Earth Syst. Dynam. Discuss., 6, 489–545, 2015
www.earth-syst-dynam-discuss.net/6/489/2015/
doi:10.5194/esdd-6-489-2015

Earth System Dynamics Discussions

This discussion paper is/has been under review for the journal Earth System Dynamics (ESD). Please refer to the corresponding final paper in ESD if available.

The Scaling LInear Macroweather model (SLIM): using scaling to forecast global scale macroweather from months to decades

S. Lovejoy, L. del Rio Amador, and R. Hébert

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

Received: 13 February 2015 – Accepted: 27 February 2015 – Published: 17 March 2015

Correspondence to: S. Lovejoy (lovejoy@physics.mcgill.ca)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

At scales of ≈ 10 days (the lifetime of planetary scale structures), there is a drastic transition from high frequency weather to low frequency macroweather. This scale is close to the predictability limits of deterministic atmospheric models; so that in GCM macroweather forecasts, the weather is a high frequency noise. But neither the GCM noise nor the GCM climate is fully realistic. In this paper we show how simple stochastic models can be developped that use empirical data to force the statistics and climate to be realistic so that even a two parameter model can outperform GCM's for annual global temperature forecasts.

The key is to exploit the scaling of the dynamics and the enormous stochastic memories that it implies. Since macroweather intermittency is low, we propose using the simplest model based on fractional Gaussian noise (fGn): the Scaling LInear Macroweather model (SLIM). SLIM is based on a stochastic ordinary differential equations, differing from usual linear stochastic models (such as the Linear Inverse Modelling, LIM) in that it is of fractional rather than integer order. Whereas LIM implicitly assumes there is no low frequency memory, SLIM has a huge memory that can be exploited. Although the basic mathematical forecast problem for fGn has been solved, we approach the problem in an original manner notably using the method of innovations to obtain simpler results on forecast skill and on the size of the effective system memory.

A key to successful forecasts of natural macroweather variability is to first remove the low frequency anthropogenic component. A previous attempt to use fGn for forecasts had poor results because this was not done. We validate our theory using hindcasts of global and Northern Hemisphere temperatures at monthly and annual resolutions. Several nondimensional measures of forecast skill – with no adjustable parameters – show excellent agreement with hindcasts and these show some skill even at decadal scales. We also compare our forecast errors with those of several GCM experiments (with and without initialization), and with other stochastic forecasts showing that even this simplest two parameter SLIM model is somewhat superior. In future, using a space–time

(regionalized) generalization of SLIM we expect to be able to exploiting the system memory more extensively and obtain even more realistic forecasts.

1 Introduction

Due to their sensitive dependence on initial conditions, the classical deterministic prediction limit of GCM's is about ten days – the lifetime of planetary sized structures (τ_w). Beyond this, the forecast weather rapidly loses any relationship with the real weather. The analogous scale (τ_ow) for near surface ocean gyres is about 1 year (Lovejoy and Schertzer, 2012b), so that even the ocean component – important in fully coupled climate models (referred to simply as GCM's below) – is poorly forecast beyond this. When using long GCM runs for making climate forecasts, we are therefore really considering a boundary value problem rather than an initial value problem (Bryson, 1997).

For these longer scales, following (Hasselmann, 1976), the high frequency weather can be considered as a noise driving an effectively stochastic low frequency system; the separation of scales needed to justify such modelling is provided by the drastic transitions at τ_w, τ_ow. In the atmosphere, the basic phenomenology behind this has been known since the earliest atmospheric spectra (Panofsky and Van der Hoven, 1955) and was variously theorized as the "scale of migratory pressure systems of synoptic weather map scale" (Van der Hoven, 1957) and later as the "synoptic maximum" (Kolesnikov and Monin, 1965). Later, it was argued to be a transition scale of the order of the lifetime of planetary structures that separated different high frequency and low frequency scaling regimes (Lovejoy and Schertzer, 1986). More recently, based on the solar-induced energy rate density, the atmospheric scale τ_w was deduced theoretically from turbulence theory (Lovejoy and Schertzer, 2010), and τ_ow was derived in Lovejoy and Schertzer (2013, Chapt. 8). The same basic picture was also confirmed in the Martian atmosphere in Lovejoy et al. (2014). Although it is only plausible at midlatitudes the competing theory from dynamical meteorology postulates that the transition scale τ_w is

the typical scale of baroclinic instabilities (Vallis, 2010); see the critique in Lovejoy and Schertzer (2013, Chapt. 8).

Independent of its origin, the transition justifies the idea that the weather is essentially a high frequency noise driving a lower frequency climate system and the idea is exploited in GCM's with long integrations as well as in Hasselman-type stochastic modelling, now often referred to as "Linear Inverse Modelling" (LIM), e.g. Penland and Sardeshmuhk (1995), Newman et al. (2003), Sardeshmukh and Sura (2009); analogous modelling is also possible at much longer time scales using energy balance models, for a review; see Dijkstra (2013). In these phenomenological models the system is regarded as a multivariate Ohrenstein–Uhlenbeck (OU) process. The basic LIM paradigm (sometimes also called the Stochastic Linear Forcing paradigm) is based on the stochastic differential equation:

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} + \omega_\mathrm{w}\right) T(t) = \sigma_\gamma \gamma(t) \tag{1}$$

where $\omega_\mathrm{w} = \tau_\mathrm{w}^{-1}$ is the "weather frequency", σ_γ is the amplitude of the forcing and $\gamma(t)$ is "δ correlated" Gaussian white noise forcing with:

$$\langle \gamma(t)\gamma(s)\rangle = \delta(t-s); \quad \langle \gamma(t)\rangle = 0 \tag{2}$$

"$\langle . \rangle$ " indicates ensemble averaging and $\delta(t-s)$ is the Dirac function. This uses the convenient physics notation; alternatively, $\gamma(t)\mathrm{d}t = \mathrm{d}W$ where W is a Wiener process. Fourier transforming Eq. (1) and using the usual rule F.T.$\left[\frac{\mathrm{d}^{H+1/2}f}{\mathrm{d}t^{H+1/2}}\right] = (i\omega)^{H+1/2}$F.T.$[f]$ where "F.T." indicates "Fourier Transform", the temperature spectrum is thus:

$$E_T(\omega) = \langle |\widetilde{T}(\omega)|^2 \rangle \approx \frac{\sigma_\gamma^2}{\omega^2 + \omega_\mathrm{w}^2} \tag{3}$$

where the tilde indicates Fourier transform, and (low and high frequencies) respectively $E_T(\omega) \approx \omega^{-\beta}$ with $\beta_l = 0$, $\beta_h = 2$. A spatial LIM model (for regional forecasting) is obtained by considering a vector each of whose components is the temperature (or other atmospheric field) at different (spatially distributed) "pixels", yielding a system of linear stochastic ordinary differential equations of integer order. A system with 20 degrees of freedom (involving > 100 empirical parameters) currently somewhat outperforms GCM's for global scale annual temperature forecasts (Newman, 2013, Table 2, Fig. 2).

The basic problem with the LIM approach, is that although we are interested in the low frequency behaviour, for LIM models it is simply white noise (put $\mathrm{d}/\mathrm{d}t = 0$ in Eq. 1) and this has no memory; by hypothesis LIM models therefore assume a priori there is no long term predictability. However, ever since Lovejoy and Schertzer (1986), there has been a growing literature (Koscielny-Bunde et al., 1998; Huybers and Curry, 2006; Blender et al., 2006; Franzke, 2012; Rypdal et al., 2013; Yuan et al., 2014) and see the extensive review in Lovejoy and Schertzer (2013) showing that the temperature (and other atmospheric fields) are scaling at low frequencies, with spectra significantly different than those of OU processes, notably with β_l in the range 0.2–0.8 with the corresponding low frequency weather regime (at scales longer than $\tau_{\mathrm{w}} \approx 10$ days) now being referred to as "macroweather" (Lovejoy, 2013). At a theoretical level, for regional forecasting, a further shortcoming of the LIM approach is that it does not respect the approximate property of space–time statistical factorization (Lovejoy and Schertzer, 2013, Chapt. 10).

While the difference in the value of β_l might not seem significant, the LIM white noise value $\beta_l = 0$, has no low frequency predictability whereas the actual values $0.2 < \beta_l < 0.8$ (depending mostly on the land or ocean location) correspond to potentially enormous predictability (see e.g. Fig. 1a–e below). Although this basic feature of "long range statistical dependency" has been regularly pointed out in the scaling literature and an attempt was already made to exploit it (Baillie and Chung, 2002; see below), the actual extent of this enhanced predictability has not been quantified before now (see however Yuan et al., 2014), it justifies the development of the new "Scaling LInear

Macroweather model" (SLIM) that we present below. We argue that even in its simplest two parameter version, that it already outperforms GCM's.

2 Linear stochastic models and fractional Gaussian noise

2.1 Linear and nonlinear stochastic atmospheric models

We have discussed the phenomenological linear stochastic models introduced in atmospheric science by Hasselman and others from 1976 onwards. Yet there is an older tradition of stochastic atmospheric modelling that can be traced back to the 1960's: stochastic cascade models for turbulent intermittency (Novikov and Stewart, 1964; Yaglom, 1966; Mandelbrot, 1974; Schertzer and Lovejoy, 1987). Significantly, these models are nonlinear rather than linear and the nonlinearity plays a fundamental role in their ability to realistically model intermittency. By the early 1980's it was realized that these multiplicative cascades were the generic multifractal processes and they were expected to be generally relevant in high dimensional nonlinear dynamical systems that were scale invariant over some range. By 2010, there was a considerable body of work showing that atmospheric cascades were anisotropic – notably with different scaling in the horizontal and vertical directions (leading to anisotropic, stratified cascades), and that this enabled cascades to operate up to planetary sizes (see the reviews Lovejoy and Schertzer, 2010, 2013). While the driving turbulent fluxes were modelled by pure cascades, the observables (temperature, wind etc.) were modelled by fractional integrals of the latter (see below): the Fractionally Integrated Flux (FIF) model. Analysis of in situ (aircraft, dropsonde), remotely sensed data, reanalyses as well as weather forecasting models showed that at least up to 5000 km, the cascade processes were remarkably accurate with statistics (up to second order) typically showing deviations of less than $\approx \pm 0.5\,\%$ with respect to the theoretical predictions (see Lovejoy and Schertzer, 2013, Chapt. 4 for an empirical review).

The success of the cascade model up to planetary scales (L_e) showed that the horizontal dynamics were dominated by the solar induced energy flux ($\varepsilon \approx 10^{-3}\,\mathrm{W\,Kg^{-1}}$ sometimes called the "energy rate density") and it implies a break in the space–time cascades at about $\tau_\mathrm{w} \approx \varepsilon^{-1/3} L_e^{2/3} \approx 10$ days discussed above. The logical next question was therefore: what happens if the model is extended in time and the cascade starts at a outer time scale much longer than τ_w? In Lovejoy and Schertzer (2013, Appendix 10A), the mathematical details of this Extended Fractionally Integrated Flux (EFIF) model were worked out, and it was shown that at frequencies below τ_w^{-1} there would a nonintermittent (near) Gaussian, (near) scaling regime with generic exponents β_l in the observed range.

Although this (temporally) extended space–time cascade based model well predicted the basic space–time weather statistics (for scales $< \tau_\mathrm{w}$) and the temporal macroweather statistics (for scales $> \tau_\mathrm{w}$), by itself, the model was not able to reproduce the *spatial* macroweather statistics that characterize climate zones, so that another even lower frequency climate process was necessary. It was proposed that – following the basic mathematical structure of the rest of the model – that this was also multiplicative in nature. This factorization hypothesis was empirically verified on macroweather temperature and precipitation data (Lovejoy and Schertzer, 2013, Chapt. 10; Lovejoy and de Lima, 2015, respectively).

We are therefore now in a position to compare and contrast LIM and the extended (nonlinear, stochastic) cascade models (EFIF) to propose Scaling LInear Macroweather model (SLIM). SLIM is therefore a linear macroweather approximation to the more general EFIF model which in the macroweather regime is only weakly nonlinear. At the scalar level – here global scale temperature time series – LIM and SLIM are very close because they are both Gaussian and can be presented as solutions to ordinary differential equations (although the latter are of fractional, not integer order). While empirically it is found that macroweather temperature probability distributions have "fat tails" – so that statistical moments of order ≈ 5 diverge (Lovejoy and Schertzer, 2013, Chapt. 5; Lovejoy, 2014a; Lovejoy and de Lima, 2015), see also Lovejoy and Schertzer

(1986) – for the (low order) statistics (e.g. near the mean and variance – first and second order), the deviations from Gaussianity are small, the intermittency exponent for the mean is low: $C_1 \approx 0.01$–0.02 ($C_1 = 0$ for Gaussian processes). The key difference between the LIM and SLIM scalar models is therefore the change in the low frequency scaling: trivial vs. long range memory. As described in a future publication, another significant difference arises when the models are extended to space–time: (regional) SLIM – but not LIM – approximately respects the statistical space–time factorization property mentioned above.

2.2 From LIM to SLIM

In this paper, we concentrate on the simplest SLIM model – for a single scalar quantity and we illustrate this by hindcasting global scale temperature series. The key change to the LIM model is thus a modification of the low frequency scaling: rather than $\beta_l = 0$ (white noise), the SLIM model has $1 > \beta_l > 0$. This can be effected by a simple extension of Eq. (1) to yield the fractional differential equation:

$$\frac{\mathrm{d}^{H+1/2}}{\mathrm{d}t^{H+1/2}} \left(\frac{\mathrm{d}}{\mathrm{d}t} + \varpi_\mathrm{w} \right) T = \sigma_\gamma \gamma(t) \tag{4}$$

where $H + 1/2$ is a fractional order of differentiation. This yields the temperature spectrum:

$$E_T(\varpi) \approx \varpi^{-(2H+1)} \frac{\sigma_\gamma^2}{(\varpi^2 + \varpi_\mathrm{w}^2)} \tag{5}$$

hence the low and high frequency SLIM exponents are: $\beta_l = 2H + 1$, $\beta_h = 2H + 3$. Note that for the global temperature series analysed below, we have $\beta_l \approx 0.6$ and $H \approx -0.2$ (see Fig. 4a and b below).

Alternatively, Eq. (4) can be solved in real space directly. First, operate on both sides of the above by $(\omega_w + \frac{d}{dt})^{-1}$ to obtain:

$$\frac{d^{H+1/2}}{dt^{H+1/2}}T = \gamma_s(t); \quad \gamma_s(t) = \sigma_\gamma \int_{-\infty}^{t} e^{-\omega_w(t-t')}\gamma(t')dt' \tag{6}$$

Since the autocorrelation of γ_s is:

$$\langle \gamma_s(t)\gamma_s(t-\Delta t)\rangle = e^{-\omega_w \Delta t}\sigma_{\gamma,s}^2; \quad \sigma_{\gamma,s}^2 = \frac{\sigma_\gamma^2}{2\omega_w} \tag{7}$$

We see that for lags $\Delta t \gg \omega_w^{-1}$ that γ_s is essentially an uncorrelated white noise: γ_s is simply γ smoothed over time scales shorter than $\tau_w = \omega_w^{-1}$.

If we are only interested in frequencies lower than ω_w, we can therefore simply solve:

$$\frac{d^{H+1/2}T}{dt^{H+1/2}} = \gamma(t) \tag{8}$$

The LIM paradigm is recovered as the special case with $H = -1/2$. Although physically, the weather scales are responsible for the smoothing at τ_w, in practice, we typically have climate data averaged at even lower resolutions: for example monthly or annually. Therefore, it is simpler to consider a "pure" process (with pure white noise forcing), and then introduce the resolution/smoothing simply as an averaging procedure.

Formally, the solution to Eq. (8) is obtained by (Riemann–Liouville) fractional integration of both sides of the equation by order $H + 1/2$:

$$T(t) = \frac{\sigma_\gamma}{\Gamma(1/2+H)} \int_{-\infty}^{t} (t-t')^{-(1/2-H)}\gamma(t')dt'; \quad -1/2 < H < 0 \tag{9}$$

(Γ is the usual gamma function). $T(t)$ is sometimes called fractional Gaussian noise (fGn). By inspection, the statistics are invariant under translations in time: $t \to t + \Delta t$ so that this process is stationary. Although basic processes of this type were first introduced by Kolmogorov (1940), since Mandelbrot and Van Ness (1968), the usual order one integral of Eq. (9) has received most of the mathematical attention: "fractional Brownian motion" (fBm). An interesting mathematical feature of fBm and fGn is that they are not semi-Martingales and hence the standard stochastic Itô and Stratatovitch calculi do not apply (see Biagini et al., 2008 for a recent mathematical review). In the present case, this is not important since we only deal with Wiener integrals (i.e. of fGn with respect to deterministic functions). The FIF model mentionned earlier has the same mathematical structure: it suffices to replace γ by a turbulent flux from a multiplicative cascade model; this overall model has the same fluctuation exponent H but is intermittent with moments other than first order potentially having quite different scaling.

While below we use simple averaging to obtain small scale convergence of fGn, for many purposes, the details of the smoothing at resolution τ are unimportant and it can be useful to define the particularly simple "truncated fGn" process:

$$T_{\tau\,\mathrm{trun}}(t) = \frac{\sigma_\gamma}{\Gamma(1/2+H)} \int_{-\infty}^{t} (t+\tau-t')^{-(1/2-H)}\gamma(t')dt'; \quad -1/2 < H < 0 \tag{10}$$

where the singular kernel is truncated at scale τ. It can be shown that for large enough lags Δt, the fluctuation and autocorrelation statistics for truncated fGn are the same as for the averaged fGn, although, when H approaches zero (from below), the convergence of the former to the latter becomes increasingly slow. In practice, the truncated model is often a convenient approximation to the slightly more complex averaged model.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

2.3 Properties of fGn

2.3.1 Definition and links to fBm

Fractional Brownian motion has received far more attention than fractional Gaussian noise and it is possible to deduce the properties of fGn from fBm. However, since we are exclusively interested in fGn, it is more straightforward to first define fGn and then – if needed – define fBm from its integral.

The canonical fractional Gaussian noise (fGn) process $G_H(t)$ with parameter H, can be defined as:

$$G_H(t) = \frac{c_H}{\Gamma(1/2+H)} \int_{-\infty}^{t} (t-s)^{H-1/2} \gamma(s) \mathrm{d}s \tag{11}$$

where c_H is a constant chosen so as to make the expression for the statistics particularly simple, see below. First, taking ensemble averages of both sides of Eq. (11) we find that the mean vanishes: $\langle G_H(t) \rangle = 0$. Now, take the average of G_H over a resolution τ:

$$G_{H,\tau}(t) = \frac{1}{\tau} \int_{t-\tau}^{t} G_H(t') \mathrm{d}t' \tag{12}$$

and define the function F_H which will be useful below:

$$F_H(\lambda) = \int_{0}^{\lambda-1} \left((1+u)^{H+1/2} - u^{H+1/2} \right)^2 \mathrm{d}u; \quad \lambda \geq 1 \tag{13}$$

with the particular value:

$$F_H(\infty) = \pi^{-1/2} 2^{-(2H+2)} \Gamma(-1-H) \Gamma(3/2+H) \tag{14}$$

and the asymptotic expression:

$$F_H(\lambda) = F_H(\infty) - \frac{(H+1/2)^2}{-2H} \lambda^{2H} + \ldots \tag{15}$$

If c_H is now chosen such that:

$$c_H = \frac{\Gamma(H+3/2)}{\left[F_H(\infty) + \frac{1}{2H+2} \right]^{1/2}} = \left(\frac{\pi}{2\cos(\pi H) \Gamma(-2H-2)} \right)^{1/2} \tag{16}$$

then we have:

$$\langle G_{H,\tau}^2 \rangle = \tau^{2H}; \quad -1 < H < 0 \tag{17}$$

This shows that a fundamental property is that in the small scale limit ($\tau \to 0$), 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 in addition $\langle G_{H,\tau}(t) \rangle = 0$, we see that sample functions $G_{H,\tau}(t)$ fluctuate about zero with successive fluctuations tending to cancel each other out; this is the hallmark of the macroweather regime.

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

$$\mathrm{d}B_{H'} = G_H(t) \mathrm{d}t; \quad H' = H+1; \quad 0 \leq H' \leq 1 \tag{18}$$

so that:

$$\Delta B_{H'}(\tau) = B_{H'}(t) - B_{H'}(t-\tau) = \int_{t-\tau}^{t} G_{H'}(t') \mathrm{d}t' = \tau G_{H',\tau}(t) \tag{19}$$

with the property:

$$\langle \Delta B_{H'}(\Delta t)^2 \rangle = \Delta t^{2H'} \tag{20}$$

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 $B_{H'}(0) = 0$, leading to the usual definition:

$$B_{H'}(t) = \frac{c_{H'}}{\Gamma(H'+1/2)}\left[\int_{-\infty}^{0}\left((t-s)^{H'-1/2} - (-s)^{H'-1/2}\right)\gamma(s)\mathrm{d}s + \frac{c_{H'}}{\Gamma(H'+1/2)}\int_{0}^{t}(t-s)^{H'-1/2}\gamma(s)\mathrm{d}s\right] \quad (21)$$

(Mandelbrot and Van Ness, 1968). Whereas fGn has a small scale divergence that can eliminated by averaging over a finite resolution τ, the fGn integral $\int_{-\infty}^{t} G_H(t')\mathrm{d}t'$ 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. (21): it eliminates this divergence at the price of imposing $B_{H'}(0) = 0$ so that fBm is nonstationary (although its increments are stationary, Eq. 19).

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 give 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 (see below); with this definition H is also useful for more general (multifractal) scaling processes. When $-1 < H < 0$, the mean at resolution τ (Eq. 12) defines the anomaly fluctuation (see below), 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. (20) 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 or-

der 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 $1/2 < H' < 1$ for the corresponding integrated process.

Some useful relations are:

$$\langle \mathrm{d}B_{H'}(t)\mathrm{d}B_{H'}(s)\rangle = \langle G_H(t)G_H(s)\rangle \mathrm{d}s\mathrm{d}t = |t-s|^{2H}\mathrm{d}s\mathrm{d}t \quad (22)$$

and:

$$\langle (B_{H'}(t_2) - B_{H'}(t_1))(B_{H'}(t_4) - B_{H'}(t_3))\rangle = \frac{1}{2}\left((t_4-t_1)^{2H'} + (t_3-t_2)^{2H'} - (t_3-t_1)^{2H'} - (t_4-t_2)^{2H'}\right) \quad (23)$$

valid for $t_1 < t_2 \le t_3 < t_4$ (e.g. Gripenberg and Norros, 1996).

The relationship Eq. (23) can be used to obtain several useful relations for finite resolution fGn. For example:

$$\langle G_{H,\tau_1}(t)G_{H,\tau_2}(t-\Delta t)\rangle = \frac{1}{2\tau_1\tau_2}\left((\Delta t+\tau_2)^{2H+2} + (\Delta t-\tau_1)^{2H+2} - \Delta t^{2H+2} - (\Delta t+\tau_2-\tau_1)^{2H+2}\right); \quad \begin{matrix}\Delta t \ge \tau \\ -1 < H < 0\end{matrix} \quad (24)$$

A convenient expression for the special case at fixed resolution $\tau = \tau_1 = \tau_2$ is:

$$R_{H,\tau}(\Delta t) = \langle G_{H,\tau}(t)G_{H,\tau}(t-\Delta t)\rangle = \frac{\tau^{2H}}{2}\left[(\lambda+1)^{2H+2} + (\lambda-1)^{2H+2} - 2\lambda^{2H+2}\right]; \lambda = \frac{\Delta t}{\tau} \quad \lambda \ge 1 \quad (25)$$

Where λ is the nondimensional lag i.e. measured in integer resolution units. This is convenient since real data is discretized in time and this shows that as long as we correct for the overall resolution factor (τ^{2H}), that the autocorrelation only depends on the nondimensional lag.

Since $H < 0$ and the large Δt limit is:

$$R_{H,\tau}(\Delta t) \approx (H+1)(2H+1)\Delta t^{2H}; \quad \Delta t \gg \tau \tag{26}$$

the autocorrelation falls off algebraically with exponent $2H$.

2.3.2 Spectrum and fluctuations

Since fGn is stationary, its spectrum is given by the Fourier transform of the autocorrelation function. The autocorrelation is symmetric: $R_{H,\tau}(\Delta t) = R_{H,\tau}(-\Delta t)$, so that for the Fourier Transform we use the absolute value of Δt. Also, we must take the limit of the autocorrelation of small resolution which is the same as using the large λ formula. In this case we obtain:

$$E(\omega) = \frac{\Gamma(3+2H)\sin \pi H}{\sqrt{2\pi}} |\omega|^{-\beta}; \quad \beta = 1 + 2H \tag{27}$$

The relation between β and H is the standard monofractal one, it is valid as long as intermittency effects are negligible i.e. if we ignore the multifractal "corrections". However, sometimes – as here for high order statistical moments – or in the case of precipitation even for low order moments - these can give the dominant contribution to the scaling.

The spectrum is one way of characterizing the variability as a function of scale (frequency), however it is often important to have real space characterizations. These are useful not only for understanding the effects of changing resolution, but also at a given time scale Δt for studying the full range of variability (i.e. statistical moments other than second order, probability distributions, etc.). Wavelets provide a general framework for defining fluctuations, we now give some simple and useful special cases.

Anomalies

An anomaly is the average deviation from the long term average and since $\langle G_H \rangle = 0$, the anomaly fluctuation over interval Δt is simply G_H at resolution Δt rather than τ:

$$(\Delta G_{H,\tau}(\Delta t))_{\text{anom}} = \frac{1}{\Delta t} \int_{t-\Delta t}^{t} G_H(t')\mathrm{d}t' = \frac{1}{\Delta t} \int_{t-\Delta t}^{t} G_{H,\tau}(t')\mathrm{d}t' = G_{H,\Delta t}(t) \tag{28}$$

Hence using Eq. (25):

$$\langle (\Delta G_{H,\tau}(\Delta t))^2_{\text{anom}} \rangle = \Delta t^{2H} \tag{29}$$

While this definition of fluctuation is fine for fGn, it is not appropriate for processes with $H > 0$ since these "wander", they do not tend to return to any long term value. Anomaly fluctuations were referred to with the less intuitive term "tendency" fluctuation in Lovejoy and Schertzer (2012a).

Differences

The classical fluctuation is the difference (the "poor man's wavelet"):

$$(\Delta G_{H,\tau}(\Delta t))_{\text{diff}} = G_{H,\tau}(t) - G_{H,\tau}(t - \Delta t) \tag{30}$$

Hence:

$$\langle (\Delta G_{H,\tau}(\Delta t))^2_{\text{diff}} \rangle = 2\tau^{2H} \left(1 + \lambda^{2H+2} - \frac{1}{2}\left((\lambda+1)^{2H+2} + (\lambda-1)^{2H+2}\right)\right); \quad \lambda = \frac{\Delta t}{\tau} \tag{31}$$

In the large Δt limit we have:

$$\langle (\Delta G_{H,\tau}(\Delta t))^2_{\text{diff}} \rangle \approx 2\tau^{2H} \left(1 - (H+1)(2H+1)\lambda^{2H}\right); \quad \lambda = \frac{\Delta t}{\tau} \gg 1 \tag{32}$$

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 Δt.

Haar fluctuations

As pointed out in Lovejoy and Schertzer (2012a), the preceding fluctuations only have variances proportional to τ^{2H} over restricted ranges of H, specifically $-1 \le H \le 0$ (anomalies), $0 \le H \le 1$ (differences), a more generally useful fluctuation (used below) is the Haar fluctuation (from the Haar wavelet, Haar, 1910). These are defined as the differences between the average of the first and second halves of the interval Δt:

$$(\Delta G_{H,\tau}(\Delta t))_{\mathrm{Haar}} = \frac{2}{\Delta t}\left[\int\limits_{t-\Delta t/2}^{t} G_{H,\tau}(t')\mathrm{d}t' - \int\limits_{t-\Delta t}^{t-\Delta t/2} G_{H,\tau}(t')\mathrm{d}t'\right] \tag{33}$$

Using Eq. (23), we obtain:

$$\langle(\Delta G_{H,\tau}(\Delta t))^2_{\mathrm{Haar}}\rangle = 4\Delta t^{2H}(2^{-2H}-1) \tag{34}$$

this indeed scales as Δt^{2H} and does not depend on the resolution τ.

2.4 Using fGn to model and forecast the temperature

Using the definition (Eq. 11) of fGn, we can define the temperature as:

$$T(t) = \sigma_T G_H(t) \tag{35}$$

(i.e. $\sigma_T = \sigma_\gamma / c_H$). Let us now introduce the integral $S(t)$:

$$S(t) = \int\limits_{-\infty}^{t} T(t')\mathrm{d}t' = \frac{1}{\Gamma(3/2+H)}\int\limits_{-\infty}^{t}(t-t')^{H+1/2}\gamma(t')\mathrm{d}t' \tag{36}$$

Since T is a fractional integral of order $1/2+H$ with respect to white noise, $S(t)$ is a fractional integral of order $3/2+H = 1/2+H'$. Strictly speaking, the above integral

diverges at $-\infty$, however this is not important since we will always take differences over finite intervals (equivalent to integrating $T(t)$ over a finite interval) and the differences will converge.

We can therefore define the resolution τ temperature as:

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

Notice that because of the divergence of $S(t)$ at $-\infty$, we did not define $S(t) = \sigma_T B_{H'}(t)$ however the differences do respect:

$$S(t)-S(t-\tau) = \sigma_T\left(B_{H'}(t)-B_{H'}(t-\tau)\right). \tag{38}$$

Using Eq. (35), the τ resolution temperature variance is thus:

$$\langle T_\tau^2\rangle = \sigma_T^2\tau^{2H} \tag{39}$$

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 $G_{H,\tau}(t)$.

2.5 Forecasts

Since an fGn process at resolution τ is the average of the increments of an fBm, process, it suffices to forecast fBm. There are four important related problems in the prediction of fBm: (a) to find the best forecast using finite past data, (b) infinite past data. The cases (1) $0 < H' < 1/2$ and (2) $1/2 < H' < 1$ (with $H' = 1+H$) must be considered separately. Since $-1/2 < H < 0$, our problem corresponds to cases 2a, 2b. Yaglom solved problem 1b in 1955 (Yaglom, 1955), Gripenburg and Norris solved 2a, 2b in 1996 (Gripenberg and Norros, 1996) and problem 1a was solved by Nuzman and Poor (2000). Hirchoren and Arantes (1998) used the Gripenburg and Norris results for $1/2 < H' < 1$ to numerically test the method adapted to fGn, see also Hirchoren and

D'attellis (1998). Although the Gripenberg and Norros (1996) result conveniently expresses the fBm predictions at time t (the "forecast horizon") directly in terms of the past series for $t \le 0$, the corresponding formulae are not simple.

The standard approach that they followed yields nontrivial integral equations (which they solved) in both the finite and infinite data cases. In what follows, we use a more straightforward method – the general method of innovations (see e.g. Papoulis, 1965, Chapt. 13) – and we obtain relatively simple results for the case with infinite past data (which is equivalent to the corresponding Gripenberg and Norros, 1996 result). In a future publication we improve on this by adapting it to the finite data case. The main new aspect of the forecasting problem with only finite data is that it turns out that not only do the most recent values (close to $t = 0$) have strong (singular) weighting, but the data in the oldest available data also have singular weightings. In the words of Gripenberg and Norris, this is because they are the "closest witnesses" of the distant past.

We now derive the forecast result for resolution τ fGn using innovations assuming that data is available over the entire past (i.e. from $t = -\infty$ to 0). Recall that the resolution τ temperature at time t is given by:

$$T_\tau(t) = \frac{S(t) - S(t-\tau)}{\tau} = \frac{c_H \sigma_T}{\tau \Gamma(H+3/2)} \left[\int_{-\infty}^{t} (t-t')^{H+1/2} \gamma(t') \mathrm{d}t' - \int_{-\infty}^{t-\tau} (t-\tau-t')^{H+1/2} \gamma(t') \mathrm{d}t' - \right] \tag{40}$$

We have used the fact that $S(t)$ in a fractional integral of order $H + 3/2$ of γ since the γ's are effectively independent random variables, they are called "innovations". If $T_\tau(t)$ is known for $t \le 0$, then the above relation can be inverted to obtain $\gamma(t)$ for $t \le 0$. If $\gamma(t)$ is known for $t \le 0$, then the minimum square (MS) estimator (circonflex) at time $t \ge \tau$ is

given by:

$$\hat{T}_\tau(t) = \frac{\hat{S}(t) - \hat{S}(t-\tau)}{\tau} = \frac{c_H \sigma_T}{\tau \Gamma(H+3/2)} \left[\int_{-\infty}^{0} (t-t')^{H+1/2} \gamma(t') \mathrm{d}t' - \int_{-\infty}^{0} (t-\tau-t')^{H+1/2} \gamma(t') \mathrm{d}t' - \right] \tag{41}$$

which depends only on $\gamma(t)$ for $t \le 0$. That this is indeed the MS estimator follows since the error E_T in this estimator is orthogonal to the estimator. To see this, note that E_T only depends on $\gamma(t)$ for $t \ge 0$:

$$E_T = T_\tau(t) - \hat{T}_\tau(t) = \frac{c_H \sigma_T}{\tau \Gamma(H+3/2)} \left[\int_{0}^{t} (t-t')^{H+1/2} \gamma(t') \mathrm{d}t' - \int_{0}^{t-\tau} (t-\tau-t')^{H+1/2} \gamma(t') \mathrm{d}t' \right] \tag{42}$$

Since

$$\left\langle \left(T_\tau(t) - \hat{T}_\tau(t) \right) \gamma(s) \right\rangle = 0; \quad t \ge 0; \quad s < 0 \tag{43}$$

and the range of integration for $\hat{T}_\tau(t)$ in Eq. (42) is $t' < 0$ whereas the range for the error E_T (Eq. 42) is $t' > 0$, $\hat{T}_\tau(t)$, E_T are clearly orthogonal: $\langle (T_\tau(t) - \hat{T}_\tau(t)) \hat{T}_\tau(t) \rangle = 0$. We can use this to obtain:

$$\langle E_T(t)^2 \rangle = \langle T_\tau(t)^2 \rangle - \langle T_\tau(t) \hat{T}_\tau(t) \rangle = \langle T_\tau(t)^2 \rangle - \langle \hat{T}_\tau(t)^2 \rangle \tag{44}$$

Using the substitution $u = -(t - \tau - t')/\tau$ in the integral Eq. (41) and the function $F_H(\lambda)$ introduced in Eq. (13), and using Eq. (16) for c_H, we obtain:

$$\langle \hat{T}_\tau(t)^2 \rangle = \sigma_T^2 \tau^{2H} \left[\frac{F_H(\infty) - F_H(\lambda)}{F_H(\infty) + \frac{1}{2H+2}} \right] \tag{45}$$

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

with $F_H(\infty)$ given in Eq. (14).

Using Eq. (44), the error variance is:

$$\langle E_T(t,\tau)^2\rangle = \langle T_\tau(t)^2\rangle - \langle \hat{T}_\tau(t)^2\rangle = \sigma_T^2\tau^{2H}\left[\frac{F_H(\lambda)+\frac{1}{2H+2}}{F_H(\infty)+\frac{1}{2H+2}}\right]$$
$$= \sigma_T^2\tau^{2H}\left[\frac{1+(2H+2)F_H(\lambda)}{1+(2H+2)F_H(\infty)}\right] \quad (46)$$

Hence, the fraction of the variance explained by the forecast, the "skill" (S_k) is:

$$S_k(\lambda) = \frac{\langle \hat{T}_\tau(t)^2\rangle}{\langle T_\tau(t)^2\rangle} = \left[\frac{F_H(\infty)-F_H(\lambda)}{F_H(\infty)+\frac{1}{2H+2}}\right]; \quad \lambda \geq 1 \quad (47)$$

Figure 1 a shows the theoretical skill as a function of H for different forecast horizons, and Fig. 1b for different forecast horizons as a function H. In Fig. 1a, dashed reference lines indicate the three empirically significant values: land ($H \approx -0.3$), global, ($H \approx -0.2$), ocean $H \approx -0.1$). In Fig. 1b, the estimated global value ($H = -0.20 \pm 0.03$, see below) is indicated in red.

This definition of skill is slightly different from the Root Mean Square Skill Score (RMSSS) that is sometimes used to evaluate GCM's (see e.g. Doblas-Reyes et al., 2013). The RMSSS is defined as one minus the ratio of the RMS error of the ensemble-mean prediction divided by the RMS temperature variation:

$$\mathrm{RMSSS} = 1 - \frac{\langle (T-\hat{T})^2\rangle^{1/2}}{\langle T^2\rangle^{1/2}} \quad (48)$$

In our case, the forecast is orthogonal to the prediction so that $\langle (T-\hat{T})^2\rangle = \langle T^2\rangle - \langle \hat{T}^2\rangle$ and we obtain:

$$\mathrm{RMSSS} = 1 - (1-S_k)^{1/2} \approx \frac{1}{2}S_k + \frac{1}{8}S_k^2 + \ldots \quad (49)$$

This shows that S_k and RMSSS are more or less equivalent skill measures both being in the range 0 to 1. However, GCM forecasts are generally *not* orthogonal to the data and for them, the RMSSS can be negative.

If the process is scaling over an infinite range in the data, but we only have access to a record of duration λ_{mem} (in "pixels") then:

$$S_{k,\lambda\mathrm{mem},\infty}(\lambda) = \left[\frac{F_H(\lambda_{\mathrm{mem}})-F_H(\lambda)}{F_H(\infty)+\frac{1}{2H+2}}\right]; \quad \lambda \geq 1 \quad (50)$$

To illustrate the potentially huge amount of memory in the climate system (especially in the ocean), we can (somewhat arbitrarily) define the memory in the system by the λ_{mem} value such that $S_{k,\lambda\mathrm{mem},\infty}(1)/S_{k,\infty,\infty}(1) = 0.9$, the result is shown in Fig. 1c. We see that over land (using $H = -0.3$), the memory estimated this way typically only goes back 15 pixels (nondimensional time steps), whereas over the ocean (using $H = -0.1$), it is 600. This means that the annual temperatures over the ocean typically have information from over 600 years in the past whereas over land, it is only 15 years. Note that these indicate the memory associated with 90 % of the skill (see Fig. 1a) and these skill levels fall off rapidly as H approaches the white noise value $H = -1/2$. We could also note that this calculation does *not* imply that we if we only had a short length of ocean data that the forecast would be terrible. This is because even if we only had 10 years of ocean data, the past from 10 years ago implicitly contains significant information from the distant past, and could in principle be exploited (see the numerical experiments in Hirchoren and Arantes, 1998).

In the real world, after the removal of the anthropogenic component (see Lovejoy and Schertzer, 2013, Fig. 4c), the scaling regime has a finite length (estimated as ≈ 100 years here), so that the memory in the process is finite. In addition, the monthly and annual resolution series that we hindcast below used memories of $\lambda = 180$, 20 units (months, years) respectively. The finite memory is easy to take into account; if the process memory extends over an interval of λ_{mem} units at resolution τ (i.e. over a time interval $t = \lambda_{\mathrm{mem}}\tau$) it suffices to integrate to λ_{mem} instead of infinity; i.e. to replace infinity by λ_{mem} in Eq. (50):

$$S_{k,\lambda\text{mem},\lambda\text{mem}}(\lambda) = \left[\frac{F_H(\lambda_{\text{mem}}) - F_H(\lambda)}{F_H(\lambda_{\text{mem}}) + \frac{1}{2H+2}}\right]; \quad \lambda_{\text{mem}} \geq \lambda \geq 1 \tag{51}$$

In Fig. 1d we show that the effect of finite memory increases strongly as H moves closer to zero, and is non negligible, even for $\lambda_{\text{mem}} = 180$, the largest used here (for the monthly series, when $H = -0.17$, the skill is reduced by 3–5 % up to $\lambda = 60$, see the bottom curves in Fig. 1d.

It is instructive to compare the skill obtained with the full memory with that if only the most recent variable $T_\tau(0)$ is used. The latter can be used either as classical persistence so that the forecast at time $t = \lambda\tau$ forecast to be equal to the present value (no change) (i.e. $\hat{T}_\tau(t) = T_\tau(0)$) or as "enhanced" persistence in which T_τ (0) is used as a linear estimator of $\hat{T}_\tau(t)$. Since the mean of the process is zero, for a single time step $t = \tau$ in the future, this is the same as the minimum square forecast made of an order 1 autoregressive model with nondimensional time step = 1: AR(1). Note however this equivalence is only for a single time step in the future, for forecasts further in the future; the AR(1) skill decays exponentially, not in a power law manner.

In persistence, the error in the forecast is simply the difference $E_T(t) = \Delta T_\tau(t) = T_\tau(t) - T_\tau(0)$, the skill is therefore $S_k = 1 - \langle \Delta T_\tau^2 \rangle / \langle T_\tau^2 \rangle$. In "enhanced persistence", the value T_τ (0) is simply considered as an estimator and the minimum square error linear estimator $\hat{T}(t)$ is only proportional to $T_\tau(0)$. A standard calculation (e.g. following (Papoulis, 1965), ch. 13) yields: $\hat{T}_\tau(t) = [\langle T_\tau(t) T_\tau(0) \rangle / \langle T_\tau(0)^2 \rangle] T_\tau(0)$ so that the term in the square brackets "enhances" the persistence value T_τ (0). Figure 1e compares the skill of the three estimators as functions of H for $\lambda = 1$ (i.e. using Eq. (25) for the autocorrelation): $\hat{T}_\tau(\tau) = (2^{2H+1} - 1) T_\tau(0)$. Whereas for $H \approx < -0.1$, classical persistence is quite poor, we see that the enhanced persistence forecast is much better.

3 Forecasting the Northern Hemisphere and global temperatures

3.1 The data and the removal of anthropogenic effects

In order to test the method, we chose the NASA GISS Northern Hemisphere and global temperature anomaly data sets, both at monthly and at annually averaged resolutions. A significant issue in the development of such global scale series is the treatment of the air temperature over the oceans which are estimated from sea surface temperatures; NASA provides two sets, the Land–Ocean Temperature Index (LOTI) and Land-Surface Air Temperature Anomalies only (Meteorological Station Data): the dT_s series. According to the site (http://data.giss.nasa.gov/gistemp/tabledata_v3/GLB.Ts+dSST.txt), LOTI provides a more realistic representation of the global mean trends than dT_s; it slightly underestimates warming or cooling trends, since the much larger heat capacity of water compared to air causes a slower and diminished reaction to changes; dT_s on the other hand it overestimates trends, since it disregards most of the dampening effects of the oceans that cover about two thirds of the Earth's surface. In order to compare the two, we used LOTI for the annual series and dT_s for the monthly series.

The prediction formulae assume that the series has the power law dependencies indicated above with RMS anomaly or Haar fluctuations following Δt^H (Eq. 34), and spectra with $\omega^{-\beta}$, with $\beta = (1 + 2H)$ (Eq. 27). However, this scaling only holds over the macroweather regime, and in the industrial epoch, anthropogenic forcing begins to dominate the low frequency variability at scales $\tau_c \approx$ 10–20 years whereas it occurs at scales $\tau_c \approx 100$ years in the pre-industrial epoch, see Lovejoy et al. (2013b) and Fig. 4d below. However, Lovejoy (2014a, b) showed that if the radiative forcing due to the observed global annually averaged CO_2 concentrations (ρ_{CO_2}) is used (proportional to $\log_2 \rho_{CO_2}$), that the "effective climate sensitivity" $\lambda_{2\times CO_2,\text{eff}}$ is quite close to the more usual "equilibrium climate sensitivity" estimated by GCM's and that the residues had statistics over the scale range 1 to ≈ 125 years that were very close to pre-industrial multiproxy statistics (see Table 2).

Therefore as a first step, using the Frank et al. (2010) data (extended to 2013 as described in Lovejoy, 2014b), we removed the anthropogenic contribution, using:

$$T(t) = T_{\text{anth}}(t) + T_{\text{nat}}(t) \tag{52}$$
$$T_{\text{anth}}(t) = \lambda_{2\times CO_2,\text{eff}} \log_2(\rho_{CO_2}(t)/\rho_{CO_2,\text{pre}}); \quad \rho_{CO_2,\text{pre}} = 277\,\text{ppm}$$

where $\rho_{CO_2,\text{pre}}$ is the pre-industrial concentration (= 277 ppm), the monthly data are shown as a function of date (Fig. 3a) and CO_2 forcing (Fig. 3b) with residues shown in Fig. 3c. The effective sensitivities are shown in Table 1a. We could note that if alternatively, the equivalent CO_2 since 1880 was used ("CO_2eq" as estimated in the IPCC AR5 report), the senstivities need only be divided by a factor 1.12, and the residues are essentially unchanged. This is because of the nearly linear relation between the actual CO_2 concentration and the estimated equivalent concentration (correlation coefficient > 0.993; see Table 1b for the SDs of the residues, T_{nat}). By using the observed CO_2 forcing as a linear surrogate for all anthropogenic effects we avoid various uncertain radiative assumptions needed to estimate CO_2 eq especially those concerning the cooling effects of aerosols which are still unsettled. As explained in (Lovejoy, 2014a), since the anthropogenic effects are linked via global economic activity, the observed CO_2 forcing is a plausible linear surrogate for all them.

From Table 1a we see that the sensitivities do not depend on the exact range over which they are estimated (columns 2–4). As we move to the present (column 4 to column 2), the sensitivities stay within the uncertainty range of the earlier estimates with the uncertainties constantly diminishing, consistent with the convergence of the sensitivities as the record lengthens. As a consequence, if we determine T_{anth} using the data only up to 1998 or up to 2013, there is very little difference: for the global data at monthly resolution, the difference in the standard deviations (SD's) of T_{nat} estimated with the different sensitivities is 0.005 K whereas at annual resolutions, it is 0.0035 K (for this period, $\Delta\log_2\rho_{CO_2} = 0.05$). These differences are larger than the estimated error in the global scale temperatures (estimated as ±0.03 K for both – the errors have very little scale dependence, Lovejoy et al., 2013a). From Table 1a, we see that there is a ≈ 30 %

difference between the global and monthly sensitivities due to the change from the LOTI (global) to d*Ts* (monthly) series the sensitivities are virtually independent of whether the data is at one month or one year resolution. We also see that the Northern Hemisphere has systematically higher sensitivities than the entire globe, this is consistent with the larger land mass in the north and the larger sensitivity of land with respect to the ocean.

An obvious criticism of the method of effective climate sensitivities is that anthropogenic forcing primarily warms the oceans and only with some lag, the atmosphere. Systematic cross-correlation analysis in Lovejoy (2014a, b) shows that while the residues are barely affected (see rows 2 and 3 in Table 2 and Lovejoy, 2014a for more on this), the values of the sensitivities are affected (see e.g. column 4 in Table 1a).

Finally, we can note that the difference between LOTI and d*Ts* temperature is primarily the sensitivities (Table 1a); that the remaining differences in the residues is mostly due to their different resolutions. From Eq. (39) we see that the ratio of RMS fluctuations in these should be λ^H where λ is the resolution ratio, here 12 months year^{-1}. Table 2 shows that the *H* estimated from the RMS values is indeed close to the *H* estimated more directly in the next subsection. This shows that the main difference between the LOTI and d*Ts* series is indeed their climate sensitivities.

In order to judge how close the residues from the CO_2 forcing (Eq. 52) are to the true natural variability, we can make various comparisons (Table 1b). Starting at the top (row 1), we see that, as shown in Lovejoy (2014a), the statistics of the resulting residues are very close to those of pre-industrial multiproxies (see also Fig. 4c below). In row 3, we see that we take the residues of the 20 year lagged temperatures, there is virtually no difference (although the sensitivities are significantly higher, see Table 1a). As further reference, (row 4), we see that it is substantially smaller than the SD of the linearly detrended series (i.e. when the residues are calculated from a linear regression with time rather than the forcing).

As further evidence that they provide a good estimate of the true natural variability, in rows 5–10 we also show the annual RMS errors of various GCM global temperature hindcasts. For example, in rows 5–6 we compare hindcasts of CMIP 3 GCM's both with and without annual data initialisation, assimilation (rows 5, 6). Without initialization (row 5), the results are half way between the CO_2 forcing residues (i.e. T_{nat}, row 2) and the SD of the linearly detrended series (row 4), i.e. the forecast is poor even for the anthropogenic part. Unsurprisingly, with annual data initialisation, assimilation (row 6) it is much better, but it is apparently still unable to do better than simply estimating the anthropogenic component. We can deduce this since the resulting RMS errors are virtually identical to the SD of the estimated T_{nat} (row 3). This conclusion is reinforced in row 7 where CMIP 3 GCM's (without data initialization) were analyzed. However, in place of annual data initialization, a complex empirical bias and variance correction scheme was implemented in order to keep the statistics of uninitialized hindcasts close to the data. We see that the resulting RMS error is virtually identical to GCM with data initialization (row 6) as well as the SD of T_{nat} (row 3). They are also very close to other GCM estimates of natural variability. These conclusions are reinforced in the 5 and 9 year "anomaly" columns. As expected – due to the averaging of the temperature in the definition of the anomalies out to the forecast horizon – the RMS error decreases. However, it is still only barely better than the T_{nat} estimates from the residues.

Very similar results are indicated in rows 8–10 for other GCM hindcast experiments, these are shown graphically in Fig. 2, which is adapted from a multimodel ENSEMBLES experiment hindcasts discussed in Garcia-Serrano and Doblas-Reyes (2012). The multimodel mean is consistently close to – but generally a little above – T_{nat} (bottom horizontal line) while remaining better than the SD of the linearly detrended temperature (top horizontal line). Also shown in Table 2 and Fig. 2 are the results of LIM, SLIM and other stochastic models, these will be discussed further in Sect. 4. For now suffice it to indicate that the SLIM model error is bounded above by the SD of T_{nat}. By using the long range memory to forecast T_{nat}, it can only do better. It thus generally improves upon the GCM's and – for two year horizons and beyond – it is better than the > 100 parameter LIM model whose 9 year forecast is essentially equivalent to a linear detrending.

3.2 Estimating H from the residues

Having estimated T_{nat} by removing the anthropogenic contribution, we may now test the quality of the scaling and estimate H. Figure 4a shows the raw spectra of the residues showing the scaling but with large fluctuations (as expected) with $\beta \approx 0.60$. We have already mentionned that the intermittency is low in this macroweather regime, indeed using exponents estimated in (Lovejoy and Schertzer, 2013), the resulting multifractal corrections to the variance are ≈ 0.01–0.02 so that we may use the monofractal relation $\beta = 1 + 2H$ which yields: $H \approx -0.20$. Slightly more accurate estimates can be obtained by averaging the spectrum over logarithmically spaced bins (Fig. 4b, and by compensating the spectrum by dividing it by the theoretical spectrum with $\beta = 0.54$ ($H = -0.17$). This figure makes the estimate $\beta = 0.20 \pm 0.06$ ($H = -0.20 \pm 0.03$) plausible. Finally, the corresponding RMS Haar fluctuations are shown in Fig. 4c, we see that they plausibly follow $H = -0.20$ out to about 100 years (the sharp drop at the largest lag is not significant: it corresponds to a single long fluctuation that is somewhat biased since some of the low frequency natural variability is also removed when T_{nat} is estimated by the method of residuals.

Also shown for reference is the GISS-E2-R millennium control run (with fixed forcings), as well as the RMS fluctuations for three pre-industrial multiproxies. We see that out to about 100 year scales, all the fluctuations have nearly the same amplitudes as functions of scale supporting the idea that T_{nat} as estimated by residuals is indeed a good estimate of the natural variability, and also confirming the estimate the global scale exponent value $H = -0.20 \pm 0.03$.

As a final comparison, Fig. 4d shows RMS Haar fluctuations for the global averages (from Fig. 4c), land only averages and from the oceans – the Pacific Decadal Oscillation (PDO). The PDO is the amplitude of the largest eigenvalue of the Pacific Sea Surface Temperature autocorrelation matrix (i.e. the amplitude of the most important Empirical Orthogonal Function: EOF). For the land only curve, notice the sharp rise for scales $>\approx 10$ years; this is the effect of the anthropogenic signal that was not removed in this

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

series. Overall we see that (roughly) for land $H \approx -0.3$, for the globe, $H = -0.2$, and for the oceans, $H = -0.1$. Figure 1a, c shows the drastic differences in memory implied by these apparently small changes in H.

4 Testing SLIM by hindcasting

4.1 The numerical approach

The theory for predicting fGn leads to the general equation for the variance of forecast error (E_T) at forecast horizon t, resolution τ, Eq. (47). In order to test the equation on the temperature residues, we can use the global and Northern Hemisphere series analyzed in the previous section and systematically make hindcasts. In this first study, we took a simple, straightforward approach based on the method of innovations. We discretised Eq. (9), which was then written as a matrix equation of the form: $T_t = \sum_{t'<t} M_{t,t'}\gamma_{t'}$ where the indices refer to the discrete time nondimensionalized by the series resolution, and $M_{tt'}$ which is the (singular) kernel from the fractional integration. The sum was over finite past of length $t_{\text{mem}} = \lambda_{\text{mem}}\tau$ units (see below) and the matrix was then inverted to yield the corresponding innovations γ_t. To make the forecast at time $t+\Delta t$ (i.e. Δt units in the future), the equation was used with an augmented kernel $M_{t+\Delta t,t'}$ with the innovation vector lengthened by appending Δt zeroes (the expectation values of the unknown future innovations) to the t_{mem} innovations that were determined in the previous step.

While our approach has the advantage of being straightforward (and it was tested on numerical simuations of fGn), in future applications improvements could be made. For example, by using a Girsanov formula, we could rewrite fGn in terms of a finite integral (see Biagini et al., 2008), and the discretised numerics would then be more accurate (this is especially important for H near the limiting values 0 and $-1/2$). Alternatively, we could use (Gripenberg and Norros, 1996) integral equation approach discretized

with a variant of the (Hirchoren and Arantes, 1998), approach which notably has the advantage of requiring less past data.

4.2 Results

In order to obtain good hindcast error statistics, it is important to make and validate as many hindcasts as possible, i.e. one for each discretised time that is available. However, due to the long-range correlations, we want to use a reasonable number of past time steps in the hindcast for memory, so that the earliest possible hindcast will be later than the earliest available data by the corresponding amount. The compromise used here consisted of dividing the 134 year series into 30 annual blocks (annual resolution) and 20 year blocks (monthly resolution). In each block in the annual series, the first 20 years were used as "memory" to develop the hindcast over the next 10 years for estimating the hindcast errors: a total of 134 – 30 = 104 forecasts were made. For the monthly series, the same procedure involved blocks of 240 months: 180 months for the memory and 60 months for the hindcast for a total of 1608 – 240 = 1368 hindcasts.

The hindcasts can be evaluated at various resolutions and forecast horizons, Eq. (47) gives the general theoretical result. The cases of special interest are the temperature hindcasts and the anomaly hindcasts with (resolutions, horizons) of (τ, $\lambda\tau$) and ($\lambda\tau$, $\lambda\tau$) respectively. The error variance ratios (R) are:

$$R_{\text{temp}} = \frac{\langle E_T(\lambda\tau,\tau)^2\rangle}{\langle E_T(\tau,\tau)^2\rangle} = 1 + (2+2H)F_H(\lambda) \tag{53}$$

and:

$$R_{\text{anom}} = \frac{\langle E_T(\lambda\tau,\lambda\tau)^2\rangle}{\langle E_T(\tau,\tau)^2\rangle} = \lambda^{2H} \tag{54}$$

Both ratios are shown in Fig. 5 along with the exact theory curves and Table 3 gives the corresonding highest resolution SDs (for both lagged and unlagged estimates of

T_{nat}, there is virtually no difference). It is seen that all the forecast error variances (global, northern, annual, monthly resolution) collapse quite well between the theory curves corresponding to $H = -0.17$ and $H = -0.23$ corresponding to $H \approx -0.20 \pm 0.03$ (although they are closer to the $H = -0.17$ curves). It is important to stress that Fig. 5 is completely nondimensional, it depends on a single parameter (H), and this parameter was estimated earlier using a quite different technique (Haar fluctuations and spectra) that had no direct relation to the property being measured (forecast skill). We have effectively used spectral and Haar analysis of scaling to determine the accuracy of forecasts using no extra information. Figure 5 has no adjustable parameters so that the agreement of the hindcast errors with theory is a particularly strong confirmation of the theory. We could add that the fact that the errors depend only on the dimensionless forecast horizon is also a consequence of the scaling, i.e. on the lack of strong characteristic time scale in the macroweather regime.

Since the anomaly errors are power laws (Eq. 54), they can conveniently evaluated on a log–log plot; see Fig. 6. Note that the RMS anomaly errors decrease with forecast horizon. The reason is that while forecasts further and further in the future loose accuracy, this loss is more than compensated by the decrease in the variance due to the lower resolution, so that the anomaly variance decreases.

4.3 Hindcast skill

Another way to evaluate the hindcasts is to determine their nondimensional skills i.e. the fraction of the variance that they explain (see the general formula Eq. 47). From the formula, we can see that the skill depends only on the nondimensional forecast horizon $\lambda = t/\tau$. Therefore the skill for forecast anomalies – i.e. the average of the forecast up to the horizon i.e. $t = \tau$, hence $\lambda = 1$, has the remarkable property of being constant, independent of the horizon. The reason is that while forecasts further and further in the future loose accuracy, this loss is exactly compensated by the decrease in the variance due to the lower resolution, so that the anomaly skill does not change. Figure 7 is another example of a nondimensional plot where the theory involves no

adjustable parameters, it shows that the theoretical prediction is well respected by the global, Northern Hemisphere annual and global resolution series. Since we estimated $H = -0.20 \pm 0.03$, it can be seen that the skill for the monthly series is nearly as high as theoretically predicted up to a year or so for the global, but up to several years for the Northern Hemisphere series. The global series has slightly lower forecast skill than theorically predicted, but is still of the order of 15 % at 10 years. Also shown is the effect of using only a finite part of the memory.

The skill in usual temperature forecasts (i.e. with fixed resolution τ, and increasing horizon $t = \lambda\tau$) is shown in Fig. 8. We see that monthly series can be predicted to nearly the theoretical limit up to about 2–3 years (≈ 5 % skill), for the annual series, this is up to about 5 years (≈ 10 % skill). Again the results are close to the $H = -0.17$ theory.

4.4 Hindcast correlations

A final way to evaluate the hindcasts is to calculate the correlation coefficient between the hindcast and the temperature:

$$\rho_{\hat{T},T}(t,\tau) = \frac{\langle \hat{T}_\tau(t) T_\tau(t)\rangle - \langle \hat{T}_\tau(t)\rangle\langle T_\tau(t)\rangle}{\langle \hat{T}_\tau(t)^2\rangle^{1/2}\langle T_\tau(t)^2\rangle^{1/2}} \tag{55}$$

Since $\langle T\rangle = 0$, the cross term vanishes; using Eq. (44) we obtain the simple result:

$$\rho_{\hat{T},T}(t,\tau) = \left(\frac{F_H(\infty) - F_H(\lambda)}{F_H(\infty) + \frac{1}{2H+2}}\right)^{1/2}; \quad \lambda = \frac{t}{\tau} \tag{56}$$

(i.e. $= S_k{}^{**1/2}$) or, asymptotically for $\lambda \gg 1$:

$$\rho_{\hat{T},T}(t,\tau) \approx 2^{H+1/2}\left(H + \frac{1}{2}\right) U^{1/2}\lambda^H; \quad \lambda \gg 1; \quad U = \frac{\sqrt{\pi}}{2\Gamma(1-H)\Gamma(\frac{3}{2}+H)} \tag{57}$$

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

In the special cases of anomalies $t = \tau$, $\lambda = 1$ and we obtain:

$$\rho_{\hat{T},T}(t,t) = \sqrt{1 + HU2^{2H+2}} \quad (58)$$

so that the correlations are constant at all forecast horizons. Over the range $-1/2 < H < 0$, the constant U is close to unity.

As in the previous hindcast error analyses, the series were broken into blocks and the forecasts were repeated as often as possible; each forecast was correlated with the observed sequence and averages were performed over all the forecasts and verifying sequences (the mean correlation given by the thick lines), Fig. 9. The uncertainty in the hindcast correlation coefficients was estimated by breaking the hindcasts into thirds: three equal sized groups of blocks with the error being given by the SD of the three about the mean (dashed lines). Also shown in Fig. 9 are the theoretical curves (Eq. 54) for $H = -0.20$, in this case the dashed lines indicate the theory for one SDs in H i.e. for $H = -0.17$, $H = -0.23$.

As predicted by Eq. (55), the anomaly correlations are relatively constant up to about 5 years for the annual data (top row), and nearly the same for the monthly data (bottom row). In addition, the northern series (blue) are somewhat better forecast than the global series (red). It can be seen that temperature forecasts (i.e. with fixed resolutions) have statistically significant correlations out to 8–9 years for the annual forecasts, out to about 2 years for the monthly global and nearly 5 years for the monthly Northern Hemisphere forecasts (bottom dashed lines). The anomaly forecasts are statistically significantly correlated at all forecast horizons. Figure 9 provides more examples of nondimensional plots with no free parameters, and again the agreement with the hindcasts validation is remarkable.

Although the results for the anomaly correlations are quite close to those of hindcasts in Garcia-Serrano and Doblas-Reyes (2012), the latter are for the entire temperature forecast, not just the natural variability as here. This means that the GCM correlations will be augmented with respect to ours due to the existence of long term anthropogenic

trends in both the data and the forecasts that are absent in ours (but even with this advantage, their correlations are not higher).

4.5 Comparison with GCM's, LIM, AR(1) and ARFIMA hindcasts

In Table 2 and Fig. 2, we have already compared GCM hindcast errors with estimates of the natural variability (T_{nat}) from the residues of a linear regression on the CO_2 radiative forcing since 1880. We found that the annual, global GCM hindcasts had errors that were close to, but generally larger than the SD of T_{nat} ($\langle T_{\mathrm{nat}}^2\rangle^{1/2}$) but smaller than the SD of the linearly detrended temperature series (the horizontal lines in Fig. 2). $\langle T_{\mathrm{nat}}^2\rangle^{1/2}$ is the RMS error of an unconditional forecast (i.e. with no knowledge of the past): $\langle T_{\mathrm{nat},\tau}^2\rangle = \langle E_T^2(\tau,\infty)\rangle$ (see Eq. 46), it is the upper bound to the hindcast errors. In Fig. 2, we see that the one-parameter stochastic hindcast (with $H = -0.2$) is somewhat better than the GCM's up to about 6 years after which it is about the same. This bolsters the hypothesis that GCM's primarily model the anthropogenic temperature change, not the natural variability whereas SLIM has some skill in forecasting the latter.

Table 2 and Fig. 2 also compare these to LIM hindcasts modelled with 20 degrees of freedom (involving > 100 parameters). We see that LIM is slightly better than SLIM for horizons up to about 2 years beyond which SLIM is better. According to the analysis in (Newman, 2013), for periods beyond about a year, the forecasts are mostly determined by the two most important Empirical Orthogonal Functions (EOF's), and their skill decays exponentially, not as a power law. From Fig. 2, their main effect seems to be to remove the long term linear trend allowing LIM to have an asymptotic RMS error roughly equal to the SD of the linearly detrended series (the upper horizontal line).

Finally, in Table 2, rows 12, 13, we have compared the errors with those of an early attempt at scaling temperature forecasts using the AutoRegressive Fractionally Integrated Moving Average process (ARFIMA) (Baillie and Chung, 2002) along with the corresponding order one AutoRegressive (AR(1)) process. Unfortunately, the forecasts were made by taking 10 year segments and in each removing a separate linear trend so that the low frequencies were not well accounted for (see the footnote to the table for more details). The AR(1) results were not so good: close to the SDs of the detrended temperatures. As expected – because it assumes a basic scaling framework – the ARFIMA results were somewhat better. Yet they are substantially worse than the other methods, probably because they did not remove the anthropogenic component first.

5 Conclusions

GCM's are basically weather models whose forecast horizons are well beyond the deterministic predictability limits, corresponding to many lifetimes of planetary scale structures: the macroweather regime. In this regime – that extends from about 10 days to ≈ 100 years (preindustrial), the weather patterns that are generated are essentially random noise. With fixed boundary conditions, GCM's therefore converge asymptotically (in a power law manner, Fig. 4c) to the their (model) climates. In order to model the low frequency variations associated with the climate proper, the GCM's must be externally forced; if the forcing is strong enough, in principle it can reverse the trend of macroweather fluctuations decreasing with increasing time scale and initiate a new climate regime where fluctuations instead increase with scale (as they do in the weather regime, see Lovejoy et al., 2013b). In the real world (pre-industrial), this occurs somewhere around 100 years and fluctuations increasing in scaling manner (but now with $H > 0$) out to ice-age time scales ($\approx$ 50–100 kyr). In addition, the real world may involve new, slow internal processes that become important at these scales.

In this view, the problem with the GCM approach is that in spite of massive improvements over the last 40 years, the weather noise that they generate isn't totally realistic nor does their climate coincide exactly with the real climate. In an effort to overcome these limitations, stochastic models have been developed that directly and more realistically model the noise and use real world data to exploit the system's memory so as to force the forecasts to be more realistic.

The main approaches that could potentially overcome these limitations are the stochastic ones. However, going back to Hasselmann (1976) these only use integer ordered differential equations, they implicitly assume that the low frequencies are white noises – and hence cannot be forecast with any skill. Modern versions – the Linear Inverse Models (LIM) add sophistication and a large number of (usually, but not necessarily) spatial parameters, but they still impose a short (exponentially correlated) memory and they focus on periods up to a few years at most. This contrasts with turbulence based nonlinear stochastic models which assume that the system is scaling over wide ranges. When they are extended to the macroweather regime, they predict low intermittency, scaling fluctuations with exponents close to those that are observed by a growing macroweather scaling literature. Contrary to their behaviour in the weather regime, in macroweather they are only weakly nonlinear, the Scaling LInear Macroweather model (SLIM) is thus an approximation to this more general Extended Fractionally Integrated Flux (EFIF) model.

In this paper, we therefore make the assumption that the low frequencies are not white noises, that they have very long memories. The simplest relevant model is of fractional Gaussian noise (fGn) process whose integral is the better known fractional Brownian motion (fBm) process. SLIM can be obtained as a solution of a fractional order generalization of the usual LIM differential equation. Although we only discuss the scalar version for single time series – here global scale temperatures – in future publications we will show how to extend SLIM to vector versions yielding regional forecasts and accounting for the different state variables (i.e. not only the temperature).

In Sect. 2, we situate the process in the mathematical literature and derive basic results for forecasts and forecast skill. These results show that a remarkably high level of skill is available in the climate system; for example for forecast horizons of one nondimensional time unit in the future (i.e. horizons equal to the resolution), the forecast skills – defined as the fraction of the variance explained by the forecast – are 15, 35, 64 % for land, the whole globe and oceans respectively (Fig. 1b; taking rough exponent values $H = -0.3, -0.2, -0.1$ respectively, Fig. 4c). To quantify the size of the memory, it can be defined as the number of nondimensional units needed to supply 90 % of the full memory of the system. Using the same empirical exponents, we found that the memory is 15, 50, 600 for land, the globe and ocean regions respectively.

The SLIM model forecasts the natural variability so that the responses to solar and volcanic forcings are implicitly included in the forecast. However, the responses to the anthropogenic forcings are not; we must therefore remove the anthropogenic component which becomes dominant at scales of 10–30 years. For this, we follow Lovejoy (2014a) who showed that the CO_2 radiative forcing is a good linear proxy for all the anthropogenic effects (including the difficult to estimate cooling due to aerosols) so that the natural variability is the residue with respect to a regression against the forcing. In Table 2, Fig. 2, we showed that the resulting SD (± 0.109 K) is very close to the RMS errors in annual, globally averaged GCM temperature hindcasts that use annual data initialisation, assimilation. Indeed, to a good approximation, all the models have errors bounded between this estimate of the natural variability and the slightly higher SD of the linearly detrended temperature series (± 0.163 K). This is true in spite of the fact that they are "optimistic" since they assume that the future volcanic and solar forcings are known in advance. The only partial exception is the stochastic LIM model (with > 100 parameters) which is only marginally better (± 0.085 K) than SLIM for forecast horizons of one to two years after which it asymptotes to the linearly detrended SD.

Using the method of innovations, we developed a new way of forecasting fGn that allows SLIM hindcasts to be made; the long-time forecast horizon RMS error is thus ± 0.109 K, the exploitation of the memory with the single parameter – the exponent $H \approx -0.20 \pm 0.03$ – reduces this to $\approx \pm 0.093$ K for one year global hindcasts so that SLIM remains better than or comparable to the multimodel GCM mean (Fig. 2).

Since this paper only deals with single time series (global scale temperatures) it is ideal for revisiting the problem of the "pause" or "slow down", "hiatus" in the warming since 1998. Lovejoy (2015) shows how SLIM hindcasts nearly perfectly predict this hiatus. However, most applications involve predicting the natural variability at regional scales. In a future publication, building on Lovejoy (2014b) we show how this can be

done and quantify the improvement that the additional information (from the regional memory) makes to the forecasts. For forecasts from months to a decade or so, the SLIM forecast results promise to be better than alternatives.

Acknowledgements. We thank C. Penland and P. Sardeshmuhk for helpful discussions. There are no conflicts of interest, this work was unfunded although L. del Rio Amador thanks Hydro-Quebec for a scholarship.

References

Ammann, C. M. and Wahl, E. R.: The importance of the geophysical context in statistical evaluations of climate reconstruction procedures, Climatic Change, 85, 71–88, doi:10.1007/s10584-007-9276-x, 2007.

Baillie, R. T. and Chung, S.-K.: Modeling and forecasting from trend-stationary long memory models with applications to climatology, Int. J. Forecasting, 18, 215–226, 2002.

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

Blender, R., Fraedrich, K., and Hunt, B.: Millennial climate variability: GCM-simulation and Greenland ice cores, Geophys. Res. Lett., 33, L04710, doi:10.1029/2005GL024919, 2006.

Bryson, R. A.: The paradigm of climatology: an essay, B. Am. Meteorol. Soc., 78, 450–456, 1997.

Dijkstra, H.: Nonlinear Climate Dynamics, Cambridge University Press, Cambridge, 357 pp., 2013.

Doblas-Reyes, F. J., Andreu-Burillo, I., Chikamoto, Y., Garcıa-Serrano, J., Guemas, V., Kimoto, M., Mochizuki, T., Rodrigues, L. R. L., and van Oldenborgh, G. J.: Initialized near-term regional climate change prediction, Nat. Commun., 4, 1715, doi:10.1038/ncomms2704, 2013.

Frank, D. C., Esper, J., Raible, C. C., Buntgen, U., Trouet, V., Stocker, B., and Joos, F.: Ensemble reconstruction constraints on the global carbon cycle sensitivity to climate, Nature, 463, 527–530, doi:10.1038/nature08769, 2010.

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

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

Gripenberg, G. and Norros, I.: On the prediction of Fractional Brownian Motion, J. Appl. Probab., 33, 400–410, 1996.

Haar, A.: Zur Theorie des orthogonalen Funktionsysteme, Math. Ann., 69, 331–371, 1910.

Hasselmann, K.: Stochastic climate models, Part I: theory, Tellus, 28, 473–485, 1976.

Hirchoren, G. A. and Arantes, D. S.: Predictors for the discrete time fractional Gaussian processes, in: Telecommunications Symposium, 1998, ITS '98 Proceedings, SBT/IEEE International, IEEE, Sao Paulo, 49–53, 9–13 August 1998.

Hirchoren, G. A. and D'attellis, C. E.: Estimation of fractal signals, using wavelets and filter banks, IEEE T. Signal Proces., 46, 1624–1630, 1998.

Huang, S.: Merging information from different resources for new insights into climate change in the past and future, Geophys. Res. Lett., 31, L13205, doi:10.1029/2004GL019781, 2004.

Huybers, P. and Curry, W.: Links between annual, Milankovitch and continuum temperature variability, Nature, 441, 329–332, doi:10.1038/nature04745, 2006.

Kolesnikov, V. N. and Monin, A. S.: Spectra of meteorological field fluctuations, Izv. Atmos. Ocean Phy.+, 1, 653–669, 1965.

Kolmogorov, A. N.: Wienershe spiralen und einige andere interessante kurven in Hilbertschen Raum, Dokl. Akad. Nauk SSSR+, 26, 115–118, 1940.

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

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

Lovejoy, S.: What is climate?, EOS T. Am. Geophys. Un., 94, 1–2, 2013.

Lovejoy, S.: Scaling fluctuation analysis and statistical hypothesis testing of anthropogenic warming, Clim. Dynam., 42, 2339–2351, doi:10.1007/s00382-014-2128-2, 2014a.

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

Lovejoy, S.: Using scaling for macroweather forecasting including the pause, Geophys. Res. Lett., submitted, 2015.

Lovejoy, S. and de Lima, M. I. P.: The joint space–time statistics of macroweather precipitation and space–time factorization, Chaos, submitted, 2015.

Lovejoy, S. and Schertzer, D.: Scale invariance in climatological temperatures and the local spectral plateau, Ann. Geophys., 4B, 401–410, 1986.

Lovejoy, S. and Schertzer, D.: Towards a new synthesis for atmospheric dynamics: space–time cascades, Atmos. Res., 96, 1–52, doi:10.1016/j.atmosres.2010.01.004, 2010.

Lovejoy, S. and Schertzer, D.: Haar wavelets, fluctuations and structure functions: convenient choices for geophysics, Nonlinear Proc. Geoph., 19, 1–14, doi:10.5194/npg-19-1-2012, 2012a.

Lovejoy, S. and Schertzer, D.: Low frequency weather and the emergence of the climate, in: Extreme Events and Natural Hazards: The Complexity Perspective, edited by: Sharma, A. S., Bunde, A., Baker, D. N., and Dimri, V. P., AGU Monographs, American Geophysical Union, Washington D.C., 231–254, 2012b.

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

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

Lovejoy, S., Schertzer, D., and Varon, D.: Do GCMs predict the climate ... or macroweather?, Earth Syst. Dynam., 4, 439–454, doi:10.5194/esd-4-439-2013, 2013b.

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

Mandelbrot, B. B.: Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier, J. Fluid Mech., 62, 331–350, 1974.

Mandelbrot, B. B. and Van Ness, J. W.: Fractional Brownian motions, fractional noises and applications, SIAM Rev., 10, 422–450, 1968.

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

Newman, M.: An empirical benchmark for decadal forecasts of global surface temperature anomalies, J. Climate, 26, 5260–5269, doi:10.1175/JCLI-D-12-00590.1, 2013.

Newman, M. P., Sardeshmukh, P. D., and Whitaker, J. S.: A study of subseasonal predictability, Mon. Weather Rev., 131, 1715–1732, 2003.
Novikov, E. A. and Stewart, R.: Intermittency of turbulence and spectrum of fluctuations in energy-disspation, Izv. Akad. Nauk. SSSR Ser. Geofiz., 3, 408–412, 1964.
Nuzman, C. J. and Poor, H. V.: Linear estimation of self-similar processes via Lamperti's transformation, J. Appl. Probab., 37, 429–452, 2000.
Panofsky, H. A. and Van der Hoven, I.: Spectra and cross-spectra of velocity components in the mesometeorlogical range, Q. J. Roy. Meteor. Soc., 81, 603–622, 1955.
Papoulis, A.: Probability, Random Variables and Stochastic Processes, Mc Graw Hill, New York, 1965.
Penland, C. and Sardeshmuhk, P. D.: The optimal growth of tropical sea surface temperature anomalies, J. Climate, 8, 1999–2024, 1995.
Rypdal, K., Østvand, L., and Rypdal, M.: Long-range memory in Earth's surface temperature on time scales from months to centuries, J. Geophys. Res.-Atmos., 118, 7046–7062, doi:10.1002/jgrd.50399, 2013.
Sardeshmukh, P. D. and Sura, P.: Reconciling non-gaussian climate statistics with linear dynamics, J. Climate, 22, 1193–1207, 2009.
Schertzer, D. and Lovejoy, S.: Physical modeling and analysis of rain and clouds by anisotropic scaling of multiplicative processes, J. Geophys. Res., 92, 9693–9714, 1987.
Smith, D. M., Cusack, S., Colman, A. W., Folland, C. K., Harris, G. R., and Murphy, J. M.: Improved surface temperature prediction for the coming decade from a global climate model, Science, 317, 796–799, 2007.
Vallis, G.: Mechanisms of climate variaiblity from years to decades, in: Stochstic Physics and Climate Modelling, edited by: Palmer, P. W. T., Cambridge University Press, Cambridge, 1–34, 2010.
Van der Hoven, I.: Power spectrum of horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour, J. Meteorol., 14, 160–164, 1957.
Yaglom, A. M.: Correlation theory of processes with random stationary nths increments (Russian) [English Transl.], Amer. Math. Soc. Trans. Ser., 8, 87–141, Mat. Sb. N. S., 37, 141–196, 1955.
Yaglom, A. M.: The influence on the fluctuation in energy dissipation on the shape of turbulent characteristics in the inertial interval, Sov. Phys. Dokl., 2, 26–30, 1966.

Yuan, N., Fu, Z., and Liu, S.: Extracting climate memory using Fractional Integrated Statistical Model: a new perspective on climate prediction, Nature Sci. Rep., 4, 6577, doi:10.1038/srep06577, 2014.

Table 1. (a) The effective climate sensitivities estimated by linear regression of $\log_2 \rho_{CO_2}$ against the temperature anomalies at monthly and annual resolutions from global and Northern Hemisphere series. The far right column shows the 20 year lagged sensitivity to (1900–2013), i.e. using $T_{\mathrm{anth},\Delta t}(t) = \lambda_{2\times CO_2,\mathrm{eff},\Delta t} \log_2(\rho_{CO_2}(t-\Delta t)/\rho_{CO_2,\mathrm{pre}})$ where $\Delta t = 20$ years.

Resolution		$\lambda 2 \times CO_2$, eff (K/doubling, no lag, 1880–2013)	$\lambda 2 \times CO_2$, eff (K/doubling, no lag, 1880–1998)	$\lambda 2 \times CO_2$, eff (K/doubling, no lag, 1880–1976)	$\lambda 2 \times CO_2$, eff (K/doubling, 20 yr lag, 1900–2013)
Monthly (dT_s)	Global	2.97 ± 0.08	2.92 ± 0.13	2.97 ± 0.25	4.29 ± 0.13
	Northern H.	3.41 ± 0.11	3.11 ± 0.17	3.10 ± 0.33	4.99 ± 0.18
Annual (LOTI)	Global	2.33 ± 0.16	2.26 ± 0.24	2.08 ± 0.48	3.73 ± 0.25
	Northern H.	2.56 ± 0.23	2.25 ± 0.34	2.41 ± 0.65	3.96 ± 0.38

Table 1. (b) The various SDs of the temperature residues (T_{nat}) after removing T_{anth} at monthly and annual resolution and the estimate of H obtained assuming perfect scaling over a factor of 12 in time scale, units (K). The results are close to the theory: $H = -0.20 \pm 0.03$.

	Monthly	Annual	$H = \log[\sigma_{T,\mathrm{yr}}/\sigma_{T,\mathrm{month}}]/\log 12$
Global	0.201	0.109	−0.24
Northern Hemisphere	0.273	0.155	−0.23

Table 2. A comparison of Root Mean Square (RMS) variances (data residues) and hindcast errors (deterministic and stochastic models) of global scale, annual temperatures. See also Fig. 2. Note that the GCM hindcasts are all "optimistic" in the sense that they use the observed volcanic and solar forcings and these would not be available for a true forecast. In comparison, the stochastic models forecast the responses to these (unknown) future forcings.

Row		1 year	5 year anomalies	9 year anomalies
	Temperature, residues			
1	Pre-industrial multiproxies (1500–1900)[a]	0.112	0.105	0.098
2	T_{nat}: Residues (1880–2013) (no lag with CO_2): $T_{anth}(t) \propto \log_2 \rho_{CO_2}(t)$	0.109	0.077[b]	0.070
3	$T_{nat,20}$: Residues from 1900–2013, 20 yr lag with CO_2: $T_{anth,\Delta t}(t) \propto \log_2 \rho_{CO_2}(t-\Delta t)$	0.108		
4	SD of the linearly detrended series 1880–2013 (residues, from a linear regression with the date)	0.163		
	Deterministic forecasts (GCM's)			
5	Without data assimilation 1983–2004 (Smith et al., 2007)	0.132	0.106	0.090
6	With data assimilation ("DePresSys") 1983–2004 (Smith et al., 2007)	0.105	0.066	0.046
7	CMIP3 simulations with bias and variance corrections 1983–2004 (Laepple et al., 2008)	0.106	0.059	0.044
8	GFDL CM2.1 (initialized yearly) cited in Newman (2013)	0.11		
9	CMIP5 multimodel ensemble (Doblas-Reyes et al., 2013) not initialized[c]		0.095	
10	CMIP5 multimodel ensemble (Doblas-Reyes et al., 2013) initialized		0.06	
	Stochastic forecasts			
11	LIM[d] (Newman, 2013)	0.085	(0.128)	(0.155)
12	Baillie and Chung (2002) ARFIMA[e]		0.132 ± 0.023	
13	Baillie and Chung (2002) AR(1)[f] forecast		0.156 ± 0.068	
14	SLIM (one parameter, Stochastic 1880–2013)[g]	0.093	0.071 (0.102)	0.067 (0.105)

[a] The average of the three multiproxies from Huang (2004), Moberg et al. (2005), Ammann and Wahl, (2007). These analyses were discussed in Lovejoy (2014a).
[b] The empirical 5 and 9 year anomaly values are close to the theoretical values $0.109 \times 5^{-0.2} = 0.079$ and $0.109 \times 9^{-0.2} = 0.070$.
[c] The results here are for a subset of the CMIP5 simulations that were run with and without annual data assimilation (initialization).
[d] Linear Inverse Modelling using, 20 eigenmodes, > 100 parameters. The errors in brackets are for the temperatures, not anomalies. Note that the 9 year LIM value is almost identical to the SD of the residues of the linear regression (fourth row of the table).
[e] ARFIMA = Autogressive Fractionally Integrated Moving Average process; this is close to the SLIM model used here. However the data and the data treatment were somewhat different. The annually, globally averaged temperatures from 1880 with a linear trend removed were used to make hindcasts over horizons of one to 10 years for the decades 1930, 1940, 1950, 1960. For each decade all the forecast errors were averaged. The value indicated here is the mean of the decade to decade mean error and the SD of that error, the errors cannot therefore be directly compared with the others. The data were from a series complied in 1986.
[f] AR(1) = AutoRegressive order 1, is equivalent to "enhanced persistence" in the preceding. The variance reduction when using ARFIMA instead of AR(1) is 29 %.
[g] The values in parentheses are for 1 year resolution temperatures.

Table 3. The hindcast SDs (in units of K) at the finest resolutions (1 month, 1 year) for natural variability temperatures obtained from the unlagged and 20 year lagged climate sensititivities. Note that the lag makes very little difference to the hindcast error variance.

Resolution		$\langle E_T(\tau,\tau)^2 \rangle^{1/2}$ No lag	$\langle E_T(\tau,\tau)^2 \rangle^{1/2}$ 20 yr lag
Monthly	Global	0.148	0.146
	Northern H.	0.214	0.209
Annual	Global	0.093	0.092
	Northern H.	0.132	0.133

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

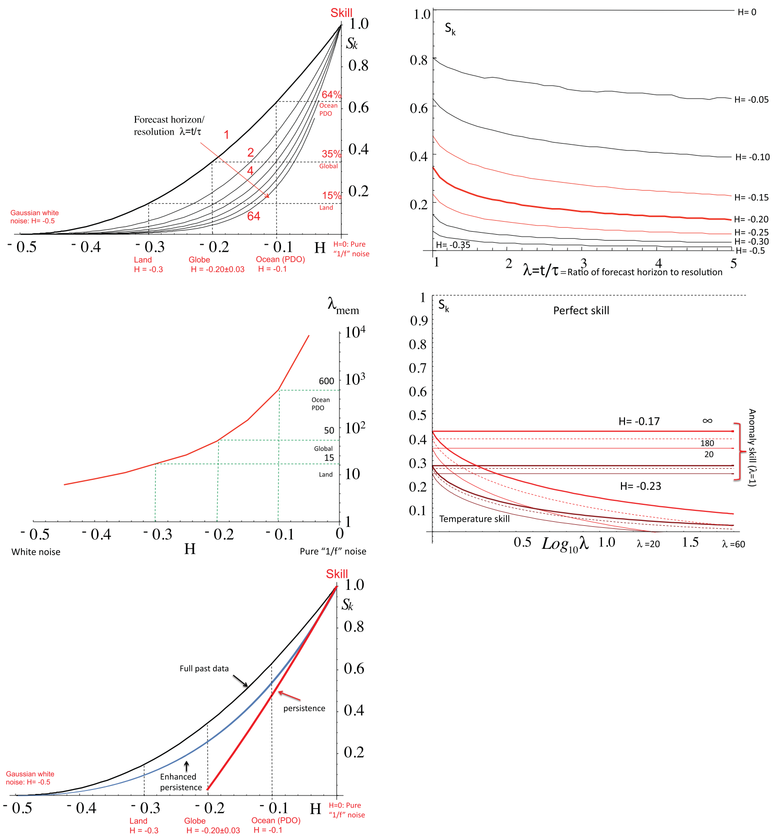

Figure 1. (**a**, upper left) Forecast skill for nondimensional forecast horizons λ = (horizon/resolution) = 1, 2, 4, 8, ..., 64 (left to right) as functions of H. For reference, the rough empirical values for land, ocean and the entire globe (the value used here, see below) are indicated by dashed vertical lines. The horizontal lines show the fraction of the variance explained (the skill, S_k, Eq. 47) in the case of a forecast of resolution τ data at a forecast horizon $t = \tau$ ($\lambda = 1$; corresponding to forecasting the anomaly fluctuation one time unit ahead). (**b**, upper right) The theoretical skill with infinite memory for various ratios of nondimensional forecast horizons λ over the range $0 > H > -0.35$ (top to bottom in steps of 0.05). The limiting value $H = -1/2$ corresponds to Gaussian white noise with zero skill. The empirically relevant range for the whole Earth ($H \approx -0.20 \pm 0.03$) is in red, thick the best estimated parameter ($H = -0.20$). (**c**, middle left) This illustrates the potentially huge memory in the climate system (especially the ocean). It gives the λ_{mem} value such that $S_{x,\lambda mem}(1)/S_{x,\infty}(1) = 0.9$. When $H = -1/2$, there is no memory and λ_{mem} is not defined, it diverges when $H \rightarrow 0$. (**d**, middle right) The theoretical skills for processes with infinite (Eq. 47) and finite memory (Eq. 49) for the empirically relevant parameter range ($H = -0.23$, brown, bottom, $H = -0.17$, red, top). The flat (constant skill) at the top are the curves for the anomaly forecasts (i.e. with forecast horizon = τ = resolution = τ so that $\lambda = 1$), the bottom curves are for constant resolution τ with forecast horizon. In each case a triplet of curves is shown corresponding to varying lengths of memories used in the forecast: infinite, 180 and 20 (the latter two corresponding to the those used for the monthly and global forecasts analysed here). (**e**, lower left) The skill of $\lambda = t/\tau = 1$ forecasts using the full memory (black, Eq. 47, from **a**), the corresponding classical persistence forecast (red), $S_k = 1-4(1-2^{2H})$ and the corresponding "enhanced persistence" result (blue; for this $\lambda = 1$ case, this is the same as the AR(1) model forecast) with $S_k = (2^{2H+1}-1)^2$. With classical persistence the skill becomes negative for $H \lesssim -0.2$, so it is not shown over the whole range.

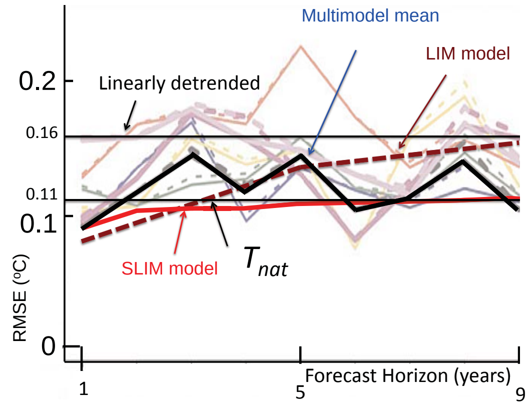

Figure 2. ENSEMBLES experiment, LIM and SLIM hindcasts for global annual temperatures for horizons 1 to 9 years. The light lines are from individual members of the ENSEMBLE experiment, the heavy line is the multimodel ensemble adapted from Fig. 4 in Garcia-Serrano and Doblas-Reyes (2012). This shows the RMSE comparisons for the global mean surface temperatures compared to NCEP/NCAR (2 m air temperatures). Horizontal reference lines indicate the SDs of T_{nat} (bottom), and of the linearly detrended temperatures (top). Also shown are the RMS error for the LIM model (from Table 1a, Newman, 2013) and the SLIM model (Sect. 4 below).

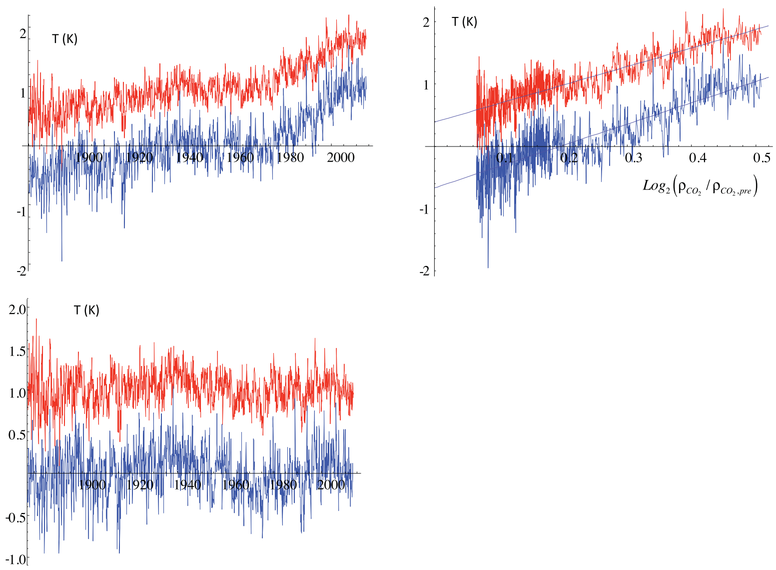

Figure 3. (**a**, upper left) The monthly surface temperature series from NASA GISS data (the monthly dTs series). Top (red) is the global average, displaced upward by 1 K for clarity, the bottom (blue) is the Northern Hemisphere series. (**b**, upper right) The same as **(a)** but for the temperatures as functions of the logarithm of the CO_2 concentration ρ_{CO_2} normalized by the preindustrial value $\rho_{CO_2,\mathrm{pre}}$ = 277 ppm (global values are displaced upward by 1 K for clarity). The regressions have slopes indicated in Table 1a, they are the effective climate sensitivities to CO_2 doubling. **(c)** The residues of the above; the estimate of the natural variability, again the global (red, top) has been shifted upward by 1 K for clarity.

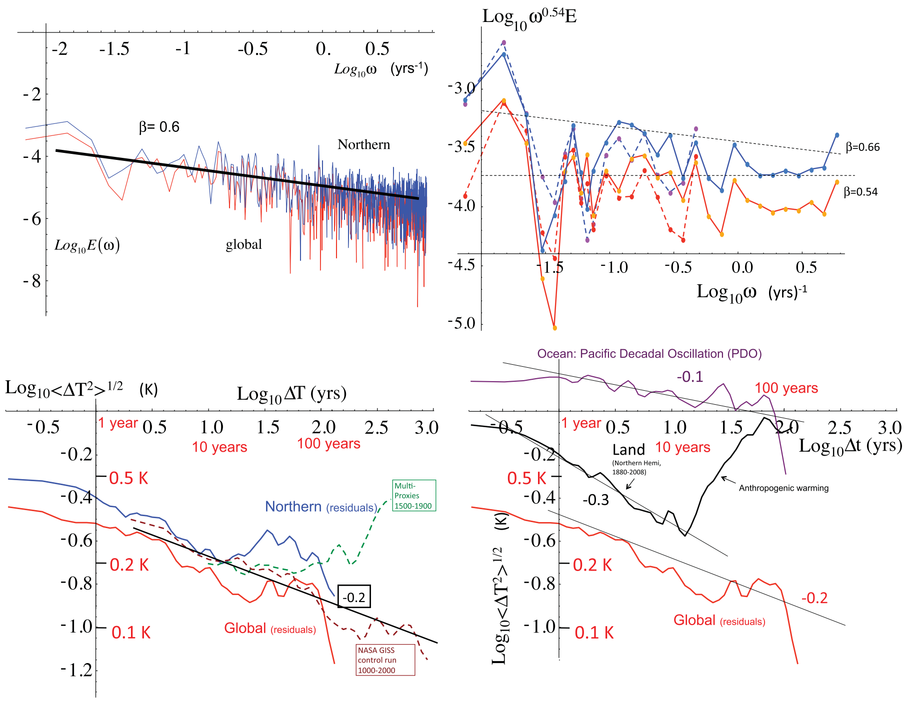

Figure 4. (**a**, upper left) The spectrum of the monthly residues for northern (blue) and global (red) data. The slope $\beta = 0.6$ is shown corresonding to the best overall estimate ($H = -0.20$). (**b**, upper right) The Northern Hemisphere (top, blue) and global (bottom, red) spectra, at monthly (solid) and annual (dashed) resolutions using the NASA GISS surface temperature series from 1880–2013. For frequencies higher than the lowest factor of ten, averages have been made over ten frequency bins per order of magnitude in scale. In addition, the spectra have been "compensated" by multiplying by $\omega^{0.54}$ so that spectra with $H = -0.23$ ($\beta = 0.54$) appear flat. The range $-0.17 < H < -0.23$ corresponding to one SD limits ($\beta = 1 + 2H$, i.e. ignoring small multifractal intermittency corrections) corresponds to $0.54 < \beta < 0.66$, the lower and upper bounding reference lines are shown as dashed. (**c**, lower left) The RMS Haar fluctuations for the northern (blue) and global (red) monthly series. Reference lines with slopes $H = -0.2$ are shown, we see that the scaling is fairly well respected up to ≈ 100 years. The raw Haar fluctuations have been multiplied by 2 (the "canonical calibration", see Lovejoy and Schertzer, 2012a) in order to bring them closer to the anomaly fluctuations. Also shown is the NASA control run and the pre-industrial multiproxies. They all agree quantitatvely very well up to about 100 years where the pre-industrial natural climate change starts to becomes important. This shows that the monthly scale residuals are almost exactly as simulated by the GISS model without any anthopogenic effects, supporting the idea that T_{nat} is a good estimate of the natural variability. (**d**, lower right) Comparisons of the RMS Haar fluctuations of global scale natural varilibty (T_{nat}) from **(c)**, with those from land only (HADCRUT3, black) and from the Pacific Decadal Oscillation (PDO, top, purple, from Lovejoy and Schertzer, 2013, Fig. 10.14). Reference lines of slopes $H = -0.1, -0.2, -0.3$ are shown close to the curves for ocean, globe and land respectively.

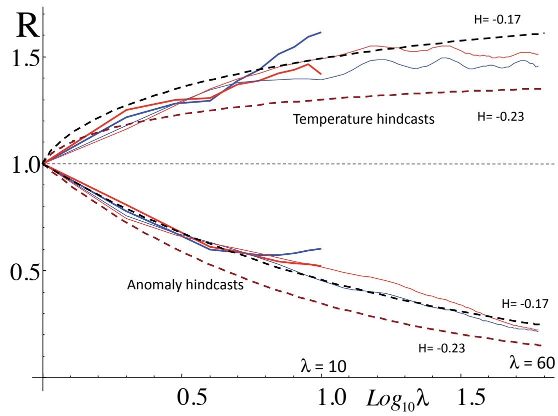

Figure 5. The dimensionless ratios (R) of the hindcast error variances to the variance at the smallest resolution and horizon (with horizon $\lambda\tau$, resolution τ (top, $R = \langle E_T(\lambda\tau,\tau)^2\rangle / \langle E_T(\tau,\tau)^2\rangle = 1 + (2+2H)F_H(\lambda)$) and anomaly, with horizon $\lambda\tau$, resolution $\lambda\tau$ (bottom, $R = \langle E_T(\lambda\tau,\lambda\tau)^2\rangle / \langle E_T(\tau,\tau)^2\rangle = \lambda^{2H}$). The red are global, the blue Northern Hemisphere, the thick, shorter curves are at annual resolution (τ = 1 yr) and the thin, longer lines are at monthly resolution (τ = 1 month). Also shown (dashed) are the theory curves for $H = -0.17$, -0.23 (top (black) and bottom (brown) of each dashed pair respectively). The data closely follow the H = -0.17 curves. The SDs of at the highest resolution $\langle E_T(\tau,\tau)^2\rangle^{1/2}$ are given in Table 3. This dimensionless plot has no adjustable parameters.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

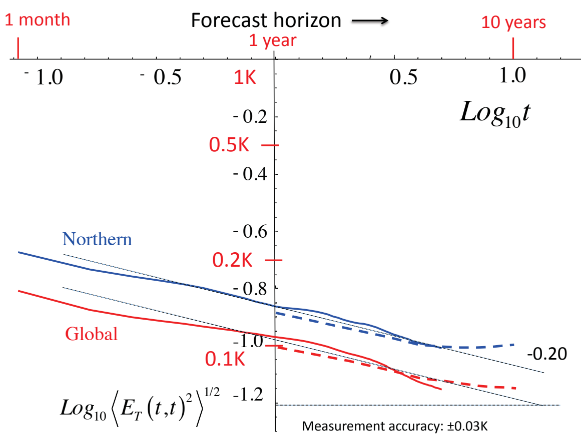

Figure 6. A log–log plot of the standard devations of the anomaly hindcasts with the theoretical reference line corresponding to $H = -0.20$. The solid lines are for the monthly data, the dashed lines for annual data, red for global, blue for Northern Hemisphere.

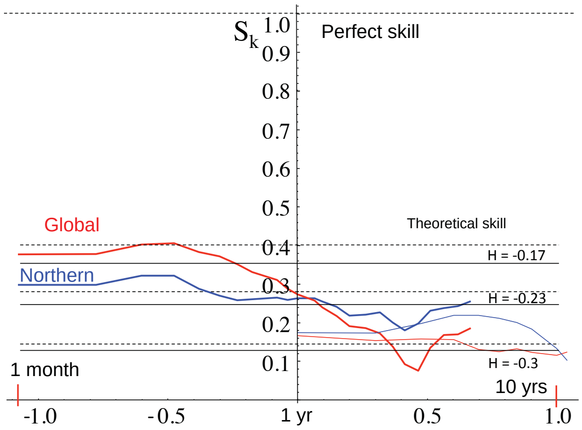

Figure 7. The anomaly forecast skill on a log-linear plot for both all series (annual thin, monthly thick, global red, Northern Hemisphere, blue). Also shown are pairs of theoretical predictions (constant skill independent of the forecast horizon) for various values of H, the top (dashed) member of the pair is for an infinite memory, the bottom solid line is for the finite memory used here: the monthly series has a memory of 180, the annual series has 20. This plot has no adjustable parameters.

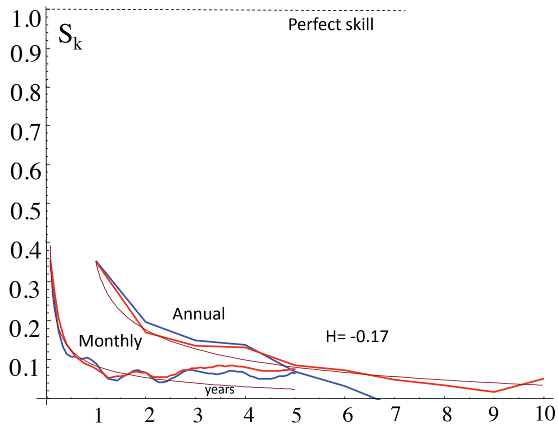

Figure 8. The forecast skill for the temperature at fixed resolutions (one month, bottom left, one year, upper right) for global (red) and Northern Hemisphere (blue) series. Also shown are the exact theoretical curves (for $H = -0.17$) that take into account the finite memories of the forecasts (20 years, 15 years annual, monthly series respectively). The raw curves were shifted a little upward so that their long-time parts were close to the theory; this is equivalent to using the theory to improve the estimate of the ensemble average skill from the single series that were available.

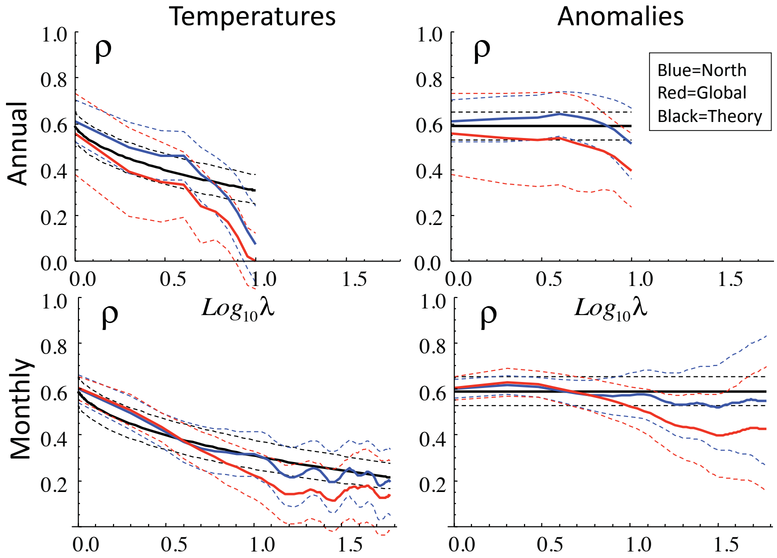

Figure 9. The nondimensional empirical correlations of the forecast temperatures (left column) and anomalies (right column), the same data as previous but with different empirical comparisons and also with comparisons with theory for $H = -0.2$ (thick black), $H = -0.17$, -0.23 top and bottom dashed black. Now note that in all cases the one SD bounds (dashed) on the empirical and theoretical curves overlap virtually throughout. The theory curves have no adjustable parameters.