


# Temporal disaggregation of intermittent rainfall

| | |
|---|---|
| Journal: | *Hydrological Sciences Journal* |
| Manuscript ID | Draft |
| Manuscript Type: | Original Article |
| Date Submitted by the Author: | n/a |
| Complete List of Authors: | Lombardo, Federico; Università degli Studi Roma Tre, Dipartimento di Ingegneria<br>Volpi, Elena; Università degli Studi Roma Tre, Dipartimento di Ingegneria<br>Koutsoyiannis, Demetris; National Technical University of Athens,<br>Serinaldi, Francesco; Newcastle University School of Civil Engineering and Geosciences; Willis Research Network |
| Keywords: | Rainfall disaggregation, Rainfall intermittency, Hurst-Kolmogorov process |
| | |

# Temporal disaggregation of intermittent rainfall

F. Lombardo*[1], E. Volpi[1], D. Koutsoyiannis[2] & F. Serinaldi[3,4]

[1] *Dipartimento di Ingegneria, Università degli Studi Roma Tre, Via Vito Volterra, 62 – 00146 Rome, Italy*
[2] *Department of Water Resources and Environmental Engineering, National Technical University of Athens, Heroon Polytechneiou 5, GR-157 80 Zographou, Greece*
[3] *School of Civil Engineering and Geosciences, Newcastle University - Newcastle Upon Tyne, NE1 7RU, UK*
[4] *Willis Research Network - 51 Lime St., London, EC3M 7DQ, UK*

* Corresponding author: e-mail: federico.lombardo@uniroma3.it

**Abstract** Single-site disaggregation models of rainfall aim at generating finer scale time series of rainfall that are fully consistent with any given coarse-scale totals. In this work, we present a disaggregation method that initially retains the formalism, the parameter set, and the generation routine of the downscaling model described by Lombardo *et al*. (2012), which generates time series with Hurst-Kolmogorov (HK) dependence structure. Then it uses an adjusting procedure to achieve the full consistency between lower- and higher-level variables without affecting the stochastic structure implied by the original downscaling model. Furthermore, we provide a general methodology to account for rainfall intermittency, which is a fundamental issue in simulation. Intermittency is quantified by the probability that a time interval is dry. Here we focus on a modelling approach of a mixed type, with a discrete description of intermittency and a continuous description of rainfall. We model the intermittent rainfall process as the product of two stochastic processes: (i) The rainfall occurrence process, which is described by a binary valued stochastic process, with the values 0 and 1 representing dry and wet conditions, respectively; (ii) The non-zero rainfall process, which is given by our disaggregation model. We study the rainfall process as intermittent with both independent and dependent occurrences, where dependence is quantified by the probability that two consecutive time intervals are dry. In either case, we provide the analytical formulations of the main statistics of our mixed-type disaggregation model and show their clear accordance with Monte Carlo simulations. An application to rainfall time series from the real world is also shown.

**Key words** Rainfall disaggregation; rainfall intermittency; Hurst-Kolmogorov process

## 1 INTRODUCTION

Rainfall is the main input to most hydrological systems. A wide range of studies concerning floods, water resources and water quality require characterization of rainfall inputs at fine time scales (Blöschl and Sivapalan 1995). This may be possible using empirical observations, but there is often a need to extend available data in terms of temporal resolution satisfying some additive property (i.e. that the sum of the values of consecutive variables within a period be equal to the corresponding coarse-scale amount) (Berne *et al*. 2004). Hence, rainfall disaggregation models are required.

Although there is substantial experience in stochastic disaggregation of rainfall to fine time scales, most modelling schemes existing in the literature are ad hoc techniques rather than consistent generalised methods (see review by Koutsoyiannis 2003a). This is mainly due to the skewed distributions and the intermittent nature of the rainfall process at fine time scales, which are severe obstacles for the application of a theoretically consistent scheme to rainfall disaggregation (Koutsoyiannis and Langousis 2011). This paper reports some progress in this respect. We propose herein a follow-up to the downscaling model by Lombardo *et al*. (2012), which is revised to include both a stochastic model accounting for intermittency and an appropriate strategy to preserve the additive property. It is reminded that the preservation of the additive property distinguishes disaggregation from downscaling, whereas both are supposed to reproduce the important statistical properties of a process of interest. This modification required to set up a disaggregation model produces a more realistic rainfall model that retains its primitive simplicity in association with a parsimonious framework for simulation. In brief, the advancements reported under the following sections include:

- Background information. A basic review with discussion about some improvements on the model structure is presented in the next section.
- Additivity constraint. Lombardo *et al*. (2012) utilize auxiliary Gaussian variables to disaggregate a given rainfall amount to a certain scale of interest by a linear generation scheme. Nevertheless, rainfall is effectively modelled by positively skewed distributions, i.e. non-Gaussian. Hence, a scale-dependent exponential transformation of the variables is used in a way that the transformed variables have lognormal distribution with some important properties (see Appendix A). However, this means that the additive property, which is one of the main attributes of the original disaggregation scheme, is lost (Todini 1980). To overcome the problem we apply an empirical correction procedure, known as “power adjusting procedure” (Section 3), to restore the full consistency of lower-level and higher-level variables. This procedure is accurate in the sense that it does not alter the original dependence structure of the synthetic time series (Koutsoyiannis and Manetas 1996).
- Intermittency. The main novelty of this paper is the introduction of intermittency in the modelling framework, which is fully general and it can be

used also when simulating mixed-type processes other than rainfall from the real world. The model by Lombardo *et al*. (2012) simulates rainfall time series without intermittency. However, the rainfall process features an intermittent character at fine (sub-monthly) time scales, and thus the probability that a time interval is dry is generally greater than zero. Generally, the analysis and modelling of rainfall intermittency relate to the study of the rainfall occurrence process. Then, we need to introduce the latter in our modelling framework. In order to achieve such an objective, we describe (Section 4) the entire rainfall process using a two-state stochastic process with a discrete (rainfall occurrences) as well as a continuous component (non-zero rainfall). Our modelling framework enables the analytical formulation of the main statistics of the intermittent rainfall process.

– Comparison to observed data. In Section 5, we show a case study in order to test the capability of our model to reproduce the statistical behaviour of real rainfall time series.

## 2 BASIC CONCEPTS AND BACKGROUND

In rainfall modelling literature, the currently dominant approach to temporal disaggregation is based on discrete multiplicative random cascades (MRCs), which were first introduced in turbulence by Mandelbrot (1974). Despite the fact that more complex scale-continuous cascade models have been introduced (see e.g. Schmitt and Marsan 2001, Schmitt 2003, Lovejoy and Schertzer 2010a, 2010b), discrete MRCs are still the most widely used approach as they are very simple to understand and apply (Paschalis *et al*. 2012). A MRC is a discrete model in scale, meaning that the scale ratio from parent to child structures is an integer number strictly larger than one. This model is multiplicative, and embedded in a recursive manner. Each step is usually associated to a scale ratio of $b = 2$ (i.e. branching number); after $k$ steps, the total scale ratio is $2^k$, and we have:

$$R_{j,k} = R_{1,0} \prod_{i=0}^{k} W_{g(i,j),i} \tag{1}$$

where $j = 1, \ldots, 2^k$ is the time step; $R_{1,0}$ is the initial rainfall intensity to be distributed over the (subscale) cells $R_{j,k}$ of the cascade, each cell being associated to a random variable $W_{g(i,j),i}$ (i.e. cascade generator, called "weight") where $g(i,j) = \left\lceil \frac{j}{2^{k-i}} \right\rceil$ denotes a function which defines the position in time at the cascade step $i$ = 0, …, $k$. All these random variables are assumed non-negative, independent and identically distributed, and satisfy the condition $\langle W \rangle = 1$ where $\langle \cdot \rangle$ denotes expectation.

Unfortunately, as detailed in Lombardo *et al.* (2012) the application of MRC models is questionable in the context of rainfall simulation. The random process underlying these models is not stationary, because its autocovariance is not a function of lag only, as it would be in stationary processes. This is simply inherent to the model structure. For example, it can be shown that for canonical MRCs we may write lagged second moments after $k$ cascade steps as:

$$\langle R_{j,k} R_{j+t,k} \rangle = \langle R_{1,0}^2 \rangle \langle W^2 \rangle^{h_{j,k}(t)} \tag{2}$$

where $t$ is the discrete-time lag; since we have $h_{j,k}(t=0) = k$ for any $j$ and $k$, then the exponent $h_{j,k}(t)$ can be calculated recursively by:

$$h_{j,k}(t) = \begin{cases} \left(h_{j,k-1}(t) + 1\right)\Theta[2^{k-1} - j - t] & j \leq 2^{k-1}, t > 0 \\ h_{2^k - j - t + 1,k}(t) & j > 2^{k-1}, t > 0 \\ h_{2^k - j + 1,k}(|t|) & t < 0 \end{cases} \tag{3}$$

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

$$\Theta[m] = \begin{cases} 0, & m < 0 \\ 1, & m \geq 0 \end{cases} \tag{4}$$

Analogous considerations apply to microcanonical and bounded MRC models. Then, from eqs. (2) and (3), it is evident that the autocovariance for a MRC model depends upon position in time $j$ and cascade step $k$. We emphasise that the nonstationarity, which is thus taken into account, is often neglected by several researchers and practitioners. The problem of nonstationarity in processes generated by discrete MRCs is indeed not new in the literature (see e.g. Mandelbrot, 1974; Over, 1995; Veneziano and Langousis, 2010). From a scientific point of view, it is not always satisfactory to model an observed phenomenon by a stationary process. Nonetheless, it is important to stress here that stationarity is also related to ergodicity, which in turn is a prerequisite to make statistical inference from data. From a practical point of view, if there is nonstationarity then ergodicity cannot hold, which forbids inference from data that represent the most reliable information in building hydrological models and making predictions (Koutsoyiannis and Montanari 2014). Even though the two concepts of ergodicity and stationarity do not coincide in general, it is usually convenient to devise a model that is ergodic provided that we have excluded nonstationarity (Montanari and Koutsoyiannis 2014, Serinaldi and Kilsby 2015).

Most of the problems of MRC models reported above might be overcome by other disaggregation methods in the literature (see e.g. Marani and Zanetti 2007, Gyasi-Agyei 2011, 2012, Pui *et al.* 2012). However, MRC models gain their popularity due to their ease of use and understanding.

We propose a model characterized by a structure equally simple as that of MRC models, but it is based on a different approach (Hurst-Kolmogorov) and it proves to be stationary. Indeed, we emphasize that this model is not a multiplicative random cascade (MRC); it exploits knowledge from an auxiliary Gaussian domain where fractional Gaussian noise (i.e. Hurst-Kolmogorov —HK— process) is generated by means of a stepwise disaggregation technique based on a random cascade structure. For a detailed theoretical and numerical comparison of this model with discrete MRCs, the reader is referred to Lombardo *et al.* (2012). In the following, we briefly outline the model in question (see also Appendix B for a step-by-step implementation procedure).

Note that most of the derivations, notation, etc., used herein follow closely the paper by Lombardo *et al*. (2012), to which the reader is referred for further details. Nevertheless, here we use a different normalizing transformation of the given rainfall amount $Z_{1,0}$ at the initial largest scale ($i = 0$). In particular, we assume $Z_{1,0}$ lognormally distributed with a given mean $\mu_0$ and variance $\sigma_0^2$, and we log-transform it into an auxiliary Gaussian variable $\tilde{Z}_{1,0}$ as follows:

$$\tilde{Z}_{1,0} = \frac{1}{\alpha(k)}\left(\log Z_{1,0} - \beta(k)\right) \quad (5)$$

where $\alpha(k)$ and $\beta(k)$ are two functions given in Appendix A, that depend on the disaggregation level $k$ of interest. Note that the transformation cannot be invariant with respect to the time scale. The functions $\alpha(k)$ and $\beta(k)$ differ from those used by Lombardo *et al*. (2012), as will be discussed later (see also Appendix A).

The auxiliary variable $\tilde{Z}_{1,0}$ obtained by eq. (5) is then disaggregated into two variables on subintervals of equal size. This procedure is applied progressively until we generate the series at the time scale of interest.

Since this is an induction technique, it suffices to describe one step. Consider the generation step in which the higher-level amount $\tilde{Z}_{j,k-1}$ is disaggregated into two lower-level amounts $\tilde{Z}_{2j-1,k}$ and $\tilde{Z}_{2j,k}$ such that:

$$\tilde{Z}_{2j-1,k} + \tilde{Z}_{2j,k} = \tilde{Z}_{j,k-1} \quad (6)$$

Thus, we generate the variable of the first subinterval $\tilde{Z}_{2j-1,k}$ only, and that of the second is then the remainder that satisfies eq. (6). At this step, we have already generated the values of previous lower-level time steps, i.e. $\tilde{Z}_{1,k}, \dots, \tilde{Z}_{2j-2,k}$, and of the next higher-level time steps, i.e. $\tilde{Z}_{j,k-1}, \dots, \tilde{Z}_{n,k-1}$ where $n = 2^{k-1}$. Theoretically, it is necessary to preserve the correlations of $\tilde{Z}_{2j-1,k}$ with all previous lower-level variables and all next higher-level variables. However, we can obtain a very good approximation if we consider correlations with only one higher-level time step behind (i.e. two lower-level time steps behind) and one ahead (Koutsoyiannis, 2002). This is particularly the case with moderate values of the Hurst coefficient, while for high values (e.g. $H \geq 0.9$) an extensive numerical investigation (not reported here) showed that we could obtain the best trade-off between model accuracy and computational burden by expanding of one higher-level time step behind and ahead the number of variables that are considered in the generation procedure. In either case, we use the following linear generation scheme:

$$\tilde{Z}_{2j-1,k} = \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{Y} + V \quad (7)$$

where $\boldsymbol{Y}$ is a vector of previously generated variables, $\boldsymbol{\theta}$ is a vector of parameters, and $V$ is a Gaussian white noise that represents an innovation term. All unknown parameters and the variance of the innovation term can be estimated applying the methodology proposed by Koutsoyiannis (2001), which yields:

$$\boldsymbol{\theta} = \{\mathrm{cov}[\boldsymbol{Y}, \boldsymbol{Y}]\}^{-1} \mathrm{cov}\left[\boldsymbol{Y}, \tilde{Z}_{2j-1,k}\right] \quad (8)$$

$$\mathrm{var}[V] = \mathrm{var}\left[\tilde{Z}_{2j-1,k}\right] - \mathrm{cov}\left[\tilde{Z}_{2j-1,k}, \boldsymbol{Y}\right] \boldsymbol{\theta} \quad (9)$$

In short, the generation step is based on eq. (7) that can account for correlations with other variables, which are the components of the vector $\boldsymbol{Y}$ above. For example, if one considers correlations with only one higher-level time step behind and one ahead, then $\boldsymbol{Y} = \left[\tilde{Z}_{2j-3,k}, \tilde{Z}_{2j-2,k}, \tilde{Z}_{j,k-1}, \tilde{Z}_{j+1,k-1}\right]^{\mathrm{T}}$ where superscript T denotes the transpose of a vector. Then, eq. (7) simplifies as follows:

$$\tilde{Z}_{2j-1,k} = a_2\tilde{Z}_{2j-3,k} + a_1\tilde{Z}_{2j-2,k} + b_0\tilde{Z}_{j,k-1} + b_1\tilde{Z}_{j+1,k-1} + V \quad (10)$$

where $a_2$, $a_1$, $b_0$ and $b_1$ are parameters to be estimated and $V$ is innovation whose variance has to be estimated as well. It can be shown that, for the HK process, the two equations above depend only on the Hurst coefficient $H$ and the variance $\sigma_0^2$ (Koutsoyiannis 2002). From eqs. (8) and (9), all unknown parameters can be estimated in terms of correlations of the form:

$$\tilde{\rho}(t) = \mathrm{corr}\left[\tilde{Z}_{2j-1,k}, \tilde{Z}_{2j-1+t,k}\right] = |t+1|^{2H}/2 + |t-1|^{2H}/2 - |t|^{2H} \quad (11)$$

The reader is pointed to eqs. (33) and (34) by Lombardo *et al*. (2012) to find the complete analytical formulation.

In the implementation of such an approach, it can be noticed that the generation procedure is affected by changes in eq. (10) that occur at the boundary of the cascade (i.e. edge effects). In practice for each cascade step, when we generate $\tilde{Z}_{2j-1,k}$ near the start or end of the cascade sequence, some elements of the vector $\boldsymbol{Y}$ may be missing. In other words, some terms of eq. (10) are eliminated when $j = 1$ or $j = 2^{k-i}$ (i.e. at the start or end of the cascade sequence, respectively), for each cascade step $i$, where $i = 0, \ldots, k$. This limits the capability of the model to reproduce the theoretical properties of the HK process. To overcome this problem, again we found a good compromise by numerical investigation. The purpose is to reduce "leakage" aberrations in the model output that are introduced by a sharp truncation of the sequence at the outer edges of the cascade. We found a good solution by simultaneously disaggregating three independent and identically distributed Gaussian variables (where $\tilde{Z}_{1,0}$ is the one in the middle), as shown in Fig. 1. We use only the synthetic series pertaining to $\tilde{Z}_{1,0}$ and discard the remainder. Then, the effects of the peripheral leakage on the main statistics are practically negligible.

Finally, the lower-level variables generated in the auxiliary (Gaussian) domain must then be transformed back to the target (lognormal) domain, so as to make them a more realistic representation of the actual rainfall process and more comparable to common MRC models. Nevertheless, it can be argued that the structure of our model is substantially different from that of MRCs.

We use the following simple exponential transformation:

$$Z_{j,k} = \exp\left(\tilde{Z}_{j,k}\right) \tag{12}$$

This transformation is simpler than that used by Lombardo *et al*. (2012) in eq. (37). In fact, we normalize the given coarse-scale total $Z_{1,0}$ by eq. (5) in order to use a simpler inverse transformation, eq. (12), at the scale $k$ of interest. This is more appropriate for a disaggregation approach resembling a top-down strategy. Note that the functions $\alpha(k)$ and $\beta(k)$ are derived to preserve some scaling properties of the auxiliary process (as shown in the Appendix A). Specifically, the mean, variance and autocorrelation after $k$ cascade steps of the actual rainfall process are given respectively by:

$$\mu_k = \langle Z_{j,k} \rangle = \mu_0 / 2^k \tag{13}$$

$$\sigma_k^2 = \mathrm{var}\left[Z_{j,k}\right] = \sigma_0^2 / 2^{2Hk} \tag{14}$$

$$\rho_k(t) = \mathrm{corr}\left[Z_{j,k}, Z_{j+t,k}\right] = \frac{\exp\left(\tilde{\sigma}_k^2 \tilde{\rho}(t)\right) - 1}{\exp(\tilde{\sigma}_k^2) - 1} \tag{15}$$

where $\tilde{\rho}(t)$ and $\tilde{\sigma}_k^2 = \mathrm{var}\left[\tilde{Z}_{j,k}\right]$ respectively denote the autocorrelation function, eq. (11), and the variance of the auxiliary Gaussian process (i.e. Hurst-Kolmogorov process), $H$ is the Hurst coefficient, $t$ is the time lag, while $\mu_0$ and $\sigma_0^2$ are, respectively, the mean and variance of the given coarse-scale total $Z_{1,0}$. Note that the autocorrelation functions of the Hurst-Kolmogorov process, $\tilde{\rho}(t)$, and the target lognormal process, $\rho_k(t)$, generally differ. Nevertheless, for small values of $\tilde{\sigma}_k^2$, as encountered in disaggregation modelling of rainfall amounts, the experimental $\rho_k(t)$ closely resembles the ideal form of $\tilde{\rho}(t)$. Specifically, in the small-scale limit of $k \to \infty$ (i.e., very small $\tilde{\sigma}_k^2$), the autocorrelation function of the target process converges to that of the Hurst-Kolmogorov process, so that $\rho_k(t) \to \tilde{\rho}(t)$.

In summary, our model assumes lognormal rainfall, and then it is reasonable to use a (scale-dependent) logarithmic transformation of variables (eq. 5) and perform disaggregation of transformed variables in a Gaussian (auxiliary) domain, thus exploiting the desired properties of the normal distribution for disaggregation schemes (Koutsoyiannis, 2003a). Therefore, we do not disaggregate rainfall by a multiplicative random cascade (MRC). Rather, we assume a Hurst-Kolmogorov process in the auxiliary domain whose characteristics are changed (by eq. 5) based on the last disaggregation step of interest. The Hurst-Kolmogorov process is effectively generated using a stepwise disaggregation approach introduced by Koutsoyiannis

(2002), which is based on a random cascade structure. Finally, the generated lower-level variables are transformed back (eq. 12) to the original lognormal domain. Our specific transformation enables to preserve the scaling properties of the Hurst-Kolmogorov process also in the target (lognormal) domain; thus, it allows to reproduce the empirically observed characteristics of rainfall time series (e.g. at the daily scale).

## 3 ADJUSTING PROCEDURE

An important drawback of the above-summarized model is that generated, back-transformed rainfall amounts, $Z_{j,k}$, generally fail to sum to the specified coarse-scale total, $Z_{1,0}$, which is a major requirement of disaggregation methods. This is what normally happens when a model is specified in terms of the logarithms of the target variables, or some other normalizing transformation. In such cases, adjusting procedures are necessary to ensure additivity constraints (Stedinger and Vogel 1984, Grygier and Stedinger 1988, 1990, Lane and Frevert 1990, Koutsoyiannis and Manetas 1996), such as:

$$Z_{1,0} = \sum_{j=1}^{n=2^k} Z_{j,k} \tag{16}$$

A relevant question is how to adjust the generated rainfall amounts without unduly distorting their marginal distribution and dependence structure. Koutsoyiannis and Manetas (1996) showed that this is possible using appropriate adjusting procedures, which preserve certain statistics of lower-level variables. In particular, here we focus on the so-called "power adjusting procedure" that can preserve the first- and second-order statistics regardless of the type of the distribution function or the covariance structure of $Z_{j,k}$. This procedure allocates the error in the additive property among the lower-level variables. Thus, it modifies the generated variables $Z_{j,k}$ ($j$ = 1, …, $2^k$) to get the adjusted ones $Z'_{j,k}$ according to:

$$Z'_{j,k} = Z_{j,k} \left( \frac{Z_{1,0}}{\sum_{j=1}^{n} Z_{j,k}} \right)^{\lambda_{j,k}/\eta_{j,k}} \tag{17}$$

where

$$\lambda_{j,k} = \frac{\sum_{i=1}^{n} \mathrm{cov}\left[Z_{j,k}, Z_{i,k}\right]}{\sum_{j=1}^{n} \sum_{i=1}^{n} \mathrm{cov}\left[Z_{j,k}, Z_{i,k}\right]} \tag{18}$$

$$\eta_{j,k} = \frac{\langle Z_{j,k} \rangle}{\sum_{j=1}^{n} \langle Z_{j,k} \rangle} \tag{19}$$

The power adjusting procedure is more effective and suitable for our modelling framework than the classical linear and proportional adjusting procedures (see e.g. Grygier and Stedinger 1988, Lane and Frevert 1990). Indeed, a weakness of the former is that it may result in negative values of lower-level variables, but rainfall variables must be positive. Conversely, the proportional procedure always results in positive variables, but it is strictly exact only in some special cases that introduce severe limitations. The power adjusting procedure has no limitations and works in any case, but it does not preserve the additive property at once. Then, the application of eq. (17) must be iterative, until the calculated sum of the lower-level variables equals the given $Z_{1,0}$. Despite converging very rapidly, iterations reduce the model speed. However, the power adjusting procedure is a useful approximate generalization of proportional procedures, which in turn have severe limitations with lognormal variables (Koutsoyiannis and Manetas 1996).

Finally, we carry out some Monte Carlo experiments for further investigation. First we generate $M = 50000$ time series assuming the same parameters as in Lombardo *et al.* 2012: $k = 7$, $\mu_k = 1$, $\sigma_k^2 = 1.5$, and $H = 0.7$; then we apply the power adjusting procedure described by eq. (17) above. Fig. 2 shows that the adjusted variables fulfil the additive property, while Fig. 3 confirms that summary statistics of the generated variables are well preserved by the adjusting procedure. The latter displays, indeed, a very good agreement between the ensemble mean $\langle Z_{j,k} \rangle = \mu_k$, standard deviation $\sigma_k$, and autocorrelogram of the adjusted variables with generated ones; the latter termed as "empirical" in the figure legend.

## 4 INTERMITTENCY

The intermittent nature of rainfall process at fine time scales is a matter of common experience. In a statistical description, this is reflected by the fact that there exists a finite nonzero probability that the value of the process within a time interval is zero (often referred to as probability dry). Intermittency results in significant variability and high positive skewness, which are difficult to reproduce by most generators (Efstratiadis *et al.* 2014). Therefore, modelling rainfall intermittency is receiving renewed research interest (Koutsoyiannis 2006, Rigby and Porporato 2010, Kundu and Siddani 2011, Schleiss *et al*. 2011, Li *et al*. 2013, Mascaro *et al*. 2013).

In essence, for modelling rainfall intermittency two strategies are commonly used. The simplest approach is to model the intermittent rainfall process as a typical stochastic process whose smallest values are set to zero values according to a specific rounding off rule (see e.g. Koutsoyiannis *et al*. 2003). The second strategy considers in an explicit manner the two states of the rainfall process, i.e. the dry and the wet state. This is a modelling approach of a mixed type with a discrete description of intermittency and a continuous description of rainfall amounts (Srikanthan and McMahon 2001). The two-state approach is preferable for our modelling framework, because it facilitates the analytical formulation of the main statistics of the intermittent rainfall process.

The rainfall occurrence process (a binary-valued stochastic process) and the rainfall depth process (a continuous-type stochastic process) can be combined to give rise to a stochastic process of the mixed type. For simplicity, we assume that the discrete and continuous components are independent of one another; therefore, we are allowed to write the intermittent rainfall as the product of those two components.

In our modelling framework, we assume to model the intermittent rainfall $X_{j,k}$ on a single time scale setting at the cascade step $k$ and discrete time $j$ (= 1, …, $2^k$) as:

$$X_{j,k} = I_{j,k} \cdot Z'_{j,k} \tag{20}$$

where $Z'_{j,k}$ denotes the continuous-type random variable pertaining to our disaggregation model (given by eq. (17)), which represents the nonzero rainfall process. Whereas, the rainfall occurrence process is represented by $I_{j,k}$ that is a discrete-type random variable taking values 0 (dry condition) and 1 (wet condition), respectively with probability $p_{0,k}$ and $p_{1,k} = 1 - p_{0,k}$. The former denotes the probability that a certain time interval is dry after $k$ cascade steps, i.e. $p_{0,k} = \Pr\{X_{j,k} = 0\}$. This is the probability dry at the scale of interest, which is an additional model parameter. Clearly, this notation reflects a stationarity assumption of rainfall occurrences, because the probability dry $p_{0,k}$ does not depend on the time position $j$ but depends only on the timescale $k$.

The above considerations imply the following relationships for the mean and variance of the mixed-type rainfall process:

$$\langle X_{j,k} \rangle = (1 - p_{0,k})\mu_k \tag{21}$$

$$\mathrm{var}[X_{j,k}] = (1 - p_{0,k})(\sigma_k^2 + p_{0,k}\mu_k^2) \tag{22}$$

338 where $\mu_k$ and $\sigma_k^2$ denote the mean and the variance of the series generated by our
339 rainfall depth model, see eqs. (13) and (14) respectively.

340 Note that eq. (20) resembles the classical intermittent lognormal $\beta$-model
341 based on MRCs (Gupta and Waymire, 1993; Over and Gupta, 1994, 1996), but it is
342 strictly embedded into our Hurst-Kolmogorov modelling framework.

343 Since we aim at modelling a family of mixed-type random variables each
344 representing the rainfall state at time steps $j$ = 1, 2, …, we need to investigate the
345 dependence structure of this particular stochastic process. In other words, we analyse
346 the pairwise dependence of two randomly chosen variables $X_{j,k}$ and $X_{j+t,k}$ separated
347 by a time lag $t$. This is accomplished through deriving the formulation of the
348 autocovariance function for the intermittent rainfall process. Recall that:

$$\mathrm{cov}[X_{j,k}, X_{j+t,k}] = \langle X_{j,k} X_{j+t,k} \rangle - \langle X_{j,k} \rangle^2 \tag{23}$$

349 where the last term of the right-hand side can be easily calculated from eq. (21), while
350 the lagged second moment $\langle X_{j,k} X_{j+t,k} \rangle$ can be expressed through the following joint
351 probabilities:

$$\begin{aligned}
p_{00,k} &= \Pr\{X_{j,k} = 0, X_{j+t,k} = 0\} \\
p_{10,k} &= \Pr\{X_{j,k} > 0, X_{j+t,k} = 0\} \\
p_{01,k} &= \Pr\{X_{j,k} = 0, X_{j+t,k} > 0\} \\
p_{11,k} &= \Pr\{X_{j,k} > 0, X_{j+t,k} > 0\}
\end{aligned} \tag{24}$$

352 By total probability theorem and eq. (20) then, we have:

$$\langle X_{j,k} X_{j+t,k} \rangle = p_{11,k} \langle Z'_{j,k} Z'_{j+t,k} \rangle = p_{11,k} \langle Z_{j,k} Z_{j+t,k} \rangle \tag{25}$$

353 where the last equality is due to the properties of the power adjusting procedure (see
354 previous section).

355 For convenience, we express the joint probability $p_{11,k}$ in terms of the
356 probability dry $p_{0,k}$ and the autocovariance of rainfall occurrences $\mathrm{cov}[I_{j,k}, I_{j+t,k}]$.
357 The latter is given by (see also Koutsoyiannis 2006):

$$\mathrm{cov}[I_{j,k}, I_{j+t,k}] = \langle I_{j,k} I_{j+t,k} \rangle - \langle I_{j,k} \rangle^2 = p_{11,k} - (1 - p_{0,k})^2 \tag{26}$$

358 The derivation of this equation is based on the relationships $\langle I_{j,k} \rangle = \langle I_{j,k}^2 \rangle = 1 - p_{0,k}$,
359 and $\langle I_{j,k} I_{j+t,k} \rangle = p_{11,k}$. Thus, from eq. (26) we obtain:

$$p_{11,k} = \left(1 - p_{0,k}\right)^2 + \mathrm{cov}\left[I_{j,k}, I_{j+t,k}\right] \quad (27)$$

Substituting eqs. (21), (25) and (27) in eq. (23), we obtain:

$$\mathrm{cov}\left[X_{j,k}, X_{j+t,k}\right] = \left(\left(1 - p_{0,k}\right)^2 + \mathrm{cov}\left[I_{j,k}, I_{j+t,k}\right]\right) \langle Z_{j,k} Z_{j+t,k} \rangle - \left(1 - p_{0,k}\right)^2 \mu_k^2 \quad (28)$$

Adding and subtracting the term $\mathrm{cov}\left[I_{j,k}, I_{j+t,k}\right]\mu_k^2$ to the right-hand side of eq. (28), we obtain:

$$\begin{aligned} &\mathrm{cov}\left[X_{j,k}, X_{j+t,k}\right] = \\ &= \left(\left(1 - p_{0,k}\right)^2 + \mathrm{cov}\left[I_{j,k}, I_{j+t,k}\right]\right) \mathrm{cov}\left[Z_{j,k}, Z_{j+t,k}\right] + \mathrm{cov}\left[I_{j,k}, I_{j+t,k}\right]\mu_k^2 \end{aligned} \quad (29)$$

Hence, we have expressed the degree of dependence of the intermittent rainfall process in terms of the dependence structures of both the rainfall occurrence and depth processes.

A more common indicator of dependence of a stochastic process is the autocorrelation coefficient:

$$\rho_{X,k}(t) = \frac{\mathrm{cov}\left[X_{j,k}, X_{j+t,k}\right]}{\mathrm{var}\left[X_{j,k}\right]} \quad (30)$$

Recalling that $\mathrm{var}\left[I_{j,k}\right] = p_{0,k}\left(1 - p_{0,k}\right)$ and substituting eqs. (22) and (29) in eq. (30), after algebraic manipulations we obtain:

$$\rho_{X,k}(t) = \frac{\left(1 - p_{0,k} + \rho_{I,k}(t) p_{0,k}\right)\rho_k(t)\sigma_k^2 + \rho_{I,k}(t) p_{0,k}\mu_k^2}{\sigma_k^2 + p_{0,k}\mu_k^2} \quad (31)$$

where $\mu_k$, $\sigma_k^2$ and $\rho_k(t)$ are given by eqs. (13), (14) and (15) respectively. The only unknown in eq. (31) is the autocorrelation function $\rho_{I,k}(t)$ of the rainfall occurrence process at a single characteristic time scale (i.e., the final disaggregation step $k$). Therefore, to quantify the degree of dependence of the intermittent rainfall process we must assume a model for the dependence structure of rainfall occurrences.

Generally, we could classify such models into three types: (i) independence, which includes the Bernoulli case, characterized by one parameter only; (ii) simple dependence, which includes Markov chains characterized by two parameters; (iii) complex dependence, characterized by more than two parameters (Koutsoyiannis 2006). For the sake of numerical investigation, hereinafter we analyse the first two modelling categories of the occurrence processes:

1. Purely random model
2. Markov chain model

It was recognized in early stages of analysis and modelling attempts that the rainfall occurrences are not independent in time, and the Markov chain model was widely adopted for discrete time representations of this process (Gabriel and Neumann 1962, Haan *et al.* 1976, Chin 1977). It was later observed, however, that Markov chain models yield unsatisfactory results for rainfall occurrences, despite being much closer to reality than the independence model (De Bruin 1980, Katz and Parlange 1998). Moreover, there exist other types of models intended to simulate more complex dependence structures that are consistent with empirical data, such as positive autocorrelation both on small scales (short-term persistence) and on large scales (long-term persistence) (see e.g. Koutsoyiannis, 2006).

Our main purpose is to generate intermittent rainfall time series at a certain time scale, which are fully consistent with a given coarse-scale total. We focus on a modelling approach of a mixed type with a discrete description of intermittency and a continuous description of rainfall amounts. By eq. (20), we introduce the intermittent character in the (back-transformed) synthetic series at the "basic scale", which is represented by the last disaggregation step. In other words, we assume to model intermittency on a single time scale setting, and then we confine our interest only to the basic scale of disaggregated series.

In summary, we generate both independent and autocorrelated (binary) time series of rainfall occurrences at the basic scale, which are then multiplied by the continuous rainfall depth time series (generated by our disaggregation model) in order to obtain the final intermittent rainfall series. Note that our intermittency model is general and allows using any type of autocorrelation function, and we use the independent and Markovian cases as simple applications of our theoretical framework for Monte Carlo experiments.

### 4.1 Random occurrences

The simplest case is to assume that the rainfall process is intermittent with independent occurrences $I_{j,k}$, which can be modelled as a Bernoulli process in discrete time. This process is characterized by one parameter only, i.e. the probability dry $p_{0,k}$. Then, we can write that:

$$\rho_{I,k}(t) = \mathrm{cov}\left[I_{j,k}, I_{j+t,k}\right] = 0 \qquad (32)$$

Substituting eq. (32) in eqs. (29) and (31), we obtain respectively:

$$\text{cov}[X_{j,k}, X_{j+t,k}] = (1 - p_{0,k})^2 \text{cov}[Z_{j,k}, Z_{j+t,k}] \tag{33}$$

$$\rho_{X,k}(t) = (1 - p_{0,k})\rho_k(t) \frac{\sigma_k^2}{\sigma_k^2 + p_{0,k}\mu_k^2} \tag{34}$$

### 4.2 Markovian occurrences

As a second example, we assume a very simple occurrence process with some correlation. In this model, the dependence of the current variable $I_{j,k}$ on the previous variable $I_{j-1,k}$ suffices to express completely the dependence of the present on the past. In other words, we assume that the state (dry or wet) in a time interval depends solely on the state in the previous interval. This is a process with Markovian dependence, which is completely determined by lag-one autocorrelation coefficient $\rho_{I,k}(1) = \text{corr}[I_{j,k}, I_{j-1,k}]$. Therefore, the occurrence process is characterized by two parameters, i.e. $p_{0,k}$ and $\rho_{I,k}(1)$. The autocorrelation of $I_{j,k}$ is (see the proof in Appendix C):

$$\rho_{I,k}(t) = \text{corr}[I_{j,k}, I_{j+t,k}] = \rho_{I,k}^{|t|}(1) \tag{35}$$

Substituting in eq. (31), we derive the autocorrelation of the entire rainfall process as:

$$\rho_{X,k}(t) = \frac{\left(1 - p_{0,k} + \rho_{I,k}^{|t|}(1) p_{0,k}\right)\rho_k(t)\sigma_k^2 + \rho_{I,k}^{|t|}(1) p_{0,k}\mu_k^2}{\sigma_k^2 + p_{0,k}\mu_k^2} \tag{36}$$

### 4.3 Numerical simulations

For the sake of illustration, we generate 10,000 time series with sample size $n = 2^{10} = 1024$ ($k$ = 10 cascade steps), Hurst coefficient $H = 0.85$, and rainfall depths with unit mean and variance $\mu_k = \sigma_k^2 = 1$. In addition, we simulate both random and Markovian occurrences with probability dry $p_{0,k} = 0.2, 0.5, 0.8$ and lag-one autocorrelation coefficient (for Markovian case) $\rho_{I,k}(1) = 0.7$. The Markovian occurrences are generated implementing Boufounos (2007) algorithm.

Figs. 4 and 5 depict, respectively, the good agreement between the theoretical and empirical probability dry and autocorrelation functions for both types of occurrence models.

Figs. 6 and 7 show autocorrelograms of the mixed-type process for various values of the probability dry, i.e. $p_{0,k} = 0, 0.2, 0.8$. Note that the case with $p_{0,k} = 0$ corresponds to the rainfall depth process. The autocorrelogram of this process is used as a benchmark to compare the influence of each occurrence model on the dependence structure of the entire process. As expected, both of our occurrence models are generally cause for decorrelation of the intermittent process with respect to the process without intermittency. Clearly, this is particularly the case for random occurrences (see Fig. 6). For Markovian occurrences (see Fig. 7), the autocorrelation is higher for small time lags than that for random occurrences, while it tends to the random case asymptotically (compare Figs. 6 and 7 for $p_{0,k} = 0.2, 0.8$).

## 5 APPLICATION TO OBSERVATIONAL DATA

In this section, we compare our model against real rainfall time series in order to show the capability of the proposed methodology to reproduce the pattern of historical rainfall data on fine timescales. The dataset consists of 30-minute rainfall time series spanning from 1995 to 2005 from a raingauge in Viterbo, Italy. For further details on the observational data, the reader is referred to Serinaldi (2010).

As the rainfall process exhibits seasonality at sub-annual timescales, we focus on rainfall records from each month of the year separately, in order for the analyses to be consistent with the stationarity requirement of our model with an acceptable degree of approximation.

As highlighted in the previous section, the dependence structure of the rainfall occurrence process appears to be non-Markovian (Koutsoyiannis, 2006). To a first approximation, we make the simplifying assumption that the autocorrelation function $\rho_{I,k}(t)$ of the binary component (intermittency) of our model is given by eq. (11), where the only parameter $H$ equals the Hurst parameter of the continuous component (rainfall depth) of our model.

Concerning the estimation of model parameters from observational data, $H$ is estimated by the LSV (Least Squares based on Variance) method as described in Tyralis and Koutsoyiannis (2011), which is applied directly to each month of the data series. As the latter represent a realization of the intermittent rainfall process, $X_{j,k}$, with mean and variance given by eqs. (21) and (22), respectively, such statistical properties can be therefore estimated directly from data. Once we easily estimate the probability dry from data, we can solve eqs. (21) and (22) for the remaining two parameters to be estimated, i.e. the mean and variance of the rainfall depth process, $Z_{j,k}$. Hence, we have a very parsimonious disaggregation model with only four parameters.

We perform 10,000 Monte Carlo experiments to disaggregate monthly totals into sub-hourly time series of intermittent rainfall at the cascade level $k = 10$ (i.e., sample size $2^{10}$). Following the procedure described in sections 2 and 3 above, we first generate correlated series of rainfall amounts, $Z'_{j,k}$, with ACF in eq. (15). Second, we generate correlated binary series of rainfall occurrences, $I_{j,k}$, with ACF in eq. (11) (for a detailed description of the algorithm, refer to Serinaldi and Lombardo (2016)). By eq. (20), we combine the outcomes of the two generation steps above to obtain the synthetic intermittent series, $X_{j,k}$, with ACF in eq. (31).

By way of example, in Figs. 8 and 9 we respectively compare the observed autocorrelograms for January 1999 and April 2003 data series against the ACFs simulated by our model. In the left and right panels of each figure, we show respectively the ACF of the occurrence (binary) process $\rho_{I,k}(t)$ and that of the intermittent (mixed) process $\rho_{X,k}(t)$. In either case, it can be noticed that the model on average fits the observed behaviour satisfactorily. Other summary statistics such as the mean, variance and probability dry of the data series are preserved by hypothesis (not shown).

In Figs. 10 and 11, we compare the historical hyetographs for January 1999 and April 2003 to a typical synthetic hyetograph generated by our model. In both cases, we can see that our model produces realistic traces of the real world hyetograph. Other than similarities in the general shapes, we showed that our model provides simulations that preserve the statistical behaviour observed in real rainfall time series.

## 6 CONCLUSIONS

The discrete MRC is the dominant approach to rainfall disaggregation in hydrological modelling literature. However, MRC models have severe limitations due to their structure, which implies nonstationarity. As it is usually convenient to devise a model that is ergodic provided that we have excluded nonstationarity, Lombardo *et al*. (2012) proposed a simple and parsimonious downscaling model of rainfall in time based on the Hurst-Kolmogorov process. This model is here revisited in the light of bringing it more in line with the properties observed in real rainfall. To this aim, we upgrade our model to produce finer-scale intermittent time series that add up to any given coarse scale total.

Our main purpose is to provide theoretical insights into modelling rainfall disaggregation in time when accounting for rainfall intermittency. Then, we propose and theoretically analyze a model that is capable of describing some relevant statistics of the intermittent rainfall process in closed forms. We combine a continuous-type stochastic process (representing rainfall amounts) characterized by scaling properties with a binary-valued stochastic process (representing rainfall occurrences) that can be characterized by any dependence structure.

In particular, we first modify our model structure according to a top-down approach. Second, since our method utilizes nonlinear transformations of the variables in the generation procedure, we need to satisfy the additive property, which is the mass conservation between lower- and higher-level variables. To accomplish this purpose, we use an accurate adjusting procedure that preserves explicitly the first- and second-order statistics of the lower-level variables. Consequently, the original downscaling model by Lombardo *et al*. (2012) now becomes a disaggregation model.

Furthermore, we account for intermittency in our modelling framework by a modelling approach of a mixed type with a discrete (binary) description of intermittency and a continuous description of rainfall amounts. Our disaggregation model gives the latter, while the former should be specified by assuming a certain rainfall occurrence model. Nevertheless, we provide general theoretical formulations for summary statistics of the mixed-type process. For illustration purposes, we assume two different models of rainfall occurrences: (i) the Bernoulli model characterized by one parameter only, and (ii) the Markov chain model characterized by two parameters. We carry out Monte Carlo experiments to emphasize the good performance of our model.

Finally, comparisons between model simulations and intermittent rainfall time series from the real world show extremely encouraging results with a very parsimonious modelling framework (just four parameters).

## REFERENCES


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

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

Boufounos, P., 2007. Generating Binary Processes with all-Pole Spectra. *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, 3, III-981–III-984.

Chin, E. H., 1977. Modeling daily precipitation occurrence process with Markov chain. *Water Resources Research*, 13 (6), 949–956.

De Bruin, H. A. R., 1980. A stochastic description of wet and dry spells in terms of an effective number of days. *Journal of hydrology*, 45 (1), 91–99.

Efstratiadis, A., Dialynas, Y., Kozanis, S. and Koutsoyiannis, D., 2014. A multivariate stochastic model for the generation of synthetic time series at multiple time scales reproducing long-term persistence. *Environmental Modelling and Software*, 62, 139–152.

Gabriel, K. R., and Neumann, J., 1962. A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quarterly Journal of the Royal Meteorological Society*, 88 (375), 90–95.

Grygier, J. C. and Stedinger, J. R., 1988. Condensed disaggregation procedures and conservation corrections for stochastic hydrology. *Water Resources Research*, 24 (10), 1574–1584.

Grygier, J. C. and Stedinger, J. R., 1990. SPIGOT, A synthetic streamflow generation software package, Technical description. School of Civil and Environmental Engineering, Cornell University, Ithaca, NY., Version 2.5.

Gupta, V. K., and Waymire, E. C., 1993. A statistical analysis of mesoscale rainfall as a random cascade. *Journal of Applied Meteorology*, 32 (2), 251–267.

Gyasi-Agyei, Y., 2011. Copula-based daily rainfall disaggregation model. *Water Resources Research*, 47, W07535.

Gyasi-Agyei, Y., 2012. Use of observed scaled daily storm profiles in a copula based rainfall disaggregation model. *Advances in Water Resources*, 45, 26–36.

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

Katz, R. W., and Parlange, M. B., 1998. Overdispersion phenomenon in stochastic modeling of precipitation. *Journal of Climate*, 11(4), 591–601.

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

Koutsoyiannis, D., 2002. The Hurst phenomenon and fractional Gaussian noise made easy. *Hydrological Sciences Journal*, 47 (4), 573–595.

Koutsoyiannis, D., 2003a. Rainfall disaggregation methods: Theory and applications. *Proceedings, Workshop on Statistical and Mathematical Methods for Hydrological Analysis*, edited by D. Piccolo and L. Ubertini, Rome, 1–23, Università di Roma "La Sapienza".

Koutsoyiannis, D., 2003b. Climate change, the Hurst phenomenon, and hydrological statistics. Hydrological Sciences Journal, 48 (1), 3–24.

Koutsoyiannis, D., 2006. An entropic-stochastic representation of rainfall intermittency: The origin of clustering and persistence. *Water Resources Research*, 42 (1), W01401, doi:10.1029/2005WR004175.

Koutsoyiannis, D., Onof, C. and Wheater, H. S., 2003. Multivariate rainfall disaggregation at a fine timescale. *Water Resources Research*, 39 (7), 1173, doi:10.1029/2002WR001600.

Koutsoyiannis, D. and Langousis, A., 2011. Precipitation. *Treatise on Water Science*, edited by P. Wilderer and S. Uhlenbrook, 2, 27–78, Academic Press, Oxford.

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

Koutsoyiannis, D. and Montanari, A., 2014. Negligent killing of scientific concepts: the stationarity case. *Hydrological Sciences Journal*, doi:10.1080/02626667.2014.959959.

Kundu, P. K. and Siddani, R. K., 2011. Scale dependence of spatiotemporal intermittence of rain. *Water Resources Research*, 47, W08522, doi:10.1029/2010WR010070.

Lane, W. L., Frevert, D. K., 1990. Applied Stochastic Techniques, User's Manual. Bureau of Reclamation, Engineering and Research Center, Denver, Co., Personal Computer Version.

Li, C., Singh, V. P. and Mishra, A. K., 2013. A bivariate mixed distribution with a heavy-tailed component and its application to single-site daily rainfall simulation. *Water Resources Research*, 49, doi:10.1002/wrcr.20063.

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

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

Lovejoy, S. and Schertzer, D., 2010a. On the simulation of continuous in scale universal multifractals, part I: spatially continuous processes. *Computers and Geoscience*, 36, 1393–1403, doi:10.1016/j.cageo.2010.04.010.

Lovejoy, S. and Schertzer, D., 2010b. On the simulation of continuous in scale universal multifractals, part II: space-time processes and finite size corrections. *Computers and Geoscience*, 36, 1404–1413, doi:10.1016/j.cageo.2010.07.001.

Mandelbrot, B. B., 1974. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *Journal of Fluid Mechanics*, 62, 331–358.

Marani, M. and Zanetti, S., 2007. Downscaling rainfall temporal variability. *Water Resources Research*, 43, W09415, doi:10.1029/2006WR005505.

Mascaro, G., Deidda, R. and Hellies, M., 2013. On the nature of rainfall intermittency as revealed by different metrics and sampling approaches. *Hydrology and Earth System Sciences*, 17, 355–369.

Montanari, A., and Koutsoyiannis, D., 2014. Modeling and mitigating natural hazards: Stationarity is immortal!. *Water Resources Research*, 50, 9748–9756, doi:10.1002/2014WR016092.

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

Over, T. M., and Gupta, V. K., 1994. Statistical analysis of mesoscale rainfall: Dependence of a random cascade generator on large-scale forcing. *Journal of Applied Meteorology*, 33 (12), 1526–1542.

Over, T. M., & Gupta, V. K., 1996. A space-time theory of mesoscale rainfall using random cascades. *Journal of Geophysical Research*, 101 (D21), 26319–26331.

Paschalis, A., Molnar, P. and Burlando, P., 2012. Temporal dependence structure in weights in a multiplicative cascade model for precipitation. *Water Resources Research*, 48, W01501, doi:10.1029/2011WR010679.

Pui, A., Sharma, A., Mehrotra, R., Sivakumar, B., and Jeremiah, E., 2012. A comparison of alternatives for daily to sub-daily rainfall disaggregation. *Journal of Hydrology*, 470, 138–157.

Rigby, J. and Porporato, A., 2010. Precipitation, dynamical intermittency, and sporadic randomness. *Advances in Water Resources*, 33, 923–932.

Roldán, J. and Woolhiser, D. A., 1982. Stochastic daily precipitation models: 1. A comparison of occurrence processes. *Water Resources Research*, 18(5), 1451–1459, doi:10.1029/WR018i005p01451.

Schleiss, M., Jaffrain, J. and Berne, A., 2011. Statistical analysis of rainfall intermittency at small spatial and temporal scales, *Geophysical Research Letters*, 38, L18403, doi:10.1029/2011GL049000.

Schmitt, F., 2003. A causal multifractal stochastic equation and its statistical properties. *European Physical Journal B*, 34, 85–98.

Schmitt, F. and Marsan, D., 2001. Stochastic equations generating continuous multiplicative cascades. *European Physical Journal B*, 20, 3–6.

Serinaldi, F., 2010. Multifractality, imperfect scaling and hydrological properties of rainfall time series simulated by continuous universal multifractal and discrete random cascade models. *Nonlinear Processes in Geophysics*, 17, 697–714, doi:10.5194/npg-17-697-2010.

Serinaldi, F. and Kilsby, C.G., 2015. Stationarity is undead: uncertainty dominates the distribution of extremes. *Advances in Water Resources*, doi: http://dx.doi.org/10.1016/j.advwatres.2014.12.013

Serinaldi, F. and Lombardo, F., 2016. A general simulation algorithm for autocorrelated binary processes. *Physical Review E*, under review.

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

Stedinger, J. R., Vogel, R. M., 1984. Disaggregation procedures for generating serially correlated flow vectors. *Water Resources Research*, 20 (1), 47–56.

Todini, E., 1980. The preservation of skewness in linear disaggregation schemes. *Journal of Hydrology*, 47, 199-214.

Tyralis, H. and Koutsoyiannis, D., 2011. Simultaneous estimation of the parameters of the Hurst-Kolmogorov stochastic process. *Stochastic Environmental Research and Risk Assessment*, 25 (1), 21–33.

Veneziano, D., and Langousis, A., 2010. Scaling and fractals in hydrology. In: B. Sivakumar ed., *Advances in data-based approaches for hydrologic modeling and forecasting*. World Scientific, Chapter 4.

## APPENDIX A

We assume that the disaggregated rainfall process at scale $k$ is given by:

$$Z_{j,k} = \exp\left(\tilde{Z}_{j,k}\right) \tag{A1}$$

Consequently, its mean $\mu_k$ and variance $\sigma_k^2$ are functions of their auxiliary counterparts $\tilde{\mu}_k$ and $\tilde{\sigma}_k^2$ as follows (recall that $\tilde{\mu}_k = \tilde{\mu}_0/2^k$ and $\tilde{\sigma}_k^2 = \tilde{\sigma}_0^2/2^{2Hk}$):

$$\begin{cases} \mu_k = \exp\left(\dfrac{\tilde{\mu}_0}{2^k} + \dfrac{\tilde{\sigma}_0^2}{2^{2Hk+1}}\right) \\ \sigma_k^2 = \exp\left(\dfrac{\tilde{\mu}_0}{2^{k-1}} + \dfrac{\tilde{\sigma}_0^2}{2^{2Hk}}\right)\left(\exp\left(\dfrac{\tilde{\sigma}_0^2}{2^{2Hk}}\right) - 1\right) \end{cases} \tag{A2}$$

Then, our primary goal is to let the process $Z_{j,k}$ follow the same scaling laws of the relevant auxiliary process $\tilde{Z}_{j,k}$, such as:

$$\begin{cases} \mu_0 = 2^k \mu_k \\ \sigma_0^2 = 2^{2Hk} \sigma_k^2 \end{cases} \tag{A3}$$

where $\mu_0$ and $\sigma_0^2$ are respectively the mean and variance of the initial rainfall amount $Z_{1,0}$ at the largest scale.

To accomplish our goal, we may write $Z_{1,0}$ as:

$$Z_{1,0} = \exp\left(\alpha(k)\tilde{Z}_{1,0} + \beta(k)\right) \tag{A4}$$

where $\alpha(k)$ and $\beta(k)$ depend on the scale $k$ of interest, and they should be derived to preserve the scaling properties in eq. (A3).

We first recall that eq. (A4) implies:

$$\begin{cases} \mu_0 = \exp\left(\beta(k) + \alpha(k)\tilde{\mu}_0 + \alpha^2(k)\dfrac{\tilde{\sigma}_0^2}{2}\right) \\ \sigma_0^2 = \exp(2\beta(k) + 2\alpha(k)\tilde{\mu}_0 + \alpha^2(k)\tilde{\sigma}_0^2)(\exp(\alpha^2(k)\tilde{\sigma}_0^2) - 1) \end{cases} \tag{A5}$$

Substituting equation (A2) in (A3), equating the latter to eq. (A5) and then taking the natural logarithm of both sides, we obtain respectively:

$$k\log 2 + \frac{\tilde{\mu}_0}{2^k} + \frac{\tilde{\sigma}_0^2}{2^{2Hk+1}} = \beta(k) + \alpha(k)\tilde{\mu}_0 + \alpha^2(k)\frac{\tilde{\sigma}_0^2}{2} \tag{A6}$$

$$\begin{aligned} &2Hk\log 2 + \frac{\tilde{\mu}_0}{2^{k-1}} + \frac{\tilde{\sigma}_0^2}{2^{2Hk}} + \log\left(\exp\left(\frac{\tilde{\sigma}_0^2}{2^{2Hk}}\right) - 1\right) = \\ &= 2\beta(k) + 2\alpha(k)\tilde{\mu}_0 + \alpha^2(k)\tilde{\sigma}_0^2 + \log(\exp(\alpha^2(k)\tilde{\sigma}_0^2) - 1) \end{aligned} \tag{A7}$$

Solving eq. (A6) we obtain:

$$\beta(k) = k\log 2 + \tilde{\mu}_0\left(\frac{1}{2^k} - \alpha(k)\right) + \frac{\tilde{\sigma}_0^2}{2}\left(\frac{1}{2^{2Hk}} - \alpha^2(k)\right) \tag{A8}$$

Substituting equation (A8) in (A7), after algebraic manipulations we have:

$$\alpha^2(k) = \frac{1}{\tilde{\sigma}_0^2} \log\left(2^{2k(H-1)}\left(\exp\left(\frac{\tilde{\sigma}_0^2}{2^{2Hk}}\right) - 1\right) + 1\right) \tag{A9}$$

Without loss of generality we assume $\alpha(k) > 0$, then we derive the following relationships for the functions $\alpha(k)$ and $\beta(k)$:

$$\begin{cases} \alpha(k) = \dfrac{1}{\tilde{\sigma}_0}\sqrt{\log\left(2^{2k(H-1)}\left(\exp\left(\dfrac{\tilde{\sigma}_0^2}{2^{2Hk}}\right) - 1\right) + 1\right)} \\ \beta(k) = k \log 2 + \tilde{\mu}_0 \left(\dfrac{1}{2^k} - \alpha(k)\right) + \dfrac{\tilde{\sigma}_0^2}{2}\left(\dfrac{1}{2^{2Hk}} - \alpha^2(k)\right) \end{cases} \tag{A10}$$

Finally, we recall that $\tilde{\mu}_0$ and $\tilde{\sigma}_0^2$ respectively denote the mean and variance of the highest-level auxiliary variable $\tilde{Z}_{1,0}$. It can be easily shown that they can be expressed in terms of the known statistics $\mu_0$ and $\sigma_0^2$ of the given rainfall amount $Z_{1,0}$ at the largest scale, such as:

$$\begin{cases} \tilde{\mu}_0 = 2^k \left(\log\dfrac{\mu_0}{2^k} - \dfrac{1}{2}\log\left(2^{2k(1-H)}\dfrac{\sigma_0^2}{\mu_0^2} + 1\right)\right) \\ \tilde{\sigma}_0^2 = 2^{2Hk}\log\left(2^{2k(1-H)}\dfrac{\sigma_0^2}{\mu_0^2} + 1\right) \end{cases} \tag{A11}$$

## APPENDIX B

We provide herein some basic instructions to improve understanding of the implementation steps of our model.

1. *Input parameters*
   - Hurst coefficient *H*: it is dimensionless in the interval (0, 1);
   - Mean $\mu_0$ and variance $\sigma_0^2$ of the rainfall amount $Z_{1,0}$ to be disaggregated in time;
   - Last disaggregation step *k*: it is assumed that the desired length of the synthetic series to be generated is $2^k$, where *k* is a positive integer;
   - Probability dry $p_{0,k}$: probability that a certain time interval is dry after *k* disaggregation steps;

   Estimating such parameters from rainfall data series is relatively straightforward (see also Koutsoyiannis 2003b). In addition, it should be emphasized that our model fitting does not require the use of statistical moments of order higher than two, which are difficult to be reliably estimated from data (Lombardo *et al.* 2014).

2. *Auxiliary domain*

   By eq. (5) we transform the initial lognormal variable $Z_{1,0}$ into the auxiliary Gaussian variable $\tilde{Z}_{1,0}$ with mean $\tilde{\mu}_0$ and variance $\tilde{\sigma}_0^2$ given by eq. (A11).

3. *Disaggregation scheme*

   This is based on a dyadic random cascade structure (see e.g. Fig. 1) such that each higher-level amount is disaggregated into two lower-level amounts satisfying the equality constraint in eq. (16). The generation step is based on eq. (7) that can account for correlations with other variables previously generated.

4. *Adjusting procedure*

   By eq. (12), we transform lower-level variables generated in the auxiliary (Gaussian) domain back to the target (lognormal) domain, but eq. (16) is not satisfied anymore. To restore full consistency, we apply the power adjusting procedure to the disaggregated series, see eq. (17).

5. *Intermittency*

   By eq. (20), we introduce the intermittent character in the (adjusted and back-transformed) synthetic series at the "basic scale", which is represented by the last disaggregation step.

**APPENDIX C**

Let rainfall occurrences, $I_{j,k}$, evolve according to a discrete-time Markov chain with state space $\{0, 1\}$. This Markov chain is specified in terms of its state probabilities:

$$\begin{cases} p_{0,k} = \Pr\{I_{j,k} = 0\} \\ p_{1,k} = \Pr\{I_{j,k} = 1\} = 1 - p_{0,k} \end{cases} \tag{C1}$$

and the transition probabilities (based on Koutsoyiannis 2006, eq. (13)):

$$\begin{cases} \pi_{00,k} = \Pr\{I_{j,k} = 0 \middle| I_{j-1,k} = 0\} = p_{0,k} + \rho_1(1 - p_{0,k}) \\ \pi_{01,k} = \Pr\{I_{j,k} = 0 \middle| I_{j-1,k} = 1\} = p_{0,k}(1 - \rho_1) \\ \pi_{10,k} = \Pr\{I_{j,k} = 1 \middle| I_{j-1,k} = 0\} = 1 - \pi_{00,k} \\ \pi_{11,k} = \Pr\{I_{j,k} = 1 \middle| I_{j-1,k} = 1\} = 1 - \pi_{01,k} \end{cases} \tag{C2}$$

where $\rho_1 = \rho_{I,k}(1)$ is the lag-one autocorrelation coefficient of the Markov chain, and $p_{0,k}$ is the probability dry. Both are model parameters. Clearly, we assume that the parameters are such that the probabilities in (C2) are all strictly positive. Then, the Markov chain is ergodic, and, therefore, it has a unique stationary distribution. Hence, we can derive its autocorrelation function (ACF).

For a Markov chain, we can say that, conditional on the value of the previous variable $I_{j-1,k}$, the current variable $I_{j,k}$ is independent of all the previous observations. However, since each $I_{j,k}$ depends on its predecessor, this implies a non-zero correlation between $I_{j,k}$ and $I_{j+t,k}$, even for lag $t > 1$. In general, conditional independence between two variables given a third variable does not imply that the first two are uncorrelated.

To derive the ACF of our process, it can be easily shown that the correlation between variables one time period apart is given by the determinant of the one-step transition matrix **P** in (C2), such that:

$$\det(\mathbf{P}) = \rho_1 = \rho_{I,k}(1) \tag{C3}$$

Similarly, the correlation between variables $t$ time periods apart is given by the determinant of the $t$-step transition matrix $\mathbf{P}[t]$, i.e.:

$$\det(\mathbf{P}[t]) = \rho_{I,k}(t) \tag{C4}$$

Recall that the Markov property yields (see Papoulis, 1991, eq. (16-114), p. 638):

$$\mathbf{P}[t] = \mathbf{P}^t \tag{C5}$$

and that the basic properties of determinants imply:

$$\det(\mathbf{P}^t) = \left(\det(\mathbf{P})\right)^t \tag{C6}$$

Substituting eqs. (C5), (C4) and (C3) in eq. (C6), we obtain eq. (35).

# FIGURES

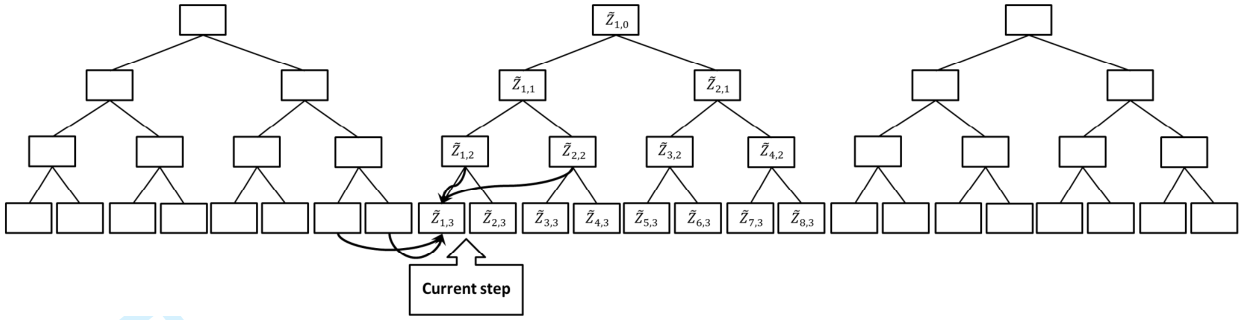


**Fig. 1:** Illustrative sketch for simulation of the auxiliary process $\tilde{Z}_{j,k}$. To eliminate "edge effects" in the generation procedure, we produce three (or five in case of $H \geq 0.9$) parallel cascades, then use only the one in the middle for simulations, and discard the remainder (adapted from Lombardo *et al*. 2012).

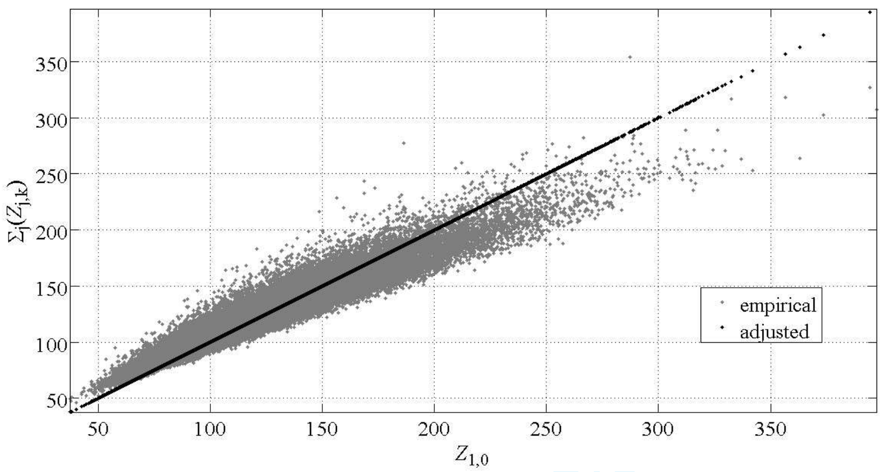


**Fig. 2**: Scatter plot of the calculated sum of lower-level variables vs. the given values of the higher-level variables $Z_{1,0}$ for all Monte Carlo experiments, where "empirical" and "adjusted" stand for original synthetic series and modified ones according to eq. (17), respectively.

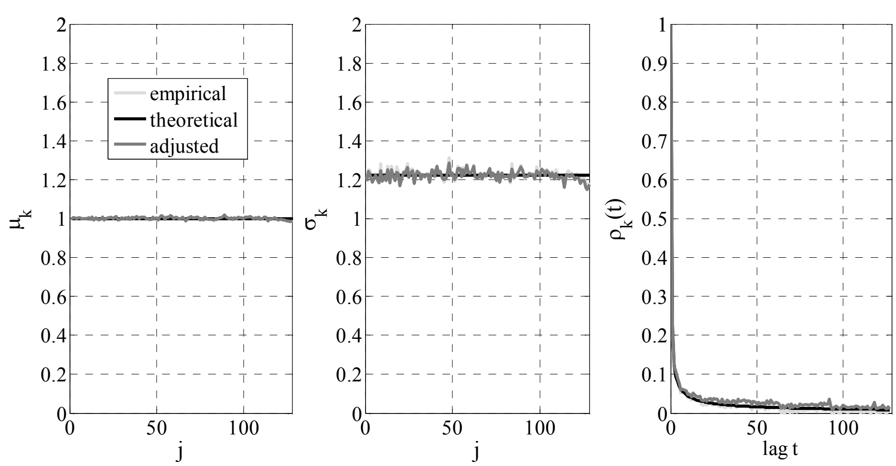

**Fig. 3**: Ensemble mean, standard deviation and autocorrelogram (from left to right, respectively) of the example disaggregation process as a function of the time position $j$ and lag $t$ after $k = 7$ cascade steps.

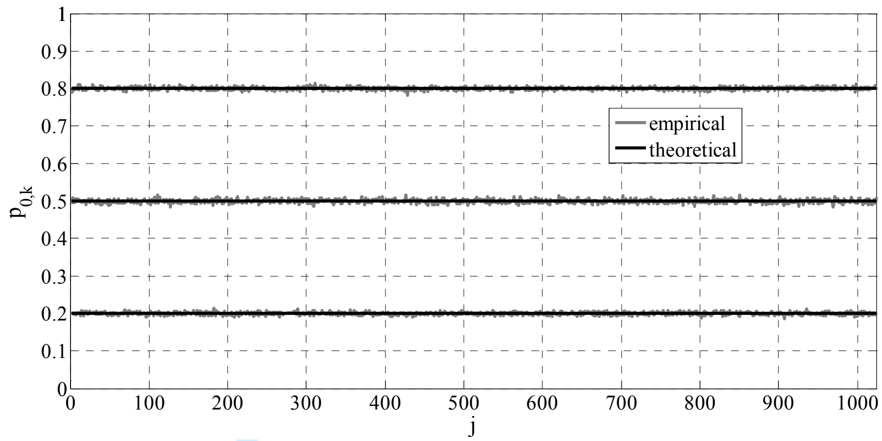


**Fig. 4**: Ensemble probability dry of three example intermittent processes (with both random and Markovian occurrences) as a function of the time position $j$ after $k = 10$ cascade steps.

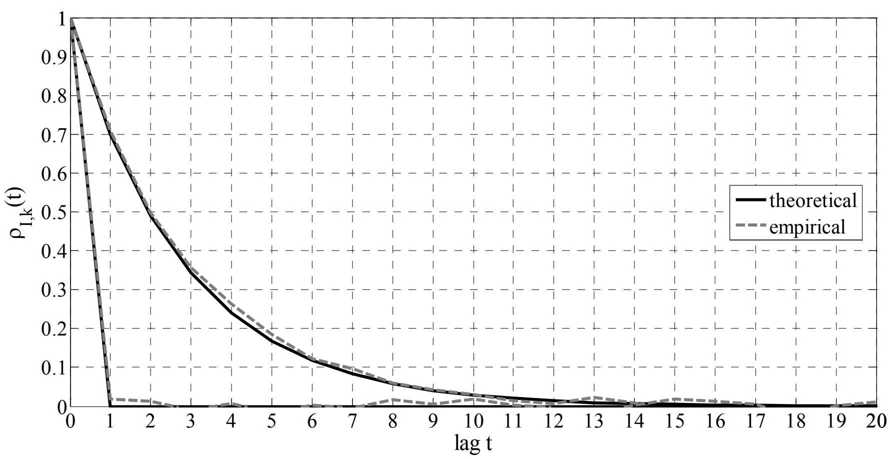


**Fig. 5:** Comparison between empirical and theoretical autocorrelation functions of random and Markovian occurrences for our simulations. Note that the former is a degenerate at zero for positive lags, according to eq. (32), while the latter follows the exponential form of eq. (35).

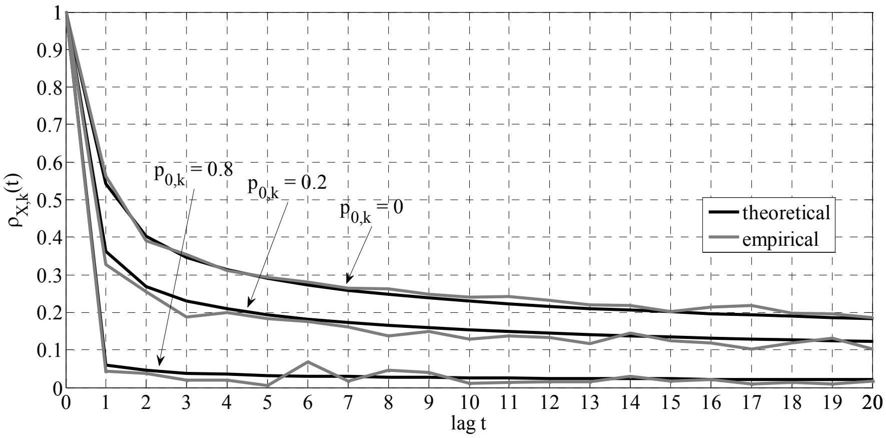


**Fig. 6**: Theoretical and empirical autocorrelograms of the entire rainfall process for three values of probability dry, i.e. $p_{0,k} = 0, 0.2, 0.8$; in case of purely random occurrences. Note that the autocorrelation function for $p_{0,k} = 0$ equals that of the rainfall depth process.

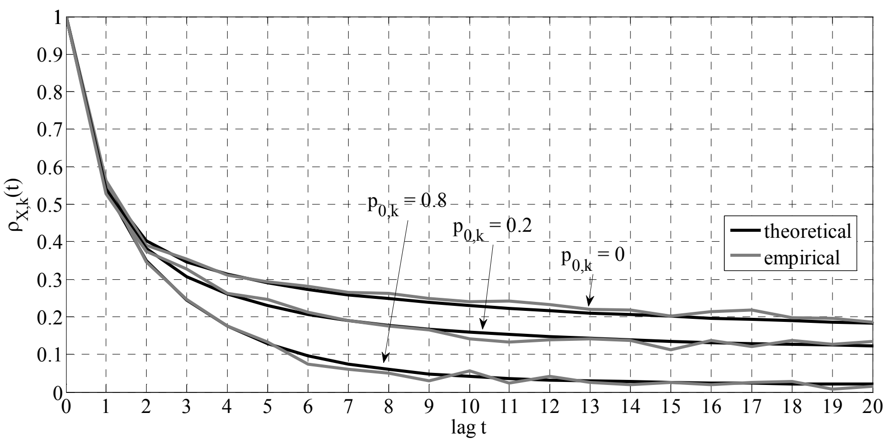


**Fig. 7**: Theoretical and empirical autocorrelograms of the entire rainfall process for three values of probability dry, i.e. $p_{0,k} = 0, 0.2, 0.8$; in case of Markovian occurrences. The autocorrelation function for $p_{0,k} = 0$ is plotted to show the impact of the occurrence model on the dependence structure of the intermittent process.

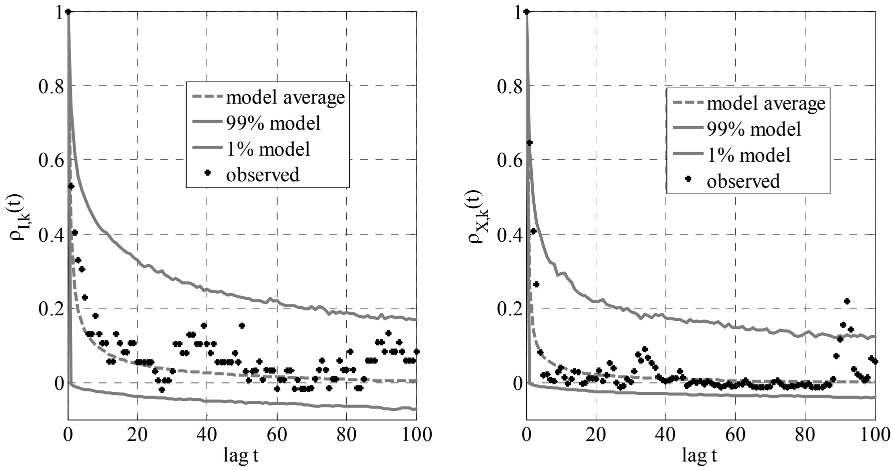


**Fig. 8**: Comparison between the simulated (average, 1st and 99th percentiles) and empirical autocorrelograms for the data series recorded at Viterbo raingauge station in January 1999. In the left and right panels, we show respectively the ACF of the occurrence (binary) process $\rho_{I,k}(t)$ and that of the intermittent (mixed) process $\rho_{X,k}(t)$. Estimated model parameters are: $\mu_k = 0.72, \sigma_k = 1.02, p_{0,k} = 0.96, H = 0.83$.

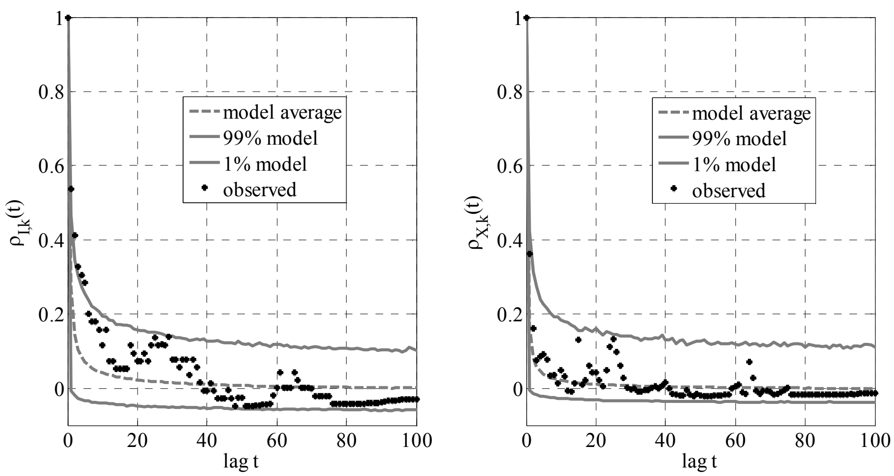


**Fig. 9**: Same as Fig. 8 for the data series recorded at Viterbo raingauge station in April 2003. Estimated model parameters are: $\mu_k = 0.61, \sigma_k = 0.65, p_{0,k} = 0.95, H = 0.7$.

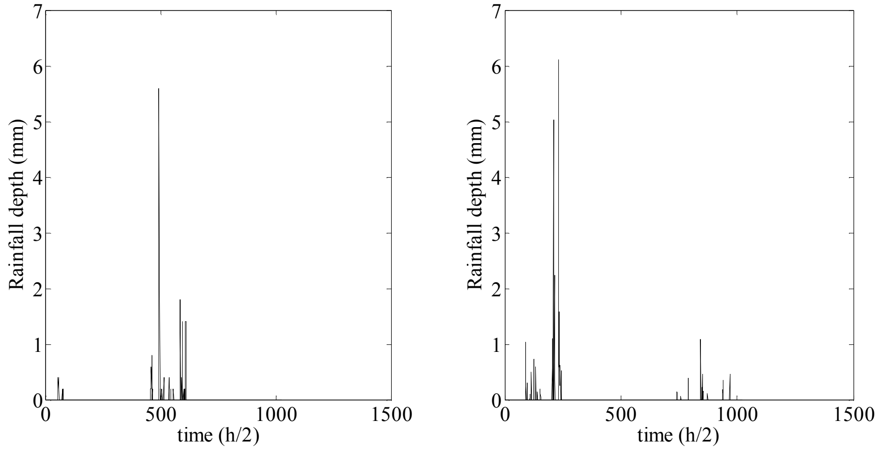


**Fig. 10**: Hyetograph of the rainfall data recorded at Viterbo raingauge station in January 1999 (left panel) along with the synthetic time series of equal length generated by our model (right panel).

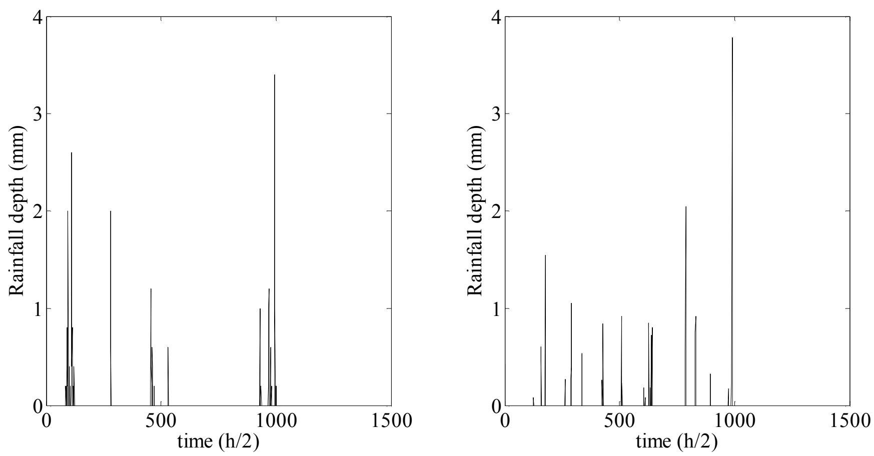


**Fig. 11**: Hyetograph of the rainfall data recorded at Viterbo raingauge station in April 2003 (left panel) along with the synthetic time series of equal length generated by our model (right panel).