

Uncertainty, entropy, scaling and hydrological stochasticity

2. Time dependence of hydrological processes and time scaling

Demetris Koutsoyiannis

Department of Water Resources, Faculty of Civil Engineering, National Technical University of Athens, Heroon Polytechniou 5, GR-157 80 Zographou, Greece (dk@itia.ntua.gr)

Abstract The well-established physical and mathematical principle of maximum entropy (ME), is used to explain the distributional and autocorrelation properties of hydrological processes, including the scaling behaviour both in state and in time. In this context, maximum entropy is interpreted as maximum uncertainty. The conditions used for the maximization of entropy are as simple as possible, i.e. that hydrological processes are non-negative with specified coefficients of variation and lag-one autocorrelation. In the first part of the study, the marginal distributional properties of hydrological processes and the state scaling behaviour were investigated. This second part of the study is devoted to joint distributional properties of hydrological processes. Specifically, it investigates the time dependence structure that may result from the ME principle and shows that the time scaling behaviour (or the Hurst phenomenon) may be obtained by this principle under the additional general condition that all time scales are of equal importance for the application of the ME principle. The omnipresence of the time scaling behaviour in numerous long hydrological time series examined in the literature (one of which is used here as an example), validates the applicability of the ME principle, thus emphasizing the dominance of uncertainty in hydrological processes.

Keywords entropy; fractional Gaussian noise; Hurst phenomenon; hydrological persistence; hydrological prediction; hydrological statistics; long-range dependence; power laws; risk; scaling; uncertainty.

Incertitude, entropie, graduation et propriétés stochastiques hydrologique

2. La dépendance de temps et d'échelle de des processus hydrologiques

Résumé Le principe bien établi à la fois physique et mathématique de l'entropie maximum (ME), est employé pour expliquer les propriétés distributionnelles et d'autocorrélation des processus hydrologiques, y compris le comportement d'échelle dans l'état et le temps. Dans ce contexte, l'entropie maximum est interprétée en tant qu'incertitude maximum. Les conditions utilisées pour la maximisation de l'entropie sont le plus simples possible, c.-à-d. que les processus hydrologiques sont non négatifs avec coefficients de variation et d'autocorrélation fixés. Dans la première pièce de l'étude, les propriétés distributionnelles marginales des processus hydrologiques et le comportement d'échelle d'état ont été étudiés. La présente deuxième partie de l'étude est consacrée aux propriétés distributionnelles communes des processus hydrologiques. Spécifiquement, on étudie la structure de dépendance temporelle qui peut résulter du principe ME et on prouve que le comportement d'échelle de temps (ou le phénomène de Hurst) peut être obtenu par ce principe dans la condition générale additionnelle que toutes les échelles de temps sont d'importance égale pour l'application du principe ME. La omniprésence du comportement d'échelle de temps des nombreuses longues séries chronologiques hydrologiques examinées dans la littérature (dont une est employée ici comme exemple), valide l'applicabilité du principe ME, de ce fait soulignant la dominance de l'incertitude dans les processus hydrologiques.

Mots clefs entropie; bruit fractionnel gaussien; phénomène de Hurst; persistance hydrologique; prévision hydrologique; statistiques hydrologiques; dépendance à longue portée; lois en puissance; risque; mise en échelle; incertitude.

First draft, September 2004; Revision, January 2005

INTRODUCTION

Prediction is very difficult, especially of the future.

Niels Bohr

In the first part of this study, it is shown that the principle of maximum entropy (ME) can explain the statistical distributions of hydrological variables. In this context, maximum entropy is interpreted as maximum uncertainty, given that in the theory of stochastic processes entropy is a measure of uncertainty or ignorance (e.g. Papoulis, 1991, p. 533). The second part of the study is devoted to joint distributional properties of hydrological processes. Specifically, it investigates the time dependence structure that may result from the ME principle and attempts to explain, based on this principle, the time scaling behaviour, which was observed in many long hydrological and meteorological time series.

To define the time scaling property, a stationary stochastic process X_i on discrete time i is considered, from which the time averaged process $X_i^{(k)}$ is formed by averaging k consecutive X_l , i.e.,

$$X_i^{(k)} := \frac{1}{k} \sum_{l=(i-1)k+1}^{ik} X_l \quad (1)$$

where $k = 1, 2, \dots$, denotes time scale (obviously, $X_i^{(1)} \equiv X_i$). The time scaling is expressed by the following equation, relating the distributional properties of the time averaged process $X_i^{(k)}$ at scale k to those of X_i at the basic scale 1:

$$(X_i^{(k)} - \mu) \stackrel{d}{=} k^{H-1} (X_i - \mu) \quad (2)$$

where the symbol $\stackrel{d}{=}$ stands for equality in distribution, μ is the mean of the process and H is a positive constant known as the Hurst coefficient (or exponent) (where $0.5 \leq H < 1$; values $H < 0.5$ are mathematically feasible but physically unrealistic). This time scaling property expresses a behaviour according to which the distribution function of a process does not change with scaling of the time scale by an integer k , except for a multiplicative factor which is a power law of the scaling factor k . A process exhibiting properties (2) can be called a simple scaling stochastic process (SSS process).

This behaviour is usually validated for the second-order properties of the process. In this case, (2) is replaced by the following set of equations (adapted from Koutsoyiannis, 2002a):

$$\begin{aligned} \sigma^{(k)} &= k^{H-1} \sigma, \quad \rho_j^{(k)} = \rho_j \approx H(2H-1) |j|^{2H-2}, \\ \gamma_j^{(k)} &\approx H(2H-1) \sigma^2 k^{2H-2} |j|^{2H-2}, \quad s^{(k)}(\omega) \approx 4(1-H) \sigma^2 k^{2H-2} (2\omega)^{1-2H} \end{aligned} \quad (3)$$

where $\sigma^{(k)}$ and σ denote the standard deviation at scales k and 1 respectively; $\rho_j^{(k)}$ denotes the lag j autocorrelation at scale k , and is independent of the scale (thus equal to ρ_j of scale 1); $\gamma_j^{(k)}$ denotes the lag j autocovariance at scale k (with $\gamma_0^{(k)} \equiv [\sigma^{(k)}]^2$); $s^{(k)}(\omega)$ denotes the power spectrum of the process at scale k and frequency ω . All equations of the set (3) are of power type, all are virtually equivalent to each other and all express one single property, the time scaling. The second and third equations of this set are given in an approximate form (which is good except for $|j| = 0$ and 1) emphasizing the power law behaviour of autocorrelation (the exact relationship for $\rho_j^{(k)}$ is given in equation (20)).

The terms ‘Hurst phenomenon’ (due to Hurst (1951), who first observed this behaviour), ‘Joseph effect’ (due to Mandelbrot, 1977, from the known biblical story) and long-range (or long-term) dependence or persistence (due to the implied high autocorrelations for high lags) have been used as alternative names for the same behaviour. Since its discovery, the scaling behaviour has been identified in several hydrological time series such as (to mention a few of the more recent studies) flows of several rivers such as Nile (Eltahir, 1996; Koutsoyiannis, 2002a), Warta, Poland (Radziejewski & Kundzewicz, 1997), Boeotikos Kephisos, Greece (Koutsoyiannis, 2003a, b), Nemunas, Lithuania (Sakalauskienė, 2003), rivers in Canada (Yue and Gan, 2004); and inflows of Lake Maggiore, Italy (Montanari et al., 1997). It was also identified in other climatological time series including wind power (Haslet & Raftery, 1989); global or point mean temperatures (Bloomfield, 1992; Koscielny-Bunde, et al., 1998; Koutsoyiannis, 2003a; Maraun et al., 2004); indexes of North Atlantic Oscillation (Stephenson et al., 2000); and tree-ring widths, which are indicators of past climate (Koutsoyiannis, 2002a).

Even though the investigation of another time series, additional to those of the previous paragraph, may be redundant and not an important addition to the literature, in order for the

paper to be self-contained a simple real world example is given here. This deals with the longest of the time series already examined in the first part of the study and simultaneously one of the longest instrumental records worldwide, the time series of the mean annual temperature of Geneva with length 228 years. Plots of the time series on scales 1, 5 and 25 are given in Figure 1a. Using classical statistical estimators, the mean of the process is 282.8 K, the standard deviation 0.67 K and the lag one autocorrelation coefficient 0.33. For comparison, a synthetic series with these statistics was generated from the autoregressive process of order 1 (AR(1) or Markov, whose details will be given later) and was plotted in Figure 1b. It is observed that the fluctuations of the processes, especially for the 25-year time scale, are much greater in the real world time series than in the synthetic series, which does not exhibit scaling behaviour. Thus, the significant fluctuations in a time series on large scales gives a first sign of the scaling behaviour.

A clearer depiction of the scaling behaviour can be done by utilizing the first of equations (3), which calls for a double logarithmic plot of standard deviation $\sigma^{(k)}$ versus time scale k . In such a plot, the Hurst behaviour is manifested as a straight line arrangement of points corresponding to different time scales, whose slope is $H - 1$. The plot for the Geneva temperature series for scales $k = 1$ to 20 is depicted in Figure 2. The empirical standard deviations were estimated by two methods, the classical statistical estimator and an estimator appropriate for SSS processes (Koutsoyiannis, 2003a). The second one increases the standard deviation at the basic scale 1 from 0.67 K to 0.71 K. This shows that even the estimation of marginal statistical properties of a hydrological variable cannot be separated from the study of its correlation in time and thus hydrological statistics should be incorporated in a unifying framework of hydrological stochastics. In addition, the theoretical standard deviations $\sigma^{(k)}$ for two models, the SSS and the Markov, have been plotted in Figure 2. Clearly, the empirical estimates of standard deviation depart significantly from the Markov model and are close to the SSS model. The average slope of the arrangement of points of empirical estimates on the logarithmic diagram is -0.2 , which means that the Hurst coefficient is 0.8.

Another depiction of the scaling behaviour is provided in Figure 3, based on the second of equations (3). Figure 3a depicts the empirical autocorrelation functions, again estimated by

two estimators, the classical and an SSS-consistent (Koutsoyiannis, 2003a). The two theoretical autocorrelation functions for the SSS and Markov models are also plotted. Again the diagram shows that empirical plots are close to the SSS model and very far from the Markov model. This confirms the scaling behaviour of the time series. In addition, Figure 3b depicts the lag one autocorrelation coefficient $\rho_1^{(k)}$ versus scale k . In an SSS process $\rho_1^{(k)}$ does not vary with scale ($\rho_1^{(k)} = \rho_1$, as shown in (3)), whereas in a Markov process it is a decreasing function of scale, plotted in Figure 3b. Clearly, the empirical estimates of lag one autocorrelation for scales 1-20 indicate that it is not a decreasing function of time scale (i.e. even for a 20-year time scale it keeps virtually the same value as in the annual time scale); again, this confirms the scaling behaviour.

The omnipresence of the Hurst phenomenon in hydrological (and other geophysical, technological and socioeconomical) time series has intrigued many to call it a mysterious phenomenon, others to “conjure it away” (to quote Klemeš, 1974) and others to propose explanations of the mechanisms that might generate it. Synopsis of older explanations and a couple of two more recent ones are given in Koutsoyiannis (2002a, 2003c, 2005). Generally, these explanations provide conditions under which the scaling behaviour might emerge, but they do not explain why these conditions are so common in nature that make the scaling behaviour be the rule rather than the exception.

In this respect, it is endeavoured in this study to link the scaling behaviour with the ME principle. The idea is that if the ME principle can result in a process with scaling behaviour, then this can be regarded as a sufficient reason for its ubiquity.

THE ENTROPY CONCEPT

A detailed presentation of the entropy concept including definitions and generalizations are given in the first part of the study. Here the elements required to study the dependence structure of a stochastic process that might represent a hydrological process are summarized and also extended to cover the notion of a stochastic process.

For a continuous random variable X taking values x with probability density function $f(x)$ satisfying

$$\int_{-\infty}^{\infty} f(x) dx = 1, \quad (4)$$

the (Shannon or extensive) entropy is by definition (e.g. Papoulis, 1991, p. 559)

$$\varphi := E[-\ln f(X)] = -\int_{-\infty}^{\infty} f(x) \ln f(x) dx \quad (5)$$

If the density $f(x)$ is defined in the interval (a, b) then application of the ME principle results in the uniform distribution in (a, b) . If any of a and b tends to $\pm\infty$, the ME principle cannot be applied unless additional constraints are imposed. The most common ones, which are also used in this paper, are the requirements for finite first and second moments, i.e.

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \mu_1 \equiv \mu \quad (6)$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \mu_2 \quad (7)$$

Application of the ME principle with constraints (4), (6) and (7) results in the normal distribution (e.g. Papoulis, 1991, p. 571; Dowson & Wragg, 1973; Tagliani, 1993, 2002a, b) with mean μ and variance $\sigma^2 = \mu_2 - \mu^2$. The maximized entropy is

$$\varphi = \ln(\sigma \sqrt{2\pi e}) \quad (8)$$

This shows that the entropy of a normally distributed variable depends only on its standard deviation, not on its mean. As discussed in detail in the first part, the fact that hydrological variables are non-negative ($x \geq 0$) implies that the ME distribution is the truncated normal distribution, in which the entropy depends on both μ and σ . Furthermore, if the variation is very high, i.e. $\sigma/\mu > 1$, then the (extensive) ME distribution does not exist. However, a generalization of the extensive entropy concept enables, even in this case, derivation of the ME distribution, which turns out to be power-type (Pareto). In the limiting case $\sigma/\mu \rightarrow 0$ the truncated normal distribution becomes identical to the normal distribution whereas in the limiting case $\sigma/\mu = 1$ both the truncated normal and the Pareto distributions are identical to the exponential distribution.

The case studies of the first part showed that temperature exhibits very low variation, even on small time scales, and thus follows virtually normal distribution on all time scales; in contrast, rainfall and runoff on small time scales (e.g. hourly, daily) exhibit high variation and thus follow Pareto distributions. As time scale becomes larger, the variation becomes lower and the ME distribution tends to be the normal distribution. In this paper it is assumed that time scales are not finer than annual and thus the normal distribution is a good approximation of the ME distribution for all processes of interest; thus no other distributions except normal are examined in this paper. This simplifies calculations as the normal distribution is preserved in aggregate time scales and, besides, the multivariate distributions, either unconditional or conditional, are normal too. The annual scale allows also the convenient hypothesis of stationarity, which was already mentioned; had the basic scale been shorter than annual, such a hypothesis would not be plausible due to sub-annual periodicity of hydrologic phenomena. In some cases, over-annual periodicities have been observed in natural phenomena, due to the 11-year sunspot cycle, or to the periodicity of El Niño and La Niña anomalies (e.g. Tomasino et al., 2004). Obviously, the stationarity assumption is not valid in such cases and the subsequent analysis would not be applicable, unless the process is appropriately transformed to remove periodicity.

With the notational convenience already described in the Introduction, the basic scale $k = 1$ is assumed to be the annual scale, so X_i represents the annual value at discrete time (year) i . Further, it is assumed that time $i = 0$ represents the present, time $i = 1, 2, \dots$ represents the future and time $i = -1, -2, \dots$ represents the past. The stationary process X_i is determined in terms of its n th order joint distribution function $F(\mathbf{x}_n)$ or the density $f(\mathbf{x}_n)$ defined as

$$F(\mathbf{x}_n) := P\{\mathbf{X}_n \leq \mathbf{x}_n\}, \quad f(\mathbf{x}_n) = \frac{\partial^n F(\mathbf{x}_n)}{\partial x_1 \cdots \partial x_n}, \quad \mathbf{x}_n = (x_1, \dots, x_n), \quad \mathbf{X}_n = (X_1, \dots, X_n) \quad (9)$$

where upper- and lower-case symbols denote respectively random variables and their values and $P\{ \}$ denotes probability. The n th order *joint entropy* is defined as (Papoulis, 1991):

$$\varphi_n := E[-\ln f(\mathbf{X}_n)] = -\int_{D_n} f(\mathbf{x}_n) \ln f(\mathbf{x}_n) d\mathbf{x}_n \quad (10)$$

where D_n is the n -dimensional space. φ_n can be interpreted as the uncertainty about the variables x_1, \dots, x_n and equals the information gained when they are observed.

The *conditional entropy of order m* of the process X_i is defined as (Papoulis, 1991)

$$\varphi_{c_m} := E[-\ln f(X_1|X_0, \dots, X_{-m+1})] \quad (11)$$

where $f(X_1|X_0, \dots, X_{-m+1})$ denotes the conditional density of X_1 given X_0, \dots, X_{-m+1} . The limit as m tends to infinity (i.e. the conditional entropy when the entire past is observed) is called simply the *conditional entropy* φ_c , i.e.,

$$\varphi_c := \lim_{m \rightarrow \infty} E[-\ln f(X_1|X_0, \dots, X_{-m+1})] \quad (12)$$

The difference of unconditional and conditional entropies, i.e.,

$$\varphi - \varphi_c =: \psi \quad (13)$$

is a non-negative number that represents the *information gain* when past and present are observed.

In the case that the process X_i is Gaussian, which is of interest here, the joint entropy of order n and the conditional entropy are given as (Papoulis, 1991, pp. 564 & 568):

$$\varphi_n = \ln \sqrt{(2\pi e)^n \delta_n}, \quad \varphi_c := \lim_{m \rightarrow \infty} \varphi_{c_m}, \quad \varphi_{c_m} = \ln \sqrt{2\pi e \delta_{m+1}/\delta_m} \quad (14)$$

where δ_n is the determinant of the covariance matrix c_n defined as

$$c_n := \text{Cov}[\mathbf{X}_n, \mathbf{X}_n] = \begin{bmatrix} \gamma_0 & \cdots & \gamma_{n-1} \\ \vdots & \ddots & \vdots \\ \gamma_{n-1} & \cdots & \gamma_0 \end{bmatrix} \quad (15)$$

Application of the ME principle in a multivariate setting can be done by maximizing either φ_n (for any n) or φ_c . In both cases, maximization of entropy with constraints (4), (6) and (7) results in a process X_i that is Gaussian white noise, i.e. all X_i are independent variables with Gaussian distribution with mean μ and variance $\sigma^2 \equiv \gamma_0 = \mu_2 - \mu^2$ (Papoulis, 1991, p. 576). This, however, is a trivial case. The situation becomes more interesting and closer to the nature of hydrological processes if temporal dependence of the process is assumed. In the simplest case, the dependence can be described by postulating a positive lag one

autocovariance $\gamma_1 \equiv \rho \gamma_0$ (where $\rho \equiv \rho_1$ is the lag one autocorrelation), so that an additional constraint is imposed:

$$E[X_i X_{i+1}] = \int_{-\infty}^{\infty} x_i x_{i+1} f(x_i, x_{i+1}) dx_i dx_{i+1} = \gamma_1 + \mu^2 \quad (16)$$

Under this additional constraint, the ME principle results in a process X_i that is Gaussian and Markovian (Papoulis, 1991, p. 577).

Thus, if nothing is known about the dependence of a process, then the ME principle results in a Gaussian white noise and if the consecutive variables are correlated then the same principle results in a Markovian Gaussian process. However, to obtain these results only the basic time scale, the annual scale, was considered. One may think that the uncertainty should be considered in other time scales, as well, since there is no reason to assume that the annual time scale is unique or more important in nature than other scales are. Thus, the ME principle should be combined with a postulate of importance of all time scales. In this case the application becomes much more difficult. Seeking a formal formulation of this postulate and application of the combined to this ME principle, a heuristic stepwise approach will be followed in the next sections.

Here, it should be noted that if the covariance function γ_j is defined at the basic time scale, then the covariance $\gamma_j^{(k)}$ at any scale k is completely determined in terms of γ_j from the following equation, which is a consequence of (1):

$$\gamma_j^{(k)} = \frac{1}{k} \sum_{i=-k+1}^{k-1} \gamma_{j+k+i} \left(1 - \frac{|i|}{k}\right) \quad (17)$$

BENCHMARK PROCESSES AND INITIAL OBSERVATIONS

Before attempting to apply the ME principle in a multivariate and multiple time scale setting, it is interesting to examine some of the simplest typical stochastic processes. Four such processes are examined in this section and are also used in subsequent sections as sort of “benchmark” processes. These are the already mentioned Markovian (AR(1)) process, the also mentioned SSS process, which, combined with the normality assumption, is identical to

the fractional Gaussian noise (FGN), the moving average process (MA), which implies the least possible autocorrelation, and a process with the highest autocorrelation which will be called the grey noise (GN).

In the Markovian or AR(1) process at the basic scale $k = 1$, the autocorrelation is

$$\rho_j = \rho^{|j|} \quad (18)$$

The unconditional entropy is given by (8) whilst the conditional entropy and the information gain are respectively

$$\varphi_c = \ln[\sigma \sqrt{2 \pi e (1 - \rho^2)}], \quad \psi = -\ln\sqrt{1 - \rho^2} \quad (19)$$

At aggregate scales, the autocorrelation and the entropy expressions become more complicated (the process is no longer AR(1)). Thus, analytical expressions are not convenient and numerical calculations based on equations (17) and (14) are preferable.

In the FGN process, the autocorrelation is independent of time scale, as already expressed in equation (3). The exact expression of autocorrelation (e.g. Koutsoyiannis, 2002a, b) is

$$\rho_j^{(k)} = \rho_j = (1/2) [(|j+1|)^{2H} + (|j-1|)^{2H}] - |j|^{2H} \quad (20)$$

where the Hurst coefficient is determined from $\rho_1 = \rho$. Combining (3) and (8), it is obtained that the unconditional entropy at scale k is

$$\varphi^{(k)} = \ln(k^{H-1} \sigma \sqrt{2\pi e}) \quad (21)$$

The conditional entropy can be estimated numerically from (14). Systematic numerical investigation for H ranging in (0.5, 1) allowed the construction of the following approximation

$$\begin{aligned} \varphi_c^{(k)} &\approx \ln\{k^{H-1} \sigma \sqrt{2 \pi e \sqrt{1 - (2H-1)^2 [0.72(H-1) + 1]}}\} \\ \psi^{(k)} &\approx -\ln\sqrt{1 - (2H-1)^2 [0.72(H-1) + 1]} \end{aligned} \quad (22)$$

which shows that the information gain is independent of the scale k . The approximation error of $\psi^{(k)}$ is smaller than $\pm 0.4\%$.

The next two benchmark processes are chosen in an attempt to establish the least and the

highest, respectively, autocorrelation function that is mathematically feasible and physically reasonable. An autocorrelation function is mathematically feasible if the implied covariance matrix c_n (defined in (15)) is positive definite for any n . Given that the processes considered here are stationary (as already justified earlier) and have a certain lag one autocorrelation ρ , it may be assumed that an autocorrelation function is physically reasonable if it is non-negative and non-increasing. The non-negativity postulate is consistent to the stationarity assumption, given that processes with over-annual periodicities, which might cause negative correlations, are not considered here. The non-increasing postulate will be replaced later by a more restrictive one.

In this sense, the least non-negative autocorrelation function is given by the MA(1) process, in which

$$\rho_0 = 1, \quad \rho_1 = \rho, \quad \rho_j = 0, |j| > 1 \quad (23)$$

This, however, is mathematically feasible when it yields positive determinants δ_n for any $n > 0$, which happens when $\rho \leq 0.5$. For $\rho > 0.5$ the least feasible non-negative autocorrelation function corresponds to a higher order MA(q). For each q , a maximum lag one autocorrelation $\rho^*(q)$ can be determined by maximizing ρ_1 subject to constraints $\delta_n > 0$. For example, numerical application of this method results in $\rho^*(1) = 0.5$, $\rho^*(2) = 0.707$, $\rho^*(3) = 0.809$, $\rho^*(4) = 0.866$ and so on. Thus, given a specific value of ρ , the minimum order q of the required MA(q) model can be determined so that $\rho^*(q-1) \leq \rho \leq \rho^*(q)$. Then, to find the least feasible non-negative autocorrelation, the exact values of the remaining nonzero autocorrelation coefficients (ρ_2, \dots, ρ_q) can be estimated by minimizing the sum $\rho_2 + \dots + \rho_q$, again subject to constraints $\delta_n > 0$. For example, for $\rho = 0.75$, the MA(1) and MA(2) models are infeasible, whereas MA(3) (and beyond) is feasible (because $\rho^*(2) = 0.707 \leq \rho = 0.75 \leq \rho^*(3) = 0.809$). Furthermore, by numerical application of the same method it is obtained that the least feasible autocorrelation function of MA(3) is $\rho_0 = 1$, $\rho_1 = 0.75$, $\rho_2 = 0.470$, $\rho_3 = 0.175$, $\rho_j = 0$ for $|j| > 3$. At aggregate scales $k > 1$, the MA(q) model yields another MA(q') with $q' \leq q$. Entropy calculations at any scale can be done numerically because derivation of analytical equations is too complicated.

The highest feasible non-increasing autocorrelation function is

$$\rho_0 = 1, \quad \rho_j = \rho, \quad |j| > 1 \quad (24)$$

It can be shown that the power spectrum of this process is constant $s(\omega) = 2(1 - \rho)$ at any frequency ω that is a rational number ($\omega = m/n$, where m and n positive integers with $m \leq n/2$). By analogy to the white noise, which has constant $s(\omega) = 2$, this process that has again constant $s(\omega)$, but smaller than 2, has been called the grey noise (GN). Clearly, GN is a non ergodic process since $\lim_{j \rightarrow \infty} \rho_j = \rho \neq 0$ (Papoulis, 1991, p. 432), whereas the other three benchmark processes described earlier are ergodic processes. This means that the statistics of the process cannot be deduced from a time series. Furthermore, if a physical process were GN, a realization of it, i.e. a single time series, could not reveal that the process is GN. A simple simulation experiment shows that a time series generated from this process behaves like white noise. In this respect, GN may not be a useful model to describe some physical phenomenon. However, it is useful as a limiting benchmark case in the context of comparison of different stochastic models.

Figure 4 depicts comparisons of the four benchmark processes in terms of the implied unconditional and conditional entropies as functions of time scale for four different values of the lag one autocorrelation, i.e., $\rho = 0, 0.25, 0.5$ and 0.75 . The time scales used for these comparisons and in all subsequent sections range from 1 to 50. It is noted that the increase of time scale increases dramatically the required calculations as in most cases these are numerical and intensive (e.g. involve computations of determinants). The conditional entropy (which is a limit for $m \rightarrow \infty$ as shown in (14)) was estimated for $m = 50$, which in all cases was found to yield sufficient approximation). With these values ($k_{\max} = 50, m = j_{\max} = 50$), it can be seen from (17) that about 2550 autocovariance terms are required for the calculations. For simplicity and without loss of generality, in all cases the variance at the basic scale was assumed to be unity.

In the case $\rho = 0$, all four benchmark processes become identical to white noise and simultaneously conditional entropy is identical to the unconditional entropy. For the other three values of autocorrelation, the following can be observed in Figure 4:

- For scales $k = 1$ and 2 , all three models result in the same unconditional entropy φ .
- For larger scales, GN corresponds to the maximum unconditional entropy followed by FGN, AR and MA.
- For scales $k = 1$ and 2 , the maximum conditional entropy φ_c is given by AR and MA, respectively.
- The GN model, which gives the highest, among the four models, unconditional entropy, simultaneously gives the least conditional entropy at almost all scales.
- For large scales, the model which gives the highest, among the four models, conditional entropy, is the FGN.
- For large scales (and with the exception of GN), the increase of autocorrelation ρ results in increase of both unconditional and conditional entropy. This may be contrary to the common perception that strong autocorrelation decreases prediction uncertainty (which is directly linked to conditional entropy), which however is correct for small time scales, e.g. 1-2.

In addition, Figure 4 manifests the difficulties in applying the ME principle. Given the antagonistic behaviour observed in different scales and different entropy types, one may think that the application of ME should involve many time scales and both unconditional and conditional entropies. For, it would be unreasonable to accept that the ME principle would result in the GN model, which on the one hand maximizes the unconditional entropy and on the other hand minimizes conditional entropy, i.e. it minimizes uncertainty in the case that the past were observed.

PARAMETRIC MAXIMIZATION OF CONDITIONAL ENTROPY

A general conclusion of the previous section is that the maximization of merely the unconditional entropy would result in the GN model at any time scale, whereas maximization of merely the conditional entropy at scales 1 and 2 result in the AR and MA models, respectively. Although these results were obtained by comparison of four benchmark processes, it can be shown that they are general. It is interesting to find the autocorrelation function that maximizes merely the conditional entropy but at large scales. To this aim and to

avoid the extremely high computational effort that would be required if all autocovariance terms were considered to be unknowns in the optimization, a parametric expression of the autocovariance function is assumed, i.e.,

$$\gamma_j = \gamma_0 (1 + \kappa \beta |j|^\alpha)^{-1/\beta} \quad (25)$$

This generalized Cauchy-type expression with parameters κ , α and β that are positive numbers was studied by Koutsoyiannis (2000) in a simpler Pareto-type form, i.e. with $\alpha = 1$, and later, in some modified forms, by Gneiting and Schlather (2004). When $\beta = 0$, (25) takes the Weibull-type form

$$\gamma_j = \gamma_0 \exp(-\kappa |j|^\alpha) \quad (26)$$

In the case examined, since $\rho_1 = \gamma_1/\gamma_0$ is fixed to ρ , (26) and (25) involve respectively one and two free parameters, because κ can be determined from

$$\kappa = \begin{cases} (\rho^{-\beta} - 1) / \beta, & \beta > 0 \\ -\ln \rho, & \beta = 0 \end{cases} \quad (27)$$

It can be seen that all four benchmark processes of the previous section can be obtained from the generalized Cauchy-type expression. Specifically, the AR(1), MA(1) and GN processes are obtained for $\beta = 0$ and respectively $\alpha = 1$, $\alpha \rightarrow \infty$ and $\alpha \rightarrow 0$. The FGN model is obtained for $\alpha/\beta = 2 - 2H$ and $\alpha \rightarrow \infty$. Furthermore, (25) yields a rich family of autocorrelation functions as it can be seen in Koutsoyiannis (2000) and Gneiting and Schlather (2004). Therefore, (25) is a proper parametric model to use in conditional entropy maximization.

For a given variance γ_0 , lag one autocorrelation ρ and scale k , the autocorrelation function $\rho_j^{(k)}$ and, consequently, the conditional entropy $\varphi_c^{(k)}$ are functions of parameters α and β . Therefore, the problem is to determine the values of α and β that maximize $\varphi_c^{(k)}$. This can be done only numerically. Examples of the optimized autocorrelation functions for $\gamma_0 = 1$, $\rho = 0.5$, and $k = 1, 2, 4, 8, 16, 32$ and 50 are shown graphically in Figure 5. In all cases, the optimized β is zero, which corresponds to Weibull-type autocorrelation. For $k = 1$ and 2 , as expected, the optimal autocorrelation functions are those of AR(1) and MA(1) processes, respectively. For larger k , the optimal autocorrelation function gradually raises from the

MA(1) case approaching and then surpassing the FGN case (Figure 5a). Similar behaviour is observed in the plots of the corresponding unconditional and conditional entropy and information gain (Figure 5b, c, d, respectively). The general conclusion of this experiment is that the maximization of conditional entropy at large time scales results in processes that are Hurst-like (with fat tails of autocorrelation functions) but not precisely scaling.

Given that in all cases depicted in Figure 5 the optimized β was zero, i.e. the resulting autocorrelation function was Weibull-type with one free parameter α as in (26), it is interesting to investigate the variation of entropies and information gain with the variation of α or equivalently, with the variation of the ratio of lag two to lag one autocorrelation which will be denoted by λ . From (26) it is obtained that $\lambda = \rho^{1-2\alpha}$, which shows that if $\rho_1 \equiv \rho$ and $\lambda = \rho_2/\rho_1$ are fixed, then α and the complete autocorrelation function are fixed, too.

This investigation is shown graphically in Figure 6 for $\rho = 0.50$ and scales $k = 1, 2$ and 50. As already mentioned, for scales $k = 1$ and 2, the unconditional entropies are constant, independent of λ . As shown in Figure 6a, for scale 50, the unconditional entropy is an increasing function of λ and thus it takes its maximum value for $\lambda = 1$ (values $\lambda > 1$ have been excluded as physically unreasonable). The conditional entropy for scale k has a maximum at a certain λ_k which is different for different scales, i.e. $\lambda_1 = 0.5$ (corresponding to the AR(1) case), $\lambda_2 = 0$ (corresponding to the MA(1) case) and $\lambda_{50} = 0.85$ (corresponding to the maximized autocorrelation function plotted in Figure 5a). Figure 6b shows the information gain as a function of λ . For scale $k = 1$, this function is concave with its minimum at $\lambda_1 = 0.5$ (corresponding to the AR(1) case). For larger scales, it is an increasing function of λ attaining its maximum at $\lambda = 1$. For some value λ^*_k (where $\lambda^*_2 = 0.77$ and $\lambda^*_{50} = 0.87$) the information gain for scale k becomes equal to that for scale 1, i.e. $\psi^{(k)} = \psi^{(1)}$, while beyond λ^*_k , $\psi^{(k)} > \psi^{(1)}$. Here one may notice that the case $\psi^{(k)} > \psi^{(1)}$ may be not physically reasonable. For, it is not reasonable to assume that observing the present and past will lead to information gain for the next k time steps (i.e. $\psi^{(k)}$) that is greater than the information gain for the next single time step (i.e. $\psi^{(1)}$). Thus, in the optimization of ME, all values $\lambda > \lambda^*_2$ should be excluded as being physically unreasonable.

Extending this thinking, the postulate for a non-increasing autocorrelation function may be

replaced by the postulate for a non-increasing information gain function, assuming that the latter postulate assures a physically reasonable autocorrelation structure. It is likely that this postulate may be produced from other considerations. As an attempt to this aim and also as a further clarification, an additional demonstration is provided with the help of the sketch in Figure 7. For simplicity, the demonstration is done for scales 1 and 2 (the generalization is direct). At these scales, the process of interest is described by the variables $X_i^{(1)}$ and $X_j^{(2)}$ with standard deviations $\sigma^{(1)}$ and $\sigma^{(2)}$, respectively, where i and j denote discrete time and (due to (1), $X_j^{(2)} = (1/2) (X_{2j-1}^{(1)} + X_{2j}^{(1)})$; e.g. $X_2^{(2)} = (1/2) (X_1^{(1)} + X_2^{(1)})$). Let $\sigma_i^{(1)}$ and $\sigma_j^{(2)}$ denote the conditional standard deviations (at times i and j and scales 1 and 2, respectively) in the case that the process has been observed at present and past times ($i, j = 0, -1, -2, \dots$). Clearly, $\sigma_i^{(1)}$ is not the same for all i , but it is zero for $i \leq 0$ and an increasing function of i for $i > 0$. As i tends to infinity, $\sigma_i^{(1)}$ tends to the unconditional $\sigma^{(1)}$, so that $\sigma_i^{(1)}/\sigma^{(1)}$ tends to unity (see Figure 7). $\sigma_j^{(2)}/\sigma^{(2)}$ behaves similarly. Intuitively, it is reasonable to accept that the aggregation of scale results in a more flat shape of σ_j/σ , i.e. that the plot of $\sigma_j^{(2)}/\sigma^{(2)}$ will be flatter than that of $\sigma_i^{(1)}/\sigma^{(1)}$. In this respect, it seems reasonable to accept that the first coordinate at scale 2 would be in between the first and second coordinate at scale 1, i.e. $\sigma_1^{(1)}/\sigma^{(1)} \leq \sigma_1^{(2)}/\sigma^{(2)} \leq \sigma_2^{(1)}/\sigma^{(1)}$. If the logarithms of the first of these inequalities are taken, it is obtained that $\ln \sigma_1^{(1)} - \ln \sigma^{(1)} \leq \ln \sigma_1^{(2)} - \ln \sigma^{(2)}$. Observing that $\ln \sigma^{(1)}$ is proportional to the unconditional entropy φ (equation (8)) and $\ln \sigma_1^{(1)}$ is proportional to the conditional entropy φ_c (equations (8) and (12) along with the normality of $f(X_1|X_0, \dots, X_{-m+1})$), and using (13), it is obtained that $\psi^{(1)} \geq \psi^{(2)}$. This is the non-increasing property of ψ for scales 1 and 2, extendable to other scales by the same thinking. Obviously, all this discussion is not a mathematical proof, nor a formulation of a mathematical principle (as is, for instance, the requirement for a positive definite autocovariance matrix discussed earlier); in contrast, violation of $\psi^{(1)} \geq \psi^{(2)}$ is mathematically feasible. Rather, this discussion is a demonstration based on physical intuition; therefore, the postulate for non-increasing ψ is put as a postulate for a physically reasonable autocorrelation function.

Coming back to Figure 5d, it may now be seen that the autocorrelation functions optimized so far (those for scales ≥ 16) do not satisfy this postulate for physical reasonability, as they

result in information gain that for small scales is an increasing function of scale.

CONSTRAINED PARAMETRIC ENTROPY MAXIMIZATION

After the discussion of the previous section, the maximization of entropy, conditional or unconditional, should be done with the additional constraint that the information gain should be a non-increasing function of time scale. A systematic numerical maximization experiment, similar to that of the previous section but with this additional constraint, is depicted in Figure 8. For small time scales, i.e. $k \leq 8$, the results are the same as in Figure 5 but for larger scales are different. Clearly, Figure 8a shows that, as the time scale used for maximization increases, the autocorrelation function approaches that of FGN model, now without surpassing it. Again, for all scales the autocorrelation functions that maximize conditional entropy correspond to $\beta = 0$. Intuitively, the FGN model can be obtained as the limit of the ME autocorrelation as the time scale of interest tends to infinity.

It is reminded that the maximization of unconditional entropy leads to the GN model for any time scale $k > 2$. However, if the constraint of non-increasing information gain is imposed, then the resulting model is the FGN. This is depicted in Figure 9 for $\rho = 0.50$. If the optimization is done for time scales $k = 4$ or $k = 8$, the same ME autocorrelation is obtained for both cases, which corresponds to $\beta = 0$ and is close to that of the FGN model. For all larger time scales, a single autocorrelation function is obtained, which corresponds to $\beta > 0$ and is practically indistinguishable from that of the FGN model. Similar results were found for other values of ρ ; Figure 10 shows the results of another experiment for $\rho = 0.75$.

In conclusion, if the constraint of non-increasing information gain is imposed and if the time scale of interest is large enough, the maximization of either the unconditional or conditional entropy leads to the FGN model. The convergence of the ME autocorrelation to the FGN model is faster in the case of maximization of the unconditional entropy.

CONSTRAINED NONPARAMETRIC ENTROPY MAXIMIZATION

All previous experiments were based on the parametric autocorrelation function (25), which was optimized for a single time scale of interest. As a final step of this investigation, a non-

parametric approach is elaborated, according to which all autocorrelation coefficients ρ_j for $j = 1$ to 50 are considered as unknowns. To avoid an unnecessarily high number of control variables, the parametric autocorrelation function (25) is kept for lags $j > 50$ and its parameters α and β are additional unknowns, so that the maximization includes $49 + 2 = 51$ control variables in total. In addition, instead of using one time scale for the maximization, a range of time scales $k = 1$ to k_{\max} is considered and the average unconditional entropy over all scales k of this range is maximized. The non-increasing information gain constraint (i.e. $\psi^{(k)} \leq \psi^{(k-1)}$ for any k) is also used. The optimization is done numerically using a widespread solver by Frontline Systems (www.solver.com) combining classical and evolutionary optimization techniques.

The results of this nonparametric optimization experiment for $\rho = 0.5$ and maximum scales $k_{\max} = 4, 8, 16, 32$ and 50 is shown in Figure 11. It is observed that if the optimization is done for scales 1 to k_{\max} , then the resulting ME model is practically indistinguishable from FGN for scales 1 to $k_{\max}/2$ and as the maximum scale of interest k_{\max} increases, the ME model tends to be identical to the FGN model. It is not reasonable to designate a special importance to some specific finite maximum scale k_{\max} , so eventually the optimization should be done for $k_{\max} \rightarrow \infty$. Although this case cannot be elaborated with numerical calculations and an analytical solution may be hard to establish, it may be conjectured, based on the numerical results shown in Figure 11, that for $k_{\max} \rightarrow \infty$ the ME model will be precisely the FGN model.

SYNOPSIS, CONCLUSION AND DISCUSSION

In the first part of this study, the principle of maximum entropy (ME), representing maximum uncertainty, was used to explain the statistical distributions met in hydrological variables. The only assumptions used for the maximization of entropy are that a hydrological variable is non-negative and possesses a certain variability, expressed by the coefficient of variation. The results of the analysis of the first part can be summarized as follows:

- (a) Maximum entropy + Low variation \longrightarrow Exponential-type (truncated normal) distribution
- (b) Maximum entropy + High variation \longrightarrow Power-type (Pareto) distribution

- (c) Maximum entropy + High variation + High return periods \longrightarrow State scaling (as an approximation)

This second part of the study is devoted to the joint distributional properties, specifically to the time dependence structure of hydrological processes, and the potential explanation of the time scaling via the ME principle. The analysis of this part considers time scales not finer than annual that are characterized by variation so low that the truncated normal distribution is virtually identical to the normal distribution. The results can be summarized as follows:

- (d) Maximum entropy + Low variation \longrightarrow Normal distribution + Time independence
 (e) Maximum entropy + Low variation + Time dependence + Dominance of a single time scale \longrightarrow Normal distribution + Markovian (short-range) time dependence
 (f) Maximum entropy + Low variation + Time dependence + Equal importance of time scales \longrightarrow Normal distribution + Time scaling (long-range dependence)

The assumption used in the second part (except in case (d)), in addition to those of the first part, is that a process exhibits time dependence expressed by a specified lag one autocorrelation coefficient which is a positive number. The maximization of entropy was done in terms of determining an unknown autocorrelation function, which should be (1) mathematically feasible, i.e. result in positive definite autocovariance matrices, and (2) physically reasonable, i.e. be non-negative and result in information gain that is a non-increasing function of time scale.

Eventually, it is hard to imagine a hydrological process without any time dependence, at least at a fine time scale, which, obviously, will be inherited to coarser scales, too. In addition, it is difficult to find justifiable reasons that make a single time scale dominant. In this respect, among the above listed cases (d)-(f), the last one (f) appears to be physically more realistic. This justifies the omnipresence of the Hurst behaviour. It should be noted that in short time series (e.g. shorter than 100 years) the Hurst behaviour may be not visible because classical statistics tend to hide it (Koutsoyiannis, 2003a).

The study of maximum entropy in this paper involves three quantities: the unconditional entropy, i.e. the uncertainty when nothing is observed about the process, the conditional entropy, i.e. the uncertainty when the past and present states of a process are observed, and

the information gain, i.e. the difference of conditional and unconditional entropies, which is always non-negative. As time scale increases, both the unconditional and conditional entropy decrease. In a scaling process, they decrease at equal rates, so that the information gain is constant, independent of the time scale. In non-scaling processes such as the Markovian, the information gain decreases rapidly with the increase of time scale. However, the constant information gain of the scaling case does not favour prediction at large scales. In fact, the constant information gain is closely related to the fact that both unconditional and conditional entropies are much higher in a scaling process than in a Markovian process or in a process of independent variables. Thus, the scaling behaviour, despite its high autocorrelation function, implies greater prediction uncertainty when large time scales are considered. It is generally believed that an accurate prediction of a process state at a long time horizon is impossible, but a prediction of the mean future conditions for this horizon is much easier. For example, it is believed that the prediction of weather in the next 30 years is impossible, but the prediction of the mean weather of the next 30 years, i.e. the climate, can be accurate. The studied entropy properties of the time scaling behaviour show that this may be wrong and that predictions of the future are difficult (as expressed in the quotation in the beginning of the Introduction) both on small and large time scales.

Several issues related to the application of the ME principle to hydrological processes should be addressed with further research. A first important issue is the study of the consequences of the ME principle on small time scales, in which the variability of a hydrological process can be high and the sub-annual periodicity makes the stationarity assumption invalid. Another related issue, also requiring further research, is the application of the ME principle in the case of over-annual periodicities, whence again the stationarity assumption is not applicable. A further important issue is the explicit use of the ME principle to stochastic generation algorithms, especially in a multivariate or even multidimensional framework. Finally, the general conclusion of this study, i.e. the dominance of the principle of maximum entropy in hydrological (and meteorological) processes must be exploited further to make a more consistent assessment of predictability of these processes and to widen our estimates of nature's uncertainty.

Acknowledgements The temperature data of Geneva are made available online by the National Climatic Data Center of the National Environmental Satellite, Data, and Information Service (<http://www1.ncdc.noaa.gov/pub/data/paleo/climate1500ad>). Thanks are due to the reviewers Bellie Sivakumar and Francesco Laio for their comments that helped improve the presentation and located some critical points that needed more careful consideration.

REFERENCES

- Bloomfield, P. (1992), Trends in global temperature, *Climatic Change* 21, 1–16.
- Dowson, D. C. & Wragg, A. (1973) Maximum-entropy distributions having prescribed first and second moments, *IEEE Transactions on Information Theory*, IT-19, 689-693.
- Eltahir, E. A. B. (1996) El Niño and the natural variability in the flow of the Nile River, *Wat. Resour. Res.* 32(1), 131–137.
- Gneiting, T., and Schlather, M. (2004) Stochastic models that separate fractal dimension and the Hurst effect, *Society for Industrial and Applied Mathematics Review*, 46(2), 269-282.
- Haslett, J. and Raftery, A. E. (1989) Space–time modelling with long-memory dependence: assessing Ireland’s wind power resource, *Appl. Statist.* 38(1), 1–50.
- Hurst, H. E. (1951) Long term storage capacities of reservoirs. *Trans. ASCE* 116, 776–808.
- Klemeš, V. (1974) The Hurst phenomenon: a puzzle? *Wat. Resour. Res.* 10(4), 675–688.
- Koscielny-Bunde, E., Bunde, A., Havlin, S., Roman, H. E., Goldreich, Y., and Schellnhuber, H.-J. (1998) Indication of a Universal Persistence Law Governing Atmospheric Variability, *Phys. Rev.* 81, 729–732.
- Koutsoyiannis, D. (2000) A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Water Resources Research*, 36(6), 1519-1533.
- Koutsoyiannis, D. (2002a) The Hurst phenomenon and fractional Gaussian noise made easy. *Hydrol. Sci. J.* 47(4), 573–595.
- Koutsoyiannis (2002b) Internal report: <http://www.itia.ntua.gr/getfile/511/2/2002HSJHurstSuppl.pdf>.
- Koutsoyiannis, D. (2003a) Climate change, the Hurst phenomenon, and hydrological statistics, *Hydrol. Sci. J.* 48(1), 3-24.
- Koutsoyiannis, D. (2003b) Hydrological statistics for engineering design in a varying climate, *EGS-AGU-EUG Joint Assembly, Geophysical Research Abstracts, Vol. 5*, Nice, April 2003, European Geophysical Society, American Geophysical Union (<http://www.itia.ntua.gr/g/docinfo/565/>).
- Koutsoyiannis, D. (2003c) A toy model of climatic variability with scaling behaviour, *Hydrofractals '03, An international conference on fractals in hydrosociences*, Monte Verita, Ascona, Switzerland, August 2003, ETH Zurich, MIT, Université Pierre et Marie Curie (<http://www.itia.ntua.gr/g/docinfo/585/>).
- Koutsoyiannis, D. (2005) Reliability concepts in reservoir design, *The Encyclopedia of Water*, edited by J. H. Lehr, Wiley, New York (in press).
- Mandelbrot, B. B. (1977) *The Fractal Geometry of Nature*, Freeman, New York, USA.
- Maraun, D., Rust, H.W., & Timmer, J. (2004) Tempting long-memory – on the interpretation of DFA results, *Nonlinear Processes in Geophysics*, 11, 495–503.
- Montanari, A., Rosso, R. and Taqqu, M. S. (1997) Fractionally differenced ARIMA models applied to hydrologic time series. *Wat.*

- Resour. Res.* 33(5), 1035–1044.
- Papoulis, A. (1991) *Probability, Random Variables, and Stochastic Processes* (third edn), 57. McGraw-Hill, New York, USA.
- Radziejewski, M. and Kundzewicz, Z. W. (1997) Fractal analysis of flow of the river Warta, *J. Hydrol.* 200, 280–294.
- Sakalauskienė, G. (2003) The Hurst Phenomenon in Hydrology, *Environmental Research, Engineering and Management*, 3(25), 16-20.
- Stephenson, D. B., Pavan, V. and Bojariu, R. (2000) Is the North Atlantic Oscillation a random walk? *Int. J. Clim.* 20, 1–18.
- Tagliani, A. (1993) On the application of maximum entropy to the moments problem, *J. Math. Phys.*, 34(1), 326-336.
- Tagliani, A. (2002a) Entropy estimate of probability densities having assigned moments: Stieltjes case, *Applied Mathematics and Computation*, 130, 201-211.
- Tagliani, A. (2002b) Maximum entropy solutions and moment problem in unbounded domains, *Applied Mathematics Letters*, 16, 519-524.
- Tomasino, M., Zanchettin, D. & Traverso, P. (2004) Long-range forecasts of River Po discharges based on predictable solar activity and a fuzzy neural network model, *Hydrological Sciences Journal*, 49(4), 673-684.
- Yue, S. and T. Y. Gan, 2004, Simple scaling properties of Canadian annual average streamflow, *Advances in Water Resources*, 27, 481-495.

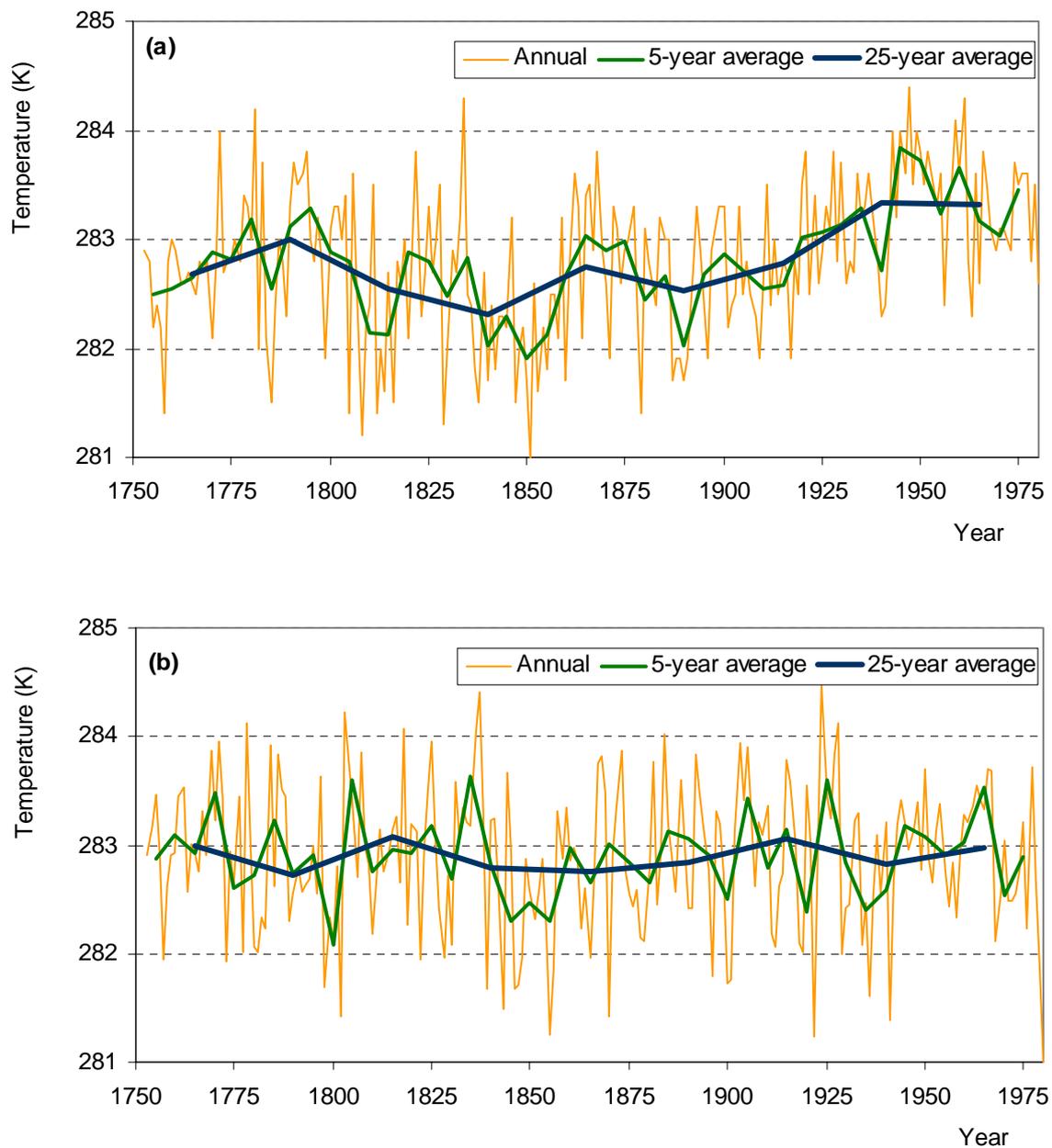


Figure 1 Plot of (a) the mean annual temperature of Geneva and, for comparison, (b) a synthetic series generated from a Markovian process with statistics same with those of the mean annual temperature of Geneva.

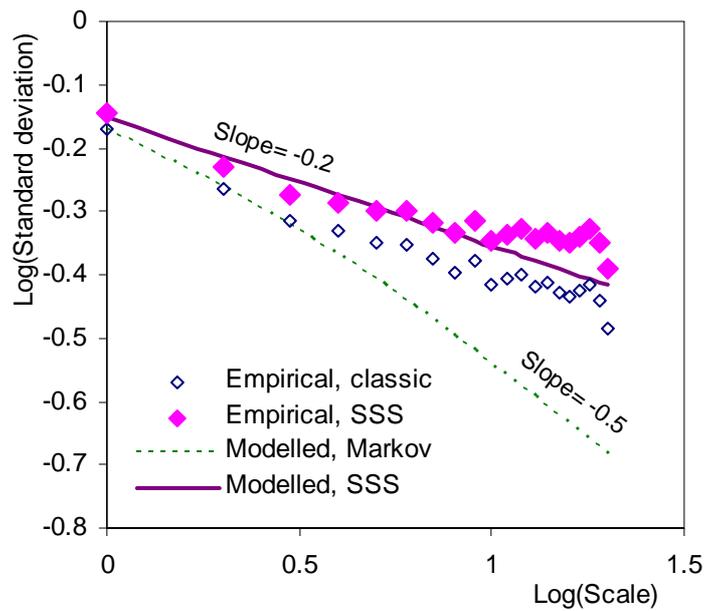


Figure 2. Logarithmic plot of standard deviation versus scale for the time series of the annual temperature of Geneva.

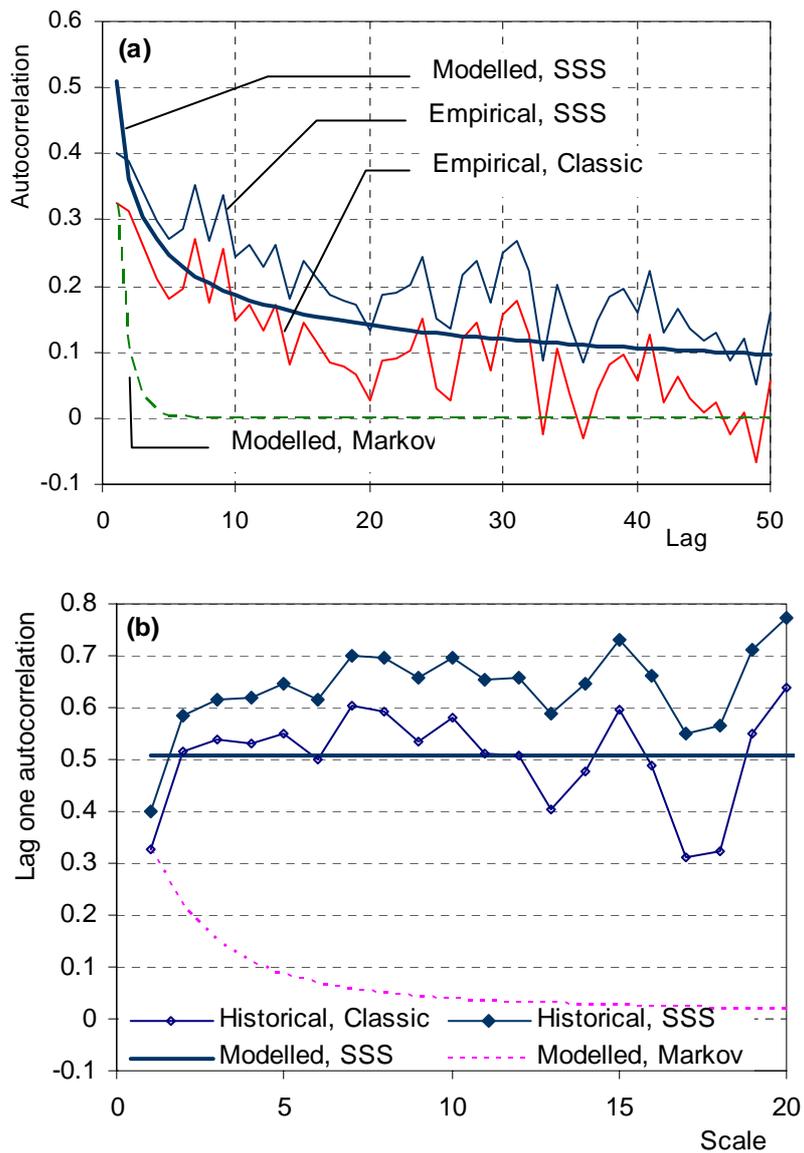


Figure 3 (a) Autocorrelation coefficient vs. lag and (b) lag one autocorrelation coefficient vs. time scale for the time series of the annual temperature of Geneva.

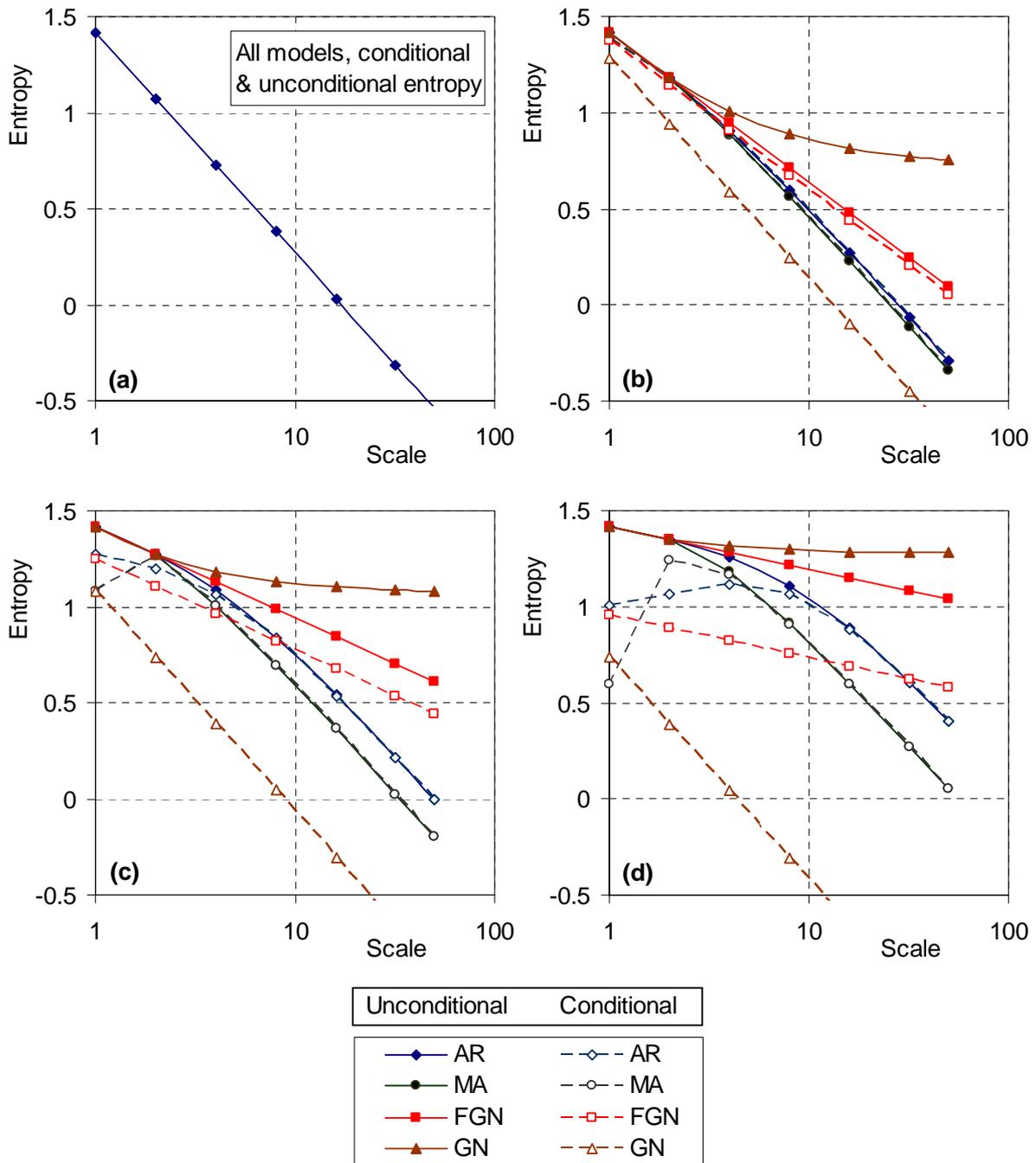


Figure 4 Comparison of the four benchmark models in terms of the implied unconditional and conditional entropies as functions of time scale assuming lag one autocorrelation (a) $\rho = 0$; (b) $\rho = 0.25$; (c) $\rho = 0.5$; and (d) $\rho = 0.75$. The autoregressive (AR) process is AR(1) whereas the moving average (MA) process is MA(1) in (b) and (c) and MA(3) in (d).

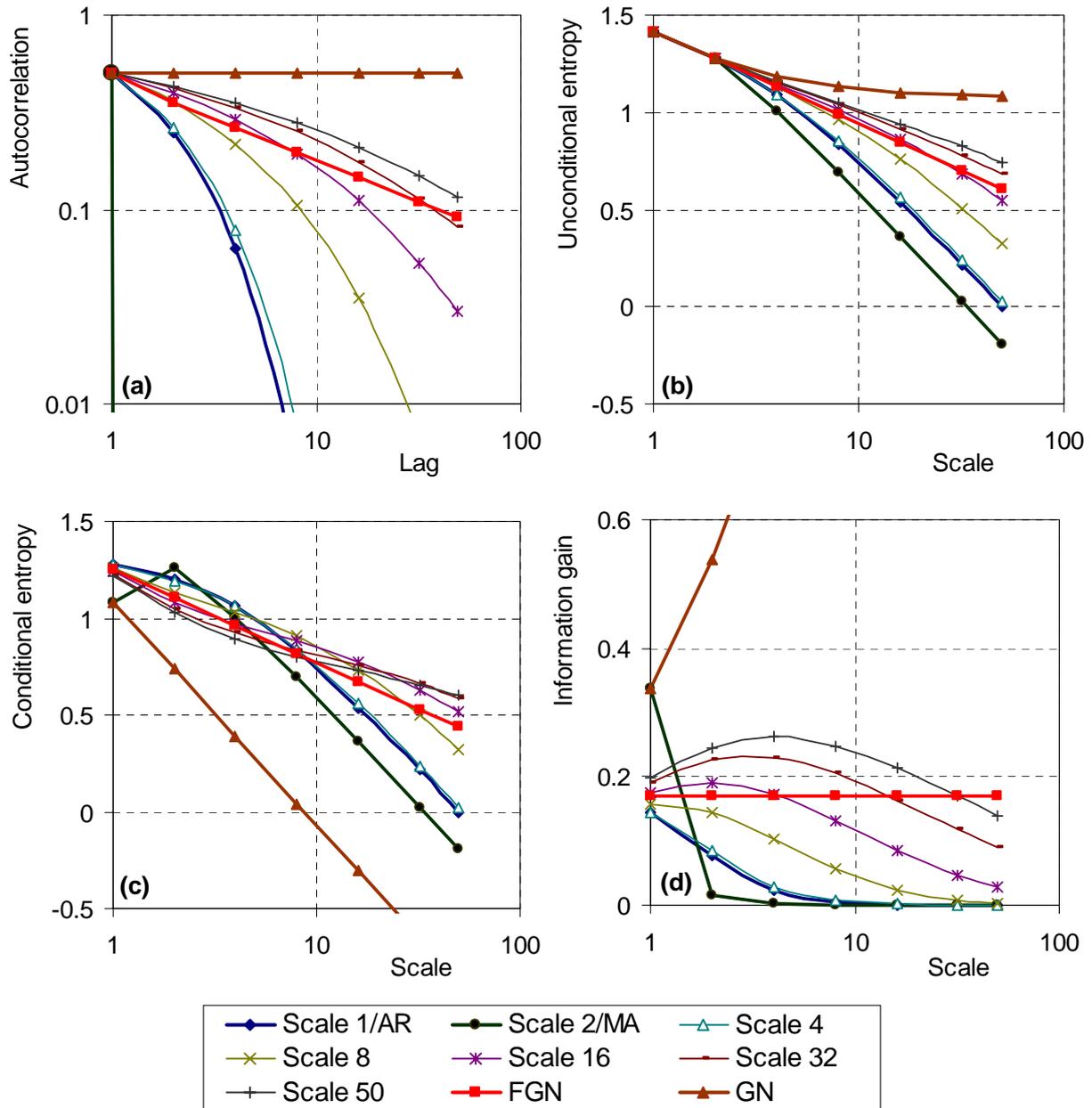


Figure 5 (a) Autocorrelation functions that maximize conditional entropy at each of the indicated time scales ($k = 1, 2, 4, 8, 16, 32$ and 50) assuming lag one autocorrelation $\rho = 0.50$. (b-d) Resulting unconditional and conditional entropy and information gain, respectively, as function of time scale k . In all panels, the relevant plots of benchmark models are also given for comparison.

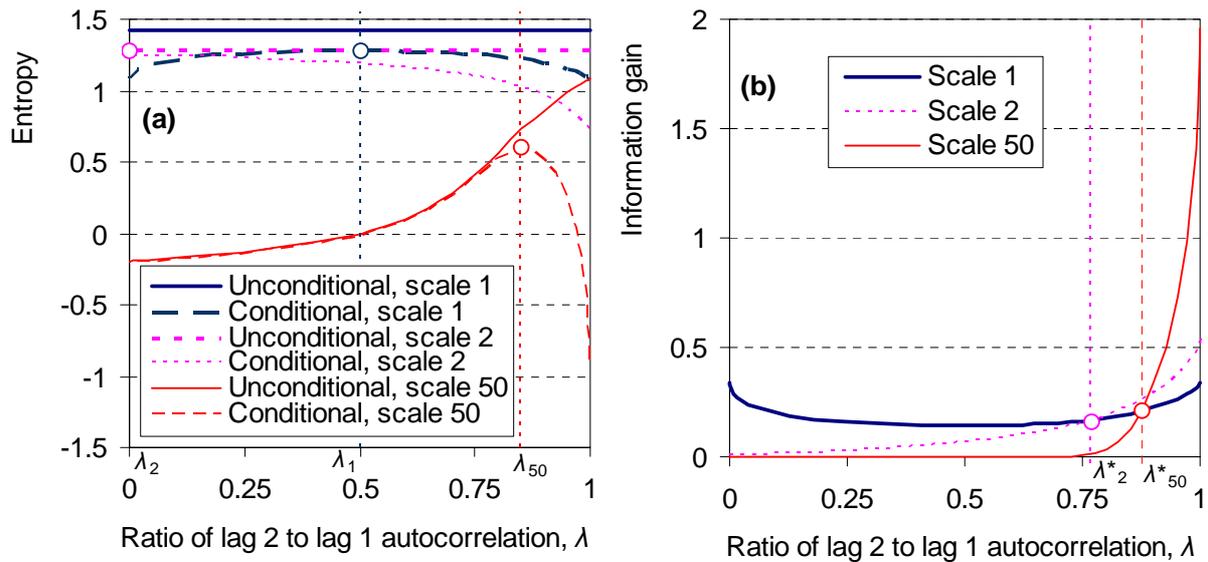


Figure 6 Variation of conditional and unconditional entropy and information gain for scales $k = 1, 2$ and 50 assuming $\rho = 0.5$ and Weibull-type autocorrelation ($\beta = 0$) with varying parameter α , versus the ratio $\lambda := \rho_2 / \rho_1$, which is determined from α .

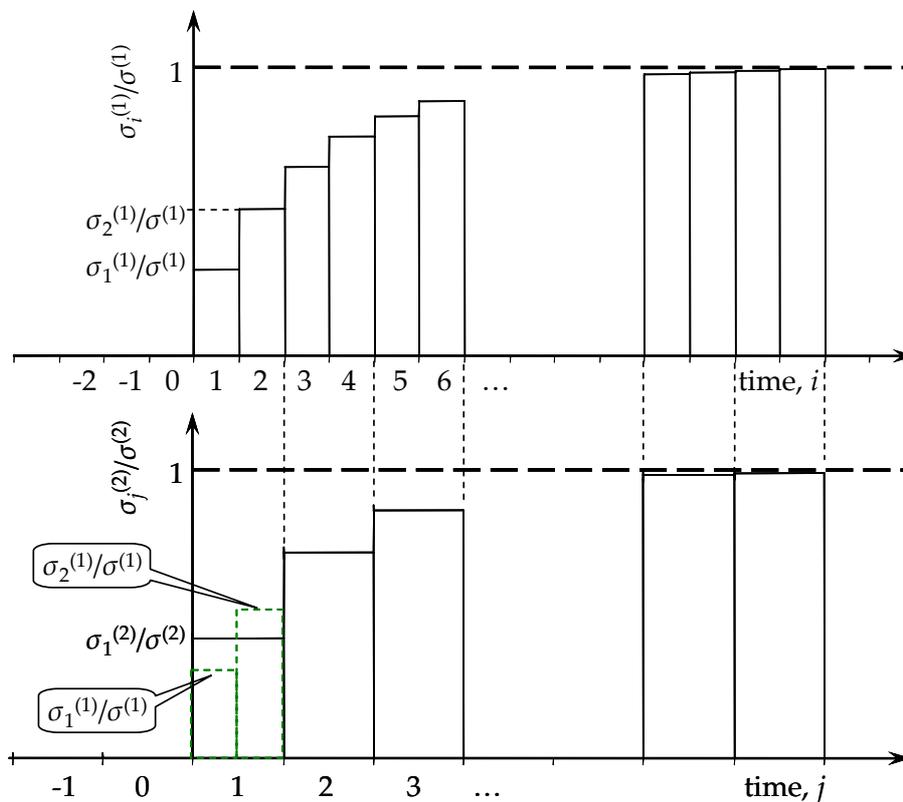


Figure 7 Explanation sketch for the postulate of non-increasing information gain.

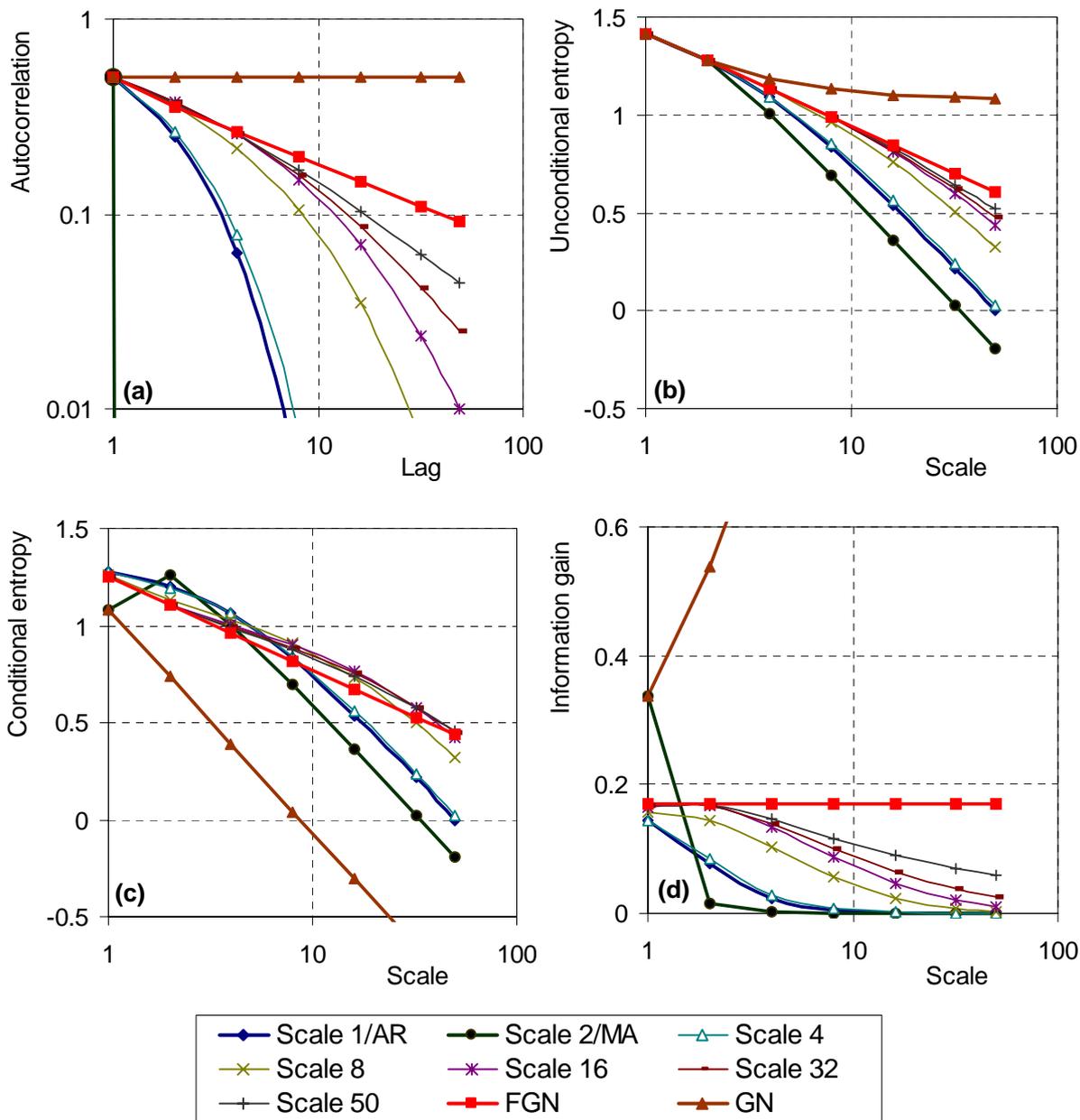


Figure 8 (a) Autocorrelation functions that maximize conditional entropy at each of the indicated time scales assuming lag one autocorrelation $\rho = 0.50$ and constraining information gain to be non-increasing function of time scale. (b-d) Resulting unconditional and conditional entropy and information gain, respectively, as function of time scale. In all panels, the relevant plots of benchmark models are also given for comparison.

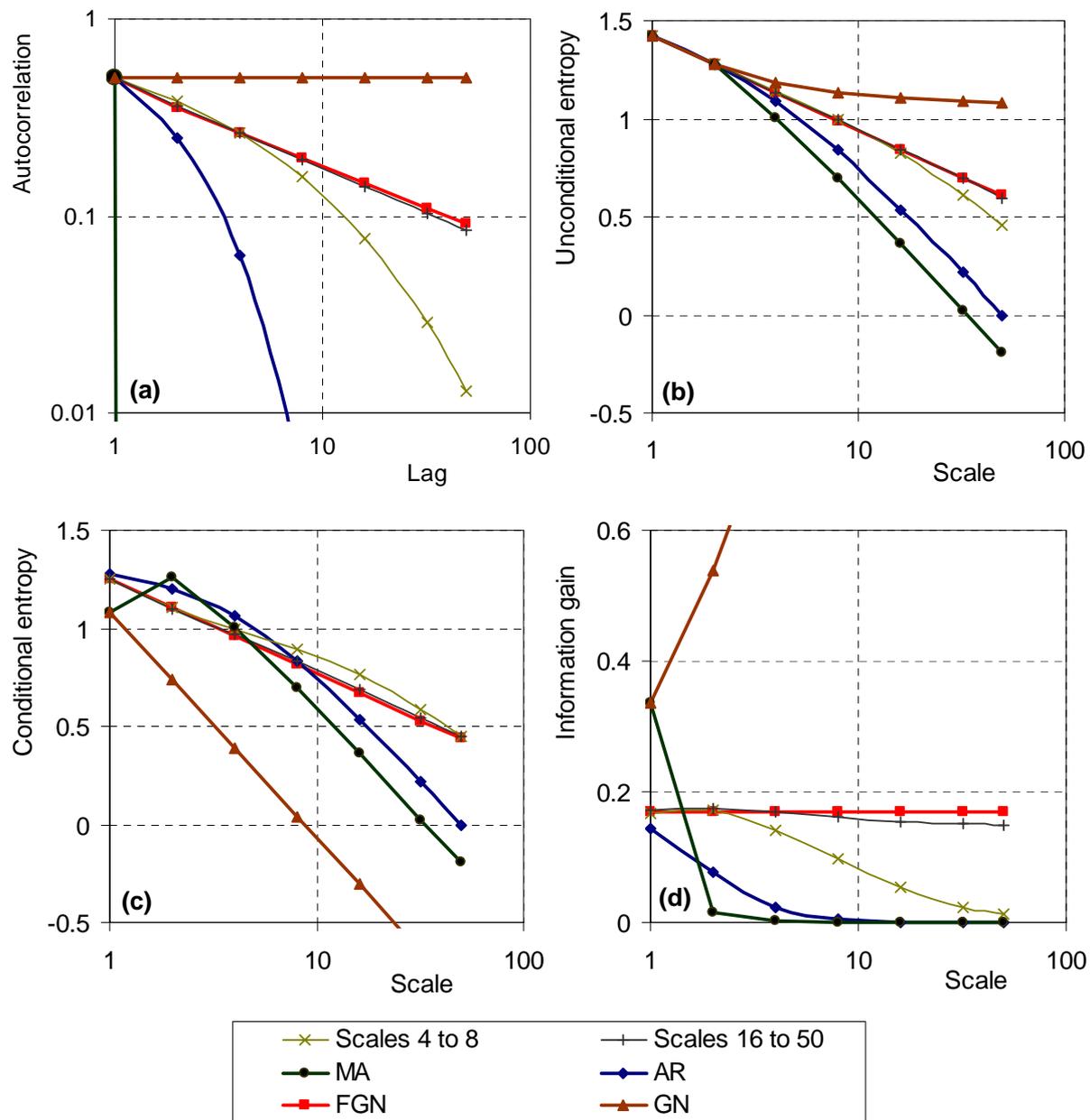


Figure 9 Similar to Figure 8 but for maximizing unconditional entropy.

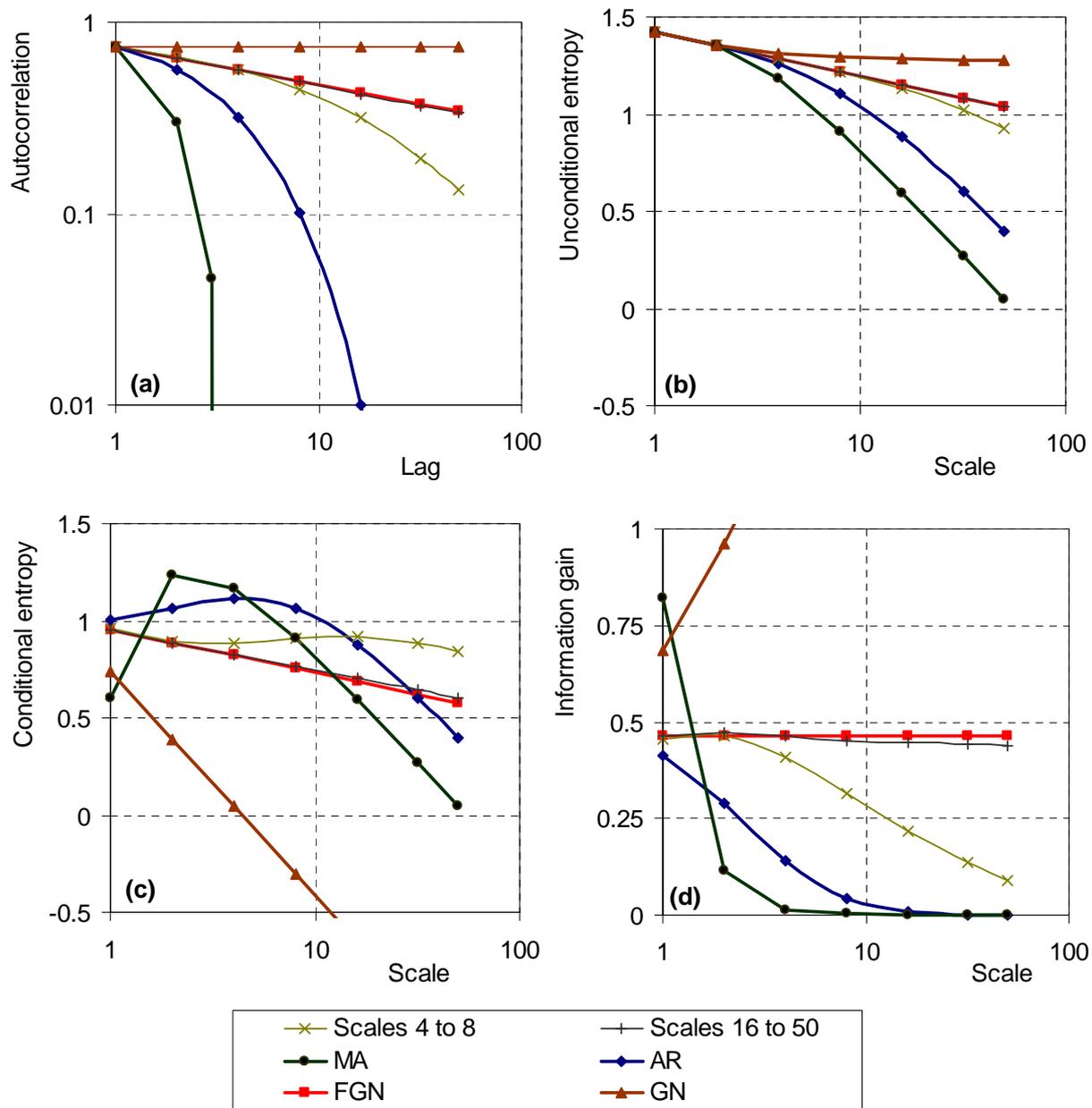


Figure 10 Same as Figure 9 but for $\rho = 0.75$.

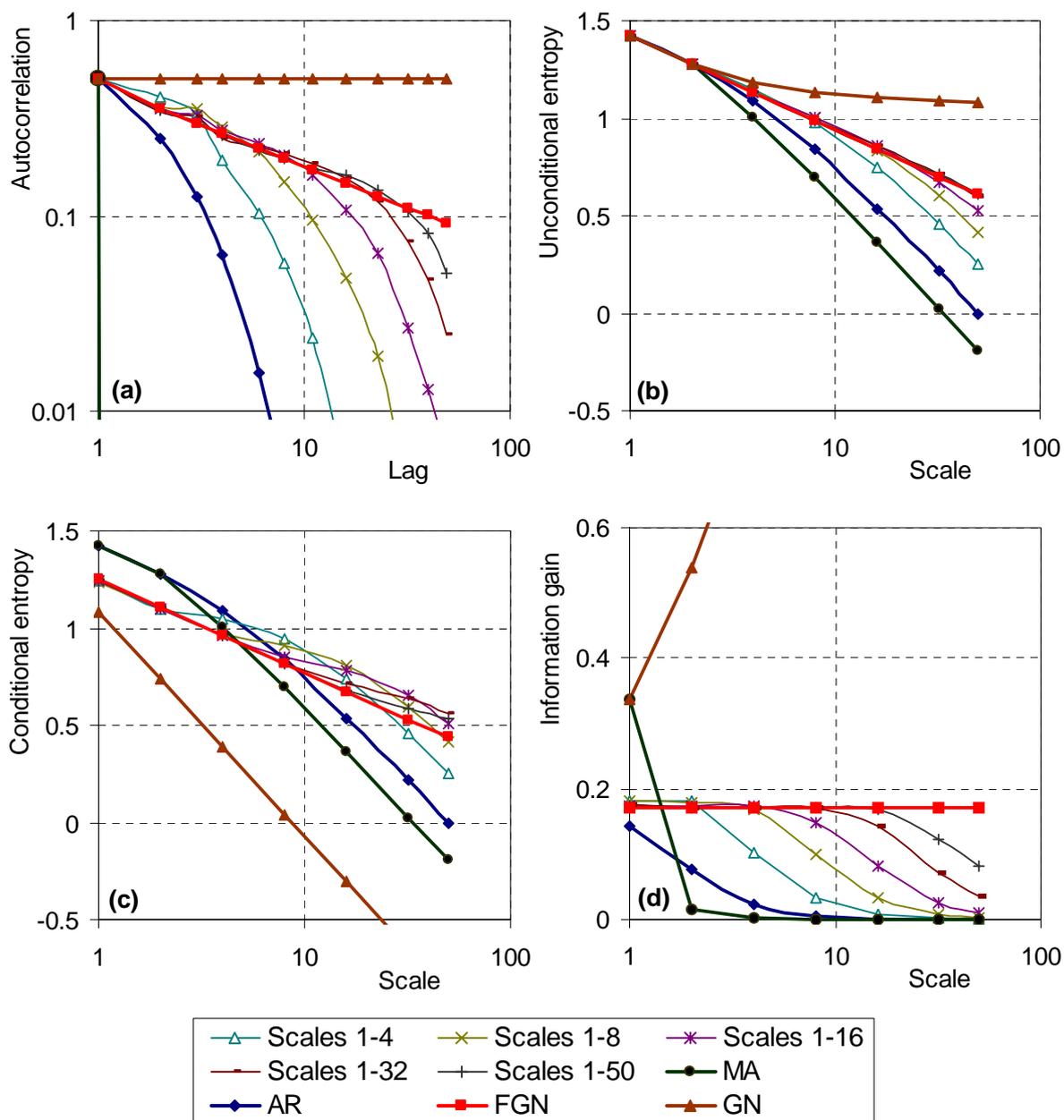


Figure 11 (a) Non-parametric autocorrelation functions that maximize unconditional entropy averaged over the indicated ranges of time scales, assuming lag one autocorrelation $\rho = 0.50$ and constraining information gain to be non-increasing function of time scale. (b-d) Resulting unconditional and conditional entropy and information gain, respectively, as function of time scale. In all panels, the relevant plots of benchmark models are also given for comparison.