Distinguishing deterministic cycles from random walk excursions in natural time series.

John Reid
P.O. Box 279
Cygnet
Tasmania 7112
Australia
johnsinclairreid@gmail.com

July 2015

Abstract

Random walks are the outcome of a summing or integrating process and as such are widespread in nature. They are manifested as power law relationships between variance density and frequency. The random walk effect may be removed by differencing and decimating the data. In this way it is possible to show that apparently deterministic cycles observed in some natural time series are, in fact, random walk excursions.

1 Introduction

In the real world, few time series can be assumed to be stationary. The integration of energy is widespread in nature and results in 'random walk' processes which are non-stationary because variance increases linearly with time. This is particularly true of climate temperature time series as pointed out by Hasselmann (1976).

In a companion paper (Reid, 2015)[13] it was shown that, by abandoning asymptotic considerations, the unwindowed periodogram is a consistent estimator of the population spectrum of a finite time series. Furthermore a generalized spectrum can be estimated even when the time series is non-stationary. This allows random-walk behaviour to be identified and eliminated.

2 Variance spectra

2.1 White spectra

By equation (20) of Reid (2015) the kth element of the periodogram spectrum is given by:

$$P_k = \frac{1}{2N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \Phi_{mn} e^{-2\pi i k(m-n)/N} \tag{1}$$

Suppose the series $\{\xi_n\}$ is unselfcorrelated and stationary, then its variance/covariance function is given by

$$\Phi_{mn} = E(\xi_m \xi_n) = \sigma^2 \delta_{mn} \tag{2}$$

where σ^2 is the variance of $\{\xi_n\}$ and δ_{mn} is the Kronecker delta function, δ_{mn}. Substituting into (1) gives

$$P_k = \frac{1}{2N^2} \sum_{n=0}^{N-1} \Phi_{nn} = \frac{\sigma^2}{2N} \tag{3}$$

The population spectrum is flat and, for this reason the series is termed 'white'.

Typically, the researcher wishes to verify that a spectral peak at a specified frequency is due to, say, a cyclical component and did not occur merely by chance. The null hypothesis is proposed that the sample, $\{x_n\}$, with spectrum, $\{P_k\}$, is white with a variance, σ^2, estimated from the sample variance, $\hat{\sigma^2}$. Confidence limits can be computed for any probability, α, from the χ^2 distribution and shown as error bars on the graphed spectrum. If an apparently random peak *at a previously specified frequency* exceeds the error bars then the null hypothesis can be rejected and the existence of a cyclical component at this frequency established at the given

level of significance, α. Note that the spectral estimate, $\{P_k\}$, has $N/2-1$ independent ordinates and so $\sim \alpha(N/2-1)$ of these will exceed the α error bars by chance.That is why the frequency abscissa should be specified in advance.

If, on the other hand the researcher wishes to test whether there are *any* cyclical components present in $\{x_n\}$, the value of N must be taken into account and the confidence limits set accordingly by choosing a more suitable value of α.

2.2 Red spectra

The presence of cyclical components is not the only reason why a spectrum should be non-white. 'Red' or 'pink' spectra are common in nature and as the outcome of a summing or integrating processes. As the name suggests such spectra show an inverse relationship between variance density and frequency. This often takes the form of a power law, viz.:

$$S = Af^{\nu} \tag{4}$$

where S is the variance density, A is a constant, f is the frequency and ν is the power law index and is negative.

For a noisy spectral estimate, and substituting $k\mathrm{d}f$ for f, (4) becomes:

$$\log \hat{S}_k = \nu \log k + \log A' + \epsilon_k \tag{5}$$

where the $\{\epsilon_k\}$ are uncorrelated with zero mean. Equation (5) is a simple first order regression equation, allowing ν to be estimated and confidence limits placed upon it.

Because frequency domain variables including spectra are zero at zero frequency, the power law relationship (4) gives rise to a spectral peak at or near the lowest frequency, $1/N\Delta t$, particularly when the spectrum is plotted using linear scales and a power law relationship would not be immediately obvious. Such a peak is likely to be wrongly interpreted as a 'cycle' when in fact it is due solely to the accumulation of random walk fluctuations. Even in the time domain such fluctuations can appear cyclic even though they are not. Unlike a truly deterministic cycle, they have no predictive power and reveal little about underlying physical processes other than the presence of an integrating mechanism.

One of the primary functions of any method of spectral analysis should be to provide a criterion for distinguishing such random walk excursions from truly cyclic behaviour. To some extent this can be done by examining the width of a given peak; if the peak is wider than two abscissas then it is unlikely to be a true cycle. There is, however, a more rigorous method.

2.3 Whitening by differencing and decimation

The simplest form of auto-regressive process is specified by

$$\xi_n = b_1 \xi_{n-1} + \epsilon_n \tag{6}$$

where b_1 is a constant coefficient and the $\{\epsilon_n\}$ are uncorrelated with zero mean. In general $b_1 \leq 1$ otherwise (6) diverges and is of little interest. When $b_1 < 1$ (6) describes an integrator with exponential decay (such as an RC integrator in electronics) with time constant $\tau = -\Delta t/\ln(b_1)$ and, when $b_1 = 1$, (6) describes a true random walk process.

Equation (6) can be written as

$$\xi = \frac{\epsilon}{1 - b_1 \mathrm{L}} \tag{7}$$

where L is the lag operator. Using z-transform methods it can be shown that

$$\Phi = \frac{\delta}{(1 - b_1 \mathrm{L}^{-1})(1 - b_1 \mathrm{L})} \tag{8}$$

When $b_1 = 1$, the right hand side of (8) is singular and Φ does not exist, i.e. the sequence is non-stationary. When $b_1 \simeq 1$, Φ exists but Φ_n may not vanish for large values of n. If τ is larger than the time-span, $N\Delta t$, of the available data we have no way of knowing whether Φ_n exists or not but the generalized spectrum will exhibit a power law trend at low frequencies. It will look 'red'.

In this event the low frequency, power law trend can be removed by forming a new sequence $\{y_n, n = 1, ..., N-1\}$ by taking the first differences of the original time series $\{x_n\}$, i.e. $y_n = x_n - x_{n-1}$ and

$$\eta = (1 - \mathrm{L})\xi \tag{9}$$

where $\{y_n\}$ is assumed to be a realization of a sequence of random variables, $\{\eta_n\}$. Note however that substituting (9) in (7) does not cancel the inconvenient $1/(1 - b_1\mathrm{L})$ factor, nor can it do so because, in general, we cannot estimate the value of the population parameter b_1 precisely.

Differencing a time series is the inverse of summing. If the initial time series is unselfcorrelated, i.e. 'white', the new sequence of first differences will be 'blue', i.e. we have a power law spectrum with index, $\nu = +2$ in (4). Its variance/covariance function will take the values $\Phi_{-1} = \Phi_1/2 = -\Phi_0$ and $\Phi_n = 0$ for $|n| > 1$, i.e. differencing has introduced an anti-correlation between neighbouring members of the new sequence. However alternate members of the sequence will remain uncorrelated since $\Phi_n = 0$ for $\mid n \mid > 1$. Therefore if the new sequence is decimated by 2, it will again be white.

The same principle can be applied when removing the red power law trend according to (9). We create a further sequences

$$y'_i = y_{2i} \quad \text{for} \quad i = 0, ...\text{Int}(N/2) \tag{10}$$

and

$$y''_i = y_{2i} \quad \text{for} \quad i = 0, ...\text{Int}(N/2) \tag{11}$$

The power law component will be removed from the spectrum of the sequence $\{y_i\}$ without a spurious blue trend being introduced. Once this is done the spectrum of either $\{y'_i\}$ or $\{y''_i\}$ can be estimated and spectral peaks tested for significance using the χ^2 distribution.

3 Application to a paleoclimate time series

3.1 Processing the data

Milankovic (1941) [10] was first to propose that the cycling of climate apparent in the geological record and known as 'the Ice Ages' could be due to secular variations in the earth's orbit around the sun. These variations are thought to give rise, in turn, to variations in solar insolation at latitudes north of 60 N and its effect on the growth and collapse of the Northern Hemisphere ice sheet . The pioneering work of Hays *et al* (1976) [6] identified clear peaks in the ocean sediment record and related them to the eccentricity, obliquity and precession of the orbit. The 'eccentricity peak' near 100 kyr dominated their spectra. However, more recently, a number of workers have questioned the astronomical origin of the 100 kyr peak, because variations in orbital eccentricity can have little effect on insolation, (e.g. Muller and MacDonald (1997) [11], Ridgwell *et al* (1999) [14], Lisiecki(2010) [9]).

The time series from the EPICA Dome C Ice Core 800 kyr Deuterium Data and Temperature Estimates (Jouzel *et al*, 2007) [8], was downloaded from the World Data Center for Paleoclimatology website. A time series of equally spaced values was generated by averaging all the values in each 653.3 year interval from 490 kyrs to the present epoch (750 samples). It is shown at the top of Figure 1. Its periodogram spectrum is shown beneath and resembles the spectra of Hayes *et al*; the classical three spectral peaks attributed to orbital cycles and listed as 'calculated' in Table 1 can be clearly seen.

Table 1: Orbital Peaks

	Eccentricity	Obliquity	Precession
	e	o	p
calculated	$.01\ kyr^{-1}$	$.0244\ kyr^{-1}$	$.043\ kyr^{-1}$
observed	-	$.024 \pm .001\ kyr^{-1}$	$.043 \pm .001\ kyr^{-1}$

The spectrum has a negative slope for index $k > 5$ and a regression line was fitted in the range $5 \leq k \leq 150$ according to equation (5) and the power law index ν found to be -2.06 ± 0.35, the confidence interval being four times the standard error. For $k \leq 5$ the spectrum is approximately white, the break-point index, k, corresponding to a period $\tau = 100kyr$ implying that $b1 \approx 0.99$ in (7) as will be discussed further below. Also displayed as vertical dashed lines in Figure 2 are the times of ice age Terminations as listed in Table 2 and a plot of solar insolation north of $60°$N due to Berger and Loutre (1991) [2].

In order to test for the statistical significance of the candidate spectral peaks listed in Table 1, a new series of first differences, $\{y_n = x_n - x_{n-1}\}$ was derived and is displayed as the upper graph in Figure 2. Decimated sequences $\{y'_n\}$ and

Table 2: Ice Age Terminations

Termination	Time (kyr BP)	Reference
V	430	Augustin *et al*[1]
IV	340	Cheng *et al* [3]
III	246	Cheng *et al* [3]
II	131	Cheng *et al* [3]
I	12	Cheng *et al* [3]

$\{y''_n\}$ were derived via (10) and (11) and their spectra plotted in the lower graph of Figure 2.

In the lower graph of Figure 2, the scales are linear since now that the power law trend has been removed logarithmic scales are no longer needed; the spectrum is now flat with sporadic peaks. Our aim is to test these peak for significance. As a null hypothesis we assume both decimated sequences, $\{y'_n\}$ and $\{y''_n\}$, are unself-correlated noise implying that each ordinate in their spectra has a χ^2 distribution. There are 120 spectral ordinates in all so that if a 95 percent confidence limit were to be used we would expect it to be exceeded by 40 or so ordinates merely by chance. In this case we used much broader confidence limit, i.e. or a 99.99 percent upper confidence limit ($\alpha = .0001$) computed from the χ^2 distribution with two degrees of freedom. This is shown as the dashed horizontal line in the lower graph of Figure 2.

3.2 Discussion

In Figure 2 (lower) the 'o' and 'p' peaks of the previous figure exceed the confidence limit but the 'e' peak does not. Thus there is strong evidence of cyclical behaviour at the frequencies of the 'o' and 'p' peaks, but there is no evidence of cyclic behaviour at the frequency of the 'e' peak. We may conclude that the 'e' peak is no more than a random walk excursion because it ceases to be significant once the random walk power law trend has been removed.

There are, in fact, no cycles with periods near 100 kyr evident in the time series itself. The last five Terminations have an average period close to 100 kyr but closer examination shows that successive differences between terminations are 90, 94, 115, and 119 kyr respectively. They are either about 90 kyr apart or about 120 kyr apart. Huybers and Wunsch (2005)[7] have shown that the ice sheets terminate every second or third obliquity cycle.

Peaks in the rate of temperature change in Figure 2 (upper) associated with the 4 most recent terminations occurred at times of rapidly increasing Northern Hemisphere insolation as noted by Cheng et al [3]. However not all times of rapidly changing NH insolation gave rise to terminations. This implies that a fully deterministic causality was not the case. Rather, only the *probability* of a termination

is increased at times of rapidly increasing NH insolation. The behaviour of global temperature at these time scales can best be described as a deterministically modulated stochastic process.

A feature of interest in Figure 1 is the almost constant variance density of the low frequency end of the log spectrum. Allowing for the inherent noisiness, for indices, $k < 5$ or so the spectrum could easily be white, i.e. the random walk $\nu = -2$ trend does not apply at low frequencies. Index $k = 5$ corresponds to an integrating time constant of $\tau = 100$kyr. Indeed visual examination of the time series in the upper panel of Figure 1 does indeed suggest a system recovering a stable equilibrium after being driven from stability by an impulse which precipitates terminations. It looks like an integrated Poisson process with an integration time of $\tau \approx 100$ kyr.

4 Application to a present-day climate time series

4.1 Processing the data

GLOBAL Land-Ocean Temperature from GISS, Hansen et al (2010) [4] was downloaded and annual means computed for the years 1880 to 2014 to create the 135 long time series which is plotted in the upper panel of Figure 3. The log spectrum is shown in the lower panel. A power law line was fitted in the range $1 \leq k \leq 21$. The spectrum for $k > 21$ appears white.

As before, first differences were found and are shown in the upper panel of Figure 4. The lower panel shows the linear spectra of the two time series obtained by decimating the first difference series by two, i.e. by forming two sequences of alternating values. Upper confidence limits at the 95 percent and 99 percent levels are shown as horizontal dashed lines.

4.2 Discussion

Only six ordinates exceed the 95 percent confidence level and no peaks are significant at the 99 percent confidence level. One in twenty ordinates can be expected to exceed the 95 percent level purely by chance, i.e. about seven ordinates out of 134. In fact that there are only six implies that there is no significant cyclical behaviour in the GISS temperature data. There is no evidence of a 'multi-decadal oscillation' or any other sort of oscillation; the differenced data is indistinguishable from random noise. Thus there is no evidence that the 'cycles' and upward trend apparent in the GISS global temperature time series shown in the upper panel of Figure 3 are anything other than random walk excursions.

4.3 A numerical experiment

It may be argued that the cycles and upward trend in the GISS time series are real but are too subtle to be detected by the methodology used here.

This is certainly a possibility. In order to test it, synthetic time series were generated with similar spectral characteristics as the GISS time series. It is a simple matter to generate 135 normally distributed random numbers with zero mean and the same variance as the GISS differenced time series. Their cumulative sum should then have the characteristics of a random walk.

However it is more complicated than that, because, as noted, the high frequency end of the spectrum of Figure 3 is nearly white (i.e. flat). This was taken into account by generating two sequences of random numbers, $\{\epsilon_n^r, n = 1, ..., 135\}$ and $\{\epsilon_n^w, n = 1, ..., 135\}$ with variances s_r^2 and s_w^2 respectively such that

$$s_r^2 = 0.15 s_d^2 \quad \text{and} \quad s_w^2 = 0.85 s_d^2 \tag{12}$$

where s_d^2 was the sample variance of the GISS difference series. A new 'red' series $\{y_n\}$ was then formed via

$$y_i = y_{i-1} + \epsilon_i^r \tag{13}$$

Finally, the synthetic time series, $\{z_n\}$, was then formed as

$$z_i = y_i + \epsilon_i^w \tag{14}$$

Six synthetic series were generated in this way are displayed in Figure 5 along with the original GISS time series from Figure 3. The spectra of the synthetic time series were averaged and the mean spectra plotted as the thin curve in the lower panel of Figure 3.

Clearly there is no difference in character between the real GISS time series and synthetic time series having similar spectra and variance. The synthetic time series show similar quasi-cycles and upward or downward trends of similar magnitude to the observed time series.

There is nothing unexpected or unusual about the GISS time series of global average temperatures.

It might be argued that the very 'redness' of the spectrum of Figure 3 is itself indicative of some unusual recent phenomenon and this has, in turn, affected the synthetic series in a similar way.

This is unlikely. Pelletier (2002) [12], using a different methodology from that used here, has shown that global temperature spectra are red or at least 'pink' on every time scale from one day to one million years. There is nothing unusual about the variance spectrum of the GISS time series of global average temperatures.

5 Conclusions

Deterministic cycles and random walk excursions may be distinguished by differencing and decimation of the original data. In this way it has been demonstrated

that, while the obliquity and precession peaks of paleoclimate time series are indeed significant deterministic cycles, the ‘eccentricity’ peak is indistinguishable from a random walk excursion. Likewise the ‘multidecadal oscillation’ of contemporary climate data is not detectable as a significant peak and there is no evidence that the recent rising trend in global temperatures is anything other than a random walk excursion.

References

[1] Augustin, L., Barbante, C., Barnes, P.R.F., Barnola, J.M., Bigler, M., Castellano, E., Cattani, O., Chappellaz, J., Dahl-Jensen, D., Delmonte, B., Dreyfus, G., Durand, G., Falourd, S., Fischer, H., Flckiger, J., Hansson, M.E., Huybrechts, P., Jugie, G., Johnsen, S.J., Jouzel, J., Kaufmann, P., Kipfstuhl, J., Lambert, F., Lipenkov, V.Y., Littot, G.C., Longinelli, A., Lorrain, R., Maggi, V., Masson-Delmotte, V., Miller, H., Mulvaney, R., Oerlemans, J., Oerter, H., Orombelli, G., Parrenin, F., Peel, D.A, Petit, J-R., Raynaud, D., Ritz, C., Ruth, U., Schwander, J, Siegenthaler, U., Souchez, R., Stauffer, B., Steffensen, J.P., Stenni, B., Stocker, T.F., Tabacco, I.E., Udisti, R., van de Wal1, R.S.W., van den Broeke, M., Weiss, J., Wilhelms, F., Winther, J-G., Eric W. Wolff, E.W and Zucchelli, M., 2004. Eight glacial cycles from an Antarctic ice core. *Nature* **429**, 623-628

[2] Berger A. and Loutre M.F., 1991. Insolation values for the climate of the last 10 million years. *Quaternary Sciences Review*, textbf10 297-317.

[3] Cheng, H., Edwards,R. L., Broecker, W. S., Denton, G. H., Kong, X., Wang, Y., Zhang, R. and X. Wang, 2009. Ice Age Terminations. *Science* **326**, 248-252

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

GISTEMP Team, 2015: GISS Surface Temperature Analysis (GISTEMP). NASA Goddard Institute for Space Studies. Dataset accessed 2015-01-21 at `http://data.giss.nasa.gov/gistemp/`.

[5] Hasselmann, K. ,1976. Stochastic climate models. Part I. Theory. *Tellus* **XXVIII**, 6, 473-485.

[6] Hays, D.J., Imbrie, J., Shackleton, N.J., 1976. Variation in the Earths Orbit: Pacemaker of the Ice Ages. *Science*, **194**, 4270, 1121-1132.

[7] Huybers, P., Wunsch, C., 2005. Obliquity pacing of the late Pleistocene glacial terminations. *Nature* **434**, 491-494.

[8] Jouzel, J., Masson-Delmotte V., Cattani, O., Dreyfus, G., Falourd, S., Hoffmann, G.,B. Minster, B., Nouet,J., Barnola, J.M., Chappellaz, J., Fischer, H., Gallet, J.C., Johnsen, S., Leuenberger, M., Loulergue, L., Luethi, D., Oerter, H., Parrenin, H., Raisbeck, G., Raynaud, D., Schilt, A., Schwander, J., Selmo, E., Souchez, R., Spahni, R., Stauffer, B., Steffensen, J.P., Stenni, B., Stocker, T.F., Tison, J.L., Werner, M., Wolff, E.W., 2007. Orbital and Millennial Antarctic Climate Variability over the Past 800,000 Years. *Science*, **317**, 5839, pp.793-797.

Web site (as accessed 13 September 2013) (`www.ncdc.noaa.gov/paleo/icecore/antarctica/domec/domec_epic_data.html`)

[9] Lisiecki, L.E., 2010. Links between eccentricity forcing and the 100,000-year glacial cycle. *Nature Geoscience* **3**, 349 352.

[10] Milankovic, M., 1941. Kanon der Erdbestrahlung und seine Anwendung auf das Eiszeitenproblem. Koniglich Serbische Akademie, Belgrade.

[11] Muller, R.A., MacDonald, G.J., 1997. Spectrum of 100-kyr glacial cycle: Orbital inclination, not eccentricity. *Proc. National Acad. Sciences USA* **94**, 8329-8334.

[12] Pelletier, J.D., 2002. *Proc. National Acad. Sciences USA* **99** suppl. 1, 2546 2553

[13] Reid, J.S., 2015. The periodogram as a consistent estimator. `www.scienceheresy.com/statisticsheresy/periodogram/periodogram.pdf`

[14] Ridgwell, A.J., Watson, A.J., Raymo, M.E., 1999. Is the spectral signature of the 100 kyr glacial cycle consistent with a Milankovitch origin? *Paleoceanography*, **14**, 437-440.

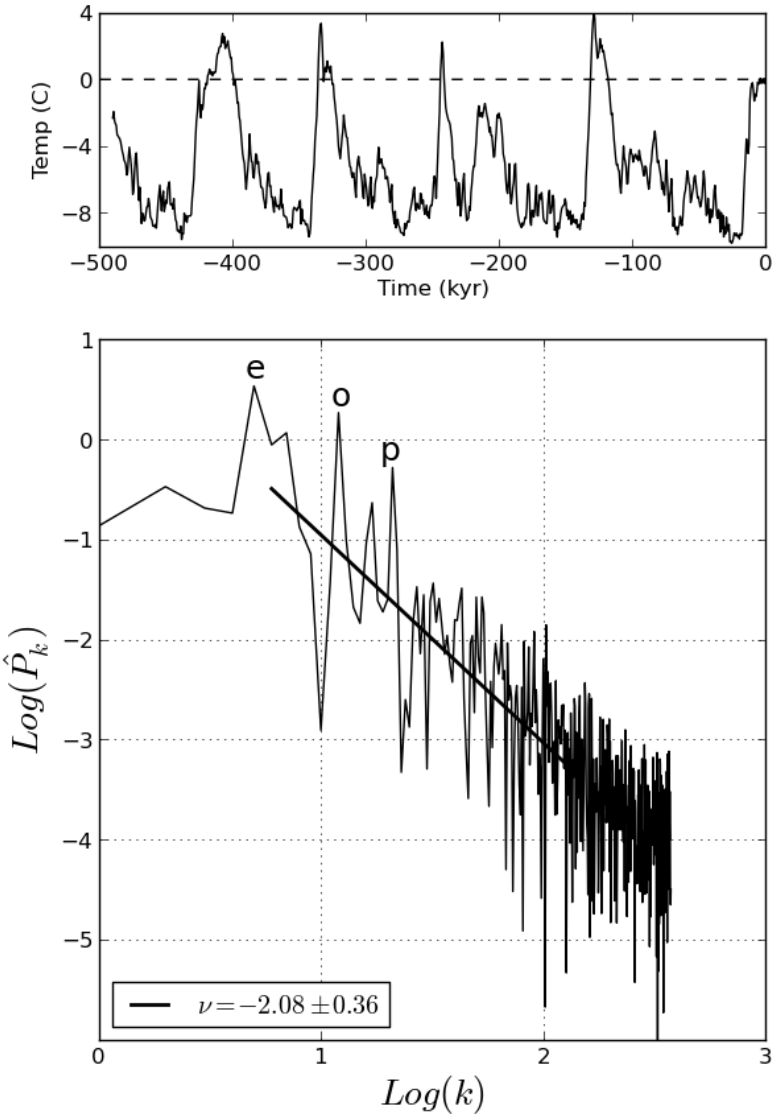

Figure 1. Upper panel: time series of ice core temperatures from EPICA Dome C Deuterium data due to Jouzel *et al* (2007)[8]. Lower panel: periodogram spectrum of this time series plotted using logarithmic scales. Peaks at the eccentricty, obliquity and precession frequencies of the Earth in its orbit are labelled 'e','o' and 'p' respectively. The power law index, ν, is the slope of the trend line as computed by conventional linear regression methods.

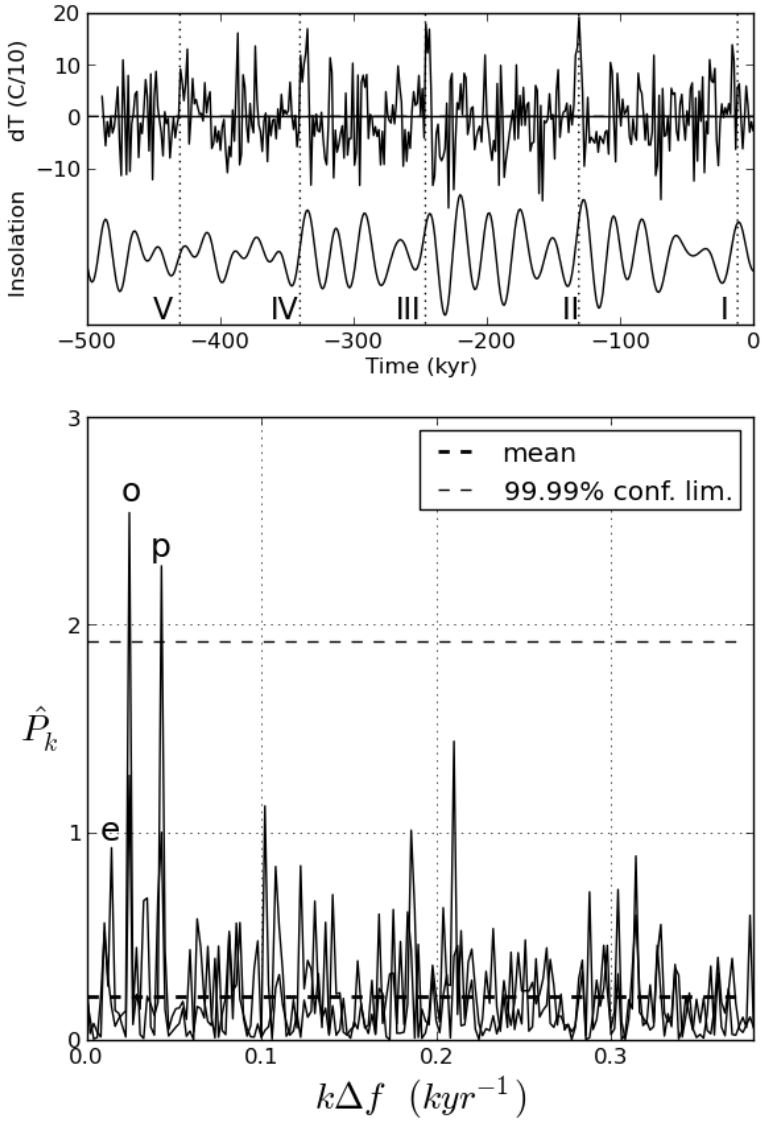

Figure 2. Upper panel: time series of first differences of the time series shown in Figure 1. The vertical dashed lines show times of Terminations I,II,III,IV [3] and V [1]. The lower curver shows insolation north of $60°$N due to Berger (1991)[2]. Lower panel: the two periodogram spectra obtained by decimation of this time series plotted using linear scales. An upper confidence limit of 99.99 percent is shown.

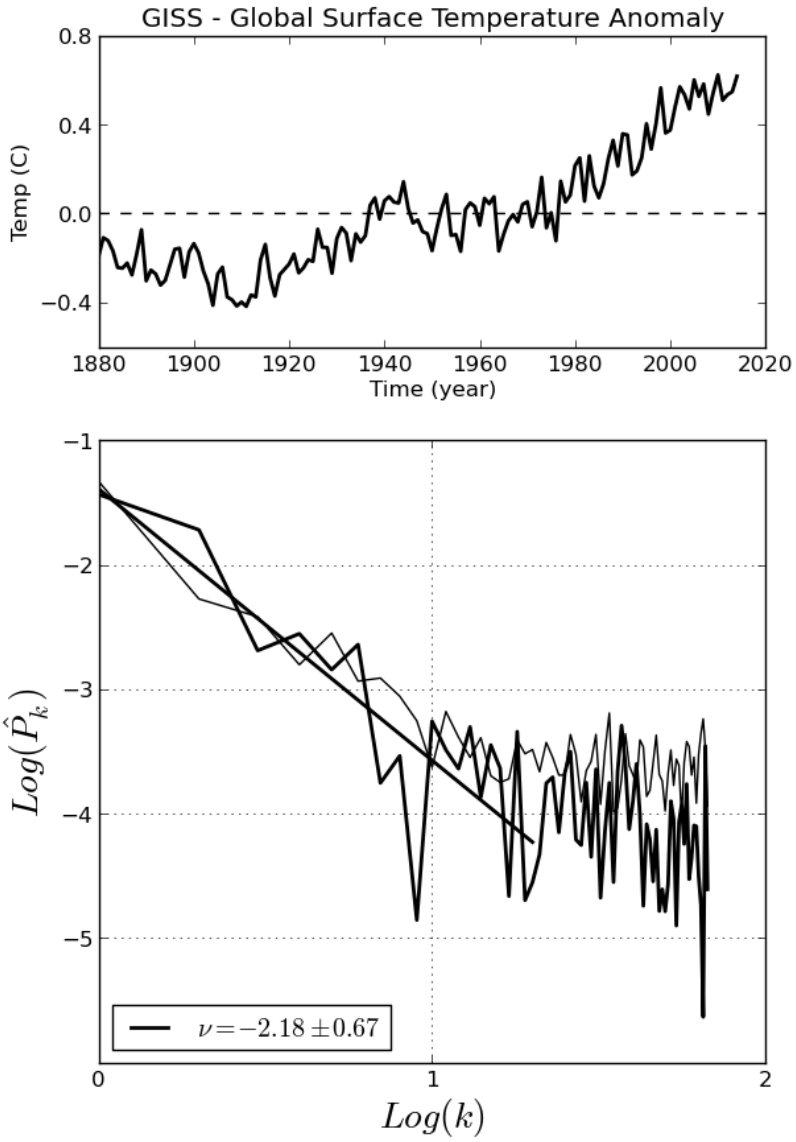

Figure 3. Upper panel: time series of the GISS global average temperature anomaly, 1880-2014, (Hansen *et al*, 2014). Lower panel: the periodogram spectrum of this time series plotted using logarithmic scales (thick line). Also shown is the mean spectrum of the six synthetic series shown in Figure 5 (thin line).

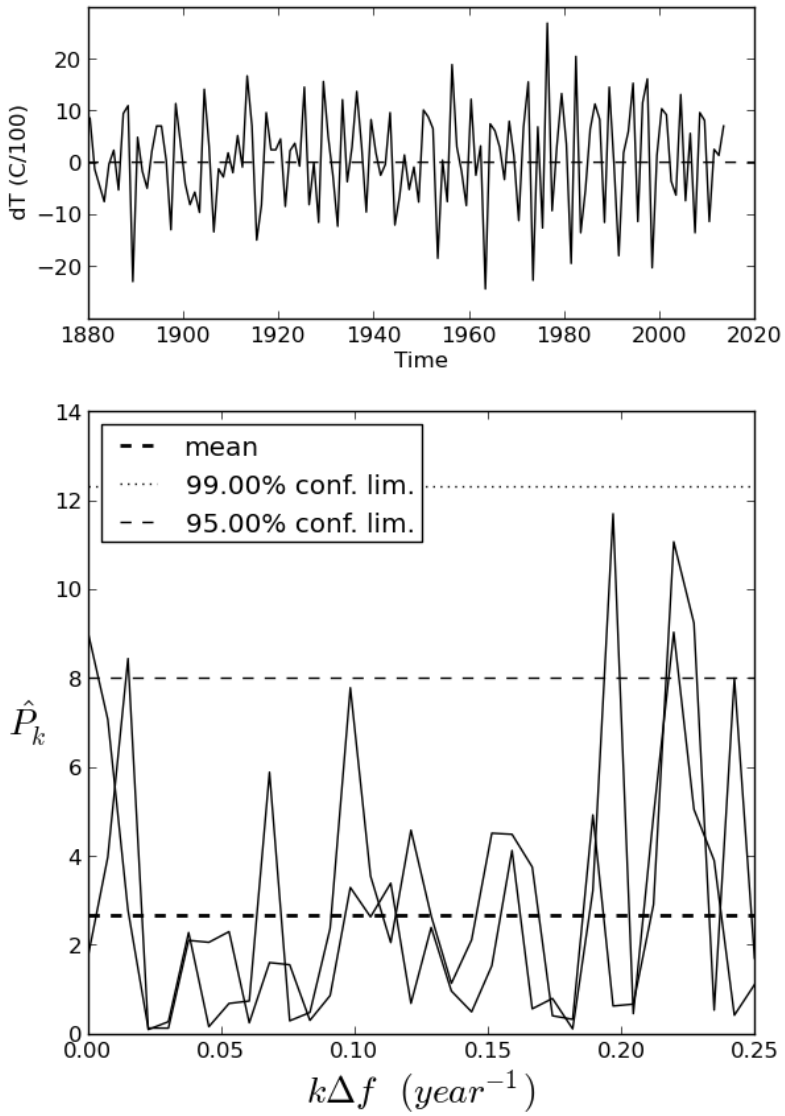

Figure 4. Upper panel: first difference of the time series shown in Figure 3. Lower panel: the two periodogram spectra obtained by decimation of this time series plotted using linear scales. Upper confidence levels of 95 percent and 99 percent are shown.

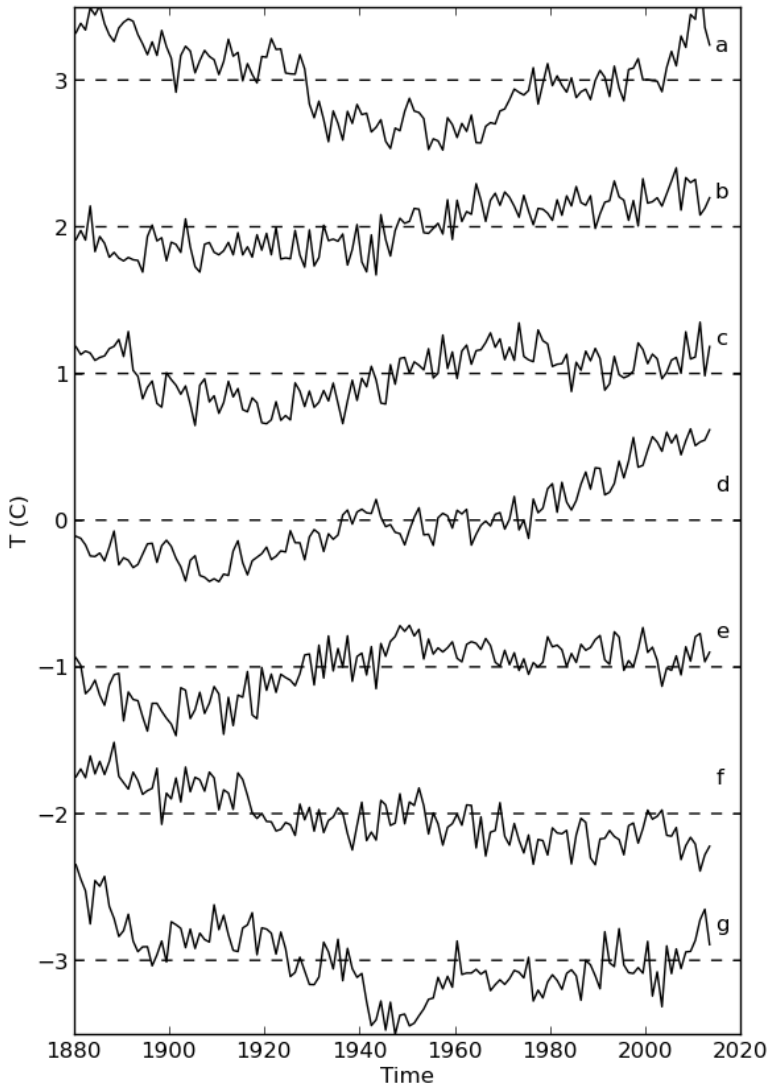

Figure 5. Six synthetic series generated by equation (14) are labelled ‘a’, ‘b’, ‘c’, ‘e’, ‘f’ and ‘g’, while ‘d’ is the original GISS global temperature anomaly time series from Figure 3. The means are displaced vertically for clarity.