ROMA TRE
UNIVERSITÀ DEGLI STUDI

SCIENZE DELL'INGEGNERIA CIVILE

SCUOLA DOTTORALE

XXV

CICLO DEL CORSO DI DOTTORATO

How to catch scaling in geophysics: some problems in stochastic modelling and inference from time series

Titolo della tesi

Federico Lombardo _____
Nome e Cognome del dottorando firma

Elena Volpi _____
Docente Guida/Tutor: Prof. firma

Aldo Fiori _____
Coordinatore: Prof. firma

Collana delle tesi di Dottorato di Ricerca
In Scienze dell'Ingegneria Civile
Università degli Studi Roma Tre
Tesi n° 46

L'arte e la scienza sono libere e libero ne è l'insegnamento.
(Costituzione della Repubblica Italiana, art. 33)

Sommario

Nel corso degli ultimi decenni lo studio dei cambiamenti nei processi geofisici ha riscosso un crescente interesse soprattutto in riferimento alle potenziali ripercussioni sulla nostra società. Lo scopo primario è di migliorare il processo di previsione di tali cambiamenti per consentire uno sviluppo sostenibile delle attività umane in un ambiente mutevole. Le variazioni dei processi geofisici sono presenti a tutte le scale temporali e sono irregolari a tal punto che la loro descrizione può certamente essere migliore in termini stocastici (casuali) che deterministici. Nella statistica classica, la casualità viene solitamente rappresentata da processi stocastici le cui variabili casuali sono indipendenti ed identicamente distribuite. Tuttavia, esiste un'ampia evidenza empirica che spesso confuta tale assunzione. Infatti, è stato osservato in più occasioni che la correlazione tra campioni sempre più distanti tra loro nel tempo decresce più lentamente non solo di quanto ovviamente ci si aspetta per campioni indipendenti ma anche rispetto al caso di dipendenza markoviana o dei modelli di tipo ARMA. Tutto ciò è coerente con il fenomeno di Hurst, che è stato infatti osservato in molte lunghe serie temporali idroclimatiche. Esso è stocasticamente equivalente ad un comportamento auto-simile della variabilità del processo alle differenti scale temporali. Di conseguenza, i cambiamenti persistenti a lungo termine sono molto più frequenti ed intensi nei processi geofisici di quanto comunemente percepito e, inoltre, gli stati futuri sono molto più incerti ed imprevedibili su lunghi orizzonti temporali rispetto alle previsioni ottenute mediante i modelli tipicamente utilizzati nella pratica. L'obiettivo della presente tesi è la descrizione dell'inferenza e della modellazione delle proprietà statistiche relative ai processi naturali che presentano un comportamento del tipo scala invariante. Dapprima si indagano le ripercussioni che tale comportamento implica in riferimento all'ingente incremento di incertezza di stima dei parametri di interesse dalle serie temporali di dati. In seguito viene proposto un modello stazionario di disaggregazione temporale della precipitazione che rispetta il fenomeno di Hurst. Tale modello è caratterizzato da una semplice struttura a cascata simile a quella dei più famosi modelli a cascata moltiplicativa di tipo discreto. Inoltre mostriamo il grande limite di questi ultimi modelli che simulano un processo intrinsecamente non stazionario a causa della loro struttura.

Abstract

During recent decades, there has been a growing interest in research activities on change in geophysics and its interaction with human society. The practical aim is to improve our capability to make predictions of geophysical processes to support sustainable societal development in a changing environment. Geophysical processes change irregularly on all time scales, and then this change is hardly predictable in deterministic terms and demands stochastic descriptions, or random. The term randomness is usually associated to stochastic processes whose samples are regarded as a sequence of independent and identically distributed random variables. This is a basic assumption of classical statistics, but there is ample practical evidence that this wish does not always become a reality. It has been observed empirically that correlations between distant samples decay to zero at a slower rate than one would expect from not only independent data but also data following classical ARMA- or Markov-type models. Indeed, many geophysical changes are closely related to the Hurst phenomenon, which has been detected in many long hydroclimatic time series and is stochastically equivalent to a simple scaling behaviour of process variability over time scale. As a result, long-term changes are much more frequent and intense than commonly perceived and, simultaneously, the future states are much more uncertain and unpredictable on long time horizons than implied by typical modelling practices. The purpose of this thesis is to describe how to infer and model statistical properties of natural processes exhibiting scaling behaviours. We explore their statistical consequences with respect to the implied dramatic increase of uncertainty, and propose a simple and parsimonious model that respects the Hurst phenomenon. In particular, we first we highlight the problems in inference from time series of geophysical processes, where scaling behaviours in state (sub-exponential distribution tails) and in time (strong time dependence) are involved. Then, we focus on rainfall downscaling in time, and propose a stationary model that respects the Hurst phenomenon. It is characterized by a simple cascade structure similar to that of the most popular multiplicative random cascade models, but we show that the latter simulate an unrealistic non-stationary process simply inherent to the model structure.

Table of contents

List of figures

List of symbols

Δ	*Time scale*
j	*Discrete time step*
$\underline{R}_\Delta$	*Adjusted range*
H	*Hurst coefficient*
t	*Continuous time*
$\underline{B}_H$	*Fractional Brownian motion*
$\underline{x}(t)$	*Instantaneous process*
$\underline{x}_j^{(\Delta)}$	*Local average process over the time window Δ at discrete time steps j*
$\underline{X}(t)$	*Cumulative process in continuous time*
$\underline{X}_j^{(\Delta)}$	*Aggregated process on a time scale Δ*
n	*Sample size*
T_o	*Observation period (in continuous time)*
T	*Observation period (in discrete time)*
μ	*Mean of a stochastic process*
$\underline{m}_q^{(\Delta)}$	*Estimator of the q-th raw moment of the local average process $\underline{x}_j^{(\Delta)}$*
$\mu_q^{(\Delta)}$	*Theoretical raw moment of order q of the local average process $\underline{x}_j^{(\Delta)}$*
$\gamma(\Delta)$	*Variance of the time-averaged process $\underline{x}_j^{(\Delta)}$ as a function of the time scale of averaging Δ*
$\sigma(\Delta)$	*Standard deviation of the time-averaged process $\underline{x}_j^{(\Delta)}$ as a function of the time scale of averaging Δ*

$\underline{g}(\Delta)$	*Sample variance*
λ	*Scale ratio λ so that $\lambda=1$ for the largest scale of interest Δ_{max}, i.e. $\lambda=\Delta_{max}/\Delta$.*
$\underline{\varepsilon}(\lambda)$	*Non-dimensional process obtained dividing the process $\underline{x}_j^{(\Delta)}$ by its mean at the largest scale Δ_{max} (or equivalently $\lambda=1$)*
$s(w)$	*Power spectrum of the continuous-time (instantaneous) process, where w is the frequency*
$s_d^{(\Delta)}(\omega)$	*Power spectrum of the discrete-time process, where $\omega = w\Delta$ is nondimensionalized frequency*
$s^{\#}(w)$	*The asymptotic slopes of the power spectrum s(w)*
$c(\tau)$	*Autocovariance function for the continuous-time process Cov[$\underline{x}(t)$, $\underline{x}(t+\tau)$], where τ is the continuous time lag*
$\rho(\tau)$	*Autocorrelation function for the instantaneous process*
$c_z^{(\Delta)}$	*Autocovariance function for the averaged process Cov[$\underline{x}_j^{(\Delta)}$, $\underline{x}_{j+z}^{(\Delta)}$] (with $c_0^{(\Delta)}=\gamma(\Delta)$), where z is the discrete time lag*
$\rho^{(\Delta)}(z)$	*Autocorrelation function for the averaged process, where z is the discrete time lag*
$\psi(w)$	*The climacogram-based pseudospectrum (CBPS)*
$\psi^{\#}(w)$	*The asymptotic logarithmic slope of CBPS*
$\underline{x}_{1,0}$	*Average rainfall intensity over time scale $f\Delta$ at the time origin (j=1), i.e. $\underline{x}_1^{(f\Delta)}$ (start vertex of the random cascade)*
μ_0, γ_0	*Mean and variance of $\underline{x}_{1,0}$*
b	*Branching number of the random cascade*
k	*Level of the cascade*
$\underline{w}_{j,k}$	*Random weight in the position j at the level k of the multiplicative cascade*

μ_w, γ_w	*Common mean and variance of the cascade weights* $\underline{w}_{j,k}$
$\underline{x}_{j,k}$	*Generic vertex of the cascade*
$\mu_{j,k}$, $\gamma_{j,k}$	*Mean and variance of* $\underline{x}_{j,k}$
$\rho_{j,k}(z)$	*The autocorrelation function for discrete-time lag z between two vertices of the cascade*
$h_{j,k}(z)$	*Exponent of formula of* $\rho_{j,k}(z)$*, which determines its dependence on the position j in the cascade*
$\underline{X}_1^{(f)}$	*Cumulative rainfall depth at the time origin (j=1) aggregated on the largest time scale f (i.e.,* $f\Delta$ *but* Δ *is omitted for convenience); it assumed log-normally distributed*
$\underline{\tilde{X}}_{1,0}$	*Auxiliary Gaussian random variable* $\underline{\tilde{X}}_1^{(f)} = \ln \underline{X}_1^{(f)}$ *of the aggregated HK process on the time scale f (start vertex of the random additive cascade)*
$\boldsymbol{\theta}$	*Vector of parameters of the HK downscaling model*
$\alpha(k)$, $\beta(k)$	*Scale-dependent functions derived to preserve the scaling properties of the actual (exponentiated) process*
$\underline{X}'_{j,k}$	*Variables modified by the power adjusting procedure*.

1. Introduction

Discipulus est prioris posterior dies.
Publilius Syrus

Change in geophysics has been studied since the birth of science and philosophy, but in modern times change has been particularly accelerated due to radical developments in demography, technology and life conditions. Therefore, during recent decades, there has been a growing interest in research activities on change in geophysics and its interaction with human society. A major example in this respect is given by the Intergovernmental Panel on Climate Change (IPCC), which was set up in 1988 by the World Meteorological Organization (WMO) and United Nations Environment Programme (UNEP) to provide policymakers with regular assessments of the scientific basis of climate change, its impacts and future risks, and options for adaptation and mitigation. Furthermore, the new scientific initiative of the International Association of Hydrological Sciences (IAHS) for the decade 2013–2022, entitled "Panta Rhei – Everything Flows" (Montanari et al., 2013), is dedicated to research activities on change in hydrology and society.

The practical purpose of all these activities is to improve our capability to make predictions of geophysical processes to support sustainable societal development in a changing environment. In order to describe the predictability of change, we adopt herein an interesting hierarchical chart by Koutsoyiannis (2013a) reported in Fig. 1.1. Change is regular in simple systems (left part of the graph), and therefore it is predictable using equations of dynamical systems (periodic or aperiodic). Nonetheless, in geophysics we are commonly interested in more complex systems with long time horizons (right part of the graph), where change is unpredictable in deterministic terms, or random. The term randomness is usually associated to stochastic processes whose samples are regarded as a sequence of independent and identically distributed random variables (pure randomness). This is a basic assumption of classical statistics, but there is ample practical evidence that this wish does not always become a reality (Beran, 1994). By the way, it has been observed empirically that correlations between distant samples decay to zero at a slower rate than one would expect from independent data or even data following classical

ARMA- or Markov-type models (Box et al., 1994). In these cases, we should assume a structured randomness. As will be seen next, the structured randomness is enhanced randomness, expressing enhanced unpredictability of enhanced multi-scale change.

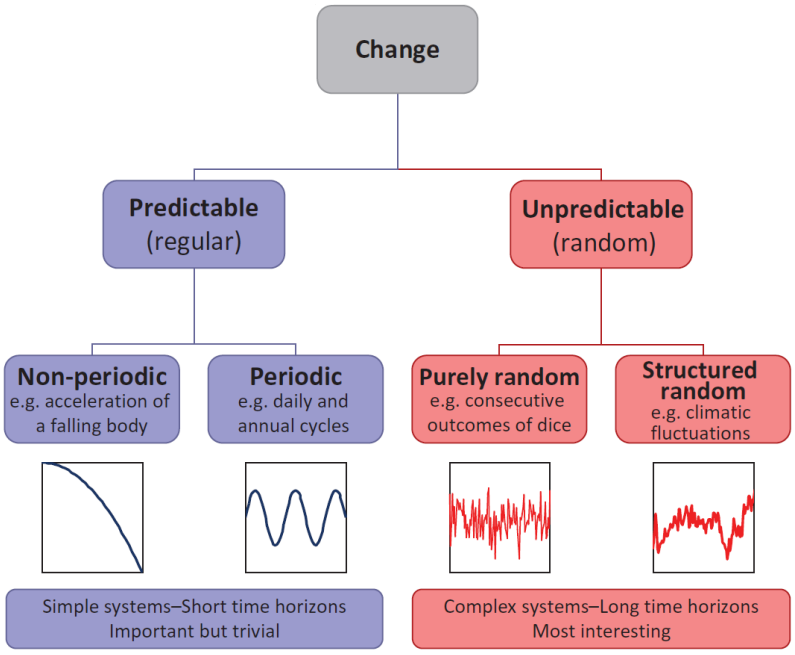


Figure 1.1 – Hierarchical chart describing the predictability of change (Koutsoyiannis, 2013a).

According to the common view, natural processes are composed of two different, usually additive, parts or components: deterministic (signal) and random (noise). This distinction implies that there is some signal that contains information, which is contaminated by a (random) noise. In this view, randomness is cancelled out at large time scales and cannot produce long-term change. In other words, we usually assume that natural changes are just a short-term "noise" superimposed on the daily and annual cycles in a scene that is static and invariant in the long run, except when an extraordinary forcing produces a long-term change. However, this view may not have a meaning in geophysics, as Nature's signs are "signals" in their entirety even though they may look like "noise". Moreover, change occurs on all time scales, from minute to geological, but our limited

senses and life span, as well as the short time window of instrumental observations, restrict our perception to the most apparent daily to yearly variations. As a result, long-term changes are much more frequent and intense than commonly perceived and, simultaneously, the future states are much more uncertain and unpredictable on long time horizons than implied by standard approaches (Markonis and Koutsoyiannis, 2013).

We endorse herein a different perspective (see also Koutsoyiannis, 2010). Randomness is simply viewed as unpredictability and coexists with determinism (which, in turn, could be identified with predictability) in the same natural process: the two do not imply different types of mechanisms or different parts or components in the time evolution, they are not separable or additive components. It is a matter of specifying the time horizon and scale of prediction to decide which of the two dominates. For long time horizons (where the specific length depends on the system), all is random – and not static.

Empirical evidence suggests that long historical hydroclimatic series may exhibit a behaviour very different from that implied by pure random models. To demonstrate this, two real-world examples are used (Koutsoyiannis, 2002). The first is a very long record: the series of standardised tree-ring widths from a palaeoclimatology study at Mammoth Creek, Utah, for the years 0–1989 (1990 values) (Graybill, 1990). The second example is the most intensively studied series, which also led to the discovery of the Hurst phenomenon (Hurst, 1951): the series of the annual minimum water level of the Nile River for the years 622–1284 A.D. (663 observations), measured at the Roda Nilometer near Cairo (Toussoun, 1925; Beran, 1994). The data values are plotted vs time for both example data sets in Figs. 1.2 and 1.3, respectively. In addition, the 5-year and 25-year averages are shown, which represent the mean aggregated processes at time scales $\Delta = 5$ and 25, respectively. For comparison, series of white noise with mean and standard deviation identical to those of standardized tree rings and annual minimum water levels are also shown. It is observed that fluctuations of the aggregated processes, especially for $\Delta = 25$, are much greater in the real-world time series than in the white noise series. Thus, the existence of fluctuations in a time series at large scales distinguishes it from random noise. When one looks only at short time periods, then there seem to be cycles or local trends. However, looking at the whole series, there is no apparent persisting trend or cycle. It rather seems that cycles of (almost) all frequencies occur, superimposed and in random sequence.

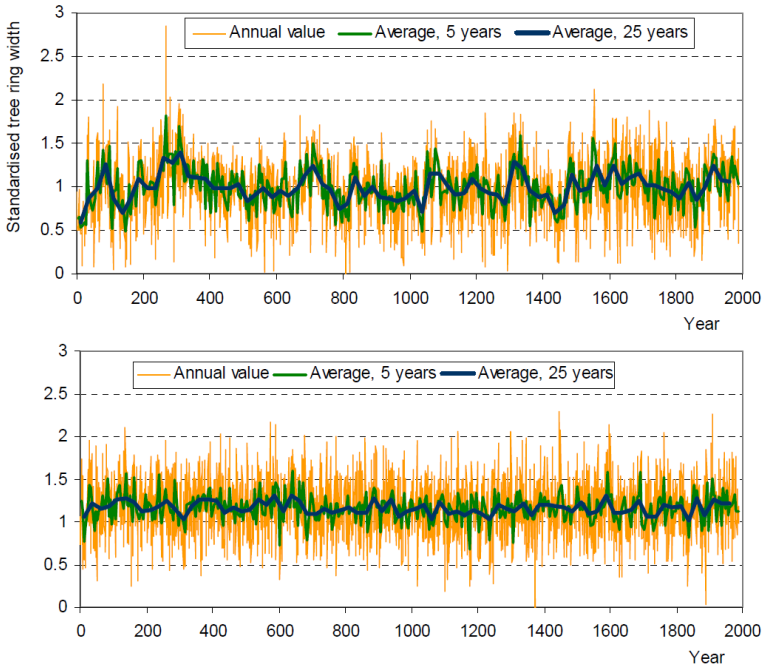


Figure 1.2 – Plot of standardized tree rings at Mammoth Creek, Utah (upper panel); white noise with same statistics (lower panel) (Koutsoyiannis, 2002).

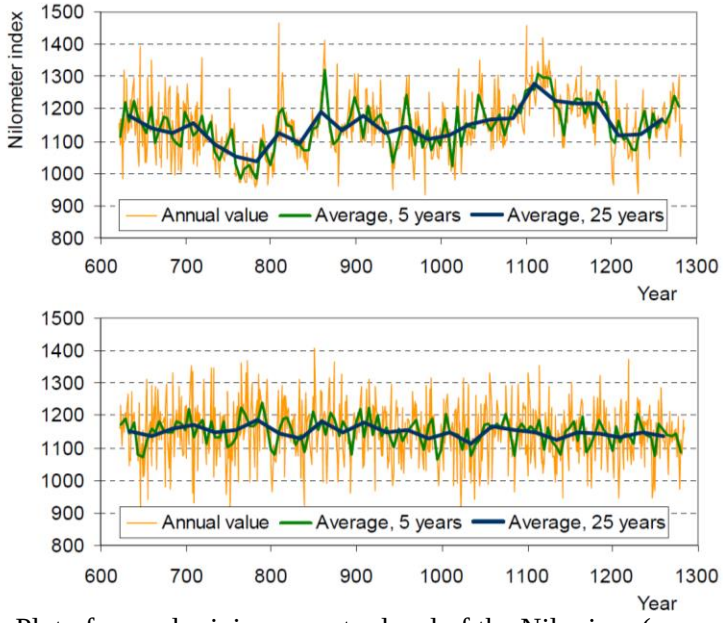


Figure 1.3 – Plot of annual minimum water level of the Nile river (upper panel); white noise with same statistics (lower panel) (Koutsoyiannis, 2002).

From the figures above, it can be noticed that in pure randomness (white noise) there are no long-term patterns; rather, the time series appear static in the long run. In real-world data, change is evident also at these scales. This change is unpredictable in deterministic terms and thus random and, more specifically, structured (or enhanced) random rather than purely random (Koutsoyiannis, 2013a).

1.1 Joseph effect and Hurst effect

The data sets in the previous section illustrated that correlations not only occur, but they also may persist for a long time. Many prominent applied statisticians and scientists recognized this many decades ago. In this section, we give a short overview on some of the important early references. This will also give rise to some principal considerations on the topic of long-range dependence.

Since ancient times, the Nile River has been known for its characteristic long-term behaviour. Long periods of dryness were followed by long periods of yearly returning floods. Floods had the effect of fertilising the soil so that in flood years the yield of crop was particularly abundant. On a speculative basis, one may find an early quantitative account of this in the Bible (Genesis 41, 29-30): "*Seven years of great abundance are coming throughout the land of Egypt, but seven years of famine will follow them*". We do not have any records of the water level of the Nile from those times. However, there are reasonably reliable historical records going as far back as 622 A.D. A data set for the years 622–1284 was discussed in the previous section (see Fig. 1.3, upper panel). It exhibits a long-term behavior that might give an "explanation" of the seven "good" years and seven "bad" years described in Genesis. There were long periods where the maximal level tended to stay high. On the other hand, there were long periods with low levels. Overall, the series seems to correspond to a stationary stochastic process, where there is no global trend. In reference to the biblical "seven years of great abundance" and "seven years of famine", Mandelbrot called this behaviour the *Joseph effect* (Mandelbrot, 1982; Mandelbrot and Wallis 1968, 1969; Mandelbrot and van Ness, 1968).

The first person to notice this behavior empirically was the British hydrologist H. E. Hurst (1951), when he was investigating the question of how to regularize the flow of the Nile River. More specifically, his discovery can be described as follows. Suppose we want to calculate the

capacity of a reservoir such that it is ideal for a given time span; assume that time is discrete and that there are no storage losses (caused evaporation, leakage, etc.). By ideal capacity, we mean that the outflow is uniform, that the water level in the reservoir is constant, and that the reservoir never overflows. Let $\underline{x}_j$ denote the inflow at time j (notice that we use the so-called Dutch convention according to which random variables are underlined; see Hemelrijk, 1966), and the partial sum $\underline{y}_\Delta=\underline{x}_1+\underline{x}_2+\ldots+\underline{x}_\Delta$ is the cumulative inflow up to time Δ, for any integer Δ. Then the ideal capacity can be shown to equal the *adjusted range* (Yevjevich, 1972):

$$\underline{R}_\Delta := \max_{1\le j\le\Delta}\left(\underline{y}_j - \frac{j}{\Delta}\underline{y}_\Delta\right) - \min_{1\le j\le\Delta}\left(\underline{y}_j - \frac{j}{\Delta}\underline{y}_\Delta\right) \tag{1.1}$$

In order to study the properties that are independent of the scale Δ, $\underline{R}_\Delta$ is standardised by the sample standard deviation $\underline{S}_\Delta$ of $\underline{x}_j$. This ratio is called the *rescaled adjusted range* or R/S-statistic. Hurst plotted the logarithm of R/S against several values of Δ. He observed that, for large values of Δ, log R/S was scattered around a straight line with a slope greater than 0.5.

This empirical finding was in contradiction to results for Markov processes, mixing processes, and other stochastic processes that were commonly used at that time. For any stationary process with short-range dependence, R/S should be asymptotically proportion al to $\Delta^{0.5}$ (Beran, 1994). Analogous considerations apply to many other geophysical records for which R/S is asymptotically proportional to Δ^H for $H > 0.5$. This is known as *Hurst effect*. Strikingly, the preeminent Soviet mathematician and physicist A. N. Kolmogorov had proposed a mathematical process that has the properties discovered 10 years later by Hurst in natural processes (Kolmogorov, 1940). Although the original name given by Kolmogorov was "Wiener's spiral", it later became more widely known by "fractional Brownian motion" or "fractional Gaussian noise" for the stationary increment process (Mandelbrot and van Ness, 1968). The latter is what we call hereinafter the Hurst-Kolmogorov (HK) process. The Hurst effect can be modelled by HK process with self-similarity (see next section) parameter $0.5 < H < 1$ (Hurst coefficient).

The reason why we prefer the term Hurst-Kolmogorov process is simple. We wish to associate the process on the one hand to Hurst, who was the first to observe and analyze the behaviour signified by this process in Nature, and on the other hand, to Kolmogorov, who was the first to point out the existence of this mathematical process. For a detailed review and

discussion of the names given to the Hurst phenomenon and its mathematical modelling, the reader is referred to Koutsoyiannis (2006).
This Hurst-Kolmogorov process is a model that simulates a stationary stochastic process.
A stochastic process $\underline{x}(t)$ is called stationary if its statistical properties are invariant to a shift of the time origin (Papoulis, 1991). This means that the processes $\underline{x}(t)$ and $\underline{x}(t+\tau)$ have the same statistics for any τ. Conversely, a process is non-stationary if some of its statistics are changing through time and their change is described as a deterministic function of time.
From a scientific point of view, it is not always satisfactory to model an observed phenomenon by a stationary process. For example, Klemeš (1974) showed that the Hurst phenomenon could be caused by non-stationarity in the mean and by random walks with one absorbing barrier. However, we would most likely want to use a stationary model as a null hypothesis unless physical considerations determined otherwise. Indeed, Koutsoyiannis (2002) offered a similar (from a practical point of view) explanation to that given by Klemeš (1974), but in a stationary setting. In essence, he assumed that the means are randomly varying on several timescales, thus regarding falling or rising trends, commonly traced in hydrological time series, as parts of large-scale random fluctuations rather than deterministic trends.

1.2 Stochastic modelling of change

In this section, we introduce stochastic processes that can be used to model data with the properties discussed in previous sections.
A simple way to understand the extreme variability of several geophysical processes over a practically important range of scales is offered by the idea that the same type of elementary process acts at each relevant scale. In a theoretical context, Kolmogorov (1940) introduced these types of processes that go under the name of "self-similar processes", which are based on a form of invariance with respect to changes of time scale. According to this idea, the part resembles the whole as quantified by so-called "scaling laws". Scaling behaviours are typically represented as power laws of some statistical properties, and they are applicable either on the entire domain of the variable of interest or asymptotically. If this random variable represents the state of a system, then we have the scaling in state, which refers to marginal distributional properties. This is to distinguish from another type of scaling, which deals with time-related

random variables: the scaling in time, which refers to the dependence structure of a process. Likewise, scaling in space is derived by extending the scaling in time in higher dimensions and substituting space for time (e.g. Koutsoyiannis et al., 2011). The scaling behaviour widely observed in the natural world (e.g. Newman, 2005) has often been interpreted as a tendency, driven by the dynamics of a physical system, to increase the inherent order of the system (self-organized criticality): this is often triggered by random fluctuations that are amplified by positive feedback (Bak et al., 1987). In another view, the power laws are a necessity implied by the asymptotic behaviour of either the survival or autocovariance function, describing, respectively, the marginal and joint distributional properties of the stochastic process that models the physical system. The main question is whether the two functions decay following an exponential (fast decay) or a power-type law (slow decay). We assume the latter to hold in the form of scaling in state (heavy-tailed distributions) and in time (long-term persistence), which have also been verified in geophysical time series (e.g. Markonis and Koutsoyiannis, 2013; Papalexiou et al., 2013). According to this view, scaling behaviours are just manifestations of enhanced uncertainty and are consistent with the principle of maximum entropy (Koutsoyiannis, 2011). The connection of scaling with maximum entropy constitutes also a connection of stochastic representations of natural processes with statistical physics. The emergence of scaling from maximum entropy considerations may thus provide theoretical background in modelling complex natural processes by scaling laws.

Since Kolmogorov's pioneering work, several researchers do not seem to have been aware of the existence or statistical relevance of such processes, until Mandelbrot and van Ness (1968) introduced them into statistics: "*By 'fractional Brownian motions' (fBm's), we propose to designate a family of Gaussian random functions defined as follows:* $\underline{B}(t)$ *being ordinary Brownian motion, and H a parameter satisfying* $0<H<1$, *fBm of exponent H* [denoted as $\underline{B}_H(t)$] *is a moving average of* $d\underline{B}(t)$, *in which past increments of* $\underline{B}(t)$ *are weighted by the kernel* $(t-s)^{H-1/2}$". As usual, t designates time, $-\infty<t<\infty$.

The increment process, $\underline{x}(t_2-t_1)=\underline{B}_H(t_2)-\underline{B}_H(t_1)$, is stationary and self-similar with parameter H, it is known as fractional Gaussian noise (i.e. Hurst-Kolmogorov process), and it is given by (see Mandelbrot and van Ness, 1968, p. 424):

$$\underline{x}(t_2 - t_1) = \frac{1}{\Gamma(H + 1/2)}\left(\int_{-\infty}^{t_2}(t_2 - s)^{H-1/2}\, d\underline{B}(s) - \int_{-\infty}^{t_1}(t_1 - s)^{H-1/2}\, d\underline{B}(s)\right) \quad (1.2)$$

which is a fractional integral in the sense of Weyl. The gamma function $\Gamma(\cdot)$ as denominator insures that, when $H - 0.5$ is an integer, a fractional integral becomes an ordinary repeated integral.

The value of the Hurst coefficient H determines three very different families of HK processes, corresponding, respectively, to: $0<H<0.5$, $0.5<H<1$, and $H=0.5$. The value 0.5 corresponds to white noise. For geophysical processes, we restrict ourselves to a discussion of HK processes positively correlated, i.e. $0.5<H<1$. Values $H<0.5$, characteristic of anti-persistence, are mathematically feasible (in discrete time) but physically unrealistic; specifically, for $0<H<0.5$ the autocorrelation for any lag is negative, while for small lags a physical process should necessarily have positive autocorrelation. High values of H, particularly those approaching 1, indicate enhanced change at large scales or strong clustering (grouping) of similar values, otherwise known as long-term persistence. In other words, in a stochastic framework and in stationary terms, change can be characterized by the Hurst coefficient.

1.3 Hurst-Kolmogorov process

The original Hurst's mathematical formulation, in terms of the so-called rescaled range, involves complexity and estimation problems as shown by Koutsoyiannis (2002). Actually, the mathematics to describe the HK process may be very simple. No more than the concept of standard deviation from probability theory is needed. Because a static system whose response is a flat line (no change in time) has zero standard deviation, we could recognize that the standard deviation is naturally related to change. To study change, however, we need to assess the standard deviation at several time scales, i.e. the relationship of the process standard deviation with the temporal scale of the process.

In order to improve understanding of Hurst-Kolmogorov process, we should describe the concept of "*local average*" of a stochastic process. Practical interest often revolves around local average or aggregates (temporal or spatial) of random variables, because it is seldom useful or necessary to describe in detail the local point-to-point variation occurring on a microscale in time or space. Even if such information were desired, it may be impossible to obtain: there is a basic trade-off between the

accuracy of a measurement and the (time or distance) interval within which the measurement is made (Vanmarke, 1983). For example, rain gauges (owing to size, inertia, and so on) measure some kind of local average of rainfall depth over time. Moreover, through information processing, "raw data" are often transformed into average or aggregate quantities such as, e.g., sub-hourly averages or daily totals.

Mathematically, let $\underline{x}(t)$ be a stationary stochastic process in continuous time t with mean $\mu=\mathrm{E}[\underline{x}]$, and autocovariance $c(\tau)=\mathrm{Cov}[\underline{x}(t), \underline{x}(t+\tau)]$, where τ is the time lag. Consider now the random process $\underline{x}_j^{(\Delta)}$ obtained by local averaging $\underline{x}(t)$ over the window Δ at discrete time steps j (=1, 2, …), defined as:

$$\underline{x}_j^{(\Delta)} = \frac{1}{\Delta}\int_{(j-1)\Delta}^{j\Delta} \underline{x}(t)\,\mathrm{d}t \qquad j=1,2,\ldots,n \tag{1.3}$$

where $n=T/\Delta$ is the number of the sample steps of $\underline{x}_j^{(\Delta)}$ in the observation period T_o, and $T=\lfloor T_o/\Delta \rfloor \Delta$ is the observation period rounded off to an integer multiple of Δ. The relationship between the processes $\underline{x}(t)$ and $\underline{x}_j^{(\Delta)}$ is illustrated in Fig. 1.4.

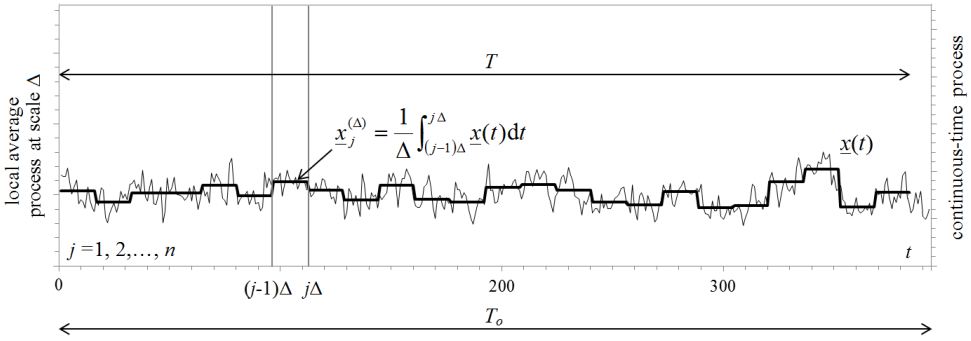


Figure 1.4 – Sketch of the local average process $\underline{x}_j^{(\Delta)}$ obtained by averaging the continuous-time process $\underline{x}(t)$ locally over intervals of size Δ.

The mean of the process $\underline{x}_j^{(\Delta)}$ is not affected by the averaging operation, i.e.:

$$\mathrm{E}\left[\underline{x}_j^{(\Delta)}\right] = \frac{1}{\Delta}\int_{(j-1)\Delta}^{j\Delta} \mathrm{E}\left[\underline{x}(t)\right]\mathrm{d}\,t = \mu \tag{1.4}$$

Let us now investigate the *climacogram* of the process $\underline{x}_j^{(\Delta)}$, which is defined to be the variance (or the standard deviation) of the time-averaged process $\underline{x}_j^{(\Delta)}$ as a function of the time scale of averaging Δ (Koutsoyiannis, 2010). The climacogram of $\underline{x}_j^{(\Delta)}$ can be calculated from the autocovariance

function $c(\tau)$ of the continuous-time process as follows (see e.g. Vanmarke, 1983, p. 186; Papoulis, 1991, p. 299):

$$\mathrm{Var}\left[\underline{x}_j^{(\Delta)}\right] = \gamma(\Delta) = \frac{2}{\Delta^2}\int_0^{\Delta}(\Delta-\tau)c(\tau)\,\mathrm{d}\tau = 2\int_0^1(1-\xi)c(\xi\Delta)\,\mathrm{d}\xi \qquad (1.5)$$

which shows that the climacogram $\gamma(\Delta)$ generally decreases with Δ and fully characterizes the dependence structure of $\underline{x}(t)$. The climacogram $\gamma(\Delta)$ and the $c(\tau)$ are fully dependent on each other; thus, the latter can be obtained by the former from the inverse transformation (see also Koutsoyiannis, 2013b):

$$c(\tau) = \frac{1}{2}\frac{\mathrm{d}^2\left(\tau^2\gamma(\tau)\right)}{\mathrm{d}\tau^2} \qquad (1.6)$$

Thus, the dependence structure of $\underline{x}(t)$ is represented either by the climacogram $\gamma(\Delta)$ or the autocovariance function $c(\tau)$. In addition, the Fourier transform of the latter, the spectral density function $s(w)$, where w is the frequency, is of common use. Selection of an analytical model for $c(\tau)$ or $s(w)$ is usually based on the quality of fit in the range of observed (observable) values of τ and w which, for reasons mentioned above, does not include the "microscale" ($\tau\to 0$ or $w\to\infty$) or in general the asymptotic behaviour. However, asymptotic stochastic properties of the processes are crucial for the quantification of future uncertainty, as well as for planning and design purposes (Montesarchio et al., 2009; Russo et al., 2006). Any model choice does imply an assumption about the nature of random variation asymptotically. Therefore, we may want this assumption (although fundamentally unverifiable) to be theoretically supported. In this context, Koutsoyiannis (2011) connected statistical physics (the extremal entropy production concept, in particular) with stochastic representations of natural processes, which are otherwise solely data-driven. He demonstrated that extremization of entropy production of stochastic representations of natural systems, performed at asymptotic times (zero or infinity) results in the Hurst-Kolmogorov process.

The HK process for local averages can be defined as a stationary stochastic process that, for any integers i and j and any time scales Δ and Λ, has the property:

$$\left(\underline{x}_j^{(\Delta)} - \mu\right) \overset{d}{=} \left(\frac{\Delta}{\Lambda}\right)^{H-1}\left(\underline{x}_i^{(\Lambda)} - \mu\right) \qquad (1.7)$$

where $\overset{d}{=}$ denotes equality in probability distributions, H is the Hurst coefficient, while μ is the mean of the process (cf. Eq. (1.4)). Thus, it can

be easily shown that the variance of $\underline{x}_j^{(\Delta)}$ (climacogram), for $i = j = \Delta = 1$, is a power law of the timescale Δ with exponent $2H–2$, such as:

$$\gamma(\Delta) = \Delta^{2H-2}\gamma(1) \tag{1.8}$$

Consequently, the autocorrelation function of $\underline{x}_j^{(\Delta)}$, for any aggregated timescale Δ, is independent of Δ (Koutsoyiannis, 2002):

$$\rho^{(\Delta)}(z) = \rho(z) = \frac{|z+1|^{2H}}{2} + \frac{|z-1|^{2H}}{2} - |z|^{2H} \tag{1.9}$$

In the discrete-time case, lag z is dimensionless. The instantaneous variance of the HK process is infinite. Therefore, HK process can be defined in continuous time by the following autocovariance function:

$$c(\tau) = \lambda(\alpha/\tau)^{2-2H} \qquad 0.5 < H < 1 \tag{1.10}$$

Thus, the autocovariance function $c(\tau)$ is a power law of the time lag τ with exponent $2H–2$, precisely the same as that of the climacogram $\gamma(\Delta)$. Consequently, it can be shown that the spectral density function $s(w)$ is also a power law of the frequency w with exponent $1–2H$. The three nominal parameters of the HK process are λ, α and H: the units of α and λ are $[\tau]$ and $[x]^2$, respectively, while H, the so-called Hurst coefficient, is dimensionless.

Substituting Eq. (1.10) in Eq. (1.5), we obtain the explicit formulation of the climacogram of the HK process as:

$$\gamma(\Delta) = \frac{\lambda(\alpha/\Delta)^{2-2H}}{H(2H-1)} \tag{1.11}$$

The climacogram contains the same information as the autocovariance function $c(\tau)$ or the power spectrum $s(w)$, because they are transformations one another. Its relationship with the latter is given by (Koutsoyiannis, 2013b):

$$\gamma(\Delta) = \int_0^\infty s(w)\frac{\sin^2(\pi w\Delta)}{(\pi w\Delta)^2}\,\mathrm{d}w \tag{1.12}$$

It has been observed that, when there is temporal dependence in the process of interest, the classical statistical estimation of the climacogram involves bias (Koutsoyiannis and Montanari, 2007), which is obviously transferred to transformations thereof, e.g. $c(\tau)$ or $s(w)$. In Chapter 2, we show how the bias in the climacogram estimation can be determined analytically and included in the estimation itself.

1.4 Climacogram

The logarithmic plot of standard deviation (or the variance) $\sigma(\Delta)$ vs. scale Δ, which has been termed climacogram in the previous section, is a very informative tool to study long-term change. Here we provide some empirical evidence. Let us take as an example the Nilometer time series described in Sect. 1 (the data are available from http://lib.stat.cmu.edu/S/beran), x_1, …, x_{663}, and calculate the sample estimate of standard deviation $\sigma(1)$, where the argument (1) indicates a time scale of 1 year. Then, we form a time series at time scale 2 (years) and calculate the sample estimate of standard deviation $\sigma(2)$:

$$x_1^{(2)} := \frac{x_1 + x_2}{2}, x_2^{(2)} := \frac{x_3 + x_4}{2}, \ldots, x_{331}^{(2)} := \frac{x_{661} + x_{662}}{2} \to \sigma(2) \qquad (1.13)$$

The same procedure is repeated with timescales Δ>2 up to scale $\Delta_{max}=\lfloor 663/10 \rfloor=66$, so that sample standard deviation can be estimated from at least 10 data values (Koutsoyiannis and Montanari, 2007):

$$x_1^{(66)} := \frac{x_1 + \cdots + x_{66}}{66}, \ldots, x_{10}^{(66)} := \frac{x_{595} + x_{660}}{66} \to \sigma(66) \qquad (1.14)$$

If the time series x_i represented a purely random process, the climacogram would be a straight line with slope –0.5, as implied by classical statistics (Beran, 1994). In real-world processes, the slope is different from –0.5, designated as H–1, where H is the so-called Hurst coefficient. This slope corresponds to the scaling law, which defines the Hurst-Kolmogorov (HK) process (see also Eq. (1.8)):

$$\sigma(\Delta) = \frac{\sigma(1)}{\Delta^{1-H}} \qquad (1.15)$$

It can be seen that if H > 1, then $\sigma(\Delta)$ would be an increasing function of Δ, which is absurd (averaging would increase variability, which would imply autocorrelation coefficients > 1). Fig. 1.5 below depicts the empirical climacogram of the Nilometer time series for time scales of averaging Δ ranging from one to 66 (years). It also provides comparison of the empirical climacogram with those of a purely random process, a Markov process and an HK process fitted to empirical data.

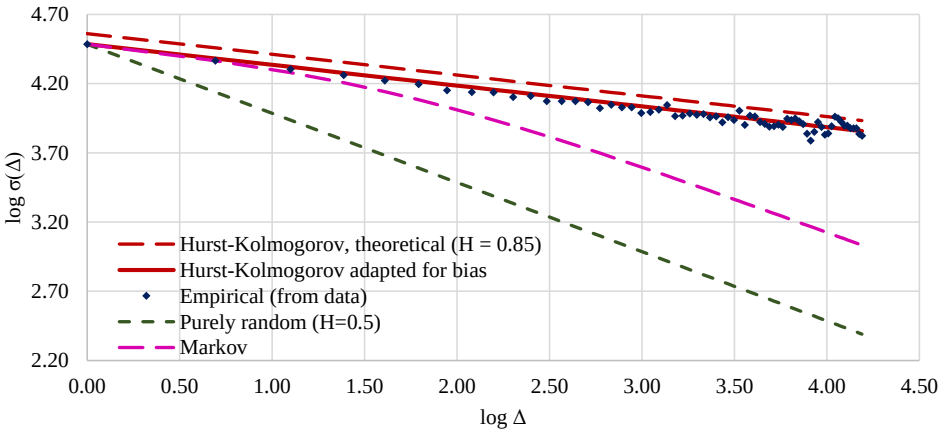


Figure 1.5 – Climacogram of Nilometer data and fitted theoretical ones of white noise (H=0.5), Markov and HK process (adapted from Koutsoyiannis, 2013a).

It should be noted that the standard statistical estimator of standard deviation σ, which is unbiased for independent samples (H = 0.5), becomes biased for a time series with HK behavior (0.5 < H < 1). It is thus essential that, if the sample statistical estimates are to be compared with the model Eq. (1.15), the latter must have been adapted for bias before the comparison (we followed the procedure given by Koutsoyiannis, 2003). Furthermore, in Fig. 1.5, we plotted also the climacogram of another stochastic process commonly used in many disciplines, i.e. the AR(1) process (autoregressive process of order 1), which is essentially a Markov process in discrete time (Box et al., 1994). For this process, the theoretical climacogram is given by (see Koutsoyiannis, 2002):

$$\sigma(\Delta) = \frac{\sigma(1)}{\sqrt{\Delta}} \sqrt{\frac{1+\rho}{1-\rho} - \frac{2\rho\left(1-\rho^{\Delta}\right)}{\Delta(1-\rho)^2}} \tag{1.16}$$

where the single parameter ρ is the lag-1 autocorrelation coefficient. This correlation implies some statistical bias in the estimation of σ from a sample, but this is negligible unless ρ is very high (close to 1). It can be seen that for large scales Δ, $\sigma(\Delta) \sim 1/\Delta^{0.5}$ and thus the climacogram of the AR(1) process behaves similarly to that of white noise, i.e. it has asymptotic slope –0.5.

In Fig. 1.5, the slope of the empirical climacogram is clearly different from –0.5, i.e. that corresponding to a purely random process and a Markov process, and is consistent with the HK behaviour with H = 0.85.

Essentially, the HK behaviour manifests that long-term changes are much more frequent and intense than is commonly perceived, and that the future states are much more uncertain and unpredictable on long time horizons (because the standard deviation is larger) than implied by pure randomness or Markov-type models.

1.5 Outline of the thesis

The following chapters expand on the approach described in the foregoing sections, with an emphasis on procedures for statistical inference and modelling. Indeed, the purpose of this thesis is neither to review the state of the art of the research related to the Hurst phenomenon, nor to give the complete mathematical details of it (see e.g. Beran, 1994). We rather aim to describe how to infer and model statistical properties of natural processes exhibiting scaling behaviours. Specifically, we investigate the dramatic increase of uncertainty in statistical estimations, and propose a simple and parsimonious model that respects the Hurst phenomenon.

Chapter 2 is concerned with the statistical implications of scaling behaviours in state (sub-exponential distribution tails) and in time (strong time dependence), which have been verified in geophysical time series. In statistical terms, this is translated in a departure from the (possibly tacit) assumptions underlying classical statistical approaches, which are commonly used in inference from time series of geophysical processes (see also Lombardo et al., 2014).

Chapter 3 deals with stochastic modelling of processes exhibiting scaling behaviour. In particular, we focus on rainfall downscaling in time. Generating finer scale time series of rainfall that are statistically consistent with any given coarse-scale totals is, indeed, an important and open issue in hydrology. We propose a stationary downscaling model, based on the HK process, which is characterized by a cascade structure similar to that of the most popular multiplicative random cascade models (see also Lombardo et al., 2012).

Finally, Chapter 4 contains some discussions and conclusions.

2. Geophysical inference

Le doute est un hommage rendu à l'espoir.
Isidore Ducasse

Due to the complexity of geophysical processes, the conducting of typical tasks, such as estimation, prediction and hypothesis testing, heavily rely on available data series and their statistical processing. The latter is usually based upon classical statistics. The classical statistical approaches, in turn, rely on several simplifying assumptions, tacit or explicit, such as independence in time and exponentially decaying distribution tails, which are invalidated in natural processes thus causing bias and uncertainty in statistical estimations. Indeed, as we showed in Chapter 1, the study of natural processes reveals scaling behaviours in state (departure from exponential distribution tails) and in time (departure from independence). Surprisingly, all these differences are commonly unaccounted for in most statistical analyses of geophysical processes, which may result in inappropriate modelling, wrong inferences and false claims about the properties of the processes studied.

In the literature, natural processes showing scaling behaviour are often classified as multifractal systems (i.e. multiscaling) that generalize fractal models, in which a single scaling exponent (the fractal dimension) is enough to describe the system dynamics. For a detailed review on the fundamentals of multifractals, the reader is referred to Schertzer and Lovejoy (2011).

Multifractal models generally provide simple power-law relationships to link the statistical distribution of a stochastic process at different scales of aggregation. All power laws with a particular scaling exponent are equivalent up to constant factors, since each is simply a scaled version of the others. Therefore, the multifractal framework provides parsimonious models to study the variability of several natural processes in geosciences, such as rainfall. Rainfall models of multifractal type have, indeed, for a long time been used to reproduce several statistical properties of actual rainfall fields, including the power-law behaviour of the moments of different orders and spectral densities, rainfall intermittency and extremes (see e.g. Koutsoyiannis and Langousis (2011) and references therein). However, published results vary widely, calling into question whether

rainfall indeed obeys scaling laws, what those laws are, and whether they have some degree of universality (Nykanen and Harris, 2003; Veneziano et al., 2006; Molnar and Burlando, 2008; Molini et al., 2009; Serinaldi, 2010; Verrier et al., 2010, 2011; Gires et al., 2012; Veneziano and Lepore, 2012; Papalexiou et al. 2013). In fact, significant deviations of rainfall from multifractal scale invariance have also been pointed out. These deviations include breaks in the power-law behaviour (scaling regimes) of the spectral density (Fraedrich and Larnder, 1993; Olsson, 1995; Verrier et al., 2011; Gires et al., 2012), lack of scaling of the non-rainy intervals in time series (Veneziano and Lepore, 2012; Mascaro et al., 2013), differences in scaling during the intense and moderate phases of rainstorms (Venugopal et al., 2006), and more complex deviations (Veneziano et al., 2006; Marani, 2003).

Multifractal signals generally obey a scale invariance that yields power law behaviours for multi-resolution quantities depending on their scale Δ. These multi-resolution quantities at discrete time steps (j = 1, 2, …), denoted by $\underline{x}_j^{(\Delta)}$ in the following, are local time averages in boxes of size Δ (see also Sect. 1.3). This is the basis of the fixed-size box-counting approach (see e.g. Mach et al., 1995). For multifractal processes, one usually observes a power-law scaling of the form:

$$\mathrm{E}\left[\left(\underline{x}_j^{(\Delta)}\right)^q\right] \propto \Delta^{-K(q)} \tag{2.1}$$

at least in some range of scales Δ and for some range of orders q. The function E[·] denotes expectation (ensemble average) and $K(q)$ is the moment scaling function. Generally, the multifractal behaviour of a physical system is directly characterized by the multiscaling exponents $K(q)$, whose estimation relies on the use of the sample q-order moments at different scales Δ and their linear regressions in log-log diagrams.

A fundamental problem in the multifractal analysis of datasets is to estimate the moment scaling function $K(q)$ from data (Villarini et al., 2007; Veneziano and Furcolo, 2009). Considerable literature has been dealing with estimation problems in the context of so-called scaling multifractal measures for three decades at least (see e.g. Grassberger and Procaccia, 1983; Pawelzik and Schuster, 1987; Schertzer and Lovejoy, 1992; Ashkenazy, 1999; Mandelbrot, 2003; Neuman, 2010). Interestingly, Mandelbrot (2003) and Neuman (2010) recognize the crucial role played by time dependence in estimating multifractal properties from finite length data. Nonetheless, herein we remain strictly within the framework of the standard statistical formalism, which is actually a novelty with

respect to the literature cited above. In this context, we highlight the problematic estimation of moments for geophysical processes, because the statistical processing of geophysical data series is usually based upon classical statistics. In many studies, it has been a common practice to neglect this problem, which is introduced when the process exhibits dependence in time and is magnified when the distribution function significantly departs from the Gaussian form, which itself is an example of an exceptionally light-tailed distribution. In their pioneering work on statistical hydrology, Wallis et al. (1974) already provided some insight into the sampling properties of moment estimators when varying the marginal probability distribution function of the underlying stochastic process. The main results of the paper agree well with those found in the following sections, but its Monte Carlo experiments were carried out under a classical statistical framework assuming independent samples.

The purpose of this Chapter is to explore, at different timescales, the information content in estimates of raw moments of processes exhibiting temporal dependence. In order for the true moments to be fully known a priori, we use synthetic examples in a Monte Carlo simulation framework. We explore processes with both normal and non-normal distributions including ones with heavy tails. We show that, even in quantities whose estimates are in theory unbiased, the dependence and non-normality affect significantly their statistical properties, and sample estimates based on classical statistics are characterized by high bias and uncertainty. In particular, statistical methods that use high order moments (> 3) are questionable (see Sect. 2.1 below). In particular, we suggest that, because of estimation problems, the use of moments of order higher than two should be avoided, either in justifying or fitting models. Nonetheless, in most problems the first two moments provide enough information for the most important characteristics of the distribution. Finally, in Sect. 2.2 we put the emphasis on autocorrelations and spectra (only involving second-order moments), and specifically study their estimation problems.

We believe this process is critical for practitioners and researchers in geophysics to gain insights into the ways they can use statistical tools reliably.

2.1 Multifractal analysis

Multifractal analysis has been used in several fields in science to characterize various types of datasets, which have been investigated by

means of the mathematical basis of multifractal theory. This is the basis for a series of calculations that reveal and explore the multiple scaling rules, if any, from datasets, in order to calibrate multifractal models. From a practical perspective, multifractal analysis is usually based upon the following steps (Lopes and Betrouni, 2009).

- Estimate the sample raw moments of different orders q over a range of aggregation scales Δ.
- Plot the sample q-moments against the scale Δ in a log-log diagram.
- Fit least-squares regression lines (one for each order q) through the data points.
- Estimate the multiscaling exponents $K(q)$ as the slopes of regression lines (see Eq. (2.1)).

The classical estimator of the q-th raw moment of the local average process $\underline{x}_j^{(\Delta)}$ is:

$$\underline{m}_q^{(\Delta)} = \frac{1}{n}\sum_{j=1}^{n}\left(\underline{x}_j^{(\Delta)}\right)^q \tag{2.2}$$

High moments, i.e. $q \geq 3$, mainly depend on the distribution tail of the process of interest. If we assume, for reasons mentioned in Sect. 1.2, scaling in state, i.e. a power-type (e.g. Pareto, see below) tail, then raw moments are theoretically infinite beyond a certain order $q_{\max}$. However, their numerical estimates from a time series by Eq. (2.2) are always finite, thus resulting in infinite biases from a practical perspective, because the estimate is a finite number while the true value is infinity. Even below $q_{\max}$, where it can be proved that the estimates are unbiased, we show that the estimation of moments can be still problematic. It is easily shown, indeed, that the expected value of the moment estimator equals its theoretical value $\mathrm{E}[(\underline{x}_j^{(\Delta)})^q]=\mu_q^{(\Delta)}$ for any timescale Δ, such as:

$$\mathrm{E}\left[\underline{m}_q^{(\Delta)}\right] = \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}\left[\left(\underline{x}_j^{(\Delta)}\right)^q\right] = \mu_q^{(\Delta)} \tag{2.3}$$

which can be used to derive the variance of the moment estimator as follows:

$$\begin{aligned}\mathrm{Var}\left[\underline{m}_q^{(\Delta)}\right] &= \mathrm{E}\left[\left(\underline{m}_q^{(\Delta)}\right)^2\right] - \mathrm{E}\left[\underline{m}_q^{(\Delta)}\right]^2 \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}\left[\left(\underline{x}_j^{(\Delta)}\right)^q\left(\underline{x}_i^{(\Delta)}\right)^q\right] - \left(\mu_q^{(\Delta)}\right)^2\end{aligned} \tag{2.4}$$

This quantity can be assumed as a measure of uncertainty in the estimation of the q-th moment of the local average process $\underline{x}_j^{(\Delta)}$. Therefore, the estimator $\underline{m}_q^{(\Delta)}$ is theoretically unbiased, because of Eq. (2.3), but involves uncertainty, quantified by Eq. (2.4), which is expected to depend on statistical properties of the instantaneous process $\underline{x}(t)$ (i.e., marginal and joint distributional properties), the averaging scale Δ, the sample size n, and the moment order q. In the next sub-sections, we show how the problems of uncertainty in statistical estimation may be extremely remarkable when using uncontrollable quantities (e.g. high order moments) to justify or calibrate stochastic models.

2.1.1 Estimation of the mean

The (unbiased) estimator of the common mean μ of the local average process $\underline{x}_j^{(\Delta)}$ is given by Eq. (2.3) for q=1:

$$\underline{m}_1^{(\Delta)} = \frac{1}{n}\sum_{j=1}^{n} \underline{x}_j^{(\Delta)} = \underline{x}_1^{(T)} \tag{2.5}$$

where T is the largest timescale of averaging multiple of Δ in a given observation period T_o (see e.g. Fig. 1.4).

Recalling the Eq. (1.5), we can provide an analytical formulation for the variance of the time-averaged process $\underline{x}_1^{(T)}$ as a function of the time scale of averaging T, which actually equals the variance of estimator of the first-order moment given in Eq. (2.5), such as:

$$\mathrm{Var}\left[\underline{m}_1^{(\Delta)}\right] = \mathrm{Var}\left[\underline{x}_1^{(T)}\right] = \gamma(T) = \frac{2}{T^2}\int_0^T (T-\tau)c(\tau)\,\mathrm{d}\tau \tag{2.6}$$

Therefore, the estimator $\underline{m}_1^{(\Delta)}$ is a function of the dependence structure of the continuous-time (autocovariance function $c(\tau)$) process $\underline{x}(t)$, and the rounded observation period T. Note that the uncertainty in the estimation of the sample mean is independent of the timescale of averaging Δ while it depends on the observation period T.

Let us now consider the Hurst-Kolmogorov process. Hence, the climacogram $\gamma(T)$ takes the form of equation (1.8), as:

$$\gamma(T) = T^{2H-2}\gamma(1) \tag{2.7}$$

In Fig. 2.1 below, we show how the temporal dependence (governed by the Hurst coefficient H for the HK process) influences the reliability of moment estimates. For simplicity and without loss of generality, we plot the ratio of $\mathrm{Var}[\underline{m}_1^{(\Delta)}]$ to $\mathrm{Var}[\underline{x}_j^{(\Delta)}]$ for Δ=1 against the scale T, which

equals the sample size $n=T$ for $\Delta=1$. As a consequence of Eq. (2.7), the ratio is given by:

$$\frac{\mathrm{Var}\left[\underline{m}_1^{(\Delta=1)}\right]}{\mathrm{Var}\left[\underline{x}_j^{(\Delta=1)}\right]} = \frac{\gamma(T)}{\gamma(1)} = n^{2H-2} \tag{2.8}$$

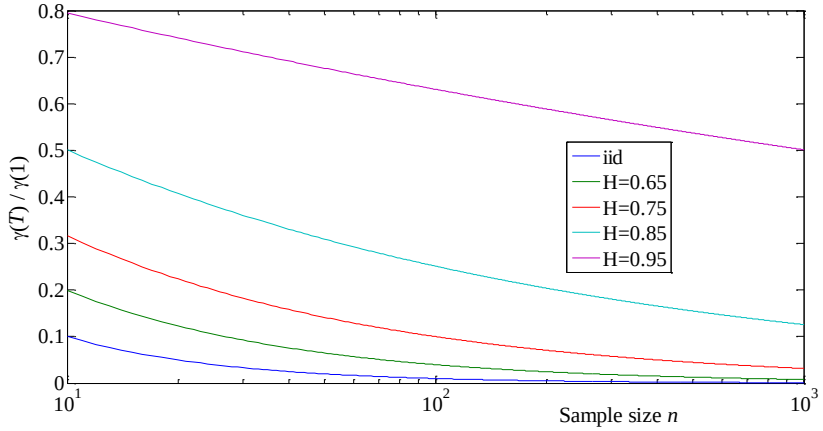


Figure 2.1 – Estimator variance of the mean of the local average process $\underline{x}_j^{(\Delta=1)}$ standardized by the process variance, i.e. $\mathrm{Var}[\underline{m}_1^{(\Delta=1)}]/\mathrm{Var}[\underline{x}_j^{(\Delta=1)}]=\gamma(T)/\gamma(1)$, plotted against the sample size $n=T$ for $\Delta=1$.

Notice that large values of H result in much higher ratio than in the iid case (which is given by $1/n$), and the convergence to the iid case is extremely slow (see Fig. 2.1). In essence, it can be argued that the greater the dependence in time, the harder it is to estimate the moment; in the sense that larger samples are required in order to obtain estimates of similar quality.

2.1.2 Estimation of higher moments

Let us now investigate the behaviour of estimators of higher order moments ($q>1$) when the underlying random process exhibits dependence in time and when changing the process marginal distribution; this can be done by Monte Carlo simulation. Specifically, we use the Gaussian distribution and three one-sided distributions whose tails are sub-exponential, i.e. heavier than the former (as observed in several geophysical processes). All synthetic time series are generated in a way to

have similar dependence structures based on the HK process, which are therefore governed by the Hurst coefficient *H*.

In this study, we estimate the performance of *q*-th moment estimators for four different common tail types (ordered from heavier to lighter): the Pareto, the lognormal, the Weibull and the Gaussian tails (see e.g. El Adlouni et al., 2008; Papalexiou et al., 2013). The Pareto and the lognormal distributions belong to the sub-exponential class (with a tail heavier than the exponential one) and are considered as heavy-tailed distributions; the Weibull can belong to both classes, depending on the values of its shape parameter (see below), while the Gaussian distribution has essentially a tail thinner than the exponential one (hyper-exponential), and it is considered as light-tailed distribution.

The Pareto is the only power-type distribution, while the rest three are of exponential type with all their moments finite. Specifically, we use the Pareto type II distribution, defined in [0, ∞), with survival function:

$$\overline{F}_{PII}(x) = P\{\underline{x} > x\} = \left(1 + \kappa \frac{x}{\beta}\right)^{-1/\kappa} \tag{2.9}$$

where β>0 is the scale parameter, and κ>0 the shape parameter. The latter, also known as the tail index, controls the asymptotic behaviour of the tail, which is given by $x^{-1/\kappa}$; as the value of κ increases the tail becomes heavier and consequently extreme values occur more frequently. For $\kappa \to 0$ the distribution tail degenerates to the exponential tail, while for $\kappa \geq 0.5$ the distribution has infinite variance. Indeed, the shape parameter κ unequivocally defines the order $q_{\max}=1/\kappa$ beyond which the *q*-th moments are theoretically infinite, i.e. $E[(\underline{x}_j^{(\Delta)})^q]=\infty$ for $q \geq 1/\kappa$; in our study we assume κ=0.2, and thus $q_{\max}$=5.

The lognormal distribution, also defined in [0, ∞), is very commonly used in geosciences and has the survival function:

$$\overline{F}_{LN}(x) = \frac{1}{2}\operatorname{erfc}\left(\ln\left(\left(\frac{x}{\beta}\right)^{\frac{1}{\kappa\sqrt{2}}}\right)\right) \tag{2.10}$$

where $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = 2/\sqrt{\pi}\int_x^\infty \exp(-t^2)\,\mathrm{d}t$ is the complementary error function, β is the scale parameter, and κ>0 is the shape parameter that controls the behaviour of the tail (notice some differences from the more typical notational convention in the literature; see Forbes et al. (2011) p. 131, for further details). Despite all its moments being

theoretically finite, the lognormal distribution is very similar in shape to a power-type distribution (Pareto), in the sense that the two distributions appear almost indistinguishable from each other for a large portion of their body (Mitzenmacher, 2004). Therefore, lognormal is regarded as a heavy-tailed distribution.

Another widely used distribution is the Weibull distribution, again defined in [0, ∞). Its survival function is a stretched exponential function (obtained by inserting a fractional power law into the exponential function), i.e.:

$$\overline{F}_W(x) = \exp\left(-\left(\frac{x}{\beta}\right)^{\kappa}\right) \tag{2.11}$$

where $\beta>0$ is the scale parameter, and the stretching exponent $0<\kappa<1$ (shape parameter) actually modifies the shape of the exponential distribution so as to obtain a heavier tail. Consequently, the Weibull distribution can be regarded as a generalization of the exponential distribution, which is recovered with $\kappa=1$. The case with $\kappa>1$ (compressed exponential function, i.e., a tail lighter than the exponential one) has less practical importance, with the notable exception of $\kappa=2$, which gives the Rayleigh distribution, closely related to the Gaussian distribution.

2.1.3 Monte Carlo simulation

As the lognormal model has been the most common in multifractal literature, we start our study from this model. For the Monte Carlo simulation we use the model introduced in the next Sect. 3.2, which follows a disaggregation approach. In that respect it resembles the discrete multifractal cascade models yet it is a fully consistent and fully controllable model, not affected by uncontrollable nonstationary issues that are typical in multifractal cascades (see Sect. 3.1). The model starts the generation from the coarsest scale and then disaggregates into finer scales applying a specific scale-dependent exponential transformation to the HK process in a way to preserve part of its scaling properties. For the Monte Carlo experiment we generate 30000 time series with sample size $n=2^{10}=1024$, unit mean, standard deviation σ = 1.29 and H=0.85. Later we will compare with the other models in a different setting, i.e. aggregating rather than disaggregating, using the same statistical properties (note that σ=1.29 is the standard deviation of the Pareto type II with unit mean and

tail index κ = 0.2). The results of the Monte Carlo simulation experiment are depicted in Figs. 2.2-2.5. Specifically, Fig. 2.2 shows the probability distribution of the natural logarithm of the ratio of q-th moment estimates to their expected values, i.e. the theoretical values following Eq. (2.3).

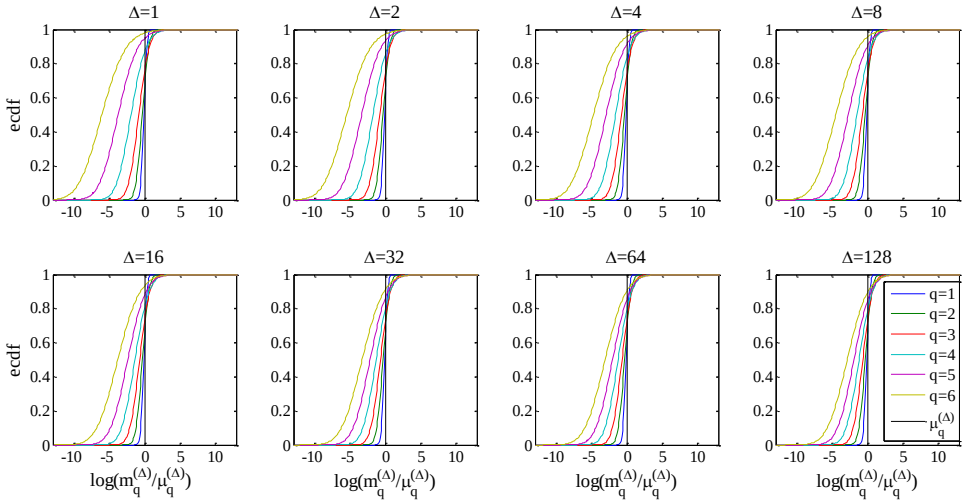


Figure 2.2 – Empirical cumulative distribution function (ecdf) of the natural logarithm of the ratio of q-th moment estimates to their expected values $E[(\underline{x}_j^{(\Delta)})^q]=\mu_q^{(\Delta)}$ when varying Δ.

It can be noticed that the information content of the sample moments strongly decreases when increasing the order q (i.e., the distribution is less concentrated around 0): only low moments have reasonably low variation, all others vary within several orders of magnitude (notice that the horizontal axis is logarithmic and spans more than 10 orders of magnitude!). Despite the sample raw moment being an unbiased estimator of the true (population) raw moment, the probability distribution of the statistical estimator is very broad and skewed. This is particularly the case for high moments. Note that the averaging scale Δ has negligible influence on the statistical characteristics of low moment estimators, while it slightly regularizes the behaviour of higher moment estimators.

In addition, in Fig. 2.3 below we show the empirical frequency distribution of the sample 5-th moment estimated from lognormal time series averaged locally over different timescales Δ. Again, here the bias is theoretically zero, but the most probable value of the moment estimate (the mode) is very different from its expected value. For example when Δ=1 (upper-left panel of Fig. 2.3), the mode of the distribution of $m_5^{(\Delta=1)}$

(green line) is almost two orders of magnitude less than the expected value (red line) and the probability of calculating from a unique sample a value equal to the mode is much greater (almost one order of magnitude) than the probability of obtaining the expected value itself. Recall that the expected value of the sample moment equals the true value of the moment, because of unbiasedness, but according to the distributions of Fig. 2.3 we can hardly expect the moment estimate from a unique sample to be close to this expected value. Increasing the averaging scale Δ reduces the difference between the mean and the mode. Nonetheless, this difference is still remarkable at large scales (see e.g. lower-right panel of Fig. 2.3).

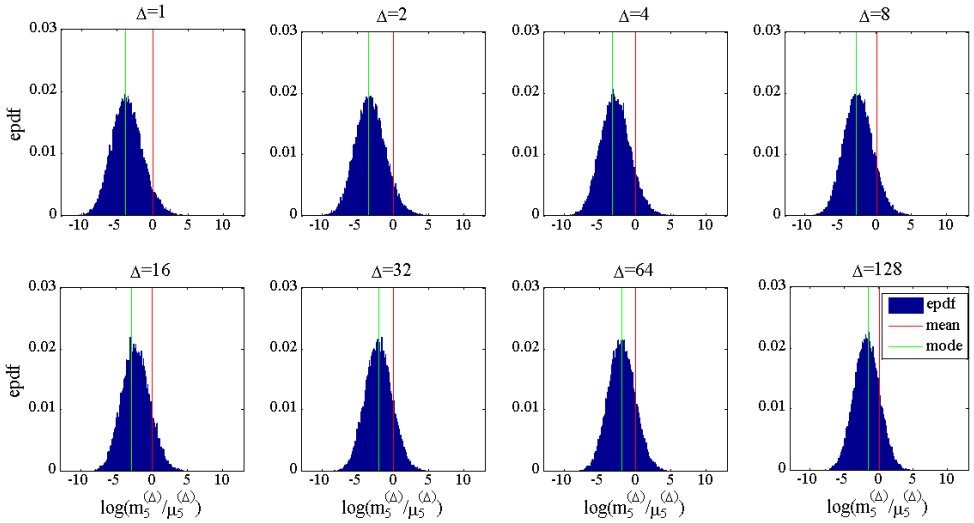


Figure 2.3 – Empirical probability density function (epdf) of the sample 5-th moment estimated from lognormal time series averaged locally over different timescales Δ.

The large difference between the mode and the expected value of the moment estimators is not the only problem. Another problem is the high estimation uncertainty. In order to illustrate the uncertainty in the moment estimation, Fig. 2.4 shows semi-logarithmic plots of the prediction intervals of the sample moments, calculated from the Monte Carlo simulations, against the moment order, for various scales Δ. The logarithmic scale on the vertical axis highlights the huge variability of estimates when the order increases. Note that the mean of raw moments (i.e., the true expected value) moves closer to the upper prediction limit for orders $q>3$, thus making the use of high moments unreliable.

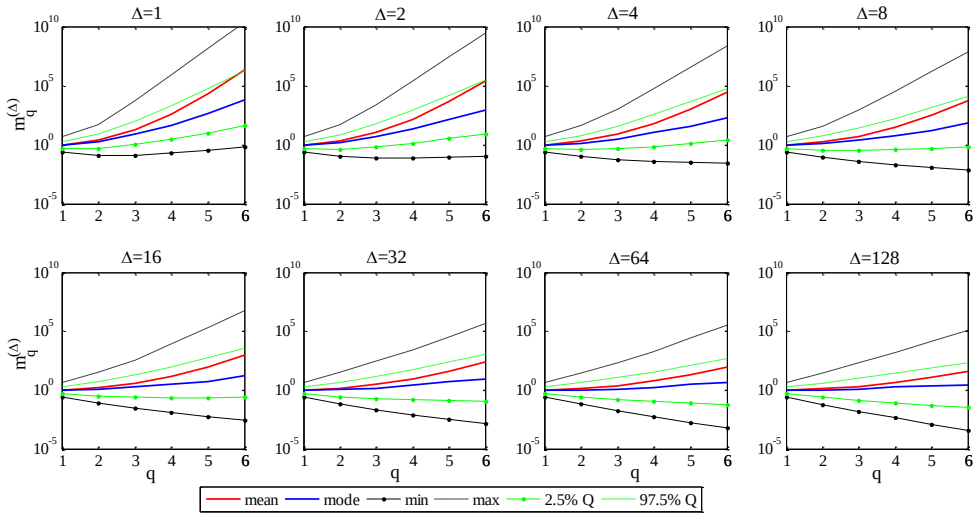


Figure 2.4 – Semi-logarithmic plots of the prediction intervals of the sample moments versus the order q for various timescales Δ, "Q" stands for quantile.

Furthermore, Fig. 2.5 depicts log-log diagrams of the prediction intervals of the sample moments against the scale of averaging Δ, for various orders q. In addition to the observations made with respect to Fig. 2.4, Fig. 2.5 shows that the increase of the averaging scale Δ has little influence on the variability of the moments, meaning that the sample size reduction is somewhat compensated by the time averaging.

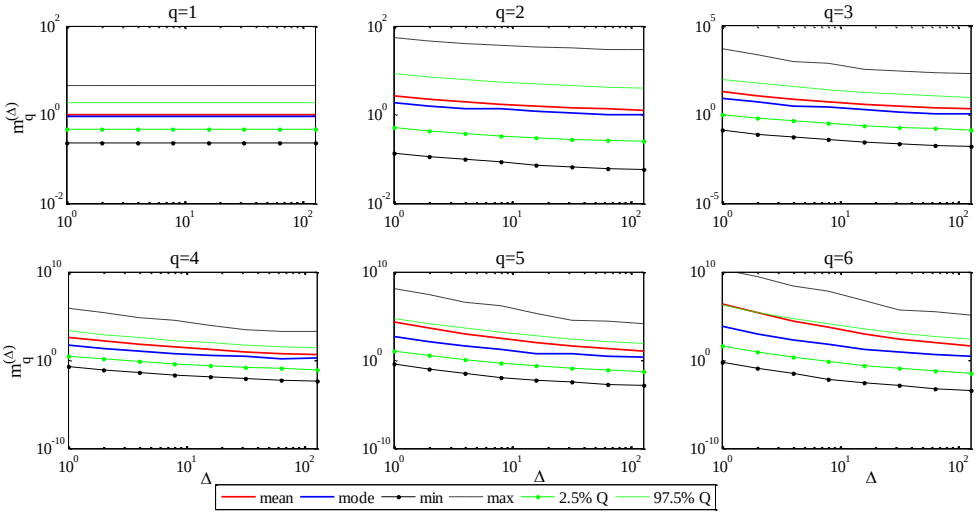


Figure 2.5 – Log-log plots of the prediction intervals of the sample moments versus the scale Δ for various orders q.

Nevertheless, it is clear that larger samples provide better estimates than smaller. For example, Meneveau and Sreenivasan (1991) propose a criterion of statistical convergence for the moments of local average processes, and find that data records of size 10^q may be sufficiently long to ensure statistical convergence for q-th order moments. However, this is not immediately straightforward in case of highly correlated data series, as we show in Fig. 2.1. To further investigate this issue accounting for the criterion of convergence above, in Fig. 2.6 we show the trend of the interquartile range (IQR) of the prediction intervals for the third (q=3) moment when increasing the sample size from 2^{10} to 2^{14} (the ensemble consists of 10000 lognormal time series for each sample size generated by our model described in Sect. 3.2). It can be noticed that the sample size should be increased more than one order of magnitude to obtain roughly a 10% improvement over the results presented in Fig. 2.4 for Δ=1.

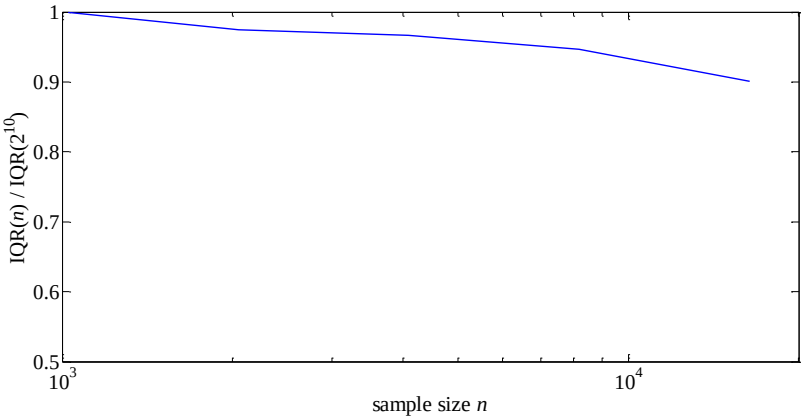


Figure 2.6 – Semi-logarithmic plot of the interquartile range (IQR) (standardized with respect to the IQR for n=2^{10}) of the prediction intervals for the third moment versus the sample size n for the lognormal series generated by our downscaling model (see Sect. 3.2).

In the second part of the Monte Carlo simulation experiment, we use a different approach, first generating at the finest scale and then aggregating into coarser scales. In this case we generate 30000 synthetic time series from the four distributions described in Sect. 2.1.2 above (ordered from heavier to lighter tail type: Pareto, lognormal, Weibull with shape parameter smaller than one and Gaussian) with characteristics same as those in the previous experiment. In this case, we investigate how the classical estimators of raw moments behave when varying the tail type of

the marginal distribution of the underlying stochastic process. To accomplish this aim, in Fig. 2.7 we plot on a semi-logarithmic scale the prediction intervals of the sample moments against the moment order (assuming Δ=1), for the four distributions. It can be seen that the tail type significantly influences the reliability of moment estimators. The heavier the distribution tail, the more uncertain the sample moments are. This is especially the case for high moments, because they depend enormously on the distribution tail and non-normality affects significantly their statistical properties. Analogous considerations apply to aggregated series (i.e., Δ>1).

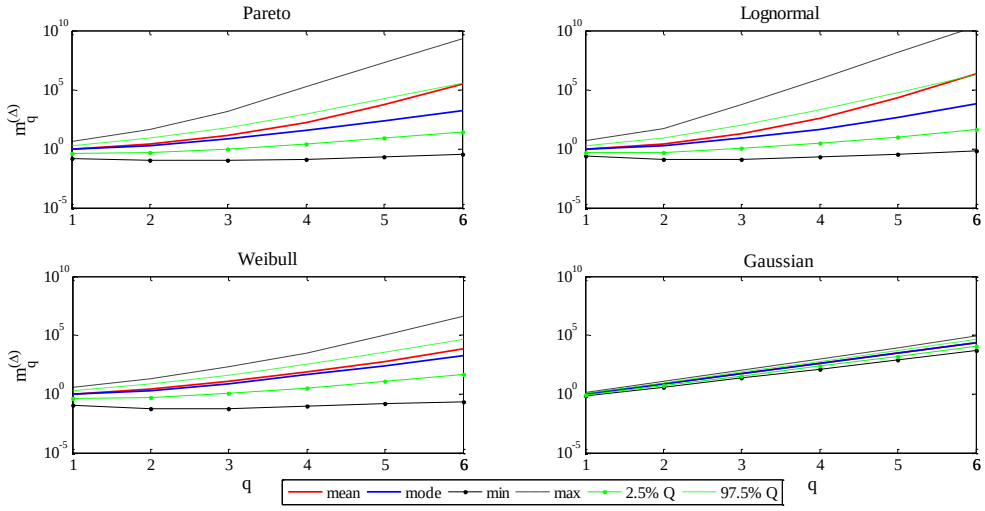


Figure 2.7 – Semi-logarithmic plots of the prediction intervals of the sample moments versus the order q for various marginal probability distributions, assuming Δ=1.

It is emphasized that the vertical axes in Fig. 2.7 span more than 10 orders of magnitude yet the prediction limits do not necessarily bracket the true value of the moment. Particularly for the Pareto distribution the true (population) values of the 5-th and 6-th moments are infinite while their statistical estimates are finite and the entire graph does not provide any hint that these high moments differ so essentially from the lower ones. Another important conclusion drawn from Fig. 8 is that the prediction limits in the case of the Gaussian distribution are dramatically narrower than in all other cases. As the Gaussian distribution has been dominating in classical statistical applications and perhaps in statistical thinking, this fact may explain why the multifractal applications were misled to neglect

the huge uncertainty of high moment estimates and its impact on modelling.

2.1.4 Empirical moment scaling function

Since the ultimate aim of a multifractal analysis is to study the scaling of raw moments, we have carried out some additional numerical investigations on the generated samples by simply taking an average slope of linear regressions of sample moments at different scales Δ in log-log diagrams (actually, this is commonly the case when dealing with real world data). Despite being not really crucial to the focus of our work (i.e. aiming to answer the question about how many raw moments we can estimate reliably), we believe it is worth exploring the variability in the estimates of the moment scaling function $K(q)$, when using the statistical tools which we cautioned against. To accomplish this purpose, we use the lognormal synthetic series generated by our downscaling model described in Sect. 3.2.

In order to estimate an empirical exponent function $K(q)$ describing the scaling of raw moments over a range of time scales, we should define the following non-dimensional quantities commonly used in the literature (e.g. de Lima and Grasman, 1999; Serinaldi, 2010). The scale ratio λ so that $\lambda=1$ for the largest scale of interest $\Delta_{\max}$, i.e. $\lambda=\Delta_{\max}/\Delta$. In our case, we assume that $\Delta_{\max}=\lfloor n/8 \rfloor=128$ where the sample size $n=1024$, so that sample moments can be estimated from at least 8 data values, while the generic aggregated scale Δ is bounded in [1, 128]. Similarly, we form the non-dimensional process $\underline{\varepsilon}(\lambda)$ dividing the local average of the continuous-time process $\underline{x}(t)$ by its mean at the largest scale $\Delta_{\max}$ (or equivalently $\lambda=1$); then:

$$\underline{\varepsilon}(\lambda)=\frac{\underline{x}_j^{\left(\frac{\Delta_{\max}}{\lambda}\right)}}{\mathrm{E}\left[\underline{x}_j^{(\Delta_{\max})}\right]}\approx\frac{\underline{x}_j^{\left(\frac{\Delta_{\max}}{\lambda}\right)}}{m};\quad \lambda=\frac{\Delta_{\max}}{\Delta} \tag{2.12}$$

where m is the temporal mean of the data series. The scaling behaviour of the process is characterized by the moment scaling function $K(q)$ as follows:

$$\mathrm{E}\left[\left(\underline{\varepsilon}(\lambda)\right)^q\right]\approx\lambda^{K(q)} \tag{2.13}$$

If $K(q)$ linearly increases with q, then the process is said to be "simple scaling", otherwise it exhibits a "multiple scaling" behaviour.

In Fig. 2.8, we graphically show how uncertainty in sample moments is reflected in the uncertainty in the estimates of scaling exponents. It can be noticed that the function $K(q)$ shows a nonlinear behaviour for the lognormal series, thus suggesting a multifractal behaviour. Analogous considerations apply to the series generated by the other Monte Carlo experiments described in Sect. 2.1.3 above (not reported here).

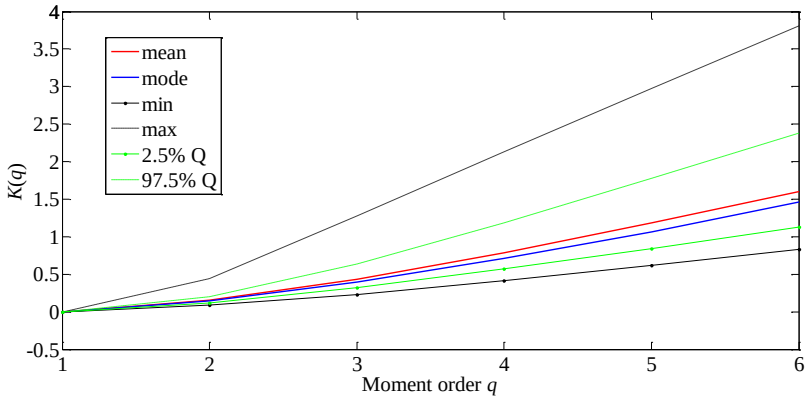


Figure 2.8 – Prediction intervals of the moment scaling function $K(q)$ versus the order q for lognormal series generated by our downscaling model (Sect. 3.2).

The prediction intervals in Fig. 2.8 spread out widely while increasing the moment order q, which is consistent with an enhancement of uncertainty. We clarify that we used the ratios of moment estimates in all calculations to compute $\underline{\varepsilon}(\lambda)$. Nonetheless, recalling that we assumed unit ensemble mean $\mu=\mathrm{E}[\underline{x}(t)]=1$ in all our Monte Carlo experiments, we found (not shown here) the same numerical results if using raw moments without taking any ratios. This is to stress that ratios of moments do not seem to play any significant role in the estimation of multiscaling exponents in our case.

It may be useful to add here some theoretical aspects. The theory of multifractals depends on the fact that raw moments obey power laws as the scale $\Delta \to 0$ (or equivalently $\lambda \to \infty$) (Falconer, 1990; Gneiting and Schlather, 2004), and so it depends on taking limits which cannot be achieved in reality. For most experimental purposes, the multifractal behaviour of a process $\underline{x}(t)$ is usually found by estimating the gradient of a graph of $\log(\mathrm{E}[(\underline{\varepsilon}(\lambda))^q])$ against $\log\lambda$ over an "appropriate" range of scales, where empirical points are closely matched by a straight line of slope $K(q)$. Being the latter an asymptotic slope, it is difficult to find the

"appropriate" range of scales to estimate $K(q)$, because we could be misled by some artificial slopes which do not indicate the multifractal behaviour of the underlying process (see e.g. Koutsoyiannis, 2013b). In addition, we should emphasize that the empirical moment scaling function $K(q)$ varies across scales for ergodic processes. The simple proof for this is given below in the special case of q=2.

According to Eqs. (2.12) and (2.13) we could write:

$$\mathrm{E}\left[\left(\underline{x}_j^{\left(\frac{\Delta_{\max}}{\lambda}\right)}\right)^2\right] \approx \lambda^{K(2)}\left(\mathrm{E}\left[\underline{x}_j^{(\Delta_{\max})}\right]\right)^2 = \lambda^{K(2)}\mu^2 \tag{2.14}$$

where μ is the mean of the process. On the other hand, we know that:

$$\mathrm{E}\left[\left(\underline{x}_j^{(\Delta)}\right)^2\right] = \gamma(\Delta) + \mu^2 \tag{2.15}$$

where $\gamma(\Delta)$ is the variance of the local average process at the scale Δ, see Eq. (1.5) in Sect. 1.3. If we assume that the process is ergodic, then we must have $\gamma(\Delta)\to 0$ as $\Delta\to\infty$ (Papoulis, 1991, p. 430).

Recalling that $\Delta=\Delta_{\max}/\lambda$, from Eqs. (2.14) and (2.15) we have:

$$\lambda^{K(2)}\mu^2 = \gamma\left(\frac{\Delta_{\max}}{\lambda}\right) + \mu^2 \tag{2.16}$$

dividing both sides by μ^2 and taking the logarithms, we obtain:

$$K(2) = \frac{\log\left(\gamma\left(\frac{\Delta_{\max}}{\lambda}\right)\Big/\mu^2 + 1\right)}{\log\lambda} \tag{2.17}$$

Clearly then, as $\lambda\to 0$ (i.e., as the scale grows to infinity $\Delta\to\infty$), the numerator $\to$ 0 and the denominator $\to \infty$. So, $K(2)$=0 asymptotically. Note that we have not made any assumption about the dependence structure or the marginal probability of the process, the only assumption is that the process is ergodic. In summary, for scales tending to infinity the $K(2)$ should tend to zero, while for scales tending to zero the $K(2)$ will take nonzero values.

Let us now investigate if empirical results shown in Fig. 2.8 agree well with the theoretical formulation of the moment scaling function $K(q)$ of our model described in the next Chapter. To this aim, in the following we derive the theoretical moment scaling function for our downscaling model based on the Hurst-Kolmogorov process.

Since the variables generated by our model are log-normally distributed, it can be shown that the generic q-order raw moment is given by (Kottegoda and Rosso, 2008, p. 216):

$$\mathrm{E}\left[\left(\underline{x}_j^{(\Delta)}\right)^q\right] = \exp\left(q\mu_{\ln\left(\underline{x}_j^{(\Delta)}\right)} + \frac{1}{2}q^2\gamma_{\ln\left(\underline{x}_j^{(\Delta)}\right)}\right) \tag{2.18}$$

where the two parameters can be determined in terms of the mean $\mu=\mathrm{E}[\underline{x}_j^{(\Delta)}]$ and the variance $\gamma(\Delta)=\mathrm{Var}[(\underline{x}_j^{(\Delta)})]$ of the local average process as follows:

$$\mu_{\ln\left(\underline{x}_j^{(\Delta)}\right)} = \log\mu - \frac{1}{2}\log\left(\frac{\gamma(\Delta)}{\mu^2}+1\right) \tag{2.19}$$

$$\gamma_{\ln\left(\underline{x}_j^{(\Delta)}\right)} = \log\left(\frac{\gamma(\Delta)}{\mu^2}+1\right) \tag{2.20}$$

As our downscaling model is based upon the Hurst-Kolmogorov process, the function $\gamma(\Delta)$ obeys the following power law (see also Eq. (1.8)):

$$\gamma(\Delta) = \gamma\Delta^{2H-2} \tag{2.21}$$

where $\gamma \equiv \gamma(\Delta=1)$ is the variance of the reference local average process $\underline{x}_j^{(\Delta=1)}$.

In order to derive the theoretical moment scaling function $K_{\mathrm{Th}}(q)$, we should investigate the following limiting behaviour (Falconer, 1990, p. 257):

$$K_{\mathrm{Th}}(q) = \lim_{\Delta\to 0}\frac{\log\left(\mathrm{E}\left[\left(\underline{x}_j^{(\Delta)}\right)^q\right]\right)}{-\log\Delta} \tag{2.22}$$

where, according to Eq. (2.18), the numerator of the right-hand side can be written as:

$$\log\left(\mathrm{E}\left[\left(\underline{x}_j^{(\Delta)}\right)^q\right]\right) = q\mu_{\ln\left(\underline{x}_j^{(\Delta)}\right)} + \frac{1}{2}q^2\gamma_{\ln\left(\underline{x}_j^{(\Delta)}\right)} \tag{2.23}$$

Substituting Eqs. (2.19) and (2.20) in the right-hand side of eq. (2.23), we obtain:

$$\log\left(\mathrm{E}\left[\left(\underline{x}_j^{(\Delta)}\right)^q\right]\right) = q\log\mu + \frac{q}{2}(q-1)\log\left(\frac{\gamma(\Delta)}{\mu^2}+1\right) \tag{2.24}$$

From Eq. (2.21) and using the properties of the logarithm, the Eq. (2.24) becomes:

$$\log\left(\mathrm{E}\left[\left(\underline{x}_j^{(\Delta)}\right)^q\right]\right)=\log\left(\mu^q\left(\frac{\gamma}{\mu^2}\Delta^{2H-2}+1\right)^{\frac{q}{2}(q-1)}\right) \tag{2.25}$$

Recalling that the Hurst coefficient is a parameter satisfying $0<H<1$, the exponent $2H-2<0$. Substituting Eq. (2.25) in Eq. (2.22), we easily obtain that the theoretical moment scaling function $K_{\mathrm{Th}}(q)$ for our downscaling model described in Sect. 3.2 is given by:

$$K_{\mathrm{Th}}(q)=q(q-1)(1-H) \tag{2.26}$$

Based on these findings, the empirical results in Fig. 2.8 do not seem to agree well with their theoretical counterparts. For example, in our case H=0.85, for q=4 the theoretical value should be $K_{\mathrm{Th}}(q)$=1.8, while the estimated mean value is about $K(q)$=0.5 in the scale range of our Monte Carlo experiments. Hence, not finding the "appropriate" range of scales, in addition to estimation problems reported in our work, may lead to remarkable underestimation of the moment scaling function.

2.1.5 Overview of key ideas

During recent decades, there has been a large raise of interest in multifractal analyses especially in the study of hydrological processes, particularly in rainfall modelling. Indeed, the multifractal framework provides parsimonious models to study the variability of several natural processes in geosciences, such as rainfall. Models following this approach require the scaling of the sample moments of different orders q, which is used in model identification and fitting. A common problem with the application of multifractal models, which in some cases may have led to incorrect results, is their disconnection from stochastic methodology and reasoning, and the (unstated) naïve consideration that statistical estimates represent the true properties of a process.

Using theoretical reasoning and Monte Carlo simulations we find that the reliability of multifractal methods which use high order moments (> 3) is questionable. In particular, we highlight the problems in inference from time series of geophysical processes. The classical statistical approaches, often used in geophysical modelling, are based upon several simplifying assumptions, tacit or explicit, such as independence in time and exponential distribution tails, which are invalidated in natural processes. Indeed, the study of natural processes reveals scaling behaviours in state (departure from exponential distribution tails) and in time (departure from

independence). While the multifractal models are based on these scaling behaviours per se, they failed to explore their statistical consequences with respect to the implied dramatic increase of uncertainty.

The following list briefly summarizes the main findings of the analyses described in previous sections.

- As natural processes are characterized by dependence in time, while classical statistics typically assumes independence, much larger samples are required in order to obtain estimates of similar reliability with classical statistics.
- Estimators of high moments whose distribution ranges over several orders of magnitude cannot support inference about a natural behaviour nor fitting of models.
- The most probable value of sample high moments (the mode) can strongly differ (by orders of magnitude) from its expected value (i.e. the true value), thus making the statistical estimate problematic even in the case of unbiasedness.
- The calculation of numerical values of high order moments is misleading as the theoretical moments may tend to infinity for high orders, while the sample estimates are always finite. Even smaller order moments can be very uncertain.
- Even if the generated process is multifractal, the sample estimates of the q-moments from a unique sample can provide misleading results.

Hence, we have shown that distribution tails heavier than the exponential one and temporal dependence result in enormously increased uncertainty and/or infinite biases from a practical perspective in raw moments. In essence, this is a warning against the blind use in geophysical time series analyses of classical statistical tools, which neglect dependence and heavy tails in distributions. Ossiander and Waymire (2000) already caution against using high moments in multifractal estimation, but their particular focus is on discrete multiplicative cascade models. Indeed, they demonstrate that the estimators of multiscaling exponents converge almost surely to the structure function of the cascade generators as the sample becomes large for all moment orders within a certain critical interval, whose upper bound is consistent with our results.

Ignorance of increased uncertainty and inattentive use of high order moments may result in inappropriate modelling, wrong inferences and false claims about the properties of the processes. Evidently, the first two moments are necessary to use in all problems as they define the most

important characteristics of the distribution, marginal (the first two moments) and joint (the second moment). Even for these two lowest moments it is important to study always their uncertainties and this only can be done in connection with a model fitted for the process of interest (as it is not possible to define uncertainty without specifying a model for the marginal distribution and dependence). The third moment is often useful as a measure of skewness but we should always be aware of its uncertainty; however use of third moment is not the only way to identify and assess the skewness of a distribution. For example in parameter estimation of three-parameter distributions, it is better to avoid the method of moments and use other fitting methods such as maximum likelihood, L-moments, etc. Moments of order > 3 should be avoided in model identification and fitting because their estimation is problematic. If we have to use them, then it is imperative to specify their uncertainty and involve this uncertainty in any type of modelling and inference.

2.2 Sampling properties of climacogram and power spectrum

The reason for fitting a statistical model to data is to make conclusions about some essential characteristics of the natural process which the data refer to. Such conclusions can be sensitive to the degree to which the datasets reflect the salient features of the process. Natural processes evolve in continuous time but their observation is inevitably made at discrete time. The observational time series formed are either series of instantaneous values of the natural phenomenon at a certain time step or aggregated quantities during this time step. In addition, the observation period is apparently a finite time period. Both time discretization and finite length may strongly affect the stochastic properties inferred from the data. In particular, time discretization distorts the stochastic properties at small time scales, while the finite length affects the properties at large time scales. Modelling of natural processes is typical made assuming discrete time and parameter estimation is usually done using classical statistical estimators which assume that observations are random samples. All these are inadequate practices and result in inappropriate and biased models. A different modelling strategy is proposed, in which the stochastic model is by definition a continuous-time process and the distortion due to discretization and finite-period observation is explicitly taken into account in model calibration. An additional benefit of the

proposed strategy is that it avoids the too artificial, often non-parsimonious, families of discrete time stochastic models (like the ARIMA(p,d,q) models).

Here we put the emphasis on autocorrelations and spectra, because they are the most extensively used concepts in the applications of stochastic processes (Papoulis, 1991). These concepts involve only second-order moments. Specifically, we focus on the power spectrum as well as the climacogram; the two are fully dependent on each other.

2.2.1 Some theoretical considerations

It may be useful to include here some theoretical aspects (see also Koutsoyiannis, 2013b).

The power spectrum of the continuous-time (instantaneous) process, denoted as $s(w)$, is twice the cosine Fourier transform of the autocovariance function of the process, while that of the discrete-time process, denoted as $s_d^{(\Delta)}(\omega)$, is twice the inverse finite cosine Fourier transform of the respective autocovariance function. The convention of the multiplying factor 2 in the Fourier transforms was adopted so that the integral of the spectrum on positive frequencies only equals the variance of the process:

$$s(w) = 2\int_{-\infty}^{\infty} c(\tau)\cos(2\pi w\tau)\,\mathrm{d}\tau \tag{2.27}$$

$$s_d^{(\Delta)}(\omega) = 2\sum_{z=-\infty}^{\infty} c_z^{(\Delta)}\cos(2\pi\omega z) \tag{2.28}$$

where $c(\tau)=\mathrm{Cov}[\underline{x}(t), \underline{x}(t+\tau)]$ and $c_z^{(\Delta)}=\mathrm{Cov}[\underline{x}_j^{(\Delta)}, \underline{x}_{j+z}^{(\Delta)}]$ (with $c_0^{(\Delta)}=\gamma(\Delta)$) denote the autocovariance functions for the continuous-time and averaged processes, respectively; the continuous time is denoted as t and the discrete time as $j = 0, 1, \ldots$; Δ is the time scale in case of averaging, the frequency (inverse time) is denoted as w while $\omega=w\Delta$ is nondimensionalized frequency. Both w and ω are real numbers ranging in $(-\infty, \infty)$ for a continuous-time process, while for a discrete-time process w ranges in $[-1/2\Delta, 1/2\Delta]$ and ω in $[-1/2, 1/2]$. As both the autocovariance function and the power spectrum are even functions, i.e. $f(x) = f(-x)$, we make all calculations for a continuous-time process in $(0, \infty)$, and for a discrete-time process in $[0, 1/2\Delta]$ for w and in $[0, 1/2]$ for ω. In the discrete-time case, both the lag z and the frequency ω are dimensionless. To make the spectrum equivalent and comparable to the continuous-time

spectrum $s(w)$, we introduce the version $s^{(\Delta)}(w)=\Delta s_d^{(\Delta)}(w\Delta)$ (with $w=\omega/\Delta$) for the averaged process, which has the same dimensions as $s(w)$.

Pursuant to the considerations above, the power spectrum of the continuous-time process is calculated from autocovariance function as:

$$s(w) = 4\int_0^\infty c(\tau)\cos(2\pi w\tau)\,\mathrm{d}\tau \tag{2.29}$$

The inverse transformation is:

$$c(\tau) = \int_0^\infty s(w)\cos(2\pi w\tau)\,\mathrm{d}w \tag{2.30}$$

The power spectrum of a discrete-time process is calculated from the autocovariance function as:

$$s_d^{(\Delta)}(\omega) = 2\gamma(\Delta) + 4\sum_{z=1}^{\infty} c_z^{(\Delta)}\cos(2\pi\omega z) \tag{2.31}$$

The inverse transformation is:

$$c_z^{(\Delta)} = \int_0^\infty s_d^{(\Delta)}(\omega)\cos(2\pi\omega\tau)\,\mathrm{d}\omega \tag{2.32}$$

Notice that, even in the discrete case, the inverse transformation is an integral, not a sum. $s^{(\Delta)}(w)$ is readily derived from its definition using Eq. (2.31).

As both the climacogram and the power spectrum are transformations of the autocovariance function, the two are also related to each other by simple transformations. Recall that the inverse formula, by which we can find the autocovariance if the climacogram is known, is derived by taking the second derivative in Eq. (1.5) using Leibniz's integral rule, which gives a formula for differentiation of a definite integral whose limits are functions of the differential variable as in Eq. (1.6). Thus, using known properties of the Fourier transform, we find:

$$s(w) = 2w^2\frac{\mathrm{d}^2}{\mathrm{d}w^2}\int_0^\infty \gamma(\tau)\cos(2\pi w\tau)\,\mathrm{d}\tau \tag{2.33}$$

which after algebraic manipulations becomes:

$$s(w) = -2\int_0^\infty (2\pi w\tau)^2\gamma(\tau)\cos(2\pi w\tau)\,\mathrm{d}\tau \tag{2.34}$$

On the other hand, combining Eqs. (1.5) and (2.30) we find:

$$\begin{aligned}\gamma(\Delta) &= 2\int_0^1 (1-\xi)\int_0^\infty s(w)\cos(2\pi w\xi\Delta)\,\mathrm{d}w\,\mathrm{d}\xi \\ &= 2\int_0^\infty s(w)\int_0^1 (1-\xi)\cos(2\pi w\xi\Delta)\,\mathrm{d}\xi\,\mathrm{d}w\end{aligned} \tag{2.35}$$

and after algebraic manipulations, we find the following equation giving directly the climacogram from the power spectrum:

$$\gamma(\Delta) = \int_0^\infty s(w) \frac{\sin^2(\pi w \Delta)}{(\pi w \Delta)^2} \mathrm{d}\, w \tag{2.36}$$

The climacogram $\gamma(\Delta)$ denotes variance and therefore should be nonnegative for any time scale Δ. It could be positive finite or even infinite for $\Delta = 0$. For (mean) ergodic processes it should necessary tend to 0 for $\Delta \to \infty$ (Papoulis, 1991, p. 429). Thus:

$$\gamma(\Delta) > 0; \gamma(\infty) = 0 \tag{2.37}$$

As the autocovariance $c(\tau)$ equals the variance for $\tau = 0$, it follows that $c(0) > 0$. For $\tau \neq 0$, $c(\tau)$ can take on negative values as well. However, $c(\tau)$ must be a positive definite function (see e.g. Stewart, 1976), a property which among other things makes it bounded from below and above by $\pm c(0)$. Ergodicity also imposes a constraint about its asymptotic behaviour (Papoulis, 1991, p. 430); in conclusion:

$$c(0) > 0; \quad |c(\tau)| \leq c(0); \quad \frac{1}{\Delta} \int_0^\Delta c(\tau) \mathrm{d}\tau \xrightarrow[\Delta \to \infty]{} 0 \tag{2.38}$$

In order for the function $c(\tau)$ to be positive definite, its Fourier transform, i.e. the power spectrum $s(w)$ should be nonnegative. Thus:

$$s(w) \geq 0 \tag{2.39}$$

Additional properties of $s(w)$ are discussed in next section.

The autocovariance $c(\tau)$ is often a nonnegative and non-increasing function. In this case $\gamma(\Delta)$ is non-increasing too. To see this, we take the derivative with respect to Δ from Eq. (1.5) and we find:

$$\gamma'(\Delta) = \frac{4}{\Delta^3} \int_0^\Delta (\tau - \Delta/2) c(\tau) \mathrm{d}\, \tau \tag{2.40}$$

The term $(\tau - \Delta/2)$ within the integral is symmetric with respect to $\Delta/2$ (negative for $\tau < \Delta/2$ and positive for $\tau > \Delta/2$). As $c(\tau)$ is nonincreasing, its values for $\tau < \Delta/2$ are greater than those for $\tau > \Delta/2$. Clearly then the negative product prevails over the positive product and thus $\gamma'(\Delta) < 0$.

2.2.2 Asymptotic properties of the power spectrum

The asymptotic slopes of the power spectrum $s(w)$ plotted in logarithmic axis vs. the logarithm of the frequency w are important properties to characterize a stochastic process. Generally, this slope is:

$$s^{\#}(w) = \frac{\mathrm{d}(\ln s(w))}{\mathrm{d}(\ln w)} = \frac{w s'(w)}{s(w)} = \frac{a(w)}{s(w)} \tag{2.41}$$

where $a(w) = ws'(w)$, and where $s'(w)$ is the derivative of $s(w)$.
We will find its asymptotic value for $w \to 0$, i.e. $s^{\#}(0)$. Note that continuous time is assumed for the process as well as the spectrum. From Eq. (2.29), the derivative is:

$$s'(w) = -4\int_0^{\infty} 2\pi\tau\, c(\tau)\sin(2\pi w\tau)\,\mathrm{d}\tau \tag{2.42}$$

We define:

$$A(w) = a(w) + s(w) = 4\int_0^{\infty} c(\tau)\left(\cos(2\pi w\tau) - 2\pi w\tau\sin(2\pi w\tau)\right)\mathrm{d}\tau \tag{2.43}$$

and we observe that:

$$A(0) = 4\int_0^{\infty} c(\tau)\,\mathrm{d}\tau = s(0) \tag{2.44}$$

From Eq. (2.43), we obtain:

$$a(0) + s(0) = s(0) \tag{2.45}$$

If $0 < s(0) < \infty$ then Eq. (2.45) simplifies to:

$$a(0) = 0 \tag{2.46}$$

and hence:

$$0 < s(0) < \infty \Rightarrow s^{\#}(0) = 0 \tag{2.47}$$

If $s(0) = 0$, then Eq. (2.46) is still valid, but the logarithmic slope $s^{\#}(0)$ given by Eq. (2.41) becomes an indeterminate quantity (0/0). This should necessarily be positive, so that $s^{\#}(w) > 0$ for $w > 0$. Thus:

$$s(0) = 0 \Rightarrow s^{\#}(0) > 0 \tag{2.48}$$

If $s(0) = \infty$ (hence $s'(0) = -\infty$, $a(0) < 0$, while $s(w)$ is continuous), then (because $s(w)$ is nonnegative) Eq. (2.45) results in:

$$\lim_{w\to 0}\left(a(w) + s(w)\right) > 0 \tag{2.49}$$

and:

$$\lim_{w\to 0}\left(\frac{a(w)}{s(w)} + 1\right) > 0 \tag{2.50}$$

Consequently, from Eq. (2.41) we obtain:

$$s(0) = \infty \Rightarrow s^{\#}(0) > -1 \tag{2.51}$$

In conclusion, the asymptotic slope in the logarithmic plot of the power spectrum for $w \to 0$ can never be lower (steeper) than –1 and more specifically it ranges as follows:

$$\begin{aligned} &s^{\#}(0) > 0; \qquad s(0) = 0 \\ &s^{\#}(0) = 0; \qquad 0 < s(0) < \infty \\ &-1 < s^{\#}(0) < 0; \quad s(0) = \infty \end{aligned} \tag{2.52}$$

The asymptotic slope for $w \to \infty$ should necessarily be non-positive, without other restrictions, i.e.:

$$s^{\#}(\infty) \le 0 \tag{2.53}$$

We often see publications reporting logarithmic slopes in empirical power spectra $s^{\#} < -1$ (e.g. $s^{\#} = -1.5$, etc.), but this should not be in contradiction with Eq. (2.51). First, we should point out that a slope $s^{\#}(w) < -1$ is mathematically and physically possible for large w. However, it is infeasible for $w \to 0$. Therefore, reported values $s^{\#} < -1$ for small w are spurious and are due to inconsistent estimation algorithms (cf. next section). Such results do not put into question the validity of Eq. (2.51) but are just invalid results. Let us prove this argument by assuming the opposite, i.e., that for frequency range $0 \le w \le \varepsilon$ (with ε however small) the logarithmic slope of the power spectrum is $s^{\#}(w) < -\beta$, or else $s(w) = \alpha w^{-\beta}$ where α and β are constants, with $\beta > 1$. We notice in Eq. (2.36) that the fraction within the integral takes significant values only for $w < 1/\Delta$ (cf. Papoulis, 1991, p. 433). Hence, assuming a scale $\Delta \gg 1/\varepsilon$, and with reference to Eq. (2.36) we may write:

$$\gamma(\Delta) = \int_0^{\infty} s(w) \frac{\sin^2(\pi w \Delta)}{(\pi w \Delta)^2} \mathrm{d}\, w \approx \int_0^{\varepsilon} \alpha w^{-\beta} \frac{\sin^2(\pi w \Delta)}{(\pi w \Delta)^2} \mathrm{d}\, w \tag{2.54}$$

On the other hand, it is easy to verify that, for $0 < w < 1/\Delta$, we have:

$$\frac{\sin(\pi w \Delta)}{\pi w \Delta} \ge 1 - w\Delta \ge 0 \tag{2.55}$$

and since $\varepsilon \gg 1/\Delta$, while the function in the integral of Eq. (2.54) is nonnegative, we can write:

$$\begin{aligned} \gamma(\Delta) &\approx \int_0^{\varepsilon} \alpha w^{-\beta} \frac{\sin^2(\pi w \Delta)}{(\pi w \Delta)^2} \mathrm{d}\, w \ge \int_0^{1/\Delta} \alpha w^{-\beta} \frac{\sin^2(\pi w \Delta)}{(\pi w \Delta)^2} \mathrm{d}\, w \\ &\ge \int_0^{1/\Delta} \alpha w^{-\beta} (1 - w\Delta)^2 \mathrm{d}\, w \end{aligned} \tag{2.56}$$

Substituing $\xi = w\Delta$ in Eq. (2.56), we find:

$$\gamma(\Delta) \ge \alpha \Delta^{\beta - 1} \int_0^1 \xi^{-\beta} (1 - \xi)^2 \mathrm{d}\, \xi \tag{2.57}$$

To evaluate the integral in Eq. (2.57) we take the limit for $r \to 0$ of the integral:

$$B(r)=\int_r^1 \xi^{-\beta}(1-\xi)^2\,\mathrm{d}\,\xi=\frac{r^{1-\beta}-1}{\beta-1}-2\frac{r^{2-\beta}-1}{\beta-2}+\frac{r^{3-\beta}-1}{\beta-3} \qquad (2.58)$$

Clearly, for $\beta > 1$ the first term of the right-hand side of Eq. (2.58) diverges for $r\to 0$, i.e., $B(0) = \infty$, and thus, by virtue of the inequality Eq. (2.57), $\gamma(\Delta)=\infty$. Therefore, the process is non-ergodic, see Eq. (2.37). It is interesting to note that, if $|\beta| < 1$, the integral in Eq. (2.54) can be evaluated to give:

$$\gamma(\Delta)=\alpha\int_0^\infty w^{-\beta}\frac{\sin^2(\pi w\Delta)}{(\pi w\Delta)^2}\,\mathrm{d}\,w=\frac{\sin(\pi\beta/2)}{(\pi\beta/2)}\,\frac{(2\pi)^\beta\alpha\Gamma(1-\beta)}{2(\beta+1)\Delta^{1-\beta}} \qquad (2.59)$$

Clearly, for $\Delta\to\infty$, the last expression gives $\gamma(\Delta)\to 0$ and thus for $|\beta| < 1$ the process is mean ergodic.

This analysis generalizes a result by Papoulis (1991, p. 434) who shows that an impulse of the power spectrum at $w = 0$ corresponds to a non-ergodic process.

In a non-ergodic process there is no possibility to infer statistical properties from the samples (as temporal averages do not represent true statistical properties). In any statistical analysis based on time series, ergodicity is necessary for the analysis to be valid. Otherwise the analysis is in vain and hence empirical results of this type are not meaningful because they contradict the basic condition on which they are based. Actually, such contradiction, when emerging from processing of data, does not suggest that a process is non-ergodic. Usually it only suggests that the algorithm used is inconsistent.

Sometimes reported slopes $s^{\#} < -1$ are interpreted as indications of non-stationarity. Such interpretations are equally invalid because even the definition of the power spectrum as a function of frequency only (as well as those of autocorrelation and climacogram as functions of lag and scale, respectively) assumes stationarity.

2.2.3 Power spectrum estimation

In this section, we focus on uncertainty in statistical estimation of power spectrum from correlated data series generated by a synthetic experiment. The next section is devoted to the climacogram estimation from the same data for comparison. Recall that the climacogram and the power spectrum are fully equivalent to each other, as well as to the autocorrelation function. We have shown in previous sections that each of these three

functions is theoretically derived by any of the other two. Therefore, a property in the spectral representation should have a one-to-one correspondence with a property in the climacogram representation. However their sampling properties may strongly different from each other.

While the power spectrum is a magnificent tool for stochastic processes, its estimation from data is problematic. To define uncertainty in statistical properties inferred from the data we need to specify a model for the underlying stochastic process. As the statistics of a standard normal process are completely determined just in terms of its climacogram, we restrict ourselves to a discussion of a stationary, standard Gaussian stochastic process defined by a Cauchy-type climacogram:

$$\gamma(\Delta) = \lambda\left(1 + (\Delta/\alpha)^{\kappa}\right)^{\frac{2H-2}{\kappa}} \tag{2.60}$$

where we have four parameters: units of α and λ are $[\Delta]$ and $[x]^2$, respectively, while H and κ are dimensionless. This model was derived by modifying one proposed by Gneiting and Schlather (2004), and its important feature is that it provides power-law correlations asymptotically. Hence, it allows explicit control of both asymptotic logarithmic slopes of the climacogram $\gamma^{\#}(\Delta)$ and the power spectrum $s^{\#}(w)$:

$$\begin{aligned} &s^{\#}(\infty) = -\kappa - 1; \quad \gamma^{\#}(0) = 0 \\ &s^{\#}(0) = 1 - 2H; \quad \gamma^{\#}(\infty) = 2H - 2 \end{aligned} \tag{2.61}$$

Note that, when taking Fourier transforms, asymptotic relationships at the origin turn into statements about the asymptotic behaviour at infinity, and vice versa.

Knowing asymptotic stochastic properties of processes is crucial for the quantification of future uncertainty, for planning and design purposes. Our primary concern is to study how these properties can be better estimated from data. To accomplish this aim, we perform a synthetic experiment by generating a time series of 1024 values from the known Cauchy-type process, assuming the following parameters: λ=1, α=10, H=0.8, κ=1. Hence, we have:

$$\begin{aligned} &s^{\#}(\infty) = -\kappa - 1 = -2; \quad \gamma^{\#}(0) = 0 \\ &s^{\#}(0) = 1 - 2H = -0.6; \quad \gamma^{\#}(\infty) = 2H - 2 = -0.4 \end{aligned} \tag{2.62}$$

Then, we compare the empirical power spectrum (see below) and the empirical climacogram (see next section) with their known theoretical counterparts.

We use the classical non-parametric approach (periodogram) because it explicitly estimates the power spectrum of the process without assuming that the process has any particular structure. In practice, we compute the periodogram from a finite-length digital sequence using the fast Fourier transform (FFT), that is why we chose $n = 2^{10} = 1024$. We consider the stochastic process defined by Eq. (2.60) with known theoretical properties, including its theoretical power spectrum, as shown on the graph (Fig. 2.9). The process is characterized by two different scaling laws, shown in its theoretical power spectrum as asymptotic slopes for frequencies $w \to 0$ and $w \to \infty$. In Eq. (2.62), we deduced these slopes theoretically, but, as we may notice in Fig. 2.9, we can hardly estimate them from data.

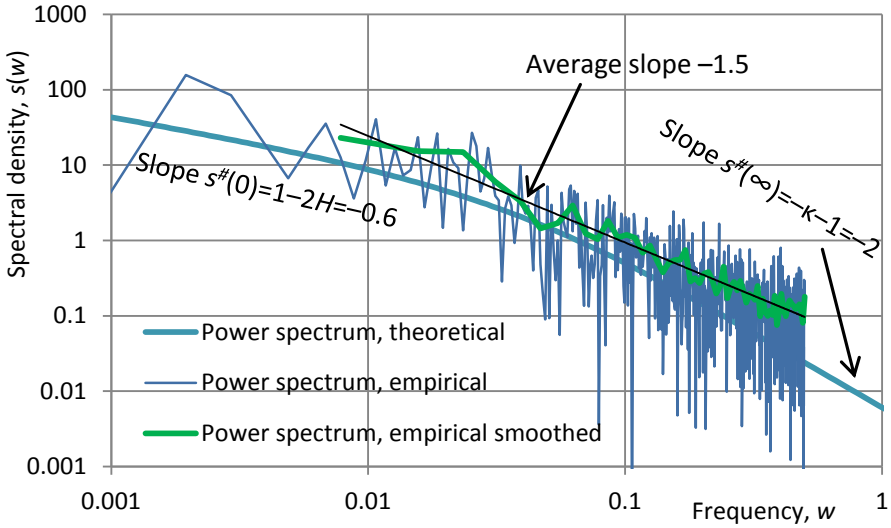


Figure 2.9 – Comparison between theoretical (true) and empirical (estimated) power spectra of a time series of 1024 values generated from the Cauchy-type process defined by Eq. (2.60).

The raw periodogram is an unbiased estimator of the power spectrum only asymptotically (i.e. shorter samples cause higher bias, even when windowing the data), and it has extremely poor variance characteristics which are not affected by the length of data used (Papoulis, 1991). The variance problem can be reduced by smoothing the periodogram. Here we show (Fig. 2.9) results for the Bartlett's method, which provides estimate of the spectrum at a given frequency by averaging the estimates from the periodograms (at the same frequency) derived from a non-overlapping

portions (segments) of the original series (here by averaging from 8 segments). We also tested the Welch's method (not shown here), which is an improvement of the Bartlett's method, and we obtained similar results. Furthermore, it can be shown that we can control the power-spectrum estimator variance by averaging more segments, but shorter segments mean larger bias; so for a fixed sample size, there is a basic trade-off between segment length, which controls the bias, and the number of segments, which controls the variance. Both bias and uncertainty in estimation may cause problems in estimating either asymptotic slopes or statistically significant peaks. In particular, the bias depends on frequency and this distorts the estimated slopes (e.g. too steep slopes, $s^{\#}(0)<-1$: unfeasible, as we demonstrated in the previous section). In addition, time discretization causes folding (i.e. symmetry of empirical power spectrum about the Nyquist frequency $w_N=1/2\Delta$); therefore the calculated slope $s^{\#}(w_N) = 0$, and it does not equal the actual asymptotic slope.

To conclude this analysis, we stress that the power spectrum, despite being very powerful in identifying strong periodicities in time series, it has some problems in identifying scaling laws and weak periodicities. Specifically, time discretization, finite length of data and data correlation alter asymptotic slopes of periodograms by introducing biases and uncertainties that are uncontrollable. Moreover, the rough shape of the periodogram may result both in false periodicities and in misleading, inaccurate or even incorrect slopes (e.g. slope > –1 for frequency → 0, which is infeasible).

2.2.4 Climacogram estimation

The procedure to estimate the climacogram from data has been described in Sect. 1.4; it is essentially concerned with the estimation of the process variance $\gamma(\Delta)$ at various scales of averaging Δ. The most common estimator of variance $\gamma(\Delta)$ of the averaged random process $\underline{x}_j^{(\Delta)}$ is the sample variance:

$$\underline{g}(\Delta)=\frac{1}{n-1}\sum_{j=1}^{n}\left(\underline{x}_j^{(\Delta)}-\underline{m}_1^{(\Delta)}\right)^2 \tag{2.63}$$

where n is the sample size and $\underline{m}_1^{(\Delta)}$ is given by Eq. (2.5). Here we discuss about an additional complication for correlated data. Namely, $\underline{g}(\Delta)$ is a biased estimator of the variance $\gamma(\Delta)$. The bias depends directly on the correlation structure (see Beran, 1994, p. 9). Only if the observations are

uncorrelated, then we obtain the well-known result that $\underline{g}(\Delta)$ is unbiased. If the data are positively correlated (as in our case, see the explanation in Sect. 1.2), then the sample variance tends to underestimate $\gamma(\Delta)$. Asymptotically the bias disappears, but the bias term converges to zero rather slowly, as n increases (Beran, 1994, p. 156). However, if we are able to estimate the correlation structure, then an unbiased estimator of $\gamma(\Delta)$ can be obtained by multiplying $\underline{g}(\Delta)$ with the corresponding estimated correction factor. Koutsoyiannis (2013b) proposed the following general equation to estimate the bias of $\underline{g}(\Delta)$:

$$\mathrm{E}\left[\underline{g}(\Delta)\right] = \frac{1}{1-1/n}\left(\mathrm{Var}\left[\underline{x}_j^{(\Delta)}\right] - \mathrm{Var}\left[\underline{m}_1^{(\Delta)}\right]\right) \tag{2.64}$$

Being $n = T/\Delta$ (see Fig. 1.4) and $\underline{m}_1^{(\Delta)} = \underline{x}_1^{(T)}$ (see Eq. (2.5)) then:

$$\mathrm{E}\left[\underline{g}(\Delta)\right] = \frac{1}{1-\Delta/T}\left(\gamma(\Delta) - \gamma(T)\right) = \eta(\Delta,T)\gamma(\Delta) \tag{2.65}$$

where the bias correction factor η is:

$$\eta(\Delta,T) = \frac{1-\gamma(T)/\gamma(\Delta)}{1-\Delta/T} \tag{2.66}$$

It becomes clear from the above equations that direct estimation of the variance $\gamma(\Delta)$ is not possible merely from the data. We need to know the ratio $\gamma(\Delta)/\gamma(T)$ and thus we should assume a stochastic model which evidently influences the estimation of $\gamma(\Delta)$. Once the model is assumed and its parameters estimated based on the data, we can expand our calculations to estimate the variance for any time scale Δ. Therefore, the important advantage of the climacogram over other common statistical tools is that its bias can be determined analytically (usually in a closed form) and included in the estimation problem.

In the case of the synthetic experiment described in previous section, we know from Eq. (2.60) the theoretical climacogram of the underlying stochastic process, so we can easily derive the bias correction factor $\eta(\Delta,T)$ from Eq. (2.66). In Fig. 2.10, we show the results.

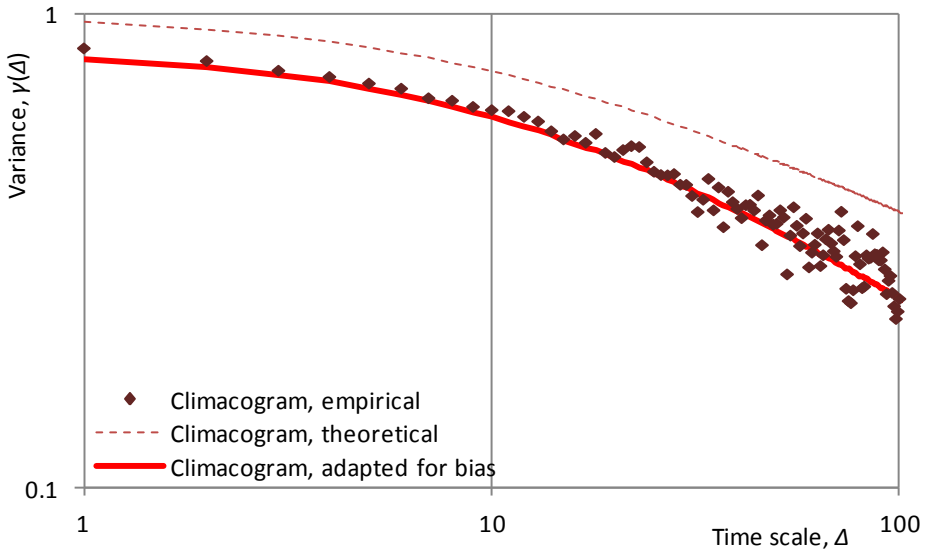


Figure 2.10 – Comparison between theoretical (true) and empirical (estimated) climacograms of a time series of 1024 values generated from the Cauchy-type process defined by Eq. (2.60).

The concept of climacogram can be used also in the frequency domain to find a substitute for the power spectrum, which has similar properties. In this context, Koutsoyiannis (2013b) proposed the climacogram-based pseudospectrum (CBPS), which we briefly describe and investigate in the following:

$$\psi(w) = \frac{2\gamma(1/w)}{w}\left(1 - \frac{\gamma(1/w)}{\gamma(0)}\right) \tag{2.67}$$

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

$$\psi(w) = \frac{2\gamma(1/w)}{w} \tag{2.68}$$

It can be shown that the value of CBPS at $w = 0$ equals that of the power spectrum, therefore $\psi(0) = s(0)$. Furthermore, the asymptotic logarithmic slopes $\psi^{\#}(w)$ of CBPS at frequencies $w\to 0$ and $w\to\infty$ follow those of the power spectrum $s^{\#}(w)$, and in most processes these slopes are precisely equal to each other. In our synthetic experiment, we have indeed:

$$\psi^{\#}(0) = s^{\#}(0) = -0.6; \quad \psi^{\#}(\infty) = s^{\#}(\infty) = -2 \tag{2.69}$$

In Fig. 2.11, we show that when the power spectrum and CBPS 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. Also, its calculation only uses the concept of variance and does not involve integral transformations (like the Fourier transform).

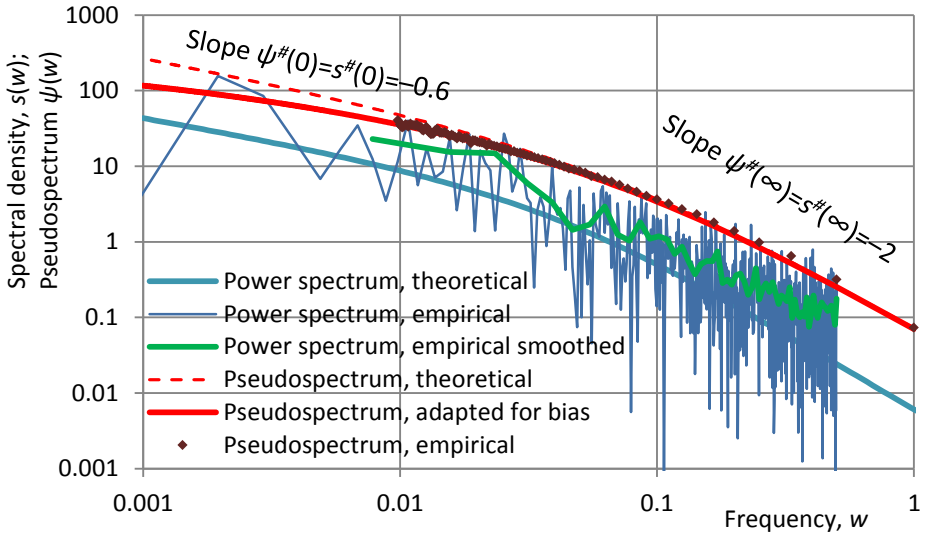


Figure 2.11 – Comparison between empirical and theoretical spectra and pseudospectra for the Cauchy-type process defined by Eq. (2.60).

2.2.5 Overview of key ideas

Geophysical processes typically evolve in continuous time but we observe and study them at discrete time. Besides, as we showed in Chapter 1, these processes commonly exhibit long-range dependence. Thus, in order to make reliable inferences about the stochastic properties of natural processes, we should always be aware of the effect of time discretization, finite record length as well as data correlation on classical statistical estimators. In particular, time discretization distorts the stochastic properties at small time scales, the finite length affects the properties at large time scales, while data correlation introduces (often uncontrollable) biases and uncertainties in statistical estimation. In Sect. 2.2, we mainly focus on second-order moments and specifically on climacograms and power spectra. Moreover, we analyse a possible substitute of the power spectrum, which is based on the concept of climacogram.

Based on a synthetic experiment for which all the stochastic properties of the underlying process are known, we compare the performances in statistical estimation of the statistical tools mentioned above.

The power spectrum, despite being very powerful in identifying strong periodicities in time series, it has some problems in identifying scaling laws and weak periodicities. Specifically, time discretization, finite length of data and data correlation alter asymptotic slopes of periodograms by introducing biases and uncertainties that are uncontrollable (see Sect. 2.2.3). Moreover, the rough shape of the periodogram may result both in false periodicities and in misleading, inaccurate or even incorrect slopes (e.g. slope > -1 for frequency $\rightarrow 0$, which is infeasible as shown in Sect. 2.2.2).

The important advantage of the climacogram over other common statistical tools is that its bias caused by the correlation structure of datasets can be determined analytically (usually in a closed form) and included in the estimation problem. However, direct estimation of climacogram is not possible merely from the data. We need to assume a stochastic model (see Sect. 2.2.4) which evidently influences the estimation of the process variance. Once the model is assumed and its parameters estimated based on the data, we can expand our calculations to estimate the variance for any time scale Δ.

The concept of climacogram can be used also in the frequency domain to find a substitute for the power spectrum, which has similar properties (e.g. the asymptotic behaviours of the two are similar). In Sect. 2.2.4, we analyze the climacogram-based pseudospectrum and show 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.

3. Rainfall downscaling

Tous pour un, un pour tous.
Alexandre Dumas

In stochastic hydrology, we often need to study natural processes at different time scales. The problems associated with the transfer of information across scales have been called scale issues (Blöschl and Sivapalan, 1995). To adequately address scale issues, we require models capable of preserving consistency across scales, i.e. both in a coarser, or higher-level, time scale and in a finer, or lower-level, time scale. These issues may arise, for instance, when coupling stochastic models of different time scales to reproduce simultaneously different important statistical properties of a hydrological process (Koutsoyiannis, 2001), e.g. the long-term and the short-term stochastic structure of precipitation (Langousis and Koutsoyiannis, 2006).

In other cases, scale issues are encountered in predictions using hydrological models, where the modelling scale may be much smaller than the observation scale; hence, we need to bridge that gap to calibrate, validate and operationally use our models. For example, when the higher-level process is the output of weather prediction models, which is given at a coarse scale, the scale discrepancy between model output and the resolution required for hydrological modelling must be resolved (e.g. Fowler et al., 2007, Groppelli et al., 2011). Furthermore, the higher-level process may be known from measurements. Specifically, when dealing with rainfall, long historical records usually come from daily rain gauges, but we need hourly or sub-hourly precipitation data in many hydrological applications. Also, the satellite rainfall data are available at a spatial scale greater than about 30 km at the Equator, and a temporal scale of 3 h, while again hydrological applications (e.g. related to flash floods) require higher resolutions (Berne et al., 2004; Koutsoyiannis and Langousis, 2011). In essence, scale issues can potentially be tackled by both disaggregation and downscaling techniques, which aim at modelling linkages across different temporal and/or spatial scales of a given process.

A natural process $\underline{x}(t)$, e.g. rainfall, is usually defined in continuous time t, but we observe or study it in discrete time as local averages $\underline{x}_j^{(\Delta)}$, which are the averages of $\underline{x}(t)$ over a fixed time scale Δ at discrete time steps j

(=1, 2, …) (see Eq. (1.3)). Furthermore, we may find useful to define here the cumulative process $\underline{X}(t)$ in continuous time:

$$\underline{X}(t) = \int_0^t \underline{x}(\zeta)\,\mathrm{d}\,\zeta \tag{3.1}$$

Its statistical properties depend on time, thus this is a non-stationary process, while $\underline{x}(t)$ is assumed to be stationary. Then, we can define the relevant process in discrete time, i.e. the aggregated process $\underline{X}_j^{(\Delta)}$ on a time scale Δ, as follows:

$$\underline{X}_j^{(\Delta)} = \underline{X}\left(j\Delta\right) - \underline{X}\left(\left(j-1\right)\Delta\right) \tag{3.2}$$

which are the stationary intervals of Eq. (3.1), the local average process given by Eq. (1.3) is recovered with $\underline{x}_j^{(\Delta)} = \underline{X}_j^{(\Delta)}/\Delta$.

Both disaggregation and downscaling refer to transferring information from a given scale (higher-level) to a smaller scale (lower-level), e.g. they generate consistent rainfall time series at a specific scale given a known precipitation measured or simulated at a certain coarser scale. The two approaches are very similar in nature but not identical to each other. Downscaling aims at producing the finer-scale rain field with the required statistics, being statistically consistent with the given field at the coarser scale. While disaggregation has the additional requirement to produce a finer scale rain field that adds up to the given coarse-scale total; thus, in this case we introduce an equality constraint to the problem in the form:

$$\underline{X}_j^{(f\Delta)} = \sum_{m=(j-1)f+1}^{jf} \underline{X}_m^{(\Delta)} = f\Delta\,\underline{x}_j^{(f\Delta)} \tag{3.3}$$

where $f\Delta$ is a time scale larger than Δ and f is a positive integer; for convenience Δ will be omitted. Then, for example $\underline{X}_1^{(f)} = \underline{X}_1 + \ldots + \underline{X}_f$ and $\underline{X}_2^{(f)} = \underline{X}_{f+1} + \ldots + \underline{X}_{2f}$.

The reader is referred to Koutsoyiannis and Langousis (2011) and the references therein for a detailed review on disaggregation and downscaling models in the literature.

This chapter focuses on the analysis of discrete random cascades for rainfall downscaling, which are characterized by a very simple structure, easy to implement and, consequently, widely applied in the literature. Hence, we compare the ensemble behaviour of two simple rainfall downscaling models based on two similar approaches: the multifractal and the Hurst-Kolmogorov. Both approaches are based on a general class of stochastic processes characterized by some invariant properties of their multivariate probability distribution under scale change, which illustrate the empirically observed scaling properties of rainfall time series.

The multifractal approach is based on the empirical detection of multifractal scale invariance of rainfall at finite, but practically important, ranges of scales (Veneziano et al., 2006). In particular, in Sect. 3.1 we focus on multiplicative random cascade (MRC) models to construct discrete multifractal fields, which are extensively used in the literature (e.g., Gupta and Waymire, 1993; Over and Gupta, 1996; Menabde and Sivapalan, 2000; Molnar and Burlando, 2005; Gaume et al., 2007; Rupp et al., 2009; Serinaldi, 2010; Licznar et al., 2011). The reason why MRC models have been so popular in the literature is that this method can parsimoniously generate complex intermittent and spiky patterns typical of rainfall time series, irrespective of whether the patterns are multifractal or not (Rupp et al., 2009).

The Hurst-Kolmogorov approach is based on the Hurst-Kolmogorov process described in Sect. 1.3. We propose a downscaling model following this approach (described in Sect. 3.2), which is a simple method to generate time series based on nonlinear transformation of stepwise linear relationship from a Gaussian random process.

3.1 Multiplicative random cascade models

Let $\underline{x}_1^{(f)}$ be the average rainfall intensity over time scale f (cf. Eq. (3.3)) at the time origin ($j = 1$); $\underline{x}_1^{(f)}$ is assumed to be a random variable with mean μ_0 and variance γ_0 of a stochastic process, which we wish to be stationary. $\underline{x}_1^{(f)}$ (for convenience $\underline{x}_{1,0}$) is then distributed over b sub-scale steps of equal size $s = f/b$ (i.e., $\underline{x}_j^{(s)}$, $j = 1, 2, ..., b$). This is accomplished through multiplying $\underline{x}_{1,0}$ by b different weights (one for each sub-scale step) $\underline{w}$ which are independent and identically distributed (iid) random variables. Moreover, their distribution is assumed to be the same for all cascade levels with mean μ_w and variance γ_w (Mandelbrot, 1974).

After repeating this procedure k times (k cascade levels; $k = 0, 1, 2, \ldots$), the resulting discrete random process at the scale of aggregation $s_k = b^{-k}f$ can be expressed as (see Fig. 3.1):

$$\underline{x}_j^{(s_k)} = \underline{x}_{j,k} = \underline{x}_{1,0} \prod_{i=0}^{k} \underline{w}_{g(i,j),i} \tag{3.4}$$

where $j = 1, 2, \ldots, b^k$ is the index of position in the series at level k; i is the index of the level of the cascade; $g(i, j)$ denotes a function which

defines the position in the series at the level i, i.e. $g(i,j)=\left\lceil \frac{j}{b^{k-i}} \right\rceil$, which is a ceiling function (Gaume et al., 2007). For $k = 0$ we have $\underline{w}_{1,0} = 1$.
For a *canonical* cascade (another common term to describe a downscaling model) the expected value of the mean process at the k-level is equal to the expected value of the process at the initial 0-level:

$$\left\langle \frac{1}{b^k}\sum_{j=1}^{b^k}\underline{x}_{j,k} \right\rangle = \left\langle \underline{x}_{1,0} \right\rangle \tag{3.5}$$

where for convenience we use < > to denote the expected value E[·] (i.e. average over the independent realizations of the stochastic process). From Eq. (3.4), we can derive the expected value of $\underline{x}_{j,k}$ as:

$$\begin{aligned} \left\langle \underline{x}_{j,k} \right\rangle &= \left\langle \underline{x}_{1,0}\prod_{i=0}^{k}\underline{w}_{g(i,j),i} \right\rangle = \left\langle \underline{x}_{1,0} \right\rangle \prod_{i=0}^{k}\left\langle \underline{w}_{g(i,j),i} \right\rangle \\ &= \left\langle \underline{x}_{1,0} \right\rangle \left\langle \underline{w} \right\rangle^{k} = \mu_0 \mu_w^k \end{aligned} \tag{3.6}$$

As a consequence of Eqs. (3.5) and (3.6):

$$\frac{1}{b^k}\sum_{j=1}^{b^k}\left\langle \underline{x}_{j,k} \right\rangle = \left\langle \underline{x}_{1,0} \right\rangle; \quad \left\langle \underline{x}_{j,k} \right\rangle = \mu_0; \quad \mu_0 \mu_w^k = \mu_0; \quad \mu_w = 1 \tag{3.7}$$

Thus, the weights $\underline{w}$ satisfy the condition $\mu_w = 1$.
For a *micro-canonical* cascade (i.e. a disaggregation model), the mean process at the k-level is equal to the process at the 0-level; this means that the following relationship (a consequence of Eq. (3.3)) holds for every pair of successive aggregation levels (k–1 and k) of the cascade:

$$\frac{1}{b}\sum_{m=b(j-1)+1}^{bj}\underline{x}_{m,k} = \underline{x}_{j,k-1} \tag{3.8}$$

where $j = 1, \ldots, b^{k-1}$ with $k > 0$. For example, if we choose $b = 2$, then:

$$\begin{aligned} &\frac{1}{2}\sum_{m=2j-1}^{2j}\underline{x}_{m,k} = \underline{x}_{j,k-1}; \\ &\underline{x}_{2j-1,k} + \underline{x}_{2j,k} = 2\underline{x}_{j,k-1}; \quad \underline{w}_{2j-1,k} = 2 - \underline{w}_{2j,k} \end{aligned} \tag{3.9}$$

Thus, the weights $\underline{w}_{j,k}$ satisfy $\mu_w = 1$ and $\underline{w} < b$ (e.g. in the case of Eq. (3.9), $\underline{w} < 2$). An important attribute of the micro-canonical model is that the distribution of $\underline{w}$ can be extracted from the data (Cârsteanu and Foufoula-Georgiou, 1996), allowing a direct examination of the associations that the weights may have with other properties of rainfall.

A graphical example of a dyadic (b = 2) multiplicative cascade for four cascade levels (k = 0, 1, 2, 3) is shown in Fig. 3.1.

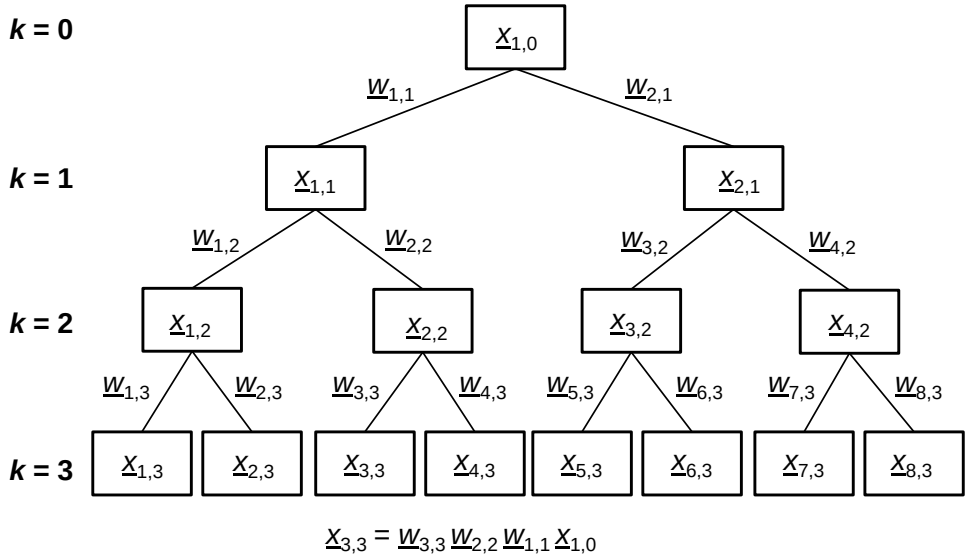


Figure 3.1 – Sketch of a dyadic (b = 2) multiplicative random cascade.

3.1.1 Downscaling model (canonical cascade)

The summary statistics of the random process $\underline{x}_{j,k}$ for a canonical cascade are derived below. Specifically, we derive the variance, $\gamma_{j,k}$, the q-th moment, $\left\langle \underline{x}_{j,k}^{q} \right\rangle$, and the autocorrelation function for discrete-time lag z, $\rho_{j,k}(z)$, of the random process at the k-level of the canonical cascade. The expected value, $\left\langle \underline{x}_{j,k} \right\rangle$, has been already given in Eq. (3.6). The variance can be expressed as follows:

$$\gamma_{j,k} = \left\langle \underline{x}_{j,k}^{2} \right\rangle - \left\langle \underline{x}_{j,k} \right\rangle^{2} = \left(\mu_0^2 + \gamma_0\right)\left(1+\gamma_w\right)^k - \mu_0^2 \tag{3.10}$$

where the second moment is given by:

$$\begin{aligned} \left\langle \underline{x}_{j,k}^{2} \right\rangle &= \left\langle \underline{x}_{1,0}^{2} \prod_{i=0}^{k} \underline{w}_{g(i,j),i}^{2} \right\rangle = \left\langle \underline{x}_{1,0}^{2} \right\rangle \prod_{i=0}^{k} \left\langle \underline{w}_{g(i,j),i}^{2} \right\rangle \\ &= \left\langle \underline{x}_{1,0}^{2} \right\rangle \left\langle \underline{w}^{2} \right\rangle^{k} = \left(\mu_0^2 + \gamma_0\right)\left(1+\gamma_w\right)^k \end{aligned} \tag{3.11}$$

Likewise, the q-th moment is:

$$\left\langle \underline{x}_{j,k}^{q} \right\rangle = \left\langle \underline{x}_{1,0}^{q} \right\rangle \left\langle \underline{w}^{q} \right\rangle^{k} \tag{3.12}$$

Finally, the correlation coefficient for lag z is given by:

$$\rho_{j,k}(z) = \frac{\left\langle \underline{x}_{j,k}\, \underline{x}_{j+z,k} \right\rangle - \left\langle \underline{x}_{j,k} \right\rangle^{2}}{\gamma_{j,k}} = \frac{\left(\mu_0^2 + \gamma_0\right)\left(1+\gamma_w\right)^{h_{j,k}(z)} - \mu_0^2}{\left(\mu_0^2 + \gamma_0\right)\left(1+\gamma_w\right)^{k} - \mu_0^2} \tag{3.13}$$

where the term $\left\langle \underline{x}_{j,k}\, \underline{x}_{j+z,k} \right\rangle$ can also be expressed as follows:

$$\left\langle \underline{x}_{j,k}\, \underline{x}_{j+z,k} \right\rangle = \left\langle \underline{x}_{1,0}^{2} \right\rangle \left\langle \underline{w}^{2} \right\rangle^{h_{j,k}(z)} \tag{3.14}$$

In Eqs. (3.13) and (3.14), exponent $h_{j,k}(z)$ (at the position $j = 1, \ldots, b^k - z$ in the cascade at level k) is bounded in $\left[0,\, k-1-\lfloor \log_2 z \rfloor\right]$ if $0 < z \leq b^k - 1$, where $\lfloor\ \rfloor$ denotes the floor function, while $h_{j,k}(z = 0) = k$, for any j and k. Assuming the cascade as a binary tree ($b = 2$), the exponent $h_{j,k}(z)$ denotes the number of vertices of the tree (excluding the start vertex $\underline{x}_{1,0}$) belonging to both simple paths leading to the vertices $\underline{x}_{j,k}$ and $\underline{x}_{j+z,k}$. The exponent $h_{j,k}(z)$ is computed as (see an explanatory sketch in Fig. 3.2):

$$h_{j,k}(z) = \begin{cases} \left(h_{j,k-1}(z)+1\right)\Theta\left[2^{k-1} - j - z\right], & j \leq 2^{k-1},\ z > 0 \\ h_{2^k - j - \tau + 1,k}(z), & j > 2^{k-1},\ z > 0 \\ h_{2^k - j + 1,k}(|z|), & z < 0 \end{cases} \tag{3.15}$$

where $\Theta[r]$ is the discrete form of the Heaviside step function, defined for a discrete variable (integer) r as:

$$\Theta[r] = \begin{cases} 0, & r < 0 \\ 1, & r \geq 0 \end{cases} \tag{3.16}$$

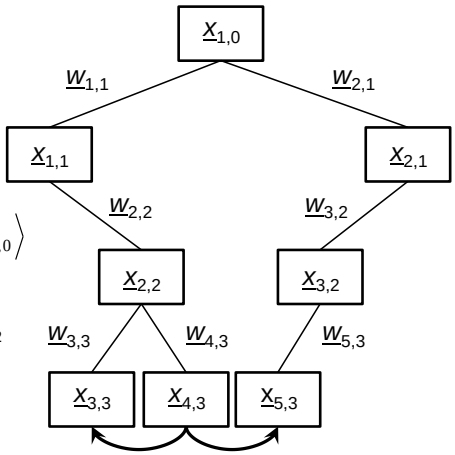
$$\left\langle \underline{x}_{j,k}\, \underline{x}_{j+z,k} \right\rangle = \left\langle \underline{x}_{1,0}^{2} \right\rangle \left\langle \underline{w}^{2} \right\rangle^{h_{j,k}(z)}$$

$\boldsymbol{h_{4,3}(z = 1) = 0}$

$$\left\langle \underline{x}_{4,3}\underline{x}_{5,3} \right\rangle = \left\langle \underline{w}_{4,3}\underline{w}_{2,2}\underline{w}_{1,1}\underline{w}_{5,3}\underline{w}_{3,2}\underline{w}_{2,1}\underline{x}_{1,0}^{2} \right\rangle = \left\langle \underline{x}_{1,0}^{2} \right\rangle$$

$\boldsymbol{h_{4,3}(z = -1) = 2}$

$$\left\langle \underline{x}_{4,3}\underline{x}_{3,3} \right\rangle = \left\langle \underline{w}_{4,3}\underline{w}_{2,2}^{2}\underline{w}_{1,1}^{2}\underline{w}_{3,3}\underline{x}_{1,0}^{2} \right\rangle = \left\langle \underline{x}_{1,0}^{2} \right\rangle \left\langle \underline{w}^{2} \right\rangle^{2}$$



$\underline{x}_{3,3} = \underline{w}_{3,3}\underline{w}_{2,2}\underline{w}_{1,1}\underline{x}_{1,0}$ $\underline{x}_{4,3} = \underline{w}_{4,3}\underline{w}_{2,2}\underline{w}_{1,1}\underline{x}_{1,0}$ $\underline{x}_{5,3} = \underline{w}_{5,3}\underline{w}_{3,2}\underline{w}_{2,1}\underline{x}_{1,0}$

Figure 3.2 – Example of computation of the exponent $h_{j,k}(z)$ for a canonical MRC. In the computation we use Eq. (3.14) and the arrows indicate the links to those variables considered.

Thus, three important considerations can be made. First, the exponent $h_{j,k}(z)$ is a function which satisfies a particular symmetry relation with respect to the position $j = 2^{k-1}$ in the dyadic cascade at level k. Second, the autocorrelation function of a canonical MRC corresponds to a non-stationary process, because it depends on the position j in the cascade (i.e. the time position) for any level k. Third, we started assuming a stationary setting of the entire process at the largest scale, then we concluded with a downscaled process that we demonstrated to be non-stationary. Consequently, it can be argued that autocorrelograms produced by canonical MRC have a physically unrealistic attitude with respect to the rainfall process.

Although the derivation of the theoretical autocorrelation function presented in Eqs. (3.13) and (3.15) is new, the problem of non-stationarity in processes generated by discrete random cascade models has been already discussed by Mandelbrot (1974, p. 356), who considered a canonical cascade with log-normal weights and a prescribed grid of eddies: "*Because the eddies were prescribed, the random function* [generated through the multiplicative scheme] *is non-stationary and discontinuous: it varies between an eddy and its neighbors, by jumps that may be very large*".

Moreover, this problem has been subsequently discussed by Over (1995), who highlighted the properties of non-stationarity (non-homogeneity) and anisotropy of the cross-moments of a discrete random cascade in a *d*-dimensional space, and by Veneziano and Langousis (2010, p. 137, Section 4.4.3.2). Hence, an important challenge is that of finding an alternative simple method to generate time series with spiky patterns typical of rainfall series and consistent with the observation at coarser scales, which is stationary. Indeed, as stated by Over (1995): "*In applications, we may find that we want a random process model that is anisotropic and non-homogeneous, but in a way that is controllable using model parameters, not simply inherent to the model, and we would most likely want to use a homogeneous and isotropic model as a null hypothesis unless physical considerations determined otherwise*" (p. 62, Section 3.4.1.1).
Thus, in Sect. 3.2 we propose a stationary downscaling model, based on the Hust-Kolmogorov process, which is characterized by a cascade structure similar to that of MRC models.

3.1.2 Example: numerical simulation

In this section, numerical simulations of a canonical MRC are carried out. For simplicity and without loss of generality, we assume μ_0=1 and γ_0=0. Thus, the summary statistics given in the previous section (Eqs. (3.6) and (3.10)-(3.13)) now become:

$$\left\langle \underline{x}_{j,k} \right\rangle = 1; \quad \left\langle \underline{x}_{j,k}^{q} \right\rangle = \left\langle \underline{w}^{q} \right\rangle^{k}$$
$$\gamma_{j,k} = \left(1+\gamma_w\right)^{k} - 1; \quad \rho_{j,k}(z) = \frac{\left(1+\gamma_w\right)^{h_{j,k}(z)} - 1}{\left(1+\gamma_w\right)^{k} - 1} \tag{3.17}$$

This example refers to weights $\underline{w}$ log-normally distributed, defined as follows (see e.g. Over and Gupta, 1996):

$$\underline{w} = b^{\sigma_N \underline{y} - \frac{\sigma_N^2 \ln b}{2}} \tag{3.18}$$

where $\underline{y}$ is a normal N(0,1) random variable; as a consequence, the variance of the weights is given by:

$$\sigma_w^2 = \exp\left(\sigma_N^2 (\ln b)^2\right) - 1 \tag{3.19}$$

whereas σ_N^2 is a parameter defining the normal N($-\sigma_N^2 \ln(b^2/2)$, $\sigma_N^2 \ln(b^2)$) random variable $\underline{v} = \ln \underline{w}$.

Monte Carlo simulations (M = 50000) have been applied to explore the ensemble behaviour of the random process, assuming e.g. k = 7 and σ_N=0.522, which gives $\gamma_{j,k}$ = 1.5 (from Eq. (3.17)). Figures 3.3 and 3.4 show respectively the ensemble mean $<\underline{x}_{j,k}>$ = $\mu_{j,k}$ and standard deviation $\sigma_{j,k}$ of the random processes as a function of the position j along the cascade level k, j = 1, 2, …, n (where $n = 2^k$ = 128). Figure 3.5 shows how the ensemble autocorrelation function $\rho_{j,k}(z)$ strongly depends on the position j in the cascade at the level k.

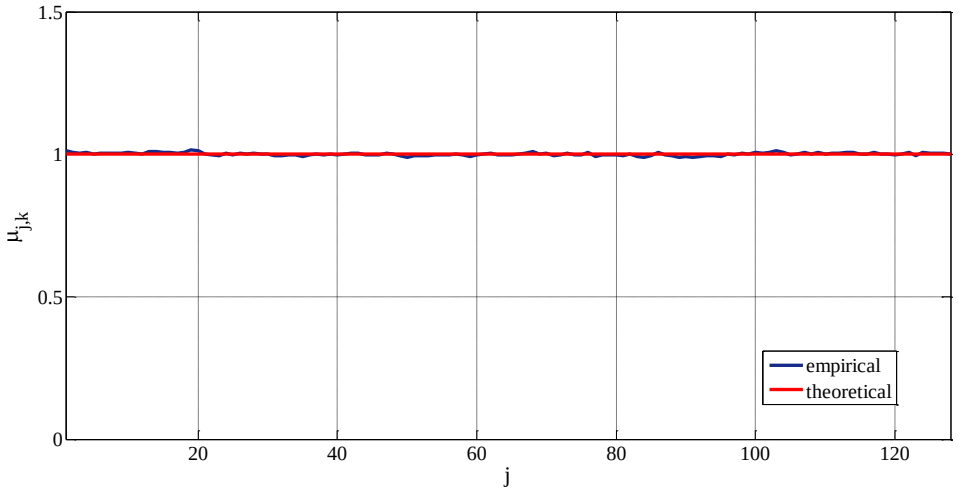


Figure 3.3 – Ensemble mean of the example MRC process as a function of the position j along the cascade level k = 7.

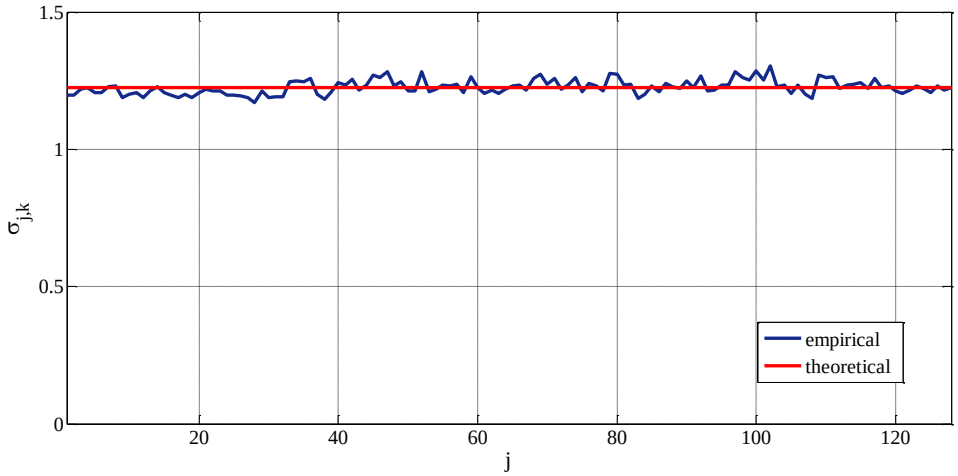


Figure 3.4 – Ensemble standard deviation of the example MRC process as a function of the position j along the cascade level k = 7.

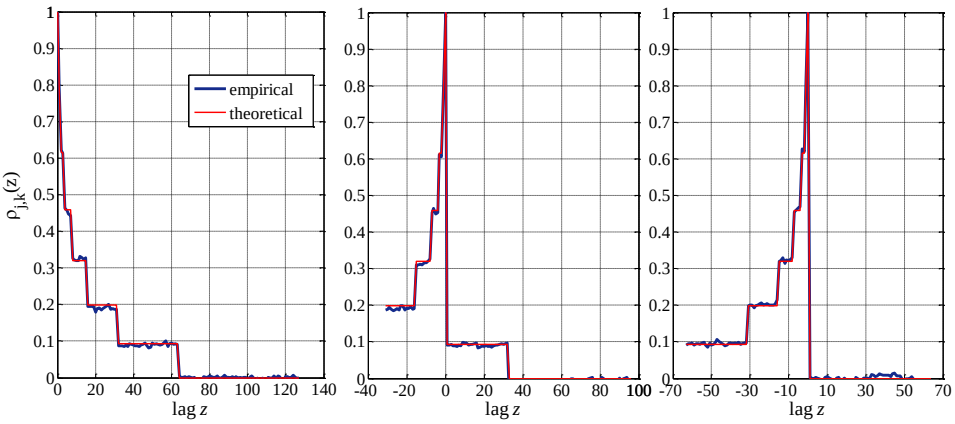


Figure 3.5 – Ensemble autocorrelation function of the example MRC process at the cascade level $k = 7$ with starting point j (for $j = 1$, $n/4$ and $n/2$, respectively, from left to right) in the considered cascade level with $n = 2^7 = 128$ elements.

In Fig. 3.6 (left), the autocorrelogram with starting point $j = n/2$ (midpoint of the cascade) is zoomed in the lag range [–5, 5] so as to illustrate that the lag 1 autocorrelation of the canonical MRC can be about 0.8 with the adjacent cell to the left and zero with the adjacent cell to the right. Moreover, if we move our simulation window just by two cells to the right, i.e. $j = n/2+2$ (see Fig. 3.6 right), then the lag 1 autocorrelation becomes about 0.8 and 0.6 with the adjacent cells to the left and to the right, respectively. These simple observations suffice to indicate how unrealistic and undesirable the stochastic structure of this model is.

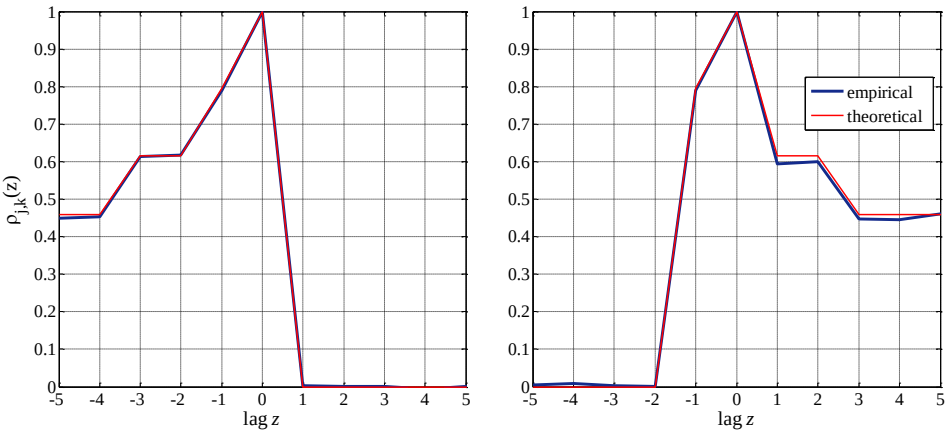


Figure 3.6 – ACF of the example MRC process at the cascade level $k = 7$ with starting point $j = n/2$ (left) and $j = n/2+2$ (right) zoomed in the lag range [–5, 5].

3.1.3 Disaggregation model (micro-canonical cascade)

In the case of a micro-canonical cascade, the summary statistics of the random process $\underline{x}_{j,k}$ can be expressed accounting for the equality constraint given in Eq. (3.8). The expected value $\langle \underline{x}_{j,k} \rangle$, the variance $\gamma_{j,k}$ and the q-moments $\langle \underline{x}_{j,k}^q \rangle$ remain the same as in the canonical case (Eqs. (3.6) and (3.10)-(3.12)), while the autocorrelation function at lag z, $\rho_{j,k}(z)$, now becomes, for $z \neq 0$:

$$\begin{aligned} \rho_{j,k}(\tau) &= \frac{\left\langle \underline{x}_{j,k}\, \underline{x}_{j+z,k} \right\rangle - \left\langle \underline{x}_{j,k} \right\rangle^2}{\gamma_{j,k}} \\ &= \frac{\left(\mu_0^2 + \gamma_0\right)\left(1+\gamma_w\right)^{h_{j,k}(z)}\left(1-\gamma_w\right) - \mu_0^2}{\left(\mu_0^2 + \gamma_0\right)\left(1+\gamma_w\right)^{k} - \mu_0^2} \end{aligned} \tag{3.20}$$

where the term $\left\langle \underline{x}_{j,k}\, \underline{x}_{j+z,k} \right\rangle$ can also be expressed as follows, if $b = 2$ (Eq. (3.9)):

$$\left\langle \underline{x}_{j,k}\, \underline{x}_{j+z,k} \right\rangle = \left\langle \underline{x}_{1,0}^2 \right\rangle \left\langle \underline{w}^2 \right\rangle^{h_{j,k}(z)} \left(2 - \left\langle \underline{w}^2 \right\rangle\right) \tag{3.21}$$

Note that, when $z = 0$, we have $h_{j,k}(z = 0) = k$, for any j and k, and the term $(1 - \gamma_w)$ in the numerator of Eq. (3.20) vanishes; consequently, we have $\rho_{j,k}(0) = 1$. As in Eqs. (3.13)-(3.14), the exponent $h_{j,k}(z)$ here also denotes the number of vertices of a binary tree (excluding the start vertex $\underline{x}_{1,0}$) belonging to both simple paths leading to the vertices $\underline{x}_{j,k}$ and $\underline{x}_{j+z,k}$. The exponent $h_{j,k}(z)$ can still be computed by Eq. (3.15). Thus, the autocorrelation function of a micro-canonical MRC again corresponds to a non-stationary process, as in the canonical case.

3.1.4 Bounded random cascades

A special form of multiplicative random cascades is the bounded random cascade (Marshak et al., 1994). Bounded cascades allow the multiplicative weights $\underline{w}$ to depend on the cascade level k and converge to unity as the cascade proceeds; this implies that the simulated random process becomes smoother on smaller scales. In the literature, bounded random cascades have been frequently applied to the stochastic fine graining of rainfall observations into high resolution data both in the canonical and microcanonical form (e.g. Menabde et al., 1997; Menabde and Sivapalan, 2000; Rupp et al., 2009; Licznar et al., 2011).

Bounded canonical cascades are constructed in the same way as the unbounded case, except that the weights $\underline{w}$ are iid only within a given cascade level, not among different levels as in the unbounded case (Menabde et al., 1997). Under these hypotheses and using the same notation as Eq. (3.14) above, the following holds:

$$\left\langle \underline{X}_{j,k}\, \underline{X}_{j+z,k} \right\rangle = \left\langle \underline{X}_{1,0}^2 \right\rangle \prod_{i=0}^{h_{j,k}(z)} \left\langle \underline{w}_{g(i,j),i}^2 \right\rangle \tag{3.22}$$

where, if $h_{j,k}(z) = 0$ (no tree vertices in common) we have $\underline{w}_{1,0} = 1$ (see Sect. 3.1). Hence, the autocorrelation function of the time series generated by bounded canonical cascades still depends on the position j in the cascade level k.

3.2 Hurst-Kolmogorov downscaling model

In this section, we analyse a simple downscaling method to generate rainfall time series based on Hurst-Kolmogorov process (defined in Sect. 1.3). The model disaggregates a fractional Gaussian noise by a dyadic additive cascade, which is then exponentially transformed to derive the actual rainfall time series that are consequently supposed to be log-normally distributed (e.g. Over, 1995).

Let $\underline{X}_1^{(f)}$ be the cumulative rainfall depth at the time origin (j=1) aggregated on the largest time scale f (see Eq. (3.3)) that is to be downscaled to a certain scale of interest. $\underline{X}_1^{(f)}$ is assumed to be a random variable with mean μ_0 and variance γ_0 of a stochastic process, which we wish to be stationary. We suppose the actual rainfall to be log-normally distributed.

Let us now introduce an auxiliary Gaussian random variable $\underline{\tilde{X}}_1^{(f)} = \ln \underline{X}_1^{(f)}$ (for convenience $\underline{\tilde{X}}_{1,0}$) of the aggregated HK process on the time scale f with mean $\tilde{\mu}_0$ and variance $\tilde{\gamma}_0$. It is well known that (see also Sect. 2.1.4):

$$\tilde{\mu}_0 = \ln \mu_0 - \frac{1}{2} \ln\left(\frac{\gamma_0}{\mu_0^2} + 1 \right) \tag{3.23}$$

$$\tilde{\gamma}_0 = \ln\left(\frac{\gamma_0}{\mu_0^2} + 1 \right) \tag{3.24}$$

$\tilde{Z}_{1,0}$ is to be disaggregated by a dyadic (b=2) additive cascade. Then, $\tilde{Z}_{1,0}$ is partitioned into two (b=2) Gaussian random variables on the time scale $s=f/2$; e.g. at the first cascade level (k=1) we have:

$$\underline{\tilde{X}}_{1,1} + \underline{\tilde{X}}_{2,1} = \underline{\tilde{X}}_{1,0} \tag{3.25}$$

Likewise, at the k-level corresponding to the scale of aggregation $s_k=2^{-k}f$, we have:

$$\underline{\tilde{X}}_{2j-1,k} + \underline{\tilde{X}}_{2j,k} = \underline{\tilde{X}}_{j,k-1} \tag{3.26}$$

Thus, it suffices to generate $\underline{\tilde{X}}_{2j-1,k}$ and then obtain $\underline{\tilde{X}}_{2j,k}$ from Eq. (3.26) above. This generic procedure resembles the well-known interpolation procedure, which is a point estimation. Thus, we can consider the following linear generation scheme (see graphical example in Fig. 3.7):

$$\underline{\tilde{X}}_{2j-1,k} = \boldsymbol{\theta}^{\mathrm{T}}\underline{\boldsymbol{Y}} + \underline{v} \tag{3.27}$$

where $\underline{\boldsymbol{Y}} = \left[\underline{\tilde{X}}_{2j-3,k}, \underline{\tilde{X}}_{2j-2,k}, \underline{\tilde{X}}_{j,k-1}, \underline{\tilde{X}}_{j+1,k-1}\right]^{\mathrm{T}}$, $\boldsymbol{\theta}$ is a vector of parameters, and $\underline{v}$ is a Gaussian white noise that represents an innovation term. Eq. (3.27) allows the generated lower-level variable $\underline{\tilde{X}}_{2j-1,k}$ to preserve autocorrelations with two earlier lower-level variables (level k) and one later higher-level variable (level k–1) (see also Koutsoyiannis, 2002).

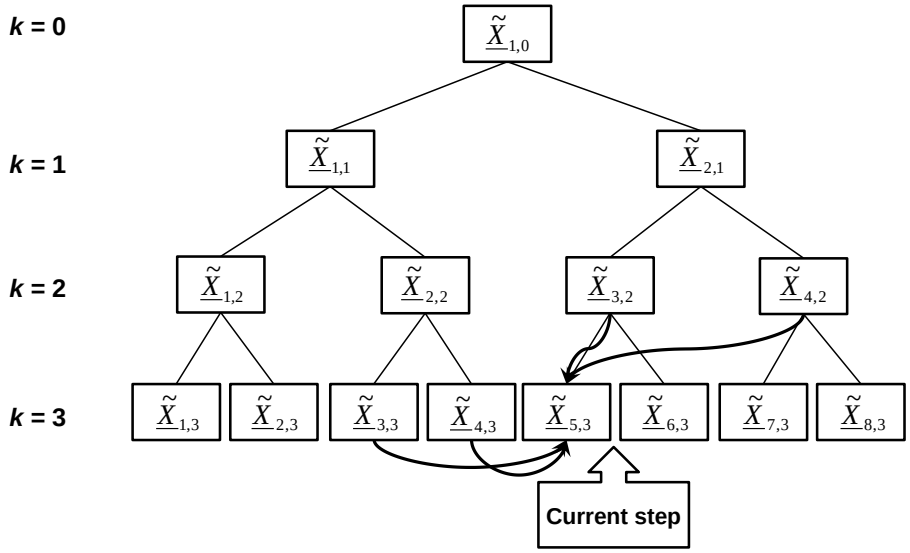


Figure 3.7 – Example of the dyadic additive cascade for four disaggregation levels (k = 0, 1, 2, 3), where arrows indicate the links to those variables considered in the current generation step (adapted from Koutsoyiannis, 2002).

Koutsoyiannis (2001) demonstrated that the vector $\boldsymbol{\theta}$ which minimizes Var[$\underline{v}$] is of the form:

$$\boldsymbol{\theta} = \left\{\mathrm{Cov}\left[\underline{\boldsymbol{Y}}, \underline{\boldsymbol{Y}}\right]\right\}^{-1} \mathrm{Cov}\left[\underline{\boldsymbol{Y}}, \underline{\tilde{X}}_{2j-1,k}\right] \tag{3.28}$$

Consequently, it can be shown that the least mean square prediction error of $\tilde{Z}_{2j-1,k}$ from $\underline{\boldsymbol{Y}}$ is the following:

$$\mathrm{Var}\left[\underline{v}\right] = \mathrm{Var}\left[\underline{\tilde{X}}_{2j-1,k}\right] - \mathrm{Cov}\left[\underline{\tilde{X}}_{2j-1,k}, \underline{\boldsymbol{Y}}\right]\boldsymbol{\theta} \tag{3.29}$$

Hence, in each disaggregation step the two lower-level variables are generated by (Eqs. (3.26)-(3.27)):

$$\begin{cases} \underline{\tilde{X}}_{2j-1,k} = a_2 \underline{\tilde{X}}_{2j-3,k} + a_1 \underline{\tilde{X}}_{2j-2,k} + b_0 \underline{\tilde{X}}_{j,k-1} + b_1 \underline{\tilde{X}}_{j+1,k-1} + \underline{v} \\ \underline{\tilde{X}}_{2j,k} = \underline{\tilde{X}}_{j,k-1} - \underline{\tilde{X}}_{2j-1,k} \end{cases} \tag{3.30}$$

Recalling that for a discrete-time Hurst-Kolmogorov we have:

$$\tilde{\rho}^{(f)}(z) = \tilde{\rho}(z) = \frac{|z+1|^{2H}}{2} + \frac{|z-1|^{2H}}{2} - |z|^{2H} \tag{3.31}$$

Parameters a_2, a_1, b_0 and b_1, and the variance of the innovation term $\underline{v}$ are estimated in terms of the correlation coefficients $\tilde{\rho}(z)$, which are independent on j and k, and of the variance of the HK process at the level k (Koutsoyiannis, 2002), as given by equations (3.32) and (3.33):

$$\begin{bmatrix} a_2 \\ a_1 \\ b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 1 & \tilde{\rho}(1) & \tilde{\rho}(2)+\tilde{\rho}(3) & \tilde{\rho}(4)+\tilde{\rho}(5) \\ \tilde{\rho}(1) & 1 & \tilde{\rho}(1)+\tilde{\rho}(2) & \tilde{\rho}(3)+\tilde{\rho}(4) \\ \tilde{\rho}(2)+\tilde{\rho}(3) & \tilde{\rho}(1)+\tilde{\rho}(2) & 2[1+\tilde{\rho}(1)] & \tilde{\rho}(1)+2\tilde{\rho}(2)+\tilde{\rho}(3) \\ \tilde{\rho}(4)+\tilde{\rho}(5) & \tilde{\rho}(3)+\tilde{\rho}(4) & \tilde{\rho}(1)+2\tilde{\rho}(2)+\tilde{\rho}(3) & 2[1+\tilde{\rho}(1)] \end{bmatrix}^{-1} \times \begin{bmatrix} \tilde{\rho}(2) \\ \tilde{\rho}(1) \\ 1+\tilde{\rho}(1) \\ \tilde{\rho}(2)+\tilde{\rho}(3) \end{bmatrix} \tag{3.32}$$

and:

$$\mathrm{Var}\left[\underline{v}\right] = \tilde{\gamma}_k \left(1 - \left[\tilde{\rho}(2), \tilde{\rho}(1), 1+\tilde{\rho}(1), \tilde{\rho}(2)+\tilde{\rho}(3)\right]\left[a_2, a_1, b_0, b_1\right]^{\mathrm{T}}\right) \tag{3.33}$$

Recalling the scaling properties of the Hurst-Kolmogorov process (see Sect. 1.3), the mean and the variance of the process at the k-level of the cascade are:

$$\tilde{\mu}_k = \left\langle \underline{\tilde{X}}_{j,k} \right\rangle = \frac{s_k}{f} \tilde{\mu}_0 = \frac{\tilde{\mu}_0}{2^k} \tag{3.34}$$

$$\tilde{\gamma}_k = \mathrm{Var}\left[\underline{\tilde{X}}_{j,k}\right] = \left(\frac{s_k}{f}\right)^{2H} \tilde{\gamma}_0 = \frac{\tilde{\gamma}_0}{2^{2Hk}} \tag{3.35}$$

where the timescale $s_k = 2^{-k} f$.

The above stepwise disaggregation approach was first introduced by Koutsoyiannis (2002), who demonstrated that it effectively generates fractional Gaussian noise, but the rainfall process (especially at the resolution needed for hydrological applications) is not Gaussian. Indeed, we apply the following specific exponentiation to the HK process to make it log-normal but preserve its scaling properties (Eqs. (3.34)-(3.35)):

$$\underline{X}_{j,k} = \exp\left(\alpha(k)\underline{\tilde{X}}_{j,k} + \beta(k)\right) \tag{3.36}$$

In other words, we assume a unique HK process in the untransformed domain, and we change the characteristics of the transformed (exponentiated) domain using different characteristics for different disaggregation steps by means of the scale-dependent functions $\alpha(k)$ and $\beta(k)$:

$$\alpha(k) = 2^{Hk}\sqrt{\frac{\ln\left(2^{2k(1-H)}\left(\exp(\tilde{\gamma}_0)-1\right)+1\right)}{\tilde{\gamma}_0}}$$
$$\beta(k) = -k\ln 2 - \tilde{\mu}_0\left(\frac{\alpha(k)}{2^k}-1\right) - \frac{\tilde{\gamma}_0}{2}\left(\frac{\alpha^2(k)}{2^{2Hk}}-1\right) \tag{3.37}$$

The proof is given in the following. These expressions of $\alpha(k)$ and $\beta(k)$ are derived to preserve the scaling properties of the process $\underline{X}_{j,k}$ at different scales of aggregation. The mean and the variance of the exponentiated process at the generic k-level of the cascade given in Eq. (3.36) are:

$$\mu_k = \exp\left(\beta(k) + \alpha(k)\tilde{\mu}_k + \alpha^2(k)\frac{\tilde{\gamma}_k}{2}\right) \tag{3.38}$$

$$\gamma_k = \exp\left(2\beta(k) + 2\alpha(k)\tilde{\mu}_k + \alpha^2(k)\tilde{\gamma}_k\right)\left(\exp\left(\alpha^2(k)\tilde{\gamma}_k\right)-1\right) \tag{3.39}$$

where $\tilde{\mu}_k$ and $\tilde{\gamma}_k$ are respectively the mean and the variance of the auxiliary process at the cascade level k, given by Eqs. (3.34) and (3.35). Substituting Eqs. (3.34) and (3.35) in (3.38) and (3.39), respectively, we obtain:

$$\mu_k = \exp\left(\beta(k) + \alpha(k)\frac{\tilde{\mu}_0}{2^k} + \alpha^2(k)\frac{\tilde{\gamma}_0}{2^{2Hk+1}}\right) \tag{3.40}$$

$$\gamma_k = \exp\left(2\beta(k) + 2\alpha(k)\frac{\tilde{\mu}_0}{2^k} + \alpha^2(k)\frac{\tilde{\gamma}_0}{2^{2Hk}}\right)\left(\exp\left(\alpha^2(k)\frac{\tilde{\gamma}_0}{2^{2Hk}}\right) - 1\right) \quad (3.41)$$

where $\tilde{\mu}_0$ and $\tilde{\gamma}_0$ are respectively the mean and the variance of the auxiliary normal variable $\underline{\tilde{X}}_1^{(f)} = \underline{\tilde{X}}_{1,0}$.

To derive the two functions $\alpha(k)$ and $\beta(k)$ we impose for the $\underline{X}_{j,k}$ process the same scaling laws of the relevant HK process ($\underline{\tilde{X}}_{j,k}$):

$$\mu_k = \left\langle \underline{X}_{j,k} \right\rangle = \frac{s_k}{f}\mu_0 = \frac{\mu_0}{2^k} \quad (3.42)$$

$$\gamma_k = \mathrm{Var}\left[\underline{X}_{j,k}\right] = \left(\frac{s_k}{f}\right)^{2H}\gamma_0 = \frac{\gamma_0}{2^{2Hk}} \quad (3.43)$$

where μ_0 and γ_0 are respectively the mean and the variance of the log-normal variables $\underline{X}_1^{(f)}$. Since we assume $\underline{X}_1^{(f)} = \exp\left(\underline{\tilde{X}}_1^{(f)}\right)$, we have $\alpha(0)=1$ and $\beta(0)=0$ and, thus, Eqs. (3.23) and (3.24) hold. Substituting Eqs. (3.23) and (3.24) in (3.42) and (3.43) respectively, we obtain:

$$\mu_k = \frac{1}{2^k}\exp\left(\tilde{\mu}_0 + \frac{\tilde{\gamma}_0}{2}\right) \quad (3.44)$$

$$\gamma_k = \frac{1}{2^{2Hk}}\exp\left(2\tilde{\mu}_0 + \tilde{\gamma}_0\right)\left(\exp\left(\tilde{\gamma}_0\right) - 1\right) \quad (3.45)$$

Equating the right-hand sides of Eqs. (3.40) and (3.41) to (3.44) and (3.45), respectively, we obtain:

$$\exp\left(\beta(k) + \alpha(k)\frac{\tilde{\mu}_0}{2^k} + \alpha^2(k)\frac{\tilde{\gamma}_0}{2^{2Hk+1}}\right) = \frac{1}{2^k}\exp\left(\tilde{\mu}_0 + \frac{\tilde{\gamma}_0}{2}\right) \quad (3.46)$$

$$\exp\left(2\beta(k) + 2\alpha(k)\frac{\tilde{\mu}_0}{2^k} + \alpha^2(k)\frac{\tilde{\gamma}_0}{2^{2Hk}}\right)\left(\exp\left(\alpha^2(k)\frac{\tilde{\gamma}_0}{2^{2Hk}}\right) - 1\right) = \frac{1}{2^{2Hk}}\exp\left(2\tilde{\mu}_0 + \tilde{\gamma}_0\right)\left(\exp\left(\tilde{\gamma}_0\right) - 1\right) \quad (3.47)$$

Solving Eqs. (3.46) and (3.47) we obtain Eq. (3.37). Then, the mean and variance of the log-normal variables $\underline{X}_{j,k}$ (actual downscaled rainfall) are given by Eqs. (3.42) and (3.43), respectively, while the autocorrelation function is given by:

$$\rho_k(z) = \frac{\exp\left(\tilde{\gamma}_k\tilde{\rho}(z)\right) - 1}{\exp\left(\tilde{\gamma}_k\right) - 1} \quad (3.48)$$

where $\tilde{\gamma}_k$ and $\tilde{\rho}(z)$ are given by Eqs. (3.35) and (3.31), respectively.

The log-normality hypothesis and our specific exponential transformation (Eqs. (3.36) and (3.37)) enable the analytical formulation of the main statistics of the actual rainfall process, given in Eqs. (3.42)-(3.43) and (3.48), which are a key element for our theoretical analysis. However, more elaborate normalizing transformations can be investigated (see, e.g. Papalexiou et al., 2011), but this is out of the scope of our analysis.

The presented model is a disaggregation model only if the random variables are Gaussian; indeed, the equality constraint in Eq. (3.26) holds. However, under the hypothesis of log-normal rainfall, we have a downscaling model, where the lower-level rainfall time series generated are only statistically consistent with the given process $\underline{X}_j^{(f)}$ at the coarser scale. The Hurst coefficient H is the only parameter of our downscaling model.

3.2.1 Example: numerical simulation

To investigate further the goodness of HK downscaling model, we explore its numerical simulations (M = 50000) as we did for the MRC downscaling model in Sect. 3.1.2. To make the two model simulations comparable, we assume the same values of summary statistics as in the MRC case, i.e. k=7, μ_k = 1 and γ_k = 1.5. Furthermore, we assume H = 0.7. Figs. 3.8 and 3.9 show, respectively, the behaviours of the ensemble mean $<\underline{x}_{j,k}>$ = $\mu_{j,k}$ and standard deviation $\sigma_{j,k}$ of the random processes as a function of the position j along the level k, j = 1, 2, …, n (where n=2^k=128).

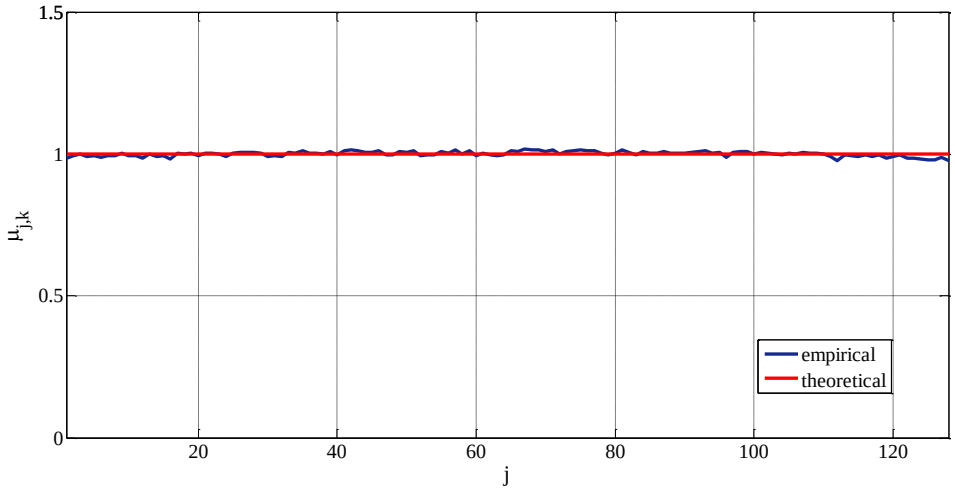


Figure 3.8 – Ensemble mean of the example HK process as a function of the position j along the cascade level k = 7.

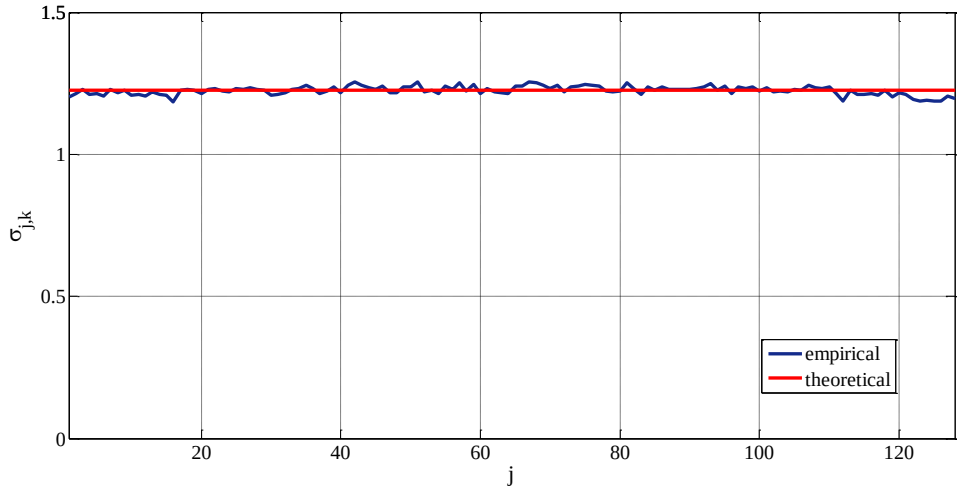


Figure 3.9 – Ensemble standard deviation of the example HK process as a function of the position j along the cascade level k = 7.

Figure 3.10 shows how, unlike the MRC case, the ensemble autocorrelation function $\rho_{j,k}(z)$ is fully independent of the position in time j in the cascade at the level k. Thus, we verified that the process corresponding to the time series generated by the HK downscaling model is stationary.

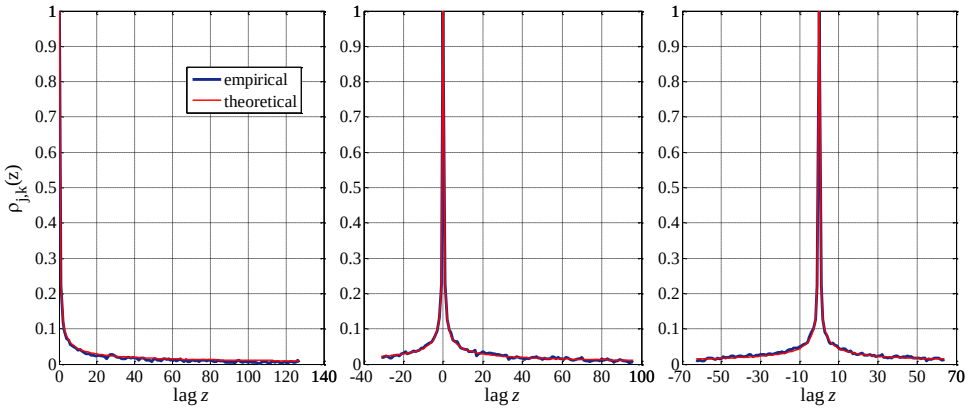


Figure 3.10 – Ensemble autocorrelation function of the example HK process at the cascade level $k = 7$ with starting point j (for $j = 1$, $n/4$ and $n/2$, respectively, from left to right) in the considered cascade level with $n = 2^7 = 128$ elements.

3.2.2 Application to an historical observed event

In this section the HK downscaling model is fitted to an historical observed event, i.e. one of the Iowa events at the 10-second timescale (event 3); for further details on the observational data, the reader is referred to Georgakakos et al. (1994). The historical hyetograph is shown in Fig. 3.11 (upper panel). It can be seen that the dataset comprises a single storm without intermittence. Thus, intermittence, despite being an important characteristic of the rainfall process, can be left out of this analysis. We aim at providing further information on the applicability of the downscaling approach based on the HK process to reproduce the pattern of rainfall time series at the 10-second resolution.

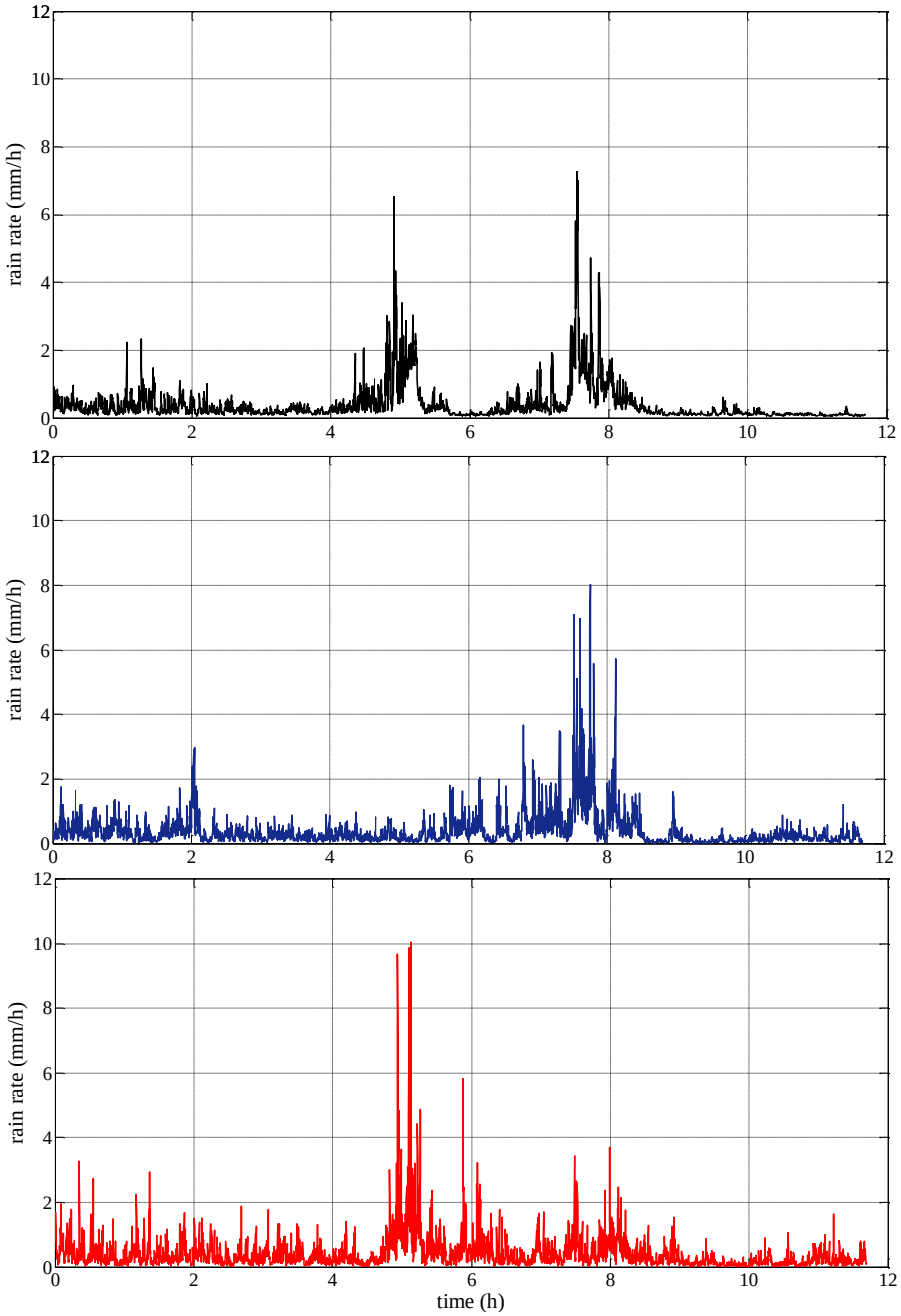


Figure 3.11 – Hyetograph of the historical rainfall event (no. 3) measured in Iowa on 30 November 1990 (upper panel; Georgakakos *et al.*, 1994) along with two synthetic time series of equal length generated by the MRC and HK models (middle and lower panels, respectively).

We estimated the HK model parameter from the real data, which is H=0.92 (see also Koutsoyiannis et al., 2007). Figure 3.12 (upper panel), depicts the climacogram (i.e. a double logarithmic plot of the standard deviation $\sigma(\Delta)$ of the aggregated process $\underline{X}_j^{(\Delta)}$ versus scale Δ) for both the real and the log-transformed datasets as a tool aiming at a multi-scale stochastic representation. It can be noticed that the two climacograms are approximately two parallel straight lines with high slopes ($H \cong 0.92$), which illustrates that the long-term persistence of the process is virtually invariant under a logarithmic transformation.

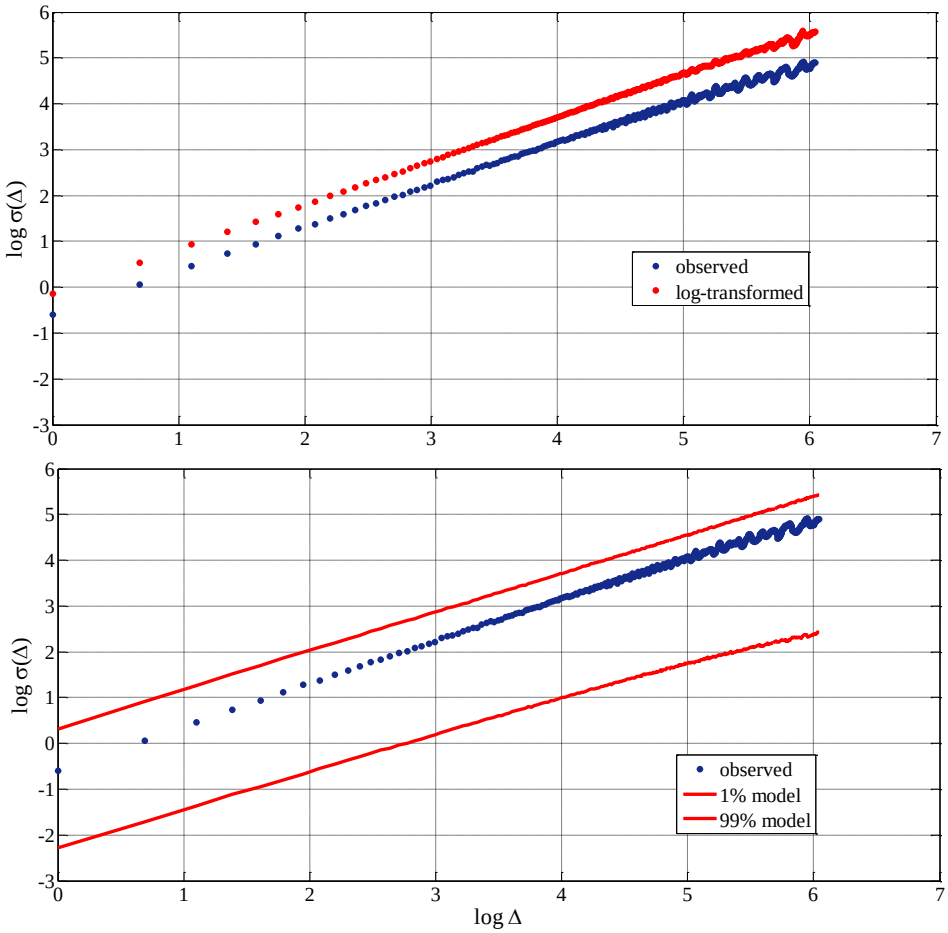


Figure 3.12 – Double logarithmic plot of the standard deviation $\sigma(\Delta)$ of the aggregated process $\underline{X}_j^{(\Delta)}$ vs. scale Δ (climacogram) for both the real and the log-transformed data of the Iowa rainfall event (upper panel); climacograms of the 1st and the 99th percentiles for the HK downscaling model (10000 Monte Carlo experiments) and for the observed rainfall event (lower panel).

We performed 10000 Monte Carlo experiments to downscale the aggregated rainfall event at the cascade level k = 13. Figure 3.12 (lower panel) shows the 1st and 99th percentiles of climacograms for the HK downscaling model to highlight the scaling behaviour of the simulated time series, which is practically consistent with the scaling properties of the observed rainfall event. Figure 3.13 depicts a comparison between the observed autocorrelogram with that simulated by our model; in particular, we plot the 1st and 99th percentiles of autocorrelation function. It can be noticed that the model on average fits the observed behaviour quite satisfactorily.

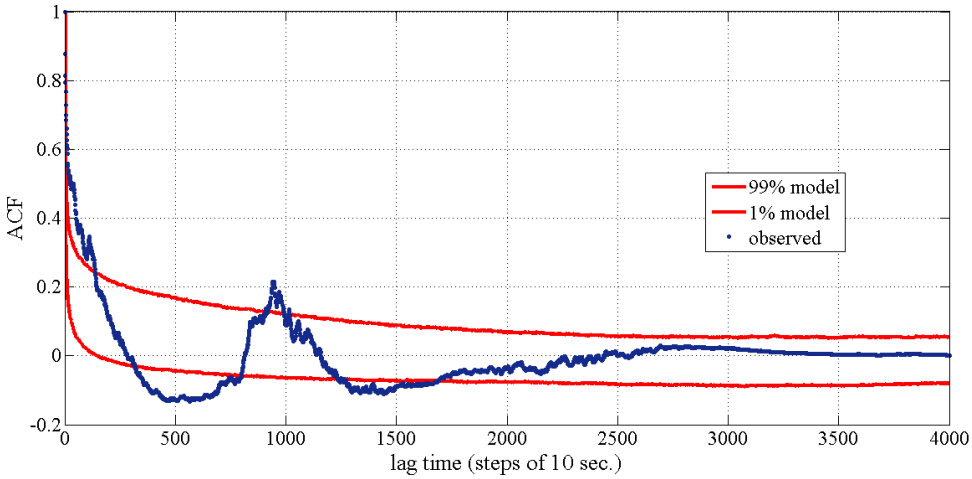


Figure 3.13 – Empirical autocorrelation function (ACF) of the Iowa rainfall event examined and 1st and 99th percentiles of ACF for the HK downscaling model.

Finally, the historical hyetograph is compared (see Fig. 3.11) to two typical synthetic hyetographs, of equal length, generated by the MRC and the HK downscaling models (the MRC model parameters were estimated from the real data imposing both the mean and the variance of the lower-level variables). We can see that both models produce realistic traces without apparent visual differences in the general shapes from each other and from the real world hyetograph (note that the models provide copies with statistical resemblance but not precise reproductions of the historical event). Despite being visually similar, the study of the details of the statistical behaviour of the two models has revealed that there are important differences.

3.3 HK disaggregation model

Generating finer scale time series of rainfall that are fully consistent with any given coarse-scale totals is an important issue in hydrology. This is commonly accomplished through disaggregation models, which let lower-level variables satisfy the equality constraint given by Eq. (3.3). In this section, we present a disaggregation method that initially retains the formalism, the parameter set, and the generation routine of the HK downscaling model described in Sect. 3.2. Then it uses an adjusting procedure to achieve the full consistency of lower-level and higher-level variables.

Our HK downscaling model is based upon a particular nonlinear transformation (see Eq. (3.36)) of the variables obtained by a stepwise disaggregation approach (see Eq. (3.30)), which generates time series with Hurst-Kolmogorov dependence structure. Unfortunately, nonlinear transformations of the variables do not preserve the additive property, which is one of the main attributes of the original disaggregation scheme. To overcome this problem, we use an empirical adjusting procedure in order to restore consistency, but such a procedure may, in turn, introduce bias in all statistics that are to be preserved. However, here we apply a particular procedure that has been proved, both theoretically and empirically, to be accurate in the sense that it preserves certain statistics of lower-level variables (Koutsoyiannis and Manetas, 1996). In other words, we modify the time series generated by our HK downscaling model in a way to be consistent with a given higher-level time series, without affecting the stochastic structure implied by our model.

In practise, we use the same linear generation scheme of Eq. (3.30) in the Gaussian auxiliary domain:

$$\begin{cases} \underline{\tilde{X}}_{2j-1,k} = \boldsymbol{\theta}^{\mathrm{T}} \underline{\boldsymbol{Y}} + \underline{v} \\ \underline{\tilde{X}}_{2j,k} = \underline{\tilde{X}}_{j,k-1} - \underline{\tilde{X}}_{2j-1,k} \end{cases} \tag{3.49}$$

In order to increase the accuracy of the model in reproducing the main statistics of the underlying stochastic process, the sequences of previous and past variables that are considered for generating each lower-level variable, and the related parameters, are expanded here with respect to the original version proposed in Eq. (3.30).

We assume $\underline{\boldsymbol{Y}} = \left[\underline{\tilde{X}}_{2j-5,k}, \underline{\tilde{X}}_{2j-4,k}, \underline{\tilde{X}}_{2j-3,k}, \underline{\tilde{X}}_{2j-2,k}, \underline{\tilde{X}}_{j,k-1}, \underline{\tilde{X}}_{j+1,k-1}, \underline{\tilde{X}}_{j+2,k-1}\right]^{\mathrm{T}}$, thus Eq. (3.49) allows the generated lower-level variable $\underline{\tilde{X}}_{2j-1,k}$ to

preserve autocorrelations with four earlier lower-level variables (level k) and two later higher-level variable (level k–1). The vector of parameters $\boldsymbol{\theta}$ is determined from Eq. (3.28) accordingly.

As stated above, when we apply our specific exponentiation (see Eq. (3.36)) to the HK process (auxiliary process) to make it log-normal (actual process), we introduce an error in the additive property, i.e., a departure of the sum of lower-level variables within a period from the corresponding higher-level total. The empirical adjusting procedure is introduced in the downscaling model in order to restore consistency, i.e. to preserve exactly the rainfall mass at the higher level of the actual process $\underline{X}_1^{(f)} = \underline{X}_{1,0}$.We use an accurate adjusting procedure to allocate the error in the additive property among the different lower-level variables: The power adjusting procedure. It is accurate because it preserves both the mean values and the variance-covariance matrix of the lower-level variables (Koutsoyiannis and Manetas, 1996). The power adjusting procedure modifies the generated lower-level variables $\underline{X}_{j,k}$ (j = 1, …, n=2^k) to get the adjusted ones $\underline{X}'_{j,k}$ according to:

$$\underline{X}'_{j,k} = \underline{X}_{j,k}\left(\underline{X}_{1,0}\Big/\sum_{j=1}^{n}\underline{X}_{j,k}\right)^{\lambda_{j,k}/\eta_{j,k}} \tag{3.50}$$

where:

$$\lambda_{j,k} = \sum_{i=1}^{n}\mathrm{Cov}\left[\underline{X}_{j,k},\underline{X}_{i,k}\right]\Big/\sum_{j=1}^{n}\sum_{i=1}^{n}\mathrm{Cov}\left[\underline{X}_{j,k},\underline{X}_{i,k}\right]^{\lambda_{j,k}/\eta_{j,k}}$$
$$\eta_{j,k} = \left\langle\underline{X}_{j,k}\right\rangle\Big/\sum_{j=1}^{n}\left\langle\underline{X}_{j,k}\right\rangle \tag{3.51}$$

This adjusting procedure does not preserve the additive property at once. Thus its application must be iterative, until the calculated sum of the lower-level variables are equal to the given $\underline{X}_{1,0}$. Due to the iterative application and the approximations made for its development, the procedure is not exact in strict sense, except for special cases (Koutsoyiannis and Manetas, 1996). However, we observed that in our case iterations converge rapidly. In addition, the power adjusting procedure may be a useful approximate generalization of the common proportional procedures retaining the advantage of returning positive values (as in our case for rainfall). Indeed, the power adjusting procedure has no limitations and it works for any type of probability distribution of lower-level variables (contrary to what is observed for common proportional adjusting procedures, which are often subject to severe

limitations). In order to show some preliminary results of our disaggregation method based on the Hurst-Kolmogorov process, we generate M = 50000 time series, as we did for the MRC and HK downscaling models in Sects. 3.1.2 and 3.2.1. To make the two model simulations comparable, we assume the same values of summary statistics as in the MRC case, i.e. k = 7, μ_k = 1 and γ_k = 1.5. Furthermore, we assume H = 0.7. Figs. 3.14 and 3.15 show, respectively, the behaviours of the ensemble mean $\langle \underline{x}_{j,k} \rangle = \mu_{j,k}$ and standard deviation $\sigma_{j,k}$ of the random processes as a function of the position j along the level k, $j = 1, 2, \ldots, 2^k$.

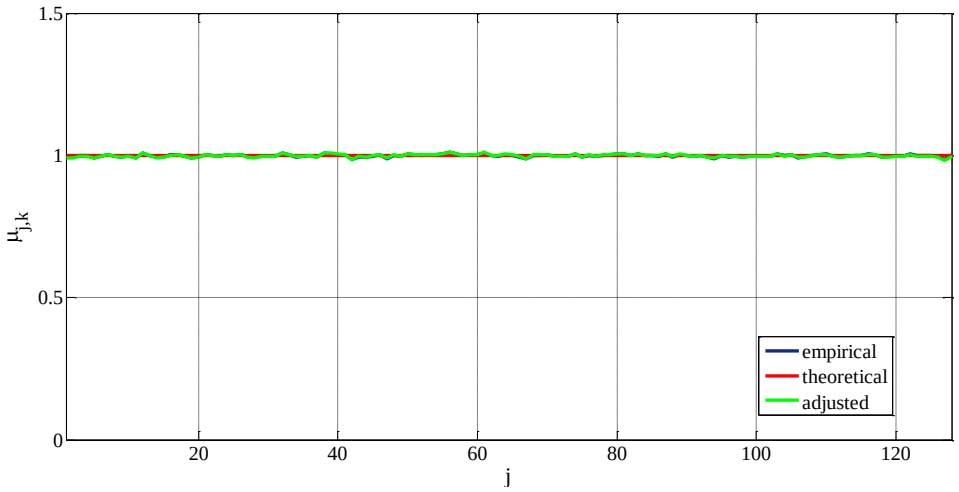


Figure 3.14 – Ensemble mean of the example HK disaggregation process as a function of the position j along the cascade level k = 7.

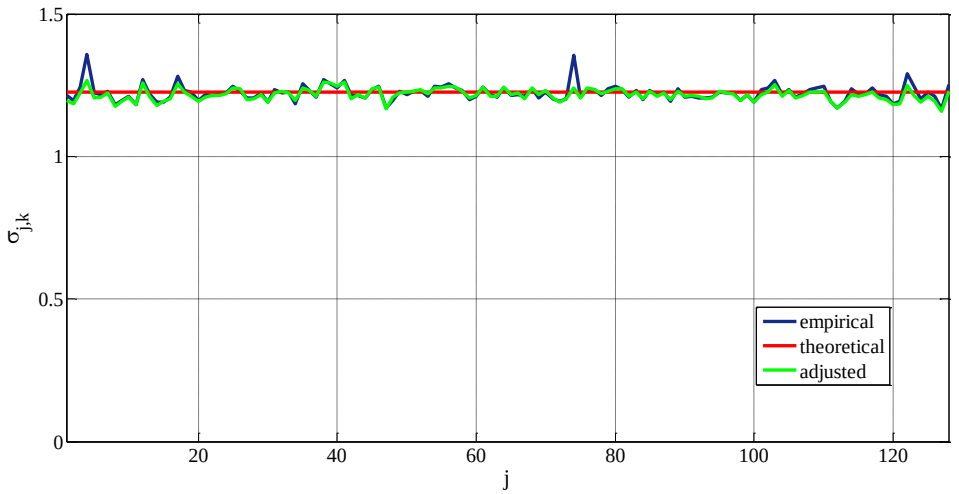


Figure 3.15 – Standard deviation of the example HK disaggregation process.

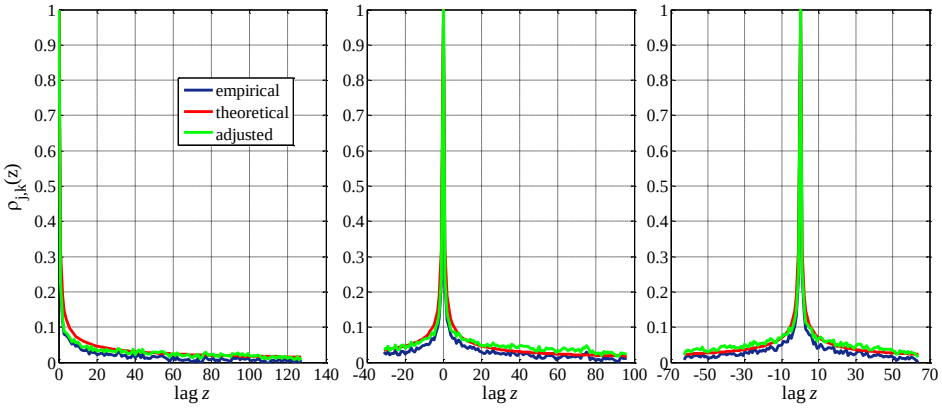


Figure 3.16 – Ensemble autocorrelation function of the example HK process at the cascade level $k = 7$ with starting point j (for $j = 1$, $n/4$ and $n/2$, respectively, from left to right) in the considered cascade level with $n = 2^7 = 128$ elements.

In Figs 3.14-3.16 the term "empirical" refers to time series generated by our downscaling model described in Sect. 3.2, while "adjusted" refers to the same time series modified by the power adjusting procedure in Eq. (3.50). All figures show that the stochastic structure implied by our downscaling model is not affected by the power adjusting procedure. However, the additive property is now fulfilled, as shown in Fig. 3.17.

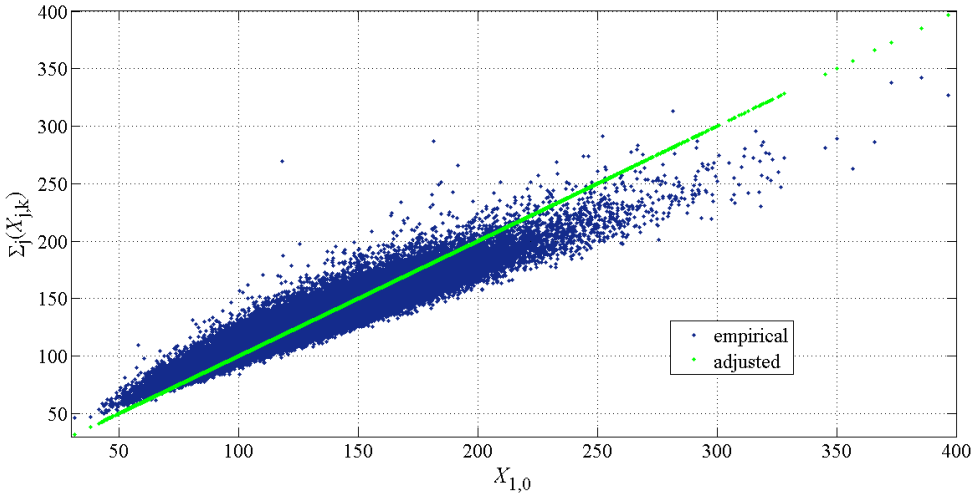


Figure 3.17 – Scatter plot of the calculated sum of the lower-level variables (before, blue, and after, green, applying the adjusting procedure) vs. the given values of the higher-level variables $\underline{X}_{1,0}$ for all Monte Carlo experiments.

3.4 Overview of key ideas

The discrete MRC has been a widely used approach of stochastic downscaling for rainfall time series. The usefulness of the discrete MRC relies on its simplicity and ability to generate time series characterized by both multifractal properties and complex intermittent and spiky patterns typical of rainfall time series.

By means of theoretical reasoning and Monte Carlo experiments, in Sect. 3.1 we show that the random process underlying the MRC model is not stationary, because its autocorrelation function is not a function of lag only, as it would be in stationary processes. Indeed, we provide a new theoretical formulation for the autocorrelation function of an unbounded canonical dyadic cascade, which is dependent on the lag, the position in time and the cascade level. As demonstrated, this undesirable violation of stationarity also extends to the micro-canonical and the bounded cascades. Consequently, MRC models cannot preserve joint statistical properties observed in real rainfall.

Mandelbrot (1974) made it clear that the structure of a discrete multiplicative cascade has problems. However, very many researchers miss this fact and treat these cascade models as if they were stationary (e.g. Menabde et al. 1997, Hingray and Ben Haha 2005, Gaume et al. 2007, Serinaldi 2010, Groppelli et al. 2011). Although fundamentally non-stationary, multiplicative random cascades were efficiently used to study the marginal and extreme distribution properties of stationary multifractal measures (see e.g. Veneziano et al. 2009 and references therein). Moreover, there exist other types of models intended to simulate multiscaling properties empirically observed in rainfall processes, which have been demonstrated to generate stationary processes, such as scale-continuous multifractal cascades (e.g., Lovejoy and Schertzer 2010a, b). However, this Chapter focuses on the analysis of discrete cascades, which are characterized by a very simple structure, easy to implement and, consequently, widely applied in the literature.

We propose and theoretically analyse an alternative downscaling approach (Sect. 3.2) based on the Hurst-Kolmogorov process, which is characterized by a simple cascade structure similar to that of MRC models, but it proves to be stationary. In its original formulation, this stepwise disaggregation approach effectively generates fractional Gaussian noise. However, the rainfall process (especially at the resolution needed for hydrological applications) is not Gaussian. Here we modified this approach to make it non-Gaussian by applying an exponential

transformation to the time series generated, so as to make it a more realistic representation of the actual rainfall process and more comparable to the MRC models. However, the logarithmic normalizing transformation, which we chose for theoretical simplicity, is not the best choice to normalize the dataset (Papalexiou et al., 2011).

Finally, we show (Sect. 3.3) some preliminary results of a disaggregation method that initially retains the formalism, the parameter set, and the generation routine of the HK downscaling model. Then it uses an adjusting procedure to achieve the full consistency of lower-level and higher-level variables. In particular, we use an accurate adjusting procedure to restore the additive property without affecting the mean values and the variance-covariance matrix of the lower-level variables.

4. Conclusions and discussion

E quindi uscimmo a riveder le stelle.
Dante Alighieri

Geophysical processes change irregularly on all time scales. Therefore, this change is hardly predictable in deterministic terms and demands stochastic descriptions. Several prominent applied statisticians and scientists recognized that many geophysical changes are closely related to the Hurst phenomenon, which has been detected in many long hydroclimatic time series and is stochastically equivalent to a simple scaling behaviour of process variability over time scale. As a result, long-term changes are much more frequent and intense than commonly perceived and, simultaneously, the future states are much more uncertain and unpredictable on long time horizons than implied by typical modelling practices. In this context, Hurst-Kolmogorov stochastic process (described in Sect. 1.3) may be the key to perceive multi-scale change and model the implied uncertainty and risk. Indeed, the reason for introducing this stationary stochastic process is that the "span of interdependence" between its random variables can be said to be infinite, thus resembling the strong interdependence between distant samples observed in many empirical studies in diverse fields of science, clearly including geophysics. The great advantage of the HK process is that it characterizes change by a single parameter (the Hurst coefficient) in a stochastic framework and in stationary terms.

The purpose of this thesis is to describe how to infer and model statistical properties of natural processes exhibiting scaling behaviours. In particular, we explore their statistical consequences with respect to the implied dramatic increase of uncertainty (Chapter 2), and propose a simple and parsimonious model that respects the Hurst phenomenon (Chapter 3).

In Chapter 2, we highlight the problems in inference from time series of geophysical processes. The classical statistical approaches, often used in geophysical modelling, are based upon several simplifying assumptions, tacit or explicit, such as independence in time and exponential distribution tails, which are invalidated in natural processes. Indeed, the study of natural processes reveals scaling behaviours in state (departure from

exponential distribution tails) and in time (departure from independence). We show that dependence in time implies that much larger samples are required in order to obtain estimates of similar reliability with classical statistics. Furthermore, we highlight the problematic estimation of moments for geophysical processes. In many studies, it has been a common practice to neglect this problem, which is introduced when the process exhibits dependence in time and is magnified when the distribution function significantly departs from the Gaussian form, which is an example of an exceptionally light-tailed distribution. We show that, even in quantities whose estimates are in theory unbiased, the dependence and non-normality affect significantly their statistical properties, and sample estimates based on classical statistics are characterized by high bias and uncertainty (see Sect. 2.1.3). In particular, statistical methods that use high order moments (> 3) are questionable. Therefore, we suggest that, because of estimation problems, the use of moments of order higher than two should be avoided, either in justifying or fitting models. Ignorance of increased uncertainty and inattentive use of high order moments may result in inappropriate modelling, wrong inferences and false claims about the properties of the processes. Evidently, the first two moments are necessary to use in all problems as they define the most important characteristics of the distribution, marginal (the first two moments) and joint (the second moment). Even for these two lowest moments it is important to study always their uncertainties and this only can be done in connection with a model fitted for the process of interest (as it is not possible to define uncertainty without specifying a model for the marginal distribution and dependence). The third moment is often useful as a measure of skewness but we should always be aware of its uncertainty; however, use of third moment is not the only way to identify and assess the skewness of a distribution. For example in parameter estimation of three-parameter distributions, it is better to avoid the method of moments and use other fitting methods such as maximum likelihood, L-moments, etc. Moments of order > 3 should be avoided in model identification and fitting because their estimation is problematic. If we have to use them, then it is imperative to specify their uncertainty and involve this uncertainty in any type of modelling and inference. Furthermore, we focus on second-order moments and specifically on autocorrelations (climacograms) and power spectra, which are the most extensively used concepts in the applications of stochastic processes. In particular, we study their estimation problems with reference to the effect

of time discretization, finite record length as well as data correlation on classical statistical estimators. We found that time discretization distorts the stochastic properties at small time scales, the finite length affects the properties at large time scales, while data correlation introduces (often uncontrollable) biases and uncertainties in statistical estimation. Based on a synthetic experiment for which all the stochastic properties of the underlying process are known, we compare the performances in statistical estimation of the climacogram and power spectrum. The power spectrum, despite being very powerful in identifying strong periodicities in time series, it has some problems in identifying scaling laws and weak periodicities. Specifically, time discretization, finite length of data and data correlation alter asymptotic slopes of periodograms by introducing biases and uncertainties that are uncontrollable (see Sect. 2.2.3). Moreover, the rough shape of the periodogram may result both in false periodicities and in misleading, inaccurate or even incorrect slopes (e.g. slope > -1 for frequency $\rightarrow 0$, which is infeasible as shown in Sect. 2.2.2). The important advantage of the climacogram over other common statistical tools is that its bias caused by the correlation structure of datasets can be determined analytically (usually in a closed form) and included in the estimation problem. However, direct estimation of climacogram is not possible merely from the data. We need to assume a stochastic model (see Sect. 2.2.4) which evidently influences the estimation of the process variance. Once the model is assumed and its parameters estimated based on the data, we can expand our calculations to estimate the variance for any time scale Δ. The concept of climacogram can be used also in the frequency domain to find a substitute for the power spectrum, which has similar properties (e.g. the asymptotic behaviours of the two are similar). In Sect. 2.2.4, we analyze the climacogram-based pseudospectrum and show 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.

In Chapter 3, we propose a stationary downscaling model of rainfall time series, based on the HK process, which is characterized by a cascade structure similar to that of the most popular multiplicative random cascade (MRC) models. Indeed, the discrete MRC has been a widely used approach of stochastic downscaling for rainfall time series. The usefulness of the discrete MRC relies on its simplicity and ability to generate time series characterized by both multifractal properties and

complex intermittent and spiky patterns typical of rainfall time series. By means of theoretical reasoning and Monte Carlo experiments, in Sect. 3.1 we show that the random process underlying the MRC model is not stationary, because its autocorrelation function is not a function of lag only, as it would be in stationary processes. Indeed, we provide a new theoretical formulation for the autocorrelation function of an unbounded canonical dyadic cascade, which is dependent on the lag, the position in time and the cascade level (see Sect. 3.1.1). As demonstrated, this undesirable violation of stationarity also extends to the micro-canonical and the bounded cascades (see Sects. 3.1.3 and 3.1.4). Consequently, MRC models cannot preserve joint statistical properties observed in real rainfall. We propose and theoretically analyse an alternative downscaling approach (Sect. 3.2) based on the Hurst-Kolmogorov process, which is characterized by a simple cascade structure similar to that of MRC models, but it proves to be stationary. In its original formulation, this stepwise disaggregation approach effectively generates fractional Gaussian noise. However, the rainfall process (especially at the resolution needed for hydrological applications) is not Gaussian. Here we modified this approach to make it non-Gaussian by applying an exponential transformation to the time series generated, so as to make it a more realistic representation of the actual rainfall process and more comparable to the MRC models. Finally, we show (Sect. 3.3) some preliminary results of a disaggregation method that initially retains the formalism, the parameter set, and the generation routine of the HK downscaling model. Then it uses an adjusting procedure to achieve the full consistency of lower-level and higher-level variables. In particular, we use an accurate adjusting procedure to restore the additive property without affecting the mean values and the variance-covariance matrix of the lower-level variables.

Future work may focus on the problems in estimating the self-similarity parameter (Hurst coefficient) of a Hurst-Kolmogorov stochastic process from data series. Indeed, for most of the estimation methods existing in the literature, it is not easy to obtain simple confidence intervals (see e.g. Beran, 1994; Tyralis and Koutsoyiannis, 2011). This makes it difficult to interpret results in an objective way. In this light, we should analyse if and how the estimation of the Hurst parameter affects the estimation of the common statistical tools studied in Chapter 2. Furthermore, we should also test the robustness of the results of our rainfall downscaling model

described in Chapter 3 in the presence of uncertainty in the estimation of the Hurst coefficient from data.

The concepts expressed in Chapter 3 should be also enriched by accounting for rainfall intermittency in the modelling framework. Indeed, the rainfall process features an intermittent character at fine timescales, and thus the probability that a time interval is dry is generally greater than zero. The capability of downscaling models to reproduce rainfall intermittency is a fundamental requirement in simulation. Therefore, we should investigate how our simple and parsimonious downscaling model may account for the variability of intermittency across timescales. Generally, the analysis and modelling of rainfall intermittency relate to the study of the rainfall occurrence process, which can be described by a binary valued stochastic process, with the values 0 and 1 representing dry and wet conditions, respectively. On the other hand, the non-zero rainfall process can be characterized for example by our disaggregation model described in Sect. 3.3. Hence, we need a modelling approach of a mixed type, with a discrete description of intermittency (varying across scales) and a continuous description of rainfall described by our disaggregation model.

Acknowledgements

There is not a more pleasing exercise of the mind than gratitude.
Joseph Addison

A little over four years ago, I started my work from which this thesis has developed. From that time until now, I met several people and did experiences that taught me a lot about science and life. This is why I am very grateful to my doctoral advisor Elena Volpi for giving me the chance to live this great opportunity for cultural growth.
I shaped the main ideas of my work during my frequent visits at the National Technical University of Athens. Indeed, I would like to thank especially Prof. Demetris Koutsoyiannis and all members of the ITIA Research Team for sharing many happy days with me and for inspiring my work.
I thank also Prof. Guido Calenda and Prof. Aldo Fiori for supporting my work with their wise advices. I warmly thank also all my colleagues of the Department of Engineering of the "Roma Tre" University of Rome for their amicable support, and in particular: Claudia Cecioni, Michele Di Lazzaro, Pietro Prestininzi, Alessandro Romano, Elisabetta Romano and Antonio Zarlenga.
Special thanks go to Francesco Napolitano, Fabio Russo, Luca Baldini, and Francesco Serinaldi for fruitful discussions.
Besides, very special thanks go to the Italian National Fire Corps for allowing me to focus on my research activities.
I also thank Valeria Montesarchio, Maura Rianna, and Elena Ridolfi for their collaboration.
Furthermore, I would like to strongly thank my family for its essential support. Finally yet importantly, I thank my beloved wife and future mom Silvia, without whom this thesis could not have been written.

Rome, 3rd of May 2014

Federico Lombardo

References

Ashkenazy, Y. The use of generalized information dimension in measuring fractal dimension of time series. *Physica A*, 271, 427-447, 1999.

Bak, P., Tang, C., and Wiesenfeld, K. Self-organized criticality: an explanation of 1/f noise. *Phys. Rev. Lett.*, 59 (4), 381-384, 1987.

Beran, J. Statistics for Long-Memory Processes, vol. 61 of Monographs on Statistics and Applied Probability. Chapman and Hall, New York, USA, 315 pp., 1994.

Berne, A., Delrieu, G., Creutin, J.-D., and Obled, C. Temporal and spatial resolution of rainfall measurements required for urban hydrology. *J. Hydrol.*, 299 (3–4), 166–179, 2004.

Blöschl, G., and Sivapalan, M. Scale issues in hydrological modelling: a review. *Hydrol. Processes*, 9, 251–290, 1995.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994) Time Series Analysis, Forecasting and Control. Prentice Hall, Upper Saddle River, New Jersey, USA, 598 pp., 1994.

Cârsteanu, A. A., and Foufoula-Georgiou, E. Assessing dependence among weights in a multiplicative cascade model of temporal rainfall. *J. Geophys. Res.*, 101 (D21), 26363–26370, 1996.

De Lima, M. I. P., and Grasman, J. Multifractal analysis of 15-min and daily rainfall from a semi-arid region in Portugal. *J. Hydrol.*, 220, 1-11, 1999.

El Adlouni, S., Bobée, B., and Ouarda, T. B. M. J. On the tails of extreme event distributions in hydrology. *J. Hydrol.*, 355 (1-4), 16-33, 2008.

Falconer, K. Fractal Geometry: Mathematical Foundations and Applications. John Wiley & Sons, Chichester, 288 pp., 1990.

Forbes, C., Evans, M., Hastings, N., and Peacock, B. Statistical Distributions. 4th. Ed., John Wiley & Sons, 212 pp., 2011.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C. Linking climate change modelling to impact studies: recent advances in downscaling techniques for hydrological modelling. *Int. J. Climatol.*, 27, 1547–1578, 2007.

Fraedrich, K., and Larnder, C. Scaling regimes of composite rainfall time series. *Tellus*, 45 A, 289-298, 1993.

Gaume, E., Mouhous, N., and Andrieu, H. Rainfall stochastic disaggregation models: Calibration and validation of a multiplicative cascade model. *Adv. Water Resour.*, 30, 1301–1319, 2007.

Georgakakos, K. P., Cârsteanu, A. A., Sturdevant, P. L., and Cramer, J. A. Observation and Analysis of Midwestern Rain Rates. *J. Appl. Meteorol.*, 33, 1433-1444, 1994.

Gires, A., Tchiguirinskaia, I., Schertzer, D., and Lovejoy, S. Influence of the zero-rainfall on the assessment of the multifractal parameters. *Adv. Water Resour.*, 45, 13-25, 2012.

Gneiting, T., and Schlather, M. Stochastic models that separate fractal dimension and the Hurst effect. *SIAM Rev.*, 46 (2), 269-282, 2004.

Grassberger, P., and Procaccia, I. Characterization of strange attractors. *Phys. Rev. Lett.*, 50 (5), 346-349, 1983.

Graybill, D. A. IGBP PAGES/World Data Center for Paleoclimatology. NOAA/NGDC Paleoclimatology Program, Boulder, Colorado, USA, 1990.

Groppelli, B., Bocchiola, D., and Rosso, R. Spatial downscaling of precipitation from GCMs for climate change projections using random cascades: A case study in Italy. *Water Resour. Res.*, 47, W03519, doi:10.1029/2010WR009437, 2011.

Gupta, V. K., and Waymire, E. C. A statistical analysis of mesoscale rainfall as a random cascade. *J. Appl. Meteorol.*, 32, 251–267, 1993.

Hemelrijk, J. Underlining random variables. *Stat. Neerl.*, 20, 1–7, 1966.

Hingray, B., and Ben Haha, M. Statistical performances of various deterministic and stochastic models for rainfall disaggregation. *Atmos. Res.*, 77, 152–175, 2005.

Hurst, H. E. Long term storage capacities of reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116, 776–808, 1951.

Klemeš, V. The Hurst Phenomenon: A puzzle?, *Water Resour. Res.*, 10(4), 675–688, 1974.

Kolmogorov, A. N. Wiener spirals and some other interesting curves in a Hilbert space. *Dokl. Akad. Nauk SSSR*, 26, 115–118, 1940.

Kottegoda, N. T., and Rosso, R. Applied Statistics for Civil and Environmental Engineers. 2nd Edn., Blackwell Publishing, 718 pp., 2008.

Koutsoyiannis, D. Coupling stochastic models of different time scales. *Water Resour. Res.*, 37 (2), 379–392, 2001.

Koutsoyiannis, D. The Hurst phenomenon and fractional Gaussian noise made easy. *Hydrolog. Sci. J.*, 47 (4), 573–595, 2002.

Koutsoyiannis, D. Climate change, the Hurst phenomenon, and hydrological statistics, *Hydrolog. Sci. J.*, 48 (1), 3–24, 2003.

Koutsoyiannis, D. Hurst, Joseph, colours and noises: The importance of names in an important natural behavior. *Niche Modeling*, 10 pages, 2006 (available at http://itia.ntua.gr/792).

Koutsoyiannis, D. A random walk on water. *Hydrol. Earth Syst. Sci.*, 14, 585–601, 2010.

Koutsoyiannis, D. Hurst-Kolmogorov dynamics as a result of extremal entropy production. *Physica A*, 390 (8), 1424-1432, 2011.

Koutsoyiannis, D. Hydrology and Change. *Hydrolog. Sci. J.*, 58 (6), 1177–1197, 2013a.

Koutsoyiannis, D.: Encolpion of stochastics: Fundamentals of stochastic processes. Department of Water Resources and Environmental Engineering – National Technical University of Athens, Greece, 2013b (available at http://itia.ntua.gr/1317).

Koutsoyiannis, D. and Langousis, A. Precipitation. In: *Treatise on water science*, edited by: Wilderer, P. and Uhlenbrook, S., Vol. 2, Oxford: Academic Press, 27–78, 2011.

Koutsoyiannis, D., and Manetas, A. Simple disaggregation by accurate adjusting procedures. *Water Resour. Res.*, 32 (7), 2105–2117, 1996.

Koutsoyiannis, D., and Montanari, A. Statistical analysis of hydroclimatic time series: Uncertainty and insights. *Water Resour. Res.*, 43 (5),

W05429, doi:10.1029/2006WR005592, 2007.

Koutsoyiannis, D., Papalexiou, S. M., and Montanari, A. Can a simple stochastic model generate a plethora of rainfall patterns? Oxford, Engineering and Physical Sciences Research Council, 2007. (available at http://itia.ntua.gr/789).

Koutsoyiannis, D., Paschalis, A., and Theodoratos, N. Two-dimensional Hurst-Kolmogorov process and its application to rainfall fields. *J. Hydrol.*, 398 (1-2), 91-100, 2011.

Langousis, A., and Koutsoyiannis, D. A stochastic methodology for generation of seasonal time series reproducing overyear scaling behaviour. *J. Hydrol.*, 322, 138–154, 2006.

Licznar, P., Łomotowski, J., and Rupp, D. E. Random cascade driven rainfall disaggregation for urban hydrology: An evaluation of six models and a new generator. *Atmos. Res.*, 99 (3–4), 563–578, 2011.

Lombardo, F., Volpi, E., and Koutsoyiannis, D. Rainfall downscaling in time: theoretical and empirical comparison between multifractal and Hurst-Kolmogorov discrete random cascades. *Hydrolog. Sci. J.*, 57 (6), 1052-1066, 2012.

Lombardo, F., Volpi, E., Koutsoyiannis, D. and Papalexiou, S. M. Just two moments! A cautionary note against use of high-order moments in multifractal models in hydrology. *Hydrol. Earth Syst. Sci.*, 18, 243–255, 2014.

Lopes, R., and Betrouni, N. Fractal and multifractal analysis: a review. *Med. Image Anal.*, 13 (4), 634-649, 2009.

Lovejoy, S., and Schertzer, D. On the simulation of continuous in scale universal multifractals, part I: spatially continuous processes. *Comput. Geosci.*, 36 (11), 1393–1403, 2010a.

Lovejoy, S., and Schertzer, D. On the simulation of continuous in scale universal multifractals, part II: space-time processes and finite size corrections. *Comput. Geosci.*, 36 (11), 1404–1413, 2010b.

Mach, J., Mas, F., and Sagués, F. Two representations in multifractal analysis. *J. Phys. A: Math. Gen.*, 28 (19), 5607-5622, 1995.

Mandelbrot, B. B. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *J. Fluid*

Mech., 62, 331-358, 1974.

Mandelbrot, B. B. The Fractal Geometry of Nature. Freeman, New York, 468 pp., 1982.

Mandelbrot, B. B. Multifractal power law distributions: Negative and critical dimensions and other "anomalies," explained by a simple example. *J. Stat. Phys.*, 110 (3-6), 739-774, 2003.

Mandelbrot, B. B., and Van Ness, J. W. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10 (4), 422-437, 1968.

Mandelbrot, B. B., and Wallis, J. R. Noah, Joseph and operational hydrology. *Water Resour. Res.*, 4 (5), 909-918, 1968.

Mandelbrot, B. B., and Wallis, J. R. Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resour. Res.*, 5 (5), 967–988, 1969.

Marani, M. On the correlation structure of continuous and discrete point rainfall. *Water Resour. Res.*, 39 (5), 1128, doi:10.1029/2002WR001456, 2003.

Markonis, Y., and Koutsoyiannis, D. Climatic variability over time scales spanning nine orders of magnitude: Connecting Milankovitch cycles with Hurst–Kolmogorov dynamics, *Surv. Geophys.*, 34 (2), 181–207, 2013.

Marshak, A., Davis, A., Cahalan, R., and Wiscombe, W. Bounded cascade models as nonstationary multifractals. *Phys. Rev.*, E49, 55–69, 1994.

Mascaro, G., Deidda, R., and Hellies, M. On the nature of rainfall intermittency as revealed by different metrics and sampling approaches. *Hydrol. Earth Syst. Sci.*, 17, 355-369, 2013.

Menabde, M., Harris, D., Seed, A., Austin, G., and Stow, D. Multiscaling properties of rainfall and bounded random cascades. *Water Resour. Res.*, 33 (12), 2823–2830, 1997.

Menabde, M., and Sivapalan, M. Modeling of rainfall time series and extremes using bounded random cascades and Levy-stable distributions. *Water Resour. Res.*, 36 (11), 3293–3300, 2000.

Meneveau, C., and Sreenivasan, K. R. The multifractal nature of turbulent energy dissipation. *J. Fluid Mech.*, 224, 429-484, 1991.

Mitzenmacher, M. A brief history of generative models for power law and lognormal distributions. *Internet Math.*, 1 (2), 226-251, 2004.

Molini, A., Katul, G. G., and Porporato, A.: Revisiting rainfall clustering and intermittency across different climatic regimes. *Water Resour. Res.*, 45, W11403, doi:10.1029/2008WR007352, 2009.

Molnar, P., and Burlando, P. Preservation of rainfall properties in stochastic disaggregation by a simple random cascade model. *Atmos. Res.*, 77, 137–151, 2005.

Molnar, P., and Burlando, P. Variability in the scale properties of high-resolution precipitation data in the Alpine climate of Switzerland. *Water Resour. Res.*, 44, W10404, doi:10.1029/2007WR006142, 2008.

Montesarchio, V., Lombardo, F., and Napolitano, F. Rainfall thresholds and flood warning: an operative case study. *Nat. Hazards Earth Syst. Sci.*, 9, 135-144, 2009.

Montanari, A., et al. "Panta Rhei – Everything Flows": Change in hydrology and society – The IAHS Scientific Decade 2013–2022. *Hydrol. Sci. J.*, 58 (6), 1256-1275, 2013.

Neuman, S. P. Apparent/spurious multifractality of data sampled from fractional Brownian/Lévy motions. *Hydrol. Processes*, 24 (15), 2056–2067, 2010.

Newman, M. E. J. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.*, 46 (5), 323-351, 2005.

Nykanen, D., and Harris, D. Orographic influences on the multiscale statistical properties of precipitation. *J. Geophys. Res.*, 108(D8), 8381, doi:10.1029/2001JD001518, 2003.

Olsson, J. Limits and characteristics of the multifractal behavior of a high-resolution rainfall time series. *Nonlin. Processes Geophys.*, 2, 23-29, 1995.

Ossiander, M., and Waymire, E. Statistical estimation theory for multiplicative cascades. *Ann. Statist.*, 28, 1533-1560, 2000.

Over, T. M. Modeling space-time rainfall at the mesoscale using random cascades. *Ph.D. Thesis*, University of Colorado, Boulder, 1995.

Over, T. M., and Gupta, V. K. A space-time theory of mesoscale rainfall using random cascades. *J. Geophys. Res.*, 101 (D21), 26319–26331,

doi:10.1029/96JD02033, 1996.

Papalexiou, S. M., Koutsoyiannis, D., and Makropoulos, C. How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrol. Earth Syst. Sci.*, 17, 851-862, 2013.

Papalexiou, S. M., Koutsoyiannis, D., and Montanari, A. Can a simple stochastic model generate rich patterns of rainfall events? *J. Hydrol.*, 411 (3–4), 279–289, 2011.

Papoulis, A. Probability, Random Variables and Stochastic Processes, 3rd edition. McGraw Hill, 666 pp., 1991.

Pawelzik, K. and Schuster, H. Generalized dimensions and entropies from a measured time series. *Phys. Rev. A*, 35 (1), 481-484, 1987.

Rupp, D. E., Keim, R. F., Ossiander, M., Brugnach, M., and Selker, J. S. Time scale and intensity dependency in multiplicative cascades for temporal rainfall disaggregation. *Water Resour. Res.*, 45, doi:10.1029/2008WR007321, 2009.

Russo, F., Lombardo, F., Napolitano, F., and Gorgucci, E. Rainfall stochastic modelling for runoff forecasting. *Phys. Chem. Earth*, 31, 1252–1261, 2006.

Schertzer, D., and Lovejoy, S. Hard and soft multifractal processes. *Physica A*, 185, 187-194, 1992.

Schertzer, D., and Lovejoy, S. Multifractals, generalized scale invariance and complexity in geophysics. *Int. J. Bifurcat. Chaos*, 21 (12), 3417-3456, 2011.

Serinaldi, F. Multifractality, imperfect scaling and hydrological properties of rainfall time series simulated by continuous universal multifractal and discrete random cascade models. *Nonlin. Processes Geophys.*, 17, 697–714, 2010.

Stewart, J. Positive definite functions and generalizations, an historical survey. *Rocky Mountain J. Math.*, 6 (3), 409-434, 1976.

Toussoun, O. Mémoire sur l'histoire du Nil. In: *Mémoires a l'Institut d'Egypte*, vol. 18, 366–404, 1925.

Tyralis, H., and Koutsoyiannis, D. Simultaneous estimation of the parameters of the Hurst-Kolmogorov stochastic process. *Stoch. Env. Res. Risk A.*, 25 (1), 21–33, 2011.

Vanmarcke, E. Random fields: Analysis and synthesis, MIT Press, Cambridge, MA., 382 pp., 1983.

Veneziano, D., and Furcolo, P. Improved moment scaling estimation for multifractal signals. *Nonlin. Processes Geophys.*, 16, 641–653, 2009.

Veneziano, D., Furcolo, P., and Iacobellis, V. Imperfect scaling of time and space-time rainfall. *J. Hydrol.*, 322 (1-4), 105-119, 2006.

Veneziano, D., and Langousis, A. Scaling and fractals in hydrology. *Advances in Data-based Approaches for Hydrologic Modeling and Forecasting*, edited by B. Sivakumar, World Scientific, Chapter 4, 145 pp., 2010.

Veneziano, D., Langousis, A., and Furcolo, P. Multifractality and rainfall extremes: A review. *Water Resour. Res.*, 42, doi:10.1029/2005WR004716, 2006.

Veneziano, D., and Lepore, C. The scaling of temporal rainfall. *Water Resour. Res.*, 48, W08516, doi:10.1029/2012WR012105, 2012.

Venugopal, V., Roux, S. G., Foufoula-Georgiou, E., and Arneodo, A. Revisiting multifractality of high-resolution temporal rainfall using a wavelet-based formalism. *Water Resour. Res.*, 42, W06D14, doi:10.1029/2005WR004489, 2006.

Verrier, S., De Montera, L., Barthès, L., and Mallet, C. Multifractal analysis of African monsoon rain fields, taking into account the zero rain-rate problem. *J. Hydrol.*, 389, 111-120, 2010.

Verrier, S., Mallet, C. and Barthès, L. Multiscaling properties of rain in the time domain, taking into account rain support biases. *J. Geophys. Res.*, 116, D20119, doi:10.1029/2011JD015719, 2011.

Villarini, G., Lang, J. B., Lombardo, F., Napolitano, F., Russo, F., and Krajewski, W. F. Impact of different regression frameworks on the estimation of the scaling properties of radar rainfall. *Atmos. Res.*, 86, 340-349, 2007.

Wallis, J. R., Matalas, N. C., and Slack, J. R.: Just a Moment! *Water Resour. Res.*, 10 (2), 211-219, 1974.

Yevjevich, V. M. Stochastic Processes in Hydrology. Water Resources Publications, Fort Collins, Colorado, USA, 276 pp., 1972.