An entropic-stochastic representation of rainfall intermittency: The origin of clustering and persistence

Demetris Koutsoyiannis

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

Abstract The well-established physical and mathematical principle of maximum entropy, interpreted as maximum uncertainty, is used to explain the observed dependence properties of the rainfall occurrence process, including the clustering behavior and persistence. The conditions used for the maximization of entropy are as simple as possible, i.e. that the rainfall processes is intermittent with dependent occurrences. Intermittency is quantified by the probability that a time interval is dry, and dependence is quantified by the probability that two consecutive intervals are dry. These two probabilities are used as constraints in a multiple scale entropy maximization framework, which determines any conditional or unconditional probability of any sequence of dry and wet intervals at any time scale. Thus, the rainfall occurrence process including its dependence structure is described by only two parameters. This dependence structure appears to be non-Markovian. Application of this theoretical framework to the rainfall data set of Athens indicates good agreement of theoretical predictions and empirical data at the entire range of scales for which probabilities dry and wet can be estimated (from one hour to several months).

Keywords clustering; entropy; hydrological persistence; Markov chains; point processes; power laws; rainfall intermittency; rainfall occurrence; risk; scaling; stochastic processes; uncertainty.

First draft, April 2005; revision, August 2005

Only entropy comes easy

Anton Chekhov

1. Introduction

1.1 Stochastic description of rainfall intermittency

The most significant peculiarity of the rainfall process, in comparison to other hydrometeorological processes, is its intermittent character on fine (sub-monthly) time scales. The analysis of intermittency and the modeling of the occurrence process of rainfall has attracted the interest of many and continues to do so. It was recognized in early stages of analysis and modeling attempts that the rainfall occurrences are not purely random. In other words, rainfall occurrence cannot be modeled (effectively) as a Bernoulli process in discrete time or, equivalently, as a Poisson process in continuous time.

Both discrete time and continuous time representations of the rainfall occurrence process, which in fact are closely related [e.g. *Foufoula-Georgiou and Lettenmaier*, 1986; *Small and Morgan*, 1986], have been investigated and are still in operational use. The most typical tool of the category of discrete time representations is the Markov chain model [*Gabriel and Neumann*, 1962; *Feyerherm and Bark*, 1964; *Hershfield*, 1970; *Todorovic and Woolhiser*, 1975; *Haan et al.*, 1976; *Chin*, 1977; *Katz*, 1977a, b; *Kottegoda and Horder*, 1980; *Roldan and Woolhiser*, 1982]. In this model, any time interval (e.g. day) can be in one of two states, dry or wet, and it is assumed that the state in a time interval depends on the state in the previous interval (see section 4 for mathematical details).

It was observed, however, that Markov chain models yield unsatisfactory results for rainfall occurrences, especially for dry intervals [*De Bruin*, 1980]. Also, the interannual variance of monthly (or seasonal) total precipitation is greater than predicted by Markov chain models, an effect termed “overdispersion” [*Katz and Parlange*, 1998]. Extended versions of the binary state Markov chains using a higher number of states may improve performance. Additional states in such model versions have been defined based on combination of states of two consecutive periods [*Hutchinson*, 1990] or on accounting for the rainfall depth of each interval [*Haan et al.*, 1976]. A more effective enhancement is to use transition probabilities

taking into account more than one previous interval, which leads to stochastic binary chains of order higher than one [*Pegram*, 1980; *Stern and Coe*, 1984; *Katz and Parlange*, 1998; *Clarke*, 1998, pp. 51-55]. In more recent developments, to take account of a long number of previous time intervals and simultaneously avoid an extremely high number of transition probabilities, it was proposed that, instead of the sequence of individual states of these intervals, one could use conditional probabilities based on aggregation of states of previous intervals [*Sharma and O'Neill*, 2002]. Similarly, one could use a discrete wetness index based on the number of previous wet intervals [*Harrold et al.*, 2003]. An extension of the Markov chain approach to multiple sites has been recently studied by *Pegram and Seed* [1998].

The dominant tools of the continuous time representation are the cluster-based point processes [*Waymire and Gupta*, 1981a, b, c]. These are essentially based on the prototype of the spatial distribution of galaxies devised by *Neyman and Scott* [1952] to describe their property of "clustering" relative to the Poisson process. With reference to storms, if they were regarded as instantaneous pulses positioned at random points in time, then the logarithm of probability that the interarrival time exceeds a value x, or the log survival function, would be proportional to x. However, empirical evidence suggests that the log survival function is a nonlinear concave function of x which indicates a tendency for clustering of rainfall events relative to the Poisson model [*Foufoula Georgiou and Lettenmaier*, 1986]. This clustering has been modeled by a cascade of two Poisson processes, corresponding to two characteristic time scales of arrivals of storms and storm cells.

The Neyman-Scott process with instantaneous pulses was the first applied to rainfall occurrence [*Kavvas and Delleur*, 1981; *Rodriguez-Iturbe et al.*, 1984], later succeeded by the Neyman-Scott rectangular pulses and the very similar Bartlett-Lewis rectangular pulses models [*Rodriguez-Iturbe et al.*, 1987]. A major problem of these models was their inability to reproduce the probability of zero rainfall at multiple time scales [*Velghe et al.*, 1994]. In this respect, *Foufoula-Georgiou and Guttorp* [1986] note that the Neyman-Scott model parameters are scale dependent and thus cannot be attributed a physical meaning. To ameliorate this, modifications of both the Neyman-Scott model [*Entekhabi et al.*, 1989] and the Bartlett-Lewis model *[Rodriguez-Iturbe et al.*, 1988; *Onof and Wheater*, 1993, 1994] were

proposed. These are in fact based on the randomization of the mean interarrival time of the one of the two Poisson processes. Evaluation and comparison of several cluster-based rectangular pulses models for rainfall are done by *Velghe et al.* [1994] and *Verhoest et al.* [1997] whereas a comprehensive review of Poisson-cluster models and the more recent developments has been provided by *Onof et al.* [2000].

Apart from rainfall occurrence, the point process models describe the entire rainfall process using one or two additional parameters. However, in some cases [e.g. *Gyasi-Agyei and Willgoose*, 1997] they have been used to simulate merely rainfall occurrences and then have been combined with other models that simulate rainfall depths. Other modeling approaches for the rainfall process (including its intermittency) are reviewed in *Srikanthan and McMahon* [2001]. More recently, multifractal simulation techniques have been also used for the modeling of the complete rainfall process [e.g. *Marshak et al.*, 1994; *Olsson*, 1996, *Menabde et al.*, 1997]. But require more analysis, particularly in their ability to reproduce the rainfall occurrence process and specifically the dry period structure at different scales.

1.2 Seeking an explanation of clustering and persistence

The terms “clustering” and “overdispersion” express the fact that rainfall occurrences are not independent in time but they are positively autocorrelated. Probably “clustering” is understood as focusing on small scales (clustering of events) whereas “overdispersion” may be more appropriate at large scales (at which the increased variance becomes more observable). “Persistence” is another term for the same behavior, which has been used in hydrology to express positive autocorrelation both on small scales (short-term persistence) and on large scales (long-term persistence). We may think that long-term persistence in the case of the binary processed examined here may not be the same as in the case of real valued processes (e.g. in the process of rainfall depth) and may not indicate a long-range dependence with the rigorous mathematical meaning [e.g. *Beran*, 1994] of the latter term. Furthermore, long-term persistence in not necessarily understood as long memory of the system. As demonstrated by *Klemes* [1974] and more recently by *Koutsoyiannis* [1992], even in the case of real valued processes, long-term persistence is more consistently understood as the effect

of climatic fluctuations on one time scale [in *Klemes*, 1974] or on several time scales [in *Koutsoyiannis*, 1992], rather than the result of some memory mechanism.

Whichever term is used, the problem we have to study in analysis and modeling of the rainfall occurrence process relates to the dependence structure of the process. Several questions have to be answered in such a study, among which are: (a) What is an appropriate yet parsimonious quantification of the dependence structure? (a theoretical question); (b) What does empirical evidence suggest about this dependence structure (an empirical question); and (c) Why does this empirically observed structure occur? (a theoretical question). All these questions are investigated in this paper but the emphasis is on the last one, the type "why". Generally, questions of this type are rarely posed and answered in hydrological stochastics. However, answers to such questions, additionally to providing explanations of natural phenomena, would help build improved models or choose the most appropriate of a spectrum of models.

With reference to the modeling approaches discussed earlier, we could claim that they do not involve type "why" questions. We could say that one model outperforms another, in terms of being more consistent with empirically observed properties, but this does not involve necessarily any explanation of the natural behavior. Before we proceed to the main type "why" question posed above, it would be useful to discuss how the performance of each of these models compare to each other and why. Generally, we could classify the models into three types: (a) independence, which includes the Bernoulli and Poisson cases, characterized by one parameter only; (b) simple dependence, which includes Markov chains and the original cluster-based models, characterized by 2-4 parameters (for merely the occurrence process); and (c) complex dependence, which includes the modified cluster-based models, characterized by 4 or more parameters. As we proceed from category (a) to (c) the performance is improved. An apparent reason for this is the increased number of parameters. For example, the five adjustable parameters of the occurrence process within the modified Bartlett-Lewis rectangular pulse model provide much more flexibility in fitting to any set of empirical probabilities, in comparison to the two independent parameters of the binary Markov chain model. More than this, it could be said that physical realism is improved as we

proceed from category (a) to (c). Category (a) is dominated by a single characteristic time scale (e.g. the inverse of the frequency of event occurrences), whereas in category (b) at least one time scale is added (due to the Markov dependence or the second Poisson process employed) and in category (c) a third time scale of fluctuation appears indirectly (due to the random fluctuation of the frequency of the second Poisson process). It seems unreasonable to assume that natural processes are dominated by a single time scale. Thus the larger the number of time scales involved in a model, the more proximity of the model to reality. As shown in *Koutsoyiannis* [2002], from a practical point of view, three time scales of fluctuation may be enough to approximate a phenomenon theoretically characterized by an infinite number of scales, such as the Hurst phenomenon.

In conclusion, the models discussed may yield better or worse descriptions of the natural rainfall occurrence process but they do not explain it, i.e. they are descriptive rather than explanatory. Here the explanation is sought via the principle of maximum entropy (ME), which is a well established principle in physics and mathematics. The notion of entropy in a stochastic context and the ME principle are briefly discussed in section 5. Recently *Koutsoyiannis* [2005a, b] successfully applied the ME principle to explain the distributional and dependence properties of hydrological processes, including the scaling behavior both in state and time. Several type "why" questions regarding behaviors of hydrological processes were considered and were all answered by a single and simple answer: each of the observed behaviors maximizes entropy for the given conditions of process variation and dependence, for the appropriate time scale or range of time scales.

The notion of entropy is used in many scientific disciplines including stochastic processes, information theory, statistics, statistical mechanics, thermodynamics, dynamical systems and fluid mechanics. Depending on the discipline, entropy may be regarded as a measure of uncertainty, information, discrimination, order or disorder, and complexity [e.g. *Georgii*, 2003]. Despite the common logical and historical root of the entropy concept in all these disciplines, the different technical variants and different views of entropy result in major difficulties and gaps in communication between the different disciplines [*Greven et al.*, 2003, p. xi]. However, it is worth noting that, in a physical context, different (though intuitively

related) to the stochastic context of the present study (described in section 5), the relevant hypothesis of Maximum Entropy Production (according to which the rate of entropy production is at a maximum), explains the long-term mean properties of the global climate system and those of turbulent fluid systems [*Ozawa et al.*, 2003]. The same hypothesis also seems to be applicable to the atmospheres and perhaps to mantle convection in planets other than Earth.

The interpretation of entropy used here is the standard one in the theory of stochastic processes, i.e. that entropy is a measure of uncertainty or ignorance [e.g. *Papoulis*, 1991]. In this respect, maximization of entropy is equivalent to maximization of uncertainty. Coming back to the question of the dependence structure observed in the rainfall occurrence process, we will endeavor to study if this structure is explained by the ME principle, assuming conditions or constraints for maximization which are as simple as possible (sections 6-8). In this respect, the scope of the present study is exploratory and explanatory. Thus, the tools suitable to this context are graphical depictions and comparisons emphasizing the general "shapes" of dependence behaviors (sections 3 and 9). No emphasis is given to operational applications, except for a discussion of the potential operational usefulness, after further research, of the results in rainfall modeling and in engineering applications (section 10).

2. Definitions, notation and basic assumptions

The rainfall occurrence process can be described by a binary valued stochastic process, with the values 0 and 1 representing dry and wet conditions, respectively. Here, a discrete time approach on multiple time scales will be considered. The scale index k is assumed to be a positive integer with the value $k = 1$ representing some basic scale δ. Without loss of generality and unless stated otherwise, in this paper the basic scale will be assumed to be the hourly scale (δ = 1 h), so, for instance, the daily scale will be represented as $k = 24$. Using this convention, the rainfall occurrence process will be denoted as $X_i^{(k)}$ with i denoting discrete time in units $k\,\delta$ (see Figure 1(a)). At the basic scale the superscript indicating scale will be omitted, i.e. $X_i \equiv X_i^{(1)}$. Obviously, the random variable $X_i^{(k)}$ will have value 0 if all its components X_j at the basic time scale are 0 and value 1 if one or more of its components are 1.

Let $p^{(k)}$ denote the probability that an interval of length k is dry, i.e., $p^{(k)} := P\{X_i^{(k)} = 0\}$. The probability that the interval is wet is $P\{X_i^{(k)} = 1\} = 1 - p^{(k)}$. Clearly, this notation implies that the probabilities do not depend on the time position i but depend only on the time scale k. This reflects a stationarity assumption that will be used throughout this paper. This, however, does not put limitations as it is always possible to confine the analysis into appropriate climatically homogeneous periods (months or seasons, thus assuming cyclostationarity rather than stationarity). The sequence $p^{(k)}$, $k = 1, 2, \ldots$, will be referred to as the probability dry sequence. At the basic scale, the superscript (1) may be omitted for convenience, so $p \equiv p^{(1)}$. Simultaneously, the symbol p will be used with one or more subscripted indices that denote states of consecutive time steps. For example, at the basic time scale, $p_0 = P\{X_t = 0\}$ (and thus $p_0 \equiv p \equiv p^{(1)}$), $p_1 = P\{X_t = 1\}$ (and thus $p_1 \equiv 1 - p \equiv 1 - p^{(1)}$), $p_{00} := P\{X_t = 0, X_{t-1} = 0\}$ (and thus $p_{00} \equiv p^{(2)}$), etc. More generally, for any scale k, any number of consecutive time intervals q (each of length k), and any indices $j_0, j_1, j_2, \ldots, j_{q-1}$ each taking values 0 and 1, the following notational convenience is used (see demonstration in Figure 1(c))

$$p^{(k)}_{j_0 j_1 \ldots j_{q-1}} := P\{X_i^{(k)} = j_0, X_{i-1}^{(k)} = j_1, X_{i-2}^{(k)} = j_2, \ldots, X_{i-q+1}^{(k)} = j_{q-1}\} \tag{1}$$

For q consecutive time intervals there are 2^q possible states and an equal number of state probabilities. The index permutations $j_0, j_1, j_2, \ldots, j_{q-1}$ characterizing each of these states, can be ordered in a convenient manner. That is, we observe that any index permutation represents the binary digit form of some integer l ranging from 0 to $2^q - 1$. Such a binary form will be denoted as $<j_0 j_1 \ldots j_{q-1}>$; for instance, if $q = 3$, and $j_0 = 0$, $j_1 = 1$ and $j_2 = 0$, then $<j_0 j_1 j_2> = <010> = 0\times2^3 + 1\times2^2 + 0 = 4 = l$. In this manner we can define the vector $\mathbf{p}_q^{(k)}$ of all possible state probabilities, i.e.,

$$\mathbf{p}_q^{(k)} := [p^{(k)}_{00\ldots00}, p^{(k)}_{00\ldots01}, \ldots, p^{(k)}_{11\ldots11}]^T \tag{2}$$

whose lth element is $p^{(k)}_{j_0 j_1 \ldots j_{q-1}}$, where $<j_0 j_1 \ldots j_{q-1}> = l - 1$ and the superscript T denotes the transpose of a vector or matrix. For example, for $k = 1$ and $q = 1, 2, 3$, we get

$$\mathbf{p}_1 = [p_0, p_1]^T, \mathbf{p}_2 = [p_{00}, p_{01}, p_{10}, p_{11}]^T, \mathbf{p}_3 = [p_{000}, p_{001}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}, p_{111}]^T \tag{3}$$

Furthermore, considering a sequence of q consecutive time steps, each of scale k, the probability of having exactly n out of q wet intervals, where $0 \le n \le q$, will be denoted as $p_{n\backslash q}^{(k)}$ (see Figure 1(c)). Obviously,

$$p_{n\backslash q}^{(k)} = \sum_{j_0, j_1, \ldots, j_{q-1}} p_{j_0 j_1 \ldots j_{q-1}}^{(k)} U(j_0 + j_1 + j_2 + \ldots + j_{q-1} = n) \tag{4}$$

where $U(B) = 1$ if logical condition B is true and $U(B) = 0$ if condition B is false.

The vector $\mathbf{p}_q^{(k)}$ represents the order q (joint) probability mass function of the process $X_i^{(k)}$ and hence the dependence structure of the process. If the process is stationary, as assumed in this study, some recursive constraints are imposed to the elements of $\mathbf{p}_q^{(k)}$. For example for $q = 3$ and $k = 1$ we can write $P\{X_t = 0, X_{t-1} = j_1, X_{t-2} = j_2\} + P\{X_t = 1, X_{t-1} = j_1, X_{t-2} = j_2\} = P\{X_{t-1} = j_1, X_{t-2} = j_2\} = P\{X_t = j_1, X_{t-1} = j_2\}$ or $p_{0 j_1 j_2} + p_{1 j_1 j_2} = p_{j_1 j_2}$. Similarly, $p_{j_1 j_2 0} + p_{j_1 j_2 1} = p_{j_1 j_2}$. Extending this for any scale and order, the following equalities hold:

$$\begin{aligned} &p_{0 j_1 \ldots j_{q-2} 0}^{(k)} + p_{0 j_1 \ldots j_{q-2} 1}^{(k)} = p_{0 j_1 \ldots j_{q-2}}^{(k)}, \quad p_{1 j_1 \ldots j_{q-2} 0}^{(k)} + p_{1 j_1 \ldots j_{q-2} 1}^{(k)} = p_{1 j_1 \ldots j_{q-2}}^{(k)} \\ &p_{0 j_1 \ldots j_{q-2} 0}^{(k)} + p_{1 j_1 \ldots j_{q-2} 0}^{(k)} = p_{j_1 \ldots j_{q-2} 0}^{(k)}, \quad p_{0 j_1 \ldots j_{q-2} 1}^{(k)} + p_{1 j_1 \ldots j_{q-2} 1}^{(k)} = p_{j_1 \ldots j_{q-2} 1}^{(k)} \end{aligned} \tag{5}$$

where the left- and right-hand sides contain elements of $\mathbf{p}_q^{(k)}$ and $\mathbf{p}_{q-1}^{(k)}$ respectively. If we sum all four equations in (5), both sides will be equal to $2\, p_{j_1 \ldots j_{q-2}}^{(k)}$, which means that one of the four equations is linearly dependent on the others. By taking all permutations $j_1 \ldots j_{q-2}$, a number $4 \times 2^{q-2}$ of such equations can be written, of which $3 \times 2^{q-2}$ are independent. Thus, if $\mathbf{p}_{q-1}^{(k)}$ is known, the number of additional parameters (probabilities) required to define $\mathbf{p}_q^{(k)}$ is $2^q - 3 \times 2^{q-2} = 2^{q-2}$. Consequently, the total number of independent parameters (probabilities) to define the order q probability mass function is $1 + 1 + \ldots + 2^{q-2} = 2^{q-1}$. For instance, for $q = 10$ the vector $\mathbf{p}_q^{(k)}$ contains $2^{10} = 1024$ elements and involves $2^9 = 512$ independent parameters.

At scale k, assuming that q past states are known, the conditional or transition probability for one time step ahead is denoted as

$$\pi_{m|j_0 j_1 \ldots j_{q-1}}^{(k)} := P\{X_{i+1}^{(k)} = m | X_i^{(k)} = j_0, X_{i-1}^{(k)} = j_1, X_{i-2}^{(k)} = j_2, \ldots, X_{i-q+1}^{(k)} = j_{q-1}\} \tag{6}$$

(see explanation in Figure 1(d)). The set of transition probabilities can be ordered in a similar manner as the state probabilities. In this way, we define the vector

$$\boldsymbol{\pi}_q^{(k)} := [\pi_{0|00\ldots00}^{(k)}, \pi_{0|00\ldots01}^{(k)}, \ldots, \pi_{0|11\ldots11}^{(k)}]^T \tag{7}$$

whose lth element is $p_{0|j_0 j_1 \ldots j_{q-1}}^{(k)}$, where $<j_0 j_1 \ldots j_{q-1}> = l - 1$. For example, for $k = 1$ and $q =$ 1 and 2 we get

$$\boldsymbol{\pi}_1 = [\pi_{0|0}, \pi_{0|1}]^T, \ \boldsymbol{\pi}_2 = [\pi_{0|00}, \pi_{0|01}, \pi_{0|10}, \pi_{0|11}]^T \tag{8}$$

Notice that the vector of transition probabilities $\boldsymbol{\pi}_q^{(k)}$ refers to transition to future state 0 only. Normally, in a Markov chain model a matrix (rather than a vector) of transition probabilities is used which contains probabilities to all future states. However in a binary Markov chain this vector suffices, because

$$\pi_{1|j_0 j_1 \ldots j_{q-1}}^{(k)} = 1 - \pi_{0|j_0 j_1 \ldots j_{q-1}}^{(k)} \tag{9}$$

The sequences of vectors of state probabilities $\mathbf{p}_q^{(k)}$ and transition probabilities $\boldsymbol{\pi}_q^{(k)}$ are closely connected to each other. From the definition of conditional probability [e.g. *Papoulis*, 1991, p. 27] we find

$$p_{m j_0 j_1 \ldots j_{q-1}}^{(k)} = \pi_{m|j_0 j_1 \ldots j_{q-1}}^{(k)} p_{j_0 j_1 \ldots j_{q-1}}^{(k)} \tag{10}$$

Using (9), we easily obtain that

$$\mathbf{p}_{q+1}^{(k)} = \begin{bmatrix} \mathbf{p}_q^{(k)} \circ \boldsymbol{\pi}_q^{(k)} \\ \mathbf{p}_q^{(k)} - \mathbf{p}_q^{(k)} \circ \boldsymbol{\pi}_q^{(k)} \end{bmatrix} \tag{11}$$

where $\circ$ denotes the Hadamard (element-by-element) product of a vector or matrix. For example, for $q = 2$ and $k = 1$,

$$\mathbf{p}_3 = [p_{000}, p_{001}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}, p_{111}]^T =$$
$$= [p_{00}\pi_{0|00}, p_{01}\pi_{0|01}, p_{10}\pi_{0|10}, p_{11}\pi_{0|11}, p_{00}(1-\pi_{0|00}), p_{01}(1-\pi_{0|01}), p_{10}(1-\pi_{0|10}), p_{11}(1-\pi_{0|11})]^T \tag{12}$$

Thus, the sequence of vectors $\mathbf{p}_q^{(k)}$ for different q is related to the sequence of $\boldsymbol{\pi}_q^{(k)}$ in a recursive manner.

The description of the dependence structure of the process in terms of $\mathbf{p}_q^{(k)}$, albeit complete, is not very efficient due to excessive number of parameters; the situation worsens more if multiple time scales k are considered. In subsequent sections it will be demonstrated how we can establish a parameter parsimonious approach with the help of the ME principle.

3. Empirical basis and indicators of dependence

As an empirical basis for this study, the data set of rainfall in Athens, Greece is used. Athens is a dry place with mean annual rainfall around 400 mm. January is the wettest month and August the driest. Characteristic indices of the rainfall occurrence process for these two months (representative of winter and summer respectively), for May and October (representative of spring and autumn respectively), as well as for the four-month (June-September) dry season and the entire year are shown in Table 1. The available data, digitized at the hourly scale, extend from 1927 to 1996. This makes a total of 70 years of measurements, which reduce to about 66 years if missing values (spread through the years) are not counted. Additional daily data exist before 1927 but were not used. For the stationarity property to hold, the data should be divided into climatically homogenous periods, which, as usual, were assumed to be the twelve months. Assuming that, in this case, the estimation of empirical probabilities can be done with some accuracy up to the scale of half a month, the range of the study time scales extends from k = 1 to 360 (= 15×24). Another option, is to consider multi-month climatic seasons. This option was also used in this study especially for the dry season. This was done to (1) include multi-month dry periods in the analysis, given that long dry periods (in fact exceeding four months – up to about 130 days) are observed in the available data set; (2) extend the observable range of time scales; and (3) acquire more accurate estimates, given that the seasonal 4-month set contains 4 times more data values than each monthly data set. In addition to the monthly and seasonal subsets, the total set of all data through all months was also examined, even though stationarity does not hold for the entire year. This was done to (1) form a picture of an "average" behavior throughout all year; (2) test if the "average" behavior is explained by the entropic framework of this study, given the exploratory and explanatory (rather than operational) character of the current analysis; and (3)

acquire even more accurate estimates, given that the annual data set contains 12 times more data values than each monthly data set. It is noted that the practice of considering all data regardless of month or season is not infrequent in hydrological applications such as analysis of extremes and construction of rainfall intensity/duration/frequency curves. As will be demonstrated later, the last task is closely connected to rainfall intermittency, so the results of the analysis at the annual basis will be useful.

A common indicator of dependence of a stochastic process is the autocorrelation coefficient ($\rho_j^{(k)}$:= Corr[$X_i^{(k)}$, $X_{i+j}^{(k)}$] for lag j and scale k) and its variation. In a multiple time scale analysis, there are two ways, virtually equivalent to each other, of studying the autocorrelation structure [*Koutsoyiannis*, 2002]: (a) to study the variation of autocorrelation with time lag j at fixed time scale k, and (b) to study the variation of autocorrelation with time scale k for fixed time lag j, usually for $j = 1$. The second approach is more convenient and suitable for the current study due to its multiple time scale setting. Denoting $\rho^{(k)} \equiv \rho_1^{(k)}$, we can easily obtain

$$\rho^{(k)} = \frac{p^{(2k)} - (p^{(k)})^2}{p^{(k)} - (p^{(k)})^2} \tag{13}$$

(The derivation of this equation is based on the obvious relationships $E[X_i^{(k)}] = E[(X_i^{(k)})^2] = p_1^{(k)} = 1 - p^{(k)}$, and $E[X_i^{(k)} X_{i+1}^{(k)}] = p_{11}^{(k)} = 1 - 2p^{(k)} + p^{(2k)}$, where $E[\ ]$ denotes expected value). Thus, the sequence $p^{(k)}$ suffices to estimate $\rho^{(k)}$ either empirically (from a given data set) or theoretically (for a given model of $p^{(k)}$).

In the radically simplifying assumption of complete independence of $X_i^{(k)}$, the following obvious relation holds

$$p^{(k)} = p^k \tag{14}$$

Thus, a single parameter p ($0 < p < 1$) suffices to describe completely the entire distribution function of the process $X_i^{(k)}$ of any order and scale. As discussed in the introduction, this simple model is not appropriate for rainfall. Notwithstanding, a comparison of the model to a real world rainfall data set is useful as it may serve as a starting point to describe the dependence structure.

As anticipated, in the independence case, combining (14) and (13) we obtain $\rho^{(k)} = 0$ regardless of k. Apparently, this is far from reality. As shown in Figure 2 (left panel), which depicts the empirical lag one autocorrelations for the Athens data set for the dry season and for the entire year, the autocorrelation is positive even for scales k as high as 3000. The arrangement of empirical points indicates a pattern decreasing with scale. It is noted for comparison that in a real valued (rather than binary) Markov process $\rho^{(k)}$ tends hyperbolically to zero for increasing scale whereas, in contrast, in fractional Gaussian noise, $\rho^{(k)}$ is kept constant [*Koutsoyiannis*, 2002]. Thus, in the case examined $\rho^{(k)}$ does not necessarily manifest a long-term persistence pattern.

Therefore, a different indicator is required here to explore the possible long-term persistence of the rainfall occurrence process. Let us try the dependence indicator

$$\tau^{(k)} := \frac{-\ln p^{(k)}}{-\ln p^{(2k)}} \tag{15}$$

which involves the same quantities as in (13), but in a simpler expression. Due to (14), in the independence case $\tau^{(k)}$ is independent of k, equal to ½. Clearly, values of $\tau^{(k)}$ in the interval (0.5, 1] indicate a positive dependence structure. As shown in Figure 2 (right panel), which depicts values of $\tau^{(k)}$ versus k for the Athens data set for dry season and for the entire year, this dependence indicator is always greater than ½. For scales smaller than around 10, the empirical $\tau^{(k)}$ decreases with k but then it seems to stabilize to a value greater than ½. The deviation of this value from ½ is assumed here as reflecting long-term persistence of the process.

Both indicators $\rho^{(k)}$ and $\tau^{(k)}$ are nonlinear expressions of probability values at two time scales; therefore, their estimation from empirical data becomes highly uncertain as estimation errors are amplified when performing the algebraic manipulations on the two time scales. (As seen in Figure 2, the plots for the dry season is more scattered than those for the entire year; in addition, the plots for each of the months, for which the empirical probabilities were estimated from 1/12 of the entire data set, are too scattered and therefore have been not included). To establish a more robust indicator of long-term persistence, we write equation (14) in the equivalent form

$$-\ln p^{(k)} = (-\ln p)\, k \tag{16}$$

It can be observed that the left side of (16) is the discrete analogue of the so called log survival function of the interarrival time in continuous time modeling using point processes [e.g. *Foufoula-Georgiou and Lettenmaier*, 1986]. (That is to say, the probability that the interarrival time exceeds k time steps equals the probability that k steps are dry, i.e. $p^{(k)}$). The difference here is that we used minus-logarithms in order for all involved quantities to be positive. Thus, in a Cartesian or in a double logarithmic plot of $-\ln p^{(k)}$ versus k the independence model will be depicted as a straight line with slope 1. Departures from this line manifests dependence of the process. Particularly, a departure from slope 1 at large scales can be interpreted as an indicator of long-term persistence.

In Figure 3, the empirical estimates of $-\ln p^{(k)}$ for several months as well as for the dry season and the entire year are plotted versus scale k in a double logarithmic diagram. The relationship (16) is also depicted in Figure 3 and it can be seen that it departs significantly from the empirical points. It can be also seen that the following inequality holds:

$$-\ln p \le -\ln p^{(k)} \le (-\ln p)\, k \tag{17}$$

where the equality case of the right-hand side corresponds to (16).

For large time scales k, an almost straight line arrangement of empirical points appears in Figure 3 in all plots referring to monthly, seasonal and annual basis. This suggests that for large k a relationship of the form

$$-\ln p^{(k)} \sim k^{\eta} \tag{18}$$

should exist, where η is a positive constant less than 1. The exponent η may serve as an appropriate indicator of long-term persistence for a binary valued process, similar to the well known Hurst exponent H, which is an indicator of the long-term persistence in a real valued process [e.g. *Koutsoyiannis*, 2002]. Obviously, there are some quantitative differences between H and η. In the case of H, the independence corresponds to the value $H = 0.5$ and the positive long-term persistence corresponds to $H > 0.5$ up to 1. In the case of η, the independence corresponds to the value $\eta = 1$ and the positive long-term persistence

corresponds to $\eta < 1$ down to 0 (as indicated by (17)). This will be clarified better in later sections.

4. The Markovian dependence case and the Markov chain model

The simplest possible dependence structure is the Markovian one. Let us recall that a process with Markovian dependence is defined from the property [e.g. *Papoulis*, 1991, p. 635]

$$P\{X_{i+l} = x_{i+l} | X_i = x_i, X_{i-1} = x_{i-1}, \ldots, X_1 = x_1\} = P\{X_{i+l} = x_{i+l} | X_i = x_i\} \tag{19}$$

which indicates that the future of the process (time $i + l$ with $l > 0$) is not influenced by the past (time $i - 1$ and earlier) when the present (time i) is known.

Because by definition $p^{(k)} = P\{X_i = 0, X_{i-1} = 0, \ldots, X_{i-k+1} = 0\} = P\{X_i = 0 | X_{i-1} = 0, \ldots, X_{i-k+1} = 0\}\, P\{X_{i-1} = 0, \ldots, X_{i-k+1} = 0\} = \pi_{0|0}\, p^{(k-1)}$ (due to the Markovian property), we may conclude that

$$p^{(k)} = \pi_{0|0}\, p^{(k-1)} = p\, \pi_{0|0}^{k-1} \tag{20}$$

or

$$-\ln p^{(k)} = -\ln p - \ln \pi_{0|0}\, (k-1) \tag{21}$$

This is a characteristic property of the Markovian process. It is easily observed that for large k, $-\ln p^{(k)} \sim k$, so that, with reference to equation (18), the exponent η is 1, as in the independence case. This suggests that the Markov chain model cannot describe the long-term persistence of the actual rainfall process.

Equation (20) can be used to estimate the parameter $\pi_{0|0}$ from any two probabilities $p^{(k)}$, say $p^{(2)}$ and $p = p^{(1)}$, in which case $\pi_{0|0} = p^{(2)} / p$. The other model parameter $\pi_{0|1}$ can be estimated from (5) (written as $p_{00} + p_{01} = p_0$ or equivalently $p^{(2)} + \pi_{0|1}\, p_1 = p_0$), from which it is found that $\pi_{0|1} = (p - p^{(2)}) / (1 - p)$. In conclusion, given the two probabilities dry for scales 1 and 2, i.e. $p^{(2)}$ and p, the probability dry sequence $p^{(k)}$ is completely determined by (20) and the dependence structure is completely determined by the Markovian assumption (19).

It is important to observe that if the property (20) holds for a single time scale then it holds for any time scale as manifested by the following equation, which is a consequence of (20):

$$p^{(j\,k)} = p^{(j)} \left(\frac{p^{(2\,j)}}{p^{(j)}} \right)^{k-1} \tag{22}$$

For $j = 1$ this becomes identical to (20) and for $j > 1$ this represents the same property for a change of the basic scale δ to $\delta' = j\,\delta$. It should be mentioned, however, that the process is not strictly Markovian on scales other than the basic scale, as it can be shown that transition probabilities, conditional on the previous period being wet, do not obey (19).

Comparison of the Markovian model with empirical data of Athens shows significant differences from reality; the model fitting was based on $p^{(2)}$ and p of the Athens data set on an hourly basic time scale. Specifically, Figure 3, shows significant departures of $-\ln p^{(k)}$ for scales 10 and above (i.e. five times greater than the largest scale used in the fitting) and insufficiency in describing the long term persistence (because it yields $\eta = 1$). This happens for all bases of analysis, monthly, seasonal and annual. Furthermore, Figure 2 shows significant departures at the seasonal and annual basis in terms of dependence indicators $\rho^{(k)}$ and $\tau^{(k)}$ at scales even lower than 10. Needless to say, however, that the Markovian model is much closer to reality than the independence model.

5. The entropy concept

If we abandon the Markovian property (the order one model) and assume a non-Markovian dependence, the probabilities that will appear in the model are too many (i.e. 2^q for a model of order q), as discussed above. The concept that we can use to reduce drastically the required number of parameters is the entropy and the appropriate tool is the ME principle.

The concept of entropy originated in the middle of the 19th century in the work of Clausius (although the name εντροπία appears already in ancient Greek, etymologized from τροπή, i.e. change, turn, drift), and was fundamental to formulate the second law of thermodynamics. In the late 19th and early 20th century, Boltzmann, then complemented by Gibbs, gave it a statistical mechanical content, showing that entropy of a macroscopical stationary state is proportional to the logarithm of the number w of possible microscopical states that correspond to this macroscopical state. Shannon generalized the mathematical form of entropy and also explored it further. At the same time, Kolmogorov founded the concept on more mathematical

grounds on the basis of the measure theory [see e.g. *Papoulis*, 1991, p. 535; *Müller*, 2003a; *Keane*, 2003; *Tsallis*, 2004.)

For a discrete random variable X taking the values x_j (j = 1,…, w) with probabilities $p_j \equiv p(x_j)$ such that

$$\sum_{j=1}^{w} p_j = 1 \tag{23}$$

the entropy (more precisely the Boltzmann-Gibbs-Shannon entropy), is by definition [e.g. *Papoulis*, 1991, p. 558] the quantity

$$\varphi := E[-\ln p(X)] = -\sum_{j=1}^{w} p_j \ln p_j \tag{24}$$

The entropy φ can be interpreted as the uncertainty about the variable X and equals the information gained when it is observed. The above definition can be extended to the case of a continuous random variable, but this is irrelevant here.

The ME principle postulates maximization of (24) subject to some constraints. It is easily seen if the only constraint is (23), then ME results in equal probabilities $p_j = 1/w$. For instance, the ME principle yields equal probabilities of 1/6 for each outcome in a die, a result that is also produced by the "principle of insufficient reason", which is attributed to Jakob Bernoulli. Conceptually, the two principles are equivalent [*Jaynes*, 1957; *Papoulis*, 1991] but the ME principle is more effective in problems involving asymmetric constraints, as will be discussed later. Also, it is more effective in problems involving stochastic processes in either an unconditional or conditional setting.

Let us consider the case of a stationary process X_i, whose distribution is determined in terms of its qth order joint probability mass function $p(\mathbf{x}_q)$ defined as

$$p(\mathbf{x}_q) := P\{\mathbf{X}_q = \mathbf{x}_q\}, \quad \mathbf{x}_q = [x_1, \ldots, x_q]^T, \ \mathbf{X}_q = [X_1, \ldots, X_q]^T \tag{25}$$

where upper- and lower-case symbols denote respectively random variables and their values. The qth order *joint entropy* is defined as:

$$\varphi_q := E[-\ln p(\mathbf{X}_q)] = -\sum p(\mathbf{x}_q) \ln p(\mathbf{x}_q) \tag{26}$$

where the sum is taken over all possible values of $\mathbf{x}_q$. The joint entropy φ_q can be interpreted as the uncertainty about the variables $X_1, \ldots, X_q$ and equals the information gained when they are observed.

The *conditional entropy of order q* of the process X_i is defined as [*Papoulis*, 1991, p. 566]:

$$\varphi_{c_q} := E[-\ln p(X_1|X_0, \ldots, X_{-q+1})] \tag{27}$$

where $p(X_1|X_0, \ldots, X_{-q+1})$ denotes the conditional probability of X_1 given the previous q items $X_0, \ldots, X_{-q+1}$. Using the so called chain rule [*Papoulis*, 1991, p. 564] it is shown that

$$\varphi_{c_q} := \varphi_q - \varphi_{q-1} \tag{28}$$

The limit as q tends to infinity (i.e. the conditional entropy when the entire past is observed) is called simply the *conditional entropy* φ_c, i.e.,

$$\varphi_c := \lim_{q \to \infty} (\varphi_q - \varphi_{q-1}) \tag{29}$$

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

$$\varphi - \varphi_c =: \psi \tag{30}$$

is a non-negative number that represents the *information gain* when past and present are observed. As will be discussed below, in a multiple scale entropy maximization framework, all the three quantities φ, φ_c and ψ should be evaluated for several scales.

6. Application to the rainfall occurrence procedure

In the following we will assume a binary chain model not necessarily Markovian, i.e. with order q that can be higher than one. Sometimes in the literature this model is referred to as a Markov chain of order q [e.g. *Clarke*, 1998, p. 46], which, however, is misleading because it does not satisfy the Markovian property (19); here the model is referred to as qth order stochastic binary chain.

The order 1 entropy of the rainfall occurrence process $X_i^{(k)}$ at scale k can be determined according to the definitions of the previous section, observing that the process has only two states 0 and 1 with probabilities $p^{(k)}$ and $1 - p^{(k)}$ respectively. In this case (24) takes the form

$$\varphi^{(k)} = -p^{(k)} \ln p^{(k)} - (1 - p^{(k)}) \ln(1 - p^{(k)}) \tag{31}$$

so $\varphi^{(k)}$ is a function of merely $p^{(k)}$. The order q entropy at scale k, estimated from (26), can be written as

$$\varphi_q^{(k)} := -(\mathbf{p}_q^{(k)})^T \ln(\mathbf{p}_q^{(k)}) \tag{32}$$

where $\ln(\mathbf{p}_q^{(k)})$ is meant as the vector whose elements are the logarithms of $\mathbf{p}_q^{(k)}$. Given $\varphi_q^{(k)}$, the conditional entropy of order q, $\varphi_{cq}^{(k)}$, the conditional entropy $\varphi_c^{(k)}$ for $q \to \infty$, and eventually the information gain $\psi^{(k)}$ are estimated respectively from (28), (29) and (30). Theoretically, to estimate the conditional entropy $\varphi_c^{(k)}$ and the information gain $\psi^{(k)}$ we need to calculate the limits for $q \to \infty$. In practice, apart from simple cases that can be confronted in an analytical manner (section 7), these quantities may be estimated numerically for a finite value of q. Numerical investigation showed that in typical cases $\varphi_{cq}^{(k)}$ converges quickly to $\varphi_c^{(k)}$ so that a value of q as low as 8 suffices to estimate $\varphi_c^{(k)}$. In addition, it was observed that the difference $(\varphi_{cq}^{(k)} - \varphi_{cq-1}^{(k)}) - \varphi_c^{(k)}$ decreases as an exponential function of q and this can be utilized to further improve the estimate of $\varphi_c^{(k)}$ using extrapolation.

As discussed above, the application of the ME principle requires some constraints to be defined. In the application to rainfall occurrence process, two conditions or equality constraints seem to be the least possible. The first is the preservation of a "mass" property of the process which can be represented by the mean of the process X_i at the basic scale, which (due to the binary nature of the process) is obviously $1 - p$. The second is the preservation of an elementary "dependence" property of the process. For this purpose the indicator $\tau \equiv \tau^{(1)}$ is used. Obviously, the preservation of both mass and dependence elementary properties, expressed as above, is equivalent to the preservation of the probabilities dry at time scales 1 and 2 (i.e. p and $p^{(2)}$). No condition will be employed for large time scales nor for the long term persistence.

7. Entropy maximization at a single time scale

For the constraints set, the unconditional entropy φ at scale 1 and order 1 is fixed and given by (31). Also, the order 2 entropy φ_2 is fixed, because $p^{(2)}$ is assumed known, and is given by (32) for $q = 2$. If we confine our interest at the basic scale $k = 1$, what we should maximize are higher order entropies φ_q at this scale. This can be done in an analytical manner, incorporating constraints by means of Lagrange multipliers and equating derivatives to zero. From (32), the derivative of φ_q with respect to any of the involved probabilities, is

$$\frac{\partial \varphi_q}{\partial p_{j_0 \dots j_{q-1}}} = -1 - \ln p_{j_0 \dots j_{q-1}} \tag{33}$$

The various $p_{j_0 \dots j_{q-1}}$cannot take arbitrary values as they are constrained by (5), which can be written in the following simpler manner by abbreviating the index permutation $j_1 \dots j_{q-2}$ with the single symbol “x”:

$$p_{0x0} + p_{0x1} = p_{0x}, \quad p_{1x0} + p_{1x1} = p_{1x}, \quad p_{0x0} + p_{1x0} = p_{x0}, \quad p_{0x1} + p_{1x1} = p_{x1} \tag{34}$$

Denoting λ_{0x}, λ_{1x}, λ_{x0} and λ_{x1} the Lagrange multipliers of the above equations (notice that the indices of λ are identical to those of p of the right-hand sides of equations) and observing that derivatives of any of the Lagrange terms with respect to any of the involved p_{ixj} is unity or zero, maximization of φ_q yields

$$\begin{aligned} -1 - \ln p_{0x0} + \lambda_{0x} + \lambda_{x0} = 0, \quad -1 - \ln p_{0x1} + \lambda_{0x} + \lambda_{x1} = 0 \\ -1 - \ln p_{1x0} + \lambda_{1x} + \lambda_{x0} = 0, \quad -1 - \ln p_{1x1} + \lambda_{1x} + \lambda_{x1} = 0 \end{aligned} \tag{35}$$

Eliminating the Lagrange multipliers we obtain

$$\ln p_{0x0} + \ln p_{1x1} = \ln p_{0x1} + \ln p_{1x0} \tag{36}$$

from which we conclude that

$$\frac{p_{0x0}}{p_{0x1}} = \frac{p_{1x0}}{p_{1x1}} = \frac{p_{0x0} + p_{1x0}}{p_{0x1} + p_{1x1}} = \frac{p_{x0}}{p_{x1}} \tag{37}$$

Equation (37) relates order q probabilities with those of order $q - 1$, and together with (34) suffices to determine all order q probabilities if those of order $q - 1$ are known. Specifically, we can directly obtain that for any i and j (each being 0 or 1)

$$p_{i\mathrm{x}j} = \frac{p_{i\mathrm{x}}\, p_{\mathrm{x}j}}{p_{\mathrm{x}}}, \tag{38}$$

where $p_{\mathrm{x}} = p_{\mathrm{x}1} + p_{\mathrm{x}0} = p_{0\mathrm{x}} + p_{1\mathrm{x}}$, i.e. represents probabilities of order $q - 2$. Now, since from the conditions set all probabilities of order 1 and 2 are known, recursive application of (38) determines all probabilities of any order, which successively maximize all φ_q.

Equation (38) can be also written as $p_{i\mathrm{x}j} \,/\, p_{\mathrm{x}j} = p_{i\mathrm{x}} \,/\, p_{\mathrm{x}}$. Since the terms of both sides represent transition probabilities, (38) can be also written as

$$\pi_{i|\mathrm{x}j} = \pi_{i|\mathrm{x}} \tag{39}$$

This shows that the transition probability does not depend on the rainfall occurrence on the most distant time interval (i.e. $\pi_{i|\mathrm{x}0} = \pi_{i|\mathrm{x}1} = \pi_{i|\mathrm{x}}$). Recursive application of (39) leads to the conclusion that the transition probability depends on one previous time interval only. In other words, the maximization of entropy at a single scale results in Markovian dependence.

8. Entropy maximization at multiple time scales

The maximization of entropy of a stochastic process at a single scale is not quite reasonable from a physical point of view. As discussed in more detail in *Koutsoyiannis* [2005b], what is needed is the application of the ME principle on a multiple time scale setting. This, however, is not easy, as it should take into account conditional and unconditional entropies at all scales, whose maximization is antagonistic. After a detailed study and examination of several options, *Koutsoyiannis* [2005b] proposed the following framework, whose successful application produced (explained) the Hurst phenomenon:

1. The quantity to be maximized is a compound measure of unconditional entropy over all scales.
2. The maximization should be done with the inequality constraint that the information gain $\psi^{(k)}$ is a non-increasing function of scale k.

Point 1 indicates that all time scales, rather than a single dominant scale, are important for entropy optimization. The compound measure could be the sum of entropies on all scales, from 1 to infinity. Numerically, this could be approximated by the sum of a finite number (m) of entropy terms, given that as time scale tends to infinity the entropy will tend to zero (since the probability dry also tends to zero). Thus, if m is large then results will not be sensitive to m. The postulate in point 2, which is an essential element of the ME procedure, has been put to avoid physically unreasonable solutions. For, violation of the non-increasing information gain $\psi^{(k)}$ would lead to illogical results in terms of predictability of the future. For example, considering scales 1 and 2, violation of $\psi^{(1)} \geq \psi^{(2)}$ would mean that observation of the present and past gives more information for the future prediction on two time steps ahead (i.e. $\psi^{(2)}$) than it gives for prediction of one time step ahead (i.e. $\psi^{(1)}$).

This framework requires the evaluation of the quantities $\varphi^{(k)}$, $\varphi_c^{(k)}$ and $\psi^{(k)}$ for a range of scales extending from 1 to a large number m (theoretically infinite). An analytical solution for this setting seems to be too difficult or even intractable. For a numerical solution, assuming that $\varphi_c^{(k)}$ is estimated for order q, the problem will depend on a number $2^{(q-1)m}$ of terms p at the basic scale. If the scales of interest extend to about $m = 3 \times 10^3$, as happens in the problem examined here, and if $q = 8$ provides an acceptable accuracy, then a number 10^{6300} of terms p is needed. In an entropy maximization framework all these terms are unknowns whose values are to be estimated by maximizing entropies $\varphi^{(k)}$ conditional on non increasing $\psi^{(k)}$. Such a maximization, however, is practically intractable. The variables p are too many and the quantities $\varphi^{(k)}$ and $\psi^{(k)}$ are nonlinear functions of the terms p, where the nonlinearity is introduced from equations (31) and (32).

A practical approximate solution that was followed here includes the following steps:

(a) Separate the effect of the probability dry $p^{(k)}$ at any scale k from all other probabilities; as implied by (31), $p^{(k)}$ determines completely the unconditional entropy $\varphi^{(k)}$ at each scale.

(b) Introduce a parametric expression of $p^{(k)}$, i.e. a mathematical expression of $p^{(k)}$ involving a number of parameters (as done in a similar situation in *Koutsoyiannis* [2005b]).

(c) For known parameters of this parametric expression, estimate $\varphi^{(k)}$ and then all involved probabilities p so as to maximize entropies $\varphi_q^{(k)}$ for progressing orders q.

(d) Given all probabilities, estimate, $\varphi_c^{(k)}$ and $\psi^{(k)}$ for all scales of interest.

(e) Locate the optimal parameters of the expression of $p^{(k)}$ by repeating steps (c) and (d) for trial values of parameters determined by a nonlinear optimization algorithm, so as to maximize the sum of $\varphi^{(k)}$ over all scales subject to the constraint of non increasing $\psi^{(k)}$.

Obviously, this technique cannot assure locating the maximum of entropy nor can it provide a mathematical proof for the derived ME solution. However, it can yield an approximation of the ME solution.

For step (b), several different expressions for $p^{(k)}$ were devised and tried; here only one will be discussed to reduce the paper length. This involves four parameters, i.e.

$$p^{(k)} = \left\{1 + (p^{-s} - 1)[1 + (\xi^{-1/\eta} - 1)(k-1)]^{\eta}\right\}^{-1/s} \tag{40}$$

where the parameter $s \geq 0$ and the other parameters p, η and ξ range in the interval [0, 1]. It can be directly verified that $p^{(1)} = p$ for any η, s and ξ; thus the parameter p is identical to the probability dry at the basic scale (hence the symbol p). To allow backward extendibility, i.e. extrapolation of the expression for time scales smaller than the basic scale down to zero (which is meaningful in the case of the rainfall occurrence process), parameters ξ and η must satisfy the inequality $1 - (\xi^{-1/\eta} - 1) \geq 0$, or $\xi \geq 1/2^{\eta}$.

Clearly, for $s > 0$, (40) implies a hyper-exponential (power-type) tail of the probability $p^{(k)}$ as scale k increases. In the special case $s = 0$ application of de l'Hôpital's rule results in

$$p^{(k)} = p^{[1 + (\xi^{-1/\eta} - 1)(k-1)]^{\eta}} \tag{41}$$

Thus, when $s = 0$ the tail is of exponential type (more specifically it could be called Weibull type as the exponent increases as a power of k; interestingly a resembling type was proposed by *De Bruin* [1980]). Equation (41) can be also written as

$$-\ln p^{(k)} = -\ln p\, [1 + (\xi^{-1/\eta} - 1)(k-1)]^{\eta} \tag{42}$$

If the probability dry is known at two scales k and l then parameter ξ can be estimated from

$$\xi = \left[\frac{(l-1)\alpha^{1/\eta} - (k-1)}{(l-2)\alpha^{1/\eta} - (k-2)}\right]^{\eta} \tag{43}$$

where

$$\alpha := \frac{[p^{(k)}]^{-s} - 1}{[p^{(l)}]^{-s} - 1}, \text{ if } s > 0; \qquad \alpha := \frac{-\ln p^{(k)}}{-\ln p^{(l)}}, \text{ if } s = 0 \tag{44}$$

It can be observed that if $k = 1$ and $l = 2$ (which is the case in our optimization framework, since p and $p^{(2)}$ are put as constraints) then $\xi = \alpha$; if, in addition, $s = 0$, then $\xi = \tau$.

For $s = 0$ and $\eta = 1$, the resulting dependence structure is Markovian, as verified by comparing the expressions (21) and (42). For $s = 0$, $\eta = 1$ and $\xi = 0.5$, the independence model emerges, as verified by comparing the expressions (16) and (42). For $s = 0$, $\eta < 1$ and large k, the resulting $-\ln p^{(k)}$ from (42) becomes proportional to k^{η}. This agrees with (18) and signifies a relatively heavy tail of $p^{(k)}$, which becomes even heavier if $s > 0$. For $s \to \infty$ and ξ determined from (43), it can be shown that $p^{(k)} \to p^{(2)}$, i.e. the probability dry becomes constant for any $k \geq 2$. For $s \to \infty$ and $\xi = 1$, (40) results in the heaviest possible tail since $p^{(k)} = p$, i.e. constant for any $k \geq 1$, which corresponds to complete dependence. Thus, parametric expression (40) can yield a very rich family of shapes, which is useful for our purpose.

For step (c), it is observed that four elements of the vectors $\mathbf{p}_q$ for each q are directly estimated from (34), given $p^{(q)}$ and the elements of the vector $\mathbf{p}_{q-1}$. For all other elements, which do not affect the unconditional entropy $\varphi^{(k)}$, it will be conjectured that their estimation can be based on the single scale optimization approach of section 7. This is equivalent to assuming Markovian structure for all other probabilities and in this sense the model constructed in this way will be called quasi-Markov. Thus, step (c) becomes easy to apply, as recursive application of (38) or (39) yields all unconditional or conditional probabilities.

To evaluate the vector $\mathbf{p}_q^{(k)}$ at scale k, normally we need to calculate the vector $\mathbf{p}_{k\,q}$ at the basic scale, whose size is $2^{k\,q}$. To avoid such exceptionally sized vectors, in the numerical approach followed, some simplifications were done. First, the entropies of interest were evaluated for geometrically progressing scales, i.e., 1, 2, 4, 8, …. Secondly, at each scale k the vectors $\mathbf{p}_q^{(k)}$ for $q = 1$ to 8 were evaluated only. Given the vectors $\mathbf{p}_q^{(k)}$ at scale k, the first four vectors $\mathbf{p}_q^{(2\,k)}$ (for scale $2\,k$ and for $q = 1$ to 4) were calculated directly from $\mathbf{p}_q^{(k)}$ and the remaining four (for $q = 5$ to 8) were calculated with the same manner as the vectors at the

basic scale (by the quasi-Markov assumption). Obviously, this introduces inaccuracies, which are not very significant, as indicated by preliminary experiments.

These ideas are illustrated with some examples depicted in Figure 4, in terms of log-probability dry ($-1\mathrm{n}\, p^{(k)}$), unconditional entropy ($\varphi^{(k)}$) and conditional entropy ($\varphi_{\mathrm{c}}^{(k)}$), and information gain ($\psi^{(k)}$) versus scale k for several assumed dependence structures. The first three examples (1, 2 and 3) correspond to the independence case with different basic probability dry p (0.5, 0.4 and 0.8, respectively). In each of the cases, the conditional and unconditional entropy are identical to each other and the information gain is zero (due to independence). Among the three cases, $p = 0.5$ (equiprobability) results in the highest entropy at the basic scale. However, the case $p = 0.8$ results in higher entropy at all other scales. So, equiprobability does not necessarily lead to ME if we consider more than one scale. In all cases, entropy is a decreasing function of scale, quickly falling off to zero.

The next three examples (4, 5 and 6) are similar to 1 in terms of the basic probability dry, which is $p = 0.5$, but assume some dependence as in all cases $p^{(2)}$ was set 0.45, much higher than $p^2 = 0.25$ that corresponds to the independence case. Example 4 corresponds to the Markov case ($\eta = 1$), example 5 to the non-Markov case (41) with $\eta = 0.6$ and $s = 0$, and example 6 to the non-Markov case (40) with $\eta = 1$ and $s = 6.5$; this is the highest possible value of s that allows backward extendibility of (40). As we move from case 1 to 4 through 6, the tails of the probability dry $p^{(k)}$ and entropies $\varphi^{(k)}$ and $\varphi_{\mathrm{c}}^{(k)}$ become fatter. Interestingly, however, case 6, which yields the highest unconditional entropy at all scales, simultaneously results in the least conditional entropy $\varphi_{\mathrm{c}}^{(k)}$ for scales up to about 40. Furthermore, case 6 results in the highest information gain $\psi^{(k)}$. At scales 1 and 2, all cases 4-6 result in precisely the same unconditional entropy $\varphi^{(k)}$, and the Markov case 4 results in the highest conditional entropy $\varphi_{\mathrm{c}}^{(k)}$. In contrast, at the largest scales the ultimate non-Markov case 6 results in the highest unconditional and conditional entropies $\varphi^{(k)}$ and $\varphi_{\mathrm{c}}^{(k)}$. It can be observed that case 6 results in increasing information gain for scales 1-4, and thus it cannot be a physically realistic model as explained earlier.

As a means to test the sensitivity of the entropies $\varphi^{(k)}$ and $\varphi_{\mathrm{c}}^{(k)}$ and the information gain $\psi^{(k)}$ on the assumption used for determination of the elements of the vectors $\mathbf{p}_q^{(k)}$ (apart from the

first elements of the vectors i.e. the probabilities dry $p^{(k)}$ which are determined from (40)) some other non Markovian options were also tried, one of which is described in the Appendix (electronic data supplement). The investigation showed that (a) the deviation of probability dry $p^{(k)}$ (first element of $\mathbf{p}_q^{(k)}$) from the Markov case is more significant than is the Markovian or non-Markovian structure of the remaining probabilities (other elements of $\mathbf{p}_q^{(k)}$), and (b) the resulting maximized parameters of (40) that determine probability dry $p^{(k)}$ are almost the same for all examined options used to estimate the remaining probabilities. This increases the confidence that the solutions obtained by the above procedure are close to maximum entropy solutions.

9. Application of the multiple scale ME framework

For known p and $p^{(2)}$ (i.e. equal to the observed frequencies of dry intervals) and assuming quasi-Markov structure, it is easily seen that the parametric model (41) depends only on two parameters, i.e. η and s, since parameter p is known and parameter ξ is estimated from (43) as a function of p, $p^{(2)}$ and s. Thus the optimization problem includes only two control variables (free parameters) and it can be easily solved using a nonlinear solver; here a widespread solver by Frontline Systems (www.solver.com) combining classical and evolutionary optimization techniques was used. The objective function was the sum of $\varphi^{(k)}$ over scales $k = 1$ to 2^{13} (=8192), which depends on $p^{(k)}$ and eventually is a function of η and s. The evaluation of the function is not difficult but can be done only numerically. More complicated is the computation of the information gain for each scale, which is required to evaluate the constraint of non-increasing $\psi^{(k)}$. This is done numerically too in the manner described in section 6 (equation (32) and below).

The application of this framework with the Athens data set was done on monthly, seasonal and annual basis; for brevity only the latter case is described in full detail. The observed values of p and $p^{(2)}$ are respectively 0.945 and 0.933 (at hourly basic scale). The maximization procedure resulted in parameter values $s = 0$ and $\eta = 0.63$. The obtained log-probability dry ($-1\mathrm{n}\, p^{(k)}$) , unconditional entropy ($\varphi^{(k)}$), conditional entropy ($\varphi_c^{(k)}$), and information gain ($\psi^{(k)}$) versus scale k are shown in Figure 5. The ME solution is marked as case 2; another four cases

are also shown for comparison, clarification and explanation of the reason why case 2 is the optimal solution. Case 1 is the Markov case ($\eta = 1$, $s = 0$), which for large scales corresponds to entropy much lower than in the ME case. In both cases 1 and 2 the constraint of non-increasing $\psi^{(k)}$ is satisfied as shown in the last panel of the figure. In case 3 we assumed again $\eta = 1$ as in case 1, but we increased s from 0 to 0.5. The resulted sum of entropy over all scales increased but the information gain became an increasing function of scale, for scales 16-128. The relevant constraint was violated and thus case 3 is not an acceptable solution. Cases 4 ($\eta = 0.55$, $s = 0$) and 6 ($\eta = 0.63$, $s = 0.5$) correspond to changes of each the parameters of the ME solution (case 2). Both cases 5 and 6 result in higher sum of entropy than in case 2. However, both violated the constraint on non-increasing information gain. Thus, they too are unacceptable solutions.

Figure 6 depicts the probability dry $p^{(k)}$ versus scale k for the above ME solution (case 2). Apart from the two empirical data values p and $p^{(2)}$ that were used to estimate the ME solution, all other empirical data values of $p^{(k)}$ for scales k up to 3072 (128 days) are also plotted in Figure 6. The agreement of the empirical data with the theoretical ME model is very good for all this wide range of scales, from which only two values were used to fit the model. As an additional model validation, the entropy maximization was repeated assuming as basic scale the daily one (so that p and $p^{(2)}$ correspond to one and two days). The resulting ME solution had negligible deviations from the case 2 that was estimated for the hourly scale. Apart from the annual plots, Figure 6 depicts also similar plots for the dry season. The entropy maximization was done in the same manner as in the annual case. The agreement of theoretical model with data is again impressive.

Furthermore, Figure 7 shows comparisons in terms of the lag one autocorrelation coefficients $\rho^{(k)}$ and dependence indicators $\tau^{(k)}$, for the entire year and the dry season. This is similar to Figure 2 except the ME results were plotted in addition. It can be shown that equation (41) implies a $\rho^{(k)}$ which, as $k \to \infty$, tends to zero exponentially (more precisely, it tends to zero as $p^{-(1-2^r)(\xi^{-1/\eta}-1)^\eta k^\eta}$) and $\tau^{(k)}$ tends to the constant value $\tau_{\min} = 2^{-\eta}$, which for $\eta < 1$ is greater than ½. In both cases, the ME results are in good agreement with real data.

Furthermore, Figure 8 provides comparisons of the log probability dry $p^{(k)}$ versus scale k, for the entire year, the dry season, as well as for four months. This is similar to Figure 3 except the ME results were plotted in addition to those of Markov and independence cases. In all cases the parameter s was found to be close to zero. Small deviations of s from zero could be attributed to numerical inaccuracies. Besides, they could not be verified with empirical data as their effect becomes significant only on very large time scales, practically out of the range of empirical points. Therefore, it was conjectured that the exponential form (41) (rather than the hyper-exponential form of (40)) applies. The parameters η obtained by the maximization procedure for $s = 0$ are shown in Table 1. As shown in Figure 8, in all cases, the ME results are in good agreement with real data, better than in the other models. Some departures appearing in some of the very last points may be due to the fact that the empirical values have been estimated from one or two samples only. The ME model captures the long term persistence, i.e., the theoretical slope (which is equal to η) agrees with the slope of the arrangement of empirical points. However, in January the model underestimates the slope of the arrangement of empirical points.

As an additional validation, the probabilities of having exactly n out of q wet days in a sequence of q days were estimated from (4) for the three theoretical models (which were fitted on the hourly scale) as well as empirically from the data. The results for $q = 2$, 4 and 8 are intercompared in Figure 9 for four months and for the entire year. The independence and Markov models (especially the former) do not agree with empirical data, whereas the ME model is more consistent with reality. Some overestimations appear for the longest wet sequences (i.e. $n = q$) and some more discrepancies appear in January.

10. Discussion

Even though the purpose of this study is exploratory and explanatory, the results found may be useful for several modeling and engineering purposes. The fact that the ME principle applied with only two empirically estimated parameters is able to describe adequately the dependence structure of the rainfall occurrence process for time scales extending to more than three orders of magnitude may help especially in parameter estimation of models and in

inferring statistical properties from observed to non observed scales. For example, in frequently met cases where only daily data are available, the ME principle can provide estimates of properties of the rainfall occurrence process at sub-daily scales.

The theoretical framework of sections 6-8 can be directly utilized in a simulation framework as it can assign transition probabilities of a binary chain model of any order, based on the sequence of probabilities dry. This chain model can then be used for simulation of rainfall occurrences directly. For a complete rainfall generator, an additional model for rainfall depth is needed. The two models may cooperate in a hybrid scheme as for instance in the study of *Gyasi-Agyei and Willgoose* [1997]. The ME principle can also help in building a model for rainfall depth, as demonstrated in *Koutsoyiannis* [2005a, b].

Particularly, the results of this study could be useful to disaggregation of rainfall into finer time scales. This concerns both the parameter estimation phase, as explained before, as well as the simulation phase. For the latter, an adaptation of the ideas studied in *Koutsoyiannis* [2001, 2002] related to the linking of the current period with previous and next periods and in *Glasbey et al.* [1995] and *Koutsoyiannis and Onof* [2001] related to the adjusting of some initially generated values will be helpful.

In addition, engineering applications such as the construction of consistent rainfall intensity/duration/frequency relationships would benefit from the ME framework. These determine the rainfall intensity $i(k, T)$ averaged over time scale k and exceeded on a return period T (i.e. with probability $1/T$ on an annual basis). Therefore, they should be more correctly called intensity/timescale/return-period relationships (the term duration is misleading). Typically, it is assumed that

$$i(k, T) = a(T)\, b(k) \tag{45}$$

where $a(T)$ and $b(k)$ are mathematical expressions, such as

$$a(T) = T^{\kappa},\ \ b(k) = (k + \theta)^{-\eta} \tag{46}$$

historically established with empirical considerations. Recently, it was attempted to give them a more theoretical foundation, based on scaling assumptions and multifractal analysis [e.g.

Veneziano and Furcolo, 2002]. However, probabilistic considerations combined with the ME principle may derive such expressions in an easy manner and simultaneously provide physically and mathematically sound explanations for the relationships. Specifically, if $Y^{(k)}$ is a random variable representing the rainfall depth at scale k, application of the ME principle in *Koutsoyiannis* [2005b] showed that this variable should have Pareto-type distribution, thus verifying and explaining an earlier investigation [*Koutsoyiannis*, 2004a, b] based on a large database of rainfall maxima worldwide. The form of $a(T)$ is then derived in the manner described in *Koutsoyiannis et al.* [1998] and is slightly different from (46) (i.e. $a(T) = T^{\kappa} - \psi$).

Now, if μ is the mean of the average intensity $Y^{(k)}/k$, which is independent of scale because $E[Y^{(k)}] \sim k$, and $\mu^{*(k)}$ is the mean of the same quantity conditional on being wet, then obviously $\mu^{*(k)} = \mu/(1 - p^{(k)})$. Assuming some similarity over scales in the distribution of $Y^{(k)}$ conditional on being wet (so that (45) should have meaning), it may be concluded that $b(k) \sim \mu^{*(k)}$. If the probability dry $p^{(k)}$ is expressed as in (41) and we approximate it taking the first two terms of its Taylor expansion (assuming small scales in order for the higher order terms of the Taylor expansion to be negligible),

$$p^{(k)} = p^{\,[(k+\theta)/(1+\theta)]^{\eta}} = 1 + [(k+\theta)/(1+\theta)]^{\eta} \ln p + \ldots \tag{47}$$

where $\theta := (2 - \xi^{-1/\eta}) / (\xi^{-1/\eta} - 1)$, then we conclude that for small scales, indeed, $b(k) \sim 1/(1 - p^{(k)}) \sim (k + \theta)^{-\eta}$ as in (46). This gives a sufficient explanation for (46). It is important that this analysis preserves the term θ, thus validating the common engineering practice for small scales, which are important for the design of urban drainage networks that have small concentration times; other methodologies based on scaling assumptions may fail to include this term and result in pure power-law behavior. It should be noted, however, that this analysis is done here on an explanatory basis. In an operational basis, the problem needs to be addressed in a more thorough and comprehensive manner, given that the assumption of similarity over scales is not valid in reality. This will be reported in another occasion.

To proceed from the explanatory character of this study to an operational utilization of its results, several research steps have to be accomplished with possible objectives:

- to develop a more accurate method of entropy maximization, given that the approach presented is only approximate; as the maximum entropy approach is essentially a statistical physics approach, numerical methods (mostly Monte Carlo methods) already established in statistical physics, such as the so called simulated annealing and the Ising model (e.g. Binder and Heermann, 1992; Newman and Barkema, 2002) should be explored;
- to test the methodological framework against a large empirical base including rainfall data sets from a wide range of climates worldwide;
- to proceed from the exploratory to a more operational character of the method, e.g. adapting the parameter estimation procedure using more than two probabilities dry, and including additional constraints referring for instance to probabilities wet.
- to integrate the rainfall occurrence model with a model of rainfall depth (as discussed above) so as to enable rainfall simulation studies, rainfall disaggregation and even consistent derivation of intensity/timescale/return-period relationships; and
- to study estimation uncertainties and especially those implied from uncertainty in the characterization of a time interval as dry or wet based on the recorded rainfall depth [*Koutsoyiannis et al.*, 2003].

11. Summary and conclusion

The well-established physical and mathematical principle of maximum entropy (ME), representing maximum uncertainty, is used to explain the observed dependence properties of the rainfall occurrence process, including the clustering behavior and persistence.

The results of the application of the ME principle can be summarized as follows:

(a) Maximum entropy + Dominance of a single time scale + Time dependence $\longrightarrow$ Markovian time dependence

(b) Maximum entropy + Equivalence of time scales + Time dependence $\longrightarrow$ Non-Markovian dependence (long-term persistence)

Simple conditions (optimization constraints) were used for entropy maximization, i.e. that the rainfall processes is intermittent with dependent occurrences. Intermittency is quantified by the probability that an interval of a certain length (basic scale) is dry, and dependence is quantified by the probability that two consecutive intervals of same length are dry. In addition, in case (b) an inequality constraint was used that the information gain when past and present is observed should be a non-increasing function of time scale, which assures physical reasonability. These constraints involve only two empirically estimated parameters, which along with the ME principle suffice to determine any conditional or unconditional probability of any sequence of dry and wet intervals at any time scale. The dependence structure in case (b) appears to be non-Markovian but not hyper-exponential (power type) in terms of probability dry versus time scale.

Application of this theoretical framework to the rainfall data set of Athens indicates good agreement of theoretical predictions and empirical data at the entire range of scales for which probabilities dry and wet can be estimated (from one hour to several months). The attitude of this application is explanatory rather than operational, even though a number of operational applications could emerge after further research.

The general conclusion of this study is the dominance of maximum entropy, or maximum uncertainty, in the rainfall occurrence process. A recent study [*Koutsoyiannis*, 2005a, b] indicated the dominance of the same principle in other hydrological and meteorological processes, and provided explanations for marginal distributional properties (from Gaussian to power-type) of variables and for the long-range dependence structure of processes (Hurst phenomenon). Thus, the results of the present study harmonize with more general results of the applicability of the ME principle in natural processes.

Acknowledgements The data for Athens became available courtesy of the National Observatory of Athens. Calculations were facilitated by numerical methods developed and made freely available by the Foxes team (http://digilander.libero.it/foxes/index.htm). The thoughtful, constructive and encouraging review comments by Christian Onof, two anonymous reviewers and the Associate Editor Richard Katz, which helped improve the analyses and presentation, are gratefully appreciated.

References

Beran, J., *Statistics for Long-Memory Processes*, Volume 61 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, New York, 1994.

Binder, K., and D. W. Heermann, *Monte Carlo Simulation in Statistical Physics*, Springer-Verlag, 1992.

Chin, E. H., Modelling daily precipitation occurrence process with Markov chains, *Water Resour. Res.*, 13(6), 959-956, 1977.

Clarke, R. T., *Stochastic Processes for Water Scientists*, Wiley, New York, 1998.

De Bruin, H. A. R., A stochastic description of wet and dry spells in terms of an effective number of days, *J. of Hydrol.*, 45, 91-99, 1980.

Entekhabi, D., I. Rodriguez-Iturbe and P. S. Eagleson, Probabilistic representation of the temporal rainfall process by the modified Neyman-Scott rectangular pulses model: Parameter estimation and validation, *Water Resour. Res.*, 25(2), 295-302, 1989.

Feyerherm, A. M., and L. D. Bark, Statistical methods for persistent precipitation patterns, *J. Appl. Meteorol.*, 4, 320-328, 1964.

Foufoula-Georgiou, E., and P. Guttorp, Compatibility of continuous rainfall occurrence models with discrete rainfall observations, *Water Resour. Res.*, 22(8), 1316-1322, 1986.

Foufoula-Georgiou, E., and D. P. Lettenmaier, Continuous-time versus discrete-time point process models for rainfall occurrence series, *Water Resour. Res.*, 22(4), 531-542, 1986.

Gabriel, K. R., and J. Neumann, A Markov chain model for daily rainfall occurrences at Tel Aviv, Royal Meteorol. Soc. Quart. J., 88, 90-95, 1962.

Georgii, H.-O., Probabilistic aspects of entropy, in *Entropy* (edited by A. Greven, G. Keller & G. Warnecke), Princeton University Press, Princeton, NJ, 2003.

Glasbey, C. A., G. Cooper, and M. B. McGechan, Disaggregation of daily rainfall by conditional simulation from a point-process model, *Journal of Hydrology*, 165(1-4), 1-9, 1995

Greven, A., G. Keller & G. Warnecke Georgii, H.-O. (editors), *Entropy*, Princeton University Press, Princeton, NJ, 2003.

Gyasi-Agyei, Y., and G. R. Willgoose, A hybrid model for point rainfall modeling, *Water Resour. Res.*, 33(7) 1699-1706, 1997.

Haan, C. T., D. M. Allen, and J. O. Street, A Markov chain model of daily rainfall, *Water Resour. Res.*, 12(3), 443-449, 1976.

Harrold, T. I., A. Sharma, and S. J. Sheather, A nonparametric model for stochastic generation of daily rainfall occurrence, *Water Resour. Res.*, 39(10), 1300, doi:10.1029/2003WR002182, 2003.

Hershfield, D. M., A comparison of conditional and unconditional probabilities for wet- and dry-day sequences, J. Appl. Meteorol., 9(5), 825-827, 1970.

Hutchinson, M. F., A point rainfall model based on a three-state continuous Markov occurrence process, *J. of Hydrol.*, 114, 125-148, 1990.

Jaynes, E. T., Information theory and statistical mechanics, *Physical Review* 106(4):620-630, 1957.

Kavvas, M. L., and J. W. Delleur, A stochastic cluster model of daily rainfall sequences, *Water Resour. Res.*, 17(4), 1151-1160, 1981.

Katz, R.W., An Application of Chain-dependent Processes to Meteorology, *J. Appl. Prob.*, 14, 598-603, 1977a.

Katz, R.W., Precipitation as a Chain-dependent process. *J. Appl. Meteorol.*, 16, 671-676, 1977b.

Katz, R.W., and M.B. Parlange, Overdispersion Phenomenon in Stochastic Modeling of Precipitation, *J. Climate*, 11, 591-601, 1998.

Keane, M., Entropy in ergodic theory, in *Entropy* (edited by A. Greven, G. Keller & G. Warnecke), Princeton University Press, Princeton, NJ, 2003.

Klemes, V., The Hurst phenomenon: A puzzle?, *Water Resour. Res.*, 10(4) 675-688, 1974.

Kottegoda, N. T., and Horder, M. A., Daily flow model based on rainfall occurrences using pulses and a transfer function, *J. of Hydrol.*, 47, 215-234, 1980.

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

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

Koutsoyiannis, D., Statistics of extremes and estimation of extreme rainfall, 1, Theoretical investigation, *Hydrol. Sci. J.*, 49(4), 575-590, 2004a.

Koutsoyiannis, D., Statistics of extremes and estimation of extreme rainfall, 2, Empirical investigation of long rainfall records, *Hydrol. Sci. J.*, 49(4), 591-610, 2004b.

Koutsoyiannis, D., Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling, *Hydrol. Sci. J.*, 50(3), 381-404, 2005a.

Koutsoyiannis, D., Uncertainty, entropy, scaling and hydrological stochastics, 2, Time dependence of hydrological processes and time scaling, *Hydrol. Sci. J.*, 50(3), 405-426, 2005b.

Koutsoyiannis, D., and C. Onof, Rainfall disaggregation using adjusting procedures on a Poisson cluster model, *J. of Hydrol.*, 246, 109-122, 2001.

Koutsoyiannis, D., D. Kozonis, and A. Manetas, A mathematical framework for studying rainfall intensity-duration-frequency relationships, *J. of Hydrol.*, 206(1-2), 118-135, 1998.

Koutsoyiannis, D., C. Onof, and H. S. Wheater, Multivariate rainfall disaggregation at a fine time scale, *Water Resources Research*, 39(7), 1173 , 1-18, 2003.

Marshak, A., A. Davis, R. Cahalan, and W. Wiscombe, Bounded cascade models as nonstationary multifractals, *Phys. Rev. E*, 49(1), 55-69, 1994.

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

Müller, I., Entropy: a subtle concept in thermodynamics, in *Entropy* (edited by A. Greven, G. Keller & G. Warnecke), Princeton University Press, Princeton, NJ, 2003.

Newman M. E. J., and G. T. Barkema, *Monte Carlo Methods in Statistical Physics*, Oxford University Press, New York, 2002.

Neyman, J., and E. L.Scott, A theory of the spatial distribution of galaxies, *Astrophysical Journal*, 116, 144-163, 1952.

Olsson, J., Scaling and fractal properties of rainfall, PhD Thesis, University of Lund, Sweden, 1996.

Onof, C. and H. S. Wheater, Modelling of British rainfall using a Random Parameter Bartlett-Lewis Rectangular Pulse Model, *J. Hydrol.*, 149, 67-95, 1993.

Onof, C. and H. S. Wheater, Improvements to the modeling of British rainfall using a Modified Random Parameter Bartlett-Lewis Rectangular Pulses Model, *J. Hydrol.*, 157, 177-195, 1994.

Onof, C., R. E. Chandler, A. Kakou, P. Northrop, H. S. Wheater, and V. Isham, Rainfall modelling using Poisson-cluster processes: a review of developments, *Stochastic Environmental Research and Risk Assessment*, 14, 384-411, 2000.

Ozawa, H., A. Ohmura, R. D. Lorenz, and T. Pujol, The second law of thermodynamics and the global climate system: A review of the maximum entropy production principle, *Rev. Geophys.*, 41(4), 1018, doi:10.1029/2002RG000113, 2003.

Papoulis, A., Probability, Random Variables, and Stochastic Processes (third edn), 57. McGraw-Hill, New York, 1991.

Pegram, G. G. S., An auto-regressive model for multi-lag Markov chains, *J. Appl. Probab.*, 17, 350–362, 1980.

Pegram, G. G. S. and A. W. Seed, The feasibility of stochastically modelling the spatial and temporal distribution of rainfields, WRC Report No. 550/1/98 to the Water Research Commission, Pretoria, South Africa, 1998.

Rodriguez-Iturbe, I., V. K. Gupta, and E. Waymire, Scale considerations in the modeling of temporal rainfall, *Water Resour. Res.*, 20(11), 1611-1619, 1984.

Rodriguez-Iturbe, I., D. R. Cox, and V. Isham, Some models for rainfall based on stochastic point processes, Proc. R. Soc. Lond., A 410, 269-298, 1987.

Rodriguez-Iturbe, I., D. R. Cox, and V. Isham, A point process model for rainfall: Further developments, Proc. R. Soc. Lond., A 417, 283-298, 1988.

Roldan, J., and D. A. Woolhiser, Stochastic daily precipitation models, 1, A comparison of occurrence processes, *Water Resour. Res.*, 18(5), 1451-1459, 1982.

Sharma, A., and R. O'Neill, A nonparametric approach for representing interannual dependence in monthly streamflow sequences, *Water Resour. Res.*, 38(7), 1100, doi:10.1029/2001WR000953, 2002.

Small, M. J., and D. J. Morgan, The relationship between a continuous-time renewal model and a discrete Markov chain model of precipitation occurrence, *Water Resour. Res.*, 22(10), 1422-1430, 1986.

Srikanthan, R., and T.A. McMahon, Stochastic generation of annual, monthly and daily climate data: A review, *Hydrology and Earth System Sciences*, 5(4), 653–670, 2001.

Stern, R. D., and R. Coe, A model fitting analysis of daily rainfall data, *J. Roy. Statist. Soc.*, A147, 1-34, 1984.

Todorovic, P., and D.A. Woolhiser, A Stochastic Model of n-day Precipitation, *J. Appl. Meteorol.*, 14, 17-24, 1975.

Tsallis, C., Nonextensive statistical mechanics: construction and physical interpretation, in *Nonextensive Entropy, Interdisciplinary Applications* (edited by M. Gell-Mann & C. Tsallis), Oxford University Press, New York, NY, 2004.

Velghe, T., P. A. Troch, F. P. De Troch, and J. Van de Velde, Evaluation of cluster-based rectangular pulses point process models for rainfall, *Water Resour. Res.*, 30(10) 2847-2857, 1994.

Veneziano, D., and P. Furcolo, Multifractality of rainfall and scaling of intensity-duration-frequency curves, *Water Resour. Res.*, 38(12), 1306, doi:10.1029/2001WR000372, 2002.

Verhoest, N., P. A. Troch, and F. P. De Troch, On the applicability of Bertlett-Lewis rectangular pulses models in the modeling of design storms at a point, *J. of Hydrol.*, 202, 108-120, 1997.

Waymire, E., and V. K. Gupta, The mathematical structure of rainfall representations, 1, A review of the stochastic rainfall models, *Water Resour. Res.*, 17(5), 1261-1272, 1981a.

Waymire, E., and V. K. Gupta, The mathematical structure of rainfall representations, 2, A review of the theory of point processes, *Water Resour. Res.*, 17(5), 1273-1285, 1981b.

Waymire, E., and V. K. Gupta, The mathematical structure of rainfall representations, 3, Some applications of the point process theory to rainfall processes, *Water Resour. Res.*, 17(5), 1287-1294, 1981c.

Table 1 Characteristic indices of the rainfall occurrence process in Athens, Greece, for representative months, for the dry season and for the entire year.

Period of year	Probability dry at scale 1 h, p	Probability dry at scale 2 h, $p^{(2)}$	Dependence indicator $\tau = -\ln p / -\ln p^{(2)}$	Exponent η of probability dry in expression (41)
January	0.891	0.865	0.796	0.52
May	0.964	0.953	0.762	0.72
August	0.995	0.993	0.785	0.88
October	0.940	0.926	0.805	0.62
Dry season	0.989	0.986	0.801	0.83
Year	0.945	0.933	0.816	0.63

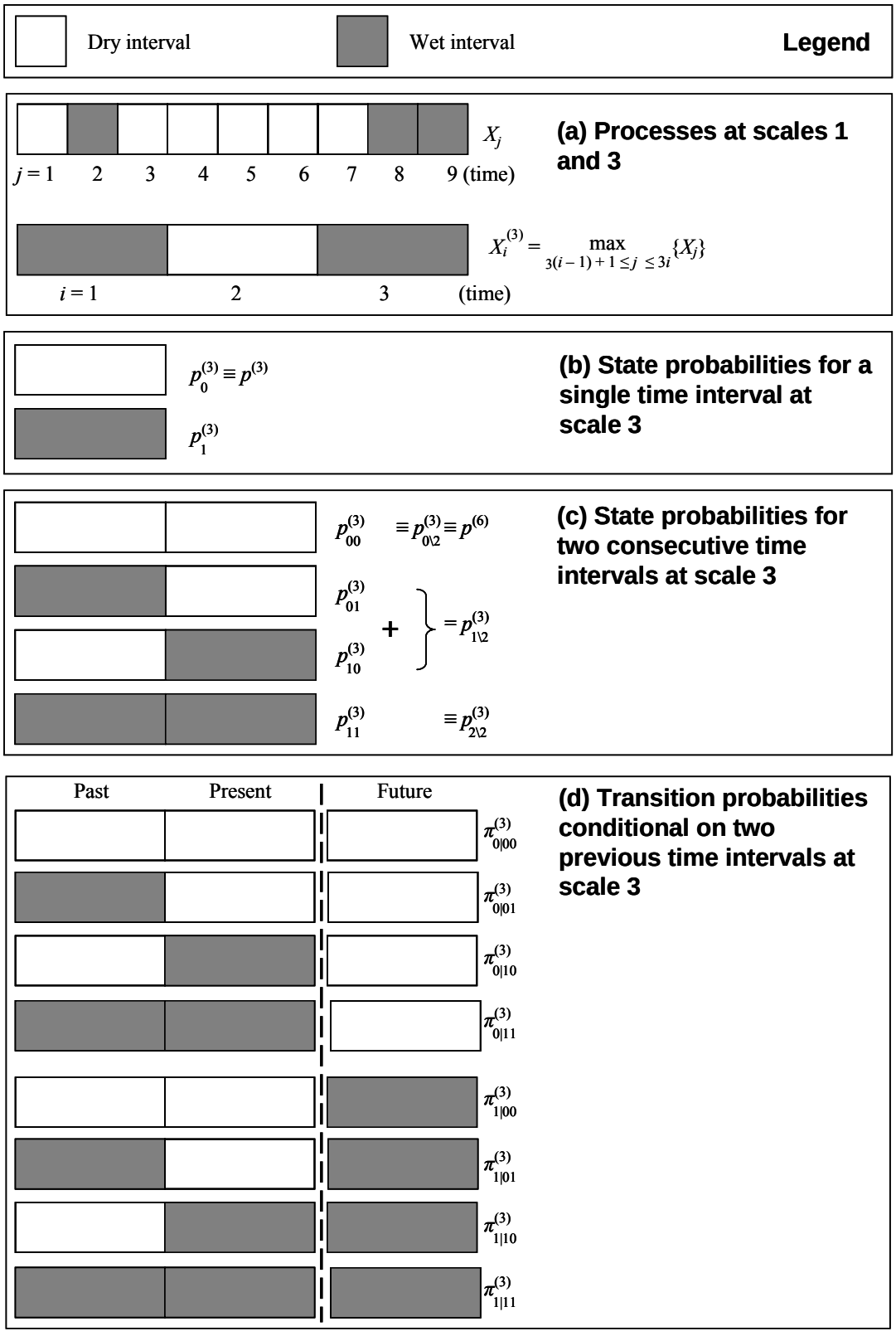

Figure 1 Sketch demonstrating the notation of processes and probabilities. Notice that the order of subscripted indices of probabilities is from most recent to oldest, which is inverse to that of the visual sequence of wet and dry intervals (from oldest to most recent).

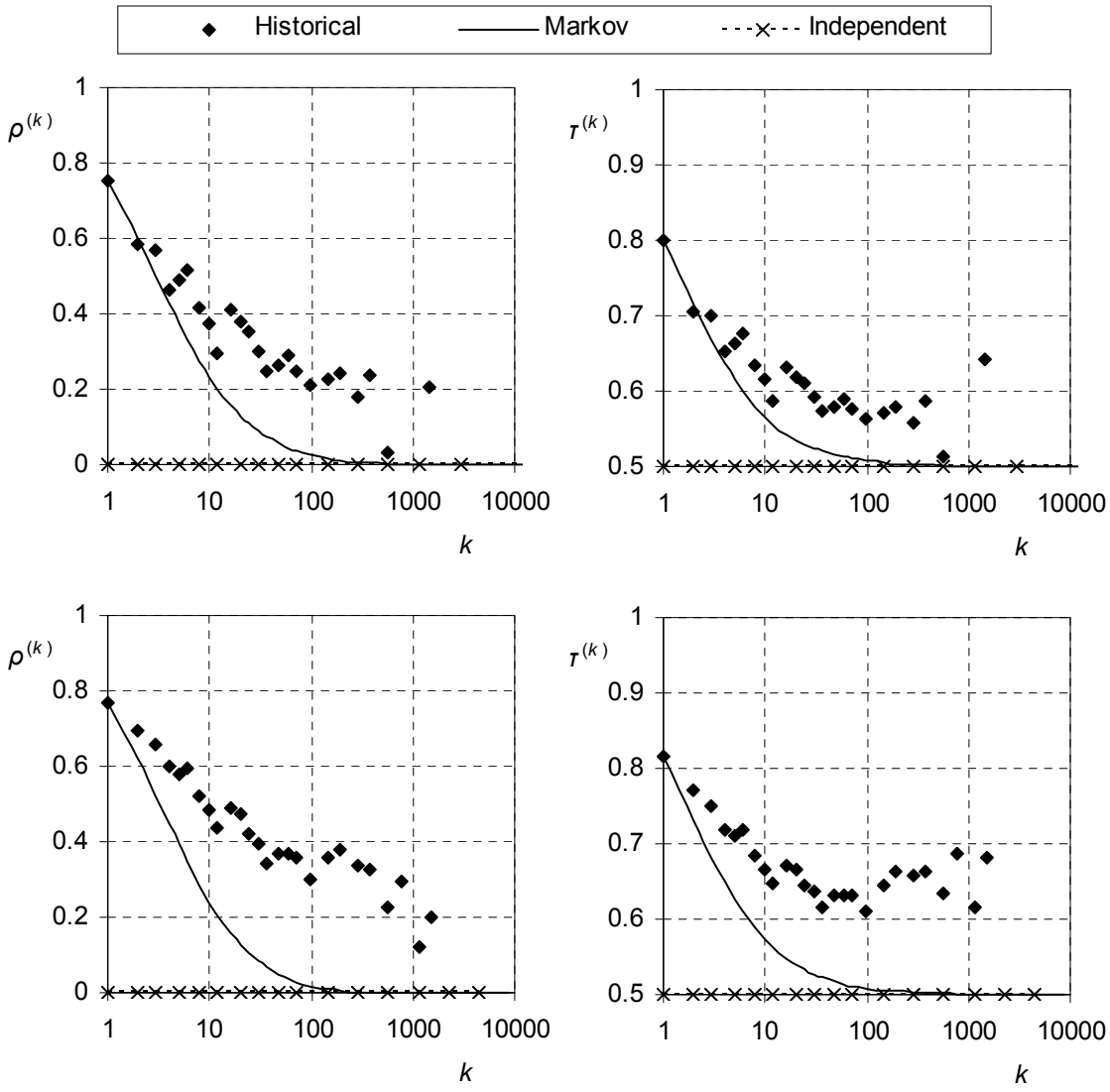

Figure 2 Lag one autocorrelation $\rho^{(k)}$ and dependence indicator $\tau^{(k)}$ vs. scale k, as estimated from the Athens rainfall data set and predicted by the models of Markov dependence and independence, (up) for the dry season and (down) for the entire year.

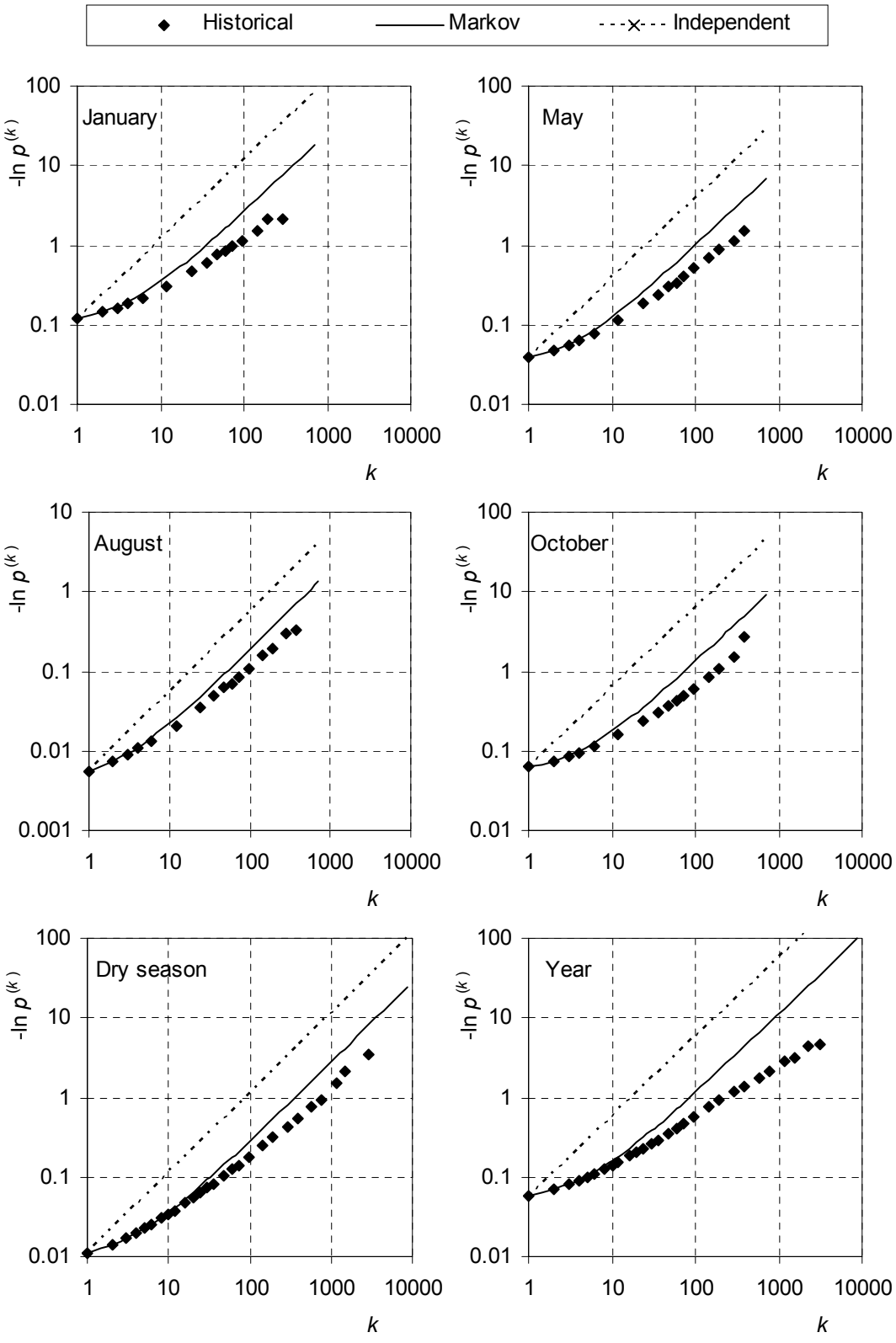

Figure 3 Log-probability dry ($-1\mathrm{n}\, p^{(k)}$) vs. scale k, as estimated from the Athens rainfall data set and predicted by the models of Markov dependence and independence, for each of the four months indicated, for the dry period and for the total year.

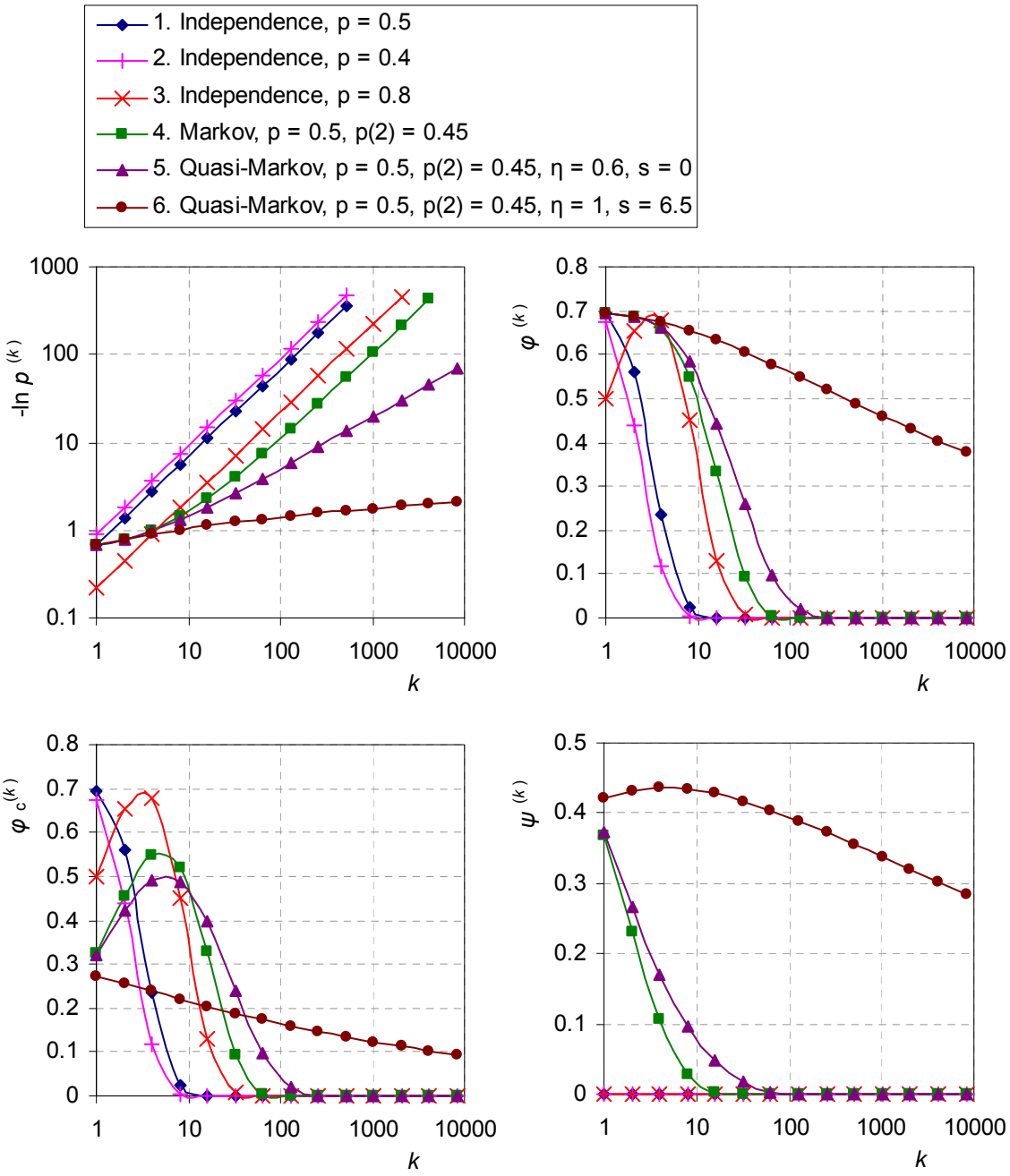

Figure 4 Log-probability dry ($-\ln p^{(k)}$), unconditional entropy ($\varphi^{(k)}$), conditional entropy ($\varphi_c^{(k)}$) and information gain ($\psi^{(k)}$) versus scale k for the six indicated hypothetical cases.

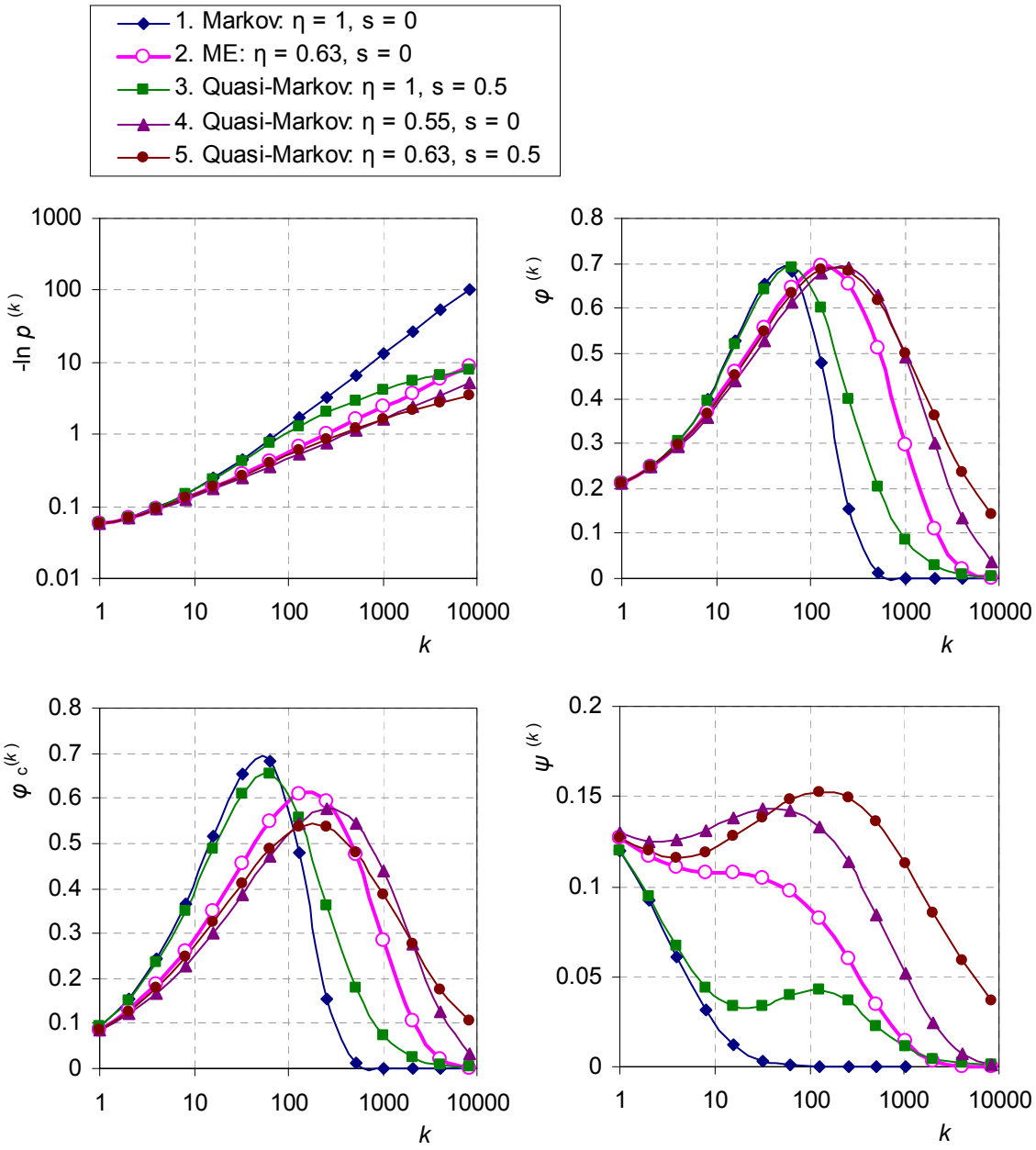

Figure 5 Log-probability dry ($-1\mathrm{n}\, p^{(k)}$), unconditional entropy ($\varphi^{(k)}$), conditional entropy ($\varphi_{\mathrm{c}}^{(k)}$) and information gain ($\psi^{(k)}$) versus scale k for the five indicated cases, all corresponding to p = 0.945 and $p^{(2)}$ = 0.933, which are the empirical probabilities dry of the Athens data set for the entire year.

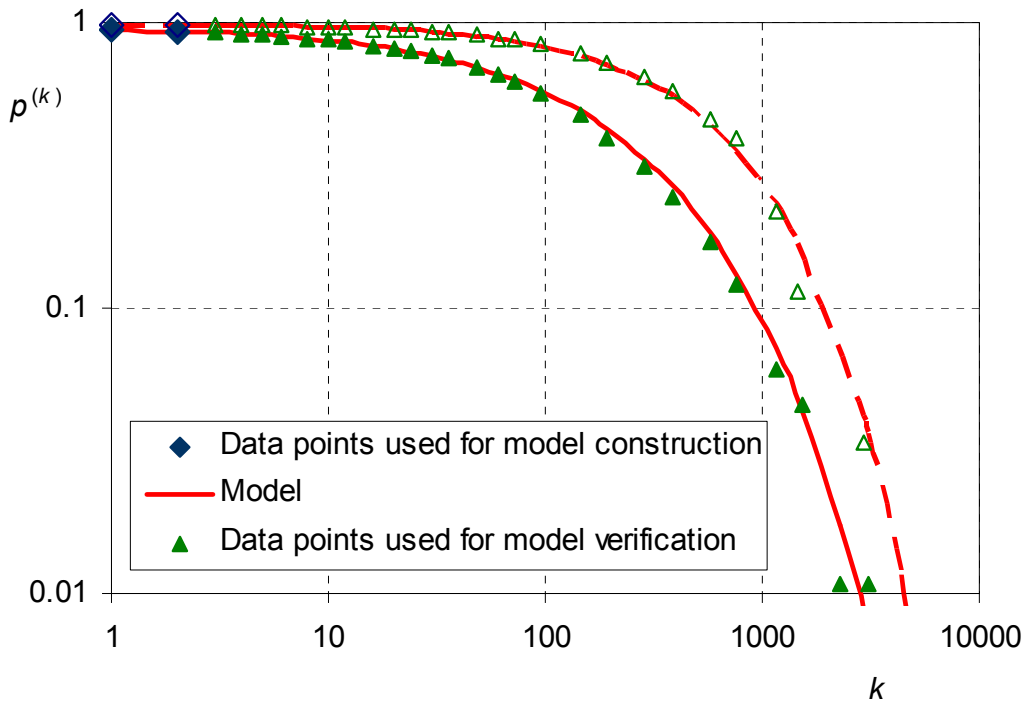

Figure 6 Probability dry $p^{(k)}$ vs. scale k, as estimated from the Athens rainfall data set and predicted by the model of maximum entropy for the entire year (full triangles and full line) and the dry season (empty triangles and dashed line).

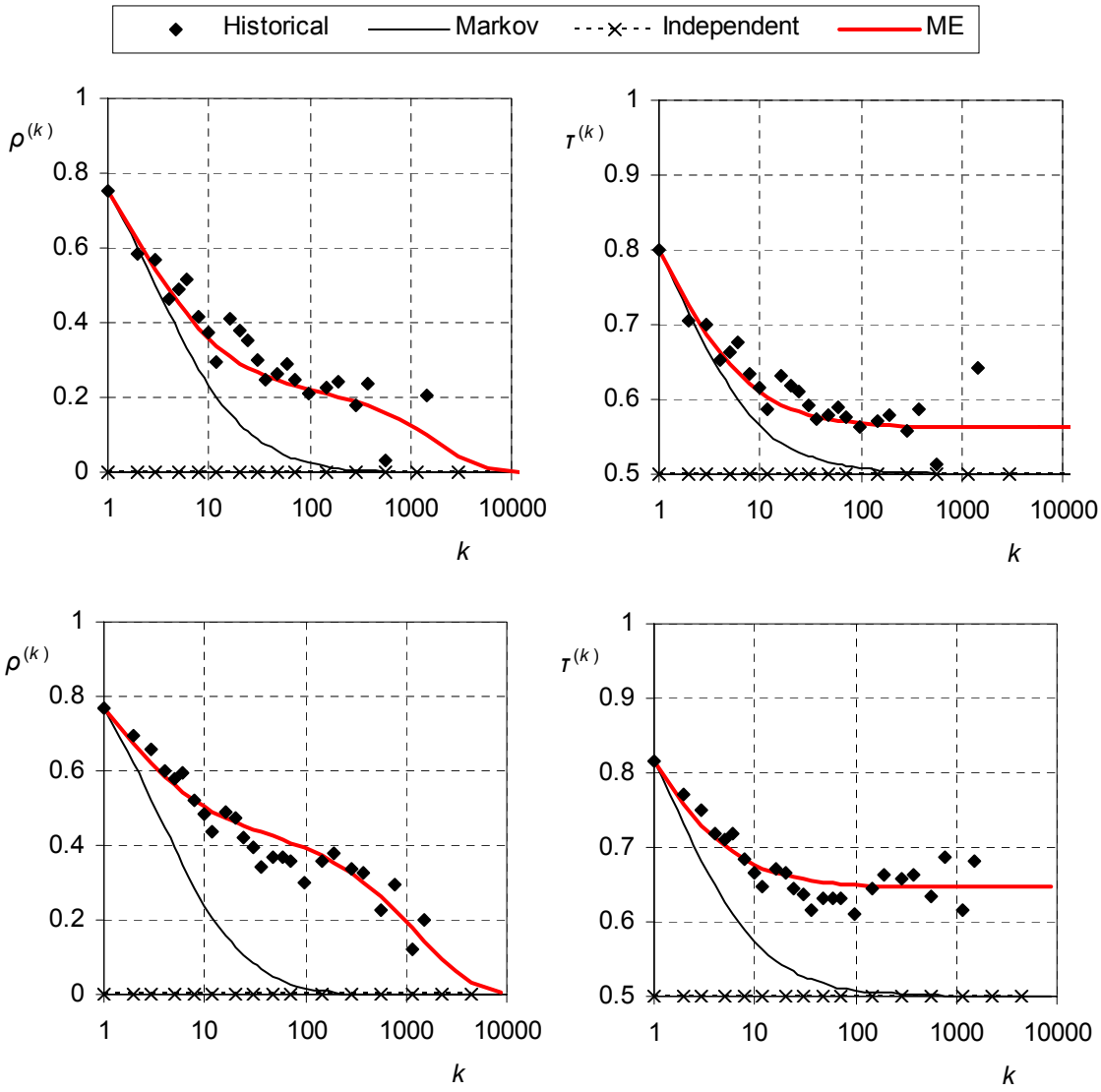

Figure 7 Lag one autocorrelation $\rho^{(k)}$ and dependence indicator $\tau^{(k)}$ vs. scale k, as estimated from the Athens rainfall data set and predicted by the models of maximum entropy, Markov dependence, and independence, (up) for the dry season and (down) for the entire year.

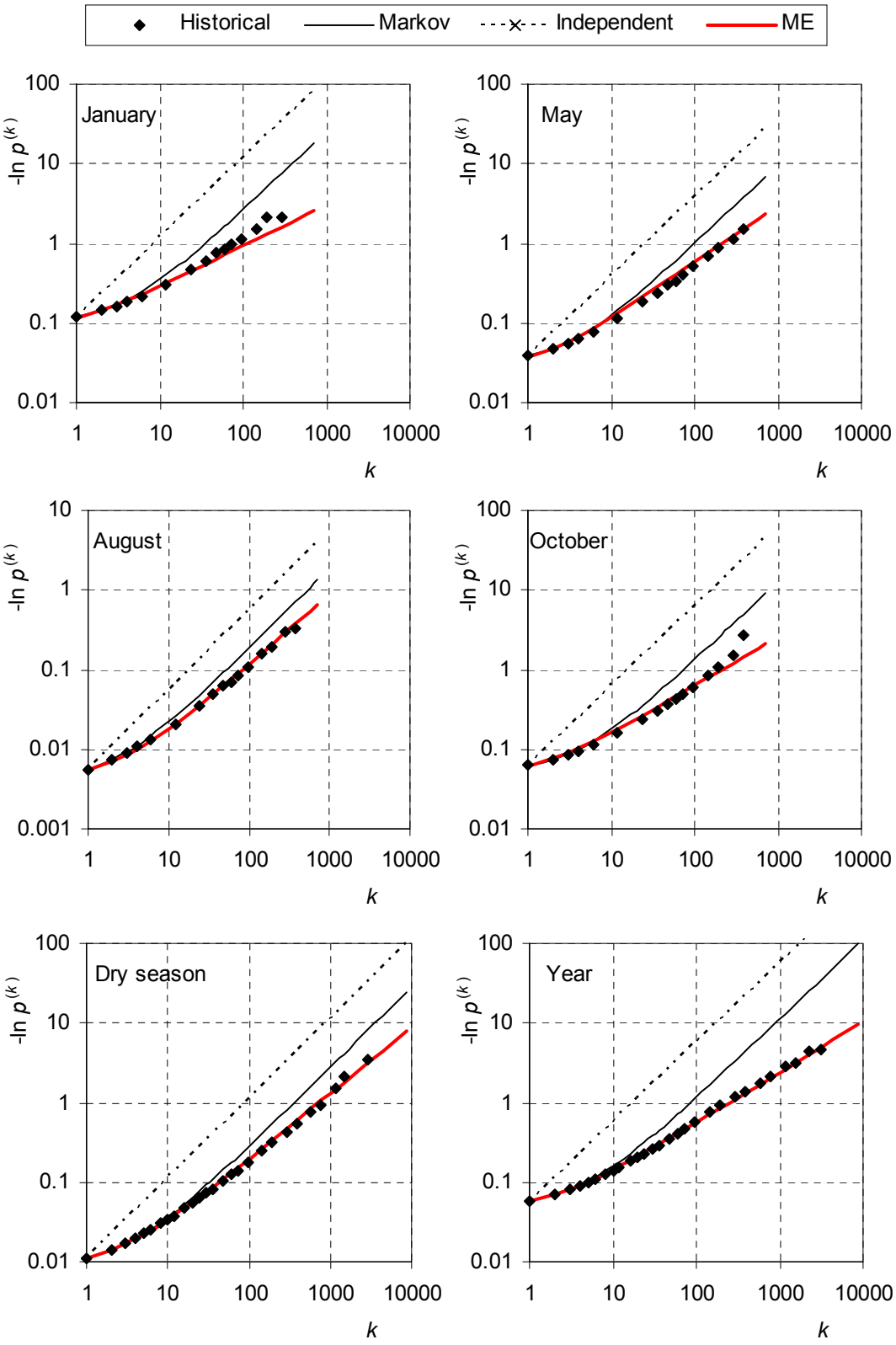

Figure 8 Log-probability dry ($-1n\, p^{(k)}$) vs. scale k, as estimated from the Athens rainfall data set and predicted by the models of maximum entropy, Markov dependence, and independence, for each of the four months indicated, for the dry season and for the total year.

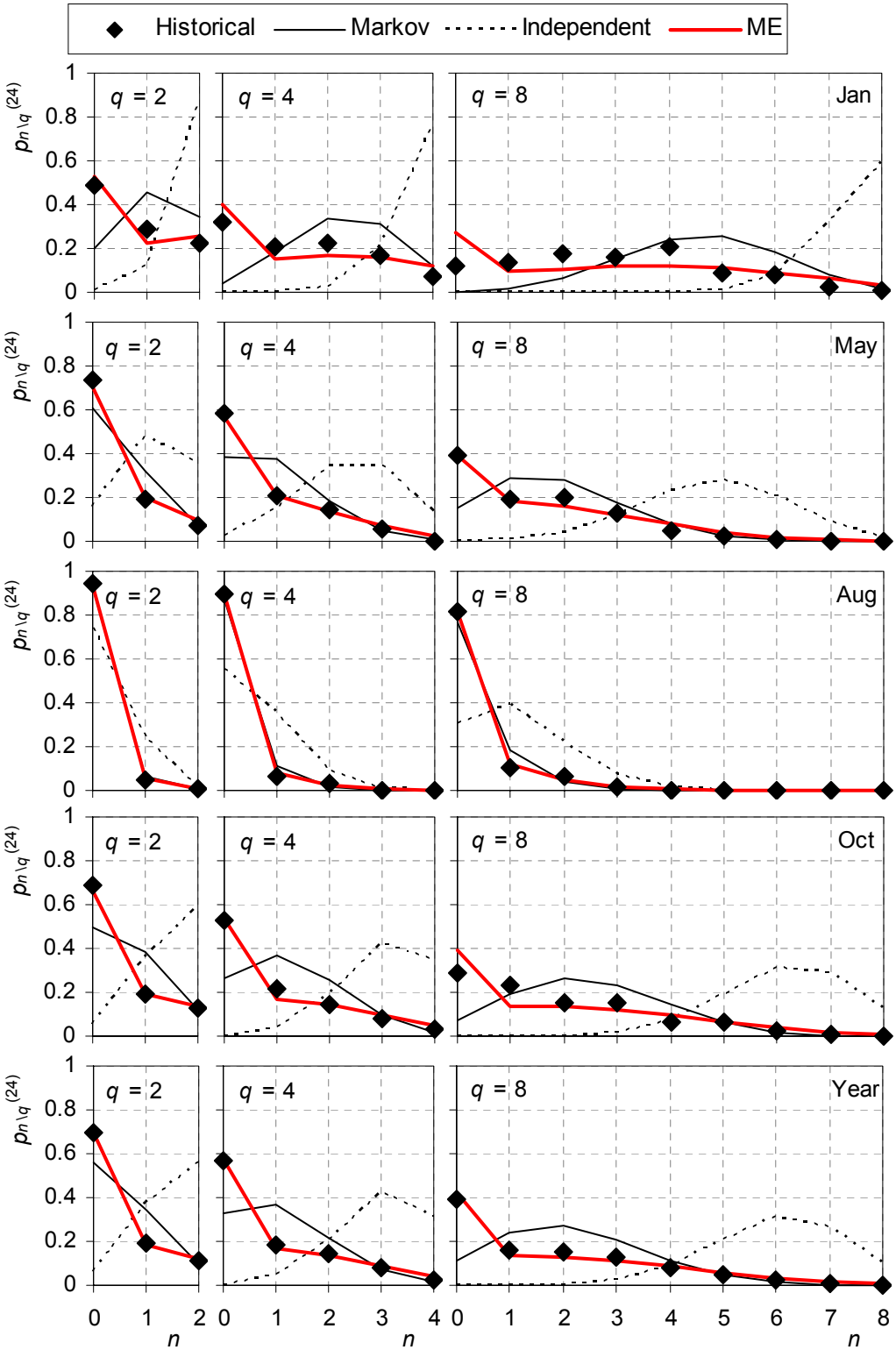

Figure 9 Probability $\boldsymbol{p}_{n\backslash q}^{(24)}$ of having n out of q wet days (scale $k = 24$) vs. n for $q = 2$, 4 and 8, as estimated from the Athens rainfall data set (for the indicated months and the entire year) and predicted by the models of maximum entropy (ME), Markov dependence and independence.