European Geosciences Union General Assembly 2012, Vienna, Austria, 22-27 April

Session HS7.5/NP8.3 Hydroclimatic stochastics: EGU2012-993

Spectrum vs Climacogram

Dimitriadis P., Koutsoyiannis D. and Markonis Y., Department of Water Resources and Environmental Engineering, National Technical University of Athens

(e-mail: pandim@itia.ntua.gr, available online: www.itia.ntua.gr)

Abstract

Two common stochastic tools, the spectrum and the climacogram are compared. Using time series from (a) a couple of simple harmonic functions, (b) synthetic data generated using a complex stochastic model, (c) a large-scale paleoclimatic reconstructions and (d) laboratory-scale measurements of turbulent velocity, we estimate the spectra (using fast Fourier transform) and climacograms. Both original and smooth versions of the spectra are used. The spectrum and the climacogram tools are compared to each other giving emphasis to each advantages and disadvantages and also, some questions regarding the interpretation and inference from the above methods, are discussed.

1. Climacogram definition

The Climacogram (Cg) comes from the Greek word climax (which means scale) and is a plot of the standard deviation SD(k) of the mean-aggregated series of the random variable Z versus the aggregated scale k (Koutsoyiannis, 2010):

$$Z_v^{(k)} = \frac{1}{k}\sum_{i=(v-1)k}^{vk} Z_i, \text{ where}$$

Z and Z_v are the random field of interest and the mean aggregated field and v is the vector index of the field indicating location in the field (lag).

The Cg is useful for detecting the long term change (or else dependence, persistence, clustering) of a process. This can be quantified through the Hurst coefficient (H = 1 – slope of the Cg in a log-log plot as scale tends to infinity). For 0<H<0.5 the process is anticorrelated, for 0.5<H<1 the process is correlated (most common case in geophysical processes) and for H=0.5 the process is purely random (zero autocorrelation, thus white noise behaviour). Long-term persistence in natural processes was first discovered by Hurst (1951) while Kolmogorov (1940) mathematically described it, working on self-similar processes in studying turbulence (Koutsoyiannis, 2011). This behaviour is also known as the Hurst phenomenon or Hurst-Kolmogorov (HK) behaviour. A stochastic process with HK behaviour is known as a Hurst-Kolmogorov process (HKp) or Fractional Gaussian noise (fGn).

2. Climacogram estimation

In an HKp the standard statistical estimator of variance is negatively biased. Below, the fGn is introduced and a method to estimate the H having assumed that the process has an HK behaviour (Tyralis & Koutsoyiannis, 2010) . The estimation of H can is done via the minimization of the square error (SE_H) of the empirical ($\gamma^{\prime(k)}$) and true ($\gamma^{(k)}$) variance over scale k.

$$\left(Z_v^{(k)} - \mu\right) =_d \left(\frac{k}{l}\right)^{-A}\left(Z_v^{(l)} - \mu\right), \text{ where}$$

k, l are any aggregation scales of the process,

μ is the mean of the process,

$=_d$ denotes equality in distribution function and

A=(1-H) is the power law exponent.

Tyralis and Koutsoyiannis (2010) method for estimating both H and σ:

$$\mathrm{SE}_{H,\sigma} = \sum_{k=1}^{K'}\left\{\mathrm{E}\left[S^{(k)}\right] - s^{(k)}\right\}^2 / k^p, \text{ where}$$

$$\mathrm{E}\left[S^{(k)}\right] = C(k;H)\gamma_0^{(k)} \text{ and } C(k;H) = \frac{N/k - (N/k)^{2H-1}}{N/k - 1},$$

$s^{(k)}$, the classical sample estimator of variance $S^{(k)}$,

$\gamma^{(k)} = k^{-B}\sigma^2$, the true variance over k, with σ^2, the true variance at k=0 and $B = 2A$,

p and k', coefficients suggested to be 2 and 10,

N, the total number of records and

E[...] denotes the expected value.

3. Energy spectrum definition

- The Fourier transform (FT) $F(w)$ (where w is usually called frequency) of a function $f(r)$ and its inverse are defined as:

$$F(w) = \int_{-\infty}^{+\infty} f(r)\, e^{-2\pi w r i}\, dr \text{ and } f(r) = \int_{-\infty}^{+\infty} F(k)\, e^{2\pi w r i}\, dk$$

- The energy spectrum (ES) $E_s(w)$ can be linked to the derivative of a system's power $E_k(s)$, with units that of $R(s)$, over frequency w:

$$E_k(s,t,T) = \int_{-T/2}^{+T/2} x^2(s,t)\, dw \underset{T\to\infty}{\overset{dT\to 0}{\leftrightarrow}} \int_{1/T}^{1/dT} E_s(w)\, dw, \text{ where } T \text{ is the length of the signal.}$$

- The autocovariance function $R(t,s)$ of a random variable $x(t)$ is defined below and has the same units as the $x(t)^2$:

$$R(t,s) = \mathrm{E}\left\{\left(x(t) - \mu_x\right)\left(x(t+s) - \mu_x\right)\right\}, \text{ where}$$

μ_x is the mean value of x and

t, s are the time vector and time-lag [T], repsectively.

Here, a stationary process of $x(t)$ and $y(t)$ is considered and thus, $R(t,s)$ depends only on s and is written as $R(s)$.

Note that the Cg is related to the autocovariance function $R(s)$ of a stationary process, through a second derivative (Koutsoyiannis 2010):

$$\text{For continuous time } R(s) = \frac{d^2\left[s^2\left[\gamma^{(s)}\right]^2\right]}{ds^2} \text{ and descrete, } R(s) = \frac{\Delta^2\left[s^2\left[\gamma^{(s)}\right]^2\right]}{\Delta s^2}.$$

4. Energy spectrum estimation

- The $E_s(w)$ of a temporal stationary process is the FT of its $R(s)$ and that it can also be expressed in terms of its FT, $F(w)$:

$$E_s(w) = \int_{-\infty}^{+\infty} R(r)\, e^{-2\pi w r i}\, dr \text{ and } E_s(w) \underset{T\to\infty}{\overset{dT\to 0}{\leftrightarrow}} \left|F(w)\right|^2 / T, \text{ where } \left|F(w)\right| \text{ is the magnitude of the FT.}$$

- The Energy spectrum (ES) is useful for detecting frequencies with high energy and energy dissipation patterns. Usually a series of observations has high fluctuations. This 'noise' is always transferred to its ES, as the FT and its inverse are bidirectional relationships. To smooth out noise, mathematical tools such as the following are used:

(a) Trend-lines, where the ES of data is fitted with a least-squares curve.

(b) Averaging methods (e.g. Welch's), where the signal is divided into segments (with or without overlapping), sometimes also applying a window-function (e.g. Bartlett's) and estimating the ES for each segment; the average of all segments' ES is then calculated.

(c) Methods of wavelet analysis, where the data are decomposed onto frequency spaces and then reconstructed using translated and scaled versions of a mother wavelet (filter).

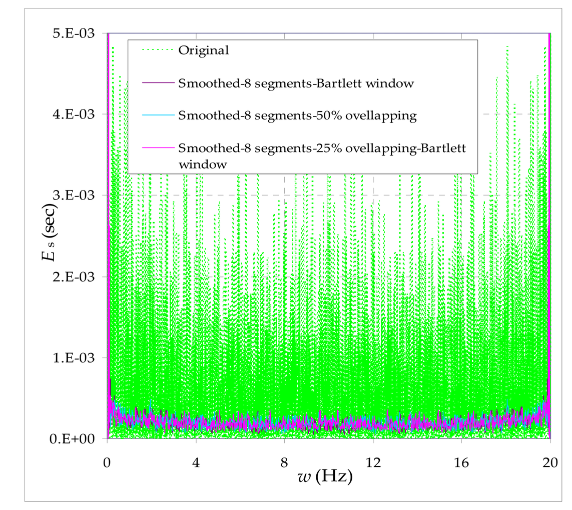


Figure 1: Energy spectrum of velocity data with different smoothing methods of type (b).

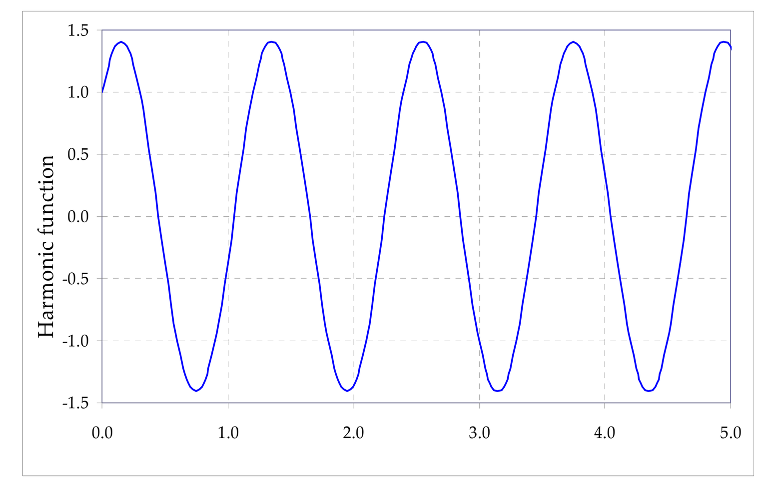
5. Simple deterministic harmonic function

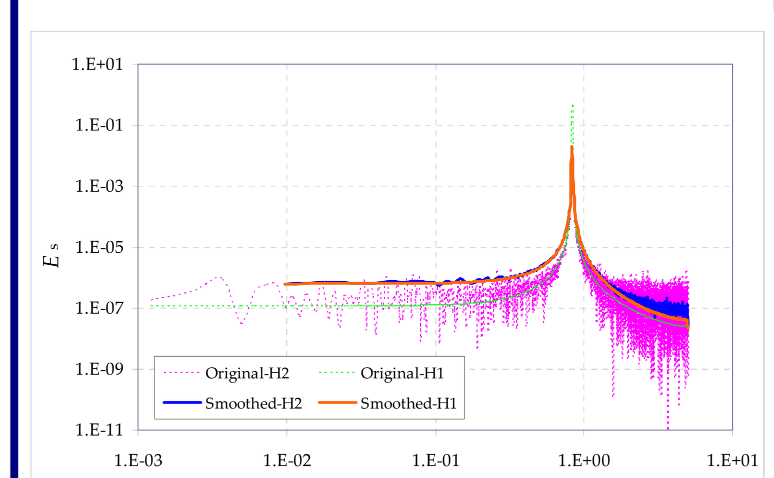
A harmonic function $f(t)$ that combines cosine and sine components is presented (series T1):

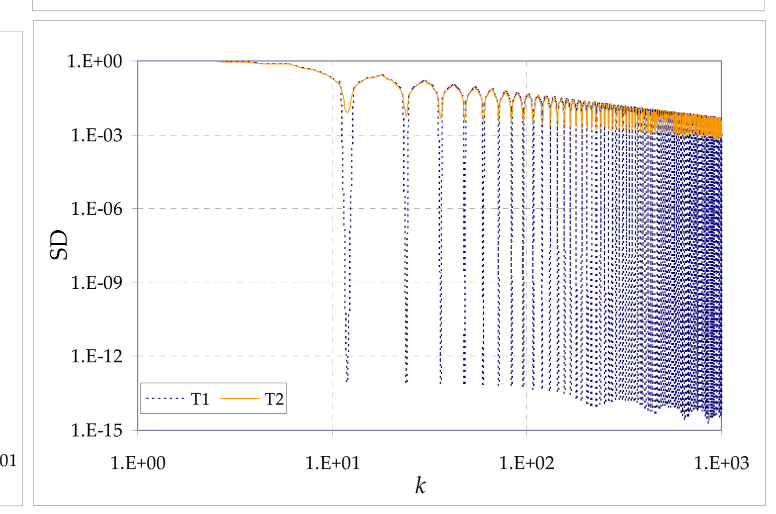
$f(t)$ = sin(2πt/1.2) + cos(2πt/1.2)

The total length of observation is 1200 units and the time step Δt is 0.1.

Moreover, random noise is added to T1 obtaining series T2, in order to see how the Cg and ES will behave.

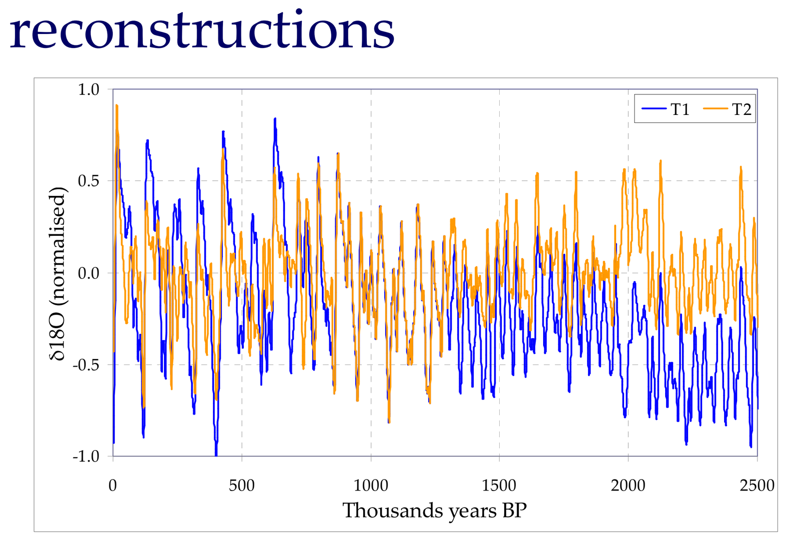


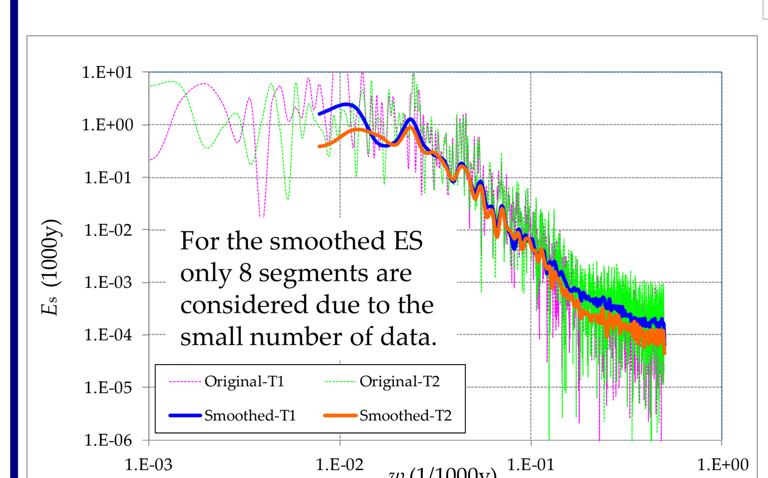


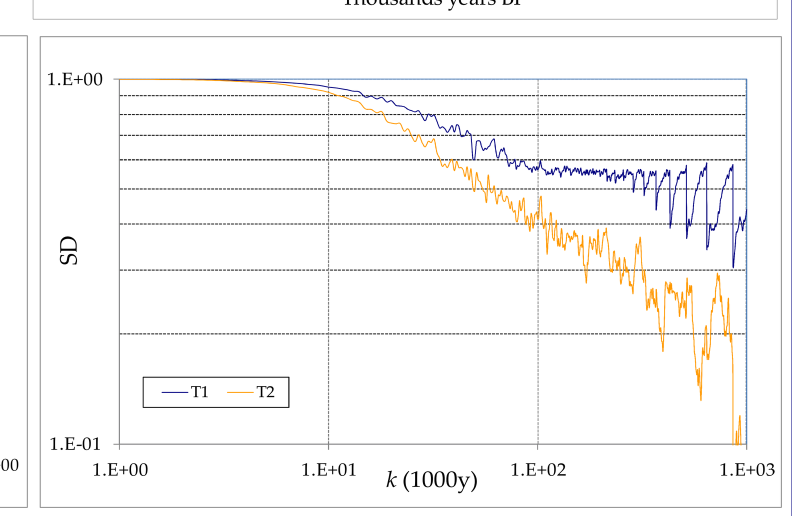


6. Large-scale paleoclimatic reconstructions

Large-scale paleoclimatic reconstruction (series T1) of global temperature is analyzed. The data are based on sediment proxy data ($\delta^{18}O$) from different locations at the Atlantic Ocean, with time resolution 1 thousand years and total length of 2.5 million years (Huybers, 2007). Also, the Huybers' reconstruction, after the harmonics corresponding to Milankovitch cycles have been removed (Markonis and Koutsoyiannis, 2010) is analyzed (series T2).

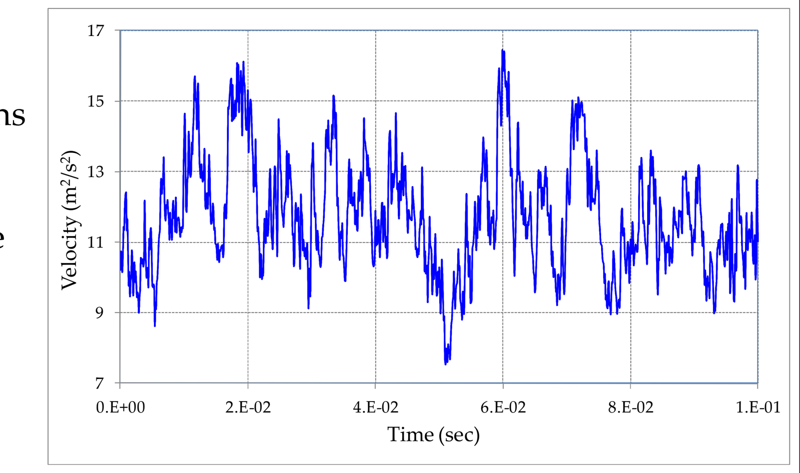


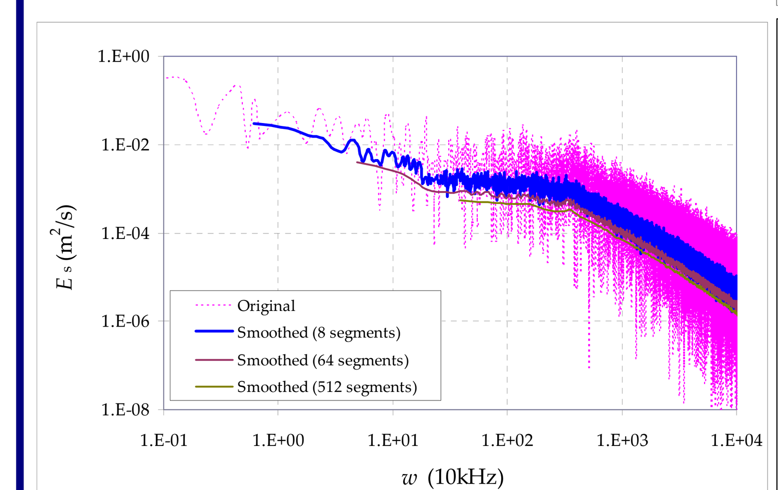


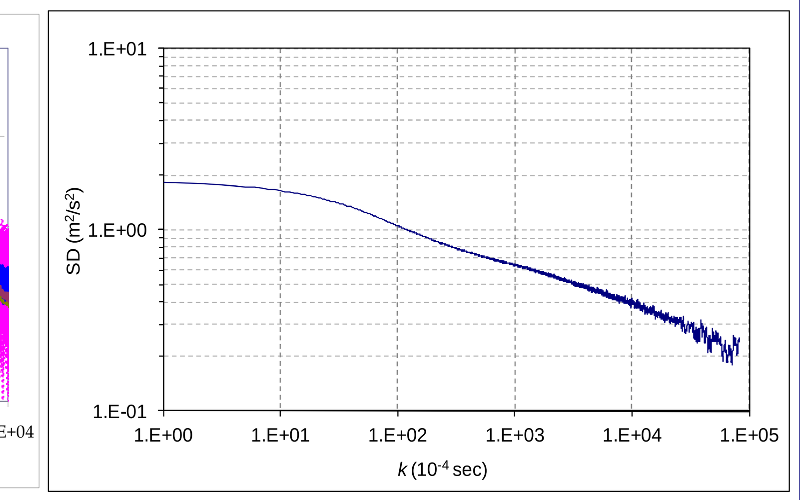


7. Laboratory-scale measurements

Timeseries shown here is from an open access dataset (http://www.me.jhu.edu/meneveau/datasets/datamap.html), provided by the Johns Hopkins University, that consists of nearly isotropic and homogeneous turbulent wind streamwise velocity data, measured by X-wire probes downstream of an active grid (Kang et al., 2002). A large range of scales is analyzed due to the fact that this dataset consists of 900000 records.

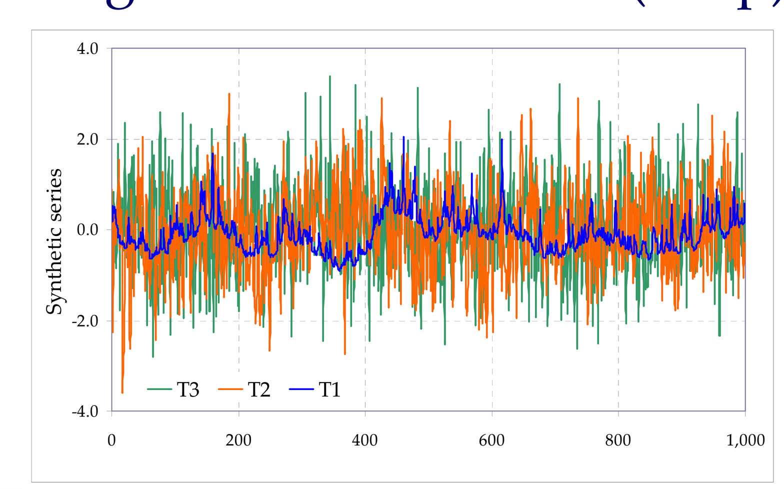


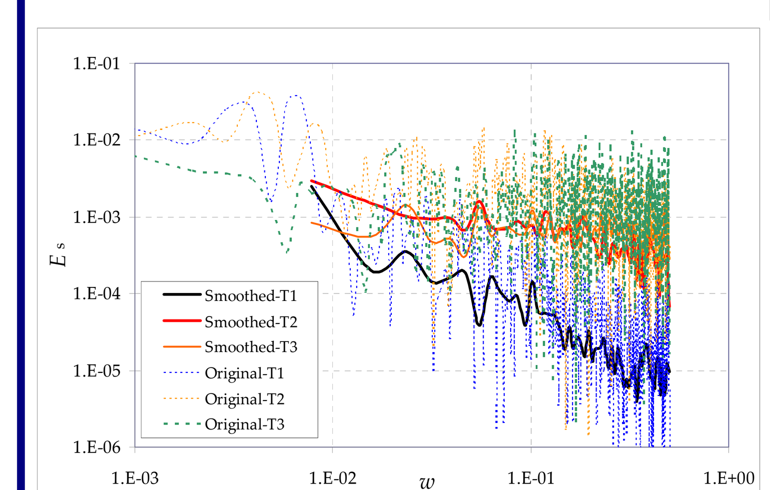


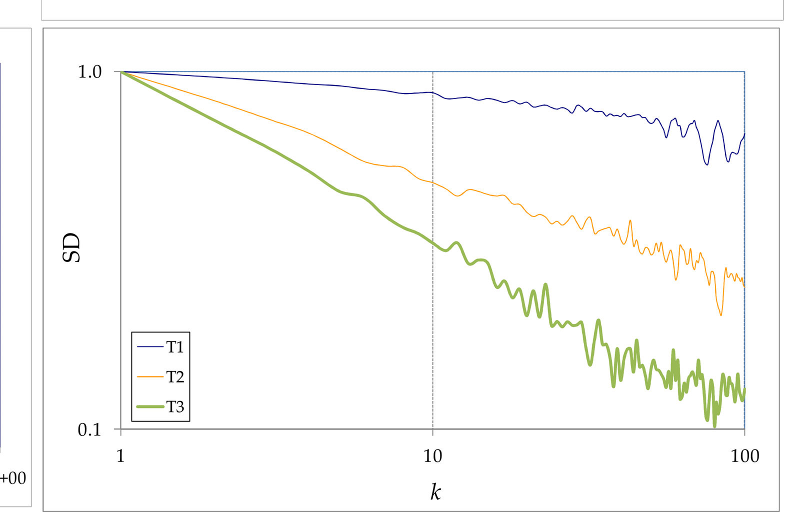


8. Synthetic data generated using a stochastic model (HKp)

Here, synthetic series are produced through the HKp, with H=0.95 (series T1), 0.68 (series T2) and 0.52 (series T3). The length of the series is only 1000 records as the purpose of this analysis is to observe how well can the Cg and ES represent the long-term persistence. The chosen range of the H corresponds to positively correlated processes (which is the most common case in geophysical processes).







9. Conclusions

1) There are many differences between ES calculated through smoothing techniques and original ES (e.g. figure 1) due to loss of information (during the smoothing process).
2) From the analysis of section 5, it can be concluded that the frequency which contains most of the energy, is apparent in the ES. Nevertheless, energy residuals in other frequencies still remain due to the numerical methods used. This may give the wrong impression of an energy dissipation rate. On the other hand, the Cg can give some information about the period of the harmonic function without any numerical residuals.
3) As the number of segments increases (e.g. section 7) the range of available frequencies decreases, thus information is being lost.
4) The long term change is different between T1 and T2 series of section 6. This result can be easily derived from the Cg but difficult to observe in the ES. Moreover, the ES fails to detect some changes in the long-term persistence, between the low and high correlated T1 and T3 synthetic series of section 8.

References

- Hurst, H.E., 'Long term storage capacities of reservoirs', *Trans. ASCE*, Vol. 116, p. 776-808, 1951.
- Kolmogorov, A.N., 'Wiener's spiral and some interesting curves in Hilbert space', *Dokl. Akad. Nauk URSS*, Vol. 26, p. 115–118, 1940.
- Tyralis H. and D. Koutsoyiannis, 'Simultaneous estimation of the parameters of the Hurst-Kolmogorov stochastic process', *Stochastic Environmental Research & Risk Assessment*, 2010.
- Markonis, Y. and Koustoyiannis, D., 'Orbital climate theory and Hurst-Kolmogorov dynamics', *11th International Meeting on Statistical Climatology, Edinburgh, Scotland,* 2010.
- Huybers P., 'Glacial variability over the last two million years: an extended depth-derive agemodel, continuous obliquity pacing, and the Pleistocene progression', *Quat. Sci. Rev.*, Vol. 26, p. 37-55, 2007.
- Kousoyiannis D., 'A random walk on water', *Hydrol. Earth Syst. Sci.*, Vol. 14, p. 585–601, 2010.