European Geosciences Union General Assembly 2013
Vienna, Austria, 7-12 April 2013

Session NP3.1: Scaling in atmospheric, terrestrial, oceanic and biogeophysical processes

Climacogram-based pseudospectrum: a simple tool to assess scaling properties

Demetris Koutsoyiannis

Department of Water Resources and Environmental Engineering

School of Civil Engineering

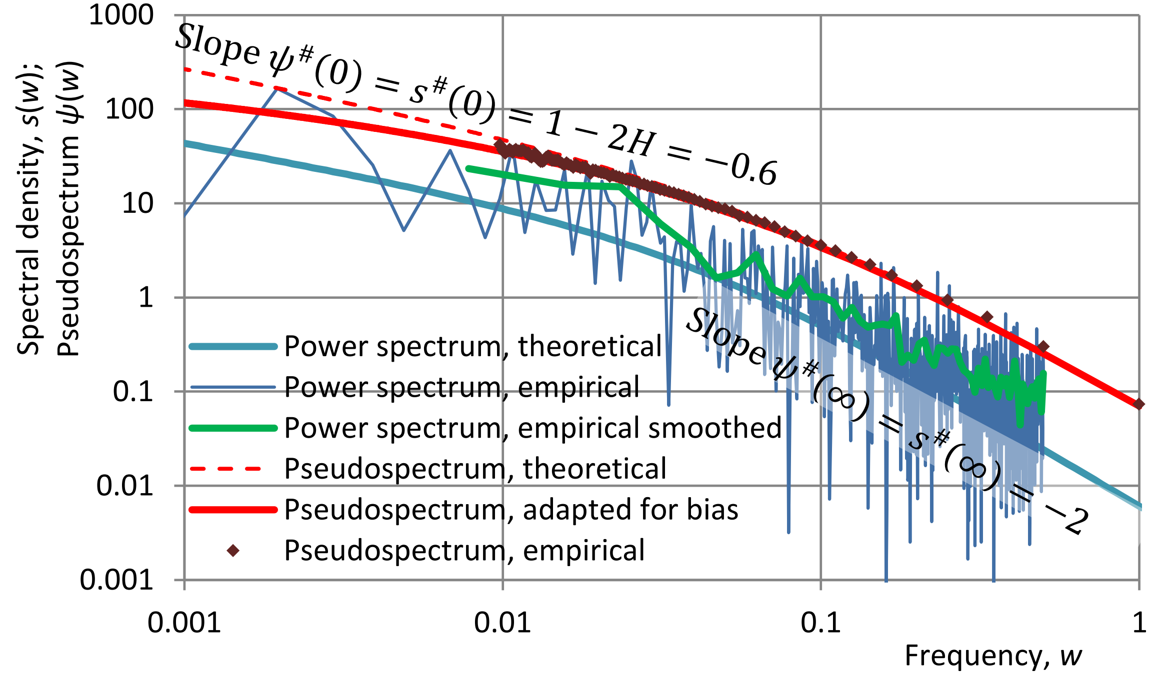
National Technical University of Athens, Greece

(dk@itia.ntua.gr, itia.ntua.gr/dk/)

Presentation available online: **itia.ntua.gr/1328/**

A first illustration

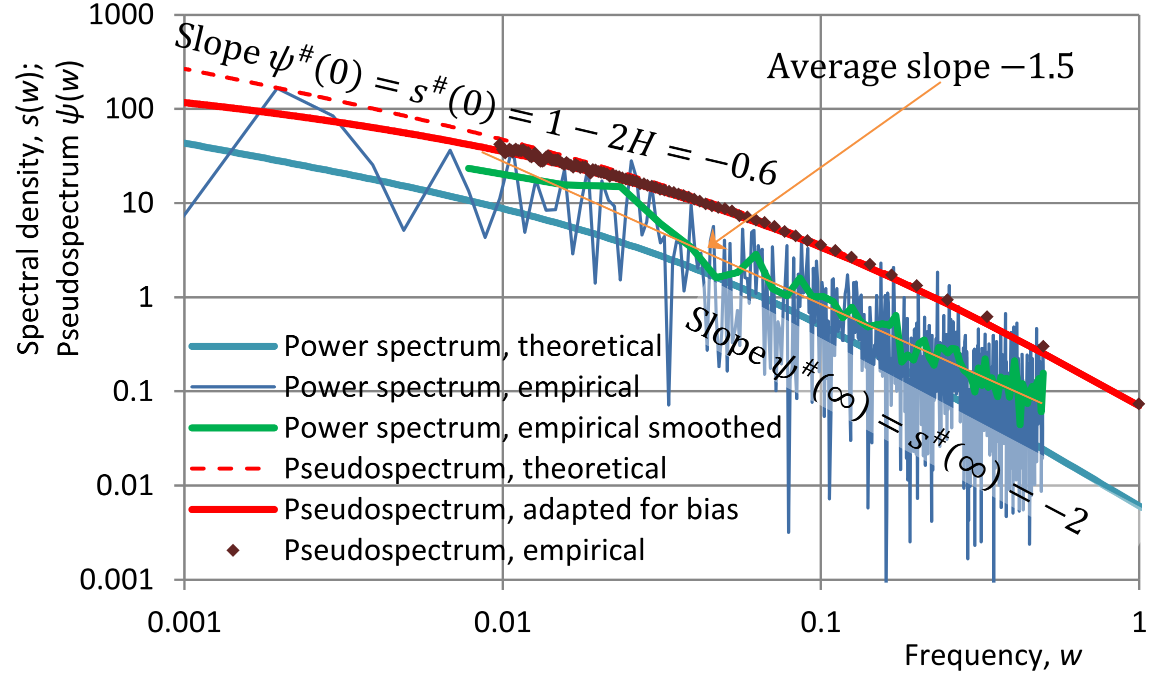
- As a first example, we consider the stochastic process with known theoretical properties, including its theoretical power spectrum, as shown on the graph.
- The process is characterized by two different scaling laws, shown in its theoretical power spectrum as asymptotic slopes for frequencies $w \to 0$ and $w \to \infty$.



- The slopes can be deduced if the stochastic properties of the process are known. But can they be estimated from data?
- Here a time series of 1024 values has been generated from the known process.
- The graph, in addition to theoretical (true) and empirical (estimated) power spectra, shows theoretical and empirical pseudospectra (explained below).

Problems in estimation of the power spectrum

- If estimated from data (the Fourier transform of the data series or its empirical autocorrelation function), the power spectrum is too rough.
- Even after smoothing (here by averaging from 8 segments) it remains too rough and inappropriate to estimate either asymptotic slopes or statistically significant peaks.



- The bias and uncertainty in estimation are uncontrollable.
- Finite sample and time discretization also cause problems in the estimation of theoretical spectrum; for example at the Nyquist frequency ($1/2D$) the calculated slope is precisely 0 ($s^{\#}(½) = 0$) and not equal to the actual asymptotic slope.
- Due to these problems, erroneous results are often reported in the literature, e.g. too steep slopes, $s^{\#}(0) < -1$ (infeasible; see Koutsoyiannis 2013), and false periodicities.

The new concept of the *climacogram-based pseudospectrum* can overcome such problems.

The empirical climacogram

- As an example, we consider the synthetic time series of the earlier illustration,

$$x_1, x_2, \dots, x_{1024} \rightarrow \hat{\gamma}(1) = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_{1024} - \bar{x})^2}{1023} \quad (1)$$

  where $\hat{\gamma}(1)$ is the sample variance whereas the argument (1) indicates time scale 1 and $\bar{x} := (x_1 + \ x_2, + \cdots + x_{1024})/1024$ is the sample average.
- We form a time series at time scale 2 and find its variance:

$$x_1^{(2)} := \frac{x_1 + x_2}{2}, x_2^{(2)} := \frac{x_3 + x_4}{2}, \dots, x_{331}^{(2)} := \frac{x_{1023} + x_{1024}}{2} \rightarrow \hat{\gamma}(2) \quad (2)$$

- We proceed forming a time series at time scale 3 and finding its variance:

$$x_1^{(3)} := \frac{x_1 + x_2 + x_3}{3}, \dots, x_{341}^{(2)} := \frac{x_{1021} + x_{1022} + x_{1023}}{3} \rightarrow \hat{\gamma}(3) \quad (3)$$

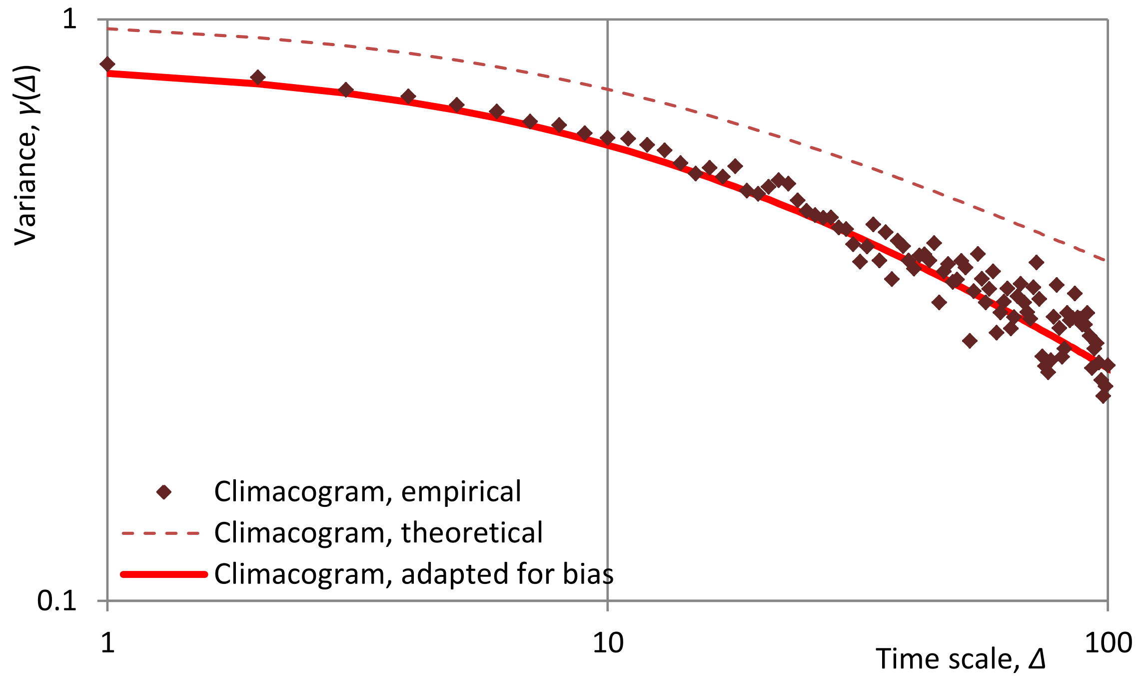
- We repeat the same procedure up to scale 102 = 1/10 of the sample size (Koutsoyiannis 2003; for larger scales the estimation is too unreliable):

$$x_1^{(102)} := \frac{x_1 + \cdots + x_{102}}{102}, \dots, x_{10}^{(102)} := \frac{x_{919} + \cdots + x_{1020}}{102} \rightarrow \hat{\gamma}(102) \quad (4)$$

- The empirical *climacogram* (Koutsoyiannis, 2010) is the logarithmic plot of the variance $\hat{\gamma}(\Delta)$ versus the time scale Δ (or that of the standard deviation $\hat{\sigma}(\Delta) = \sqrt{\hat{\gamma}(\Delta)}$ vs. Δ; a contraction of the former logarithmic plot by 2).

The theoretical climacogram

- For a stochastic process $\underline{x}(t)$ at continuous time t, the averaged process on time scale Δ at discrete time i is
  $\underline{x}_i^{(\Delta)} := \frac{1}{\Delta}\int_{(i-1)\Delta}^{i\Delta}\underline{x}(\xi)\mathrm{d}\xi$ (5)
- The theoretical climacogram is the variance
  $\gamma(\Delta) \coloneqq \mathrm{Var}\left[\underline{x}_i^{(\Delta)}\right]$ (6)
- In our example,
  $\gamma(\Delta) = \frac{\lambda}{(1+\Delta/\alpha)^{2-2H}}$ (7)
  where α = 1 [time], λ = 1 $[x]^2$ and H = 0.8 [-] (Hurst parameter).



- The empirical climacogram $\hat{\gamma}(\Delta)$ is an estimate of the theoretical one $\gamma(\Delta)$, but not an unbiased one. The bias is calculated from the model properties:

$$\mathrm{E}[\hat{\gamma}(\Delta)] = \eta(\Delta, T)\gamma(\Delta) \text{ where } \eta(\Delta, T) = \frac{1-\gamma(T)/\gamma(\Delta)}{1-\Delta/T} \quad (8)$$

where $T \coloneqq nD$, n the sample size and D the spacing (Koutsoyiannis, 2011).

Relationships of climacogram, autocovariance and power spectrum

- The climacogram, the autocovariance function and the power spectrum of a process are transformations one another.
- The climacogram $\gamma(\Delta) \coloneqq \mathrm{Var}\left[\underline{x}_i^{(\Delta)}\right]$ and the autocovariance function $c(\tau) \coloneqq \mathrm{Cov}[\underline{x}(t), \underline{x}(t+\tau)]$ of a continuous time process are interrelated as follows:

$$\gamma(\Delta) = 2\int_0^1 (1-\xi) c(\xi\Delta) \mathrm{d}\xi \;\leftrightarrow\; c(\tau) = \frac{1}{2}\frac{\mathrm{d}^2\left(\tau^2\gamma(\tau)\right)}{\mathrm{d}\tau^2} \tag{9}$$

- The power spectrum $s(w)$ and the autocovariance function $c(\tau)$ of a continuous time process are interrelated as follows:

$$s(w) \coloneqq 4\int_0^\infty c(\tau)\cos(2\pi w\tau)\,\mathrm{d}\tau \;\leftrightarrow\; c(\tau) = \int_0^\infty s(w)\cos(2\pi w\tau)\,\mathrm{d}w \tag{10}$$

- The slope of the logarithmic plot of power spectrum, which is of particular interest in identifying scaling properties, is defined as:

$$s^{\#}(w) = \frac{\mathrm{d}(\ln s(w))}{\mathrm{d}\,(\ln w)} = \frac{w\,s'(w)}{s\,(w)} \tag{11}$$

See details in Koutsoyiannis (2013).

The climacogram-based pseudospectrum (CBPS)

- A substitute of the power spectrum which has similarities in its properties, is the climacogram-based pseudospectrum (CBPS) defined as

$$\psi(w) \;:= \frac{2\,\gamma(1/w)}{w}\left(1 - \frac{\gamma(1/w)}{\gamma(0)}\right) \tag{12}$$

- In processes with infinite variance ($\gamma(0) = c(0) = \infty$) the CBPS simplifies to

$$\psi(w) = \frac{2\,\gamma(1/w)}{w} \tag{13}$$

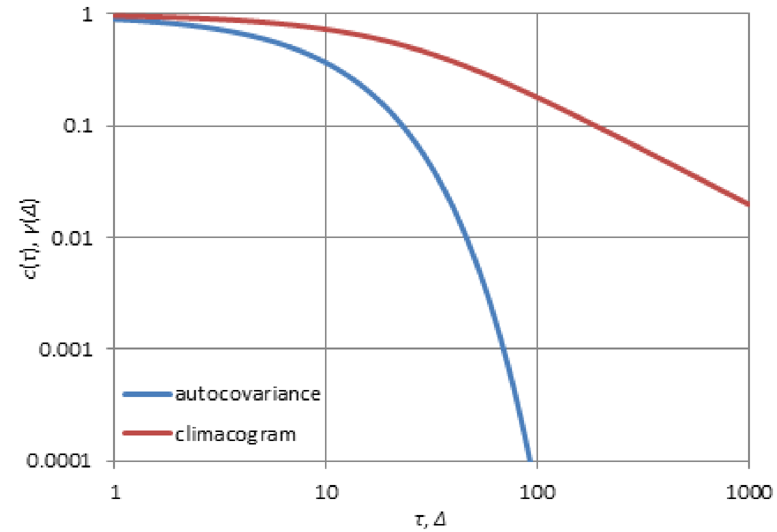
- The CBPS value of at $w = 0$ equals that of the power spectrum (indeed from (9) and (10) we obtain $\psi(0) = s(0) = \Delta\,\gamma(\Delta)|_{\Delta\to\infty} = 4\int_0^\infty c(\tau)\mathrm{d}\tau$).
- Furthermore, the asymptotic slopes $\psi^{\#}(w)$ of CBPS at frequencies (or resolutions) $w \to 0$ and ∞ follow those of the power spectrum $s^{\#}(w)$ and in most processes the asymptotic slopes are precisely equal to each other.
- At frequencies where the power spectrum has peaks, the CBPS has troughs (negative peaks).
- In contrast to the empirical periodogram, the empirical $\psi(w)$ is pretty smooth.

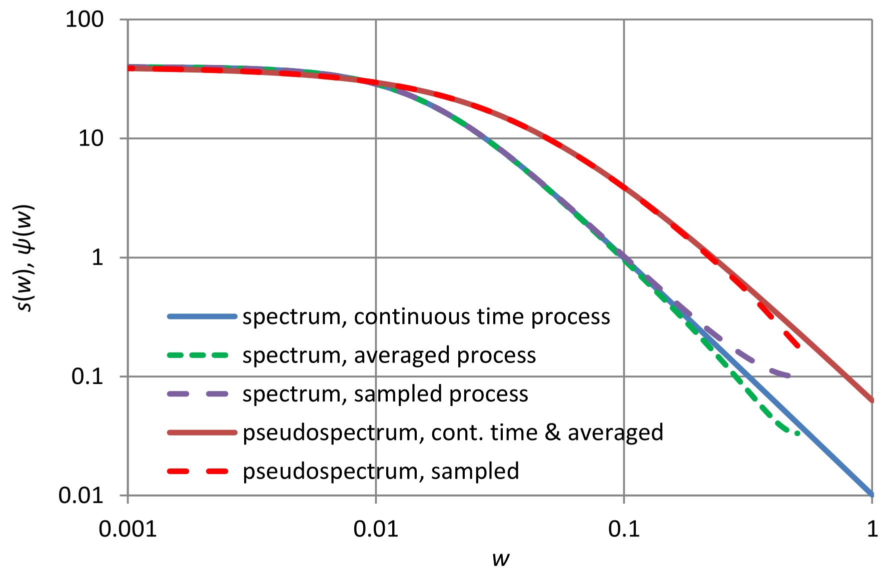
See details in Koutsoyiannis (2013).

Example 1: The Markov process

Variance of instantaneous process	$\gamma = \gamma(0) = c(0) = \lambda$
Variance at scale Δ (Climacogram)	$\gamma(\Delta) = \frac{2\lambda}{\Delta/\alpha}\left(1 - \frac{1-e^{-\Delta/\alpha}}{\Delta/\alpha}\right)$
Autocovariance function for lag τ	$c(\tau) = \lambda \mathrm{e}^{-\tau/\alpha}$
Power spectrum for frequency w	$s(w) = \frac{4\alpha\lambda}{1+(2\pi\alpha w)^2}$
Asymptotic slopes	$\psi^{\#}(0) = s^{\#}(0) = 0$ $\psi^{\#}(\infty) = s^{\#}(\infty) = -2$ $\gamma^{\#}(\infty) = 2\sigma^{\#}(\infty) = -1$ $\gamma^{\#}(0) = \sigma^{\#}(0) = 0$
Parameter values used	$\lambda = 1$, $\alpha = 10$, and the spacing is $D = 1$, resulting in $\rho = 0.905$.

- The theoretical power spectra of derived discrete-time processes (discretized either by averaging at a time scale D or by sampling at spacing D) fail to capture the slopes for $w > 1 / 10D$, while for $w = 1 / 2D$ they give a slope which is precisely zero.
- The pseudospectrum performs better in identifying the asymptotic slopes.

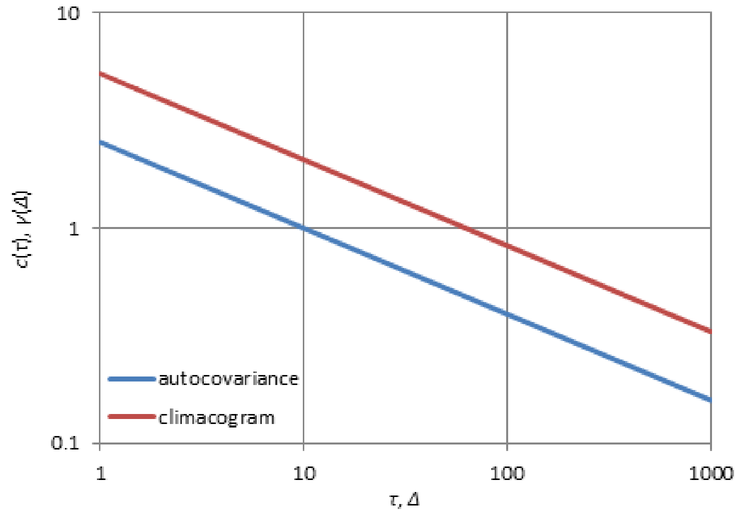


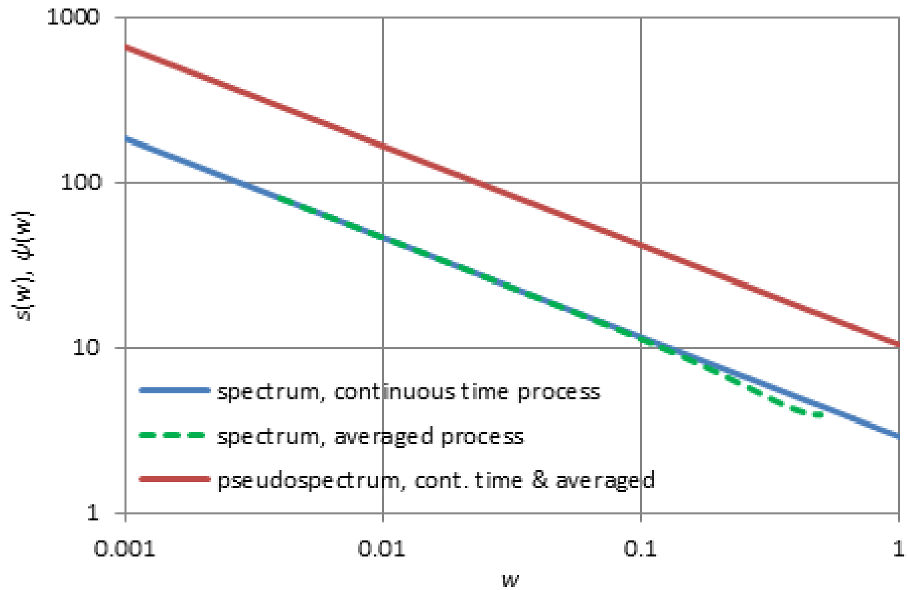


Example 2: The Hurst-Kolmogorov (HK) process

Variance of instantaneous process	$\gamma = \gamma(0) = c(0) = \infty$
Variance at scale Δ (Climacogram)	$\gamma(\Delta) = \frac{\lambda(\alpha/\Delta)^{2-2H}}{H(2H-1)}$
Autocovariance function for lag τ	$c(\tau) = \lambda(\alpha/\tau)^{2-2H}$ $(0.5 \le H < 1)$
Power spectrum for frequency w	$s(w) = \frac{4\alpha\lambda\,\Gamma(2H-1)\sin(\pi H)}{(2\pi\alpha w)^{2H-1}}$
Asymptotic slopes	$\psi^{\#}(w) = s^{\#}(w) = 1 - 2H$ $\gamma^{\#}(\Delta) = 2\sigma^{\#}(\Delta) = 2H - 2$
Parameter values used	$\lambda = 1$, $\alpha = 10$, $H = 0.8$; the spacing is $D = 1$.

- The model parameters are in essence two, i.e. H and $(\lambda\, \alpha^{2-2H})$. Here the formulation has three nominal parameters for dimensional consistency: the units of α and λ are $[\tau]$ and $[x]^2$, respectively, while H is dimensionless.
- For $0.5 \le H < 1$ the process is called a persistent process; it has often been used with $0 < H < 0.5$, being called an antipersistent process, but this is inconsistent with physics (A proper antipersistent process is discussed in Example 4).

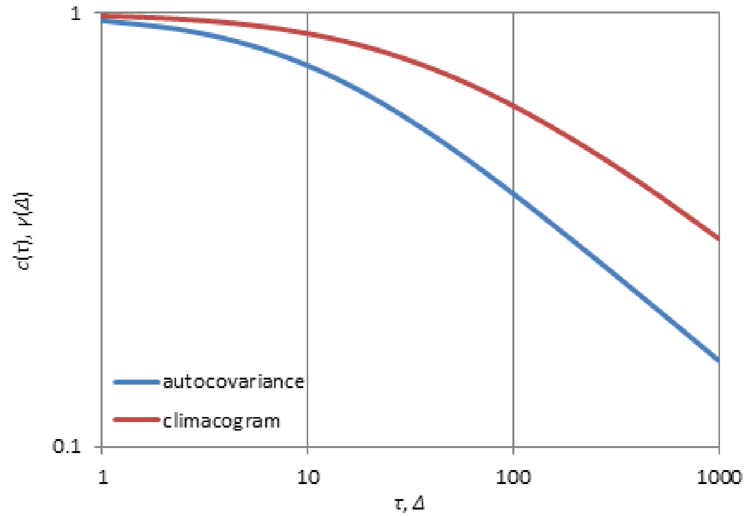


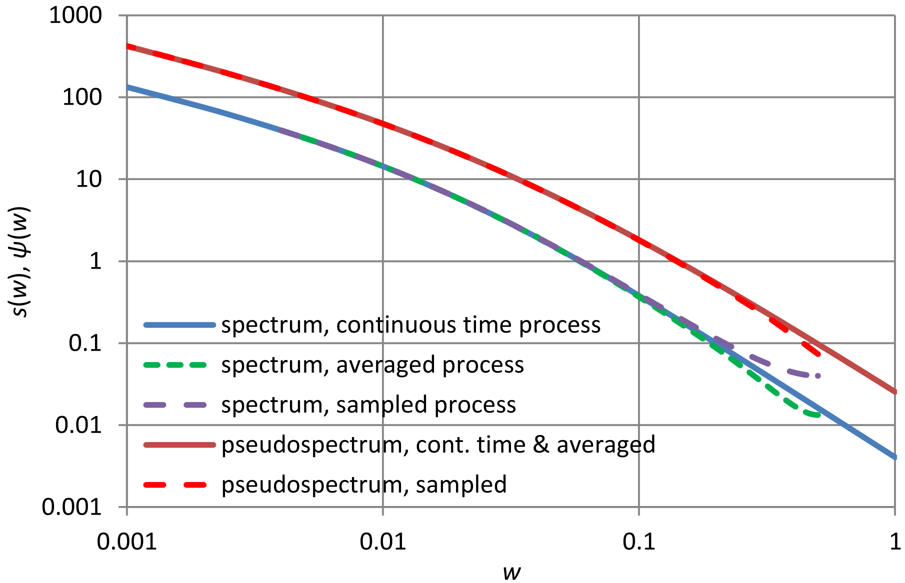


Example 3: A modified finite-variance HK process

Variance of instantaneous process	$\gamma = \gamma(0) = c(0) = \lambda$
Variance at scale Δ (Climacogram)	$\gamma(\Delta) = \frac{\lambda(\alpha/\Delta)}{H(2H-1)} \times \left(\frac{\alpha}{\Delta}\left(1+\frac{\alpha}{\Delta}\right)^{2H} - \frac{\alpha}{\Delta} - 2H\right)$
Autocovariance function for lag τ	$c(\tau) = \frac{\lambda}{(1+\tau/\alpha)^{2-2H}}$
Power spectrum for frequency w	$s(w)$: closed expression too complex
Asymptotic slopes	$\psi^{\#}(0) = s^{\#}(0) = 1 - 2H$ $\psi^{\#}(\infty) = s^{\#}(\infty) = -2$ $\gamma^{\#}(\infty) = 2\sigma^{\#}(\infty) = 2H - 2$ $\gamma^{\#}(0) = \sigma^{\#}(0) = 0$
Parameter values used	$\lambda = 1$, $\alpha = 10$, $H = 0.8$; the spacing is $D = 1$.

- The asymptotic slopes of the power spectrum are both nonzero and different in the cases $w \to 0$ and $w \to \infty$. The slopes of the pseudospectrum are identical with those of the spectrum.

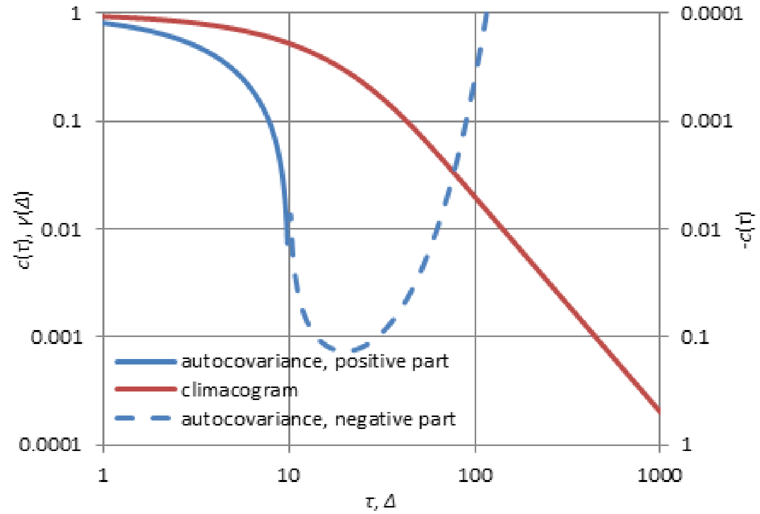


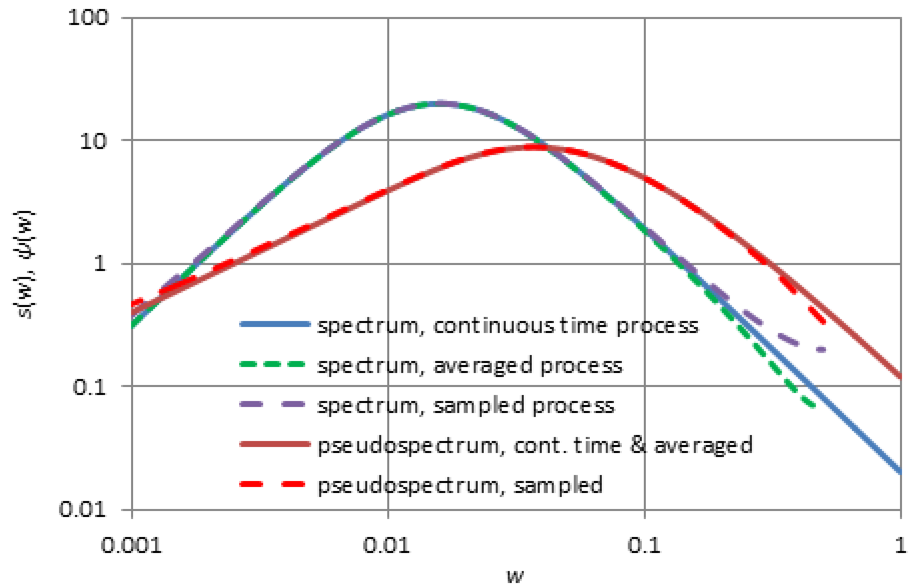


Example 4: A simple antipersistent process

Variance of instantaneous process	$\gamma = \gamma(0) = c(0) = \lambda$
Variance at scale Δ (Climacogram)	$\gamma(\Delta) = \frac{2\lambda}{\Delta/\alpha}\left(\frac{1-e^{-\Delta/\alpha}}{\Delta/\alpha} - e^{-\Delta/\alpha}\right)$
Autocovariance function for lag τ	$c(\tau) = \lambda(1-\tau/\alpha)\mathrm{e}^{-\tau/\alpha}$
Power spectrum for frequency w	$s(w) = 8\lambda\alpha\left(\frac{2\pi\alpha w}{1+(2\pi\alpha w)^2}\right)^2$
Asymptotic slopes	$\psi^\#(0) = 1;\ s^\#(0) = 2$ $\psi^\#(\infty) = s^\#(\infty) = -2$ $\gamma^\#(\infty) = 2\sigma^\#(\infty) = -2$ $\gamma^\#(0) = \sigma^\#(0) = 0$
Parameter values used	$\lambda = 1$, $\alpha = 10$; the spacing is $D = 1$.

- The condition making the process antipersistent is that $4\int_0^\infty c(\tau)\mathrm{d}\tau = \psi(0) = s(0) = 0$ (while of course $c(0) = \gamma(0) = \lambda > 0$).
- For $\tau > \alpha$, the autocovariance is consistently negative—but for small τ it is positive.
- Antipersistence is manifested in the positive slopes in power spectrum and pseudospectrum. Clearly, these slopes are positive only for low w.

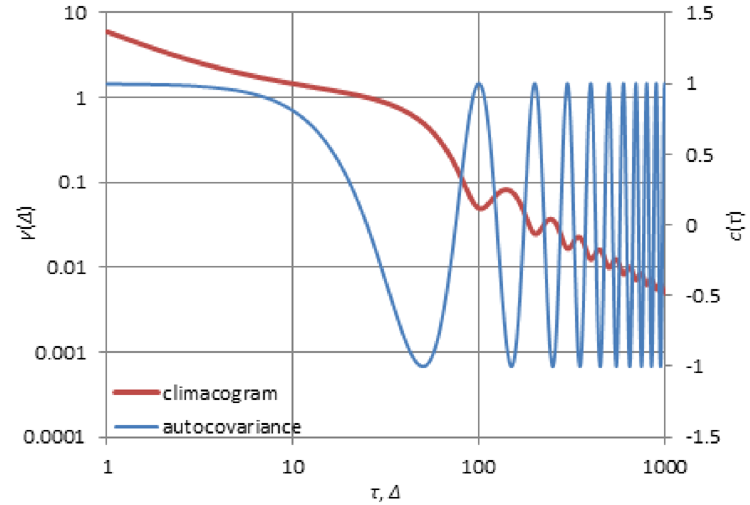


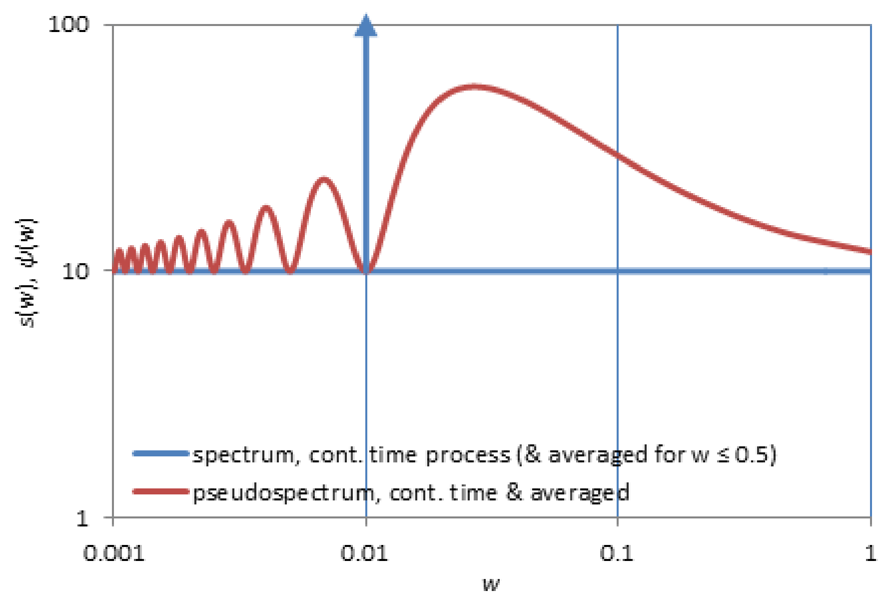


Example 5: A periodic process with white noise

Variance of instantaneous process	$\gamma = \gamma(0) = c(0) = \infty$
Variance at scale Δ (Climacogram)	$\gamma(\Delta) = \frac{\lambda_1}{\Delta/\alpha} + \lambda_2 \text{sinc}^2\left(\frac{\Delta}{\alpha}\right)$
Autocovariance function for lag τ	$c(\tau) = \lambda_1 \delta(\tau/\alpha) + \lambda_2 \cos(2\pi\, \tau/\alpha)$
Power spectrum for frequency w	$s(w) = 2\lambda_1\alpha + \lambda_2\alpha\,\delta(\alpha w - 1)$
Asymptotic slopes for $\lambda_1 > 0$ [and for $\lambda_1 = 0$; but not valid for $s^\#(\,)$]	$\psi^\#(0) = s^\#(0) = 0\ [+1]$ $\psi^\#(\infty) = s^\#(\infty) = 0\ [-3]$ $\gamma^\#(\infty) = 2\sigma^\#(\infty) = -1[-2]$ $\gamma^\#(0) = \sigma^\#(0) = -1\ [0]$
Parameter values used	$\lambda_1 = 0.05, \lambda_2 = 1, \alpha = 100$

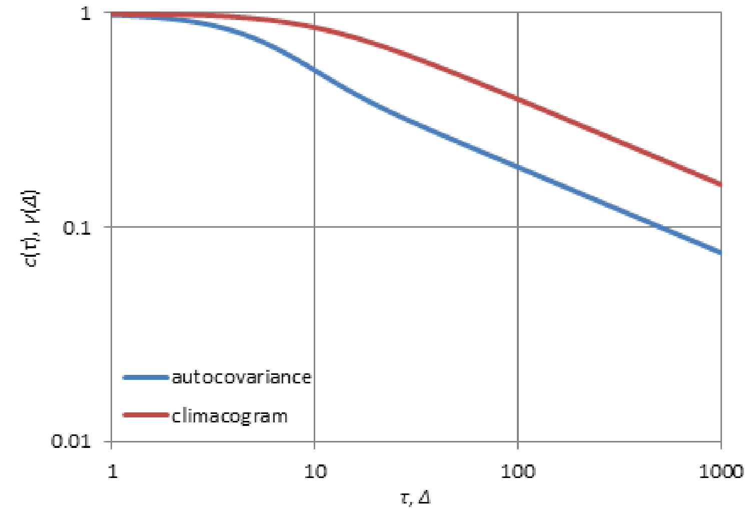
- $\delta(x)$ is the Dirac delta function while $\text{sinc}(x) \coloneqq \sin(\pi x)/\pi x$.
- Strictly speaking, the periodic component is a deterministic rather than a stochastic process. In this respect, the process should be better modelled as a cyclostationary one. However, the fact that the autocorrelation is a function of the lag τ only, allows the process to be treated as a typical stationary stochastic process.



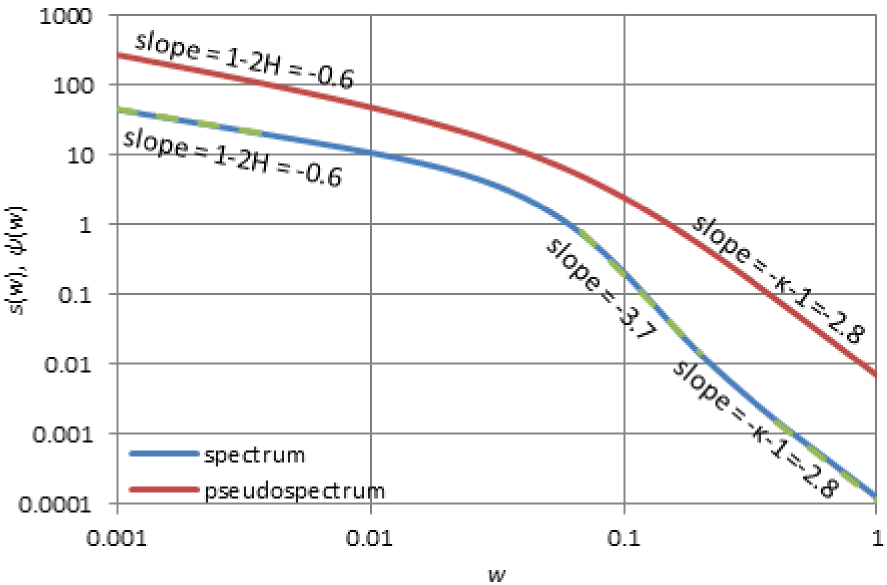


Example 6: A process with Cauchy-type climacogram

Variance of instantaneous process	$\gamma = \gamma(0) = c(0) = \lambda$
Variance at scale Δ (Climacogram)	$\gamma(\Delta) = \frac{\lambda}{(1+(\Delta/\alpha)^{\kappa})^{\frac{2-2H}{\kappa}}}$
Autocovariance function for lag τ	$c(\tau)$: expression too complex
Power spectrum for frequency w	$s(w)$: expression too complex
Asymptotic slopes	$\psi^{\#}(0) = s^{\#}(0) = 1 - 2H$ $\psi^{\#}(\infty) = s^{\#}(\infty) = -\kappa - 1$ $\gamma^{\#}(\infty) = 2\sigma^{\#}(\infty) = -2 + 2H$ $\gamma^{\#}(0) = \sigma^{\#}(0) = 0$
Parameter values used	$\lambda = 1$, $\alpha = 10$, $H = 0.8$, $\kappa = 1.8$



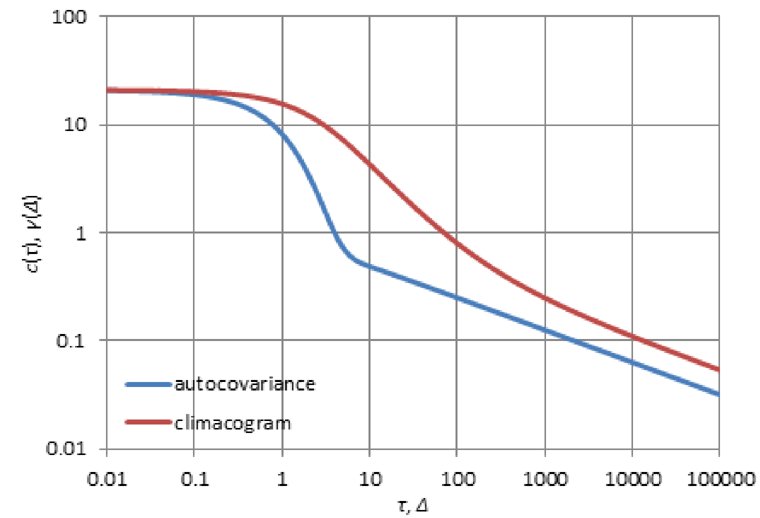
- The process was derived by modifying one proposed by Gneiting and Schlather (2004).
- The important feature of this process is that it allows control of both asymptotic slopes.
- The asymptotic slopes of the pseudospectrum are identical with those of the spectrum.
- An intermediate steep slope that appears in the power spectrum is artificial and does not indicate a scaling behaviour.

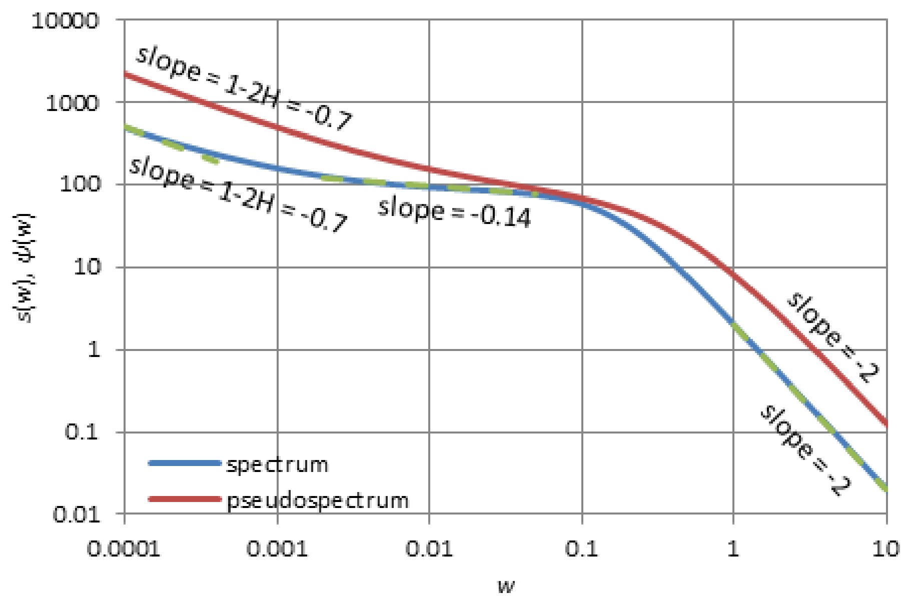


Example 7: A composite long-range and short-range dependence

Variance of instantaneous process	$\gamma = \gamma(0) = c(0) = \lambda_1 + \lambda_2$
Variance at scale Δ (Climacogram)	$\gamma(\Delta)$: expression too complex (sum from examples 1 and 2)
Autocovariance function for lag τ	$c(\tau) = \frac{\lambda_1}{(1+\tau/\alpha)^{2-2H}} + \lambda_2 \mathrm{e}^{-\tau/\alpha}$
Power spectrum for frequency w	$s(w)$: expression too complex
Asymptotic slopes	$\psi^{\#}(0) = s^{\#}(0) = 1 - 2H$ $\psi^{\#}(\infty) = s^{\#}(\infty) = -2$ $\gamma^{\#}(\infty) = 2\sigma^{\#}(\infty) = -2 + 2H$ $\gamma^{\#}(0) = \sigma^{\#}(0) = 0$
Parameter values used	$\lambda_1 = 1, \lambda_2 = 20, \alpha = 10, H = 0.85$

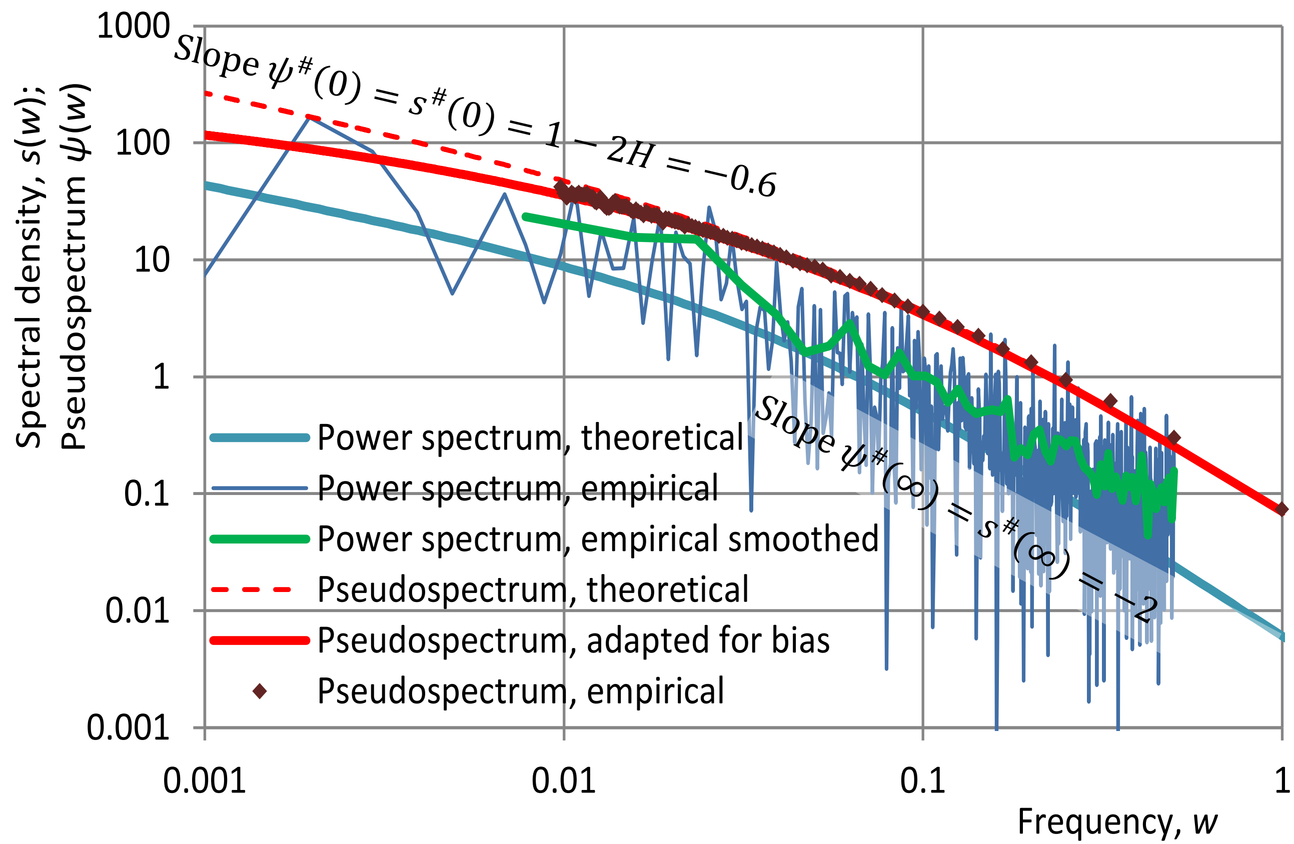
- Again the asymptotic slopes of the pseudospectrum are identical with those of the spectrum.
- As in the previous example, an intermediate slope appears in the power spectrum (in this case a mild one). Again this is artificial, here imposed by the Markov process, and does not indicate a scaling behaviour.





Example 8: The initial example: process defined by (7)

- Examples 1-7 have been focused on the theoretical power spectrum and pseudospectrum and have shown that:
  - (a) the asymptotic behaviours of the two are similar;
  - (b) the pseudospectrum is less affected by discretization.
- Example 8 adds information from data.



- It shows that when the power spectrum and pseudospectrum are estimated from data, the latter is much smoother and its bias is a priori known, thus enabling a more direct and accurate estimation of slopes and fitting on a model.

Conclusions

- The power spectrum is very powerful in identifying strong periodicities in time series. However, it has some problems in identifying scaling laws and weak periodicities, as:
  - Discretization and finite length of data alter asymptotic slopes;
  - The rough shape of the periodogram may result in:
    - misleading, inaccurate or even incorrect slopes (e.g. slope > –1 for frequency → 0, which is infeasible);
    - false periodicities;
  - Biases and uncertainties are uncontrollable, particularly when the periodogram is smoothed;
  - False detection of artificially induced scaling areas is likely.
- The climacogram-based pseudospectrum has an asymptotic behaviour similar to that of the power spectrum and offers some advantages such as:
  - Its calculation is very easy: it only uses the concept of variance and does not involve integral transformations (like the Fourier transform);
  - It is smooth;
  - Its biases and uncertainties are smaller and easy to determine;
  - Its asymptotic slopes are determined more accurately from data.

References

- Gneiting, T., and M. Schlather, Stochastic models that separate fractal dimension and the Hurst effect, *Society for Industrial and Applied Mathematics Review*, 46 (2), 269-282, 2004.
- Koutsoyiannis, D., Climate change, the Hurst phenomenon, and hydrological statistics, *Hydrological Sciences Journal*, 48 (1), 3–24, 2003.
- Koutsoyiannis, D., A random walk on water, *Hydrology and Earth System Sciences*, 14, 585–601, 2010.
- Koutsoyiannis, D., Hurst-Kolmogorov dynamics as a result of extremal entropy production, *Physica A: Statistical Mechanics and its Applications*, 390 (8), 1424–1432, 2011.
- Koutsoyiannis, D., *Encolpion of stochastics: Fundamentals of stochastic processes*, 30 pages, Department of Water Resources and Environmental Engineering – National Technical University of Athens, Athens, 2013 (itia.ntua.gr/1317/).