

Hurst-Kolmogorov dynamics as a result of extremal entropy production

Demetris Koutsoyiannis

Department of Water Resources and Environmental Engineering, National Technical University of Athens, Greece (dk@itia.ntua.gr – <http://www.itia.ntua.gr/dk>)

Abstract. It is demonstrated that extremization of entropy production of stochastic representations of natural systems, performed at asymptotic times (zero or infinity) results in constant derivative of entropy in logarithmic time and, in turn, in Hurst-Kolmogorov processes. The constraints used include preservation of the mean, variance and lag-1 autocovariance at the observation time step, and an inequality relationship between conditional and unconditional entropy production, which is necessary to enable physical consistency. An example with real world data illustrates the plausibility of the findings.

Keywords. Stochastic processes, Long-term persistence, Hurst-Kolmogorov dynamics, Entropy, Uncertainty, Extremal entropy production

1. Introduction

Variational principles, such as the principle of extremal time for determining the path of light (a generalization of Hero's and Fermat's principles) and the principle of extremal action for determining the path of simple physical systems, are well established in physics. In complex systems, entropy maximization is a principle that can determine the thermodynamic equilibrium of a system. Historically, in classical thermodynamics, entropy was introduced as a conceptual and rather obscure notion via a rather circular logic*. Later, within statistical thermophysics, entropy has been given an austere and clear probabilistic meaning by Boltzmann and Gibbs, while Shannon used essentially the same entropy definition to describe the information content, which he also called entropy at von Neumann's suggestion [1]. Despite having the same name, information entropy and thermodynamic entropy are commonly regarded as two distinct notions†. One of the strongest proponents of this view is E. T. Jaynes [2]‡, who, notably, introduced the principle of maximum entropy as a principle for

* Cf. the classical definition $dS = dQ/T$, where S , Q and T denote entropy, heat and temperature, respectively; the definition is applicable in a reversible process, which in turn is one in which $dS = dQ/T$.

† This was also suggested by a reviewer of this paper.

‡ Jaynes [2] wrote: “*We must warn at the outset that the major occupational disease of this field is a persistent failure to distinguish between the information entropy, which is a property of any probability distribution, and the experimental entropy of thermodynamics, which is instead a property of a thermodynamic state as defined, for example by such observed quantities as pressure, volume, temperature, magnetization, of some physical system. They should never have been called by the same name; the experimental entropy makes no reference to any probability distribution, and the information entropy makes no reference to thermodynamics. Many textbooks and research papers are flawed fatally by the author's failure to distinguish between these entirely different things, and in consequence proving nonsense theorems.*”

inference. However, this distinction has been disputed by others: the “*information entropy [is] shown to correspond to thermodynamic entropy*” [1] and the two become precisely identical, if Boltzmann’s constant is taken equal to unity. Perhaps the fact that the information entropy is a dimensionless quantity, while the thermodynamic entropy has units (J/K), has been a major obstacle to see that the two are identical notions*. But this is an historical accident, related to the arbitrary introduction of temperature scales, which Boltzmann respected, thus multiplying the probabilistic entropy by the constant bearing his name. This has been made crystal clear now [3]† and it has been recognized that, given Shannon’s entropy definition, “*in principle we do not need a separate base unit for temperature, the kelvin*” and in essence, thermodynamic entropy, like the information entropy, is dimensionless [4].

Here we regard the thermodynamic and information entropy identical notions, for which we use the term Boltzmann-Gibbs-Shannon (BGS) entropy whose definition is given in Equation (3) below. Likewise, we regard the principle of maximum entropy both as a physical principle to determine equilibrium thermodynamic states of natural systems and as a logical principle to make inference about natural systems (in an optimistic view that our logic in making inference about natural systems could be consistent with the behavior of the natural systems).

Application of the principle of maximum entropy is rather simple when time is not involved in the problem studied, which is the case in systems in equilibrium. In studying systems out of the equilibrium, time should necessarily be involved [e.g., 5] and the paths of the systems should be inferred. The more recently developed, within non-equilibrium thermodynamics, extremal principles of entropy production attempt to predict the most likely path of system evolution [6,7,8]. Central among the latter are Prigogine’s minimum entropy production principle [9] and Ziegler’s maximum entropy production principle [10], which are not in contradiction [11].

Extremization of entropy or its production rate involves a set of constraints, which impose macroscopic physical restrictions in the probabilistic concept of entropy. For example, in the kinetic theory of gases [e.g.,12], if $f(\mathbf{p})d\mathbf{p}$ is the number of molecules with momenta between \mathbf{p} and $\mathbf{p} + d\mathbf{p}$ in a certain macroscopic volume, then the zeroeth moment (the integral of $f(\mathbf{p})$ over the entire volume) is the total number of molecules and should be constant due to

* The fact the kelvin is just a convention rather than a fundamental physical unit is reflected in the currently discussed proposal for redefining the kelvin scale indirectly, based on Boltzmann’s constant k (specifically, the kelvin, unit of thermodynamic temperature, is such that the Boltzmann constant is exactly $1.380\ 65 \times 10^{-23}$ J/K [4]); had k been chosen equal to one, without dimensions, the kelvin scale would not be needed at all).

† Atkins [3] writes: “*Although Boltzmann’s constant k is commonly listed as a fundamental constant, it is actually only a recovery from a historical mistake. If Ludwig Boltzmann had done his work before Fahrenheit and Celsius had done theirs, then ... we might have become used to expressing temperatures in the units of inverse joules... Thus, Boltzmann’s constant is nothing but a conversion factor between a well-established conventional scale and the one that, with hindsight, society might have adopted*”.

conservation of mass. Also, the first moment is constrained to a constant value due to the law of conservation of (macroscopic) momentum, and the second marginal moment is constrained by the law of conservation of energy (since $\|\mathbf{p}\|^2$ is proportional to the kinetic energy). Likewise, joint moments of momenta represent deviatoric stresses [13].

In studying very complex systems, e.g. the Earth climatic system or parts thereof, it may be difficult to establish a relationship of macroscopic and microscopic quantities. However, the entropy can still be determined in terms of the probability density function of merely macroscopic quantities, and constraints in entropy extremization could also be formulated on the basis of assigning a conceptual meaning to macroscopic properties (moments) of the process of interest. The values of these moments can be determined from a statistical sample, rather than deduced from fundamental physical laws. Thus, the first and second marginal moments (mean and variance) represent a mean level of the process and the intensity of fluctuations around the mean, respectively, whereas a joint moment of time lagged variables (autocovariance) represents time dependence. In natural processes, for short lags, the dependence (expressed in terms of autocovariance) must necessarily be positive because values at adjacent times should be close to each other.

In geophysical processes, the available length of observation records are usually short and restrict the reliable statistical estimation to the lowest marginal moments (mean, variance) and the smallest lags for autocovariances. The asymptotic stochastic properties of the processes, i.e. the marginal distribution tails and the tails of the autocovariance functions, can hardly be estimated from available records. However, these asymptotic properties are crucial for the quantification of future uncertainty, as well as for planning and design purposes. This work tries to connect statistical physics (the extremal entropy production concept, in particular) with stochastic representations of natural processes, which are used as modeling tools in planning and design purposes. Extremal entropy production may provide a theoretical background in such stochastic representations, which are otherwise solely data-driven. The focus of the study is the asymptotic behavior of the autocovariance function. The principal question is whether the autocovariance decays following an exponential or a power-type law. The two different behaviors are also known as short-term and long-term persistence, respectively. A typical representative of the former behavior is a Markov process, while one of the latter is the Hurst-Kolmogorov (HK) process (named so after Hurst [14], who pioneered its detection in geophysical time series, and Kolmogorov [15], who proposed the mathematical process), else known as fractional Gaussian noise [16]. Both these processes will be fully described below.

2. Basic notions

Let \underline{x}_i^A be a stationary stochastic process at discrete time $i = 0, 1, \dots$ (where time 0 denotes the present), representing a natural process x_i^A . (Notice that underlined quantities denote random variables, whereas regular mathematical quantities are not underlined). As natural processes

evolve in continuous time, it is natural to assume that \underline{x}_i^Δ are stationary increments, at time steps of equal size Δ , of a cumulative process $\underline{z}(t)$ in continuous time t , i.e., $\underline{x}_i^\Delta := \underline{z}(i\Delta) - \underline{z}((i-1)\Delta)$. Observing the natural process x_i^1 at time step, say, $\Delta = 1$ we can estimate from a statistical sample the following quantities:

$$E[\underline{x}_i^1] =: \mu, \text{Var}[\underline{x}_i^1] := E[(\underline{x}_i^1 - \mu)^2] =: \gamma, \text{Cov}[\underline{x}_i^1, \underline{x}_{i+1}^1] := E[(\underline{x}_i^1 - \mu)(\underline{x}_{i+1}^1 - \mu)] = \rho \gamma \quad (1)$$

where $E[\]$ denotes expectation and ρ is the lag-1 autocorrelation coefficient. The three equalities in (1) form the simplest possible set of constraints for multivariate entropy extremization. Alternatively, these constraints could be formulated in terms of the cumulative process $\underline{z}(t)$ (assuming $\underline{z}(0) = 0$) as:

$$E[\underline{z}(1)] = \mu, \text{Var}[\underline{z}(1)] = \gamma, \text{Var}[\underline{z}(2)] = 2(1 + \rho) \gamma \quad (2)$$

The problem we investigate here is the determination of the multivariate (joint) distribution of the process \underline{x}_i^Δ (or equivalently, $\underline{z}(t)$) by extremizing a proper entropic metric. It is known [e.g., 17] that, for a single random variable \underline{x} (rather than a stochastic process), extremization of the BGS entropy using the first two of the above constraints results in the normal (Gaussian) distribution of \underline{x} . Also, for many random variables \underline{x}_i with given correlation matrix, the BGS entropy extremizing distribution is the multivariate Gaussian [17]. We note that the BGS entropy definition may be insufficient for certain phenomena, where generalized entropies are necessary to consider, which result in different entropy extremizing distributions [18,19]; again, however, an appropriate nonlinear transformation (derived from the generalized entropy definition) may make the process Gaussian. For simplicity, in this study we adhere to the BGS entropy and use the above classical results to assert that the joint distribution of $\underline{z}(t)$ (as well as \underline{x}_i^Δ) of any order is Gaussian. The focus of the study is not the marginal distribution but the dependence structure (autocovariance) of \underline{x}_i^Δ and this can be readily inferred if the variance of $\underline{z}(t)$ is known for any time t . We recall that the BGS entropy of a single random variable \underline{z} with probability density function $f(z)$ is defined as*:

$$\Phi[\underline{z}] := E[-\ln f(\underline{z})] = -\int_{-\infty}^{\infty} f(z) \ln f(z) dz \quad (3)$$

and that for a Gaussian variable the entropy depends on its variance γ only and is given as [17]

$$\Phi[\underline{z}] = (1/2) \ln(2\pi e \gamma) \quad (4)$$

* Here we have used the older notation of entropy by Φ instead of the newer S and H , which in stochastics are reserved for standard deviation and Hurst coefficient, respectively. The quantity $\ln f(z)$ in (3) is meant, more precisely, as $\ln [f(z)/m(z)]$ (e.g. Ref. [2], p. 375) where $m(z)$ is meant as the Lebesgue density, i.e. equal to 1 times the unit $[z]^{-1}$, so that $\Phi[\underline{z}]$ is dimensionless.

Furthermore, for the stochastic process $\underline{z}(t)$ we can define entropy production as the time derivative $\Phi[\underline{z}(t)] := d\Phi[\underline{z}(t)]/dt$. For a Gaussian process, by virtue of (4), the entropy production is

$$\Phi[\underline{z}(t)] = [d\gamma(t)/dt] / 2\gamma(t) \quad (5)$$

We will use the notion of extremal entropy production to infer the variance of $\underline{z}(t)$, which in turn determines the entire dependence structure of \underline{x}_t^A , using the constraints (2) (or (1)). We clarify that, while other studies make an assumption on the type of dependence (e.g. long-range dependence [20]), here we attempt to derive, rather than assume, the dependence structure.

3. Methodology

We start our investigation from a linear Markovian process in continuous time, because of its simplicity and mathematical convenience, also assuming (without loss of generality) zero mean μ (we will remove this assumption in section 5). We note that any dependence structure of a stochastic process can be obtained by the sum of an infinite number of Markovian processes independent to each other [cf. 21]. Furthermore, the sum of a finite number of independent Markovian processes can provide good approximations of any dependence structure on a finite range of time scales. For example, in practical problems, an HK process can be approximated using as few as three Markovian processes [22].

A linear Markovian process $\underline{q}(t)$ in continuous time, else known as Ornstein–Uhlenbeck (OU) process [17,23], with zero mean is defined by

$$(1/k) d\underline{q} + \underline{q} dt = \sigma \underline{v} dt \quad (6)$$

where $k > 0$ and $\sigma > 0$ are parameters, and \underline{v} is the white noise process (i.e., $\underline{v} dt = d\underline{w}$ where \underline{w} is the Wiener process) with mean $E[\underline{v}(t)] = 0$ and covariance function $\text{Cov}[\underline{v}(t), \underline{v}(s)] = \delta(t - s)$, whereas $\delta(\cdot)$ is the Dirac delta function. The quantity $1/k$ has units of time and represents a characteristic time scale of the process, whereas the quantity σ^2 represents a characteristic variance. The linear differential equation (6) is easily solved to give

$$\underline{q}(t) = \underline{q}(0) e^{-kt} + k \sigma e^{-kt} \int_0^t \underline{v}(\zeta) e^{k\zeta} d\zeta \quad (7)$$

and has mean and variance, respectively,

$$E[\underline{q}(t)] = e^{-kt} E[\underline{q}(0)], \text{Var}[\underline{q}(t)] = \text{Var}[\underline{q}(0)] e^{-2kt} + (k \sigma^2/2) (1 - e^{-2kt}) \quad (8)$$

The cumulative OU process is

$$\underline{z}(t) = \int_0^t \underline{q}(\zeta) d\zeta = \underline{q}(0) (1 - e^{-kt})/k + \sigma \int_0^t \underline{v}(\zeta) [1 - e^{-k(t-\zeta)}] d\zeta \quad (9)$$

and has mean and variance, respectively,

$$E[\underline{z}(t)] = E[\underline{q}(0)] (1 - e^{-kt})/k \quad (10)$$

$$\text{Var}[\underline{z}(t)] = \text{Var}[\underline{q}(0)] (1 - e^{-kt})^2 / k^2 + \sigma^2 [2kt + 1 - (2 - e^{-kt})^2] / (2k) \quad (11)$$

Two cases are most interesting to consider: the *unconditional* case, in which no initial condition is known, and the *conditional* case where the present and the entire past of $q(t)$, $t \leq 0$, are known (observed). In the former case, stationarity in $\underline{q}(t)$ demands that $E[\underline{q}(t)] = E[\underline{q}(0)] = 0$ and $\text{Var}[\underline{q}(t)] = \text{Var}[\underline{q}(0)]$. Substituting this into (8), we find $\text{Var}[\underline{q}(0)] = k \sigma^2/2 =: \omega$, so that in the unconditional (stationary) case we have (by virtue of (11); see also [17]):

$$E[\underline{q}(t)] = E[\underline{z}(t)] = 0, \text{Var}[\underline{q}(t)] = \omega, \gamma(t) := \text{Var}[\underline{z}(t)] = \frac{2(kt + e^{-kt} - 1)}{k^2} \omega \quad (12)$$

In the conditional case, since the process is Markovian, only the most recent observed value $\underline{q}(0) = q(0)$ matters in conditioning on the past, and also $\text{Var}[\underline{q}(0)] = 0$. Thus,

$$E[\underline{q}(t)|q(0)] = q(0) e^{-kt}, E[\underline{z}(t)|q(0)] = q(0) (1 - e^{-kt})/k, \text{Var}[\underline{q}(t)|q(0)] = (1 - e^{-2kt}) \omega, \\ \gamma_C(t) := \text{Var}[\underline{z}(t)|q(0)] = \frac{2kt + 1 - (2 - e^{-kt})^2}{k^2} \omega \quad (13)$$

By means of $\gamma(t)$ and $\gamma_C(t)$ in (12) and (13) and by virtue of (4) we can determine, at any time t , the unconditional and conditional entropy of $\underline{z}(t)$. Likewise, by virtue of (5), we can determine the unconditional and conditional entropy production, which are, respectively,

$$\Phi'[\underline{z}(t)] = \frac{k(1 - e^{-kt})}{2(kt + e^{-kt} - 1)}, \Phi'_C[\underline{z}(t)] = \frac{k(1 - e^{-kt})^2}{2kt + 1 - (2 - e^{-kt})^2} \quad (14)$$

We observe that in both cases the entropy production is proportional to k , the inverse characteristic time scale, depends on the dimensionless time kt , and is independent of the characteristic variance ω . We also observe that the limit of both $\Phi'[\underline{z}(t)]$ and $\Phi'_C[\underline{z}(t)]$ for $t \rightarrow 0$ is ∞ and for $t \rightarrow \infty$ is 0. Consequently, any composite process determined as a sum of Markovian processes will also have the same limits. This does not enable comparison of the asymptotic behaviour of different processes (dependence structures). On the other hand, the asymptotic values of entropy production for $t \rightarrow 0$ and ∞ are the most important for model comparisons and eventually for extremization of entropy production. Otherwise, comparisons of different models at any specified finite time t would involve a degree of subjectivity and arbitrariness for the choice of t . In a similar situation, Koutsoyiannis [24] used an average entropy over a range of time scales.

Here, we follow a different method to tackle the infinite and zero entropy production. We first observe that for any time t an inequality relationship between any two stochastic processes 1 and 2, i.e. $\Phi'[\underline{z}_1(t)] \leq \Phi'[\underline{z}_2(t)]$ is preserved if we replace entropy production Φ' with the following quantity, which we will call entropy production in logarithmic time (EPLT):

$$\varphi[\underline{z}(t)] := \Phi'[\underline{z}(t)] t \equiv d\Phi[\underline{z}(t)] / d(\ln t) \quad (15)$$

Therefore, extremization of entropy production is equivalent to extremization of EPLT. In the OU model the unconditional and conditional EPLTs are, respectively,

$$\varphi[\underline{z}(t)] = \frac{kt(1 - e^{-kt})}{2(kt + e^{-kt} - 1)}, \varphi_C[\underline{z}(t)] = \frac{kt(1 - e^{-kt})^2}{2kt + 1 - (2 - e^{-kt})^2} \quad (16)$$

These quantities depend on the dimensionless time kt only and their limits are finite, i.e. $\varphi[\underline{z}(0)] = 1$, $\varphi_C[\underline{z}(0)] = 3/2$, and $\varphi[\underline{z}(\infty)] = \varphi_C[\underline{z}(\infty)] = 1/2$.

As stated above, we can materialize a stochastic process with an arbitrary dependence structure as the sum of a number m of OU processes independent to each other, i.e. $\underline{z}(t) := \sum_{j=1}^m \underline{z}_j(t)$, where $\underline{z}_j(t)$ has characteristic variance ω_j and characteristic time scale $1/k_j$. The variances of $\underline{z}(t)$ will then be

$$\gamma(t) = \sum_{j=1}^m 2(k_j t + e^{-k_j t} - 1) \omega_j / k_j^2, \gamma_C(t) = \sum_{j=1}^m [2k_j t + 1 - (2 - e^{-k_j t})^2] \omega_j / k_j^2 \quad (17)$$

and the EPLTs

$$\varphi[\underline{z}(t)] = \frac{t}{\gamma(t)} \sum_{j=1}^m (1 - e^{-k_j t}) \omega_j / k_j, \varphi_C[\underline{z}(t)] = \frac{t}{\gamma_C(t)} \sum_{j=1}^m (1 - e^{-k_j t})^2 \omega_j / k_j \quad (18)$$

Clearly, the characteristic variances ω_j are weights of each of the constituents $\underline{z}_j(t)$ of the composite process $\underline{z}(t)$ and the EPLTs of $\underline{z}(t)$ depend in a nonlinear manner on all these weights.

4. Application

The EPLT extremization problem is formulated as

$$\begin{aligned} &\text{extremize } \varphi[\underline{z}(t)] \text{ or } \varphi_C[\underline{z}(t)] \text{ for } t \rightarrow 0 \text{ or } t \rightarrow \infty \\ &\text{subject to constraints of Equation (2)} \end{aligned} \quad (19)$$

The control variables (unknown quantities) to be determined are the weights ω_j for specified inverse characteristic time scales k_j . By appropriate choice of k_j , analytical solutions may be possible, but here we tackle the optimization problem numerically. The numerical procedure is quite general, so that it can host constraints more than those contained in (2) (e.g. more than one autocovariance value), but here we adhere to the simplest possible set of constraints for parsimony. At the same time, the framework is quite simple and, thus, the extremization could be performed by widespread all-purpose optimization tools (here we used the generalized reduced gradient tool by Frontline Systems, v3.5, implemented as a Solver in Excel).

In our numerical framework we investigate times t in the range 2^{-10} to 2^{10} (a range spanning approximately 6 orders of magnitude) and we use inverse characteristic time scales

k_j in the range $k_{\min} = k_1 = 2^{-15}$ to $k_{\max} = 2^{15}$, with $k_j = 2 k_{j-1}$ (a range spanning approximately 9 orders of magnitude). The unknown quantities are the $m = 31$ characteristic variances ω_j . These are determined by extremizing (maximizing or minimizing) the EPLTs for the lowest and highest times of the range considered (2^{-10} to 2^{10}). These are supposed to approximate the limits for $t \rightarrow 0$ and $t \rightarrow \infty$, respectively. We use the three equality constraints in (2) and, to force determination of a limit rather than a local optimum for time 2^{-10} or 2^{10} , we set an additional inequality constraint that the standard deviation of the EPLT values at five consecutive time scales nearest to 2^{-10} or 2^{10} is lower than a small value ε .

We performed several numerical optimizations of this type and we generalized the results which proved to be very simple. An example is shown in Figure 1 in terms of the extremizing weights ω_j for each inverse characteristic time scale k_j , and in Figure 2 in terms of the resulting EPLTs. This example corresponds to constraints $\mu = 0$, $\gamma = 1$ and $\rho = 0.543$; the latter value was chosen so as to correspond to an OU process with $k = 1$. Two characteristic processes emerge by EPLT extremization. The first is a single OU process, in which all ω_j are zero except one, namely that corresponding to $k = 1$. The single OU process *maximizes* both EPLTs for $t \rightarrow 0$, with $\varphi[\underline{z}(0)] = 1$ and $\varphi_C[\underline{z}(0)] = 3/2$. The same process is the one that *minimizes* both EPLTs for $t \rightarrow \infty$, with $\varphi[\underline{z}(\infty)] = \varphi_C[\underline{z}(\infty)] = 1/2$.

The second extremizing process, denoted in the figures as the Hybrid process, has again all ω_j zero except two, i.e. those corresponding to $k_{\min} = 2^{-15}$ and $k_{\max} = 2^{15}$. This *minimizes* both EPLTs for $t \rightarrow 0$, with $\varphi[\underline{z}(0)] = \varphi_C[\underline{z}(0)] = 1/2$ and *maximizes* EPLTs for $t \rightarrow \infty$, with $\varphi[\underline{z}(\infty)] = 1$ and $\varphi_C[\underline{z}(\infty)] > 1$.

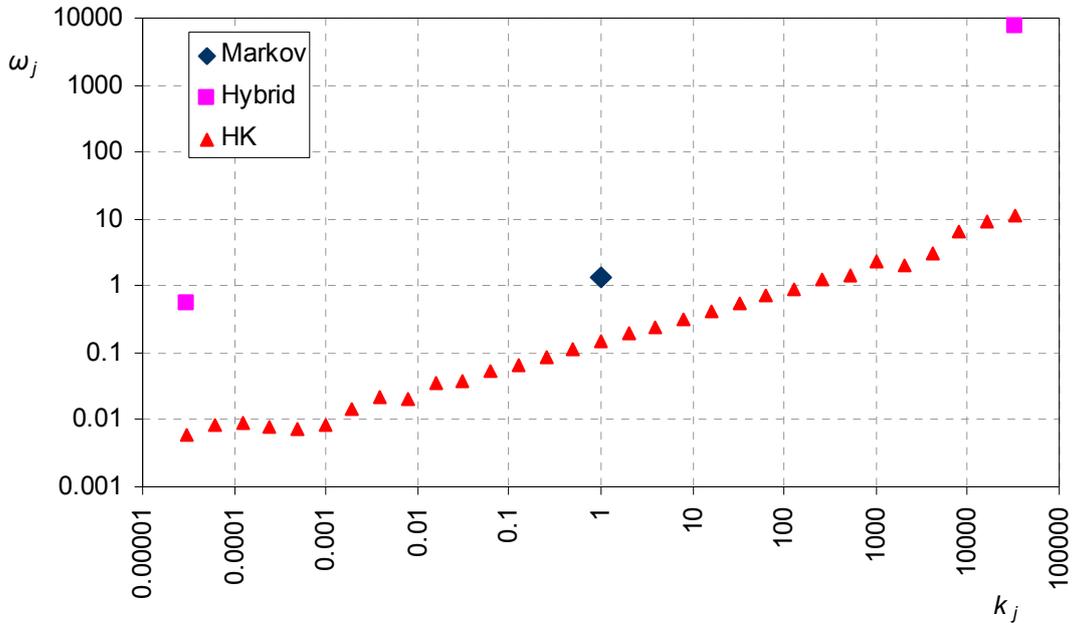


Figure 1 Entropy extremizing characteristic variances (weights of the composite processes) ω_j versus the inverse characteristic time scales k_j , as obtained by the numerical framework for the three solutions discussed in text.

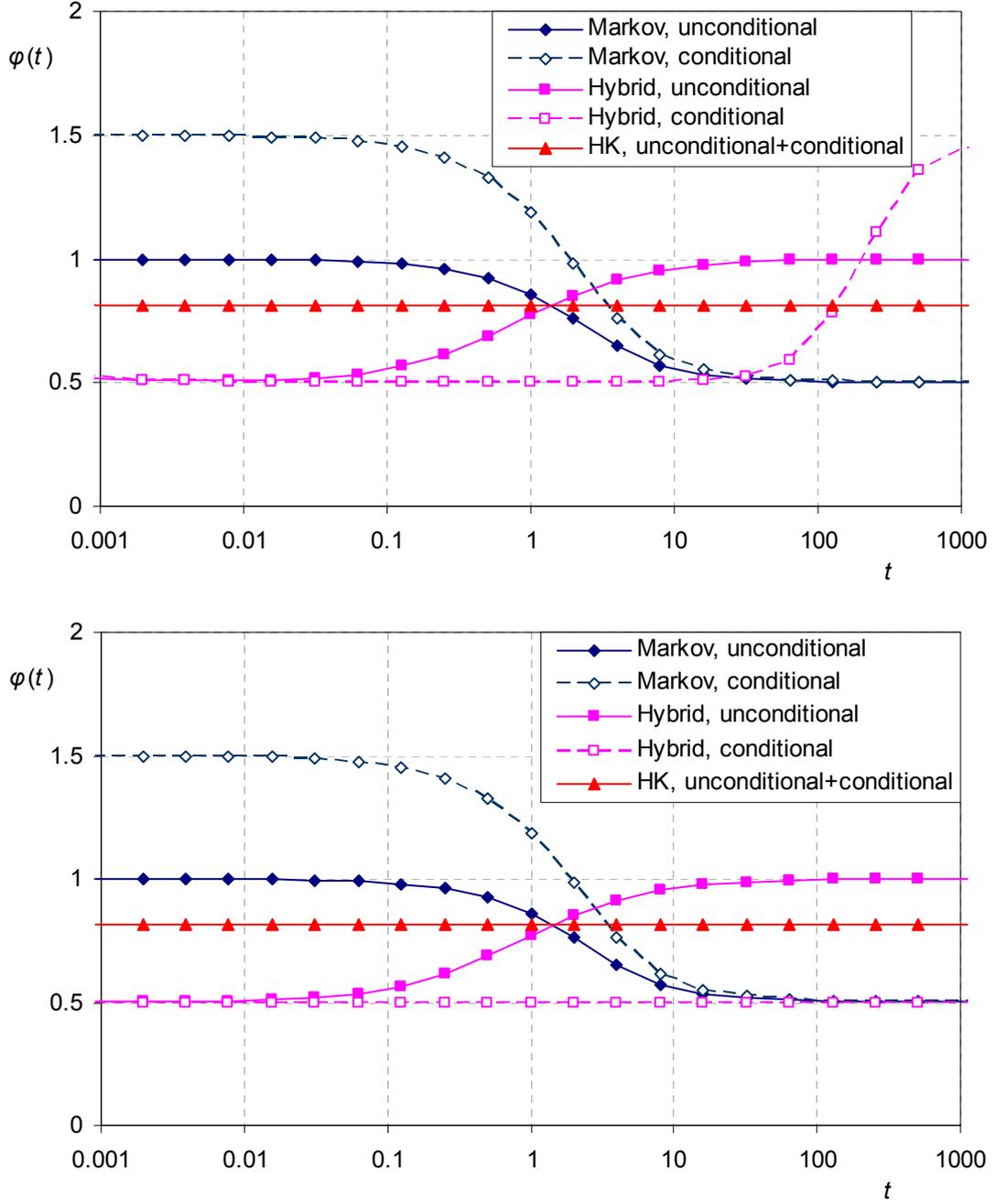


Figure 2 Entropy production in logarithmic time $\varphi(t) \equiv \varphi[\underline{z}(t)]$ versus time t in the three extremizing solutions discussed in text and shown in Figure 1; (upper) as determined by the numerical framework; (lower) analytical models.

However, it is readily understood that the latter numerical solution is artificially affected because of the use of finite minimum and maximum scales. We can thus infer that the precise analytical solution would be the one in which $k_{\min} \rightarrow 0$ and $k_{\max} \rightarrow \infty$. It can be derived from (12) and (13) that, as $k_{\max} \rightarrow \infty$, $\gamma(t) = \gamma_c(t) \rightarrow 2\omega_{\max}t/k_{\max} = at$ (where $a = 2\omega_{\max}/k_{\max}$), whereas, as $k_{\min} \rightarrow 0$, $\gamma(t) \rightarrow \omega_{\min} t^2 = bt^2$ (where $b = \omega_{\min}$) and $\gamma_c(t) = 0$. Thus, in the analytical extremizing solution, the variance of $\underline{z}(t)$ should be the sum $\gamma(t) = at + bt^2$ and the

conditional variance should be $\gamma_C(t) = at$. Given constraints (2), we can calculate the parameters a and b as $a = 2\gamma(1) - \gamma(2)/2 = (1 - \rho)\gamma$ and $b = -\gamma(1) + \gamma(2)/2 = \rho\gamma$. Its EPLTs are thus $\varphi[\underline{z}(t)] = (1 - \rho + 2\rho t) / (2 - 2\rho + 2\rho t)$ (so that $\varphi[\underline{z}(0)] = 1/2$ and $\varphi[\underline{z}(\infty)] = 1$), and $\varphi_C[\underline{z}(t)] = 1/2$ (constant). These analytical solutions have been depicted in Figure 2 (lower panel), where it can be seen that they are indistinguishable from the numerical ones except for $\varphi_C[\underline{z}(t)]$ for very large times ($t > 2^5$), for which the numerical solution indicates values $> 1/2$ while the analytical solution is exactly $1/2$.

Are these mathematically optimal results physically meaningful? It seems that both have problems. The OU process heavily depends on the time scale (here assumed equal to 1) in which the constraints in (2) were formulated; a different OU process would perhaps be obtained if the three constraints were formulated at a different time scale. The problem here emerges because the constraints were quantified based on induction (observational data as explained in the introduction) rather than deduction. It is not reasonable to expect that the actual dependence structure of a process might be affected by the choice of the time scale of observation. The Hybrid process has a more fundamental problem: As shown in Figure 2, the conditional entropy production is lower than the unconditional one at all times t . This contradicts physical realism: Naturally, by observing the present state of a process (at $t = 0$), the future entropy is reduced. In other words, the conditional entropy should be lower than the unconditional for $t > 0$, whereas as $t \rightarrow \infty$ conditional and unconditional entropies should tend to be equal. However, this cannot happen if the entropy production is consistently lower in the conditional than in the unconditional case. The problem arises here because of the fact that noises, such as the constituents of the Hybrid process, are just mathematical constructs, whereas Nature does not produce noises—rather it produces uncertainty, i.e. entropy [25]. As will be seen in section 5, the Hybrid process has additional problems related to its observability.

For this reason, to reinstate physical plausibility we impose an inequality constraint $\varphi_C[\underline{z}(t)] \geq \varphi[\underline{z}(t)]$, in addition to the equality constraints (2). In this case, the Hybrid model is replaced by one in which the inequality constraint becomes binding. Our numerical framework resulted in an extremizing solution whose weights ω_j are all nonzero as shown in Figure 1 (the solution marked as “HK”) and whose resulting EPLTs are shown in Figure 2. The same solution is extremizing for both $t \rightarrow 0$ and $t \rightarrow \infty$ and for both the conditional and unconditional cases. In this solution,

$$\varphi_C[\underline{z}(t)] = \varphi[\underline{z}(t)] = H \quad (20)$$

(independent of time). Despite being constant, this solution *minimizes* both EPLTs for $t \rightarrow 0$ and *maximizes* them for $t \rightarrow \infty$. The small roughness of the curve shaped by the weights ω_j , as a function of the inverse scale k_j , in Figure 1 may be due to numerical effects, but these effects do not seem to disturb the constancy of EPLTs in Figure 2.

Under constant EPLT equal to H , combining (5) and (15) we obtain $(1/2) d(\ln \gamma)/d(\ln t) = H$, or $d\gamma/\gamma = 2H dt/t$. This results in

$$\gamma(t) = t^{2H} \gamma(1) \quad (21)$$

which is readily recognized as the variance function of the cumulative HK process. In effect, (21) can serve as a definition of the cumulative HK process, while the constant H is the well known Hurst coefficient. The fact that H in (21) is identical to the EPLT gives it a sound physical interpretation. Based on a numerical approximation from [24], the conditional variance of the HK process is

$$\gamma_c(t) = t^{2H} c \gamma(1), \quad c := 1 - (2H - 1)^2 [0.72(H - 1) + 1] \quad (22)$$

Unlike the OU and Hybrid processes, the HK extremizing solution does not involve apparent inconsistencies with physical reality.

5. A real world example

To examine the consistency with reality of each of the three stochastic processes extremizing entropy production we present an example with real world data. We chose one of the longest available instrumental geophysical records, the mean annual temperature of Vienna, Austria (WMO station code: 11035 WIEN/HOHE_WAR; coordinates: 48.25°N, 16.37°E, 209 m) which comprises 235 years of data (1775-2009) available online from the Royal Netherlands Meteorological Institute (<http://climexp.knmi.nl>). Part of the record (141 years; 1851-1991) has been included in the Global Historical Climatology Network (GHCN). Comparison of the GHCN time series (available from the same web site), which has undergone consistency checking and correction, with the original one shows that adjustments are minor (less than $\pm 0.25^\circ$ at the annual level), which increases confidence to the longer original data set. Both time series are shown in Figure 3 (upper panel). The Gaussian distribution fits very well the time series, whose skewness is almost zero.

The EPLT cannot be calculated directly from the time series but the entropy $\Phi[z(t)]$ can, for a certain range of t , based on (4) and using sample estimates $g(t)$ of $\gamma(t)$. Denoting the total observation period as T (235 years) and assuming a varying time step Δ we form samples of \underline{x}_i^Δ with size $n = T/\Delta$ (more precisely, the floor of T/Δ). Here we varied Δ from 1 to 23 years, so that the sample size be at least 10. For each Δ , the sample mean is

$$\bar{\underline{x}}^\Delta \equiv \underline{z}(T) (\Delta/T) \quad (23)$$

which is an unbiased estimator of the true mean μ , and the sample variance is

$$g(\Delta) = \frac{1}{T/\Delta - 1} \sum_{i=1}^{T/\Delta} (\underline{x}_i^\Delta - \bar{\underline{x}}^\Delta)^2 \quad (24)$$

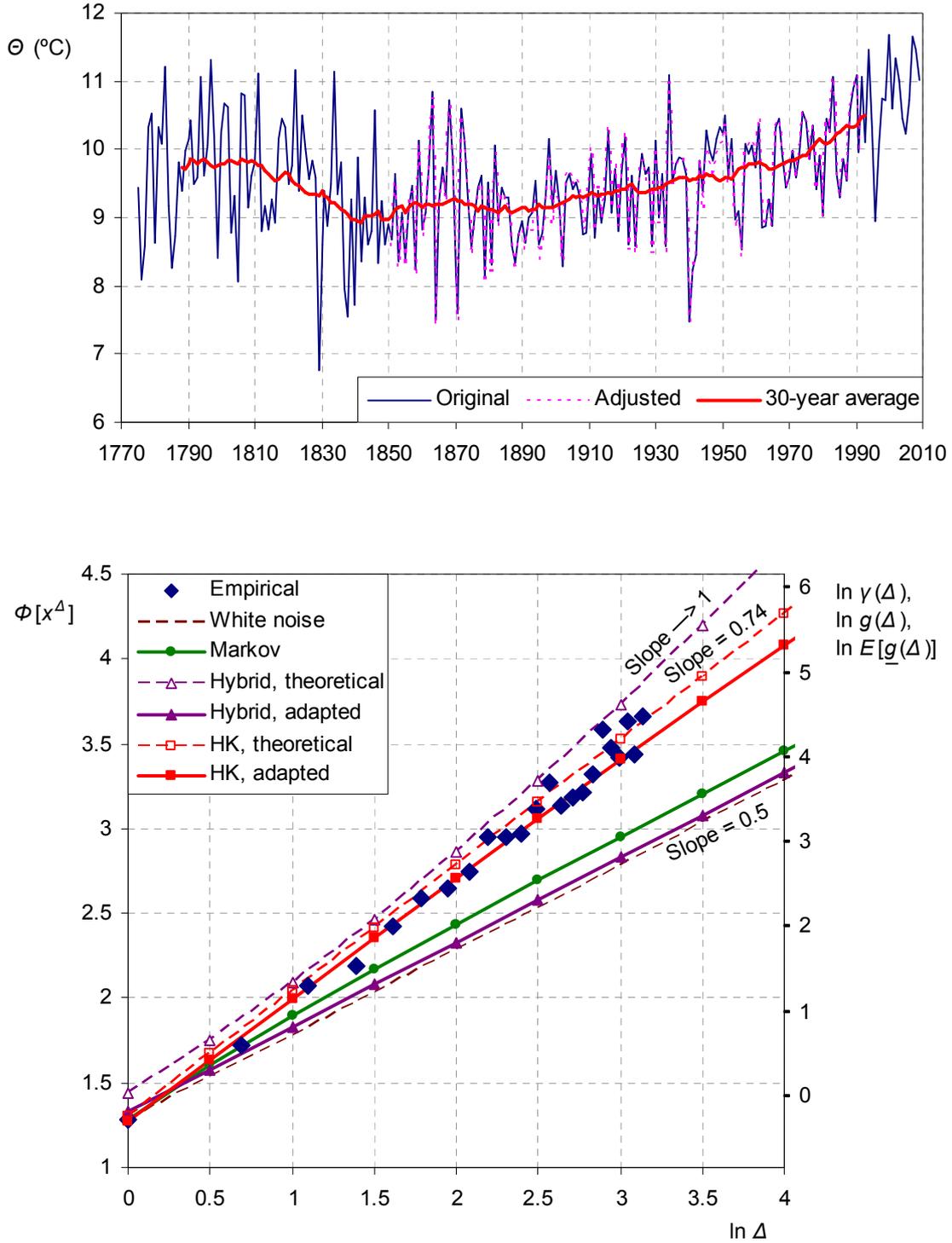


Figure 3 (Upper) The time series of the annual temperature (Θ) in Vienna, Austria; (lower) the entropy $\Phi[x^\Delta]$, along with the corresponding variances at time step Δ , versus $\ln \Delta$, as estimated from the time series using the classical statistical variance estimate $g(\Delta)$ and as predicted by the three extremizing models theoretically ($\gamma(\Delta)$) and after adaptation for bias ($E[g(\Delta)]$).

In classical statistics, where x_t^Δ are independent identically distributed, $g(\Delta)$ is an unbiased estimator of the true variance $\gamma(\Delta)$. However, if consecutive x_t^Δ are dependent, the following equation holds true:

$$E[g(\Delta)] = \frac{T/\Delta}{T/\Delta - 1} \{\text{Var}[\underline{x}_i^{\Delta}] - \text{Var}[\underline{x}^{\Delta}]\} \quad (25)$$

where $\text{Var}[\underline{x}^{\Delta}] = (\Delta/T)^2 \text{Var}[\underline{z}(T)]$ and $\text{Var}[\underline{x}_i^{\Delta}] = \gamma(\Delta)$. This results in $E[g(\Delta)] = c(\Delta, T) \gamma(\Delta)$, where

$$c(\Delta, T) := \frac{1 - (\Delta/T)^2 \gamma(T)/\gamma(\Delta)}{1 - \Delta/T} \quad (26)$$

represents a bias adaptation factor, which equals 1 (no bias) only when $\gamma(\Delta)/\gamma(T) = \Delta/T$, i.e. in white noise. Depending on the dependence structure, the bias adaptation factor may be quite smaller than 1 (although this is often missed in the literature). This is particularly the case in the Hybrid model (Figure 3, lower panel), where $E[g(\Delta)]$ (plotted as a continuous line with triangles) is almost indistinguishable from the true variance $\gamma(\Delta)$ of the white noise instead of being close to the true variance of the Hybrid model (notice that there is a one-to-one association between variance and entropy, shown in figure). This constitutes the *observability problem* of the Hybrid model mentioned above.

In contrast, in the Markovian (OU) model (also shown in Figure 3, lower panel), the bias adaptation is negligible (the two curves before and after adaptation are indistinguishable to each other), i.e. the factor $c(\Delta, T)$ is very close to unity. On the other hand, the OU model implies the lowest *theoretical* entropy for large scale Δ . After adaptation for bias, as shown in Figure 3 (lower panel), the Hybrid model entails even lower *observed* entropy, despite that it entails the highest *theoretical* entropy.

In the HK model the adaptation factor is closer to unity than in the Hybrid model (the theoretical and adapted entropy curves are close to each other) but not as close as in the OU model. In this respect, the HK model represents a good balance of observability on the one hand, expressed by high $c(\Delta, T)$, and, on the other hand, high entropy for large scales, expressed by $\gamma(\Delta)$ or even by the ratio $\gamma(T)/\gamma(\Delta)$ (since $T \gg \Delta$). In the HK model the adaptation factor $c(\Delta, T)$ is maximized for $H = 0.5$ and the ratio $\gamma(T)/\gamma(\Delta)$ is maximized for $H = 1$, in which however $c(\Delta, T) = 0$. The aforesaid balance is expressed by the fact that the product of the two (i.e., $c(\Delta, T) \gamma(T)/\gamma(\Delta)$) is maximized for H between 0 and 1 (specifically, for $H = 1 - (1/2) \ln 2 / \ln(T/\Delta)$), as can be easily verified).

To compare the three models with reality, the proximity of empirical points (corresponding to sample estimates $g(\Delta)$) with the model should be assessed on the basis of the predicted curves $E[g(\Delta)]$ of the models, rather than their theoretical variances $\gamma(\Delta)$. The former are designated in Figure 3 (lower panel) as “adapted” whereas the latter are designated as “theoretical”. It is seen in the figure that the HK model is the only one that agrees with reality. The curves $E[g(\Delta)]$ of the other two models lie far apart from empirical points. It is noted that the consistency of the HK model with reality has been detected in numerous studies using geophysical, biological, technological and even economical time series (e.g. [22-25,26,27,28, 29,30,31] and references therein) and this makes the findings of this study more physically plausible.

6. Conclusion and discussion

It is demonstrated that extremization of entropy production of stochastic representations of natural systems, performed at asymptotic times (zero or infinity) and using simple constraints referred to finite times, at which a process is observed, results in constant derivative of entropy in logarithmic time, which in turn results in Hurst-Kolmogorov processes and long-term persistence. One eminent characteristic of the derivation is its parsimony, in terms of both constraints and physical principles used. Specifically, it was demonstrated that no other notions (e.g. self-organized criticality, scale invariance, etc.) in addition to entropy extremization are necessary to explain the emergence of the Hurst-Kolmogorov behavior. An example with real world data, which is in agreement with a large body of studies that have detected long-term persistence in long time series of a wide range of processes, illustrates the plausibility of the findings.

These findings connect statistical physics (the extremal entropy production concept, in particular) with stochastic representations of natural processes. Extremal entropy production may provide a theoretical background in such stochastic representations, which otherwise are solely data-driven. A theoretical background in stochastic representations is important also for the reason that, as noted by Koutsoyiannis and Montanari [31], merely statistical arguments do not suffice to verify or falsify the presence of long-term persistence in natural processes. The practical consequences of these findings may be significant because stochastic representations of processes are used as modeling tools particularly for the estimation of future uncertainty for planning and design purposes with long time horizons. The emergence of maximum entropy (i.e., maximum uncertainty) for large time horizons, as demonstrated here, should be considered seriously in planning and design studies, because otherwise the uncertainty would be underestimated and the constructions undersized. The relevance of the last point may be even wider, given the current scientific and public interest on long-term predictions.

Acknowledgments I gratefully thank three reviewers for their encouraging comments and constructive suggestions, which helped me to substantially improve the presentation.

References

-
- [1] H. S. Robertson, *Statistical Thermophysics* (Prentice Hall, Englewood Cliffs, NJ, 1993).
 - [2] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge Univ. Press, 728 pp., 2003).
 - [3] P. Atkins, *Four Laws that Drive the Universe* (Oxford Univ. Press, 131 pp., 2007).
 - [4] J. Fischer *et al.*, Report to the CIPM on the implications of changing the definition of the base unit Kelvin (International Committee for Weights and Measures, http://www.bipm.org/wg/CCT/TG-SI/Allowed/Documents/Report_to_CIPM_2.pdf, Retrieved 2010-11-28).
 - [5] A. Porporato *et al.*, Irreversibility and fluctuation theorem in stationary time series, *Phys. Rev. Lett.*, 98, 094101 (2007).

-
- [6] D. Kondepudi and I. Prigogine, *Modern Thermodynamics* (Wiley, Chichester, 1998).
- [7] T. Pujol and J. E. Llebot, *Q. J. R. Meteorol. Soc.*, 125, 79-90 (1999) .
- [8] H. Ozawa *et al.*, *Rev. Geophys.*, 41(4), 1018, doi:10.1029/2002RG000113 (2003).
- [9] I. Prigogine, *Bulletin de la Classe des Sciences, Academie Royale de Belgique*, 31, 600-606 (1945).
- [10] H. Ziegler, in *Progress in Solid Mechanics* ed. by I.N. Sneddon and R. Hill, vol. 4 (North-Holland, Amsterdam, 1963).
- [11] L.M. Martyushev and V.D. Seleznev, *Phys. Rep.*, 426, 1-45 (2006).
- [12] W. Pauli, *Statistical Mechanics* (Dover, New York, 1973).
- [13] I. Müller, in *Entropy*, ed. by A. Greven, G. Keller and G. Warnecke, Ch. 5, 79–105 (Princeton University Press, Princeton, New Jersey, USA, 2003).
- [14] H.E. Hurst, *Trans. Am. Soc. Civil Engrs.*, 116, 776–808 (1951).
- [15] A.N. Kolmogorov, *Dokl. Akad. Nauk URSS*, 26, 115–118 (1940).
- [16] B.B. Mandelbrot and J. W. Van Ness, *SIAM Rev.*, 10 (4), 422-437 (1968).
- [17] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd edn. (McGraw-Hill, New York, 1991).
- [18] C. Tsallis, *Journal of Statistical Physics*, 52(1), 479-487 (1988).
- [19] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*, Springer (2009).
- [20] F. Bouchet *et al.*, *Physica A*, 389, 4389-4405 (2010).
- [21] B. B. Mandelbrot, A fast fractional Gaussian noise generator. *Wat. Resour. Res.* 7 (3), 543–553 (1971).
- [22] D. Koutsoyiannis, *Hydrol. Sci. J.*, 47 (4), 573–595 (2002).
- [23] S. Karlin and H. M. Taylor, *A Second Course in Stochastic Processes* (Academic Press, Boston, 1981).
- [24] D. Koutsoyiannis, *Hydrol. Sci. J.*, 50 (3), 405–426 (2005).
- [25] D. Koutsoyiannis, *Hydrol. Earth Sys. Sci.*, 14, 585–601 (2010).
- [26] T. Lux, *Appl. Econ. Lett.*, 1996, 3, 701–706 (1996).
- [27] E. Koscielny-Bunde *et al.*, *Phys. Rev. Lett.*, 81(3), 729–732 (1998).
- [28] W.E. Leland *et al.*, *IEEE/ACM Trans. Networking*, 2(1), 1-15 (1994).
- [29] A. Bunde *et al.*, *Phys. Rev. Lett.*, 85(17), 3736- 3739 (2000).
- [30] A. Bunde and S. Havlin, *Physica A*, 314, 15-24 (2002).
- [31] D. Koutsoyiannis and A. Montanari, *Water Resour. Res.*, 43 (5), doi:10.1029/2006WR005592 (2007).