

American Geophysical Union Fall Meeting 2014

San Francisco, USA, 15-19 December 2014

Session (NG41A-3737): Nonlinear Geophysics; Scale and Scaling in the Atmosphere, Ocean, Hydrosphere, Natural Hazards, and Climate

Presenting date: Thursday, 18/12/2014, 08:00 AM - 12:20 PM , Moscone South-Poster Hall

Using multiple stochastic tools in identification of scaling in hydrometeorology

Dimitriadis P.* and D. Koutsoyiannis**

Department of Water Resources and Environmental Engineering National Technical University of Athens

*pandim@itia.ntua.gr

**presenting author

1. Introduction - Abstract

The identification and quantification of stochastic scaling laws has been an important task in modelling of hydrometeorological processes. Stochastic tools such as the power spectrum, autocovariance function, structure and climacogram have been among the most powerful. However, the common practice of using solely one of them may lead to process misinterpretation. We introduce a methodology that compares these stochastic tools and seeks the optimal one for different scales in terms of minimizing fitting errors. For validation and illustration purposes, we apply this methodology to various fundamental stochastic processes, such as Markovian, Hurst-Kolmogorov (HK) and Cauchy type ones. For each one, we produce Gaussian synthetic time series, we estimate the uncertainty of their expected values and finally, we conclude upon the ones with the smallest uncertainty. Furthermore, we apply this method to a real case time-series of high resolution turbulent flow velocities.

2. Definitions and notations

Here, we adopt the Dutch convention for the notation and a climacogram-based stochastic framework described in Koutsoyiannis (2013) and Dimitriadis et al. (under review) along with the assumption $D = \Delta > 0$, where D is the time interval between two observations of the continuous time process $\underline{x}(t)$ and Δ is the time window that corresponds to each observation. The discrete time process $\underline{x}_i^{(\Delta)}$, can then be calculated from $\underline{x}(t)$ as:

$$\underline{x}_i^{(\Delta)} = \frac{\int_{(i-1)\Delta}^{i\Delta} \underline{x}(\xi) d\xi}{\Delta} \quad (1)$$

where $i \in [1, n]$ is an index representing discrete time, n is the total number of observations and $T \in [0, \infty)$ is the time length of observations.

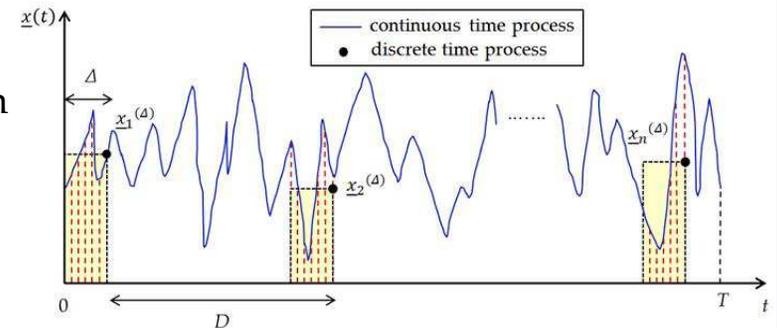


Figure 1: An example of a continuous time process sampled at time intervals D for a total period T and with instrument response time Δ (Source: Dimitriadis et al., under review).

3. Stochastic tools (autocovariance-based)

Table 1: Autocovariance definition and expressions for a process in continuous and discrete time, along with the properties of its estimator and the variogram and power-spectrum for the continuous time only (the rest can be easily derived from eq. 3-5).

Type	Autocovariance	
continuous	$c(\tau) := \text{Cov}[\underline{x}(t), \underline{x}(t + \tau)]$	(2)
	where $\tau \in \mathbb{R}$ is the lag for a continuous time process (in time units)	
discrete	$c_d^{(\Delta)}(j) := \text{Cov}[\underline{x}_i^{(\Delta)}, \underline{x}_{i+j}^{(\Delta)}] = \frac{\Delta^2 [j^2 \gamma(j\Delta)]}{2\Delta [j^2]}$	(3)
	$= \frac{1}{2} \left((j+1)^2 \gamma((j+1)\Delta) + (j-1)^2 \gamma((j-1)\Delta) - 2j^2 \gamma(j\Delta) \right)$	
	where $j \in \mathbb{Z}$ is the lag for the process at discrete time (dimensionless)	
classical estimator	$\hat{c}_d^{(\Delta)}(j) = \frac{1}{\zeta(j)} \sum_{i=1}^{n-j} \left(\underline{x}_i^{(\Delta,D)} - \frac{1}{n} \left(\sum_{l=1}^n \underline{x}_l^{(\Delta)} \right) \right) \left(\underline{x}_{i+j}^{(\Delta,D)} - \frac{1}{n} \left(\sum_{l=1}^n \underline{x}_l^{(\Delta)} \right) \right)$	(4)
	where $\zeta(j)$ is usually taken as: n or $n-1$ or $n-j$	
expectation of classical estimator	$E[\hat{c}_d^{(\Delta)}(j)] = \frac{1}{\zeta(j)} \left((n-j)c_d^{(\Delta)}(j) + \frac{j^2}{n} \gamma(j\Delta) - j\gamma(n\Delta) - \frac{(n-j)^2}{n} \gamma((n-j)\Delta) \right)$	(5)
continuous Variogram	$v(\tau) := c(0) - c(\tau)$	(6)
continuous Power-spectrum	$s(w) := 4 \int_0^\infty c(\tau) \cos(2\pi w \tau) d\tau$	(7)
	where $w \in \mathbb{R}$ is the frequency for continuous time (in inverse time units).	

In this and the following section, we present the most common and powerful stochastic tools following Dimitriadis et al. (under review) categorization into autocovariance-based ones: autocovariance, 2nd structural function (else known as variogram) and power-spectrum; and climacogram-based ones: climacogram (i.e., variance of the averaged process versus averaging time scale, introduced by Koutsoyiannis, 2013), climacogram-based 2nd structural function (CBSF) and climacogram-based spectrum (CBS; introduced by Koutsoyiannis, 2013). In Tables 1 and 2, we present the definition for a stochastic stationary process in continuous and discrete time, a classical estimator and its corresponding estimation for the autocovariance and climacogram. For the rest of the examined stochastic tools, we present only the expressions in continuous time.

4. Stochastic tools (climacogram-based)

Table 2: Climacogram definition and expressions for a process in continuous and discrete time, along with the properties of its estimator and the CBSF and CBS for the continuous time only (the rest can be easily derived from eq. 9-11).

Type	Climacogram	
continuous	$\gamma(m) := \frac{\text{Var}\left[\int_t^{t+m} \underline{x}(\xi) d\xi\right]}{m^2} = \text{Var}\left[\int_0^m \underline{x}(\xi) d\xi\right] / m^2$	(8)
	where $m \in \mathbb{R}^+$ is the scale for a continuous time process (in time units)	
discrete	$\gamma_d^{(\Delta)}(k) := \frac{\text{Var}\left[\sum_{l=1}^k \underline{x}_l^{(\Delta,D)}\right]}{k^2} = \gamma(k\Delta)$	(9)
	where $k \in \mathbb{N}$ is the dimensionless scale for a discrete time process	
classical estimator	$\hat{\gamma}_d^{(\Delta)}(k) = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{1}{k} \left(\sum_{l=k(i-1)+1}^{ki} \underline{x}_l^{(\Delta)} \right) - \frac{\sum_{l=1}^n \underline{x}_l^{(\Delta)}}{n} \right)^2$	(10)
expectation of classical estimator	$\mathbb{E}\left[\hat{\gamma}_d^{(\Delta)}(k)\right] = \frac{1 - \gamma_d^{(\Delta)}(n)/\gamma_d^{(\Delta)}(k)}{1 - k/n} \gamma_d^{(\Delta)}(k)$	(11)
continuous CBSF	$\xi(m) := \gamma(0) - \gamma(m)$	(12)
continuous CBS	$\psi(w) := \frac{2\gamma(1/w)}{w} \left(1 - \frac{\gamma(1/w)}{\gamma(0)} \right)$	(13)

It is important to note that in most cases (e.g., Dimitriadis and Koutsoyiannis, under review and Dimitriadis et al., under review), the climacogram and autocovariance are useful for investigating the large-scale behaviour of the process (e.g., determine if there is an exponential or a power-law decay in large scales), the variogram and CBSF for the small-scale behaviour (e.g., to estimate the fractal dimension of the process) and the power-spectrum and CBS for the intermediate-scale behaviour (e.g., to test the validity of K41 theory in case of a turbulent isotropic process). For an illustrative example of such analysis see the application in sections 11-13 (which is based on the work of Dimitriadis et al., under review).

5. Stochastic processes and important scaling parameters

Here, we present the powered exponential process (abbreviated as PEX), a generalized HK process (gHK) and the simple processes of Markov and HK, respectively (all expressed via autocovariance):

$$c(\tau) = e^{-(|\tau|/q)^a} \quad (14)$$

$$c(\tau) = (1 + |\tau|/q)^{2H-2} \quad (15)$$

$$c(\tau) = e^{-|\tau|/q} \quad (16)$$

$$c(\tau) = |\tau|^{2H-2} \quad (17)$$

Moreover, a useful mathematical tool for investigating scaling laws is the negative logarithmic derivative (abbreviated as NLD). For any function $f(x)$ its NLD is denoted as $f^\#(x)$ and defined as:

$$f^\#(x) := -\frac{d \ln(f(x))}{d \ln x} = -\frac{x}{f(x)} \frac{df(x)}{dx} \quad (18)$$

Also, we introduce two of the most important scaling parameters in hydrometeorology, the Hurst coefficient and fractal dimension (Dimitriadis et al., under review):

$$H := 1 + \frac{1}{2} \lim_{\tau \rightarrow \infty} c^\#(\tau) \quad (19)$$

$$F := 2 - \frac{1}{2} \lim_{\tau \rightarrow 0} v^\#(\tau) \quad (20)$$

6. Process identification issues:

Markov vs PEX in small scales (fractal dimension?)

An important characteristic of a process is the fractal parameter a (in case we choose a PEX process for the small scales). However, it is not always easy to estimate its value, since a Markov process ($\alpha = 1$) can exhibit similar NLD at the small scales with a PEX process ($\alpha \neq 1$) for a value of the q parameter. Here, we estimate the α parameter of the PEX process by equalizing its true NLD value (between scales 2 and 3) with that of a Markov process (Fig. 2) for a known q , based on climacogram's classical estimator. Close results were estimated for the NLD between scales 1 and 2 as well as 3 and 4. As we can see from Fig. 2, the bias related to these processes is negligible for the first scales and thus, the results are expected to be independent of n .

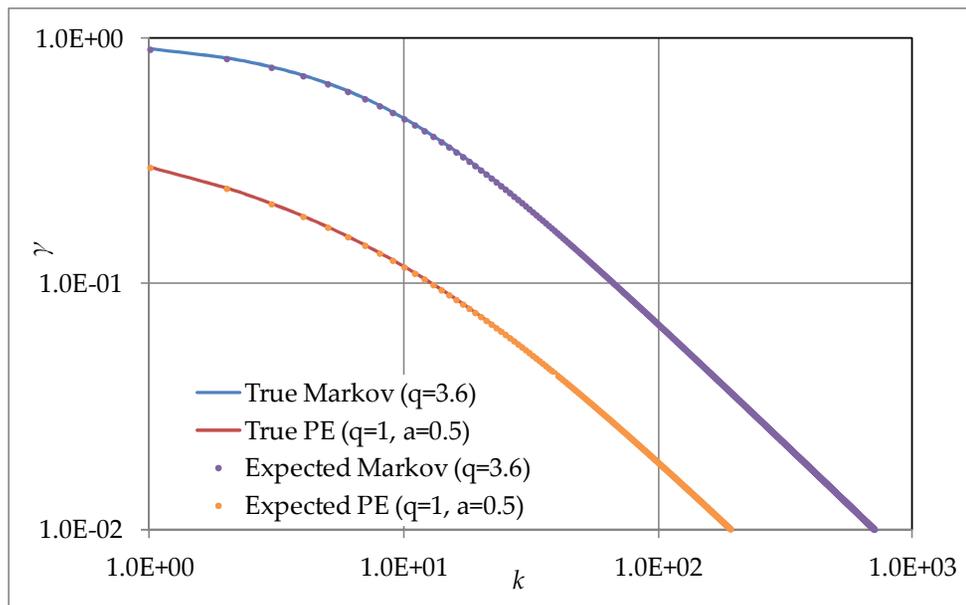


Figure 2: Comparison between a PEX process with $\alpha = 0.5$ ($F = 1.75$) and $q = 1.0$ and a Markov process with $q = 3.6$.

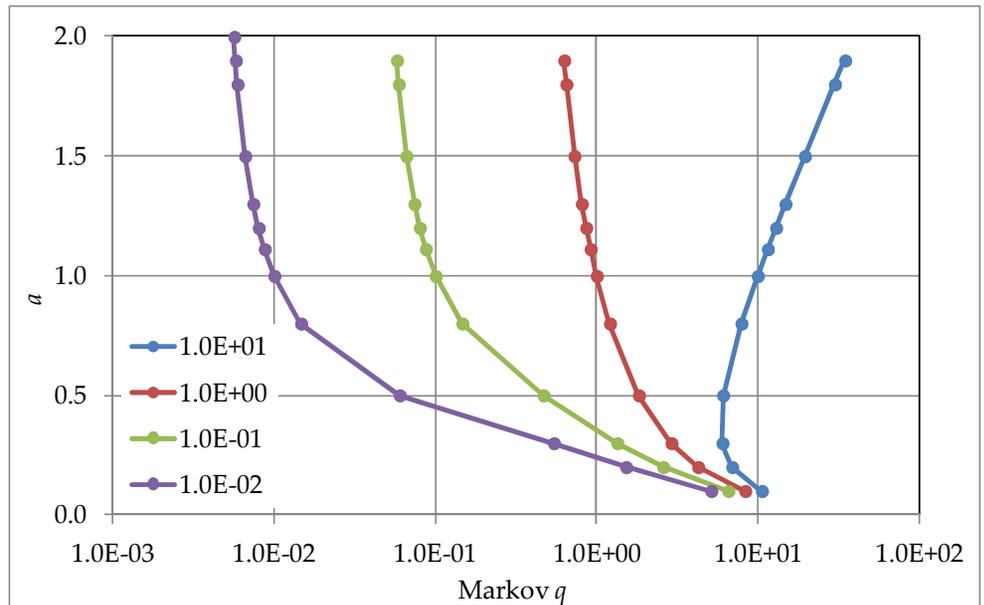


Figure 3: Estimated power parameter of a PEX process for various values of the q parameter of a PEX and a Markov process.

7. Process identification issues (cont.):

Markov vs HK (with $H > 0.5$) in large scales (HK behaviour?)

Another significant decision concerning the type of selected process is whether exhibits an HK behaviour. Again, there can be more than one processes that could exhibit an HK behaviour at a range of scales. To illustrate this, we estimate the q parameter of a Markov process by equalizing the NLD of its expected value (at scale equal to 90% of n) with that of an HK process for a known H (Fig. 5), based on climacogram's classical estimator. Similar results were derived at scales equal to 80% and $\sim 100\%$ of n . The curves in Fig. 5 can be well approximated from:

$$H \approx 0.5 e^{2.25 q/n} \leftrightarrow q \approx \frac{n}{2.25} \ln(2H) \quad (21)$$

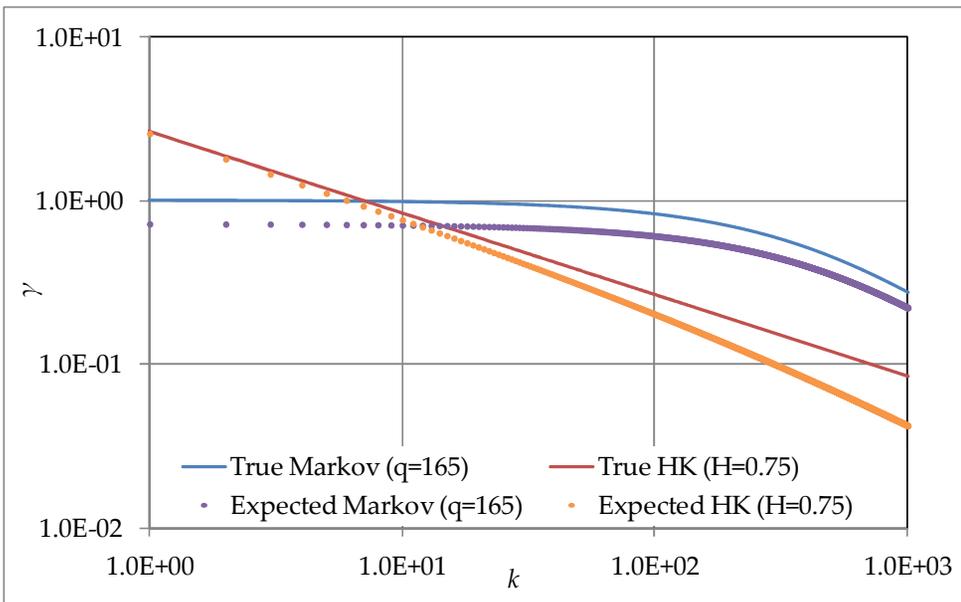


Figure 4: Comparison between a Markov process with $q = 165$ and an HK process with $H = 0.75$.

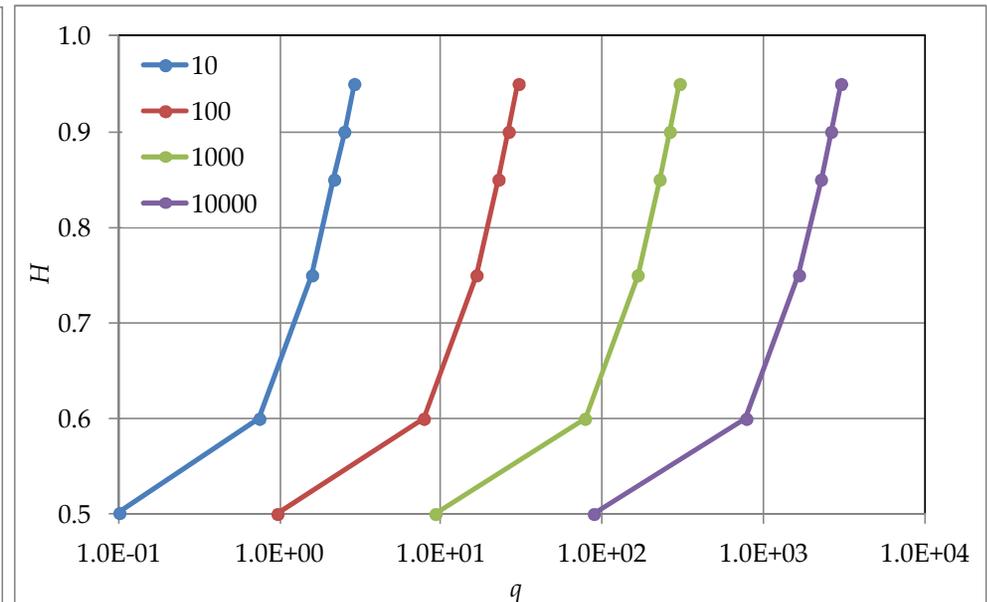


Figure 5: Estimated Hurst coefficients of an HK process for various values of the q parameter of a Markov process and for various n .

8. Process parameters quantification

Here, we will describe a methodology for the parameter estimation of a process. Despite the fact that all aforementioned stochastic tools are linked to each other through the equations in Tables 1 and 2, the statistical uncertainty they produce is not the same. In that sense, for each process we should often have to choose a different tool to calculate its parameters. Below, we investigate the performance of the estimators of climacogram, autocovariance and power spectrum for a Markov process (with $q = 10$ and $n = 10^3$). For their evaluation we use mean square error (MSE) expressions as shown in the equations below (Dimitriadis and Koutsoyiannis, under review):

$$\varepsilon = \frac{E[(\hat{\theta} - \theta)^2]}{\theta^2} = \varepsilon_v + \varepsilon_b \quad (22)$$

$$\varepsilon_v = \text{Var}[\hat{\theta}]/\theta^2 \quad (23)$$

$$\varepsilon_b = (\theta - E[\hat{\theta}])^2/\theta^2 \quad (24)$$

where θ is the true value of a statistical characteristic (i.e. climacogram, autocovariance, power spectral density and NLDs thereof) of the process.

Although we have the expressions for the expected values of the examined stochastic tools (Tables 1 and 2), we lack of analytical expressions for their variance, and thus, we adopt a Monte Carlo method by producing 10^4 synthetic timeseries of a Markov process with $q = 10$ (following the algorithms and schemes in Dimitriadis and Koutsoyiannis, under review). In Fig. 6a, we show the true, discrete and expected values for the examined process for the three examined tools. We observe that a Markov process can be easily misinterpreted with a random process if it is not analyzed using the autocovariance (in large scales or small frequencies, since both climacogram and power spectrum give the same NLDs as in the case of a random process). This is also an example that highlights the need to use multiple tools when analyzing a timeseries.

9. Process parameters quantification (cont.)

In Fig. 6b, we present the resulted MSE following the previous analysis. As we can see, for the specific process, the climacogram attributes the smaller MSE (among the examined tools) and thus, in case we choose to apply this process, we should estimate its parameters based on the climacogram classical estimator.

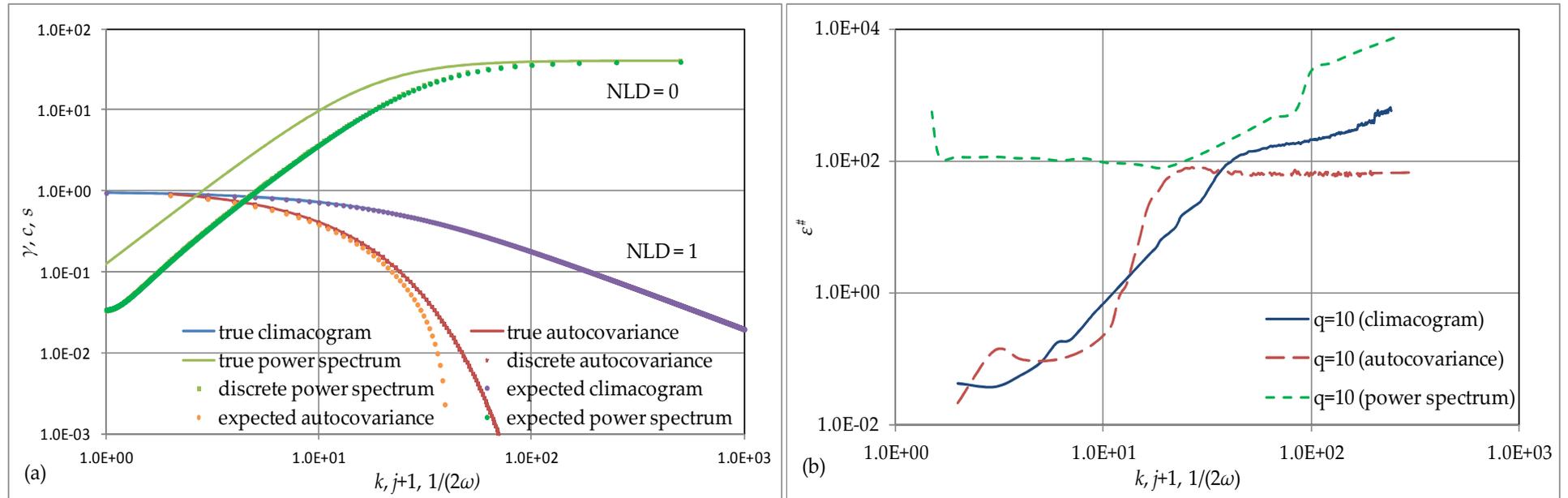


Figure 6: (a) True, discrete and expected values for the climacogram, autocovariance and power spectrum for a Markov process with $q = 10$ ($n = 10^3$); (b) the overall MSE for each stochastic tool for the examined Markov process.

It is interesting to note that, based on the analysis of Dimitriadis and Koutsoyiannis (under review) where various cases of HK, gHK and Markov process were examined, the overall MSE between those three stochastic tools followed approximately the inequality below in all cases:

$$E \left[\left(\hat{\gamma} - \gamma \right)^2 \right] / \gamma^2 \approx E \left[\left(\hat{c}_d^{(\Delta)} - c_d^{(\Delta)} \right)^2 \right] / c_d^{(\Delta)2} \approx E \left[\left(\hat{s}_d^{(\Delta)} - s_d^{(\Delta)} \right)^2 \right] / s_d^{(\Delta)2} \quad (25)$$

10. Process parameters quantification (cont.)

In Fig. 7, we present the MSE of the climacogram, autocovariance and power spectrum for various cases of Markov, HK and gHK processes (taken from the analysis of Dimitriadis and Koutsoyiannis, under review). As we can see, the climacogram attributes the smaller MSE and thus, in case we choose to apply this process, we should estimate its parameters based on the climacogram's classical estimator.

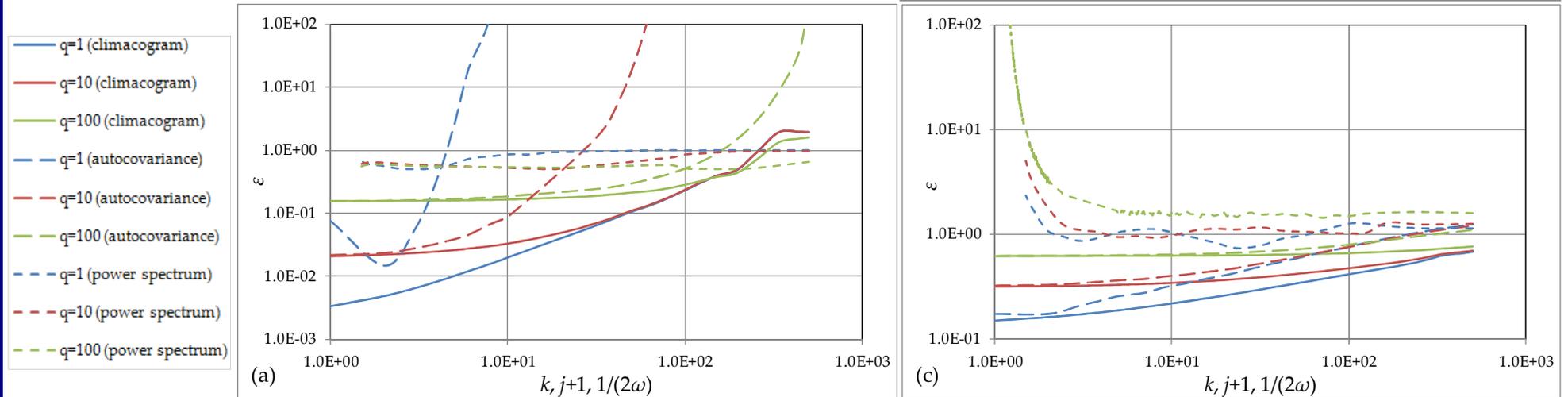


Figure 7: (a) True, discrete and expected values for the climacogram, autocovariance and power spectrum for a Markov process with $q = 10$ ($n = 10^3$); (b) the overall MSE for each stochastic tool for the examined Markov process.

11. Application

In conclusion, we first need to decide on the process we are going to use. As we showed in the previous sections, there could be more than one processes with similar values for certain ranges of scale. However, we can narrow our choices by adopting the principle of parsimony. After we built our model, we decide on the stochastic tools we want to use and we follow the proposed methodology in the previous sections to incorporate the statistical uncertainty produced by each tool. Finally, we can verify our null hypothesis on the chosen process by checking that its parameters were indeed calculated using the tool with the smallest MSE. For more information on the above methodology, see in Dimitriadis and Koutsoyiannis (under review).

The above rules of thumb were applied in a turbulent wind velocity timeseries measured by X-wire probes downstream of an active grid at the direction of the flow (Kang et al., 2003). The selected process is the following (Fig. 8):

$$c(\tau) = \lambda_1 e^{-(|\tau|/q_1)^a} + \lambda_2 (1 + |\tau|/q_2)^{-b} \quad (26)$$

We used all the examined stochastic tools but we focused in the CBS for the small and large scales related parameters (i.e., the fractal parameter α and Hurst coefficient H) and the variogram for the rest (since these tools have the smallest MSE in Fig. 10b). The fitted parameters (Dimitriadis et al., under review) are: $\lambda_1 = 0.422 \text{ m}^2/\text{s}^2$ and $\lambda_2 = 0.592 \text{ m}^2/\text{s}^2$, $q_1 = 19.6 \text{ ms}$ and $q_2 = 1.45 \text{ ms}$, $a = 1.4$ ($F = 1.3$) and $H = 0.84$.

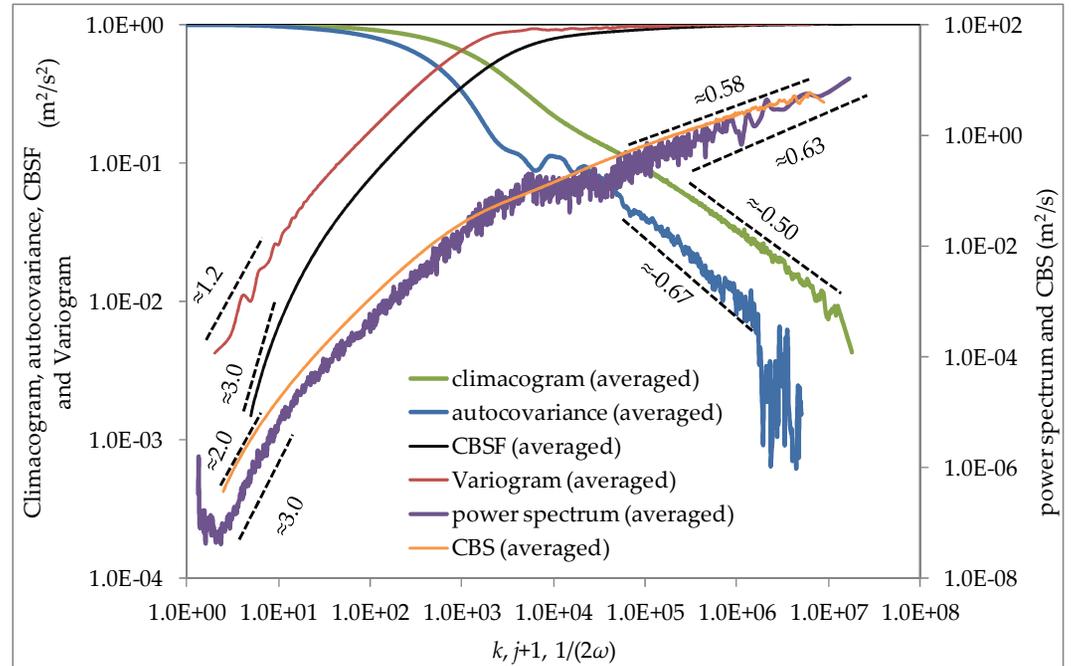


Figure 8: Empirical averaged values for all the examined stochastic tools along with their log-log derivatives (Source: Dimitriadis et al., under review).

12. Application (cont.)

Here, we show the results from the model (Fig. 9) as well as the total MSE for each examined stochastic tool (Fig. 10) which verifies the use of CBS for the small and large scales and the variogram for the intermediate ones. Note that the variogram and CBSF are not appropriate for the large scales and similarly, the autocovariance and climacogram for the small scales, since their log-log derivatives tend to 0 at these ranges, independently of the process (fixed boundary).

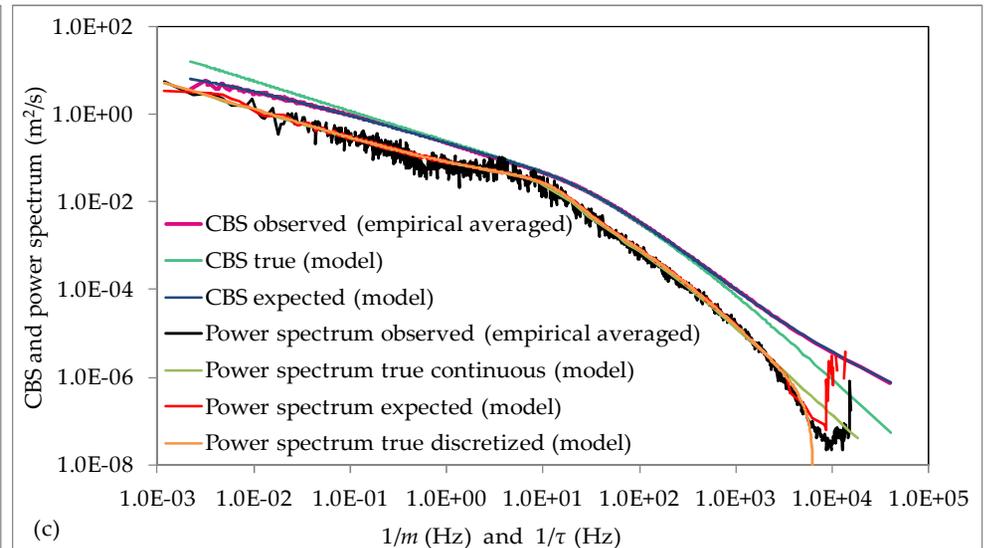
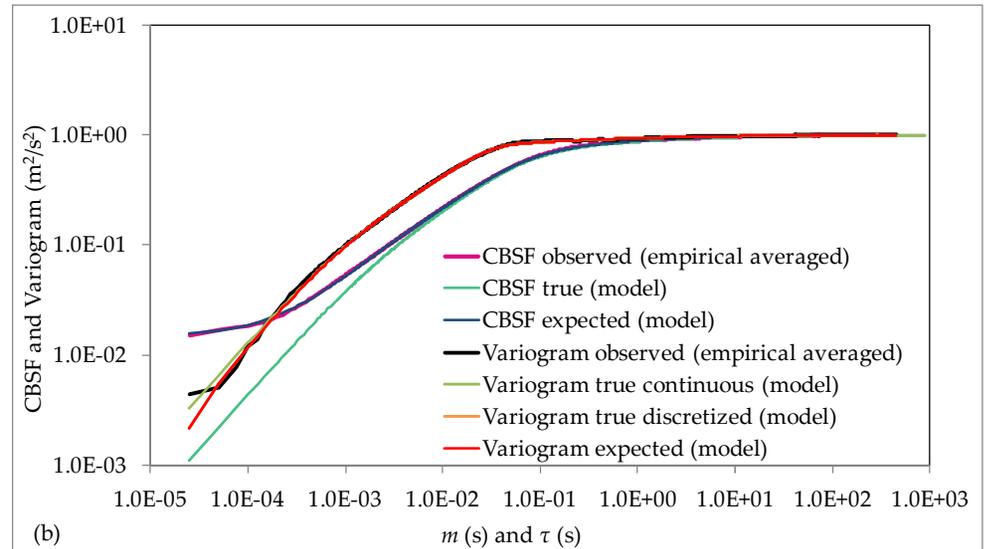
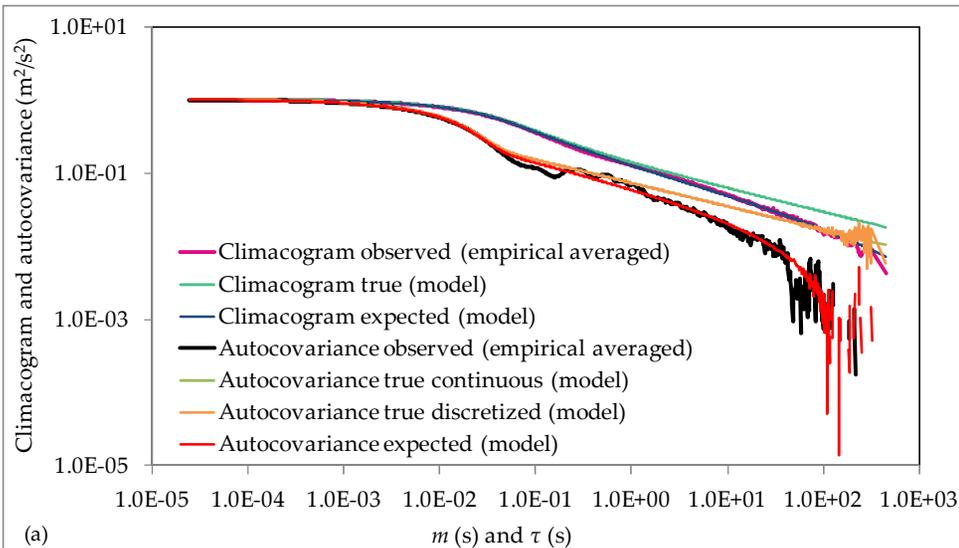


Figure 9: Empirical averaged values from data, true continuous and discretized as well as expected values from the model for (a) the climacogram and autocovariance, (b) the variogram and CBSF and (c) the power spectrum and CBS (Source: Dimitriadis et al., under review).

13. Application (cont.)

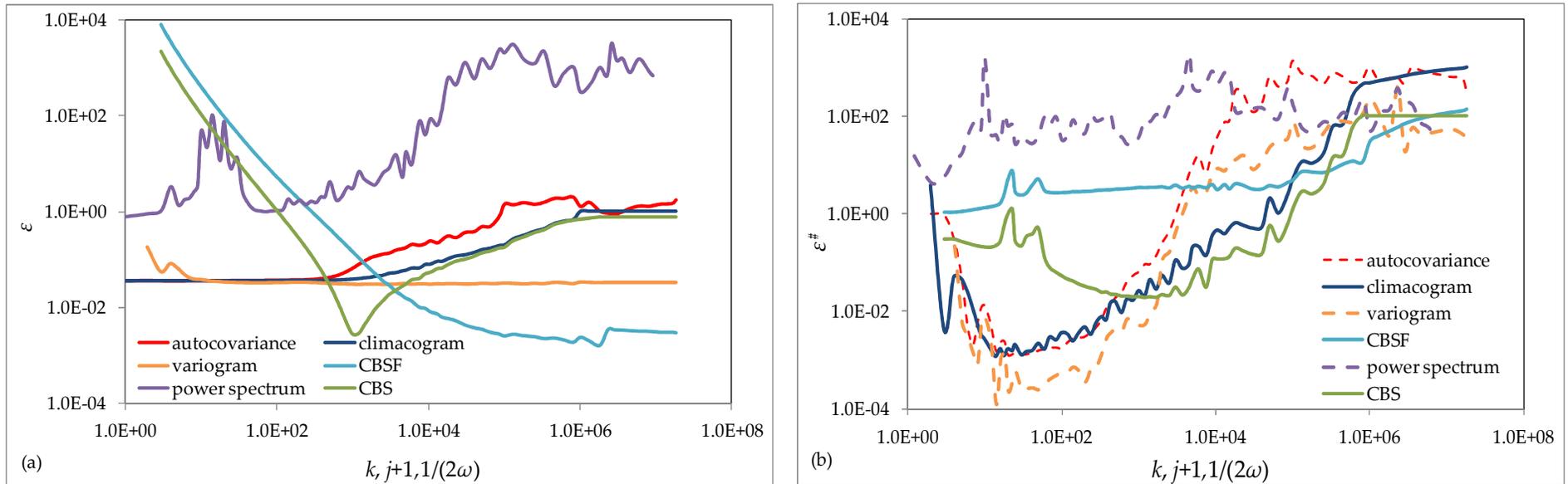


Figure 10: MSE of (a) regular values (i.e., ϵ) and (b) NLDs (i.e., $\epsilon^\#$) for the climacogram, autocovariance, variogram, CBSF, power spectrum and CBS.

One of the most important features of the above analyses was to highlight the advantages of investigating a process from the perspective of multiple stochastic tools, incorporating in this way the varying statistical uncertainty produced in different scales.

References

- Dimitriadis P. and D. Koutsoyiannis, under review. Climacogram vs. autocovariance and power spectrum in stochastic modelling for Markovian and Hurst-Kolmogorov processes.
- Dimitriadis P., D. Koutsoyiannis and P. Papanicolaou, under review. Climacogram-based stochastic modelling of isotropic turbulence.
- Koutsoyiannis, D., 2013. Encolpion of stochastics: Fundamentals of stochastic processes, 12 pages, *Department of Water Resources and Environmental Engineering, National Technical University of Athens, Athens*.
- Kang H.S., S. Chester and C. Meneveau, 2003. Decaying turbulence in an active-grid-generated flow and comparisons with large-eddy simulation, *J. Fluid Mech.* 480, p. 129-160.