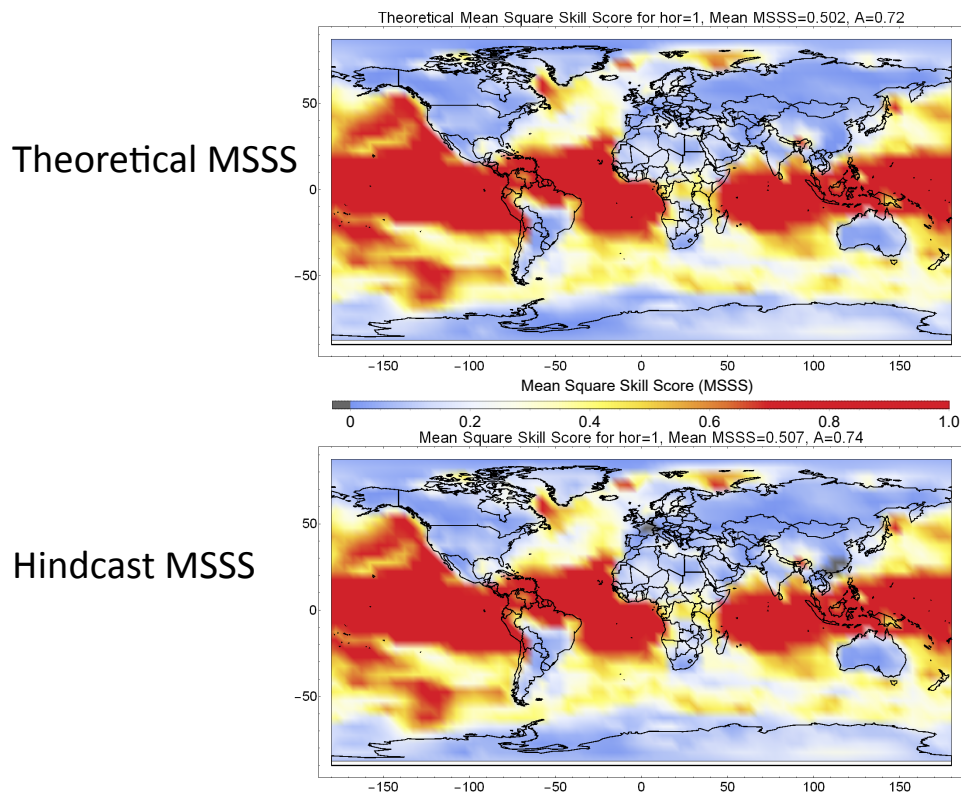
Theoretical MSSS

Hindcast MSSS

Fig. 15: Theoretical (top) versus empirical (bottom) hindcast skill for 1 month hindcasts using Period Sep, 1980 - Dec, 2015. Reference: NCEP Reanalysis. The theory and practice are very close. Reproduced from [*Del Rio Amador*, 2017].

### 4.3 StocSIPS-CanSIPS comparison

The previous section reminded us that real world forecasts must estimate, remove and separately forecast the nonmacroweather low frequencies, the higher frequency internal fGn-like component. The overall model, including this "pre-processing" is called the Stochastic Seasonal to Interannual Prediction System (StocSIPS, see the website: http://www.physics.mcgill.ca/StocSIPS/), it is comparable in scope to the Canadian Seasonal to Interannual Prediction System (CanSIPS, [*Merryfield et al.*, 2011]) and the European Seasonal to Interannual Prediction System (EuroSIPS, http://www.ecmwf.int/en/forecasts/documentation-and-support/long-range/seasonal-forecast-documentation/eurosip-user-guide/multi-model), but of course is based directly on a stochastic rather than a deterministic-stochastic (GCM) model. Indeed, according to the World Meteorological Organization (WMO) site (http://www.wmo.int/pages/prog/wcp/wcasp/gpc/gpc.php), there are twelve international "producing" centres; StocSIPS based at McGill would be the 13th. Although completely unfunded, since April 2016, it has provided operational monthly, seasonal and annual temperature forecasts at 5º resolution.

   As the previous section showed, SLIMM can forecast GCM control runs to nearly their theoretical stochastic predictability limits.  However, we must evaluate the full StocSIPS system (preprocessing plus SLIMM) and compare it with conventional approaches.  We singled out CanSIPS, which since 2010 is the institutional product of Environment Canada, for particularly close comparisons.  Every month, CanSIPS makes monthly through annual temperature and precipitation forecasts; the publically available maps are only over Canada, but we accessed the global products and made global hindcast comparisons since 1980.  The CanSIPS products are based on "multimodel ensemble" consisting of 10 realizations of the CanCM3 and 10 realizations of the CanCM4 GCM.

   Before continuing, recall the method by which GCM's currently produce macroweather forecasts.  The first step is the initialisation; when CanSIPS is initialized it uses reanalyses from the European Centre for Medium-range Weather Forecasts (ECMWF) and these are data-model "hybrids" obtained by assimilating meteorological observations into the ECMWF GCM.   The problem is that both the reanalyses and CanSIPS have their own different climatologies so that the latter cannot directly ingest the ECMWF reanalyses, instead, the ECMWF initial values must be converted into ECMWF anomalies.  These anomalies are used to determine the CanSIPS initial values, the "actuals".  The process can be symbolically written as:

$$T_{CanSIPS}\left(\underline{r},t\right) = \overline{T}_{i(t),CanSIPS}\left(\underline{r}\right) + T'_{CanSIPS}\left(\underline{r},t\right)$$
$$T_{ECMWF}\left(\underline{r},t\right) = \overline{T}_{i(t),ECMWF}\left(\underline{r}\right) + T'_{ECMWF}\left(\underline{r},t\right)$$

(53)

where the overbar represents the climatological temperature $\overline{T}_i\left(\underline{r}\right)$ at position $\underline{r}$, for the month number $i$ = 1, 2, …12 and the primes indicate the anomalies which are functions of both position and time ($i(t)$ denotes the month number of time $t$).   The conventional way to define $\overline{T}_i\left(\underline{r}\right)$ is to use the averages over the previous thirty $i$th months (at each location/pixel $\underline{r}$).    Aside from the annual cycle (that was deliberately ignored in section 2.2), the anomalies differ from the internal variability because they are based on temporal rather than ensemble averages.  A significant consequence is that the anomalies have contributions from both internal and externally forced variability.

   CanSIPS is thus initialized $T_{CanSIPS}\left(\underline{r},0\right)$ using the ECMWF anomaly at time $t$ = 0:

$$T_{CanSIPS}\left(\underline{r},0\right) = \overline{T}_{i(0),CanSIPS}\left(\underline{r}\right) + T'_{ECMWF}\left(\underline{r},0\right)$$

(54)

   The forecasts $\hat{T}_{CanSIPS}\left(\underline{r},t\right)$  (at $t$>0, indicated with circonflex) are then made using the 20 member CanSIPS ensemble followed by complex (and time consuming) post-processing that primarily correct for the "model drift" and poor climate sensitivity.  "Model drift" refers to the tendency of model temperatures (even in control runs) to display low frequency variations that are usually attributed to slow (mostly ocean) processes, artefacts that are not fully "balanced" when the model is initialized.  Since the model does not have perfect representation of the sensitivity to anthropogenic effects, the corresponding systematic errors also contribute a

further low frequency "drift". Both are removed (to some extent) using hindcasts over the previous 5 year period in an attempt to estimate (and remove) spurious linear trends ([*Merryfield et al.*, 2011]). Unfortunately, five years is too short to properly estimate the trend (the true trends are buried in the macroweather noise until a decade or so in scale, see fig. 5d) so that the internal 5 year variability is thus spuriously removed in the post-processing.
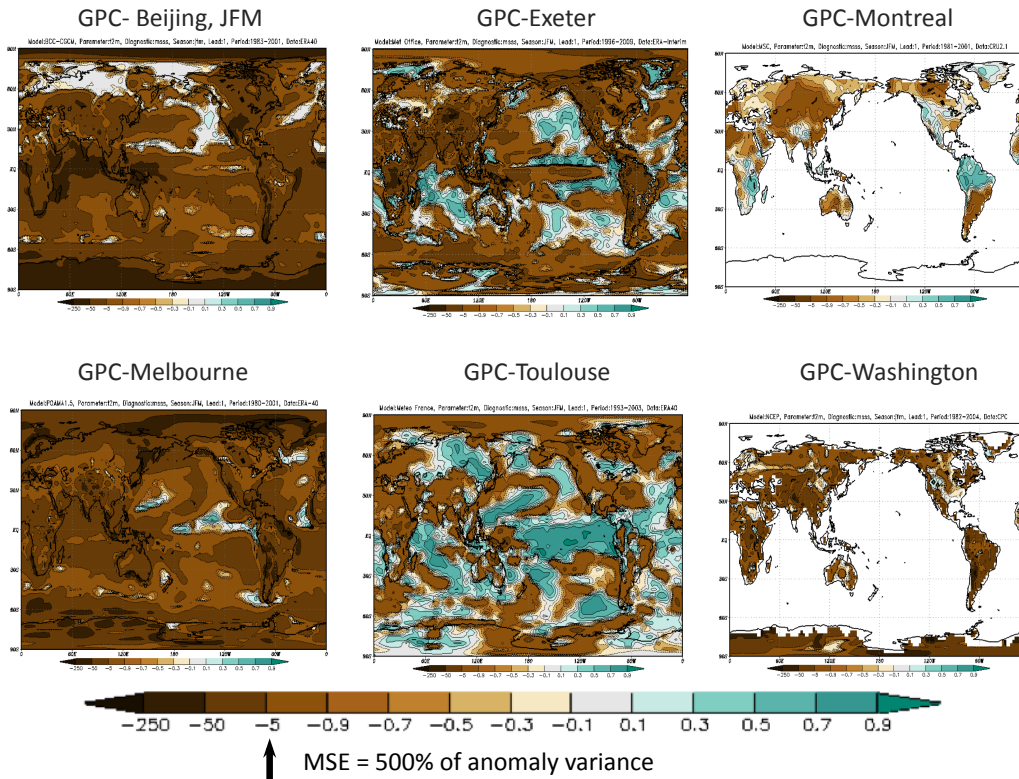


Fig. 15: The MSSS, shown for the actuals and estimated from hindcasts from six of the 12 "producing centres", adapted from the WMO web site (accessed in April 2016). To aid in the interpretation, an example is given by the black arrow: when the MSSS = -5, the Mean Square Error (MSE) is 5 times the amplitude of the anomaly variance. It can be seen that actuals' error variances are typically several times the anomaly variances leading to significant negative skill over most of the earth. Reproduced from [*Del Rio Amador*, 2017].
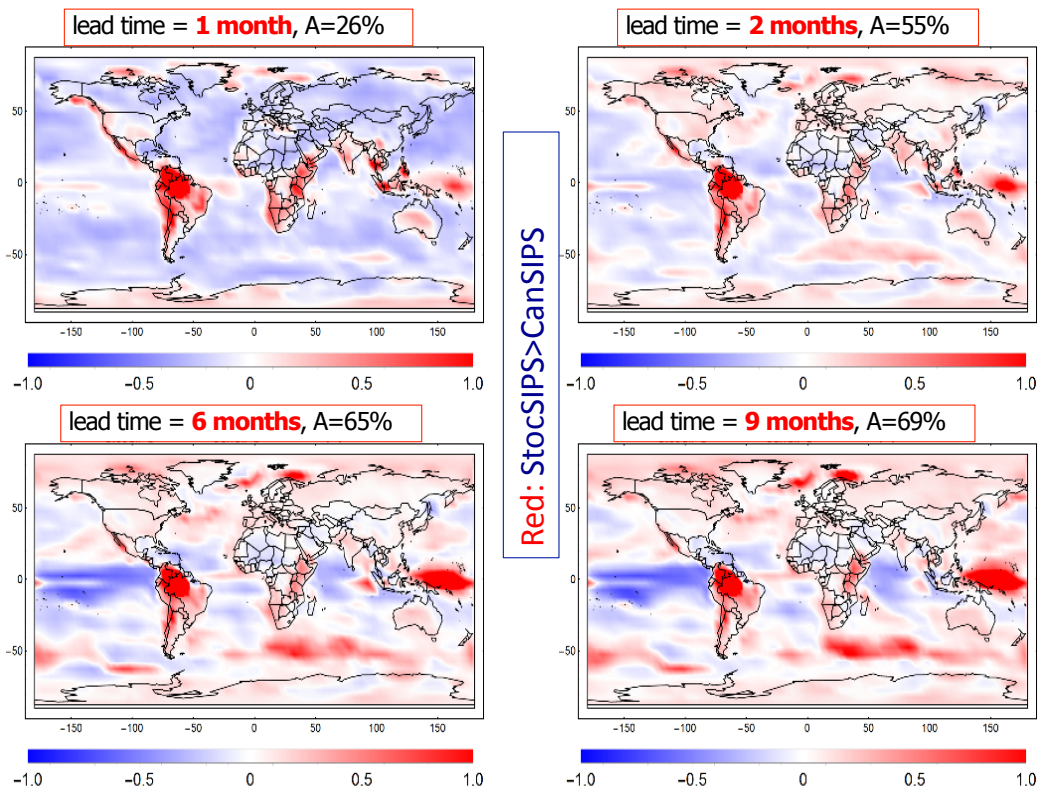
Fig. 16: The differences in MSSS for CanSIPS and StocSIPS anomaly hindcasts over the period 1980-2010 for lead times of 1, 2, 6, 9 months, red indicates regions over which StocSIPS has higher skill. It may be seen that for two months and longer, this is over most of the globe. StocSIPS' increased skill is particularly noticeable over land, probably due to the fact that the CanSIPS ocean model is still within its deterministic predictability limit of 1 -2 years. Reproduced from [*Del Rio Amador*, 2017].
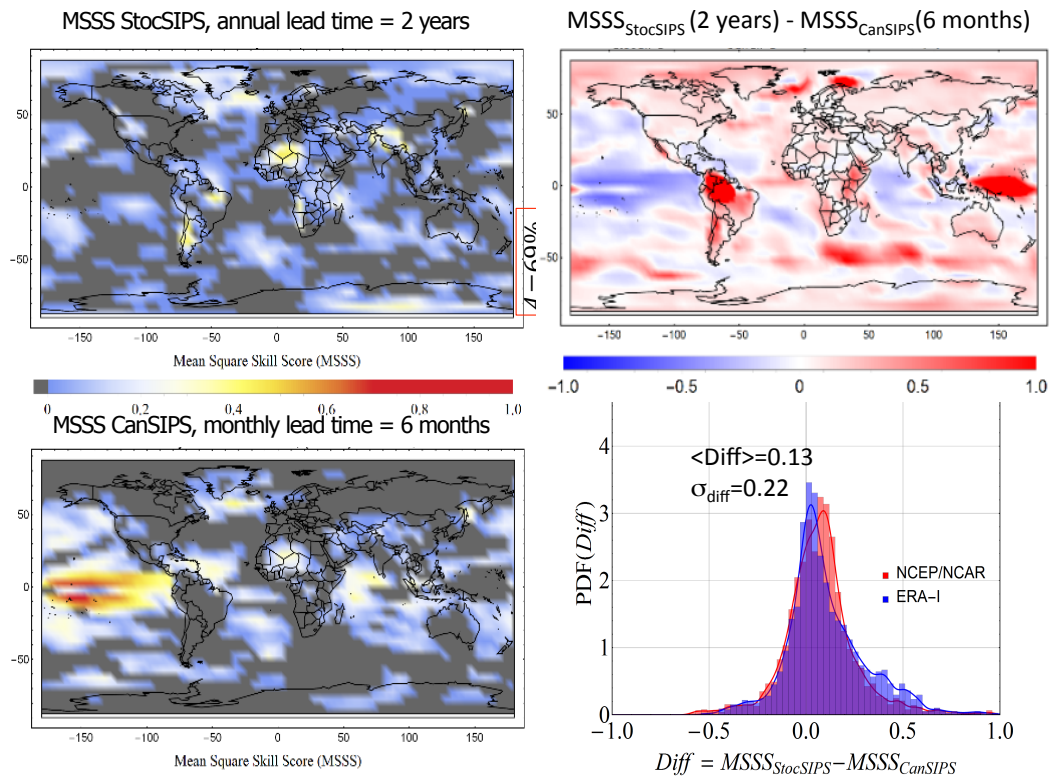
Fig.17: A comparison of the MSSS for StocSIPS at 2 year lead times (top left) and CanSIPS at 6 months (bottom left). The map of their differences (top right) and histogram of the differences lower right using both the ECMWF interim reanalyses (ERA-I, red) and NCEP reanalyses (blue) show that the 2 year StocSIPS forecast is somewhat better than the CanSIPS six month forecast. Reproduced from [*Del Rio Amador*, 2017].
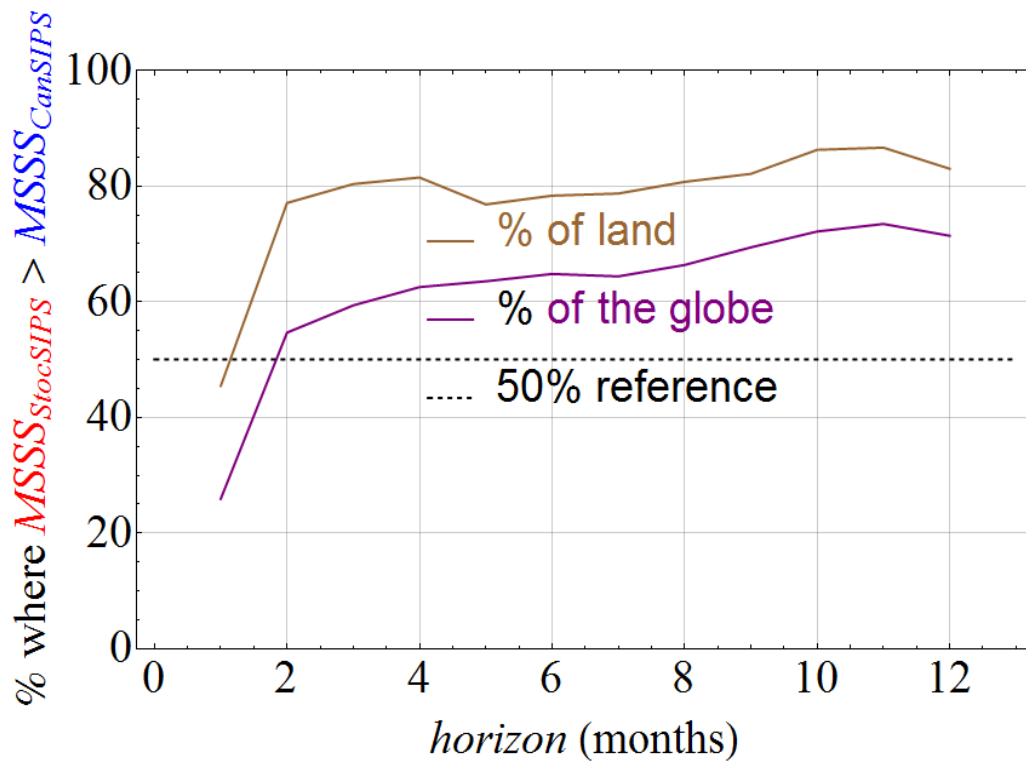
Fig. 18: The relative skill of StocSIPS and CanSIPS anomaly hindcasts (1980-2010) over the globe and over land only showing that StocSIPS' relative advantage increase systematically with lead time and is particularly strong over the oceans. Reproduced from [*Del Rio Amador*, 2017].

In spite of these manipulations, the final result $\hat{T}_{CanSIPS}(\underline{r},t)$ - i.e. an "actual" - is seriously in error as can be seen in fig. 15: which shows that the actuals' error variance is typically several times larger than the anomaly variance. Due to this, the publically available macroweather forecasts are of the anomalies $\widehat{T}'_{CanSIPS}(\underline{r},t) = \hat{T}_{CanSIPS}(\underline{r},t) - \bar{T}_{i(t),CanSIPS}(\underline{r},t)$. For these anomalies, the comparison with StocSIPS is much closer, see fig. 16. The figure shows that even for anomalies over most of the globe, for two months and longer, StocSIPS has higher skill. StocSIPS' increased skill is particularly noticeable over land, probably due to the fact that the CanSIPS ocean model is still within its deterministic predictability limit of 1 - 2 years making its ocean forecast reasonably accurate. This impression in bolstered by fig. 17 which compares CanSIPS at 6 months and StocSIPS at 2 years (the skill are comparable), and also by fig. 18 that shows that StocSIPS' relative advantage grows with lead time and is particularly strong over land.

Although we have not discussed it in this review, StocSIPS actually provides forecasts the probability distributions (both mean, discussed up until now, and the

standard deviation about the mean).  This can be used for various probabilistic forecasts.  For example, fig. 19a,b show a typical seasonal forecast and its validation. In fig. 19a we see that the StocSIPS anomaly forecasts generally follow the data better than CanSIPS.  In fig. 19b, we see that for this location and date, that the StocSIPS forecast was both more accurate and less uncertain than the CanSIPS forecast.  This was true for both the actuals and the anomalies.  This can be seen since not only is the dashed red StocSIPS mean closer to the NCEP validation (dashed black) than the CanSIPS dashed blue but the uncertainties (the spreads in the probability densities) was narrower for the StochSIPS forecast.   Other probabilistic forecasts that can readily be produced by StocSIPS include tercile forecasts: i.e. the probabilities of the forecast temperature being below, above or equal to the local climatology; see the StocSIPS site for examples.
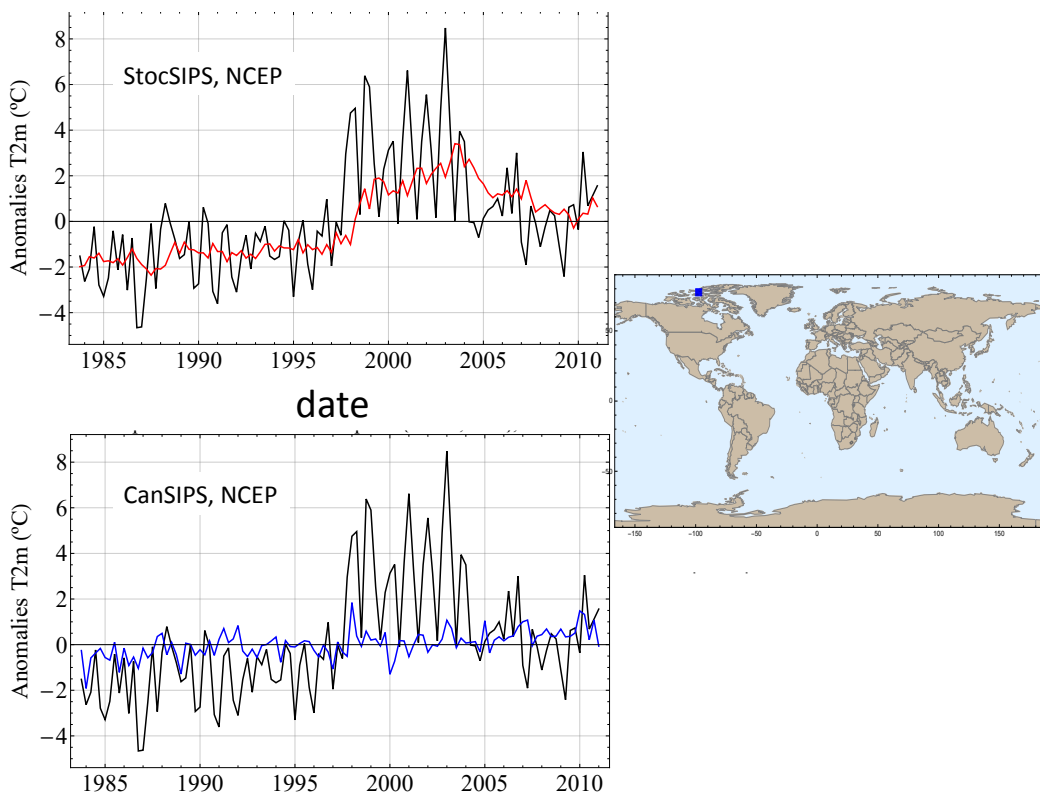


Fig. 19a:  StocSIPS (top, red) and CanSIPS (bottom, blue) seasonal forecasts, 2 seasons ahead for temperature anomalies at 97.5W, 77.5N (see blue point on the map at right).    The forecasts are compared with the NCEP reanalysis anomalies (black) that are calculated with respect to the period 1980-2010.  It can be seen that StocSIPS is much closer to the data (see also fig. 19b).
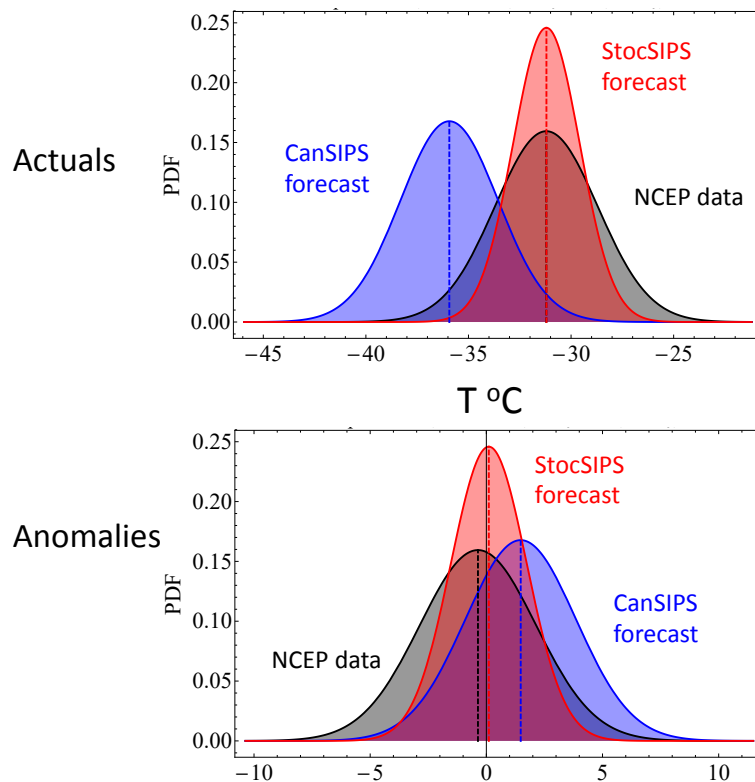
Fig. 19b:  The histograms of seasonal forecasts, two seasons ahead for DJF (2009-2010) using data up to $t = 0 =$ JJA 2009, location the same as in fig. 19a (top actuals, bottom, anomalies, StocSIPS in red, CanSIPS in blue, NCEP data in black).   The dashed black lines are the NCEP validation data for DJF, the black probability density curves show the spread of the climatological variations based on past NCEP reanalyses (1981-2010), the variability is thus placed around the observed DJF temperature.   The StocSIPS and CanSIPS dashed lines (red and blue) are their respective forecasts for DJF, the curves represent the estimated uncertainties in the forecast.    For both actuals and anomalies StocSIPSs forecasts are sharper - their probability density functions, (PDFs) are narrower and more peaked; they are also more accurate since the red dashed lines (the StocSIPS forecasts) are closer to validation data (the black dashed lines).

## 5. Conclusions:

Over the last decades, it has become increasingly clear that at weather scales, atmospheric dynamics are governed by both deterministic laws of continuum mechanics and by stochastic turbulence laws.  Although the GCM equations do not acknowledge the existence of atoms or molecules, they are nevertheless compatible with statistical mechanics.  Similarly they are also believed to be compatible with the turbulence laws obeying them quite accurately.   Over the same period, the GCM approach has - with the development of ensemble forecasting and stochastic

parametrisations - itself evolved into a stochastic one, making it tempting to make weather forecasts directly using the turbulence laws. However in the weather regime the latter are highly intermittent, involving vector multifractal processes, and the corresponding mathematical prediction problem has yet to be solved: the GCM approach to weather prediction is thus the only one currently available.

The situation is radically different at time scales beyond the GCM deterministic predictability limit - in macroweather. On the one hand GCM's have large errors associated with unrealistic model climatologies, especially poor representations of the annual cycle, and they also display model drift and unrealistic sensitivities to anthropogenic effects. On the other hand, macroweather "turbulence" (the extension of turbulence models to the macroweather regime), have low intermittency so that Gaussian models are useable approximations. In addition, a new symmetry: statistical space-time factorization essentially decouples space and time so that mean square predictions can conveniently be made for each spatial location independently. Physically this means that even though strong spatial correlations exist (including "teleconnections"), if one has a long enough history at a given point, this spatial information is also in the series and using data at other spatial locations as co-predictors does not necessarily improve the forecast. The factorisation is not exact and does not necessarily apply to other series such as climate indices so that there may be future scope for finding co-predictors and improve the skill.

The ideal testing ground for this approach is in GCM control runs since this is closest to pure fGn. We found that the ScaLIng Macroweather Model (SLIMM) based on a fractional Gaussian noise (fGn) model applied to temperatures from GCM control runs (i.e. pure macroweather processes, no changes in external forcings) is nearly able to attain the maximum theoretical stochastic predictability limit, verifying that GCM's well obey the macroweather laws upon which SLIMM is based and raising the possibility that these stochastic predictability limits are true GCM limits. With respect to usual stochastic forecasts based on exponential correlations (AutoRegressive, Linear Inverse Modelling), the radically new feature of SLIMM is its exploitation of the huge long range memory. The SLIMM prediction kernel thus has singular weighting to both the most recent data as well as the most ancient data since the latter contain the maximum information of the distant past.

Applying SLIMM to real data requires preprocessing to remove non-macroweather processes in particular low frequency anthropogenic effects as well as – for regional forecasts - the annual cycle. The overall resulting system (preprocessing plus SLIMM) is the STOChastic Seasonal to Interannual Prediction System (StocSIPS). We compared StocSIPS with one of the leading GCM macroweather products: CanSIPS. Even without any co-predictors or other use of spatial correlations, we showed that StocSIPS was much superior to CanSIPS for forecasting "actuals": this was due to StocSIPS ability to essentially forecast the climatology (especially the annual cycle). However, even for anomaly forecasts, StoSIPS was superior to CanSIPS for lead times of two month or longer and its relative advantage grew with lead time, it was particularly important over land where for two months and longer it was superior over ≈80% or more of the earth's land surface.

Aside from its increased skill, StocSIPS has other advantages. For example, at the moment, seasonal forecasts for the city of Montreal (or other localized region) are highly indirect. First data from all atmospheric fields from all over the world must be assimilated. Then the model - on grids typically several hundred kilometers across -  is integrated forward in time. Anomalies are calculated, and post-processing is performed to make low frequency corrections for known biases. Finally, the Montreal temperature anomaly is estimated by "downscaling" from the large pixel scale to the local city scale. This can be done either using sophisticated (but complex) nested regional models (of GCM type), or via ad hoc statistical methods based on local climatology. In contrast, if long enough (preferably several decades) of monthly or seasonal data are available, then StocSIPS simply removes the low frequencies (including the annual cycle), separately forecasts the anomalies and low frequencies, and adds them to produce the forecast. The overall saving in computational speed is estimated to be of the order of $10^7$ (about $10^5$ to $10^6$ for global forecasts on $5^o X 5^o$ grids). Finally, StocSIPS directly forecasts the conditional ensemble average, i.e. effectively the results of an infinite ensemble whereas CanSIPS uses only 20 members.

StocSIPS can be directly extended to other fields such as wind or precipitation which instead are known to have macroweather statistics roughly satisfying the SLIMM requirements (low intermittency temporal macroweather scaling with -1/2<$H$<0 and space time statistical factorization (SSTF), [*Lovejoy and de Lima*, 2015] and fig. 9a). But StocSIPS main advantage may be its ability to directly forecast other fields, such as insolation, wind power or degree-days, that can currently only be very indirectly forecast by GCM's. Other future extensions of StocSIPS could include drought indices and the prediction of extremes.

## 6. Acknowledgements

## 7. References

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.

Biagini, F., Y. Hu, B. Øksendal, and T. Zhang (2008), *Stochastic Calculus for Fractional Brownian Motion and Applications*, Springer-Verlag, London.

Chen, W., S. Lovejoy, and J. P. Muller (2016), Mars' atmosphere: the sister planet, our statistical twin, *J. Geophys. Res. Atmos.*, *121* doi: 10.1002/2016JD025211.

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.

Del Rio Amador, L. (2017), The Stochastic Seasonal to Interannual Prediction System, McGill University, Montreal.

Garcıa-Serrano, J., and F. J. Doblas-Reyes (2012), On the assessment of near-surface global temperature and North Atlantic multi-decadal variability in the ENSEMBLES decadal hindcast, *Clim. Dyn.* , *39*, 2025–2040 doi: DOI 10.1007/s00382-012-1413-1.

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

Guemas, V., F. J. Doblas-Reyes, I. Andreu-Burillo, and M. Asif (2013), Retrospective prediction of the global warming slowdown in the past decade *Nature Cli. Change*, *3*, 649-653.

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

Hebert, R., S. Lovejoy, and A. de Vernal (2017), A Scaling Model for the Forced Climate Variability in the Anthropocene, *(in preparation for Climate Dynamics)*.

Hébert, R., and S. Lovejoy (2015), The runaway Green's function effect: Interactive comment on "Global warming projections derived from an observation-based minimal model" by K. Rypdal, *Earth System Dyn. Disc.* , *6*, C944–C953.

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

Hirchoren, G. A., and C. E. D'attellis (1998), Estimation Of Fractal Signals, Using Wavelets And Filter Banks, *IEEE Transactions On Signal Processing*, *46*(6), 1624-1630.

Kolesnikov, V. N., and A. S. Monin (1965), Spectra of meteorological field fluctuations, *Izvestiya, Atmospheric and Oceanic Physics*, *1*, 653-669.

Lean, J. L., and D. H. Rind (2008), How natural and anthropogenic influences alter global and regional surface temperatures: 1889 to 2006, *Geophys. Resear. Lett.* , *35*, L18701 doi: 10.1029/2008GL034864.

Lilley, M., S. Lovejoy, D. Schertzer, K. B. Strawbridge, and A. Radkevitch (2008), Scaling turbulent atmospheric stratification, Part II: empirical study of the the stratification of the intermittency, *Quart. J. Roy. Meteor. Soc.*, DOI: 10.1002/qj.1202.

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

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

Lovejoy, S. (2015a), A voyage through scales, a missing quadrillion and why the climate is not what you expect, *Climate Dyn.*, *44*, 3187-3210 doi: doi: 10.1007/s00382-014-2324-0.

Lovejoy, S. (2015b), Using scaling for macroweather forecasting including the pause, *Geophys. Res. Lett.*, *42*, 7148–7155 doi: DOI: 10.1002/2015GL065665.

Lovejoy, S. (2017), How accurately do we know the temperature of the surface of the earth? , *Clim. Dyn., (in press)*.

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 (2007), Scale, scaling and multifractals in geophysics: twenty years on, in *Nonlinear dynamics in geophysics*, edited by J. E. A.A. Tsonis, Elsevier.

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

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

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: 10.1063/1.4927223.

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

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 Macroweather Model (SLIMM): using scaling to forecast global-scale macroweather from months to Decades, *Earth Syst. Dynam.*, *6*, 1–22 doi: http://www.earth-syst-dynam.net/6/1/2015/,  doi:10.5194/esd-6-1-2015.

Lovejoy, S., A. F. Tuck, S. J. Hovde, and D. Schertzer (2007), Is isotropic turbulence relevant in the atmosphere?, *Geophys. Res. Lett.*, *doi:10.1029/2007GL029359*, L14802.

Lovejoy, S., D. Schertzer, M. Lilley, K. B. Strawbridge, and A. Radkevitch (2008), Scaling turbulent atmospheric stratification, Part I: turbulence and waves, *Quart. J. Roy. Meteor. Soc., 134*, 277-300 doi: DOI: 10.1002/qj.201.

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

Merryfield, W. J., B. Denis, J.-S. Fontecilla, W.-S. Lee, S. Kharin, J. Hodgson, and B. Archambault (2011), The Canadian Seasonal to Interannual Prediction System  (CanSIPS)  An overview of its design and operational implementation*Rep.*, 51pp pp, Environment Canada.

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.

Panofsky, H. A., and I. Van der Hoven (1955), Spectra and cross-spectra of velocity components in the mesometeorlogical range, *Quarterly J. of the Royal Meteorol. Soc.*, *81*, 603-606.

Papoulis, A. (1965), *Probability, Random Variables and Stochastic Processes*, Mc Graw Hill, New York.

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., and S. Lovejoy (2014), Atmospheric waves as scaling, turbulent phenomena, *Atmos. Chem. Phys.* , *14*, 3195–3210 doi: doi:10.5194/acp-14-3195-2014.

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.

Radkevitch, A., S. Lovejoy, K. B. Strawbridge, D. Schertzer, and M. Lilley (2008), Scaling turbulent atmospheric stratification, Part III: emplrical study of Space-time stratification of passive scalars using lidar data, *Quart. J. Roy. Meteor. Soc.*, DOI: 10.1002/qj.1203.

Ragone, F., V. Lucarini, and F. Lunkeit (2015), A new framework for climate sensitivity and prediction: a modelling perspective, *Climate Dyn.*, 1-13 doi: DOI 10.1007/s00382-015-2657-3.

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

Rypdal, K. (2015), Global warming projections derived from an observation-based minimal model, *Earth Syst. Dynam. Discuss.*, *6*, 1789-1813 doi: 10.5194/esdd-6-1789-2015.

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.

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 (1995), From scalar cascades to Lie cascades: joint multifractal analysis of rain and cloud processes, in *Space/time Variability and Interdependance for various hydrological processes*, edited by R. A. Feddes, pp. 153-173, Cambridge University Press, New-York.

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.

Schertzer, D., I. Tchiguirinskaia, S. Lovejoy, and A. F. Tuck (2012), Quasi-geostrophic turbulence and generalized scale invariance, a theoretical reply, *Atmos. Chem. Phys.*, *12*, 327-336 doi: 10.5194/acp-12-327-2012.

Schertzer, D., S. Lovejoy, F. Schmitt, Y. Chigirinskaya, and D. Marsan (1997), Multifractal cascade dynamics and turbulent intermittency, *Fractals*, *5*, 427-471.

Schmidt, G. A., D. T. Shindell, and K. Tsigaridis (2014), Reconciling warming trends, *Nature GeoSci.*, *7*, 158-160.

Schwartz, S. E. ( 2012), Determination of Earth's transient and equilibrium climate sensitivities from observations over the twentieth century: strong dependence on assumed forcing., *Surv. Geophys.*, *33*, 745-777.

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.

Suckling, E. B., E. Hawkins, G. Jan van Oldenborgh, and J. M. Eden (2016), An empirical model for probabilistic decadal prediction: A global analysis, *Climate Dynamics (submitted)*.

Tennekes, H. (1975), Eulerian and Lagrangian time microscales in isotropic turbulence. , *J. Fluid Mech*, *67* 561-567.

Vallis, G. (2010), Mechanisms of climate variaiblity from years to decades, in *Stochastic Physics and Climate Modelliing*, edited by P. W. T. Palmer, pp. 1-34, Cambridge University Press, Cambridge.

Van der Hoven, I. (1957), Power spectrum of horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour, *Journal of Meteorology*, *14*, 160-164.

Zeng, X., and K. Geil ( 2017), Global warming projection in the 21st Century based on an observational data driven model, *Geophys. Res. Lett. , (in press)*.