# A theoretically consistent stochastic cascade for temporal disaggregation of intermittent rainfall

**F. Lombardo[1], E. Volpi[1], D. Koutsoyiannis[2], and 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: Federico Lombardo (federico.lombardo@uniroma3.it)

**Key Points:**

- Parsimonious generation of fine-scale time series of intermittent rainfall with prescribed dependence structure
- Modeling approach of mixed type in stationary setting, with discrete description of intermittency and continuous description of rainfall
- Analytical formulation of autocorrelation for mixed process without making any assumption about dependence structure or marginal probability

## Abstract

Generating fine-scale time series of intermittent rainfall that are fully consistent with any given coarse-scale totals is a key and open issue in many hydrological problems. We propose a stationary disaggregation method that simulates rainfall time series with given dependence structure, wet/dry probability, and marginal distribution at a target finer (lower-level) time scale, preserving full consistency with variables at a parent coarser (higher-level) time scale. We account for the intermittent character of rainfall at fine time scales by merging a discrete stochastic representation of intermittency and a continuous one of rainfall depths. This approach yields a unique and parsimonious mathematical framework providing general analytical formulations of mean, variance, and autocorrelation function (ACF) for a mixed-type stochastic process in terms of mean, variance, and ACFs of both continuous and discrete components, respectively. To achieve the full consistency between variables at finer and coarser time scales in terms of marginal distribution and coarse-scale totals, the generated lower-level series are adjusted according to a procedure that does not affect the stochastic structure implied by the original model. To assess model performance, we study 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 analytical formulations of main statistics of our mixed-type disaggregation model and show their clear accordance with Monte Carlo simulations. An application to rainfall time series from real world is shown as a proof of concept.

## 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. Both disaggregation and downscaling models refer to transferring information from a given scale (higher-level) to a smaller scale (lower-level), e.g. they generate consistent rainfall time series at a specific scale given a known precipitation measured or simulated at a certain coarser scale. The two approaches are very similar in nature but not identical to each other. Downscaling aims at producing the finer-scale time series with the required statistics, being statistically consistent with the given variables at the coarser scale, while disaggregation has the additional requirement to produce a finer scale time series that adds up to the given coarse-scale total.

Although there is substantial experience in stochastic disaggregation of rainfall to fine time scales, most modeling schemes existing in the literature are ad hoc techniques rather than consistent general methods [see review by *Koutsoyiannis*, 2003a]. Disaggregation models were introduced in hydrology by the pioneering work of *Valencia and Schaake* [1973], who proposed a simple linear disaggregation model that is fully general for Gaussian random fields without intermittency. However, the skewed distributions and the intermittent nature of the rainfall process at fine time scales 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. Our model exploits the full generality and theoretical consistency of linear disaggregation schemes proposed by *Valencia and Schaake* [1973] for Gaussian random variables, but it generates intermittent time series with lognormal distribution that are more consistent with the actual rainfall process at fine time scales.

The following sections expand on a stochastic approach to rainfall disaggregation in time, with an emphasis on the analytical description of a model of the mixed (discrete-continuous) type. Firstly, we generate lognormal time series of rainfall depths with prescribed mean, variance and autocorrelation function (ACF) based on fractional Gaussian noise (fGn), also known as Hurst-Kolmogorov (HK) process [*Mandelbrot and Van Ness*, 1968]. Note that the lognormality hypothesis and our specific normalizing transformation (see next section) enable the analytical formulation of the main statistics of the rainfall depth process. Secondly, we obtain the intermittent rainfall process by multiplying the synthetic rainfall depths above by user-specified binary sequences (i.e., rainfall occurrences) with given mean and ACF. The resulting stochastic model is of the mixed type and we derive its summary statistics in closed forms.

We propose herein an evolution of the downscaling model by *Lombardo et al*. [2012], which is upgraded and revised to include both a stochastic model accounting for intermittency and an appropriate strategy to preserve the additive property. The preservation of the additive property distinguishes indeed disaggregation from downscaling. 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.
- Intermittency. The main novelty of this paper is the introduction of intermittency in the modeling framework, which is fully general and it can be used when simulating mixed-

type processes other than rainfall from the real world. 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 modeling of rainfall intermittency relate to the study of the rainfall occurrence process. Then, we need to introduce the latter in our modeling framework. In order to achieve such an objective, in Section 3, we describe the entire rainfall process using a two-state stochastic process comprising a discrete and a continuous component accounting for rainfall occurrences and non-zero rainfall, respectively. Our modeling framework enables the analytical formulation of the main statistics of the discrete-continuous rainfall process.

- Additivity constraint. We utilize auxiliary Gaussian variables to disaggregate a given rainfall amount to a certain scale of interest by means of the linear generation scheme proposed by *Koutsoyiannis* [2002]. Nevertheless, rainfall is effectively modelled by positively skewed distributions, i.e. non-Gaussian. Hence, then an exponential transformation of the variables is used in a way that the transformed variables follow a 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 linear disaggregation scheme, is lost [*Todini*, 1980]. To overcome the problem we apply an empirical correction procedure, known as "power adjusting procedure" (Section 4), 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].

- Monte Carlo experiments and comparison to observed data. In Section 5 and 6, we show respectively some Monte Carlo experiments and a case study in order to test the capability of our model to reproduce the statistical behavior of synthetic and real rainfall time series. We conclude our work with Section 7, where we give an overview on the key ideas and briefly discuss the applicability aspects of our approach.

## 2 Basic concepts and background

In rainfall modeling 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]. MRCs are discrete models in scale, meaning that the scale ratio from parent to child structures is an integer number strictly larger than one. These models are multiplicative, and embedded in a recursive manner. Each step is usually associated to a scale ratio of $b = 2$ (i.e. branching number); after $m$ cascade steps ($m = 0, 1, 2, \ldots$), the total scale ratio is $2^m$, and we have:

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

where $j = 1, \ldots, 2^m$ is the index of position (i.e., time step) in the series at the cascade step $m$, and $i$ is the index of the cascade step. $R_{1,0}$ denotes the initial rainfall intensity to be distributed over the (subscale) cells $R_{j,m}$ of the cascade, each cell being associated to a random variable $W_{g(i,j),i}$ (i.e. cascade generator, called "weight") where $g(i,j) = \left\lceil \frac{j}{2^{m-i}} \right\rceil$ denotes a ceiling function which defines the position in time at the cascade step $i = 0, \ldots, m$ [see e.g. *Gaume et al.*, 2007]. 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. A graphical example of a dyadic ($b$ = 2) multiplicative cascade with four cascade steps ($m$ = 0, 1, 2, 3) is shown in Fig. 1.

As detailed by *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 due to the model structure. For example, it can be shown that for MRCs we may write lagged second moments after $m$ cascade steps as:

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

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

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

where $\Theta[n]$ is the discrete form of the Heaviside step function, defined for an integer $n$ as:

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

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 emphasize that several researchers and practitioners often neglect this nonstationarity, which is simply inherent to the model structure. 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 conceptual 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. In fact, ergodicity is a topic dealing with the relationship between statistical averages and sample averages, which is a central problem in the estimation of statistical parameters in terms of real 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*, 2015]. 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; *Pui et al.*, 2012; *Efstratiadis et al.*, 2014]. 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 and it proves to be stationary. Indeed, we emphasize that this model is not a MRC; for a detailed theoretical and numerical comparison of this model with discrete MRCs, the reader is referred to *Lombardo et al.* [2012].

Our rainfall disaggregation model (see also Appendix B for a step-by-step implementation procedure) exploits knowledge from an auxiliary Gaussian domain where fGn is generated by means of a stepwise disaggregation approach based on a random cascade structure. Then, we assume the given rainfall amount $Z_{1,0}$ at the initial largest scale (*m* = 0) to be lognormally distributed with a given mean $\mu_0$ and variance $\sigma_0^2$, and we log-transform it into an auxiliary Gaussian variable $\tilde{Z}_{1,0}$ with mean $\tilde{\mu}_0$ and variance $\tilde{\sigma}_0^2$ given by eq. (A11), as follows:

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

where $\alpha(k)$ and $\beta(k)$ are two functions given in eq. (A10), that depend on a given disaggregation step $m = k$, which is the last disaggregation step of interest. Hence, it is assumed that the desired length of the synthetic series to be generated is $2^k$, where $k$ is a given positive integer. The functions $\alpha(k)$ and $\beta(k)$ are introduced to preserve some scaling properties of the auxiliary Gaussian process, as then better described in 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 $\tilde{Z}_{j,m-1}$ is disaggregated into two lower-level amounts $\tilde{Z}_{2j-1,m}$ and $\tilde{Z}_{2j,m}$ such that (see explanatory sketch in Fig. 2, where $j = 3$ and $m = 3$):

$$\tilde{Z}_{2j-1,m} + \tilde{Z}_{2j,m} = \tilde{Z}_{j,m-1} \tag{6}$$

Thus, we generate the variable of the first subinterval $\tilde{Z}_{2j-1,m}$ 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. $\tilde{Z}_{1,m}, \dots, \tilde{Z}_{2j-2,m}$, and of the next higher-level time steps, i.e. $\tilde{Z}_{j,m-1}, \dots, \tilde{Z}_{s,m-1}$ where $s = 2^{m-1}$. Theoretically, it is necessary to preserve the correlations of $\tilde{Z}_{2j-1,m}$ 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 two lower-level time steps behind and one higher-level time step ahead [*Koutsoyiannis*, 2002]. This is particularly the case if we wish to generate fGn with moderate values of the Hurst parameter

$H \in (0,1)$. In our work we are interested in positively correlated processes, therefore $0.5 < H < 1$. The fGn reduces to white noise for $H = 0.5$.

Even though the scheme sketched in Fig. 2 is already good for most practical purposes, if we wish to generate highly correlated time series, i.e. with high values of the Hurst parameter (e.g. $H \geq 0.9$), then we could expand the number of variables that are considered in the generation procedure. An extensive numerical investigation (not reported here) showed that we obtain the best trade-off between model accuracy and computational burden if we consider two more lower-level time steps behind and one more higher-level time step ahead with respect to the sketch in Fig. 2.

In either case, we use the following linear generation scheme:

$$\tilde{Z}_{2j-1,m} = \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{Y} + V \tag{7}$$

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 $\boldsymbol{\theta}$ and the variance of the innovation term $V$ needed to solve eq. (7) can be estimated applying the methodology proposed by *Koutsoyiannis* [2001] that is based on a generalized mathematical proposition, which ensures preservation of marginal and joint second-order statistics and of linear relationships between lower- and higher-level variables:

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

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

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. In the example of Fig. 2, we consider correlations with two lower-level time steps behind and one higher-level time step ahead, then $\boldsymbol{Y} = \left[\tilde{Z}_{2j-3,m}, \tilde{Z}_{2j-2,m}, \tilde{Z}_{j,m-1}, \tilde{Z}_{j+1,m-1}\right]^{\mathrm{T}}$ where superscript T denotes the transpose of a vector. Hence, eq. (7) simplifies as follows:

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

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. From eqs. (8) and (9), all unknown parameters can be estimated in terms of fGn correlations, which are independent of $j$ and $m$:

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

Therefore, in this case we can write eq. (8) as follows:

$$\begin{bmatrix} a_2 \\ a_1 \\ b_0 \\ b_1 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & \tilde{\rho}(1) & \tilde{\rho}(2)+\tilde{\rho}(3) & \tilde{\rho}(4)+\tilde{\rho}(5) \\ \tilde{\rho}(1) & 1 & \tilde{\rho}(1)+\tilde{\rho}(2) & \tilde{\rho}(3)+\tilde{\rho}(4) \\ \tilde{\rho}(2)+\tilde{\rho}(3) & \tilde{\rho}(1)+\tilde{\rho}(2) & 2[1+\tilde{\rho}(1)] & \tilde{\rho}(1)+2\tilde{\rho}(2)+\tilde{\rho}(3) \\ \tilde{\rho}(4)+\tilde{\rho}(5) & \tilde{\rho}(3)+\tilde{\rho}(4) & \tilde{\rho}(1)+2\tilde{\rho}(2)+\tilde{\rho}(3) & 2[1+\tilde{\rho}(1)] \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\rho}(2) \\ \tilde{\rho}(1) \\ 1+\tilde{\rho}(1) \\ \tilde{\rho}(2)+\tilde{\rho}(3) \end{bmatrix}$$

For the fGn, we can write $\tilde{\sigma}_m^2 = \mathrm{var}\left[\tilde{Z}_{2j-1,m}\right] = \tilde{\sigma}_0^2/2^{2Hm}$, where $\tilde{\sigma}_0^2 = \mathrm{var}\left[\tilde{Z}_{1,0}\right]$ [*Koutsoyiannis*, 2002]. Then eq. (9) becomes:

$$\mathrm{var}[V] = \tilde{\sigma}_m^2 \left(1 - [\tilde{\rho}(2), \tilde{\rho}(1), 1+\tilde{\rho}(1), \tilde{\rho}(2)+\tilde{\rho}(3)][a_2, a_1, b_0, b_1]^{\mathrm{T}}\right) \qquad (13)$$

Then, the two equations above depend solely on the Hurst parameter $H$ and the variance $\tilde{\sigma}_0^2$ given by eq. (A11).

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, see Fig. 2). In practice for each cascade step, when we generate $\tilde{Z}_{2j-1,m}$ 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 at the start or end of the cascade sequence, for each cascade step $m$, where $m = 0, \ldots, k$. To overcome this "edge" problem, 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. 3. 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 disaggregated series with the desired length $2^k$ generated in the auxiliary (Gaussian) domain must then be transformed back to the target (lognormal) domain (actual rainfall) by the following simple exponentiation:

$$Z_{j,k} = \exp\left(\tilde{Z}_{j,k}\right) \tag{14}$$

This transformation is simpler than that used by *Lombardo et al.* [2012]. In fact, we normalize the given coarse-scale total $Z_{1,0}$ by eq. (5) in order to use a simpler inverse transformation, eq. (14), at the scale $k$ of interest. This is more appropriate for a disaggregation approach resembling a top-down strategy. As shown in Appendix A, the mean, variance and ACF of the disaggregated rainfall process so obtained are given respectively by:

$$\mu_k = \langle Z_{j,k} \rangle = \mu_0 / 2^k \tag{15}$$

$$\sigma_k^2 = \mathrm{var}\left[Z_{j,k}\right] = \sigma_0^2 / 2^{2Hk} \tag{16}$$

$$\rho_k(t) = \mathrm{corr}\left[Z_{j,k}, Z_{j+t,k}\right] = \frac{\exp\left(\tilde{\sigma}_k^2 \tilde{\rho}(t)\right) - 1}{\exp(\tilde{\sigma}_k^2) - 1} \tag{17}$$

where $\tilde{\rho}(t)$ and $\tilde{\sigma}_k^2 = \mathrm{var}[\tilde{Z}_{j,k}]$ respectively denote the ACF (eq. 11) and the variance of the auxiliary Gaussian process (i.e., HK process or fGn), $H$ 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 ACFs of the HK 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 modeling 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 linear disaggregation schemes [*Koutsoyiannis*, 2003a]. Indeed, we simulate a fGn in the auxiliary domain whose characteristics are modified (by eq. 5) based on the last disaggregation step of interest $k$, in order to obtain (by eq. 14) $2^k$ variables in the lognormal (target) domain with the desired statistical properties given by eqs. (15)-(17).

## 3 Introducing 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, modeling 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 modeling 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 modeling 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 modeling 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 can write the intermittent rainfall as the product of those two components.

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

$$X_{j,k} = I_{j,k} \cdot Z_{j,k} \tag{18}$$

where $Z_{j,k}$ denotes the continuous-type random variable pertaining to our disaggregation model (given by eq. (14)), 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 $k$ disaggregation 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:

$$\langle X_{j,k} \rangle = \left(1 - p_{0,k}\right)\mu_k \tag{19}$$

$$\mathrm{var}\left[X_{j,k}\right] = \left(1 - p_{0,k}\right)\left(\sigma_k^2 + p_{0,k}\mu_k^2\right) \tag{20}$$

where $\mu_k$ and $\sigma_k^2$ denote the mean and the variance of the series generated by the rainfall depth model (see eqs. (15) and (16), respectively).

Note that eq. (18) resembles the classical intermittent lognormal *β*-model based on MRCs [*Gupta and Waymire*, 1993; *Over and Gupta*, 1994, 1996], but it is more general and embedded into our Hurst-Kolmogorov modeling framework.

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

$$\mathrm{cov}\left[X_{j,k}, X_{j+t,k}\right] = \langle X_{j,k} X_{j+t,k} \rangle - \langle X_{j,k} \rangle^2 \tag{21}$$

where the last term of the right-hand side can be calculated from eq. (19), while the lagged second moment $\langle X_{j,k} X_{j+t,k} \rangle$ can be expressed through the following joint probabilities:

$$\begin{aligned} p_{00,k} &= \Pr\left\{X_{j,k} = 0, X_{j+t,k} = 0\right\} \\ p_{10,k} &= \Pr\left\{X_{j,k} > 0, X_{j+t,k} = 0\right\} \\ p_{01,k} &= \Pr\left\{X_{j,k} = 0, X_{j+t,k} > 0\right\} \\ p_{11,k} &= \Pr\left\{X_{j,k} > 0, X_{j+t,k} > 0\right\} \end{aligned} \tag{22}$$

Therefore, by total probability theorem and eq. (18), we have:

$$\langle X_{j,k} X_{j+t,k} \rangle = p_{11,k} \langle Z_{j,k} Z_{j+t,k} \rangle \tag{23}$$

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

$$\mathrm{cov}\left[I_{j,k}, I_{j+t,k}\right] = \langle I_{j,k} I_{j+t,k} \rangle - \langle I_{j,k} \rangle^2 = p_{11,k} - \left(1 - p_{0,k}\right)^2 \tag{24}$$

The derivation of this equation is based on the relationships $\langle I_{j,k} \rangle = \langle I_{j,k}^2 \rangle = 1 - p_{0,k}$, and $\langle I_{j,k} I_{j+t,k} \rangle = p_{11,k}$. Thus, from eq. (24) we obtain:

$$p_{11,k} = \left(1 - p_{0,k}\right)^2 + \mathrm{cov}\left[I_{j,k}, I_{j+t,k}\right] \tag{25}$$

Substituting eqs. (19), (23) and (25) in eq. (21), it follows:

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

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. (26), yields:

$$\begin{aligned} &\operatorname{cov}[X_{j,k}, X_{j+t,k}] = \\ &= \left(\left(1 - p_{0,k}\right)^2 + \operatorname{cov}[I_{j,k}, I_{j+t,k}]\right) \operatorname{cov}[Z_{j,k}, Z_{j+t,k}] + \operatorname{cov}[I_{j,k}, I_{j+t,k}]\mu_k^2 \end{aligned} \tag{27}$$

Hence, eq. (27) expresses 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:

$$\rho_{X,k}(t) = \frac{\operatorname{cov}[X_{j,k}, X_{j+t,k}]}{\operatorname{var}[X_{j,k}]} \tag{28}$$

Recalling that $\operatorname{var}[I_{j,k}] = p_{0,k}(1 - p_{0,k})$ and substituting eqs. (20) and (27) in eq. (28), after algebraic manipulations we obtain:

$$\rho_{X,k}(t) = \frac{\left(1 - p_{0,k} + \rho_{I,k}(t)p_{0,k}\right)\rho_k(t)\sigma_k^2 + \rho_{I,k}(t)p_{0,k}\mu_k^2}{\sigma_k^2 + p_{0,k}\mu_k^2} \tag{29}$$

where $\mu_k$, $\sigma_k^2$ and $\rho_k(t)$ are given by eqs. (15), (16) and (17) respectively. The only unknown in eq. (29) is the ACF $\rho_{I,k}(t)$ of the rainfall occurrence process at the finer characteristic time scale (i.e., the final disaggregation step *k*). When deriving the theoretical ACF in eq. (29), note that we have not made any assumption about the dependence structure or the marginal probability of the process; the only assumption is that the process is stationary. Eq. (29) is fully general and new, to the best of our knowledge; it can be used to derive the theoretical ACF of a mixed-type stochastic process in terms of its discrete and continuous components (provided they are independent of one another).

In order 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].

In early stages of analysis and modeling attempts, the Markov chain model was widely adopted for discrete time representations of rainfall occurrences, recognizing that they are not independent in time [*Gabriel and Neumann*, 1962; *Haan et al*., 1976; *Chin*, 1977]. However, later studies observed 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]. For the sake of numerical investigation, hereinafter we analyze the first two modeling categories of the occurrence processes:

1. Purely random model.
2. Markov chain model.

To summarize, we believe it is worth repeating here a short overview on some of the key ideas of our model. A continuous model (described in Section 2) to generate finer-scale time series of lognormal rainfall depths with HK-like dependence structure, and an arbitrary binary model (e.g. Bernoulli, Markov, etc.) to simulate rainfall intermittency are combined by eq. (18) to give rise to a complete rainfall disaggregation model characterized by mean, variance, and ACF as in eqs. (19), (20) and (29), respectively. The preservation of the additive property is guaranteed by applying eq. (36) to the generated series (see next section). The intermittent component refers exclusively to the target scale, and is combined with the continuous component at that scale. Note that mean and variance in eqs. (19) and (20) are independent of the specific model, while the ACF in eq. (29) relies on the dependence structures of both the continuous and binary components. In the following, we show how this ACF specializes for intermittent components with Bernoulli and Markov structures.

### 3.1 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:

$$\rho_{I,k}(t) = \mathrm{cov}\left[I_{j,k}, I_{j+t,k}\right] = 0 \tag{30}$$

Substituting eq. (30) in eqs. (27) and (29), we obtain respectively:

$$\mathrm{cov}\left[X_{j,k}, X_{j+t,k}\right] = \left(1 - p_{0,k}\right)^2 \mathrm{cov}\left[Z_{j,k}, Z_{j+t,k}\right] \tag{31}$$

$$\rho_{X,k}(t) = \left(1 - p_{0,k}\right)\rho_k(t)\frac{\sigma_k^2}{\sigma_k^2 + p_{0,k}\mu_k^2} \tag{32}$$

### 3.2 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 $\rho_{I,k}(1) = \mathrm{corr}\left[I_{j,k}, I_{j-1,k}\right]$. 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):

$$\rho_{I,k}(t) = \mathrm{corr}\left[I_{j,k}, I_{j+t,k}\right] = \rho_{I,k}^{|t|}(1) \tag{33}$$

Substituting in eq. (29), we derive the autocorrelation of the entire rainfall process as:

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

## 4 Adjusting procedure

A shortcoming of the above-summarized model is that generated, back-transformed rainfall amounts, $Z_{j,k}$, generally fail to sum to the coarse-scale total, $Z_{1,0}$, which is a major requirement of disaggregation methods. Therefore, analogous considerations apply to the corresponding intermittent rainfall process $X_{j,k}$, where the coarse-scale total $X_{1,0} = \left(1 - p_{0,k}\right) Z_{1,0}$ is known. 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:

$$X_{1,0} = \sum_{j=1}^{s=2^k} X_{j,k} \tag{35}$$

A relevant question is how to adjust the generated rainfall time series 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 $X_{j,k}$. This procedure allocates the error in the additive property among the lower-level variables. Thus, it modifies the generated variables $X_{j,k}$ ($j = 1, \ldots, 2^k$) to get the adjusted ones $X'_{j,k}$ according to:

$$X'_{j,k} = X_{j,k} \left( \frac{X_{1,0}}{\sum_{j=1}^{s} X_{j,k}} \right)^{\lambda_{j,k}/\eta_{j,k}} \tag{36}$$

where

$$\lambda_{j,k} = \frac{\sum_{i=1}^{s} \mathrm{cov}\left[X_{j,k}, X_{i,k}\right]}{\sum_{j=1}^{s} \sum_{i=1}^{s} \mathrm{cov}\left[X_{j,k}, X_{i,k}\right]} \tag{37}$$

$$\eta_{j,k} = \frac{\langle X_{j,k} \rangle}{\sum_{j=1}^{s} \langle X_{j,k} \rangle} \tag{38}$$

The power adjusting procedure is more effective and suitable for our modeling 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, whereas 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. (36) must be iterative, until the calculated sum of the lower-level variables equals the given $X_{1,0}$. Although iterations slightly reduce the model speed, the power adjusting procedure greatly outperforms the other procedures in terms of accuracy.

## 5 Numerical simulations

A Monte Carlo simulation is carried out to assess model performance and analytical results under controlled conditions. We generate 10000 time series assuming the following parameters to model rainfall depths as described in Section 2: $k = 10$, $\mu_0 = 1024$, $\sigma_0 = 362.04$, and $H = 0.85$. Then, according to eqs. (15) and (16), we obtain the lower-level series, $Z_{j,k}$, of size $s = 2^k = 1024$, and unit mean and variance $\mu_k = \sigma_k^2 = 1$. In order to simulate rainfall occurrences described in Section 3, we generate binary sequences, $I_{j,k}$, with Markovian dependence structure by implementing Boufounos (2007) algorithm with three different values of probability dry $p_{0,k} \in \{0.2, 0.5, 0.8\}$ and the lag-one autocorrelation coefficient $\rho_{I,k}(1) = 0.7$ as an additional model parameter. Then, the three mixed-type (intermittent) processes, $X_{j,k}$, are derived by applying eq. (18) to the synthetic series of $Z_{j,k}$ and $I_{j,k}$ for each value of $p_{0,k}$. Finally, we apply the adjusting procedure in eq. (36) to let the generated variables $X_{j,k}$ satisfy the additivity constraint in eq. (35).

According to eqs. (19), (20) and the values of $p_{0,k}$ given above, the simulated intermittent rainfall processes have mean $\langle X_{j,k} \rangle \in \{0.8, 0.5, 0.2\}$ and variance $\mathrm{var}\left[X_{j,k}\right] \in \{0.96, 0.75, 0.36\}$. Fig. 4 shows that the adjusted variables fulfil the additive property, while Fig. 5 confirms that summary statistics of the generated variables are well preserved by the adjusting procedure.

Figs. 6 and 7 show empirical vs. theoretical ACFs of two different mixed-type processes assuming respectively purely random and Markovian occurrences, $I_{j,k}$, with the same parameters as above (clearly, for random occurrences we have $\rho_{I,k}(1) = 0$). Note that both figures also depict the case with null probability dry, i.e. $p_{0,k} = 0$, which corresponds to the rainfall depth process, $Z_{j,k}$. The ACF of the latter is used as a benchmark to compare the two figures together in order to investigate the influence of each occurrence model on the dependence structure of the entire process, $X_{j,k}$. As expected, both of our occurrence models are generally cause for decorrelation of the intermittent process with respect to the process without intermittency. This is particularly the case if we model rainfall occurrences by a white noise as in 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 $p_{0,k} \in \{0.2, 0.8\}$).

## 6 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 time scales, 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 Section 3, the dependence structure of the rainfall occurrence process appears to be non-Markovian (not shown). 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 model calibration on observational data, the Hurst parameter $H$ 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 30-minute rainfall time series. As this represents a realization of the lower-level intermittent rainfall process, $X_{j,k}$, with mean and variance given by eqs. (19) and (20), respectively, such statistical properties can be therefore estimated directly from data. Once the probability dry, $p_{0,k}$, is derived from data, we can solve eqs. (19) and (20) for the remaining two parameters to be estimated, i.e. the mean $\mu_k$ and variance $\sigma_k^2$ of the rainfall depth process, $Z_{j,k}$ (the higher-level counterparts $\mu_0$ and $\sigma_0^2$ are easily derived from eqs. (15) and (16)). For simplicity, here it is assumed that the desired length $s$ of the synthetic series to be generated is $s = 2^{10}$, i.e. $k = 10$, which is similar to sample sizes of the monthly data series under consideration (i.e., number of 30-minute intervals in each month). However, the model works equally well (not shown) if one increases $s$ to the next power of 2 and then discards the redundant generated items before performing the adjusting procedure. Hence, we have a very parsimonious disaggregation model in time with only four parameters: $k$, $\mu_0$, $\sigma_0$, and $H$.

We perform 10000 Monte Carlo experiments following the procedure described in sections above. First, we generate correlated series (Section 2) of rainfall amounts, $Z_{j,k}$, with ACF in eq. (17). Second, we generate correlated binary series of rainfall occurrences, $I_{j,k}$, with ACF in eq. (11) (for a detailed description of the simulation algorithm, refer to *Serinaldi and Lombardo* [2017]). The outcomes of the two generation steps above are therefore combined by eq. (18) to obtain the synthetic intermittent series, $X_{j,k}$, with ACF in eq. (29). Finally, we apply to $X_{j,k}$ the procedure in eq. (36) to get the adjusted process, $X'_{j,k}$, that satisfies the additive property in eq. (35).

By way of example, Figs. 8 and 9 respectively compare the observed autocorrelograms for January 1999 and April 2003 data series against the ACFs simulated by our model. The ACF of the occurrence (binary) process $\rho_{I,k}(t)$ and that of the intermittent (mixed) process $\rho_{X,k}(t)$ are shown in the left and right panels of each figure, respectively. In either case, the model fits on average the observed behavior satisfactorily. Other summary statistics such as the mean, variance and probability dry of the data series are preserved by construction (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 behavior observed in real rainfall time series.

## 7 Conclusions

The discrete MRC is the dominant approach to rainfall disaggregation in hydrological modeling 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 amounts 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 modeling 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. The model combines 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 adopt to a top-down approach resulting in a modular modeling strategy comprising a discrete (binary) description of intermittency and a continuous description of rainfall amounts. A stochastic process with lognormal random variables and Hurst-Kolmogorov dependence structure gives the latter, while the former is based on a user-specified model of rainfall occurrences. We provide general theoretical formulations for summary statistics of the mixed-type process as functions of those of the two components. We stress that these relationships are fully general and hold true for whatever stationary mixed process independently of the specific form of the continuous and discrete components. For illustration purposes, it is shown how formulae specialize for 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. Monte Carlo experiments confirm the correctness of the analytical derivations and highlight the good performance of the proposed model under controlled conditions.

Since our method utilizes nonlinear transformations of the variables in the generation procedure, the additivity constraint between lower- and higher-level variables, i.e. the mass conservation, is not satisfied and must be restored. For this purpose, we use an accurate adjusting procedure that preserves explicitly the first- and second-order statistics of the generated intermittent rainfall. Consequently, the original downscaling model by *Lombardo et al*. [2012] now becomes a disaggregation model.

Intermittent rainfall time series from the real world are compared with simulations drawn from a very parsimonious four-parameter version of the proposed model, confirming its remarkable potentiality and accuracy in reproducing marginal distributions, correlation structure, intermittency, and clustering.

In order to make our stationary disaggregation model an operational tool, we need to account for seasonal fluctuations observed in historical rainfall records at sub-annual time scales. To a first approximation, *Marani* [2003] suggests assuming that different stationary stochastic processes generate the rainfall records from each month of the year. Hence, we should estimate twelve sets of model parameters and then perform simulations for the entire year accordingly.

Finally, our work provides a theoretically consistent methodology that can be applied to disaggregate actual rainfall (or model outputs) at fine time scales, which can be used in several fields that have been significant catalysts for the development of recent hydrological research. In fact, a wide range of studies concerning e.g. climate change impacts, resilience of urban areas to hydrological extremes, and online prediction/warning systems for urban hydrology require accurate characterization of rainfall inputs at fine time scales [*Lombardo et al*., 2009; *Fletcher et al*., 2013; *Tabari et al*., 2016; *McCabe et al*., 2017]. Hence, complete rainfall disaggregation methods with solid theoretical basis together with reliable data series are crucial to meet these needs.

**Appendix A**

We assume that the disaggregated rainfall process at the last disaggregation step $k$ 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 $\tilde{\sigma}_k^2$ of the fractional Gaussian noise (fGn) as follows:

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

In fact, recall that $\tilde{\mu}_k = \tilde{\mu}_0/2^k$ and that for the fGn we can write $\tilde{\sigma}_k^2 = \tilde{\sigma}_0^2/2^{2Hk}$, where $0 < H < 1$ is the Hurst coefficient [*Mandelbrot and Van Ness*, 1968].

Then, our primary goal is to let the target process $Z_{j,k}$ follow analogous scaling rules to those of the auxiliary process $\tilde{Z}_{j,k}$. In other words, we want the following laws to hold true for the target process $Z_{j,k}$:

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

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), but rainfall models require positively correlated processes, therefore $0.5 < H < 1$.

   – 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.

   – Mean $\mu_0$ and variance $\sigma_0^2$ of the rainfall amount $Z_{1,0}$ to be disaggregated in time, which are related to their counterparts of the higher-level intermittent rainfall $X_{1,0}$ by eqs. (19) and (20).

   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. 2) such that each higher-level amount is disaggregated into two lower-level amounts satisfying the additivity constraint in eq. (6). The generation step is based on eq. (7) that can account for correlations with other variables previously generated. By eq. (14), we transform lower-level variables generated in the auxiliary (Gaussian) domain back to the target (lognormal) domain, but the additive property is not satisfied anymore.

4. *Intermittency*

   By eq. (18), we introduce the intermittent character in the (back-transformed) synthetic series at the "basic scale", which is represented by the last disaggregation step *k*.

5. *Adjusting procedure*

   To ensure the full consistency between lower- and higher-level variables, we apply the power adjusting procedure to the disaggregated intermittent series. Then, the additive property is restored without modifying the summary statistics of the original variables.

## 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. (33).

**Acknowledgments**

Rainfall data used in this study are available from authors upon request. Two eponymous reviewers, Marco Marani and Salvatore Grimaldi, and an anonymous reviewer are gratefully acknowledged for their constructive comments that helped to substantially improve the paper. We especially thank the editor Alberto Montanari for his encouraging comments and support.

**References**

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

Blöschl, G., and M. Sivapalan (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., Y. Dialynas, S. Kozanis, and D. Koutsoyiannis (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.

Fletcher, T. D., H. Andrieu, and P. Hamel (2013), Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art, *Advances in Water Resources*, 51, 261–279.

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

Gaume, E., N. Mouhous, and H. Andrieu (2007), Rainfall stochastic disaggregation models: calibration and validation of a multiplicative cascade model, *Advances in Water Resources*, 30, 1301–1319.

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

Grygier, J. C., and J. R. Stedinger (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 E. C. Waymire (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.

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

Katz, R. W., and M. B. Parlange (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., C. Onof, and H. S. Wheater (2003) Multivariate rainfall disaggregation at a fine timescale, *Water Resources Research*, 39(7), 1173, doi:10.1029/2002WR001600.

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

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

Koutsoyiannis, D., and A. Montanari (2015), Negligent killing of scientific concepts: the stationarity case, *Hydrological Sciences Journal*, 60(7-8), 1174–1183.

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

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

Li, C., V. P. Singh, and A. K. Mishra (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., V. Montesarchio, F. Napolitano, F. Russo, and E. Volpi (2009), Operational applications of radar rainfall data in urban hydrology, *IAHS-AISH Publication*, 327, 258–266.

Lombardo, F., E. Volpi, and D. Koutsoyiannis (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., E. Volpi, D. Koutsoyiannis, and S. M. Papalexiou (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 D. Schertzer (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 D. Schertzer (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.

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

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

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

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

McCabe, M. F., M. Rodell, D. E. Alsdorf, D. G. Miralles, R. Uijlenhoet, W. Wagner, A. Lucieer, R. Houborg, N. E. C. Verhoest, T. E. Franz, J. Shi, H. Gao, and E. F. Wood (2017), The Future of Earth Observation in Hydrology, *Hydrology and Earth System Sciences Discussion*, doi:10.5194/hess-2017-54.

Montanari, A., and D. Koutsoyiannis (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 V. K. Gupta (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., and V. K. Gupta (1996), A space-time theory of mesoscale rainfall using random cascades, *Journal of Geophysical Research*, 101(D21), 26319–26331.

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

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

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

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

Schleiss, M., J. Jaffrain, and A. Berne (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 D. Marsan (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 C. G. Kilsby (2015), Stationarity is undead: uncertainty dominates the distribution of extremes, *Advances in Water Resources*, 77, 17–36.

Serinaldi, F. and F. Lombardo (2017), A general simulation algorithm for autocorrelated binary processes, *Physical Review E*, 95(2), 023312.

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

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

Tabari, H., R. De Troch, O. Giot, R. Hamdi, P. Termonia, S. Saeed, E. Brisson, N. Van Lipzig, and P. Willems (2016), Local impact analysis of climate change on precipitation extremes: are high-resolution climate models needed for realistic simulations?, *Hydrology and Earth System Sciences*, 20, 3843–3857.

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

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

Valencia, D., and J. C. Schaake (1973), Disaggregation processes in stochastic hydrology, *Water Resources Research*, 9(3), 580–585, doi:10.1029/WR009i003p00580.

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

**Figure captions**

Figure 1. Sketch of a dyadic (b = 2) multiplicative random cascade.

Figure 2. Sketch of the disaggregation approach for generation of the auxiliary Gaussian process. Grey boxes indicate random variables whose values have been already generated prior to the current step. Arrows indicate the links to those of the generated variables that are considered in the current generation step [adapted from Koutsoyiannis, 2002].

Figure 3. Illustrative sketch for simulation of the auxiliary process $\tilde{Z}_{j,m}$. 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.

Figure 4. Scatter plot of the calculated sum of disaggregated variables $X_{j,k}$ (see eq. (35)) vs. the corresponding values of the higher-level variables $X_{1,0}$, generated with model parameters $k = 10$, $\mu_0 = 1024$, $\sigma_0 = 362.04$, $H = 0.85$, $p_{0,k} = 0.2$ and $\rho_{I,k}(1) = 0.7$ for all 10000 Monte Carlo experiments. "Empirical" and "adjusted" stand for original synthetic series and modified ones according to eq. (36), respectively.

Figure 5. Ensemble mean, standard deviation and autocorrelogram (from left to right, respectively) of the example disaggregation process $X_{j,k}$ as a function of the time step $j$ and lag $t$. Same simulations as in Fig. 4. Note the clear consistency between summary statistics of the original process $X_{j,k}$ and those of the adjusted process $X'_{j,k}$. The theoretical values of the statistics are given respectively by eq. (19) for the mean, the square root of eq. (20) for the standard deviation, and eq. (34) for the ACF of Markovian occurrences.

Figure 6. Theoretical and empirical autocorrelograms of the entire rainfall process, $X_{j,k}$, for three values of probability dry, i.e. $p_{0,k} \in \{0.2, 0.5, 0.8\}$, in case of purely random occurrences. The theoretical ACF of the process $X_{j,k}$ is derived from eq. (32) for random occurrences. Note that the ACF for $p_{0,k} = 0$ equals that of the rainfall depth process, $Z_{j,k}$.

Figure 7. Theoretical and empirical autocorrelograms of the entire rainfall process for three values of probability dry, i.e. $p_{0,k} \in \{0.2, 0.5, 0.8\}$, in case of Markovian occurrences. The theoretical ACF of the process $X_{j,k}$ is derived from eq. (34) for Markovian occurrences. The autocorrelation function for $Z_{j,k}$ (i.e., $p_{0,k} = 0$) is used as a benchmark to compare the Figs. 6 and 7 together in order to investigate the influence of each occurrence model on the dependence structure of the entire process, $X_{j,k}$.

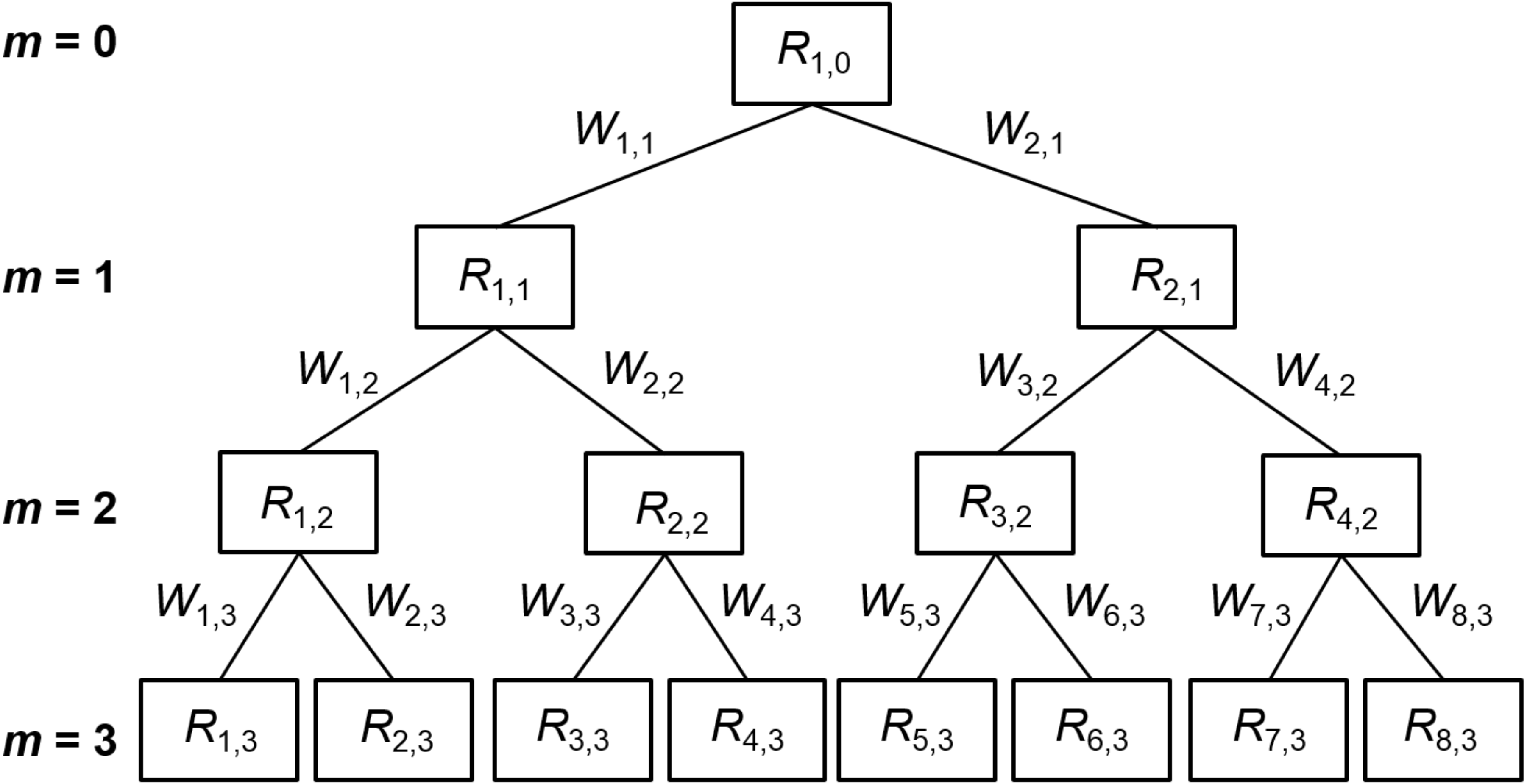
Figure 8. Comparison between the simulated (average, 1st and 99th percentiles) and empirical autocorrelograms for the data series recorded at Viterbo raingauge station in January 1999. In the left and right panels, we show respectively the ACF of the occurrence (binary) process $\rho_{I,k}(t)$ and that of the intermittent (mixed) process $\rho_{X,k}(t)$. Estimated model parameters are: $\mu_0 = 736.3, \sigma_0 = 320.2, p_{0,k} = 0.96, H = 0.83$.

Figure 9. Same as Fig. 8 for the data series recorded at Viterbo raingauge station in April 2003. Estimated model parameters are: $\mu_0 = 626.7, \sigma_0 = 83.8, p_{0,k} = 0.95, H = 0.7$.

Figure 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).

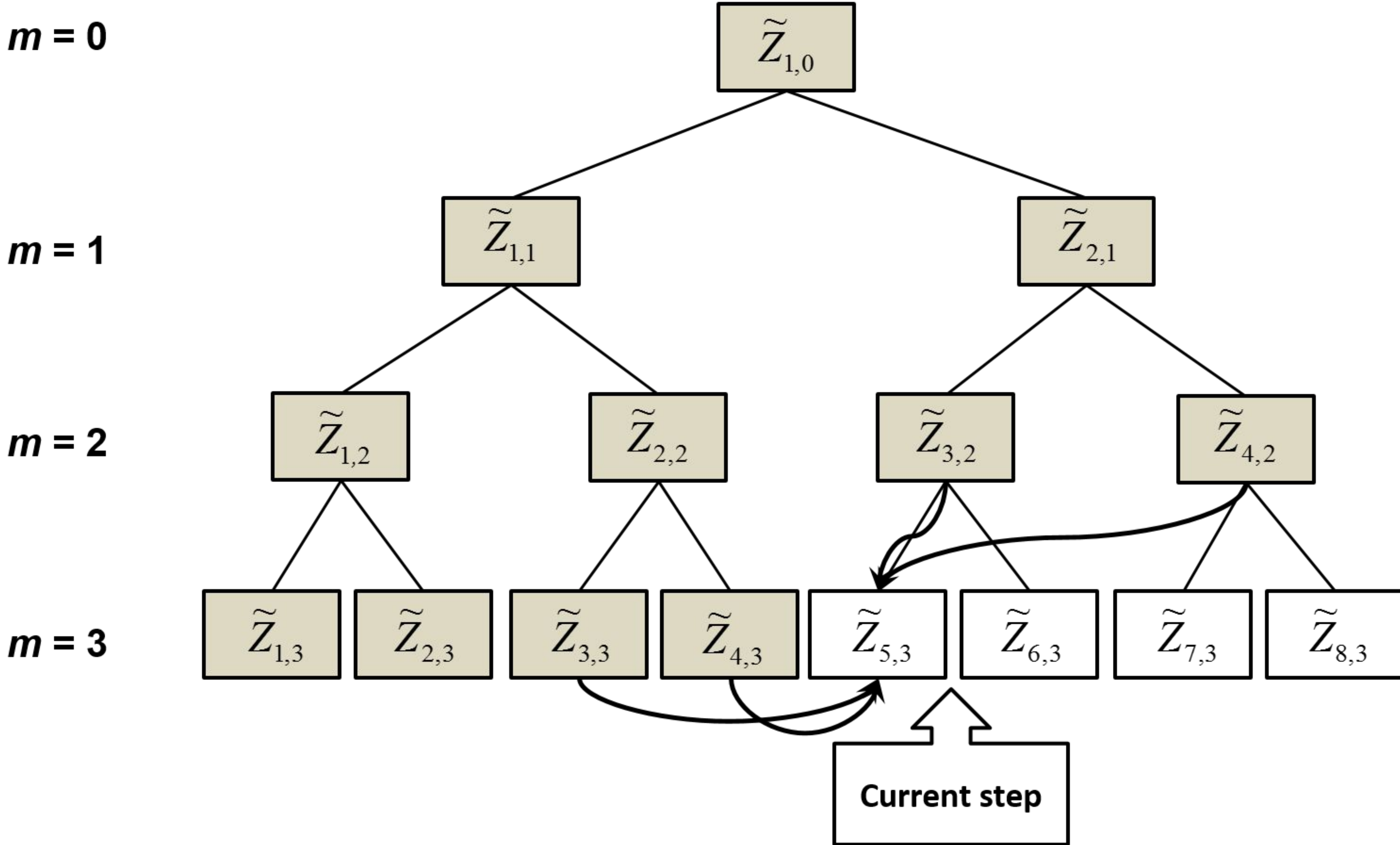
Figure 11. Hyetograph of the rainfall data recorded at Viterbo raingauge station in April 2003 (left panel) along with the synthetic time series of equal length generated by our model (right panel).

**Figure 1.**




$$R_{3,3} = W_{3,3}\, W_{2,2}\, W_{1,1}\, R_{1,0}$$

**Figure 2.**



m = 0
m = 1
m = 2
m = 3
$\tilde{Z}_{1,0}$
$\tilde{Z}_{1,1}$
$\tilde{Z}_{2,1}$
$\tilde{Z}_{1,2}$
$\tilde{Z}_{2,2}$
$\tilde{Z}_{3,2}$
$\tilde{Z}_{4,2}$
$\tilde{Z}_{1,3}$
$\tilde{Z}_{2,3}$
$\tilde{Z}_{3,3}$
$\tilde{Z}_{4,3}$
$\tilde{Z}_{5,3}$
$\tilde{Z}_{6,3}$
$\tilde{Z}_{7,3}$
$\tilde{Z}_{8,3}$
Current step

**Figure 3.**

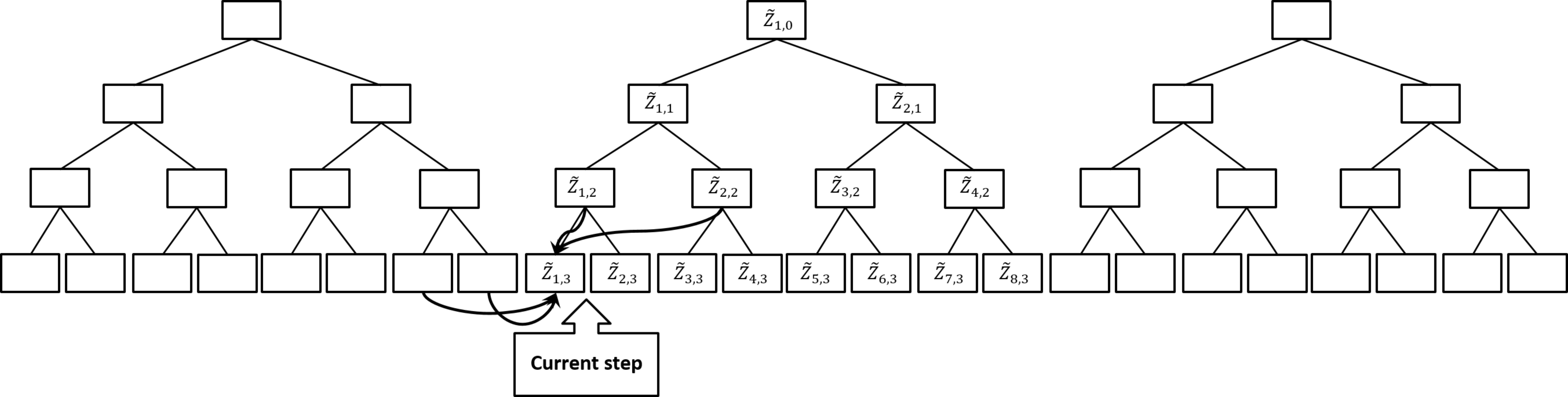
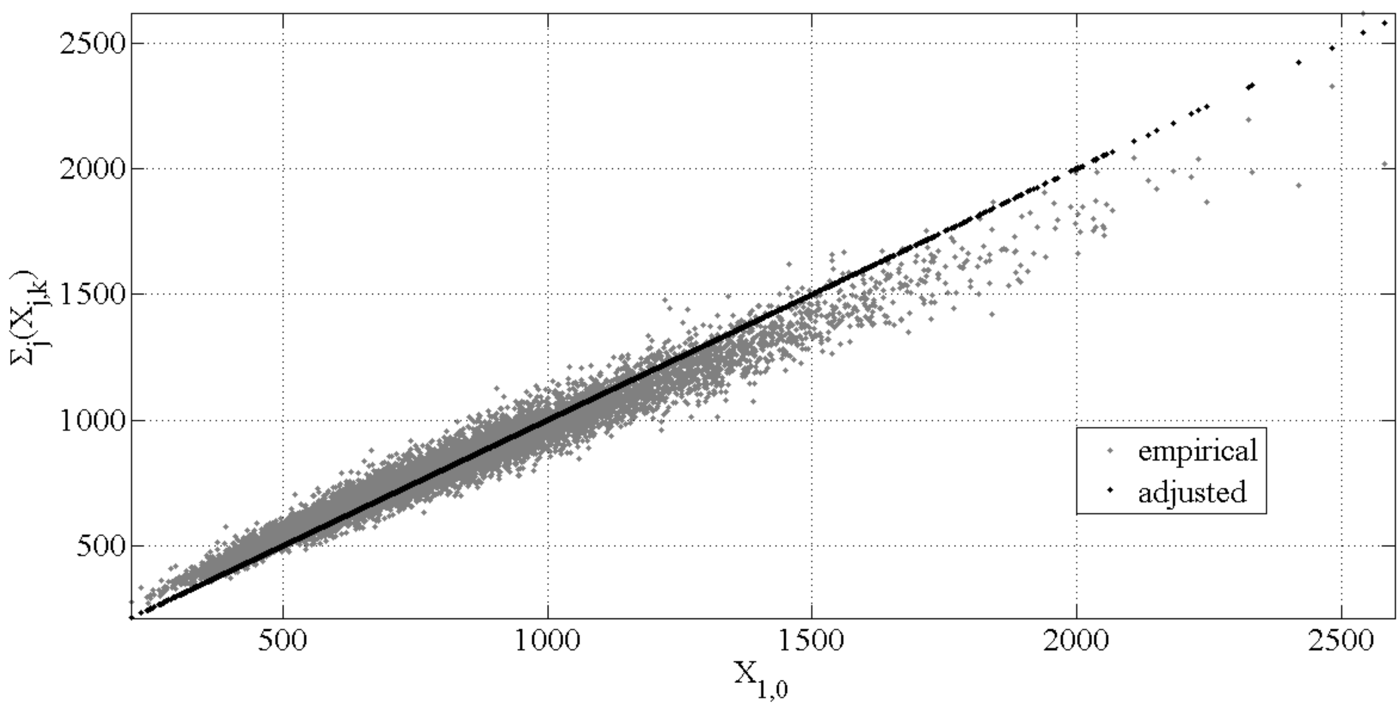
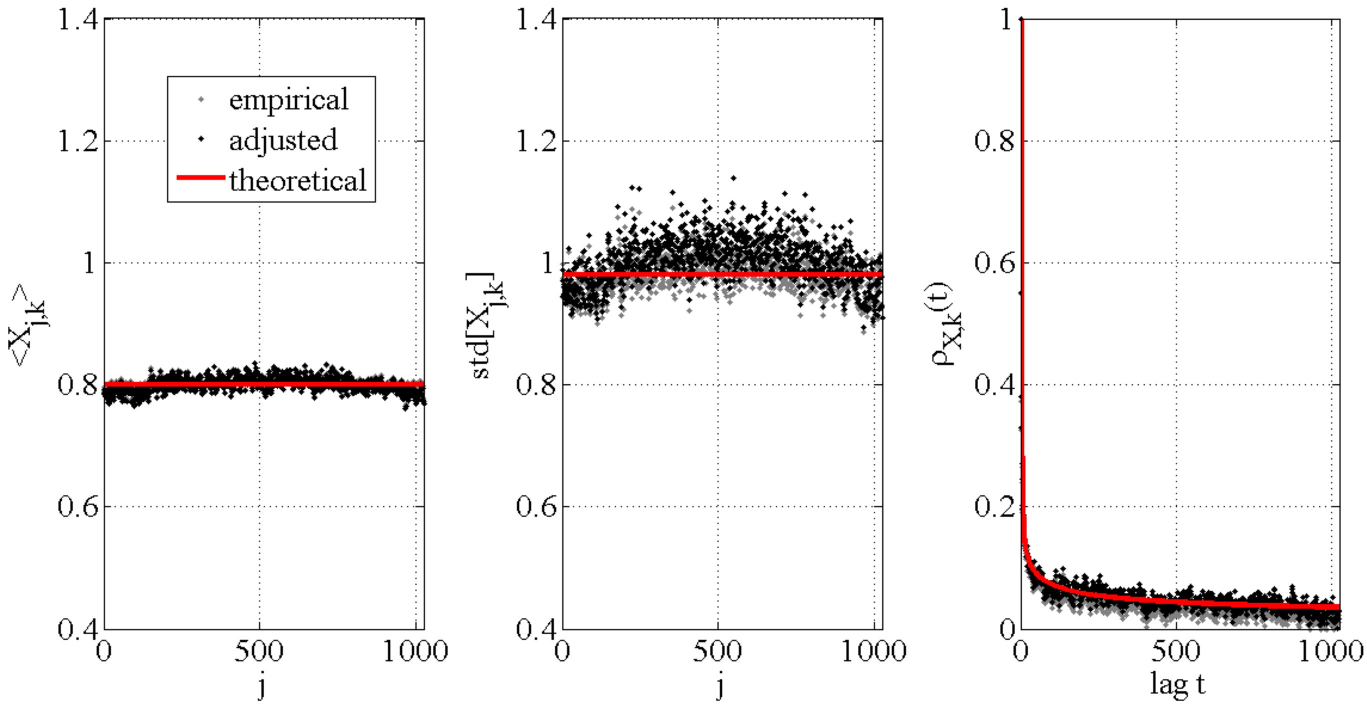

$\tilde{Z}_{1,0}$
$\tilde{Z}_{1,1}$
$\tilde{Z}_{2,1}$
$\tilde{Z}_{1,2}$
$\tilde{Z}_{2,2}$
$\tilde{Z}_{3,2}$
$\tilde{Z}_{4,2}$
$\tilde{Z}_{1,3}$
$\tilde{Z}_{2,3}$
$\tilde{Z}_{3,3}$
$\tilde{Z}_{4,3}$
$\tilde{Z}_{5,3}$
$\tilde{Z}_{6,3}$
$\tilde{Z}_{7,3}$
$\tilde{Z}_{8,3}$
Current step

**Figure 4.**



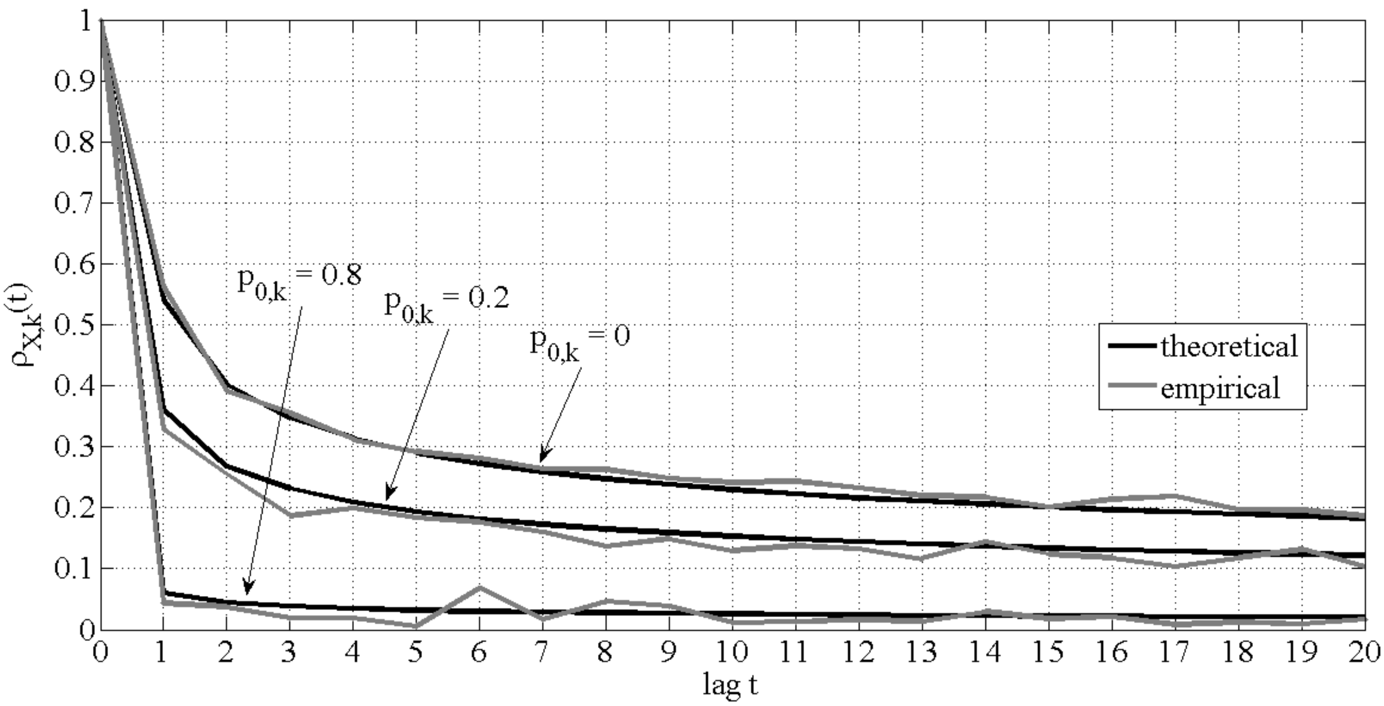
$\Sigma_j(X_{j,k})$
2500
2000
1500
1000
500
500
1000
1500
2000
2500
$X_{1,0}$
empirical
adjusted

**Figure 5.**



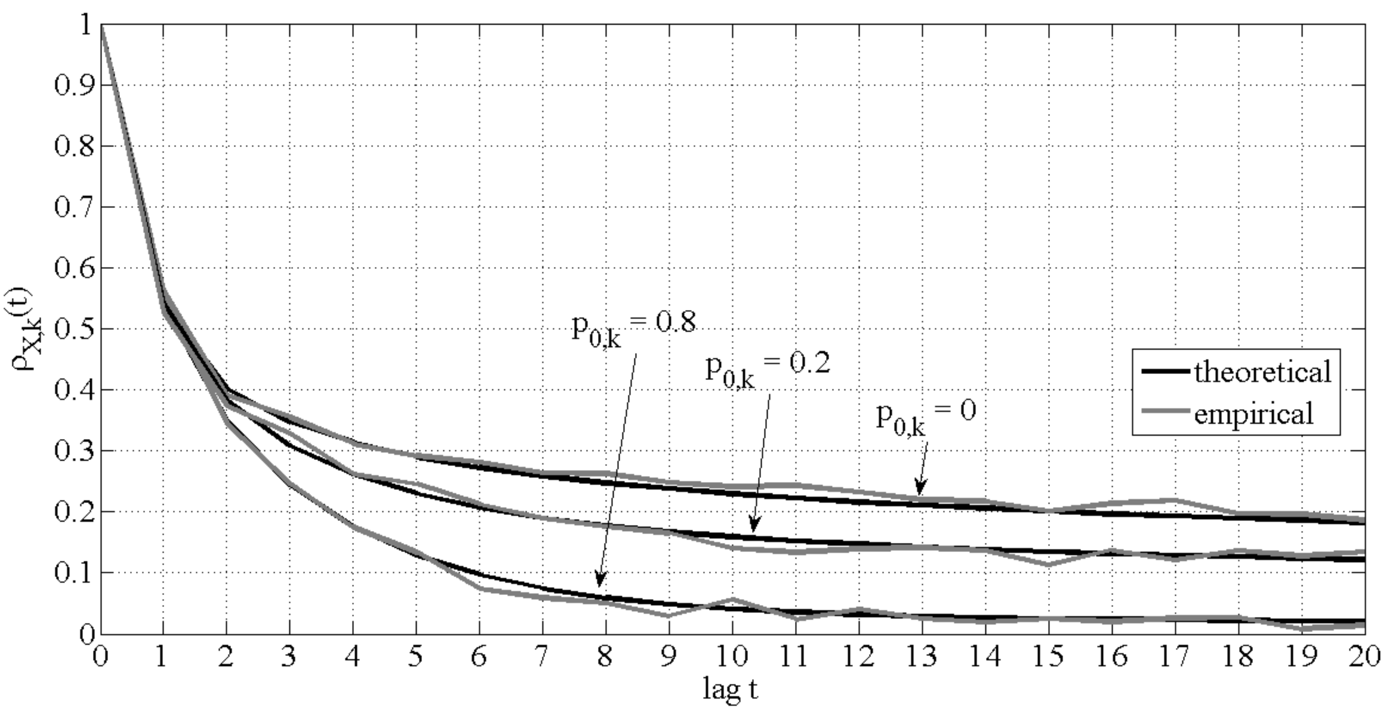
1.4
1.2
1
0.8
0.6
0.4
0
500
1000
j
<X_{j,k}>
empirical
adjusted
theoretical
std[X_{j,k}]
ρ_{X,k}(t)
0.2
lag t

**Figure 6.**



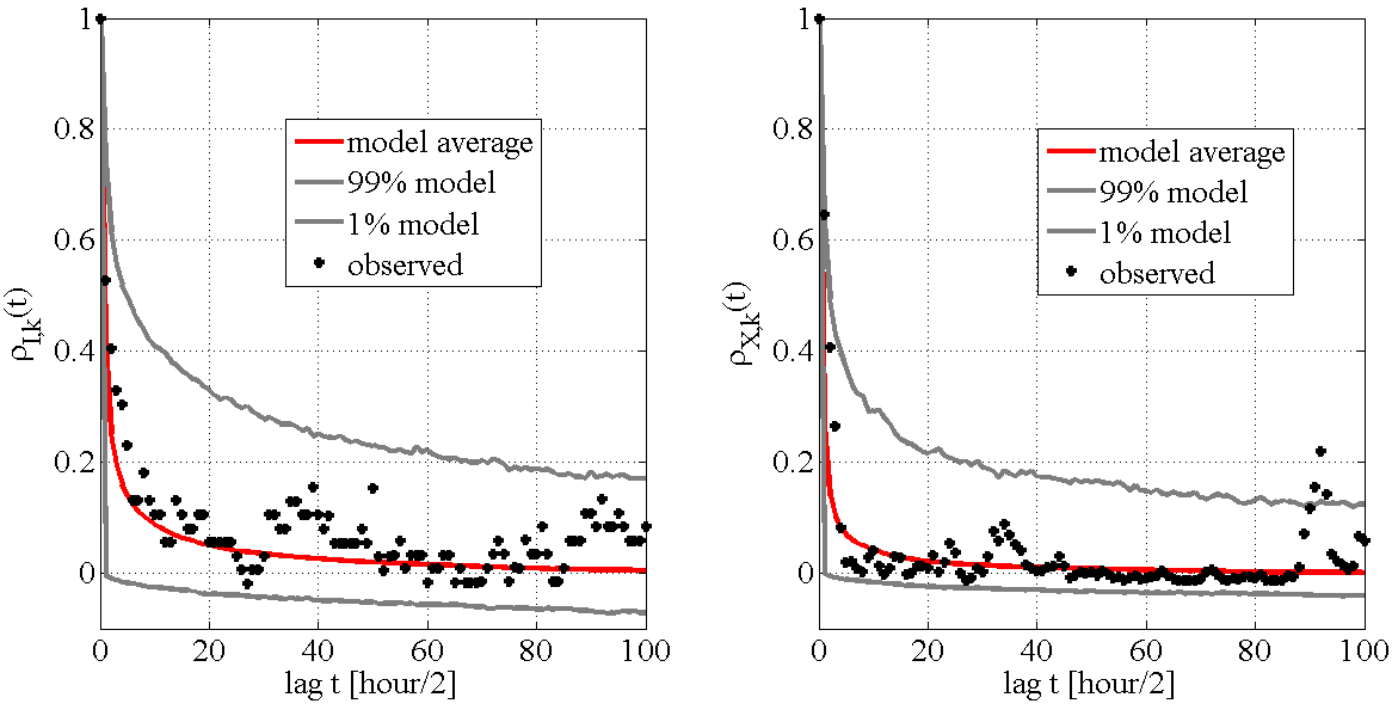
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
lag t
$\rho_{X,k}(t)$
$p_{0,k}$ = 0.8
$p_{0,k}$ = 0.2
$p_{0,k}$ = 0
theoretical
empirical

**Figure 7.**



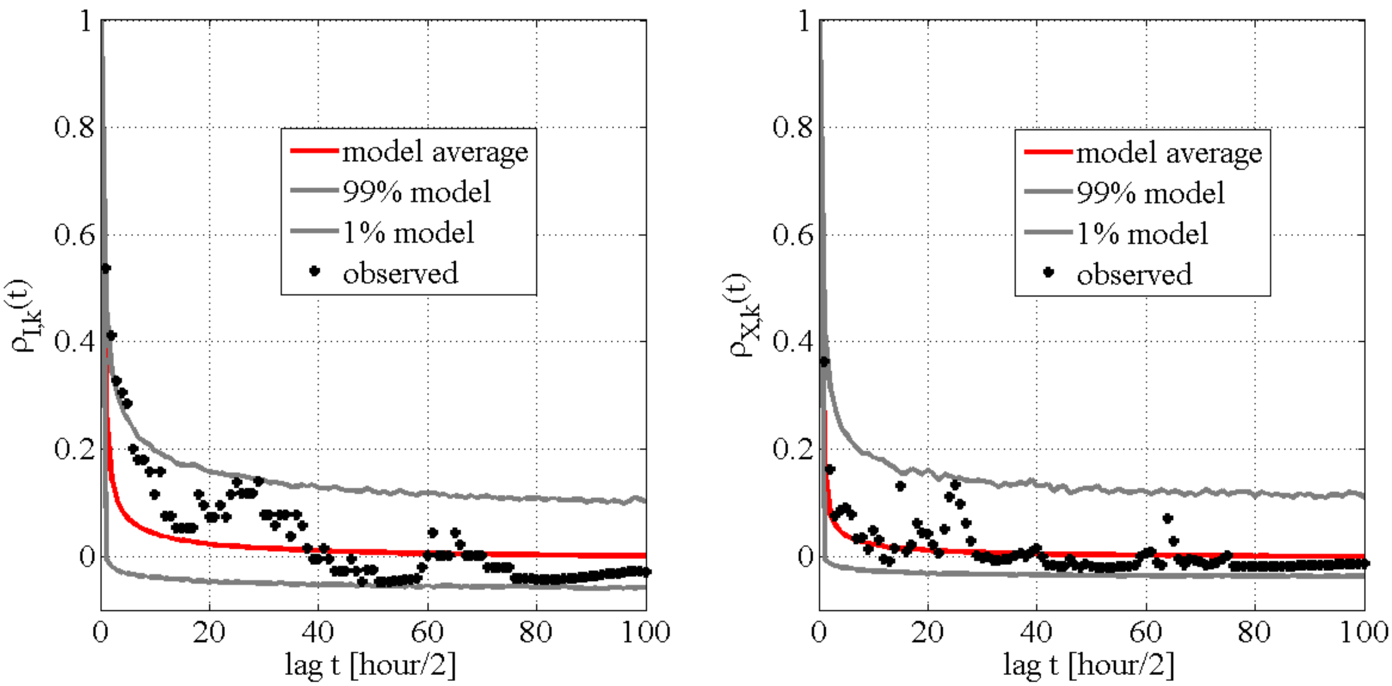
$\rho_{X,k}(t)$
lag t
$p_{0,k} = 0.8$
$p_{0,k} = 0.2$
$p_{0,k} = 0$
theoretical
empirical

**Figure 8.**



model average
99% model
1% model
observed
$\rho_{I,k}(t)$
lag t [hour/2]
model average
99% model
1% model
observed
$\rho_{X,k}(t)$
lag t [hour/2]

**Figure 9.**



model average
99% model
1% model
observed
$\rho_{I,k}(t)$
$\rho_{X,k}(t)$
lag t [hour/2]
lag t [hour/2]
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100

**Figure 10.**

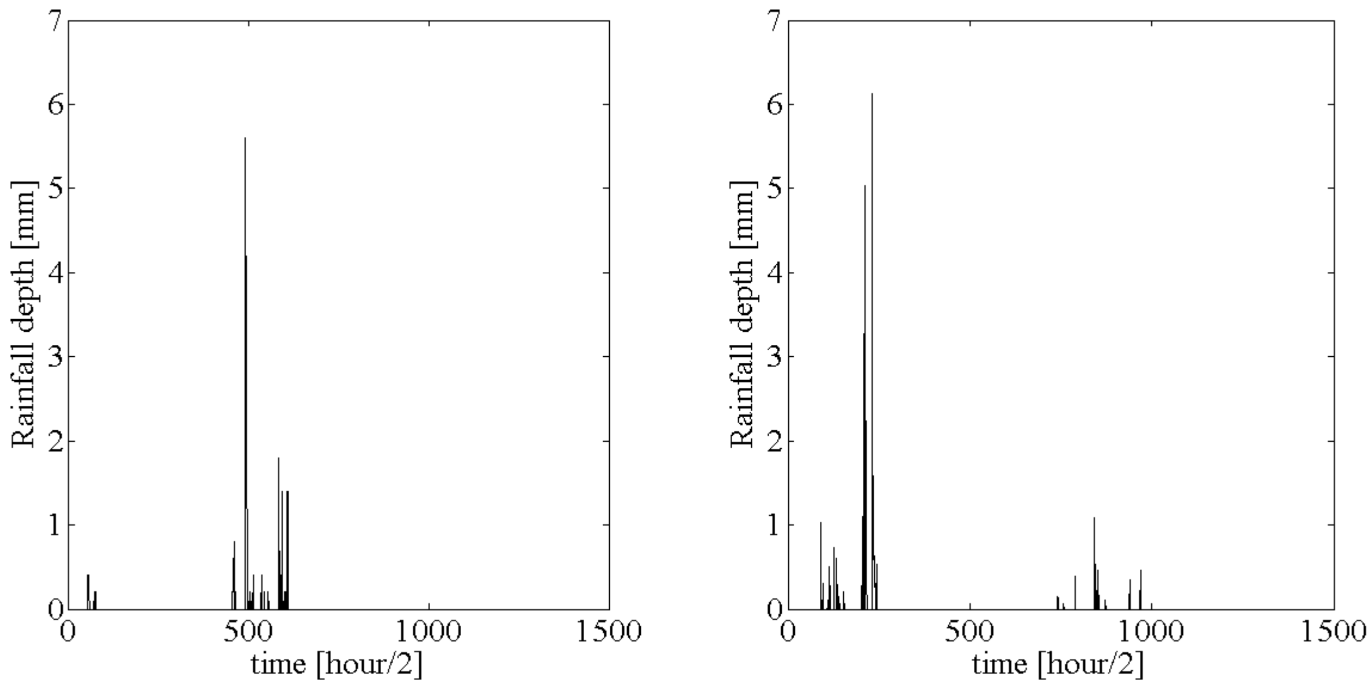
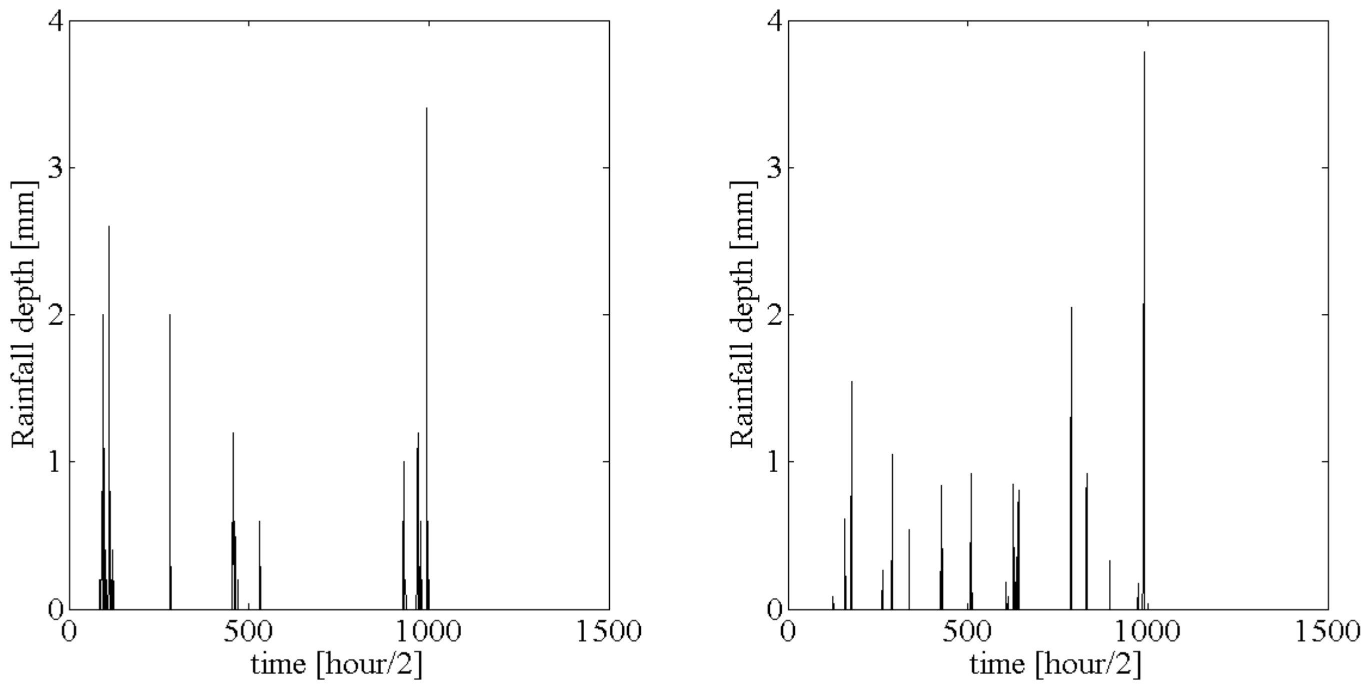

Rainfall depth [mm]
time [hour/2]
Rainfall depth [mm]
time [hour/2]

**Figure 11.**


Rainfall depth [mm]
time [hour/2]
Rainfall depth [mm]
time [hour/2]