This is a pre-proof version of Lovejoy, S., 2014: Scaling fluctuation analysis and statistical hypothesis testing of anthropogenic warming, **Climate Dyn**. DOI 10.1007/s00382-014-2128-2.

Scaling fluctuation analysis and statistical hypothesis testing of anthropogenic warming

S. Lovejoy,

Physics, 3600 University st.,

Montreal, Que. H3A 2T8

Canada

Email: lovejoy@physics.mcgill.ca

Tel: 514-398-6537

Fax: 514-398-6482

Abstract

Although current global warming may have a large anthropogenic component, its quantification relies primarily on complex General Circulation Models (GCM's) assumptions and codes; it is desirable to complement this with empirically based methodologies. Previous attempts to use the recent climate record have concentrated on "fingerprinting" or otherwise comparing the record with GCM outputs. By using CO_2 radiative forcings as a linear surrogate for all anthropogenic effects we estimate the total anthropogenic warming and (effective) climate sensitivity finding: $\Delta T_{anth} = 0.87 \pm 0.11$ K,

$\lambda_{2x,CO2,eff}$ = 3.08 ±0.58 K. These are close the IPPC AR4, AR5 values ΔT_{anth} = 0.74±0.18 K and $\lambda_{2x,CO2}$ = 1.5 - 4.5 K (equilibrium) climate sensitivity and are independent of GCM models, radiative transfer calculations and emission histories. We statistically formulate the hypothesis of warming through natural variability by using centennial scale probabilities of natural fluctuations estimated using scaling, fluctuation analysis on multiproxy data. We take into account two nonclassical statistical features - long range statistical dependencies and "fat tailed" probability distributions (both of which greatly amplify the probability of extremes). Even in the most unfavourable cases, we may reject the natural variability hypothesis at confidence levels > 99%.

1. Introduction

Well before the advent of General Circulation Models (GCM's), [*Arrhenius*, 1896], proposed that greenhouse gases could cause global warming and he even made a surprisingly modern quantitative prediction. Today, GCM's are so much the dominant tool for investigating the climate that debate centers on the climate sensitivity to a doubling of the CO_2 concentration which - whether "equilibrium" or "transient" - is defined as a purely theoretical quantity being accessible only through models. Strictly speaking - short of a controlled multicentennial global scale experiment - it cannot be empirically measured at all. A consequence is that not enough attention has been paid to directly analyzing our ongoing uncontrolled experiment. For example, when attempts are made to test climate sensitivity predictions from the climate record, the tests still rely on GCM defined "fingerprints" (e.g. [*Santer et al.*, 2013] or the review in section 9.2.2 of 4th Assessment Report (AR4) of the International Panel on Climate Change (IPCC) or on other comparisons of the record with GCM outputs (e.g. [*Wigley et al.*, 1997], [*Foster and Rahmstorf*, 2011]). This situation can easily lead to the impression that complex GCM codes are indispensible for inferring connections between greenhouse gases and global warming. An unfortunate side effect of this reliance on models is that it allows GCM skeptics to bring into question the anthropogenic causation of the warming. If only for these reasons, it is desirable to complement model based approaches with empirically based methodologies.

But there is yet another reason for seeking non-GCM approaches: the most convincing demonstration of anthropogenic warming has not yet been made – the statistical comparison of the observed warming during the industrial epoch against the null hypothesis for natural variability. To be as rigorous as possible, we must demonstrate that the probability that the current warming is no more than a natural fluctuation is so low that the natural variability may be rejected with high levels of confidence. Although the rejection of natural variability hypothesis would not "prove" anthropogenic causation, it would certainly enhance it's credibility. Until this is done, there will remain some legitimate grounds for doubting the anthropogenic provenance of the warming. Such statistical testing requires knowledge of the probability distributions of natural fluctuations over roughly centennial scales (i.e. the duration of the industrial epoch). To achieve this using GCM's one would need to construct a statistical ensemble of realistic pre-industrial climates at centennial scales. Unfortunately the GCM variability at these (and longer) scales under natural (especially solar and volcanic) forcings is still the object of active research (e.g. "Millennium" simulations). At present, the variability at these long time scales is apparently somewhat underestimated ([*Lovejoy*, 2013]) so that it is premature to use GCM's for this purpose. Indeed, at the moment, the only way of estimating the centennial scale natural variability is to use observations (via multicentennial length multiproxies) and a (modest) use of scaling ideas.

The purpose of this paper is thus to establish an empirically based GCM-free methodology for quantifying anthropogenic warming. This involves two parts. The first part is to estimate both the total amplitude of the anthropogenic warming and the (empirically accessible) "effective" climate sensitivity. It is perhaps surprising that this is apparently the first time that the latter has been directly and simply estimated from surface temperature data. Two innovations were needed. First, we used a stochastic approach that combines all the (nonlinear) responses to natural forcings as well as the (natural) internal nonlinear variability into a single global stochastic quantity $T_{nat}(t)$ that thus takes into account all the natural variability. In contrast, the anthropogenic warming ($T_{anth}(t)$) is treated as deterministic. The second innovation is to use the CO_2 radiative forcing as a surrogate for all anthropogenic forcings. This includes not only the relatively well understood warmings due to the other long lived Green House Gases (GHG's) but also the poorly understood cooling due to aerosols. The use of the CO_2 forcing as a broad surrogate is justified by the common dependence (and high correlations) between the various anthropogenic effects due to their mutual dependencies on global economic activity (see fig. 2 a, b below).

The method employed in the first part (section 2) leads to conclusions not very different from those obtained from GCM's and other approaches. In contrast, the main part of the paper (section 3), outlines the first attempt to statistically test the null hypothesis using the statistics of centennial scale natural fluctuations estimated from pre-industrial multiproxies. To make the statistical test strong enough, we use scaling ideas to parametrically bound the tails of the extreme fluctuations using extreme ("fat-tailed", power law) probability distributions and we scale up the observed distributions from 64 to 125 years using a scaling assumption. Even in the most unfavourable cases, we may reject the natural variability hypothesis at confidence levels > 99%. These conclusions are robust because they take into account two nonclassical statistical features which greatly amplify the probability of extremes - long range statistical dependencies and the fat tails.

2. A stochastic approach:

2.1 A simple stochastic hypothesis about the warming

Within the scientific community, there is a general consensus that in the recent epoch (here, since 1880) that anthropogenic radiative forcings have dominated natural ones so that solar and volcanic forcings and changes in land use are relatively unimportant in explaining the overall warming. This conclusion applies to centennial scales but by using fluctuation analysis on global temperatures it can be extended to somewhat shorter time scales (≈20-30 years for the global average temperature [*Lovejoy et al.*, 2013b]).

Let us therefore make the hypothesis that anthropogenic forcings are indeed dominant (skeptics may be assured that this hypothesis will be tested and indeed quantified in the following analysis). If this is true, then it is plausible that they do not significantly affect the type or amplitude of the natural variability so that a simple model may suffice:

$$T_{globe}(t) = T_{anth}(t) + T_{nat}(t) + \varepsilon(t) \tag{1}$$

T_{globe} is the measured mean global temperature anomaly, T_{anth} is the deterministic anthropogenic contribution, T_{nat} is the (stochastic) natural variability (including the responses to the natural forcings) and ε is the measurement error. The latter can be estimated from the differences between the various observed global series and their means; it is nearly independent of time scale [*Lovejoy et al.*, 2013a] and sufficiently small (≈ ±0.03 K) that we ignore it.

While eq. 1 appears straightforward, it requires a few comments. The first point is that the anthropogenic contribution $T_{anth}(t)$ is taken to be deterministic whereas the natural variability $T_{nat}(t)$ is assumed to be stochastic. The second point is that this definition of $T_{nat}(t)$ includes the responses to both volcanic, solar and any other natural forcings so that $T_{nat}(t)$ does *not* represent pure "internal" variability. While at first sight this may seem reasonable, it is actually quite different from the usual treatments of solar and volcanic forcings and the corresponding responses which are deterministic and where stochasticity is restricted to ("pure") internal variability (see e.g. [*Lean and Rind*, 2008]). One of the reasons for the classical approach is that there is enough data to allow one to make "reconstructions" of past forcings. If they can be trusted, these hybrid model - data products allow GCM's to model and isolate the corresponding responses. However, we suspect that another reason for these deterministic treatments – especially in the case of volcanic forcing – is that the intermittency of the process is so large that it is often assumed that the generating process could not be stationary. If it were true that solar and volcanic processes were nonstationary then their statistics would have to be specified as functions of time. In this case, little would be gained by lumping them in with the internal variability which even in the presence of large anthropogenic forcing - is quite plausibly stationary since – as assumed in GCM climate modelling –the effect of anthropogenic forcings is essentially to change the boundary conditions but not the internal dynamics.

However, it is quite likely that the basic solar and terrestrial stochastic processes responsible for variable solar output and volcanic activity are unchanged over the last millennium, yet that the corresponding stochastic realizations of these processes are highly intermittent, scaling and multifractal giving a spurious appearance of nonstationarity (multifractals have nonclassical scaling behviours: unlike quasi-Gaussian processes, each statistical moment is characterized by a different exponent and there are strong resolution dependencies). While the basic analyses were presented in [*Lovejoy and Schertzer*, 2012c] we revisit and reanalyze them here. Consider fig. 1a which shows the [*Gao et al.*, 2008] volcanic reconstruction from 500 – 2000 A.D. along with three realizations of a multifractal process with identical statistical parameters (estimated by the analysis of the reconstructions in [*Lovejoy and Schertzer*, 2012c]), calibrated so that the overall process (but not each realization!) has the observed mean. It is very hard to distinguish the reconstruction from the three independent realizations. Since by construction, the multifractal process is stationary, this strongly supports the hypothesis that the mechanism behind terrestrial volcanism during the last 1500 years has not changed. Similar conclusions apply to the solar output (excluding the 11 year cycle) although - since its intermittency is much smaller- this is perhaps less surprising. Further support for this comes from the fluctuation analysis in fig. 1b which compares the RMS fluctuations of the reconstruction over the (mostly) pre-industrial period 1500-1900 and the industrial period 1880-2000 with the RMS fluctuations of the corresponding multifractal simulations. We see that although the amplitude of the industrial period fluctuations is a factor ≈ 2 lower than for the pre-industrial period, that this is well within what is expected due to the (very high) natural variability of volcanic processes (note that the fluctuations isolate the variability as a function of time scale, they are independent of the absolute level of the forcing; for more analysis, see [*Lovejoy and Schertzer*, 2012c] and [*Lovejoy et al.*, 2014]). Finally, fig. 1c shows the corresponding analyses for the volcanic reconstruction as well as two solar reconstructions, with the same basic conclusions: they may all be considered stationary and there is nothing unusual about the statistics in the recent epoch when compared to the pre-industrial epoch. In any event, we shall see below that eq. 1 can be justified ex-post-facto by empirically estimating T_{nat} and verifying directly that it has the same industrial and pre industrial statistics.

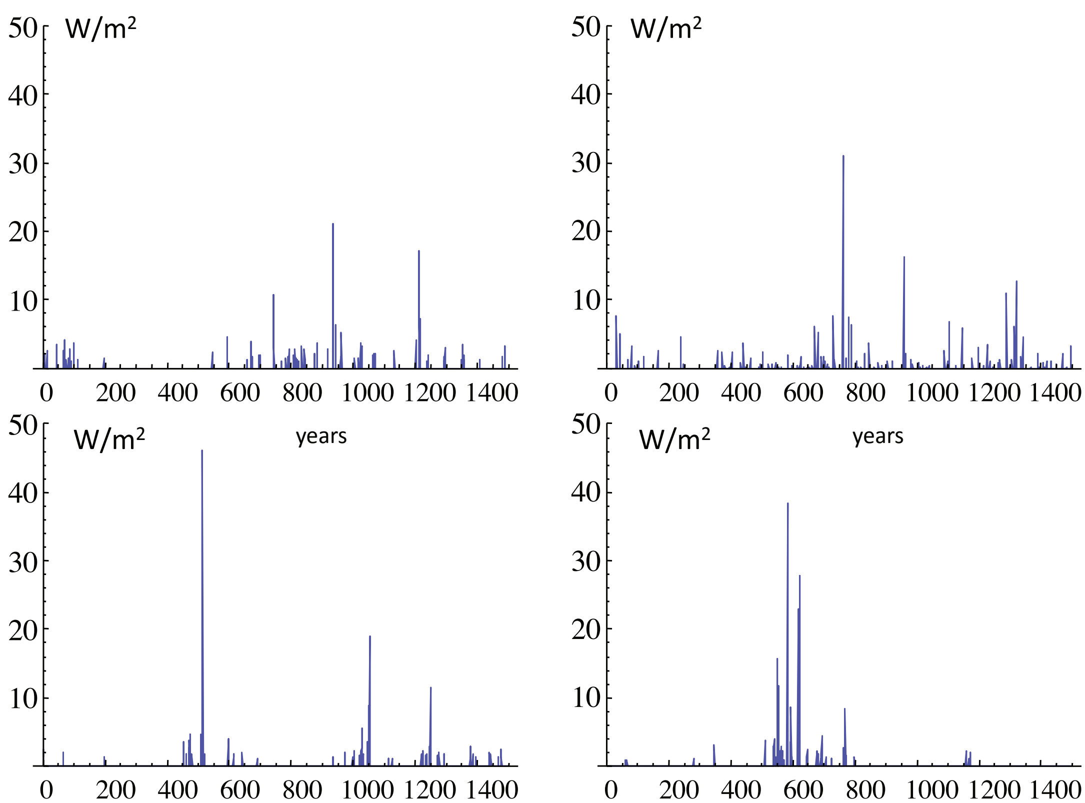

Fig. 1a: The 1500 year [*Gao et al.*, 2008] volcanic reconstruction of the radiative forcing (over the period 500 – 2000 A.D.) along with three multifractal simulations with

the measured parameters (C_1 = 0.2, H = -0.3, α = 1.8; parameters estimated in [*Lovejoy and Schertzer*, 2012c]). The simulations differed only by their random seeds and were calibrated to have the same average forcing value (0.15 W/m^2). The fact that the reconstruction is essentially indistinguishable from these statistically stationary multifractal simulations strongly supports the hypothesis that the basic volcanism responsible for eruptions over this period is constant. The reconstruction is in the upper right, the others are "fakes".

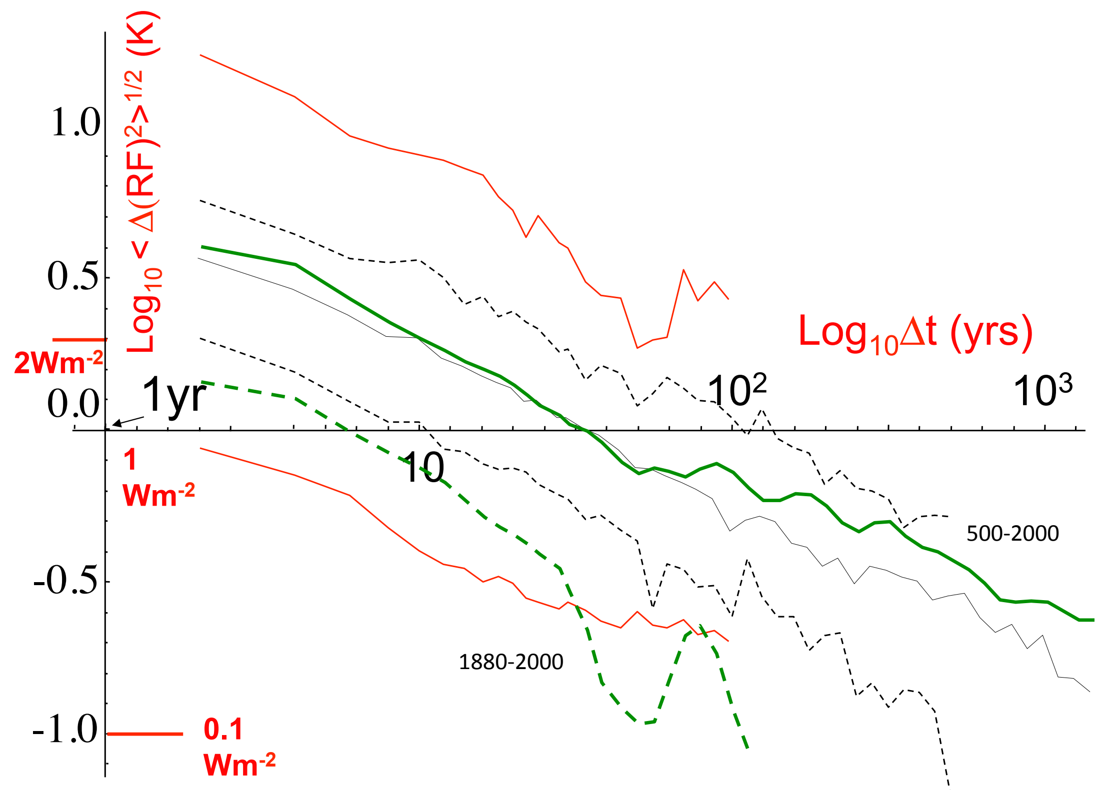

Fig. 1b: The RMS fluctuations for the [*Gao et al.*, 2008] reconstruction (green, thick) for the period 500-2000 (solid) and 1880-2000 (dashed; see fig. 1c for the slightly different curve for the period 1500-1900). The fluctuations over a lag Δt are defined by the difference of the average over the first and second halves of the interval ("Haar" fluctuations, see section 3.1). Also shown is the ensemble average (thin black line) of ten realizations of the multifractal process with the fig. 1a parameters. The thin dashed black lines indicate the one standard deviation bounds of the log of the RMS fluctuations estimated from the realization to realization variability for 500 year simulated segments. The thin red lines are for the bounds for 100 year segments (they are wider since the variability is less averaged out than for the 500 year bounds).

The wide bounds indicated by the one standard deviation limits show that the variability of the process is so large that in spite of the fact that the RMS amplitude of the volcanic forcing over the industrial period is roughly a factor ≈ 2 lower than over the pre-industrial period (compare the dashed and solid green lines), that it is nevertheless generally within the one standard deviation bounds (red) of the stochastic multifractal process (i.e. the dashed green line generally lies between the thin red lines).

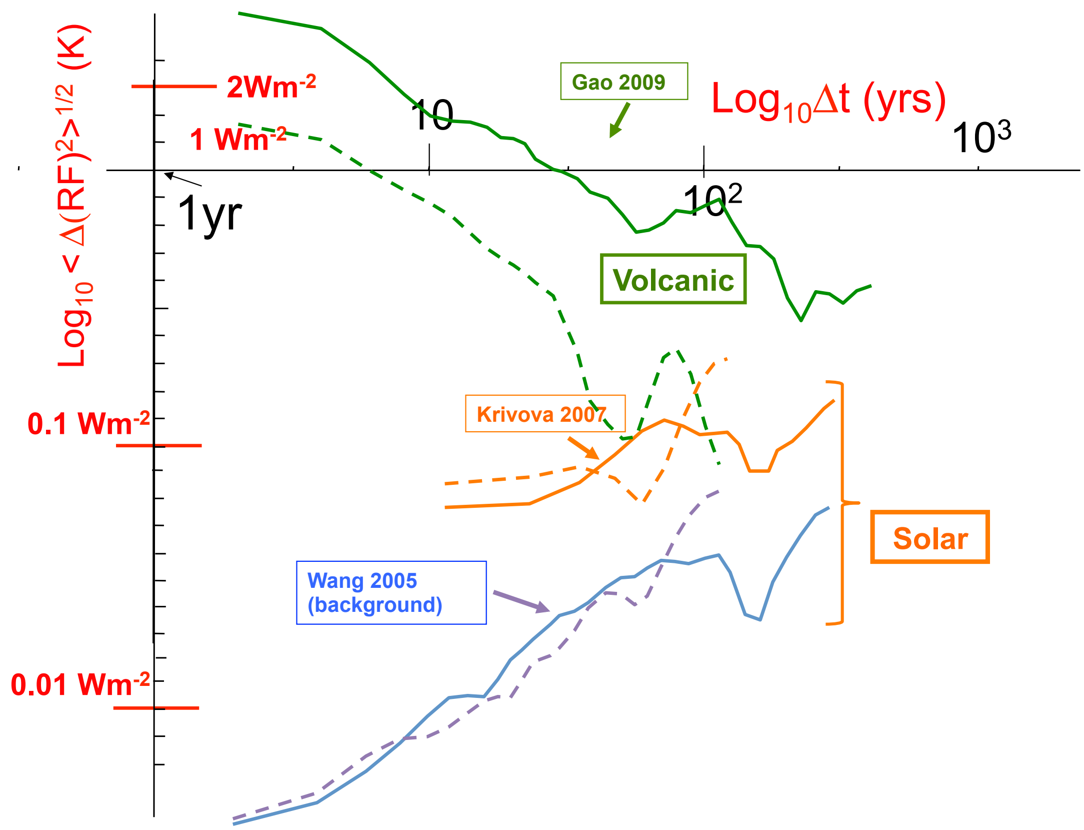

Fig. 1c: The RMS radiative forcing fluctuations for the [*Gao et al.*, 2008], volcanic reconstruction (since 1500) as well as the same from sunspot based solar reconstructions [*Wang et al.*, 2005], [*Krivova et al.*, 2007] (from 1610). The full lines are for the period up to 1900, the dashed lines for the period since 1880. One can see that the industrial and preindustrial solar fluctuations are of nearly the same. In contrast, the amplitude of the volcanic forcing fluctuations have decreased by a factor ≈2 in the recent period (note that this does not imply a change in the amplitude of the forcing itself). For a more complete analysis of the fluctuations over the whole period, see [*Lovejoy and Schertzer*, 2012c].

2.2 CO_2 radiative forcing as a linear surrogate for anthropogenic effects

The first step in testing eq. 1 is to empirically estimate T_{anth}. The main contribution is from CO_2, for which there are fairly reliable reconstructions from 1880 as well as from reliable in situ measurements from Mauna Loa and Antarctica from 1959. In addition, there is general agreement about its radiative forcing (R_F) as a function of concentration ρ_{CO2}:

$$R_{F,CO_2} = R_{F,2xCO_2} \log_2\left(\rho_{CO_2} / \rho_{CO_2,pre}\right); \quad R_{F,2xCO_2} = 3.7W/m^2 \quad ; \rho_{CO_2,pre} = 277\,ppm \qquad (2)$$

where $R_{F,2xCO2}$ is the forcing for CO_2 doubling; the basic logarithmic form is a semi-analytic result from radiative transfer models, the values of the parameters are from the AR4. Beyond CO_2, the main other anthropogenic forcings are from other long-lived greenhouse gases (warming) as well as the effect of aerosols (cooling). While the reconstruction of the global GHG forcing since 1880 is reasonably well estimated, that is not the case for aerosols which are short lived, poorly mixed (regionally concentrated), and whose effects (especially the indirect ones) are poorly understood (see below).

However, all the key anthropogenic effects are functions of economic activity, the CO_2 levels provide a convenient surrogate for the latter (over the period 1880-2004, $\log_2\rho_{CO2}$ varies by only ≈0.5 – half an octave in ρ_{CO2} - so that ρ_{CO2} and $\log_2\rho_{CO2}$ are linear to within ±1.5% and there is not so much difference between using ρ_{CO2} or $R_{F,CO2}$ as a surrogate). The strong connection with the economy can be seen using the recent [*Frank et al.*, 2010] CO_2 reconstruction from 1880-2004 to estimate $\log_2\left(\rho_{CO_2} / \rho_{CO_2,pre}\right)$, fig. 2a shows its correlation with the global Gross Domestic Product (GDP; correlation coefficient $r_{RFCO2,GDP}$ = 0.963). Also shown is the annual global production of sulfates which is a proxy for the total (mostly sulfate) aerosol production. The high correlation coefficient ($r_{RFCO2,sulfate}$ = 0.983) indicates that whatever cooling effect the aerosols have, that they are likely to be roughly linear in $\log_2\left(\rho_{CO_2} / \rho_{CO_2,pre}\right)$. Also shown in the figure (using data from [*Myhre et al.*, 2001]), is the total forcing of all GHG's (including CO_2); we find the very high correlation $r_{RFCO2RF,GHG}$ = 0.997. This justifies the simple strategy adopted here of considering $R_{F,CO2}$ to be a well measured linear surrogate for $R_{F,anth}$ (i.e. the two are considered to be equal to within a constant factor).

Concentrating on the total GHG radiative forcing ($R_{F,GHG}$) as well as the total anthropogenic RF (including aerosols, $R_{F,anth}$) we present fig. 2b. We see that $R_{F,\mathrm{CO2}}$ and $R_{F,\mathrm{GHG}}$ are closely related with regressions yielding:

$$R_{F,GHG} = -0.190 \pm 0.019 + (1.793 \pm 0.027) R_{F,CO_2} \tag{3}$$

(as in fig. 2a, $r_{RFCO2RF,GHG}$ = 0.997) so that $R_{F,\mathrm{CO2}}$ may be considered "enhanced" by the other GHG by ≈ 79%. Although ozone, biomass and other effects contribute, the main additional contribution – and uncertainty - in the total anthropogenic $R_{F,anth}$, is from the direct and indirect cooling effects of aerosols, and is still under debate. Recent estimates (for both effects) are ≈ -1.2 (AR4), -1.0 W/m^2, [*Myhre*, 2009] and ≈ -0.6 W/m^2, [*Bauer and Menon*, 2012] (all with large uncertainties). Using the Mauna Loa estimate for ρ_{CO2} in 2012 (393.8 ppm, http://co2now.org/), these estimates can be compared to ≈ 1.9 W/m^2 for CO_2 and ≈ 3.1 W/m^2 for all GHG (the above relation). Using the $R_{F,anth}$ data in [*Myhre et al.*, 2001] we obtain:

$$R_{F,anth} = 0.034 \pm 0.033 + (0.645 \pm 0.048) R_{F,CO_2} \tag{4}$$

with $r_{CO2,anth}$ = 0.944 (fig. 2b). This is tantamount to assuming -1.5 W/m^2 for aerosol cooling at the end of the [*Myhre et al.*, 2001] series (1995). If the most recent cooling estimates (*Bauer and Menon*, 2012]) are correct (-0.6 W/m^2), the amplitude of the cooling is diminished by 60%, so that in eq. 4 we obtain a proportionality constant ≈ 1.25 rather than 0.645.

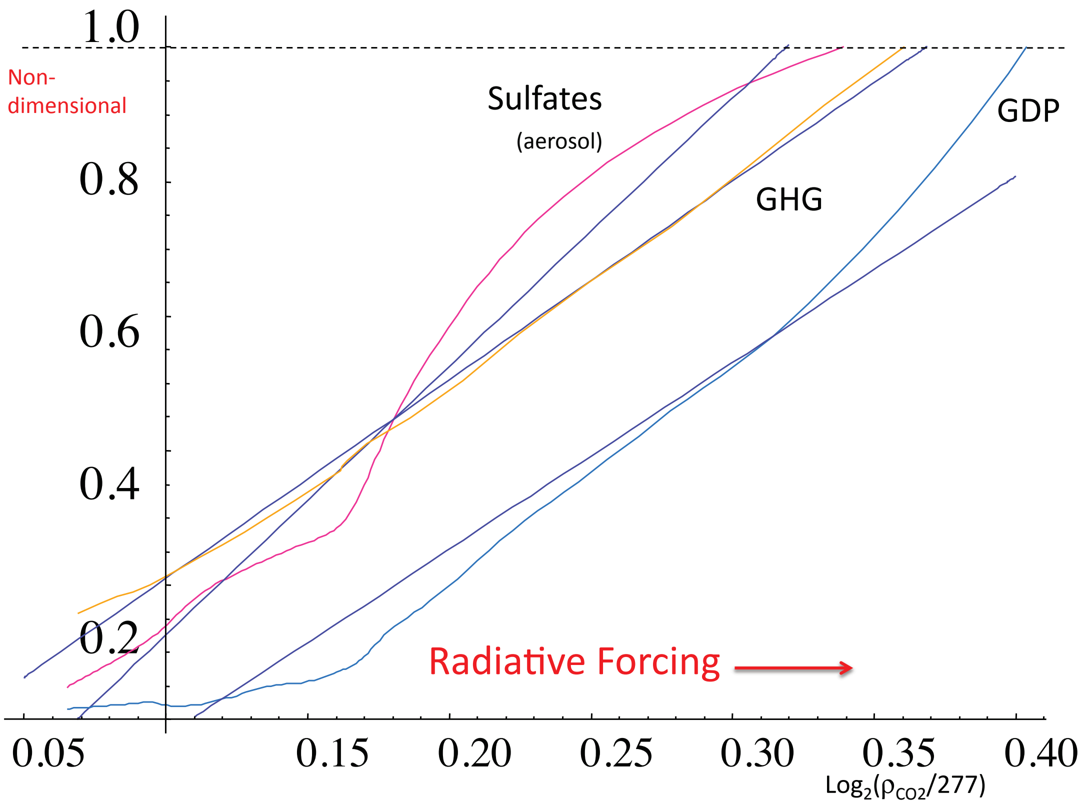

Fig. 2a: This shows the annual world sulfate aerosol production from 1880-1990 (top, pink, from [*Smith et al*., 2004]), the total Greenhouse Gas radiative forcing from 1880-1995 (orange, from [*Myhre et al*., 2001], including CO_2), and the world Gross Domestic Product (GDP, 1880-2000, blue, from J. Bradford DeLong of the Department of Economics, U.C. Berkeley: http://holtz.org/Library/Social%20Science/Economics/Estimating%20World%20GDP%20by%20DeLong/Estimating%20World%20GDP.htm) all nondimensionalized by their maximum values ($6.9x10^7$ metric tons/yr, 2.29 W/m^2, \$$4.1x10^{13}$ respectively). The regression lines have slopes corresponding to an increase of $2.8x10^8$ metric tons of sulfate for each CO_2 doubling, and an increase of GHG forcing by 6.63 W/m^2 for each CO_2 doubling, an increase of GDP by \$$1.1x10^{14}$ for every CO_2 doubling. The correlation coefficients are 0.983, 0.997, 0.963 for sulfate production, total GHG forcing and GDP respectively.

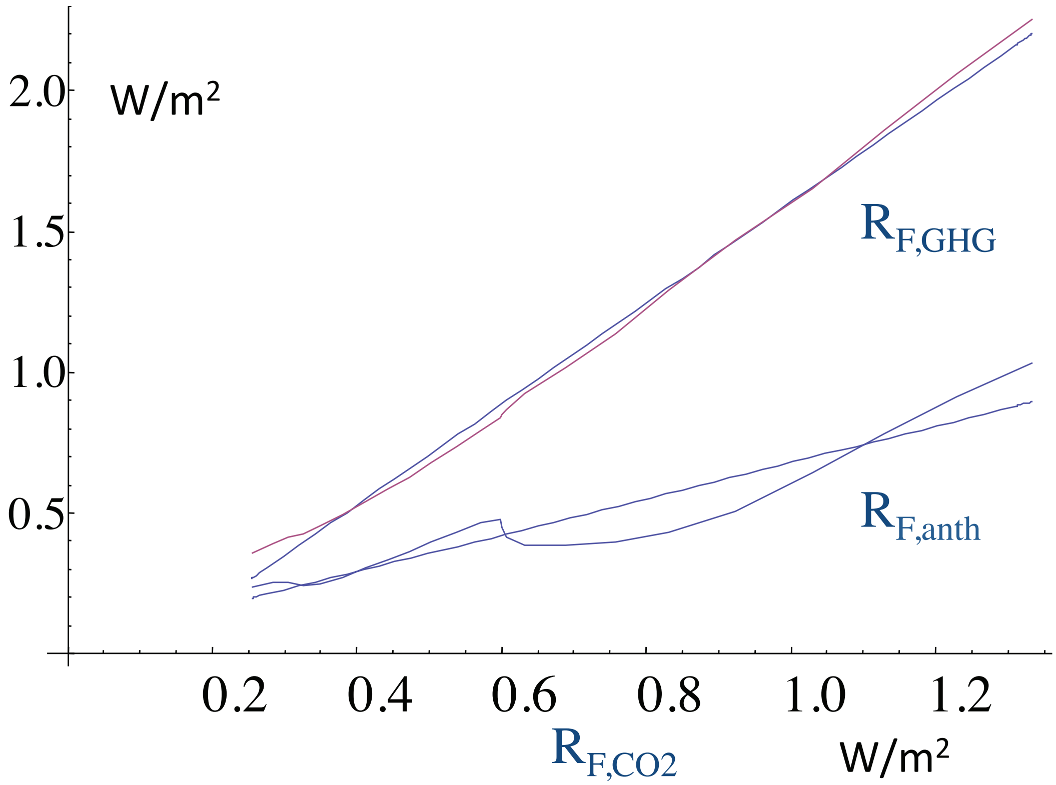

Fig. 2b: Over the period 1880-1995, the relationship between the radiative forcing of CO_2 ($R_{F,CO2}$), the radiative forcing of all the long lived Greenhouse Gases (including CO_2: $R_{F,GHG}$) and the total radiative forcing of all the anthropogenic emission including aerosols; data from [*Myhre et al.*, 2001]. For reference, current (2012) R_{FCO2} is estimated as ≈1.9 W/m^2. The slopes and correlation coefficients are: 1.79 and 0.997 (top) and 0.645 and 0.944 (bottom).

2.3 The instrumental data and the effective climate sensitivity

If we take $R_{F,CO2}$ to be a well-measured linear surrogate for $R_{F,anth}$ (i.e. $T_{anth} \propto R_{F,CO_2}$) we can define the "effective" climate sensitivity λ to a doubling of CO_2 by:

$$T_{anth}(t) = \lambda_{2xCO2,eff} \log_2\left(\rho_{CO_2}(t)/\rho_{CO_2,pre}\right) \tag{5}$$

In order to empirically test eq. 1, it therefore suffices to perform a regression of $T_{globe}(t)$ against $\log_2\left(\rho_{CO_2}(t)/\rho_{CO_2,pre}\right)$; the slope yields $\lambda_{2xCO2,eff}$ and the residues $T_{nat}(t)$. As mentioned above, empirical estimates of the annually, globally averaged surface temperatures do not perfectly agree with each other, the differences between the series may be used to quantify the uncertainty in the estimates. For example, in this analysis, we used data over the period 1880 – 2008 from three sources: the NOAA NCDC (National Climatic Data Center) merged land, air and sea surface temperature dataset (abbreviated NOAA NCDC below), on a 5^{o}x5^{o} grid [*Smith et al.*, 2008], the NASA GISS (Goddard Institute for Space Studies) dataset [*Hansen et al.*, 2010] (from 1880 on a 2^{o}x 2^{o}) and the HadCRUT3 dataset [*Rayner et al.*, 2006] (on a 5^{o}x5^{o} grid), and as mentioned earlier, these series only agree to within about ±0.03 K even at centennial scales. There are several reasons for the differences: HadCRUT3 is a merged product created out of the HadSST2 Sea Surface Temperature (SST) dataset and its companion dataset of atmospheric temperatures over land, CRUTEM3 [*Brohan et al.*, 2006]. Both the NOAA NCDC and the NASA GISS data were taken from http://www.esrl. noaa.gov/psd/; the others from http://www.cru.uea. ac.uk/cru/data/temperature/. The NOAA NCDC and NASA GISS are both heavily based on the Global Historical Climatology Network [*Peterson and Vose*, 1997], and have many similarities including the use of sophisticated statistical methods to smooth and reduce noise. In contrast, the HadCRUT3 data are less processed, with corresponding advantages and disadvantages. Analysis of the space-time densities of the measurements shows that they are sparse (scaling) in both space and time [*Lovejoy and Schertzer*, 2013]. Even without other differences between the data sets, this strong sparseness means that we should not be surprised that the resulting global series are somewhat dependent on the assumptions about missing data.

The mean and standard deviation of the $T_{globe}(t)$ series is shown in fig. 3a as functions of $\log_2\left(\rho_{CO_2}(t)/\rho_{CO_2,pre}\right)$; the result is indeed quite linear with slope equal to the effective climate sensitivity to CO_2 doubling. We find:

$$\lambda_{2x,CO2,eff} = 2.33\pm0.22 \text{ K} \qquad (6)$$

(note that for the northern hemisphere only, $\lambda_{2x,CO2,eff} = 2.59\pm0.25$ K so that hemispheric differences are not very large). For 5 year averages for 1880-2004 (the CO_2 from the reconstruction) and 1959-2004 (using the mean of the instrumental Mauna Loa and Antarctica CO_2), the correlation coefficients are respectively $r_{RFCO2,T}$ = 0.920, 0.968. Note that this simple direct estimate of λ_{2xCO2} can be compared with several fairly similar but more complex analyses (notably multiple regressions which include CO_2), see [*Lean and Rind*, 2008], [*Muller et al.*, 2013]. By use of the proportionality constants between $R_{F,anth}$ and $R_{F,CO2}$ we can estimate the effects of a pure CO_2 doubling. For the strongly cooling aerosols ([*Myhre et al.*, 2001]) we obtained 0.645 (eq. 4) whereas for the weakly cooling [*Bauer and Menon*, 2012], aerosols we obtained 1.25. These lead to the pure CO_2 doubling estimates $\lambda_{2x,CO2,pure}$ = 3.61±0.34 and 1.86±0.18 K respectively.

If we plot the temperatures in the usual way as functions of time, we obtain figs. 3b, c where we also show the anthropogenic contribution estimated with $\lambda_{2x,CO2,eff}$ from eq. 6 and T_{anth} from eq. 5. It follows the temperatures very well, and we can already see that the residues ($T_{nat}(t)$) are fairly small. Using these estimates of the anthropogenic contribution, we can estimate the total change in temperature as T_{anth} =0.85±0.08 over the entire industrial period (see the discussion below). Note that the same methodology can be used to analyze the postwar cooling and the recent "pause" in the warming; this is the subject of current work in progress.

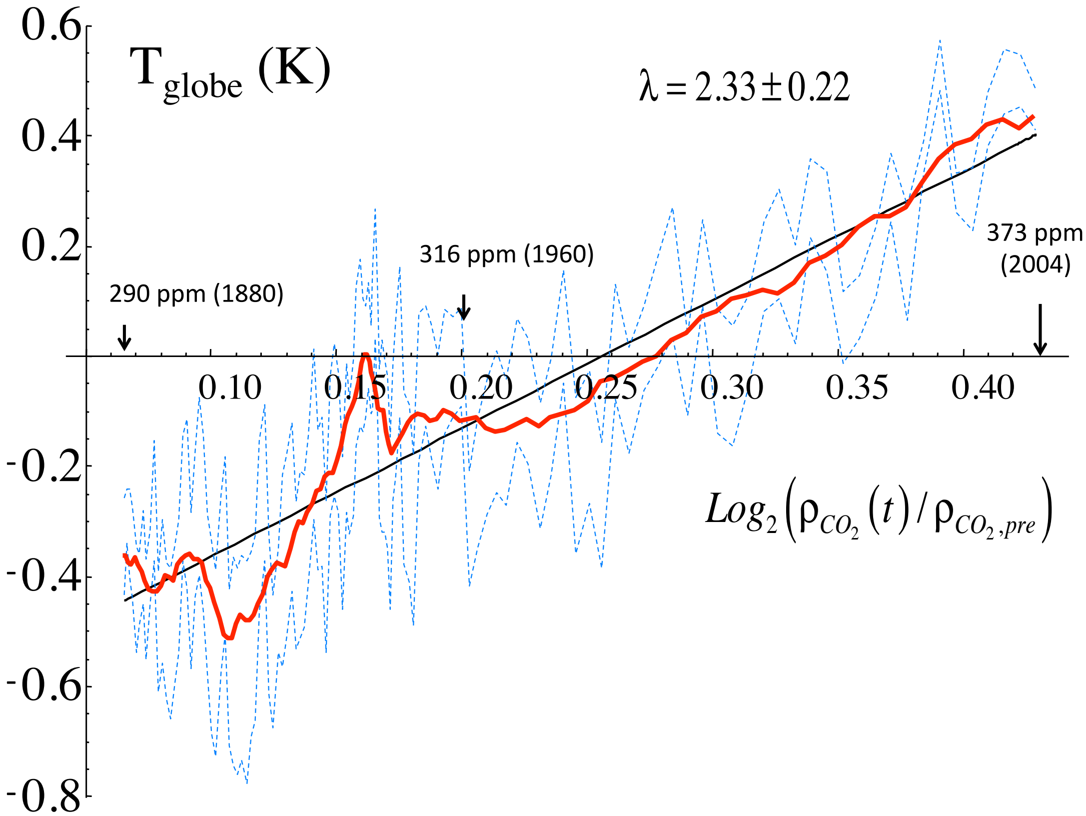

Fig. 3a: The mean global temperature estimated from NASA-GISS, NOAA NCDC, HADCrut3 data bases as a functions of the logarithm of the mean CO_2 concentration from [*Frank et al.*, 2010]. The dashed lines represent the one standard deviation variations of the three series at one year resolution, the thick line is the mean with a five year running average. Also shown is the linear regression with the effective climate sensitivity to CO_2 doubling: 2.33 ±0.22 K.

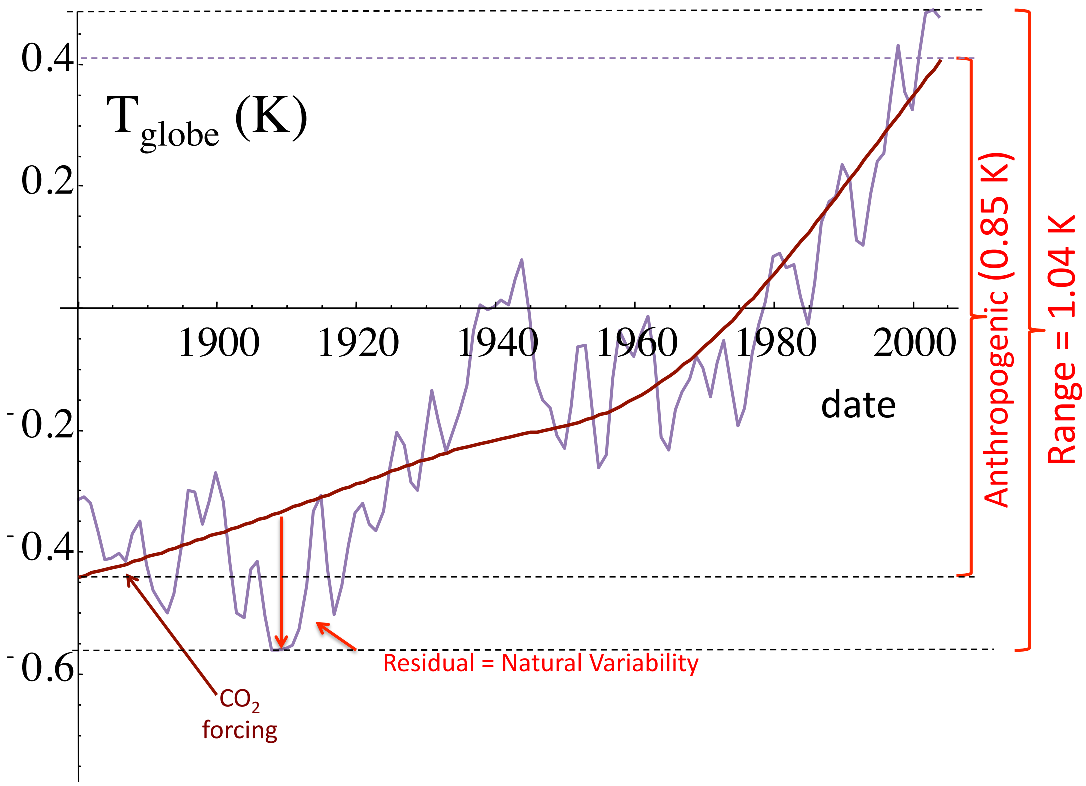

Fig. 3b: Five year running average of the average temperature. The brown line is the estimate of $T_{anth}(t)$ from eq. 6 with λ_{2xCO2} = 2.33 and the difference (residue) is the estimate of the natural variability $T_{nat}(t)$. Also shown in the regression of the latter with time (straight line) as well the overall estimates ΔT_{anth} =0.85±0.08 for the unlagged relation and the overall range $\Delta T_{\mathrm{globe,range}}$ =1.04±0.03 K which presumably bounds ΔT_{anth}.

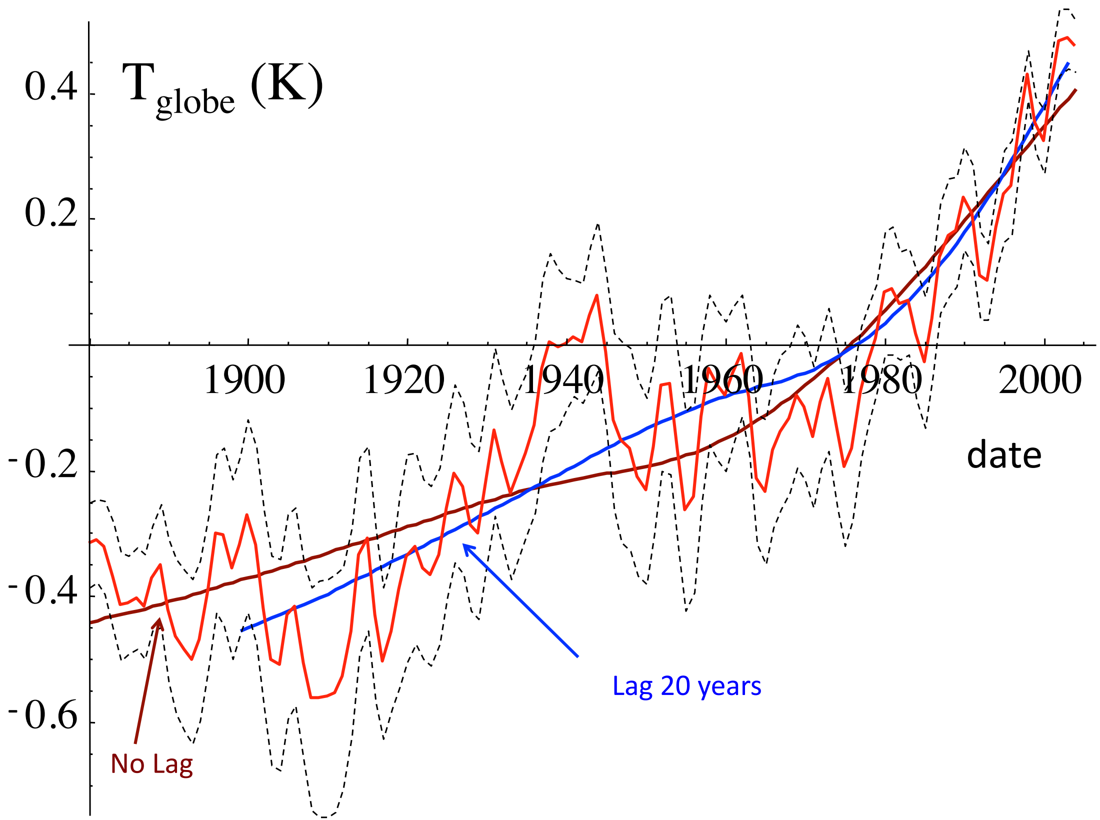

Fig. 3c: The comparison of the mean global temperature series (red), one standard deviation limits (dashed, all from the three surface series discussed above, all with a five year running average), compared with the unlagged (brown, corresponding to fig. 3a) and 20 year lagged (blue) estimates obtained from $\log_2\rho_{CO2}$ versus T_{globe} regressions as discussed in the text.

2.4 The time Lagged sensitivities

It may be objected that the most immediate consequence of R_F is to warm the oceans [*Lyman et al.*, 2010] so that we expect a time lag between the forcing and atmospheric response, for example, with GCM's [*Hansen et al.*, 2005] finds a lag of 25- 50 years, and [*Lean and Rind*, 2008] empirically find a lag of 10 years (of course, the situation is not quite so simple due to feedbacks). By considering the time lagged cross correlation between $R_{F,CO2}$ and T_{globe} (fig. 4) it is found that the cross correlations are so high (with maximum 0.94) that the maximally correlated lag is not well pronounced. To clarify this, we also calculated the corresponding curves for the cross correlation of the temperature fluctuations (ΔT, differences) at a five year resolution. The fluctuations are more weakly correlated than with the temperatures themselves so that this is a bit more sensitive to varying lags. In all cases, we can see that the maximum is roughly between a lag of zero and 20 years. However, the effective climate sensitivity to doubling CO_2 increases from 2.33±0.22 (zero lag) to 3.82±0.54 with a 20 year lag (see fig. 3c for a comparison with the zero lag anthropogenic and empirical global temperatures). If we use a Bayesian approach and assign equal a priori probabilities to all the lags between zero and 20 years, then we obtain the estimate $\lambda_{2x,CO2,eff} = 3.08 \pm 0.58$ K which is (unsurprisingly) quite close to the ten year lag value (fig. 4). Note that we could use a general linear relation between forcings and responses using Green's functions, but this would require additional assumptions and is not necessary at this point.

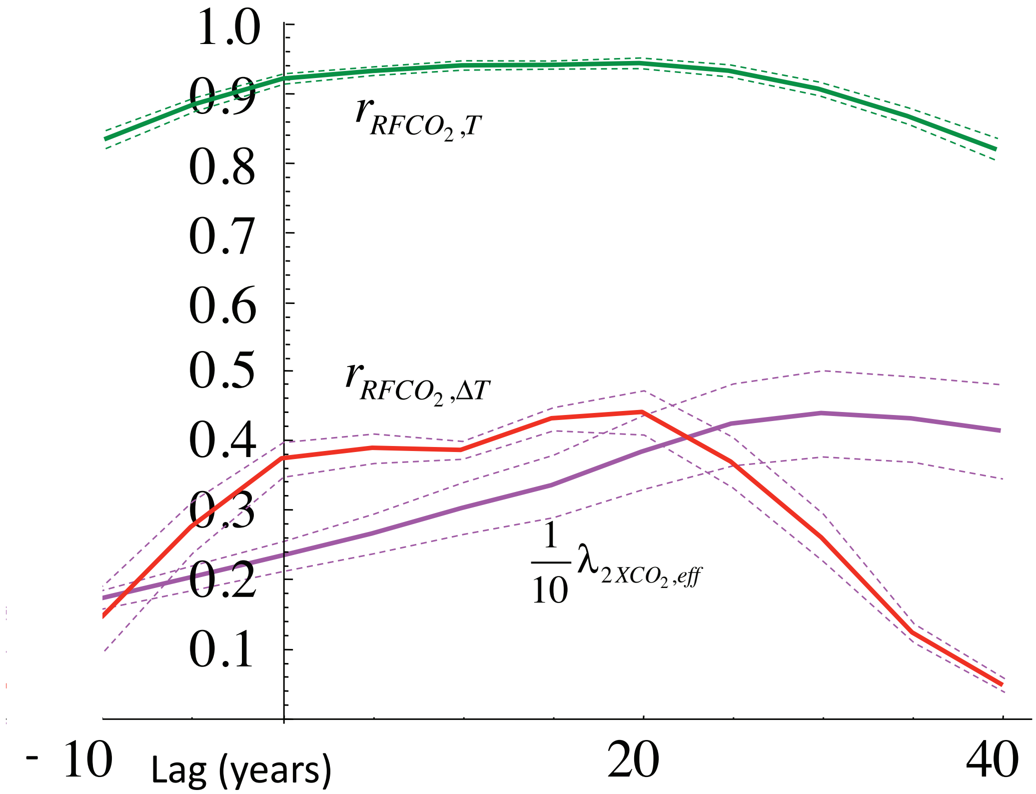

Fig. 4: The green curve is the cross correlation coefficient of the lagged R_{FCO2} (from the CO_2 reconstruction of [*Frank et al.*, 2010]) and the global mean temperatures (averaged at 5 year resolution) with dashed lines indicating one standard deviation variations (as estimated from the three global mean temperature series). As can be seen, the cross correlations are so high that the maximally correlated lag is not well pronounced. To bring out the maximum more clearly, we also calculated (red) the corresponding curves for the cross correlation of the fluctuations (differences) of five year averages. We can see that the maximum is roughly between zero and lag 20 years. However, the effective climate sensitivity to doubling CO_2 (purple, divided by 10) increases from 2.33±0.22 (zero lag) to 3.82±0.54 with a 20 year lag.

2.5. Effective and equilibrium Climate sensitivities

Our estimate of $\lambda_{2x,CO2,eff}$ has the advantage of being not only independent of GCM's, but also with respect to assumptions about radiative transfer, historical (non CO_2) GHG and aerosol emission histories. However, $\lambda_{2x,CO2,eff}$ is an "effective" sensitivity both because it uses CO_2 as a surrogate for all the anthropogenic R_F, and also because it is not a usual "equilibrium climate sensitivity" defined as "the equilibrium annual global mean temperature response to a doubling of equivalent atmospheric CO_2 from pre-industrial levels" (AR4). Since only GCM's can truly attain "equilibrium" (and this only asymptotically in a slow power law manner [*Lovejoy et al.*, 2013a]), this climate sensitivity is really a theoretical / model concept that can at best only be approximated with real world data. From an empirical point of view, whereas the effective climate sensitivity is the actual sensitivity to our current (uncontrolled) experiment, the equilibrium and transient sensitivities are the analogues for various (impractical) controlled experiments.

Because of the differences in the definitions of climate sensitivity, it would be an exaggeration to claim that we have empirically validated the model based results, even though our value $\lambda_{2xCO2,eff}$ = 3.08±0.58 (taking into account the uncertainty in the lag) is very close to literature values (c.f. the AR5 range 1.5- 4.5 K, the AR4 range 2 - 4.5 K, and the value 3±1.5 K adopted by the National Academy of Sciences (1979) and the AR1 – 3 reports). It is not obvious whether effective or equilibrium sensitivities are more relevant for predicting the temperature rise in the 21st century.

3. Statistical analysis

3.1 The stationarity of the residuals T_{nat} and comparison with the pre-industrial T_{nat}

While the linearity of fig. 3a, c is encouraging (even impressive), its interpretation as representing an anthropogenic component is only credible if the residuals ($T_{\mathrm{nat}}(t)$) have statistics very similar to those of T_{globe} in pre-industrial epochs (when T_{anth} = 0) so that as hypothesized in eq. 1, they could all be realizations of the same stochastic process. As a first confirmation of this, in the top two curves of fig. 5 we plot both T_{globe} and T_{nat} estimated from the residuals (i.e. $T_{nat}(t) = T_{globe}(t) - \lambda_{2xCO2,eff} \log_2\left(\rho_{CO_2}(t)/\rho_{CO_2,pre}\right)$). Even without any formal statistical analysis, we see - as expected - that whereas T_{globe} is clearly increasing, T_{nat} is roughly flat. However, for eq. 1 to be verified, we also require that the residuals have similar statistics to the preindustrial fluctuations when $T_{anth} = 0$ and $T_{globe} = T_{nat}$. In order to establish this, we must use multiproxy reconstructions which are the only source of annual resolution preindustrial global scale temperatures.

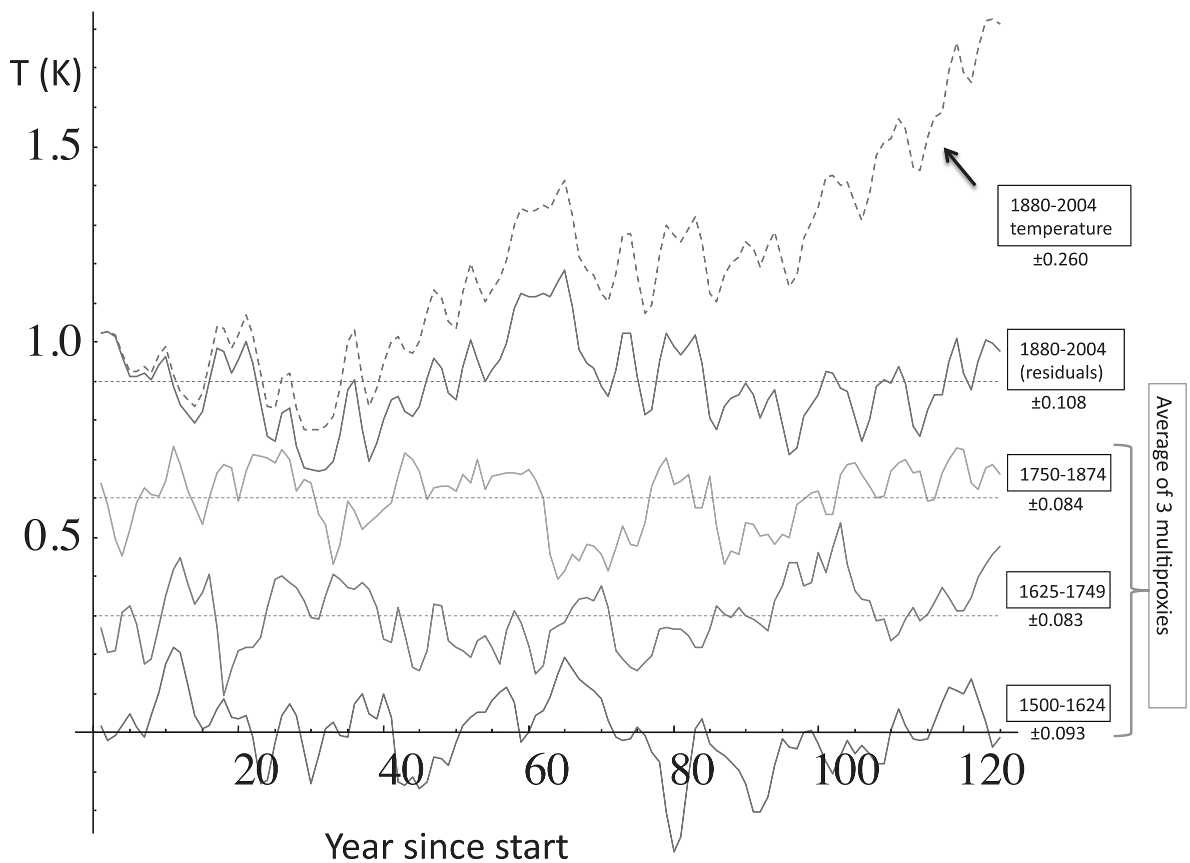

Fig. 5: The three lower curves are the means of the three multiproxies discussed in the text over three consecutive 125 year periods starting in the year 1500 with their

standard deviations indicated. Each segment had its overall mean removed and was displaced by 0.3K in vertical for clarity. The fourth curve from the bottom is from the (unlagged) residuals with respect to the CO_2 regression in fig. 3a (1880-2004). The top (dashed) curve is the annual resolution mean temperature. Whereas the curves from the three multiproxy epochs are quite similar to the residuals in the recent epoch, the actual recent epoch temperature shows a fairly systematic increase.

Following the analysis in [*Lovejoy and Schertzer*, 2012a], the more recent (mostly post 2003) multiproxies (those developed after 2003) were argued to be more faithful to the low frequency (multicentennial) variability. In particular, when compared to ice core paleotemperatures the low frequencies in [*Huang*, 2004], [*Moberg et al.*, 2005] and [*Ljundqvist*, 2010] were found to be more realistic with fluctuations starting to increase in amplitude for $\Delta t > \approx 100$ years (preindustrial). However, one of these series ([*Ljundqvist*, 2010]) was at 10 year resolution and was not suited for the present study which required annual series. It was therefore replaced by the [*Ammann and Wahl*, 2007] update of the original [*Mann et al.*, 1998] reconstruction which although having somewhat smaller multicentennial variability was statistically not too different (see fig. 6 for a comparison of the probability distributions of the differences at lags of one year). This shows that at one year resolution, fluctuations from the different multiproxies have nearly the same probability distributions although with slightly different amplitudes (c.f. the left-right shift on the log-log plot). Changes in the amplitude arise due to varying degrees of spatial averaging so that - given the different types and quantities of data contributing to each multiproxy - these amplitude differences are not surprising (see [*Lovejoy and Schertzer*, 2013]). In the figure we also see the residuals of the unlagged estimate of T_{nat}. At this scale the residuals have slightly larger variability (see the comparison of the standard deviations as functions of scale in fig. 7), although after $\Delta t \approx 4$ years, it falls within the epoch to epoch variations of the mean of the multiproxies.

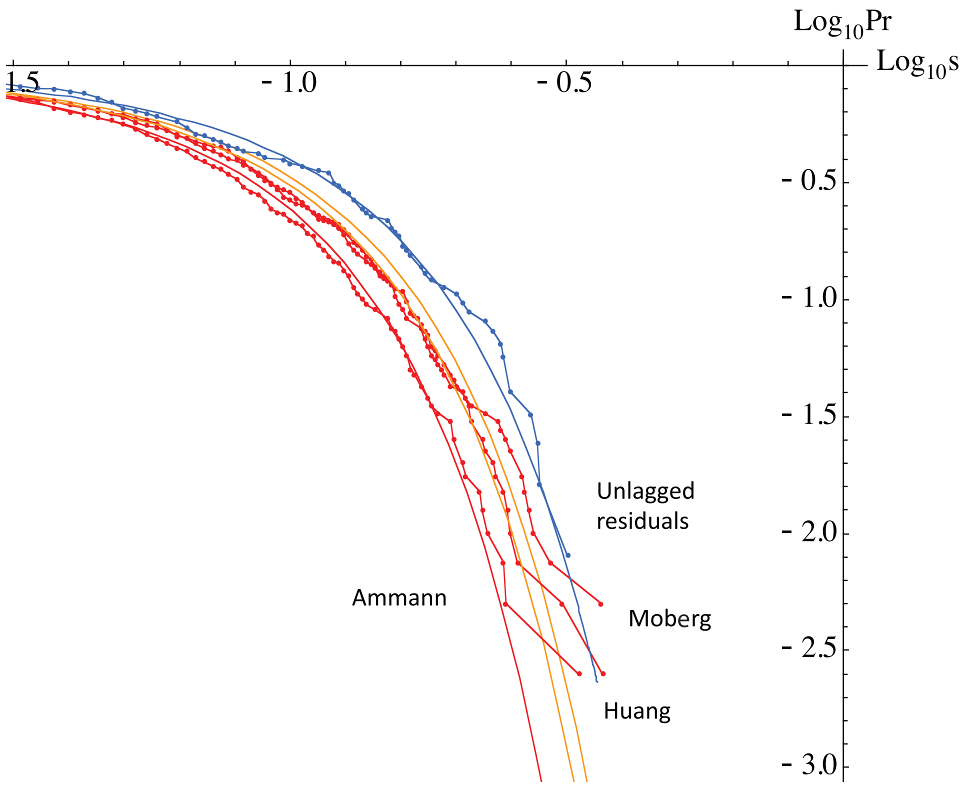

Fig. 6: The temperature differences for Δt = 1 year for the three multiproxies (red, 1500-1900) compared with the (unlagged) residuals from fig. 1. "Pr" indicates $Pr(\Delta T>s)$ which is the probability that a random temperature difference ΔT exceeds a fixed threshold s. The smooth curves are the Gaussians with the same standard deviations. We see that the multiproxies are quite close to each other – although with some small variations in amplitude - about 10% between each curve - but not much in shape.

We can now make a first comparison between the industrial epoch residuals and the pre-industrial anomalies; see the bottom three curves in fig. 5. To mimick the 125 year industrial period, the multiproxies were divided into 3x125 pre-industrial periods (1500-1624, 1625-1749, 1750-1875) as shown, each with its overall mean removed. We see that while the industrial epoch temperatures increase strongly as functions of time, that the amplitudes and visual appearances of the residuals and the multiproxies are strikingly similar.

We now turn to the problem of making this similitude quantitative. The traditional way to characterize the variability over a wide range of scales is by spectral analysis. It is typically found that climate spectra are dominated by red noise "backgrounds" and over wide ranges, these are roughly power laws (scaling) indicating that over the range, there is no characteristic scale and (in general) that there are long range statistical dependencies (e.g. correlations; see [*Lovejoy*, 2014] for recent overview and disucssion). However spectral analysis has disadvantages, the most important of which is that its interpretation is not as straightforward as real-space alternatives. This has lead to the development of wavelets and other methods of defining fluctuations (e.g. Detrended Fluctuation Analysis [*Peng et al.*, 1994]). However [*Lovejoy and Schertzer*, 2012b] shows that the simple expedient of defining fluctuations over intervals Δt by the differences in the means over the first and second halves of the interval ("Haar fluctuations") is particularly advantageous since unlike differences - which on (ensemble) average cannot decrease – Haar fluctuations can both increase and decrease. The critical distinction between increasing and decreasing fluctuations corresponds to a spectral exponent greater or less than $\beta = 1$ (ignoring small intermittency corrections). In regions where the Haar flucutations increase they are proportional to differences, in regions where they decrease, they are proportional to averages so that the interpretation is very straightforward.

3.2 Fluctuation analysis of the industrial residuals and preindustrial multiproxies

In figure 7, first note the comparison of the RMS difference fluctuations of the three surface series (1880-2008) with those of the three multiproxies (1500-1900). Up until Δt $\approx$10 years they are quite close to each other (and slowly decreasing), then they rapidly diverge with the RMS preindustrial differences ($\sigma_{\Delta t}$) remaining roughly constant ($\sigma_{\Delta t} \approx 0.20\pm0.03$) until about 125 years. Fig. 8 shows the corresponding figure for the Haar fluctuations. Again we find that the industrial and preindustrial curves are very close up to $\approx$ 10 years followed by a divergence due to the high decadal and longer scale industrial period variability. Note that the preindustrial Haar fluctuations decrease slowly until $\approx$ 125 years. When we consider the RMS residuals we find they are mainly within the one standard deviation error bars of the epoch to epoch multiproxy variability so that removing the anthropogenic contribution gives residuals T_{nat} with statistics close to those of the pre-industrial multiproxies (fig. 8).

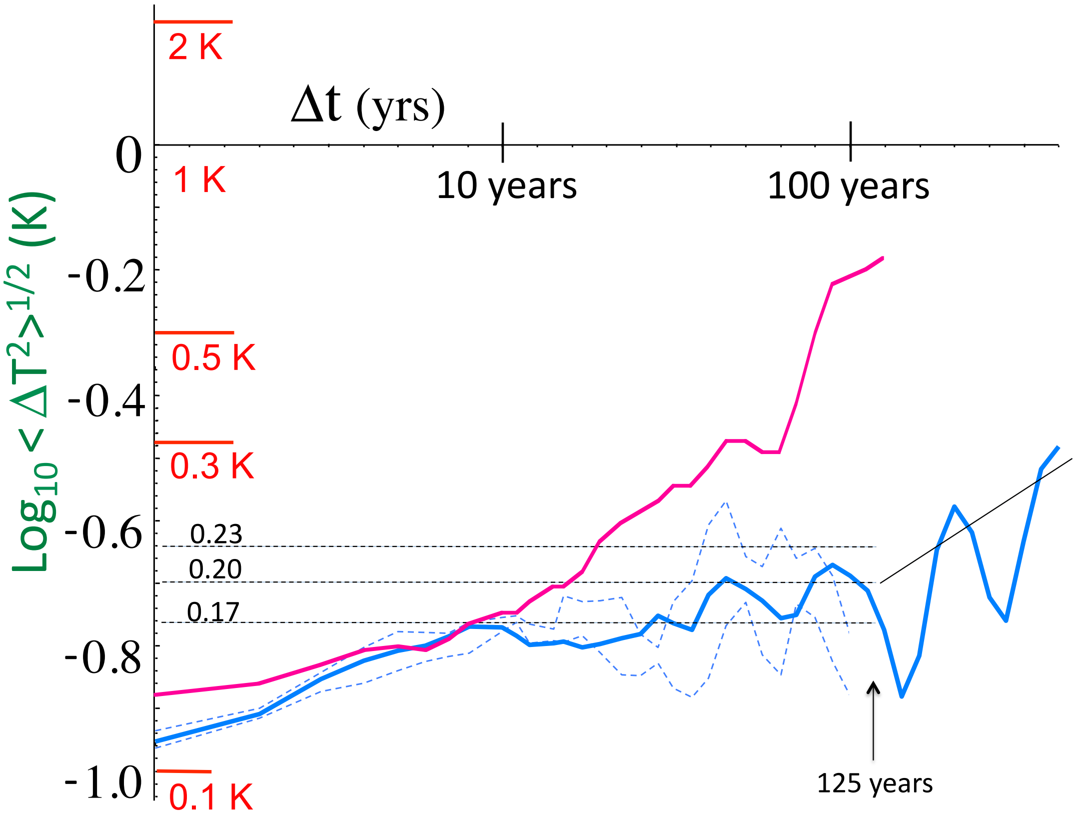

Fig. 7: The root mean square difference fluctuations for the mean of the three global surface series (top right, magenta, 1880-2004; from [*Lovejoy and Schertzer*, 2012a]); in the notation of section 3; $\sigma_{\Delta t}$. The corresponding (long blue) curve is for the northern hemisphere multiproxies from 1500-1900 and the dashed lines show the one standard deviation error bars estimated from the three 125 year epochs indicated in fig. 5 indicating the epoch to epoch variability. For periods less than about 10 years the fluctuations are roughly the same so that there is no significant difference in the northern hemisphere and global multiproxies. The increase in the beyond 10 years is due to global warming in the recent period.

For the (preindustrial) multiproxies we see that between ≈ 10 and 125 years, the RMS differences are ≈ constant, this is expected due to the slight decrease of the Haar fluctuations (fig. 8) over this range, see the appendix for a discussion. The solid line at the right has a slope 0.4; it shows the increase in the variability in the climate regime. From the graph at 125 years the RMS difference may be estimated as 0.20±0.03 K.

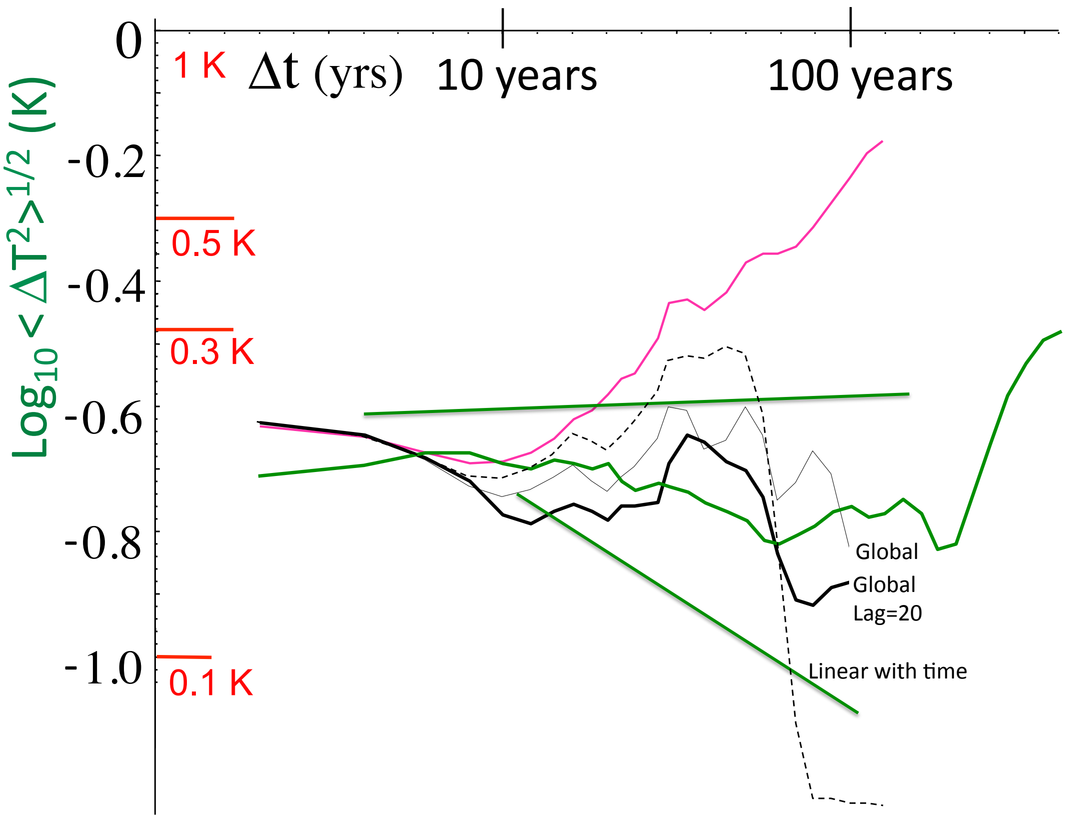

Fig. 8: The RMS Haar fluctuations for the surface series (magenta, top) and the multiproxies from 1500-1900 (long, thick green) with the green straight lines showing (roughly) the one standard deviation error bars estimated from the three 125 year epochs (1500-1624, 1625-1749, 1750-1874) indicated in fig. 5. The difference between the preindustrial multiproxies and industrial epoch surface temperatures is due to global warming. These are compared with the residuals from 1880-2004 obtained after subtracting the anthropogenic contribution obtained from the regression in fig. 3a (thin black line), from the corresponding residuals for a twenty year lag between forcing and temperature (thick black line), and for a linear CO_2 concentration versus temperature relation (dashed line). Both the lagged and unlagged $\log_2\rho_{CO2}$ residuals are generally within the one standard deviation limits, although the 20 year lagged residuals are closer to the mean.

The Haar fluctuations were multiplied by a "calibration" factor = 2 so that they would be close to the difference fluctuations (fig. 7). Note that a straight line slope H corresponds to a power law spectrum exponent 1+2H so that a flat line has spectrum $E(\omega) \approx \omega^{-1}$, and hence long range statistical dependencies (Gaussian white noise has slope -0.5). The roughly linear decline of the multiproxy variability to about $\Delta t \approx 125$ years is the (fluctuation cancelling, decreasing) macroweather regime, the rise beyond it, the "wandering" climate regime [Lovejoy, 2013].

3.3 Estimating the probability that the warming is due to natural variability

Regressing $R_{F,CO2}$ against the global mean temperature leads to satisfactory results in the sense that the residuals and preindustrial multiproxies are plausibly realizations of the same stochastic process. However, this result is not too sensitive to the exact method of estimating T_{anth} and T_{nat} - the 20 year lagged residuals are a bit better although using simply a linear regression of T_{globe} against time is substantially worse; see fig. 8. From the point of view of determining the probability that the warming is natural, the key quantity is therefore the total anthropogenic warming $\Delta T_{ant} = T_{ant}(2004) - T_{ant}(1880)$. Using the $\log_2\rho$ method (fig. 3a) we find $\Delta T_{anth} \approx 0.85\pm0.08$ K and with a 20 year lag $\approx0.90\pm0.13$ K (the zero lag northern hemisphere value is 0.94 ± 0.09 K). With a Bayesian approach, assuming equal a priori probabilities of any lag between zero and twenty years, we obtain $\Delta T_{anth} \approx 0.87 \pm 0.11$; for comparison, for the linear in time method, we obtain $\approx0.75\pm0.07$ K (essentially the same as the AR4 estimate which used a linear fit to the HadCRUT series over the period 1900 - 2004). We can also estimate an upper bound - the total range $\Delta T_{\mathrm{globe,range}} = Max(\Delta T_{globe}) \approx 1.04\pm0.03$ K so that (presumably) $\Delta T_{\mathrm{anth}} < \Delta T_{\mathrm{globe,range}}$.

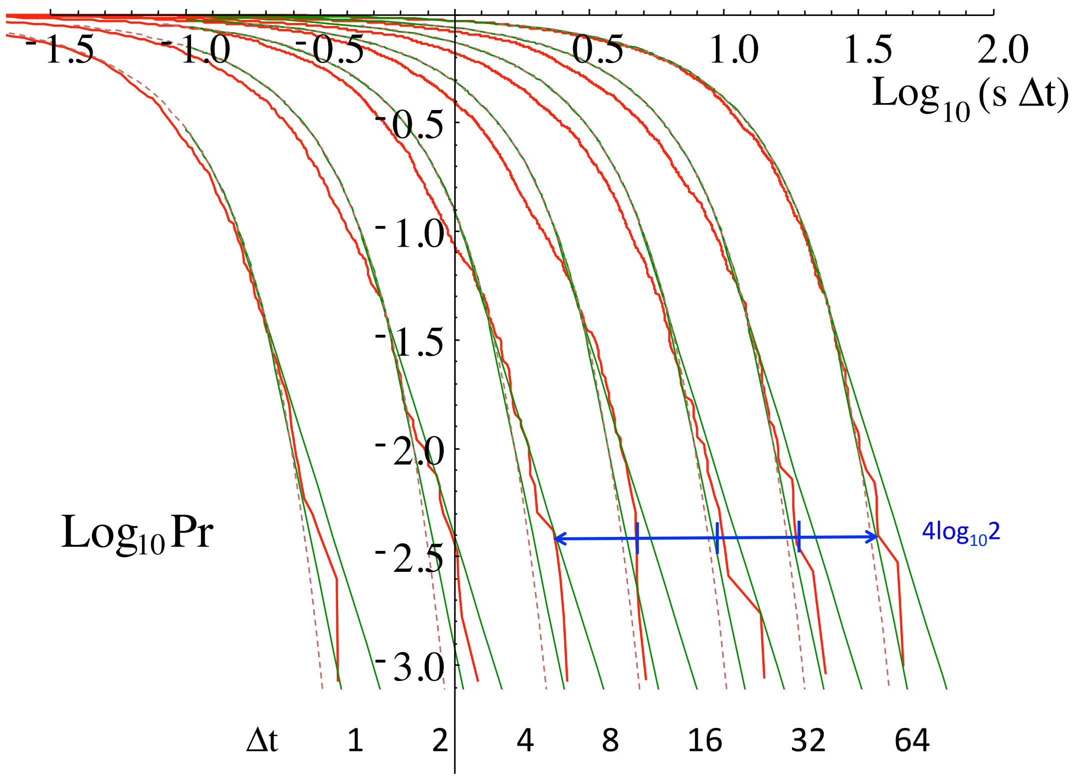

Fig. 9: This shows the total probability of random absolute pre 1900 temperature differences exceeding a threshold s (in K), using all three multiproxies to increase the sample size (compare this to fig. 6 which shows that the distribution are very similar in form for each of the multiproxies). To avoid excessive overlapping, the latter were compensated by multiplying by the lag Δt (in years, shifting the curves to the right successively by $\log_{10}2 \approx 0.3$), the data are the pooled annual resolution multiproxies from 1500-1900. The blue double headed arrow shows the displacement expected if the difference amplitudes were constant for 4 octaves in time scale (corresponding to negative H for Haar fluctuations, H =0 for differences, see fig. 7 for the standard deviations each octave is indicated by a vertical tick mark on the arrow). The (dashed) reference curves are Gaussians with corresponding standard deviations and with (thin, straight) tails (Pr ≈<3%) corresponding to bounding s^{-4} and s^{-6} behaviors.

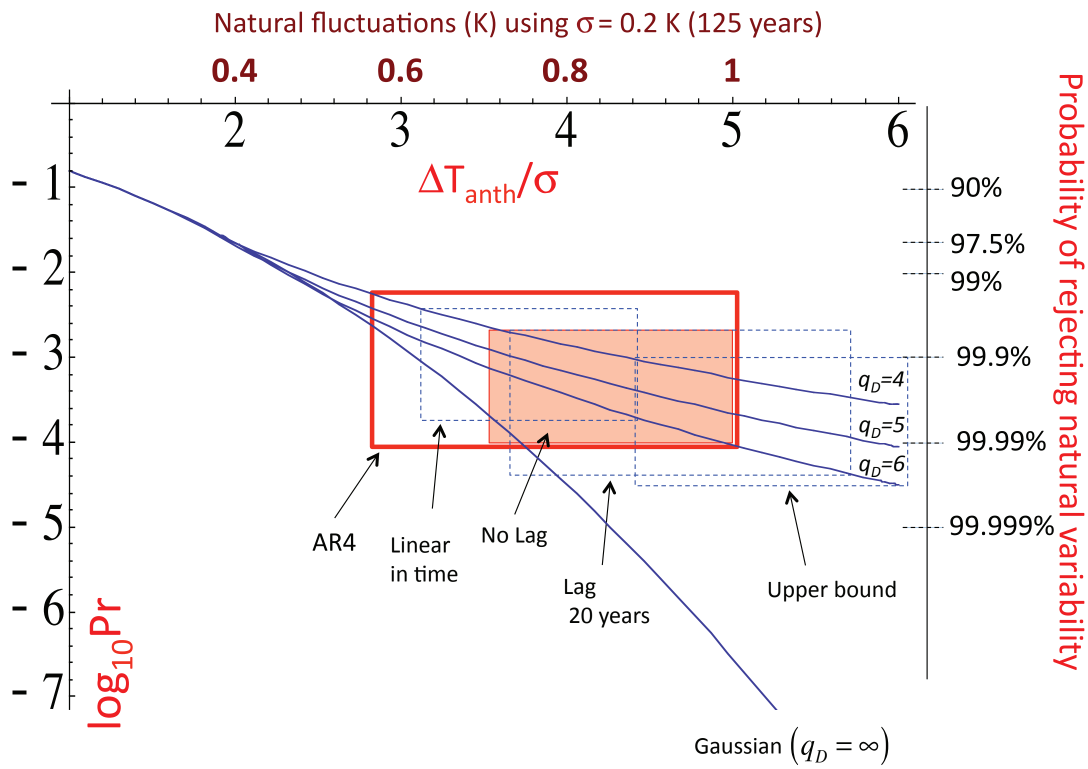

Fig. 10: The probability of anthropogenic warming by ΔT_{anth} as functions of the number of standard deviations for the five cases discussed in the text. Also shown for reference is the equivalent temperature fluctuation using the mean standard deviation at 125 years. The vertical sides of the boxes are defined by the one standard deviation limits of ΔT_{anth} /σ, the horizontal sides by the q_D =4 (upper) and q_D = 6 (lower) limits; the middle curve (q_D = 5) is the mean (most likely) exponent. The classical statistical hypothesis (Gaussian, corresponding to $q_D = \infty$) is indicated for reference. The AR4 ΔT_{anth} =0.74 ± 0.18 is indicated by the thick red line and using $\log_2\rho_{CO2}$ as a surrogate for the RF followed by linear regression (ΔT_{anth} =0.85 ± 0.08) is shown in the filled orange box. The other cases are shown by dashed lines: $\log_2\rho_{CO2}$ but with a 20 year lag, linear regression of T_{globe} against time and the upper bound on ΔT_{anth} = 1.04±0.03.

We now estimate the probability distribution of the temperature differences from the multiproxies first over the shorter lags with reliable estimates of extremes (up to Δt = 64 years, fig. 9), and then using the scaling of the distributions and RMS fluctuations to deduce the form at Δt = 125 years, (see the appendix). We find the 125 year RMS temperature difference ($\left\langle \Delta T(125)^2 \right\rangle^{1/2} = \sigma_{125}$ = 0.20±0.03 K (fig. 7). Theoretically, spatial and temporal scaling are associated with probabilities with power law "fat" tails (i.e. $Pr(\Delta T>s) \approx s^{-q_D}$ for the probability of a fluctuation exceeding a threshold s; q_D is an exponent), hence in fig. 10 we compare q_D = 4, 6 and $q_D = \infty$ (a pure Gaussian). We see that the former two values bracket the distributions (including their extremes) over the whole range of large fluctuations (the extreme 3%).

Stated succinctly, our statistical hypothesis on the natural variability is that its extreme probabilities (Pr <3%) are bracketed by a modified Gaussian with q_D between 4 and 6 and with standard deviation (and uncertainties) given by the scaling of the multiproxies in fig. 7: σ_{125} = 0.20±0.03 K. For large enough probabilities (small s), the modified Gaussian is simply a Gaussian, but below a probability threshold (above a critical threshold s_{qD}) the logarithmic slope is equal to $-q_D$; i.e. it is a power law (see the appendix for details). With this, we can evaluate the corresponding probability bounds for various estimates of ΔT_{anth}. These probabilities are conveniently displayed in fig. 10 by boxes. For example, the AR4 ΔT_{anth} = 0.74 ± 0.18 K (thick red box) yields a probability (p): 0.009% < p < 0.6% whereas the (unlagged) $\log_2\rho_{CO2}$ regression (filled red box) yields 0.0009% < p < 0.2% and the 20 year lag (dashed blue) yields 0.002% < p < 0.2%, the northern hemisphere yields 0.009% < p < 0.1% with most likely values (using q_D = 5) of 0.08%, 0.08%, 0.03%, 0.03% respectively. In even the most extreme cases, the hypothesis that the observed warming is due to natural variability may be rejected at confidence levels 1-p >99%, and with the most likely values, at levels >99.9%. The other cases considered do not alter these conclusions (fig. 10).

4. Conclusions

Two aspects of anthropogenic global warming are frequent sources of frustration. The first is the lack of a quantitative theory of natural variability with which to compare the observed warming ΔT_{anth}, the second is the near exclusive reliance on GCM's to estimate it. In this paper we have argued that since ≈ 1880, anthropogenic warming has dominated the natural variability to such an extent that straightforward empirical estimates of the total warming can be made. The one favoured here - using CO_2 radiative forcing (R_F) as a surrogate for all anthropogenic R_F - gives both effective sensitivities $\lambda_{2xCO2,eff}$ and total anthropogenic increases ΔT_{anth} (3.08±0.58 K and 0.87±0.11 K) comparable to the AR4, AR5 estimates (1.5 - 4.5 K and 0.74±0.18 K for the slightly shorter period 1900-2005). The method was justified because we showed that over a wide range of scales, the residuals have nearly the same statistics as the preindustrial multiproxies. An additional advantage of this approach is that it is independent of many assumptions and uncertainties including radiative transfer, GCM and emission histories. The main uncertainty is the duration of the lag between the forcing and the response.

Whether one estimates ΔT_{anth} using the empirical method proposed here, or using a GCM based alternative, when ΔT_{anth} is combined with the scaling properties of multiproxies we may estimate the probabilities as functions of time scale and test the hypothesis that the warming is due to natural variability. Our statistical hypothesis – supported by the multiproxy data - is that due to the scaling - there are long range correlations in the temperature fluctuations coupled with nonclassical "fat tailed" probability distributions which bracket the observed probabilities. Both effects lead to significantly higher probabilities than would be expected from classical "scale bound" (exponentially decorrelated) processes and/or with "thin" (e.g. Gaussian or exponential) tails. However, even in the most extreme cases, we are still able to reject the natural variability hypothesis with confidence levels >99% - and with the most likely values - at levels >99.9%. Finally, fluctuation analysis shows that the variability of the recent period solar forcing was close to preindustrial levels (at all scales), and that volcanic forcing variabilities were a factor 2-3 times weaker (at all scales), so that they cannot explain the warming either.

While no amount of statistics will ever prove that the warming is indeed anthropogenic, it is nevertheless difficult to imagine an alternative.

Appendix: Scaling modified Gaussians with fat tails:

In fig. 9 we showed the empirical probability distributions ($\Pr(\Delta T>s)$, for the probability of a random (absolute) temperature difference ΔT exceeding a threshold s for time lags Δt increasing by factors of 2. Note that we loosely use the expression "distribution function" to mean $\Pr(\Delta T>s)$. This is related to the more usual "cumulative distribution function" (CDF) by: CDF = $\Pr(\Delta T<s)$ so that $\Pr(\Delta T>s)$ = 1- CDF. Two aspects of fig. 9 are significant; that the first is their near scaling with lag Δt: the shapes change little, this is the type of scaling expected for a monofractal "simple scaling" process, i.e. one with weak multifractality (as discussed in [*Lovejoy and Schertzer*, 2013], over these time scales, the parameter characterizing the intermittency near the mean, $C_1 \approx 0$ so that this is a reasonable approximation).

This implies that there is a nondimensional distribution function $P(s)$:

$$P(s) = \Pr\left(\frac{\Delta T(\Delta t)}{\sigma_{\Delta t}} > s\right); \quad \sigma_{\Delta t} = \left\langle \Delta T(\Delta t)^2 \right\rangle^{1/2}$$

$\sigma_{\Delta t}$ is the standard deviation. Due to the temporal scaling, we have $\sigma_{\lambda\Delta t} = \lambda^H \sigma_{\Delta t}$ where H is the fluctuation exponent and $P(s)$ is independent of time lag Δt. From fig. 9 it may be seen that as predicted by the RMS fluctuations ($\sigma_{\Delta t}$, fig. 7), $H \approx 0$. This is a consequence of the slight decrease in the RMS Haar fluctuation (with exponent $H_{Haar} \approx$ -0.1; fig. 8). Unlike the Haar fluctuation, the ensemble mean RMS differences cannot decrease but simply remain constant until the Haar fluctuations begin to increase again (compare figs. 7, 8, beyond $\Delta t \approx 125$ years).

The second point to note is that the lag invariant distribution function $P(s)$ has roughly a Gaussian shape for small s, whereas for large enough s, it is nearly algebraic. This can be simply modelled as:

$$P_{q_D}(s) = \begin{matrix} P_G(s); & s < s_{qD} \\ P_G(s_{qD})\left(\dfrac{s}{s_{qD}}\right)^{-q_D}; & s \geq s_{qD} \end{matrix}$$

where $P_G(s)$ is the cumulative distribution function for the absolute value of a unit Gaussian random variable. The simple way of determining s_{qD} used here is to define s_{qD} as the point at which the logarithmic derivative of P_G is equal to $-q_D$ so that the plot of $\log P_{qD}$ versus $\log s$ is continuous:

$$\left.\frac{d\log P_G(s)}{d\log s}\right|_{s=s_{qD}} = -q_D$$

this is an implicit equation for the transition point s_{qD}.

In actual fact the only part of the model that is used for the statistical tests is the extreme large s "tail" which fig. 9 empirically shows could be bracketed between:

$$P_{qD1}(s) < P(s) < P_{qD2}(s); \quad q_{D1} > q_{D2}; \quad s > s_{qD1} > s_{qD2}$$

(with $q_{D1} = 6$, $q_{D2} = 4$) hence the Gaussian part of the model is not very important, it only serves to determine the transition point s_{qD}. In any case, for the extremes we can see from the figure that this bracketing is apparently quite well respected by the empirical distributions.

Acknowledgements:

P. Dubé, the president of the Quebec Skeptical Society, is thanked for helping to motivate this work. This work was unfunded, there were no conflicts of interest.

References

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

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

Bauer, S., and E. Menon (2012), Aerosol direct, indirect, semidirect, and surface albedo effects from sector contributions based on the IPCC AR5 emissions for preindustrial and present-day conditions, *J. Geophys. Resear. (Atmos.)*, *117* doi: 10.1029/2011JD016816.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850, *J. Geophysical Research*, *111*, D12106 doi: doi:10.1029/2005JD006548

Foster, G., and S. Rahmstorf (2011), Global temperature evolution 1979–2010, *Environ. Res. Lett.* , *6* 044022 (044028pp) doi: 10.1088/1748-9326/6/4/044022.

Frank, D. C., J. Esper, C. C. Raible, U. Buntgen, V. Trouet, B. Stocker, and F. Joos (2010), Ensemble reconstruction constraints on the global carbon cycle sensitivity to climate, *Nature*, *463 (28)* doi: doi:10.1038/nature08769.

Gao, C. G., A. Robock, and C. Ammann (2008), Volcanic forcing of climate over the past 1500 years: and improved ice core-based index for climate models, *J. Geophys. Res.*, *113*, D23111 doi: 10.1029/2008JD010239.

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

Hansen, J., et al. (2005), Earth's Energy Imbalance: Confirmation and Implications, *Science*, *308 no. 5727* 1431-1435 doi: 10.1126/science.1110252.

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

Krivova, N. A., L. Balmaceda, and S. K. Solanski (2007), Reconstruction of solar total irradiance since 1700 from the surface magnetic field flux, *Astron. and Astrophys*, *467*, 335-346 doi: 10.1051/0004-6361:20066725.

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

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

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

Lovejoy, S. (2014), A voyage through scales, a missing quadrillion and why the climate is not what ou expect, *Climate Dyn.*, *(submitted, 2/14)*.

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

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

Lovejoy, S., and D. Schertzer (2012c), Stochastic and scaling climate sensitivities: solar, volcanic and orbital forcings, *Geophys. Res. Lett.* , *39*, L11702 doi: doi:10.1029/2012GL051871.

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

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

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

Lovejoy, S., C. A. Varotsos, and M. N. Efstathiou (2014), Scaling analyses of forcings and outputs of a simplified Last Millennium climate model, *J. Geophys. Res.*, *(submitted, March 2014)*.

Lyman, J. M., S. A. Good, V. V. Gouretski, M. Ishii, G. C. Johnson, M. D. Palmer, D. M. Smith, and J. K. Willis (2010), Robust warming of the global upper ocean, *Nature*, *465*, 334–337 doi: 10.1038/nature09043.

Mann, M. E., R. S. Bradley, and M. K. Hughes (1998), Global-scale temperature patterns and climate forcing over the past six centuries, *Nature*, *392*, 779-787.

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

Muller, R. A., J. Curry, D. Groom, R. Jacobsen, S. Perlmutter, R. Rohde, A. Rosenfeld, C. Wickham, and J. Wurtele (2013), Decadal Variations in the Global Atmospheric Land Temperatures *(Berkeley Earth project, unpublished)*.

Myhre, G. (2009), Consistency Between Satellite-Derived and Modeled Estimates of the Direct Aerosol Effect, , *Science*, *325 (5937)* 187-190 doi: DOI: 10.1126/science.1174461.

Myhre, G., A. Myhre, and F. Stordal (2001), Historical evolution of radiative forcing of climate, *Atmos. Environ.* , *35*, 2361-2373.

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

Peterson, T. C., and R. S. Vose (1997), An overview of the Global Historical Climatology Network temperature database, *Bull. Amer. Meteorol. Soc.*, *78*, 2837-2849 doi: doi:10.1175/1520-0477.

Rayner, N. A., P. Brohan, D. E. Parker, C. K. Folland, J. J. Kennedy, M. Vanicek, T. Ansell, and S. F. B. Tett (2006), Improved analyses of changes and uncertainties in marine temperature measured in situ since the mid-nineteenth century: the HadSST2 dataset. , *J. Climate*, *19*, 446-469.

Santer, B. D., et al. (2013), Identifying human influences on atmospheric temperature, *Proceedings of the National Academy of Sciences*, *110*(1), 26-33 doi: 10.1073/pnas.1210514109.

Smith, S. J., E. Conception, A. R., and L. J. (2004), Historical Sulfur Dioxide Emissions 1850-2000: Methods and Results*Rep.*, 16 pp, U.S. Department of Energy.

Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore (2008), Improvements to NOAA's Historical Merged Land-Ocean Surface Temperature Analysis (1880-2006), *J. Climate*, *21*, 2283-2293.

Wang, Y.-M., J. L. Lean, and N. R. J. Sheeley (2005), Modeling the Sun's magnetic field and irradiance since 1713, *Astrophys J.* , *625*, 522–538.

Wigley, T. M. L., P. D. Jones, and S. C. B. Raper (1997), The observed global warming record: What does it tell us?, *Proc. Natl. Acad. Sci. USA*, *94*, 8314–8320.